



**Mickael Santos  
Creolo**

**Conceção e Projeto de Data Warehouse: Olhar o  
Futuro do Ensino Básico e Secundário em Portugal**





**Mickael Santos  
Creolo**

**Conceção e Projeto de Data Warehouse: Olhar o  
Futuro do Ensino Básico e Secundário em Portugal**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Sistemas de Informação, realizada sob a orientação científica do Prof. Doutor Óscar Mortágua Pereira, Professor Auxiliar do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro e da Prof. Doutora Maribel Yasmina Santos, Professora Associada com Agregação da Universidade do Minho.



## **o júri**

presidente

**Prof. Doutor José Manuel Matos Moreira**  
Professor Auxiliar do Departamento de Eletrónica, Telecomunicações e Informática da  
Universidade de Aveiro

vogais

**Prof. Doutor Marco Paulo Amorim Vieira**  
Professor Auxiliar da Faculdade de Ciências e Tecnologia da Universidade de Coimbra

**Prof. Doutor Óscar Mortágua Pereira**  
Professor Auxiliar do Departamento de Eletrónica, Telecomunicações e Informática da  
Universidade de Aveiro



## **agradecimentos**

Quero agradecer ao Prof. Doutor Óscar Mortágua Pereira e à Prof. Doutora Maribel Yasmina Santos que foram os meus orientadores no trabalho. Agradeço também à minha família e amigos pelo apoio que me têm dado.





**palavras-chave**

Data Warehouse, Data Mart, ETL, Modelo Multidimensional, Escolas, Ranking, Avaliação.

**resumo**

O ensino básico e secundário são duas etapas fundamentais na formação académica da população portuguesa. Durante essas etapas o desempenho dos alunos é analisado para verificar como evolui o ensino em Portugal. O desempenho dos alunos é analisado tendo em conta as notas e as percentagens de aprovações. As análises feitas são básicas uma vez que não são analisadas as componentes que influenciam o desempenho escolar dos alunos. Neste sentido o objetivo do trabalho é o desenvolvimento de um projeto conceptual de um sistema de Data Warehousing. No projeto é definido um modelo multidimensional que permite analisar as notas ou percentagens de aprovações considerando diversas componentes tais como escola, disciplina, características dos alunos e contexto familiar. O projeto especifica e caracteriza os processos de extração, transformação e carregamento com base nas fontes de dados. Obtive a colaboração de duas escolas que forneceram ficheiros de dados, por isso foram consideradas duas fontes de dados, uma para cada escola. O projeto especifica também a forma como o Data Warehouse é preenchido com dados. Não foi feita nenhuma implementação do sistema de Data Warehouse uma vez que as escolas apenas forneceram ficheiros exemplo. O projeto foi planeado para uma empresa de consultadoria privada que fornece prestação de serviços para as escolas aderentes.



**keywords**

Data Warehouse, Data Mart, ETL, Multidimensional Model, Schools, Ranking, Evaluation.

**abstract**

Basic and secondary education are two vital stages in the academic education of the Portuguese population. During these steps the academic performance of students is analyzed to verify how the education has evolved in Portugal. The students' academic performance is analyzed based on the grades and 'Pass' grades percentages. These analysis were considered basic because the components that influence the academic performance of students were not analyzed. Therefore, this work's aim is to develop a conceptual project of a Data Warehousing system. A multidimensional model is defined in the project and it allows the analysis of the students' grades or 'Pass' grades percentages taking into account several components such as the school, the subjects, the student characteristics and the family context. The project specifies and characterizes the processes of extraction, transformation and load based on the data sources. I obtained the collaboration of two schools that provided the data files, so two data sources were considered, one for each school. The project also specifies how the Data Warehouse is populated with data. The Data Warehouse system was not implemented because the schools only provided example files. The project was planned for a private consulting company that provides services for the associated schools.



## Índice

Lista de Figuras.....	v
Lista de Tabelas .....	vii
1. Introdução .....	1
1.1. Objetivos .....	1
1.2. Organização do Documento.....	2
2. Enquadramento Conceptual e Tecnológico .....	3
2.1. Situação Atual .....	3
2.1.1. Tratamento dos Dados.....	3
2.1.2. Aplicações utilizadas .....	8
2.2. Data Warehouse.....	9
2.2.1. Conceitos utilizados em Data Warehouse .....	9
2.2.2. Operações para analisar um cubo.....	11
2.2.3. Representações Relacionais.....	12
2.3. Proposta de Solução.....	15
2.4. Trabalhos realizados na área do Ensino.....	17
3. Caracterização e Planeamento.....	21
3.1. Análise e Caracterização dos Recursos Disponíveis .....	21
3.1.1. Recursos Operacionais.....	21
3.1.2. Recursos Materiais.....	22
3.1.3. Recursos Humanos.....	22
3.2. Estudo e Caracterização dos Resultados Pretendidos .....	22
3.2.1. Ranking de Escolas .....	22
3.2.2. Efeito do contexto familiar no desempenho escolar dos alunos.....	23
3.2.3. Efeito do escalão social no desempenho escolar dos alunos .....	23
3.2.4. Comparar o desempenho escolar dos alunos por sexo .....	23
3.2.5. Analisar o desempenho escolar dos alunos com necessidades educativas especiais .....	24
3.2.6. Relacionar a idade dos alunos com o ano escolar .....	24
3.3. Especificação dos Requisitos do Sistema .....	24
3.3.1. Requisitos Funcionais.....	24
3.3.2. Requisitos Operacionais.....	26
3.3.3. Requisitos Temporais.....	27

3.4.	Planeamento e Gestão de Projeto .....	27
3.5.	Aprovação do Sistema e Validação do Projeto .....	29
3.5.1.	Validade do Sistema Proposto .....	29
3.5.2.	Validade do Projeto de Desenho e Conceptualização .....	30
3.6.	Foco do Problema .....	30
4.	Modelo Multidimensional.....	31
4.1.	Matriz .....	31
4.2.	Data Mart .....	32
4.3.	Granularidade.....	32
4.4.	Dimensões.....	32
4.4.1.	Dimensão <i>Ano Letivo</i> .....	33
4.4.2.	Dimensão <i>Ano Escolar</i> .....	34
4.4.3.	Dimensão <i>Escola</i> .....	35
4.4.4.	Dimensão <i>Características do Aluno</i> .....	37
4.4.5.	Dimensão <i>Contexto Familiar</i> .....	38
4.4.6.	Dimensão <i>Estado de Avaliação</i> .....	40
4.4.7.	Dimensão <i>Disciplina</i> .....	40
4.5.	Tabelas de Fatos.....	41
4.5.1.	Tabela de Fatos <i>Matricula</i> .....	41
4.5.2.	Tabela de Fatos <i>Notas do 5º Ano ao 9º Ano</i> .....	42
4.5.3.	Tabela de Fatos <i>Notas do 10º Ano ao 12º Ano</i> .....	44
4.6.	Revisão e Aprovação .....	45
4.7.	Modelo Logico do Data Mart .....	46
4.8.	Análise das Fontes de Dados.....	47
4.8.1.	Disponibilidade da Informação .....	49
4.8.2.	Qualidade da Informação das Fontes .....	52
4.8.3.	Mapeamento dos dados das fontes para o Data Mart .....	52
4.9.	Revisão e Aprovação do Modelo desenvolvido .....	57
4.10.	Estratégia de Extração, Transformação e Carregamento dos Dados.....	57
4.10.1.	Extração.....	57
4.10.2.	Transformação .....	57
4.10.3.	Carregamento .....	58
4.11.	Estimativa de Dimensionamento .....	58

4.12.	Revisão e Aprovação final da análise multidimensional .....	61
5.	Estratégia de Implementação .....	63
5.1.	Bases de Dados.....	63
5.2.	Manipulação dos dados na base de dados SA .....	64
5.2.1.	Processos de Extração e Transformação dos dados .....	64
5.2.2.	Processos de Carregamento dos dados na base de dados <i>SA_DataMart</i> .....	78
5.3.	Estratégia de Carregamento do Data Warehouse .....	90
5.3.1.	Processos de Extração e Transformação dos dados das fontes.....	91
5.3.2.	Processos de Carregamento do Data Mart .....	91
6.	Discussão e Conclusão.....	93
6.1.	Discussão .....	93
6.2.	Conclusão .....	93
	Referências.....	95
Anexo I.	Fonte 1.....	97
	Fonte 1 – Ficheiro Excel da categoria <i>F1_Alunos</i> com a lista de alunos num ano letivo .....	97
	Fonte 1 – Ficheiro Excel da categoria <i>F1_Pais</i> com as habilitações e profissões dos pais num ano letivo.....	99
	Fonte 1 – Ficheiro Excel da categoria <i>F1_Escalões</i> com os alunos com um determinado escalão social.....	99
	Fonte 1 – Ficheiro Excel da categoria <i>F1_Turmas</i> para uma turma do ensino básico.....	101
	Fonte 1 – Ficheiro Excel da categoria <i>F1_Turmas</i> para uma turma do ensino secundário .....	103
Anexo II.	Fonte 2.....	105
	Fonte 2 – Ficheiro Excel da categoria <i>F2_Alunos</i> com a lista de alunos de uma turma .....	105
	Fonte 2 – Ficheiro Excel da categoria <i>F2_Escalões</i> com os alunos com um determinado escalão social.....	107
	Fonte 2 – Ficheiro Excel da categoria <i>F2_Turmas</i> para uma turma do ensino básico.....	109
	Fonte 2 – Ficheiro Excel da categoria <i>F2_Turmas</i> para uma turma do ensino secundário .....	111
Anexo III.	Base de dados <i>Escolas</i> – Modelo Lógico .....	113





## Lista de Figuras

Figura 1 – Cubo [16] .....	10
Figura 2 - Star Schema [16] .....	12
Figura 3 - Snowflake Schema [16] .....	14
Figura 4 – Notas no Exame de Matemática por Escolas do Agrupamento .....	16
Figura 5 – Número de pais por categoria de habilitação .....	16
Figura 6 – Arquitetura do Data Warehouse .....	27
Figura 7 – Diagrama da dimensão Ano Letivo.....	34
Figura 8 – Diagrama da dimensão <i>Ano Escolar</i> .....	35
Figura 9 - Diagrama da dimensão <i>Escola</i> .....	37
Figura 10 - Diagrama da dimensão <i>Características do Aluno</i> .....	38
Figura 11 - Diagrama da dimensão <i>Contexto Familiar</i> .....	39
Figura 12 - Diagrama da Dimensão <i>Estado de Avaliação</i> .....	40
Figura 13 - Diagrama da dimensão <i>Disciplina</i> .....	41
Figura 14 - Diagrama da Tabela de Fatos <i>Matrícula</i> .....	42
Figura 15 – Diagrama da Tabela de Fatos <i>Notas do 5º Ano ao 9º Ano</i> .....	44
Figura 16 – Diagrama da Tabela de Fatos <i>Notas do 10º Ano ao 12º Ano</i> .....	45
Figura 17 - Estrela <i>Matrícula</i> .....	46
Figura 18 - Estrela <i>Notas do 5º Ano ao 9º Ano</i> .....	46
Figura 19 - Estrela <i>Notas do 10º Ano ao 12º Ano</i> .....	47
Figura 20 – Diagrama com as bases de dados criadas .....	63
Figura 21 - Tabela para a inserção de dados dos alunos e dos pais.....	65
Figura 22 - Tabela para a inserção de dados dos Escalões dos Alunos .....	69
Figura 23 – Tabelas para a inserção de Disciplinas, Notas e Estado de Avaliação dos Alunos .....	70
Figura 24 – Tabelas para a integração das Características dos Alunos .....	76
Figura 25 – Query para fazer joins das tabelas <i>Aluno</i> e <i>Escalão</i> .....	77
Figura 26 - Tabelas para a integração dos dados dos Pais .....	78
Figura 27 - Query para inserir dados na tabela <i>PaisCorrigido</i> .....	78
Figura 28 – Tabela para o carregamento da <i>SAANL_AnoLetivo</i> .....	79
Figura 29 - Tabela para o carregamento da <i>SAANE_AnoEscolar</i> .....	80
Figura 30 - Tabela para o carregamento da <i>SAESC_Escola</i> .....	80
Figura 31 - Tabelas para o carregamento da <i>SACAL_CaracteristicasAluno</i> .....	82

Figura 32 - Tabelas para o carregamento da <i>SACFA_ContextoFamiliar</i> .....	83
Figura 33 - Tabelas para o carregamento da <i>SAESA_EstadoAvaliacao</i> .....	84
Figura 34 - Tabelas para o carregamento da <i>SADIS_Disciplina</i> .....	84
Figura 35 - Tabelas para o carregamento da <i>SAMAT_Matricula</i> .....	85
Figura 36 - Query para obter as Alternate Key de cada dimensão da <i>SAMAT_Matricula</i> .....	86
Figura 37 - Tabelas para o carregamento da <i>SANT1_Nota5_9</i> .....	87
Figura 38 - Query para obter as Alternate Key de cada dimensão da <i>SANT1_Nota5_9</i> .....	87
Figura 39 – Tabelas para o carregamento da <i>SANT2_Nota10_12</i> .....	89
Figura 40 – Query para obter as Alternate Key de cada dimensão da <i>SANT2_Nota10_12</i> .....	89
Figura 41 - Diagrama de sequência para o carregamento do Data Warehouse.....	90
Figura 42 – Ficheiro Excel da <i>Fonte 1</i> com a lista de alunos no ano letivo 2013/2014.....	97
Figura 43 - Ficheiro Excel da <i>Fonte 1</i> com as habilitações e profissões dos pais no ano letivo 2013/2014.....	99
Figura 44 – Ficheiro Excel da <i>Fonte 1</i> com os alunos que têm escalão A no 12º ano do ano letivo 2013/2014.....	99
Figura 45 – Ficheiro Excel da <i>Fonte 1</i> do 9º A no ano letivo 2013/2014.....	101
Figura 46 – Ficheiro Excel da <i>Fonte 1</i> do 11º A no ano letivo 2013/2014.....	103
Figura 47 – Ficheiro Excel da <i>Fonte 2</i> com a lista de alunos do 10º A no ano letivo 2013/2014...	105
Figura 48 – Ficheiro Excel da <i>Fonte 2</i> com os alunos que têm escalão A no 10º ano do ano letivo 2013/2014.....	107
Figura 49 – Ficheiro Excel da <i>Fonte 2</i> do 6º A no ano letivo 2013/2014.....	109
Figura 50 – Ficheiro Excel da <i>Fonte 2</i> do 12º A no ano letivo 2013/2014.....	111
Figura 51 - Modelo lógico da base de dados <i>Escola</i> .....	113

## Lista de Tabelas

Tabela 1 - Matriz .....	31
Tabela 2 - Dimensão Ano Letivo.....	33
Tabela 3 – Dimensão <i>Ano Escolar</i> .....	34
Tabela 4 - Dimensão <i>Escola</i> .....	35
Tabela 5 - Dimensão Características do Aluno.....	37
Tabela 6 - Dimensão <i>Contexto Familiar</i> .....	39
Tabela 7 - Dimensão <i>Estado de Avaliação</i> .....	40
Tabela 8 - Dimensão <i>Disciplina</i> .....	41
Tabela 9 - Tabela de Fatos <i>Matrícula</i> .....	42
Tabela 10 - Tabela de Fatos de <i>Notas do 5º Ano ao 9º Ano</i> .....	43
Tabela 11 - Tabela de Fatos de <i>Notas do 10º Ano ao 12º Ano</i> .....	44
Tabela 12 - Mapeamento Logico das Fontes para o Data Mart.....	53
Tabela 13 - Estimativa do número de alunos no ensino básico e secundário em Portugal.....	58
Tabela 14 - Estimativa do número de disciplinas.....	59
Tabela 15 - Número de registos por aluno para cada tabela de fatos.....	59
Tabela 16 – Número de registos para cada tabela de dimensão.....	60
Tabela 17 – Origem dos dados para a tabela <i>Aluno</i> .....	66
Tabela 18 – Origem dos dados para a tabela <i>Pais</i> .....	68
Tabela 19 – Origem dos dados para a tabela <i>Escalao</i> .....	69
Tabela 20 – Origem dos dados para a tabela <i>Nota5_9</i> .....	72
Tabela 21 – Origem dos dados para a tabela <i>Nota10_12</i> .....	73
Tabela 22 – Origem dos dados para a tabela <i>EstadoAvaliacao</i> .....	75



## 1. Introdução

O ensino básico e secundário são duas etapas fundamentais na formação académica da população portuguesa. Durante essas etapas o desempenho dos alunos é analisado para verificar de que forma evolui o ensino em Portugal. O desempenho dos alunos em Portugal é analisado com base nas notas que os alunos têm às disciplinas. São também feitas contagens dos alunos que são aprovados num determinado ano escolar. O tratamento estatístico feito nas análises é básico não determinando as componentes que influenciaram esses resultados. Para melhorar o desempenho escolar dos alunos é importante observar quais são as componentes que influenciam esse desempenho. Para analisar essas componentes as escolas têm de fornecer dados que são suscetíveis de influenciar o desempenho escolar dos alunos.

Com esta perspetiva o objetivo principal do trabalho foi o desenvolvimento de um projeto conceptual de um sistema de Data Warehousing para uma empresa de consultoria privada. Foi definido o nome fictício Empresa, para identificar a empresa de consultoria privada. O desenvolvimento do projeto foi feito com base na metodologia de Kimball [1].

Os sistemas de Data Warehousing permitem fazer a integração de dados provenientes de diferentes fontes. Obtive a colaboração de duas escolas que forneceram ficheiros Excel para verificar o esquema dos ficheiros. Foram consideradas duas fontes de dados, uma para cada escola. Apesar dos ficheiros estarem no formato Excel em ambas as fontes, esses ficheiros têm um esquema diferente em cada fonte. A Empresa com base nas análises feitas no Data Warehouse poderá fazer prestação de serviços às escolas que pretenderem aderir.

### 1.1. Objetivos

O sistema de Data Warehousing vai permitir à Empresa analisar os dados das escolas de forma integrada, de maneira a encontrar as componentes que influenciem o desempenho escolar dos alunos. Com os resultados das análises a Empresa vai poder prestar serviço às escolas. O Data Warehouse tem como objetivo dar suporte à tomada de decisões para melhorar o desempenho dos alunos nas escolas. Neste trabalho o foco do problema está na análise do desempenho escolar dos alunos com base nas notas e percentagens de aprovações tendo em conta as seguintes componentes:

- Escola;
- Disciplina;

- Escalão social dos alunos;
- Características de cada aluno, nomeadamente o sexo e idade;
- Habilitações e profissões dos pais de cada aluno.

No projeto é definido um modelo multidimensional que permite analisar as notas ou percentagens de aprovações considerando as componentes descritas. O projeto especifica e caracteriza os processos de extração, transformação e carregamento com base nas fontes de dados. O projeto especifica também a forma como o Data Warehouse é preenchido com dados. Não foi feita nenhuma implementação do sistema de Data Warehouse uma vez que as escolas apenas forneceram ficheiros exemplo.

## 1.2. Organização do Documento

Para além da introdução, este documento apresenta mais quatro capítulos. O capítulo 2 tem o Enquadramento Conceptual e Tecnológico. O capítulo 3 tem a caracterização e planeamento do sistema de Data Warehousing. O capítulo 4 apresenta a modelação multidimensional do Data Warehouse. O capítulo 5 define a forma como é implementado o sistema de Data Warehousing. O capítulo 6 apresenta a discussão e as conclusões do trabalho.

## 2. Enquadramento Conceptual e Tecnológico

Este capítulo começa por apresentar a situação atual do ensino escolar em Portugal, nomeadamente a forma como é feita a gestão escolar nas escolas e as aplicações utilizadas para gerir a informação de uma escola. A seguir são apresentados alguns conceitos utilizados em Data Warehouse. Depois é apresentada uma proposta de solução. Por fim são apresentados trabalhos realizados na área do ensino.

### 2.1. Situação Atual

O sistema escolar está dividido no ensino básico e no ensino secundário. O ensino básico tem três ciclos. O 1º ciclo engloba os anos escolares do 1º ano ao 4º ano, o 2º ciclo engloba os anos escolares do 5º ano ao 6º ano, o terceiro ciclo engloba os anos escolares do 7º ano ao 9º ano. O secundário engloba os anos escolares do 10º ano ao 12º ano. No ensino básico, o plano curricular é comum em cada ano escolar. No ensino secundário existem vários cursos em que cada um tem o seu plano curricular. Os cursos estão organizados nas seguintes áreas [2]:

- Cursos científicos-humanísticos;
- Cursos artísticos especializados;
- Cursos profissionais;
- Cursos científicos-humanísticos do ensino recorrente.

Ao longo do percurso escolar os alunos têm exames nacionais. As disciplinas de língua portuguesa e matemática têm exames nacionais no 4º ano, 6º ano e 9º ano. No secundário existem exames nacionais no 11º ano e 12º ano de escolaridade. Os alunos fazem exames a algumas disciplinas dependendo do curso que escolheram. Os exames nacionais têm duas fases. Os alunos podem ir à segunda fase se reprovarem à disciplina na primeira fase ou se pretenderem fazer melhoria [3]. Existem também exames para alunos com estatuto autoproposto. Um aluno com estatuto autoproposto tem de realizar exames às disciplinas no qual não teve positiva, de maneira a acabar o ano escolar ou ciclo escolar. A nota a cada uma das disciplinas é obtida pela nota do exame respetivo [4].

#### 2.1.1. Tratamento dos Dados

No final de cada ano letivo podem ser feitos relatórios de avaliação interna pelas escolas, relatórios de avaliação externa pela Inspeção Geral da Educação e Ciência e relatórios de avaliação

de exames nacionais por um júri nacional de exames [5] [6]. As duas escolas do qual obtive colaboração forneceram relatórios de avaliação. Para além dos relatórios, existem rankings de escolas feitos por jornais. Os jornais Público e Expresso fazem rankings das escolas portuguesas, com base na média dos resultados obtidos nos exames nacionais, em cada escola [7] [8].

### Relatórios de Avaliação Interna

As escolas podem estar inseridas num agrupamento escolar, pelo que os relatórios internos podem ser feitos para uma escola ou para as escolas do agrupamento. As equipas de avaliação interna são formadas por docentes das escolas avaliadas. Nos relatórios de avaliação interna feitos pelas escolas são analisados os seguintes pontos:

- Resultados dos alunos no ano letivo anterior ou num conjunto de anos letivos anterior;
- População Escolar;
- Recursos Humanos;
- Candidaturas ao Ensino Superior.

Em relação aos resultados dos alunos são analisados dados por ciclo escolar. No 1º ciclo são analisados as percentagens de aprovações dos alunos no 1º ano, 2º ano, 3º ano e 4º ano. No 4º ano são feitas análises às percentagens de aprovações nos exames nacionais de Língua Portuguesa e Matemática. No 2º ciclo e no 3º ciclo as notas finais vão de 1 a 5. São feitas análises às percentagens de positivas por disciplina e em cada disciplina são analisados as percentagens de 1,2,3,4 e 5. É analisado também a média por disciplina e feita a comparação com a média nacional. São feitas também análises à percentagem de alunos que transitaram para o próximo ano escolar. No 6º ano e no 9º ano são feitos exames nacionais a Língua Portuguesa e Matemática. Para cada um dos exames no ano de escolaridade respetivo, são analisadas as médias obtidas e a percentagem de positivas. A média de exame é comparada com a média de frequência da disciplina. São analisados também as percentagens de 1,2,3,4 e 5 nos respetivos exames. No secundário as notas finais vão de 0 a 20. Como já referi anteriormente no secundário existem vários cursos que um aluno pode escolher. Para cada curso podem ser analisados as percentagens de aprovações em cada ano escolar do secundário. São feitas análises às percentagens de positivas por disciplina. Por cada disciplina é analisada a média e comparada com a média nacional. Para cada exame nacional no secundário é analisado a média de exame e é comparada com a média interna da disciplina. Para além disso são analisadas as percentagens de positivas em cada exame. Quando o relatório de avaliação interna é feito para um agrupamento escolar, são feitas comparações entre os resultados



obtidos pelas escolas do agrupamento. Os resultados das escolas do agrupamento são também comparados com os resultados obtidos a nível nacional.

Relativamente à população escolar é analisado o número de alunos por ciclo escolar. No secundário é analisado o número de alunos por curso. São analisados também o número de alunos por disciplina, por sexo, por nacionalidade e por escalão social. Para os pais são analisados dados sobre habilitações e profissões que os pais podem ter. Nas habilitações são feitas contagens do número de pais por categoria de habilitação e nas profissões são feitas contagens do número de pais por categoria de profissão. As categorias de profissão consideradas são Quadros Dirigentes, Quadros Superiores, Quadros Intermédios, Trabalhadores não Qualificados e Outros. As categorias de habilitação consideradas são Doutoramento e Mestrado, Licenciatura e Bacharelato, Secundário e 3º Ciclo, Básico (1º e 2º ciclos) e Sem Habilitações/Desconhecida.

Nos Recursos Humanos são feitas análises relativamente ao pessoal docente e não docente. Em relação ao pessoal docente é analisado o número de docentes por faixa de idade e por faixa de tempo de serviço e é analisado o número de docentes por habilitação. São feitas as mesmas análises para o pessoal não docente.

Nas candidaturas ao Ensino Superior são analisados o número de candidatos e o número de colocados na primeira fase e na segunda fase. São analisados também os cursos e universidades mais procurados pelos alunos.

#### **Relatórios de Avaliação Externa**

Como já foi referido anteriormente os relatórios de avaliação externa são feitos pela Inspeção Geral da Educação e Ciência no fim de cada ano letivo [5]. Cada relatório de avaliação externa tem a avaliação feita às escolas de um agrupamento escolar ou a uma escola sem agrupamento. Um relatório de avaliação externa é feito com base em análises tendo em conta dados de contexto, taxas de transição ou conclusão e disciplinas com exame.

Nos dados de contexto são feitas as seguintes análises:

- Número de alunos e turmas;
- Número de alunos por escalão social;
- Número de alunos por naturalidade;
- Número de alunos por filiação;
- Número de funcionários por idade e tempo de serviço.

Em relação ao número de alunos e turmas, para cada ano escolar e ciclo escolar são analisados o número de alunos, o número de turmas, a idade média dos alunos e a percentagem de alunos do sexo feminino. No número de alunos por escalão social, para cada ano escolar e ciclo escolar são analisados o número de alunos com escalão A, B e sem escalão. No número de alunos por naturalidade, para cada país são analisados o número de alunos no ensino básico e no ensino secundário. No número de alunos por filiação são feitas análises relativamente às habilitações dos pais e às profissões dos pais. Nas habilitações dos pais, para cada habilitação são feitas contagens do número de pais e do número de mães com essa habilitação no ensino básico e secundário. Nas profissões dos pais, para cada área de profissão são feitas contagens do número de pais e do número de mães com essa área de profissão no ensino básico e secundário. No número de funcionários por idade e tempo de serviço, para cada faixa de idades e faixa de tempo de serviço são analisados o número de funcionários docentes. O mesmo tipo de análise é feito para os funcionários não docentes.

As taxas de transição ou conclusão são analisadas por cada ano escolar de um ciclo escolar. Por exemplo no primeiro ciclo são analisadas as percentagens de transição e conclusão para o 1º ano, 2º ano, 3º ano e 4º ano. As análises são feitas de forma semelhante para os outros ciclos. Os resultados obtidos pelo agrupamento ou por uma escola sem agrupamento são comparados com os resultados obtidos a nível nacional. Nas disciplinas com exame são feitas análises para as disciplinas de Língua Portuguesa e Matemática no 4º ano, 6º ano e 9º ano. No 4º ano são analisados para cada exame as percentagens de positivas e são comparados com os resultados nacionais. No 6º ano são analisados para cada exame as percentagens de positivas e as percentagens de 1,2,3,4 e 5, e são comparados com os resultados nacionais. No 9º ano para cada exame é analisada a média de exame e a média interna da disciplina. No 9º ano são feitos os mesmos tipos de análise que no 6º ano. No secundário existem exames nacionais no 11º ano e no 12º ano. Para cada disciplina com exame, é analisado a média interna da disciplina e a média de exame. A média de exame da disciplina é comparada com a média de exame nacional.

#### **Relatórios de Avaliação do Júri Nacional de Exames**

Para cada ano letivo é feito um relatório de avaliação pelo Júri Nacional de Exames [6]. O relatório analisa dados relativamente aos exames nacionais que decorrem no 4º ano, 6º ano, 9º ano e no secundário. Existem duas fases para fazer exame. No 4º ano, 6º ano e 9º ano, existem exames às disciplinas de Língua Portuguesa e Matemática. Para os exames nesses anos escolares são analisados o número total de provas em escolas públicas e escolas privadas e o número total

de provas por sexo para cada disciplina. No 4º Ano a escala das provas vai de A a E. São analisados as percentagens de provas com A,B,C,D e E por sexo para cada disciplina, o número de provas com A,B,C,D e E por região e por sexo para cada disciplina e as percentagens de A,B,C,D e E para escolas públicas e privadas para cada disciplina. No 6º ano e no 9º ano a escala vai de 1 a 5. São analisados as percentagens de 1,2,3,4 e 5 para cada disciplina, a média de exame por sexo para cada disciplina, as percentagens de 1,2,3,4,5 por região e por sexo para cada disciplina e a média de exame por região a cada disciplina. No ensino secundário existem exames no 11º ano e 12º ano. Para os exames são analisados o número total de provas em escolas públicas e escolas privadas por fase de exame, número total de provas por sexo e por fase, número de provas realizadas e classificação média por fase para cada disciplina, diferença entre a classificação interna e a classificação de exame para as 8 disciplinas com maior número de alunos, média de classificação interna e de exame por região para cada disciplina e média de exame para as escolas públicas e privadas para as disciplinas de Português e Matemática A.

#### **Rankings de escolas feito pelo jornal Público e pelo jornal Expresso**

Por cada ano letivo, os jornais Público e Expresso fazem rankings das escolas [7] [8]. Os rankings são feitos com base nos resultados obtidos nos exames nacionais em cada escola. Nos jornais Público e Expresso são feitos rankings para o 6º ano, 9º ano e secundário. O jornal Expresso tem também rankings para o 4º ano. Todos os rankings têm para cada escola, informação do nome da escola, concelho, tipo de escola (pública ou privada), número de provas realizadas e média obtida nas provas. No jornal Público para os rankings do 9º ano e secundário, para cada escola pública, existe o contexto socioeconómico no qual a escola está inserida. As escolas públicas estão agrupadas em quatro contextos. O contexto socioeconómico menos favorecido tem o valor 1 e o mais favorecido tem o valor 4. O contexto socioeconómico para cada escola, é calculado tendo em conta o nível de escolaridade de ambos os pais, a profissão do pai e a percentagem de alunos com ação social escolar. Os rankings no jornal Público podem ser filtrados pelo concelho e tipo de escola. No 9º ano e no secundário os rankings podem também ser filtrados por contexto socioeconómico. Nos rankings do jornal Expresso para além da informação já descrita, existe também para cada escola o distrito, a percentagem de alunos com subsídio escolar, a percentagem da variação de médias entre o ano atual e o ano anterior e o ranking dessa percentagem com o número de posições que a escola sobe ou desce. Os rankings no jornal Expresso podem ser filtrados pelo distrito, concelho, tipo de escola e número de exames. O filtro relativamente ao número de exames permite

definir se apenas são mostradas as escolas com mais de 50 exames realizados ou se são mostradas todas as escolas.

### 2.1.2. Aplicações utilizadas

As escolas têm aplicações fornecidas por empresas que permitem fazer a gestão escolar. Duas dessas empresas são a Truncatura e a J.P.M & Abreu, LDA [9] [10]. A Truncatura disponibiliza aplicações para várias áreas de gestão numa escola. A seguir estão algumas dessas aplicações:

- WinGA [11];
- TProfessor [12];
- Módulo de Estatística [13];
- Módulo de Gráficos [13].

O *WinGA* é uma aplicação que permite fazer a gestão de alunos numa escola. A aplicação tem funcionalidades para lançamento de notas e faltas, inscrição nos exames, gestão de matrículas dos alunos e gestão de pautas [11]. A aplicação *TProfessor* é dedicada ao professor e permite, entre outros, registar as faltas dos alunos, registo de avaliações e marcação de testes. Esta aplicação permite também aos diretores de turma fazerem a gestão das faltas dos alunos [12]. O *Módulo de Estatística* permite analisar dados relativamente a alunos matriculados por ano escolar, idade, sexo, língua estrangeira, curso, disciplina, entre outros [13]. O *Módulo de Gráficos* permite visualizar e imprimir gráficos de notas relativas aos anos escolares a decorrer no ano letivo atual por período, ano escolar, disciplina e sexo [13]. A empresa J.P.M & Abreu, LDA tem uma aplicação que permite fazer a gestão de alunos. Essa aplicação tem o nome *Alunos*. A aplicação tem funcionalidades para a marcação de faltas dos alunos, lançamento de notas, gestão de matrículas, comunicações aos encarregados de educação e marcação de exames. A aplicação *Alunos* permite imprimir listagens tais como pautas de avaliação, assiduidade, gráficos de sucesso ou insucesso escolar, listas de turmas e impressos de matrícula. As aplicações da Truncatura e a aplicação *Alunos* da J.P.M & Abreu, LDA permitem exportar dados para o sistema MISI [14]. O sistema MISI é um sistema de informação que armazena dados das escolas. A recolha desses dados é feita através de aplicações de gestão escolar certificados pelo ministério de educação e ciência como é o caso das aplicações da Truncatura e da aplicação *Alunos* da J.P.M & Abreu, LDA. Com base nos dados recolhidos pelo sistema MISI o Ministério da Educação e Ciência pode fazer estatísticas com esses dados e com base nessas estatísticas podem ser feitos relatórios de avaliação [14].

## 2.2. Data Warehouse

Um Data Warehouse é um repositório de dados que é utilizado para dar suporte à tomada de decisões com base na análise dos dados [15] [16]. Um Data Warehouse normalmente tem dados provenientes de diversas fontes heterogéneas [15] [16]. Segundo Bill Inmon, um dos pioneiros em Data Warehousing, um Data Warehouse tem as seguintes características: [16].

- É orientado ao assunto;
- É integrado;
- É variante no tempo;
- É não volátil.

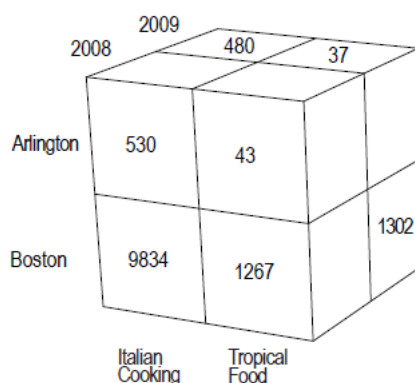
Um Data Warehouse é orientado ao assunto uma vez que é planeado em torno de assuntos importantes para um negócio, em que esses assuntos possam ser analisados [16]. Um Data Warehouse é integrado uma vez que recebe dados provenientes de fontes heterogéneas e esses dados têm de ser integrados de maneira a que a informação fique consistente [16]. Um Data Warehouse é variante no tempo uma vez que mostra a evolução dos dados ao longo tempo e não apenas os dados mais recentes [16]. Um Data Warehouse é não volátil uma vez que os dados não podem ser removidos nem alterados. A única modificação acontece no carregamento de novos dados [16].

### 2.2.1. Conceitos utilizados em Data Warehouse

Existem conceitos aplicados em Data Warehouse. A seguir são apresentados os conceitos de Cubo, Dimensões, Fatos e Medidas [16].

#### Cubo

Um cubo é uma estrutura multidimensional que permite a captura e análise dos dados sobre várias perspetivas. Um cubo pode ter várias dimensões que representam as perspetivas do cubo. Numa dimensão podem ser definidas hierarquias. Uma hierarquia permite na dimensão analisar os dados com o nível de detalhe desejado. Um cubo é composto de células em que cada uma delas é identificada através de interseções entre as dimensões. Uma célula não vazia representa um fato. Um fato pode ter várias medidas associadas. Na Figura 1 está um exemplo de um cubo [16]. O cubo tem três dimensões que são o *Tempo*, *Cidade* e *Livro*. A medida no exemplo é a contagem das vendas. Por exemplo em 2008 na cidade de Boston foram vendidos 1267 livros.

**Figura 1 – Cubo [16]**

### Dimensões

As dimensões são fundamentais nas Bases de dados multidimensionais, uma vez que permitem criar perspectivas num cubo que tenham interesse para análise [16]. As dimensões permitem selecionar dados e agrupar dados. Nas dimensões podem ser também definidas hierarquias. Uma hierarquia é composta por vários níveis em que cada um deles permite ter um nível de detalhe dos dados com interesse para as análises. Uma hierarquia permite agrupar os dados segundo a análise que se pretende fazer [16]. A seguir está um exemplo de uma hierarquia:

- País -> Continente.

A hierarquia apresentada tem dois níveis. O nível mais baixo da hierarquia é o País. Os valores do nível País são agrupados em valores do nível Continente. Quanto mais baixo for o nível da hierarquia maior é o nível de detalhe dos dados. Se considerarmos que a medida é a contagem da população, o nível Continente permite ver essa população por continente e apresenta menos detalhe que o nível País uma vez que este permite ver a população por país.

### Fatos e medidas

Os fatos resultam da combinação de valores das dimensões. Se numa combinação particular resulta uma célula não vazia, essa célula representa um fato. Um fato pode ter várias medidas. Uma medida é uma propriedade de um fato. Por exemplo se o fato for uma venda, duas medidas possíveis podem ser o preço de venda e a quantidade vendida [16].

### 2.2.2. Operações para analisar um cubo

As bases de dados multidimensionais suportam a análise de grandes volumes de dados [16]. A seguir são apresentados operações que podem ser feitas num cubo:

- ROLL-UP;
- DRILL-DOWN;
- DRILL-OUT;
- SLICE AND DICE;
- DRILL-ACROSS.

A operação *ROLL-UP* permite passar de informação mais detalhada para informação menos detalhada numa hierarquia [16]. Por exemplo se o analista estiver a ver a população portuguesa por freguesia e quiser ver a população portuguesa por concelho pode fazer *ROLL-UP*.

A operação *DRILL-DOWN* é a operação inversa de *ROLL-UP* ou seja permite passar de informação menos detalhada para informação mais detalhada [16]. Voltado ao exemplo dado se o analista estiver a ver a população portuguesa por concelho e quiser obter a população por freguesia pode fazer *DRILL-DOWN*.

A operação *DRILL-OUT* permite ter mais detalhe na análise acrescentando uma ou mais dimensões a essa análise [16]. Voltando ao exemplo se o analista estiver a ver a população portuguesa por concelho e quiser ver a população por concelho e por sexo pode fazer *DRILL-OUT* acrescentado a dimensão sexo.

A operação *SLICE AND DICE* permite fazer análises a um subconjunto do cubo. A operação *SLICE AND DICE* pode ser dividida em *SLICE* e *DICE*. Voltando ao exemplo se o analista pretender analisar a população portuguesa das freguesias de um concelho, basta selecionar nessa dimensão o concelho que pretende observar. Neste caso o analista utilizou a operação *SLICE*. A operação *DICE* permite fazer análises com base em operações mais complexas como por exemplo analisar a população portuguesa das freguesias para dois concelhos no ano 2012 [16].

A operação *DRILL ACROSS* permite fazer análises combinando vários cubos [16]. Por exemplo considerando um supermercado que vende produtos na internet e na própria loja. Um cubo tem informação das vendas na loja e outro tem informação das vendas na internet. A operação *DRILL ACROSS* permite combinar os dois cubos para fazer análises, como por exemplo ver o número total de vendas para cada produto do supermercado, juntando as vendas feitas na internet com as vendas na loja.

### 2.2.3. Representações Relacionais

Em bases de dados relacionais existem duas formas principais de representar uma dimensão [16]:

- Star Schema – Esquema em Estrela;
- Snowflake Schema – “Esquema em flocos de neve”.

#### Star Schema

O *Star Schema* tem uma tabela de dimensão por cada dimensão e uma tabela de fatos. Uma dimensão tem uma coluna por cada nível de hierarquia e para cada um dos níveis tem uma coluna para cada propriedade do nível respectivo. Uma dimensão tem também uma coluna que permite guardar os valores da chave primária. A tabela de fatos tem uma coluna para cada medida. Tem também uma coluna por cada dimensão que exista. Essas colunas formam a chave primária da tabela de fatos e cada uma delas representa uma chave estrangeira que referencia a chave primária da dimensão respectiva. Um registo na tabela de fatos tem os valores das chaves estrangeiras que referenciam os valores das chaves primárias de cada dimensão e os valores de cada medida. Na Figura 2 está um exemplo de um *Star Schema*.

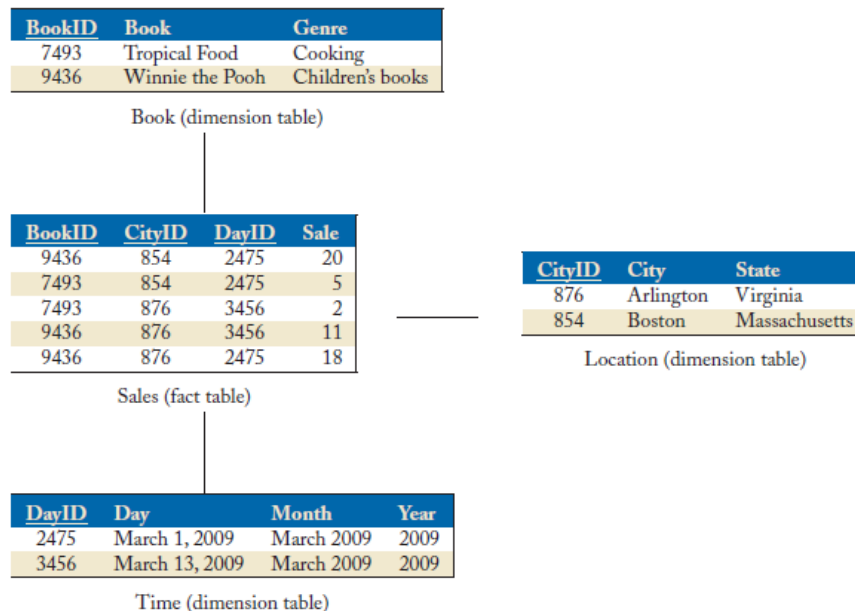


Figura 2 - Star Schema [16]

O *Star Schema* da Figura 2 tem três dimensões e a tabela de fatos. As dimensões são *Book* (*Book dimension*), *Location* (*Location Dimension*) e *Time* (*Time Dimension*). A tabela de fatos é a tabela *Sales* (*Sales fact table*). Cada uma das dimensões está associada a uma tabela com mesmo



nome da dimensão. A tabela *Book* permite guardar registos de livros, a tabela *Location* permite guardar registos de localizações e a tabela *Time* permite guardar registos temporais. Cada tabela de dimensão tem a primeira coluna com valores da chave primária e tem colunas que caracterizam a dimensão. Por exemplo a coluna *BookID* da tabela *Book* guarda os valores da chave primária dessa tabela. Tem também duas colunas, uma para guardar o nome de cada livro (*Book*) e outra para guardar a categoria de cada livro (*Genre*). As outras dimensões estão organizadas de forma semelhante. Na tabela de fatos *Sales*, as três primeiras colunas são chaves estrangeiras em que cada uma delas guarda os valores das chaves primárias da dimensão respetiva. A coluna *BookID* guarda valores da chave primária da coluna *BookID* da tabela *Book*. A coluna *CityID* guarda valores da chave primária da coluna *CityID* da tabela *Location*. A coluna *DayID* guarda valores da chave primária da coluna *CityID* da tabela *Time*. As três colunas da tabela de fatos *Sales* que são chaves estrangeiras, formam a chave primária. Essa tabela tem apenas uma medida representada pela coluna *Sales*. Essa medida guarda em cada registo a contagem das vendas. Cada registo da tabela de fatos *Sales* guarda uma combinação particular dos valores das três chaves estrangeiras e o número de vendas para essa combinação. Por exemplo no primeiro registo da tabela de fatos *Sales* foram feitas 20 vendas. Com as chaves estrangeiras consegue-se observar que foram feitas 20 vendas do livro “Winnie the Pooh” na cidade “Arlington” e no dia 1 de março de 2009. As tabelas de dimensão no *Star Schema* não estão normalizadas. Por exemplo na tabela *Time* são necessários 31 registos para o mês de março do ano de 2009 e nesses 31 registos o mês (*Month*) e o ano (*Year*) tem sempre o mesmo valor.

### Snowflake Schema

O *Snowflake Schema* tem uma tabela de fatos como no *Star Schema*. A diferença está nas dimensões uma vez que o *Snowflake Schema* tem várias tabelas para cada dimensão. Cada dimensão tem uma tabela por nível de hierarquia. Desta forma as dimensões estão normalizadas uma vez que não existe redundância de dados. Cada tabela das dimensões tem uma chave primária. Ao contrário do *Star Schema*, no *Snowflake Schema* existem também chaves estrangeiras nas dimensões. Cada tabela de nível mais baixo tem uma chave estrangeira com a referência da chave primária da tabela do nível a seguir. Cada tabela de uma dimensão tem uma coluna para cada propriedade do nível da dimensão. A tabela do nível mais elevado de uma dimensão, não tem chave estrangeira. A tabela de fatos tem uma coluna para cada medida. Tem também uma coluna por cada dimensão que exista. Essas colunas formam a chave primária da tabela de fatos e cada uma delas representa uma chave estrangeira que referencia a chave primária da tabela de nível mais

baixo da dimensão respetiva. Um registo na tabela de fatos tem os valores das chaves estrangeiras que referenciam os valores das chaves primárias de cada tabela de nível mais baixo das dimensões e os valores de cada medida. Na Figura 3 está um exemplo de um *Snowflake Schema* [16].

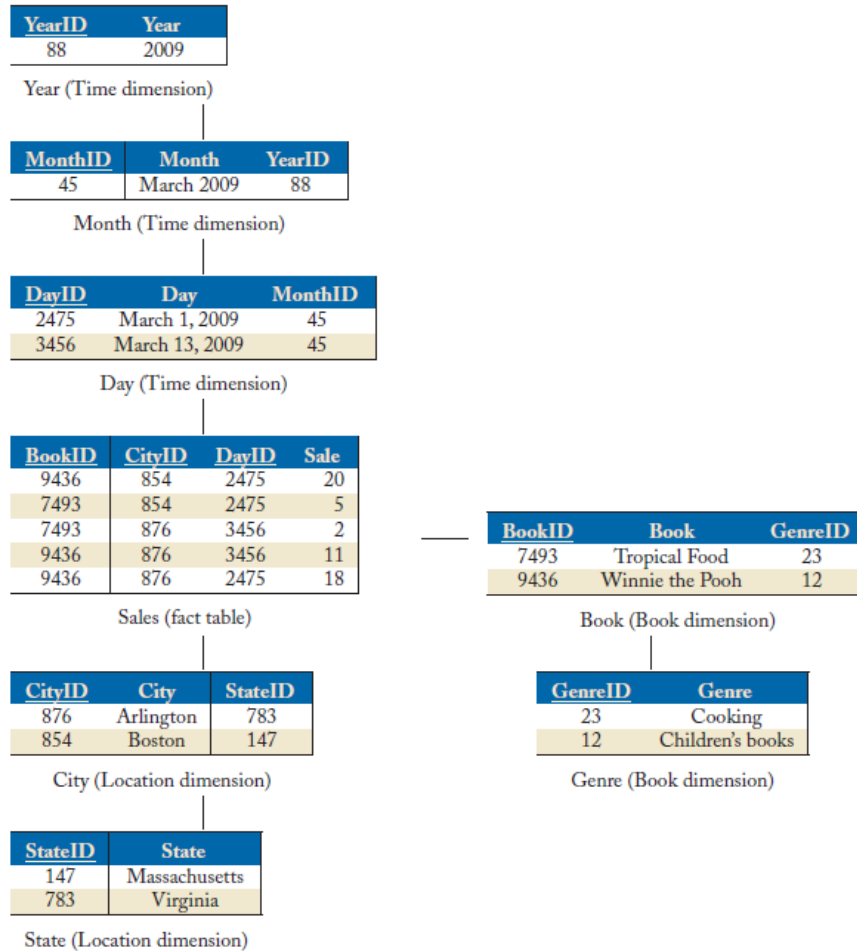


Figura 3 - Snowflake Schema [16]

O *Snowflake Schema* da Figura 3 permite guardar a mesma informação que o *Star Schema* da Figura 2. A diferença está nas dimensões uma vez que no *Snowflake Schema* cada dimensão tem uma tabela por cada nível de hierarquia. A dimensão *Book* tem duas tabelas que são *Book* e *Genre*. A dimensão *Location* tem duas tabelas que são *City* e *State* e a dimensão *Time* tem três tabelas que são *Day*, *Month* e *Year*. Cada tabela de uma dimensão tem a primeira coluna com valores da chave primária e tem uma coluna por cada propriedade do nível da dimensão. As tabelas de dimensão têm na última coluna os valores das chaves estrangeiras que referenciam os valores das chaves primárias da tabela de dimensão do nível a seguir. A tabela de nível mais elevado não tem chaves estrangeiras. Por exemplo na dimensão *Book*, a coluna *BookID* da tabela *Book* guarda os valores da chave primária dessa tabela. Tem também uma coluna para guardar o nome de cada livro (*Book*).

Na última coluna tem os valores da chave estrangeira que referenciam os valores da chave primária da tabela *Genre*. A tabela *Genre* guarda os valores da chave primária dessa tabela na coluna *GenreID*. Tem também uma coluna para guardar as categorias de livros que existe (*Genre*). A tabela *Genre* não tem uma coluna com chave estrangeira uma vez que representa o nível mais elevado da hierarquia da dimensão *Book*. As outras dimensões estão organizadas de forma semelhante. Na tabela de fatos *Sales*, as três primeiras colunas são chaves estrangeiras em que cada uma delas guarda os valores das chaves primárias da tabela de nível mais baixo da dimensão respetiva. A coluna *BookID* guarda valores da chave primária da coluna *BookID* da tabela *Book*. A coluna *CityID* guarda valores da chave primária da coluna *CityID* da tabela *City*. A coluna *DayID* guarda valores da chave primária da coluna *DayID* da tabela *Day*. As três colunas da tabela de fatos que são chaves estrangeiras, formam a chave primária da tabela de fatos. Tal como no *Star Schema* a tabela de fatos *Sales* tem uma medida representada pela coluna *Sales*. Essa medida guarda em cada registo a contagem das vendas. Cada registo da tabela de fatos *Sales* guarda uma combinação particular dos valores das três chaves estrangeiras e o número de vendas para essa combinação. Por exemplo no primeiro registo da tabela de fatos *Sales* foram feitas 20 vendas. Com as chaves estrangeiras consegue-se observar que foram feitas 20 vendas do livro “Winnie the Pooh” na cidade “Arlington” e no dia 1 de março de 2009. Para ver a categoria do livro “Winnie the Pooh” é necessário observar qual é o valor da chave estrangeira na coluna *GenreID*. Com essa chave é possível observar na tabela *GenreID* que a categoria do livro “Winnie the Pooh” é “Children’s books”. Nas outras dimensões a informação é obtida de forma semelhante. As tabelas de dimensão no *Snowflake Schema* estão normalizadas ao contrário das mesmas dimensões no *Star Schema*. Por exemplo na tabela *Time* no *Star Schema*, são necessários 31 registos para o mês de março do ano de 2009 e nesses 31 registos o mês (*Month*) e o ano (*Year*) tem sempre o mesmo valor. Nas tabelas da mesma dimensão no *Snowflake Schema*, para guardar valores do mês de março de 2009 apenas é necessário um registo na tabela *Year* com o ano de 2009 e um registo na tabela *Month* com o mês de março de 2009. Os registos para o mês e para o ano não se repetem no *Snowflake Schema*.

### 2.3. Proposta de Solução

Como já referi anteriormente as duas escolas do qual obtive colaboração forneceram relatórios de avaliação feitos nas respetivas escolas. Nas Figura 4 e na Figura 5 estão gráficos obtidos de um relatório de avaliação interna de uma escola.

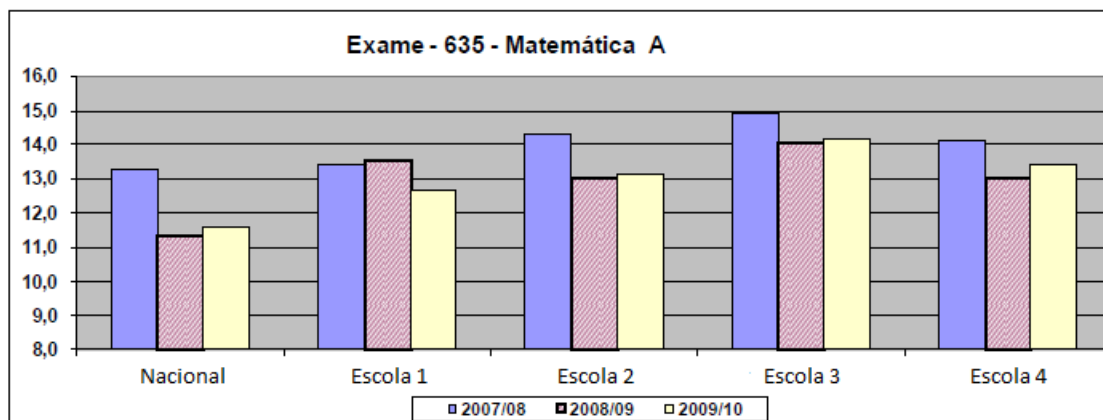


Figura 4 – Notas no Exame de Matemática por Escolas do Agrupamento

O gráfico da Figura 4 mostra a média de exame a matemática obtida nas escolas do agrupamento e a nível nacional para os anos letivos 2007/2008, 2008/2009 e 2009/2010. O gráfico só consegue mostrar as notas obtidas pelas escolas do agrupamento a uma disciplina e não se pode ver por exemplo o desempenho escolar nesses 3 anos letivos uma vez que estes são mostrados de forma separada. O gráfico da Figura 5 mostra o número de pais de alunos por categoria de habilitação no secundário de uma escola. As categorias de habilitação da Figura 5 são:

- A – Doutoramento e Mestrado;
- B – Licenciatura, Bacharelato;
- C – Secundário e 3.º ciclo;
- D – Básico (1.º e 2.º ciclos);
- E – Sem habilitações/Desconhecida.

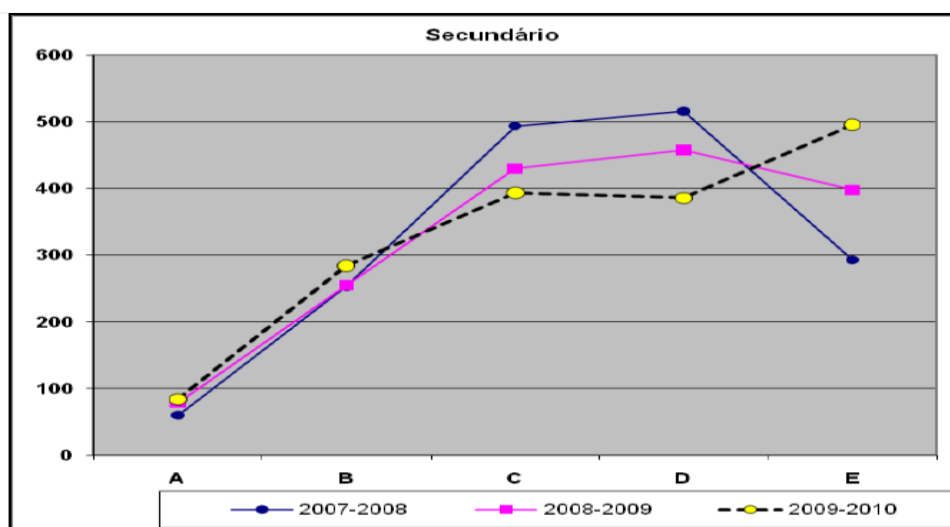


Figura 5 – Número de pais por categoria de habilitação

O gráfico mostra o número de pais por categoria de habilitação nos anos letivos 2007/2008, 2008/2009 e 2009/2010. O gráfico mostra apenas contagens sem verificar em que medida as categorias de habilitação influenciam no desempenho escolar dos alunos.

Os gráficos da Figura 4 e da Figura 5 são gráficos típicos presentes nos relatórios de avaliação. Existem outros gráficos nesses relatórios em que o tratamento dos dados é feito de forma semelhante que nos gráficos apresentados. Como nos relatórios de avaliação não é possível observar as componentes que influenciam o desempenho escolar dos alunos, a minha proposta de solução é o desenvolvimento de um projeto conceptual de um sistema de Data Warehousing para a Empresa, com o objetivo de analisar as componentes que influenciam o desempenho escolar dos alunos. Com base nas análises a Empresa poderá prestar serviços às escolas aderentes.

## 2.4. Trabalhos realizados na área do Ensino

Neste ponto são apresentados alguns trabalhos realizados na área do ensino. Em cada um dos trabalhos são descritos os sistemas utilizados e as análises que são feitas.

O artigo “School Performance Evaluation in Portugal: A Data Warehouse Implementation to Automate Information Analysis” apresenta a implementação de um Sistema de Data Warehouse para analisar as performances entre escolas e para analisar as performances internas de cada escola [17]. Foram utilizados dados reais provenientes do Ministério da Educação e de uma escola (Colégio Paulo VI de Gondomar). O objetivo final do sistema é fornecer informações que ajudam os gestores das escolas a tomar decisões. Existem algumas diferenças entre as componentes consideradas para o sistema de Data Warehouse descrito no artigo e o sistema de Data Warehouse do trabalho. O sistema de Data Warehouse descrito no artigo apenas foi pensado para guardar notas de disciplinas com exame ao contrário do sistema de Data Warehouse do trabalho. O sistema descrito no artigo não considera as componentes habilitações e profissões dos pais e escalão social dos alunos. Por fim, o sistema descrito no artigo considera a componente professor ao contrário do sistema de Data Warehouse do trabalho.

O artigo “Use Data Warehouse and Data Mining to Predict Student Academic Performance in Schools: A Case Study (Perspective Application and Benefits)” propõe um modelo que é aplicado em Data Warehouse e utiliza técnicas de Data Mining para prever o desempenho dos alunos nas escolas [18]. O modelo proposto foi modelado com base nos dados que existem numa escola (“BINUS International School na Indonésia”). O Data Warehouse foi projetado com base na metodologia de Kimball e permite que sejam feitas análises através de técnicas de Data Mining. As técnicas de Data Mining são utilizadas para prever o estado de avaliação de um aluno (aprovado ou

não aprovado) com base no seu desempenho escolar passado. O Data Warehouse descrito no artigo permite guardar as notas dos testes e notas finais de cada disciplina ao contrário do sistema de Data Warehouse do trabalho que apenas permite guardar as notas finais de cada disciplina. O Data Warehouse descrito no artigo permite também guardar o número de presenças dos alunos nas aulas. Por fim no Data Warehouse descrito no artigo não são consideradas algumas componentes que possam influenciar o desempenho escolar dos alunos tais como habilitações e profissões dos pais e escalão social dos alunos.

O artigo “SBIAES — Business intelligence system for analysis of access to higher education: The case of the Polytechnic Institute of Coimbra” apresenta o desenvolvimento de um sistema de business intelligence para analisar as candidaturas de acesso ao ensino superior. O sistema foi desenvolvido para o Instituto Politécnico de Coimbra [19]. O sistema é composto por um Data Warehouse e uma aplicação cliente que permite fazer análises no Data Warehouse. O sistema integra os dados das bases de dados dos candidatos ao Concurso Nacional de Acesso (CNA), que colocaram em algumas das opções de candidatura uma escola do Instituto Politécnico de Coimbra. O Data Warehouse foi desenvolvido com base na metodologia de Kimball e permite guardar quais são as universidades e cursos preferidos dos alunos, as escolas secundárias e cursos secundários de onde vem os alunos e o número de vagas de cada curso do ensino superior. Os dados foram obtidos das bases de dados dos candidatos ao CNA disponibilizados pela Direção Geral do Ensino Superior (DGES). O sistema desenvolvido permite aos decisores do Instituto Politécnico de Coimbra tomar decisões com base nos resultados das análises feitas. O sistema descrito no artigo permite fazer análises com base nas candidaturas ao ensino superior e é um exemplo de aplicação de sistemas de Data Warehouse no ensino superior.

O artigo “Data Warehouse Based Evaluation of Students' Achievements in Information Systems Education” apresenta a análise de dados dos resultados dos alunos nos testes e projetos realizados. Para analisar esses dados foi implementado um Data Warehouse [20]. Foram recolhidos dados de cursos na área de sistemas de informação de uma universidade da Servia (“University of Novi Sad, Technical faculty "Mihajlo Pupin", Zrenjanin, Serbia”). O Data Warehouse permite guardar as notas dos testes e dos projetos que cada aluno têm no curso respetivo. As análises no Data Warehouse foram feitas através da utilização de uma ferramenta OLAP. O Data Warehouse descrito no artigo permite analisar dados relativamente a alunos que estão no ensino superior enquanto que o sistema de Data Warehouse do trabalho permite analisar dados dos alunos do ensino básico (2º e 3º ciclo) e secundário. No Data Warehouse do artigo foi considerado a componente do professor ao contrário do sistema de Data Warehouse do trabalho. Por fim, no Data Warehouse descrito no

artigo não são consideradas algumas componentes que possam influenciar o desempenho escolar dos alunos tais como habilitações e profissões dos pais, escalão social dos alunos, sexo dos alunos e idade dos alunos.





### 3. Caracterização e Planeamento

Para implementar um sistema de Data Warehousing é necessário saber o que se pretende analisar e os resultados que se pretendem obter. Para isso é importante fazer o levantamento dos recursos disponíveis, estudar e caracterizar os resultados pretendidos. A fase de caracterização e planeamento é fundamental para obter um bom sistema de Data Warehousing que permita fazer as análises pretendidas.

#### 3.1. Análise e Caracterização dos Recursos Disponíveis

Nos recursos disponíveis é necessário analisar quais são os recursos operacionais, os recursos materiais e os recursos humanos.

##### 3.1.1. Recursos Operacionais

Como já foi referido anteriormente, obtive a colaboração de duas escolas. Foram feitas reuniões com membros das escolas para verificar a informação existente. Depois de verificada a informação que existe nas escolas foi necessário selecionar o tipo de informação que se pretende analisar e verificar que informação poderia ser disponibilizada pelas escolas. Como existem duas escolas foram consideradas duas fontes de dados, uma para cada escola. Cada uma das escolas representa um agrupamento escolar, por isso em cada fonte de dados existe informação das escolas do agrupamento respetivo. As fontes de dados de ambas as escolas estão em ficheiros Excel. Apesar de ambas as escolas fornecerem ficheiros em Excel, esses ficheiros têm um esquema diferente em cada fonte. As fontes de dados têm informação sobre as escolas, características dos alunos e contexto familiar. Relativamente às escolas existe informação sobre as notas dos alunos em cada disciplina organizados por anos letivos e anos escolares e o estado de avaliação dos alunos (Aprovado ou Não Aprovado). Em relação às características dos alunos existe para cada aluno, informação sobre o sexo, data de nascimento, se tem necessidades educativas especiais e em cada ano letivo se tem algum escalão social. Em relação ao contexto familiar existe informação sobre as habilitações e profissões do pai e da mãe. São conhecidas duas fontes em que cada uma tem um esquema diferente. Num contexto real em que existam várias escolas que adiram ao serviço da Empresa é possível que existam fontes de dados adicionais com diferentes esquemas e diferentes tipos de formato.

### 3.1.2. Recursos Materiais

Cada escola tem uma aplicação para enviar os ficheiros de dados para o Data Warehouse da Empresa. Sendo assim a escola tem de ter computadores para aceder à aplicação e ter rede para enviar os ficheiros.

### 3.1.3. Recursos Humanos

Cada escola tem de ter uma pessoa responsável pelo envio dos ficheiros de dados para a Empresa.

## 3.2. Estudo e Caracterização dos Resultados Pretendidos

Para que o sistema de Data Warehousing da Empresa permita fazer as análises desejadas, é necessário saber quais são os resultados pretendidos.

Os resultados pretendidos são os seguintes:

- Rankings de escolas;
- Efeito do contexto familiar no desempenho escolar dos alunos;
- Efeito do escalão social no desempenho escolar dos alunos;
- Comparar o desempenho escolar dos alunos por sexo;
- Analisar o desempenho escolar dos alunos com necessidades educativas especiais;
- Relacionar a idade dos alunos com o ano escolar.

Pretende-se analisar os resultados mencionados no tempo, através do ano letivo ou de um triénio que representa uma faixa de 3 anos letivos.

### 3.2.1. Ranking de Escolas

Os rankings de escolas podem ser obtidos através das percentagens de aprovações dos alunos e das notas dos alunos. Em relação às percentagens de aprovações pretende-se obter rankings de escolas ordenados de forma decrescente pela percentagem de aprovações. Em relação às notas pretende-se obter rankings de escolas para uma determinada disciplina ordenado de forma decrescente pela média da disciplina. Podem existir dois tipos desses rankings:

- Ranking em que a média é calculada a partir de notas finais dos alunos;
- Ranking em que a média é calculada a partir de notas de exame do aluno, em disciplinas com exame.

Os rankings podem também ser obtidos para um determinado ano escolar ou ciclo escolar. Com os rankings das escolas podem ser feitas comparações de desempenho escolar entre as escolas.

### 3.2.2. Efeito do contexto familiar no desempenho escolar dos alunos

Neste ponto pretende-se analisar se existe uma relação entre as habilitações ou áreas de profissão que os pais podem ter com o desempenho escolar dos alunos. Essa análise pode ser feita através de percentagens de aprovações ou notas. Em relação às percentagens de aprovações pretende-se ver se habilitações elevadas correspondem a melhores percentagens de aprovações, analisando a habilitação da mãe e a habilitação do pai. Por exemplo pretende-se verificar se a habilitação Ensino Superior do pai corresponde a uma percentagem de aprovações mais elevada. Em relação às notas pretende-se ver para uma determinada disciplina se habilitações elevadas correspondem a melhores médias nessa disciplina. Pretende-se obter resultados semelhantes para as áreas de profissão dos pais.

Os resultados descritos podem também ser obtidos para um determinado ano escolar ou ciclo escolar.

### 3.2.3. Efeito do escalão social no desempenho escolar dos alunos

Neste ponto pretende-se analisar em que medida existe uma relação entre os escalões sociais dos alunos com o desempenho escolar dos alunos. Esta análise pode ser feita através de percentagens de aprovações ou notas. Em relação às percentagens de aprovações pretende-se verificar se a percentagem de aprovações é maior nos alunos com escalão, e nos alunos com escalão pretende-se analisar se a percentagem de aprovações é maior em alunos com escalão A. Em relação às notas pretende-se verificar se a média para uma determinada disciplina é maior nos alunos com escalão, e nos alunos com escalão pretende-se analisar se a média é maior em alunos com escalão A.

Os resultados descritos podem também ser obtidos para um determinado ano escolar ou ciclo escolar.

### 3.2.4. Comparar o desempenho escolar dos alunos por sexo

Neste ponto pretende-se analisar o desempenho escolar dos alunos de sexo masculino e compara-los com o desempenho escolar dos alunos do sexo feminino. Esta análise pode ser feita através de percentagens de aprovações ou notas. Em relação às percentagens de aprovações

pretende-se verificar se esta é mais elevada em alunos de sexo masculino ou em alunos de sexo feminino. Em relação às notas pretende-se ver para uma determinada disciplina se a média é maior em alunos de sexo masculino ou em alunos de sexo feminino.

Os resultados descritos podem também ser obtidos para um determinado ano escolar ou ciclo escolar.

### **3.2.5. Analisar o desempenho escolar dos alunos com necessidades educativas especiais**

Neste ponto pretende-se analisar o desempenho escolar dos alunos com necessidades educativas especiais. Esta análise pode ser feita através de percentagens de aprovações ou notas. Em relação às percentagens de aprovações pretende-se verificar se esta é elevada ou baixa em alunos com necessidades educativas especiais. Em relação às notas pretende-se ver para uma determinada disciplina se a média obtida por alunos com necessidades educativas especiais é elevada ou baixa.

Os resultados descritos podem também ser obtidos para um determinado ano escolar ou ciclo escolar.

### **3.2.6. Relacionar a idade dos alunos com o ano escolar**

Neste ponto pretende-se analisar a percentagem de alunos por idade para um determinado ano escolar. O objetivo é observar para um determinado ano escolar a percentagem de alunos com idades superiores ao que era suposto nesse ano. Por exemplo no 5º ano de escolaridade se houver 20% de alunos com 13 anos significa que esses alunos já reprovaram na escola.

## **3.3. Especificação dos Requisitos do Sistema**

A especificação dos requisitos que o Data Warehouse deve satisfazer é importante para ir de encontro com os resultados pretendidos. Para isso é necessário especificar os requisitos funcionais, operacionais e temporais. Nos requisitos funcionais são detalhadas as análises que se pretendem obter do Data Warehouse. Nos requisitos operacionais é feita uma descrição dos recursos físicos e lógicos necessários para o funcionamento do sistema. Nos requisitos temporais é definida a melhor janela temporal para o carregamento do Data Warehouse.

### **3.3.1. Requisitos Funcionais**

Para que o Data Warehouse permita fazer as análises pretendidas é necessário que os requisitos funcionais estejam bem especificados.

Para analisar o desempenho escolar dos alunos é necessário ter em conta:

- Notas dos alunos
  - 5º Ano ao 9º Ano – Notas de 1 a 5
    - *CF* – classificação interna anual a uma disciplina;
    - Exame – classificação de exame a uma disciplina;
    - *CFD* – classificação final anual a uma disciplina. Se a disciplina não tem exame *CFD* é igual a *CF*. Se a disciplina tem exame *CFD* é obtido a partir de *CF* e *Exame*.
  - 10º Ano ao 12º Ano – Notas de 0 a 20
    - *CF* – classificação interna anual de uma disciplina;
    - *CIF* – classificação média interna a uma disciplina no ano escolar de conclusão. Nas disciplinas bienais e trienais o *CIF* é a média aritmética simples das classificações internas anuais (*CFs*) [21]. Numa disciplina anual o *CIF* é igual ao *CF*. Nas disciplinas que não estão em ano de conclusão, não existe *CIF*;
    - Exame – classificação de exame a uma disciplina;
    - *CFD* – classificação média final a uma disciplina no ano escolar de conclusão. Se a disciplina não tem exame *CFD* é igual a *CIF*. Se a disciplina tem exame, *CFD* é obtido a partir de *CIF* e *Exame*. Nas disciplinas que não estão em ano de conclusão, não existe *CFD*.
- Contagens de matrículas – Uma matrícula representa a frequência de um aluno numa escola e num determinado ano letivo.

Para analisar o desempenho escolar dos alunos foram consideradas as seguintes perspetivas:

- Ano Escolar – cada ano escolar é definido por:
  - Ano escolar;
  - Ciclo escolar.
- Escola – cada escola é definida por:
  - Nome da escola;
  - Agrupamento escolar;
  - Tipo de Escola (Pública ou Privada);
  - Freguesia, Conselho e Distrito.
- Características do Aluno – as características que um aluno pode ter são:

- Sexo;
- Idade;
- Escalão Social;
- Se tem ou não necessidades educativas especiais.
- Contexto Familiar – cada contexto familiar é definido por:
  - Habilitação do pai;
  - Habilitação da mãe;
  - Área de profissão do pai;
  - Área de profissão da mãe.
- Estado de Avaliação: cada estado de avaliação é definido por:
  - Nome (Aprovado ou não aprovado).
- Disciplina – cada disciplina é definida por:
  - Nome da disciplina.

Para além das perspetivas descritas, existe uma perspetiva temporal de maneira a que os dados possam estar organizados no tempo. A perspetiva temporal é a seguinte:

- Ano Letivo – cada ano letivo é definido por:
  - Ano letivo;
  - Triénio – faixa de 3 anos letivos.

Como podemos ver os dados estão organizados de forma anual e por triénio.

### 3.3.2. Requisitos Operacionais

Para o desenvolvimento de um sistema de Data Warehousing não basta especificar os requisitos funcionais uma vez que é necessário garantir que o sistema vai funcionar corretamente. Para isso é necessário definir os requisitos operacionais. São necessários recursos de hardware e de software. Em termos de hardware é necessário ter um ou mais servidores que suportem o Data Warehouse. Em termos de software os servidores têm de ter sistema operativo e um sistema de base de dados. Na Figura 6 está a arquitetura geral da instalação física do Data Warehouse.

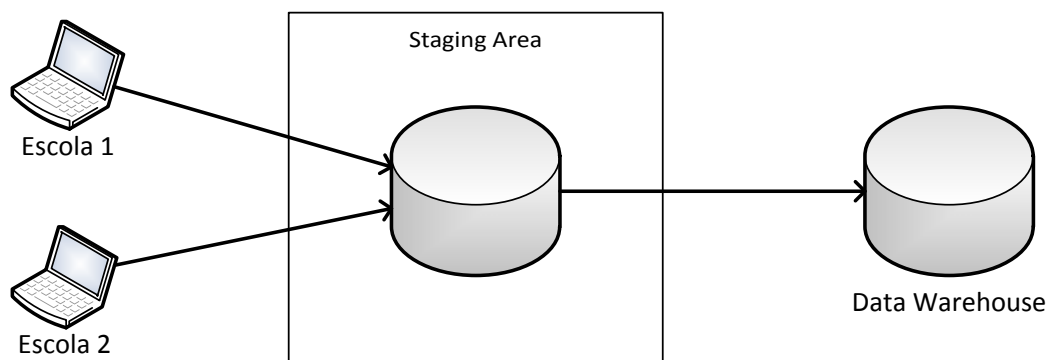


Figura 6 – Arquitetura do Data Warehouse

Cada escola tem uma aplicação que permite enviar ficheiros de dados necessários para analisar as componentes que influenciam o desempenho escolar dos alunos nas escolas. Os dados são extraídos para a *Staging Area* onde é feito o tratamento da informação de maneira a que esta possa ser inserida de forma correta no Data Warehouse.

### 3.3.3. Requisitos Temporais

O sistema de Data Warehousing da Empresa tem atualizações anuais uma vez que as escolas só têm as notas dos alunos no final de cada ano letivo. O carregamento do Data Warehouse deve ser feito após o final do ano letivo quando a Empresa receber os dados das escolas aderentes.

## 3.4. Planeamento e Gestão de Projeto

Para obter os resultados pretendidos do sistema de Data Warehousing é necessário fazer o planeamento das fases do projeto desde dos objetivos pretendidos até ao carregamento do Data Warehouse. A seguir estão descritas as fases do projeto:

- Análise e caracterização dos recursos disponíveis
  - Analisar para cada escola se os dados enviados estão corretos;
  - Analisar para cada escola se não existe falta de dados.
- Estudo e Caracterização dos Resultados Pretendidos
  - Estudo das componentes que influenciam as percentagens de aprovações dos alunos;
  - Relacionar a idade dos alunos com o ano escolar;
  - Estudo das componentes que influenciam as notas dos alunos.
- Especificação dos requisitos do sistema

- Caracterização dos requisitos funcionais do sistema de maneira a permitir análises tendo em conta:
  - Escola;
  - Anos Letivo;
  - Anos Escolar;
  - Disciplina;
  - Contexto Familiar;
  - Características do Aluno;
  - Estado de Avaliação.
- Caracterização dos Requisitos operacionais para instalar o Data Warehouse;
- Caracterização dos Requisitos temporais de maneira a definir a melhor janela temporal para carregar o Data Warehouse.
- Modelação multidimensional do Sistema
  - Caracterização das estrelas;
  - Caracterização das tabelas de fatos nomeadamente das medidas e da granularidade;
  - Caracterização das Dimensões.
- Especificação do processo ETL
  - Extract
    - Estudo da forma como vai ser feita a extração dos dados dos ficheiros de dados das escolas para a “Staging Area”;
  - Transform
    - Integração dos dados das duas fontes;
    - Definição de chaves primárias surrogate keys (SK) a partir de chaves alternativas (AK).
  - Load
    - Estudar os mecanismos de carregamento das estrelas.

Para que as fases do projeto e suas tarefas possam ser executadas corretamente é importante fazer uma boa distribuição de tarefas ao longo do projeto. Para isso as tarefas têm de ser distribuídas pelos elementos da equipa do projeto de maneira a evitar duplicação de tarefas. Saber



quem executa cada tarefa permite também responsabilizar os membros da equipa. As responsabilidades devem ser organizadas entre os elementos da equipa do projeto respeitando os seguintes pontos:

- Caracterização dos resultados pretendidos dos dados das escolas;
- Especificação do modelo multidimensional;
- Especificação da arquitetura da base de dados;
- Especificação do sistema ETL.

### 3.5. Aprovação do Sistema e Validação do Projeto

Por muito bom que seja a conceptualização e planeamento do sistema de Data Warehousing a validação deste depende de vários fatores tais como:

- Custos de implementação;
- Custos de manutenção;
- Utilidade real do sistema.

Nos pontos referidos o mais difícil de prever é a utilidade real do sistema uma vez que só com a utilização deste se poderá verificar se é útil ou não. Para que o sistema tenha sucesso é necessário existirem escolas que adiram e que os resultados obtidos a partir do sistema sejam uma mais-valia para essas escolas.

#### 3.5.1. Validade do Sistema Proposto

O sistema de Data Warehousing proposto permite dar suporte à tomada de decisões relacionadas com as componentes que influenciam o desempenho escolar dos alunos nas escolas. O Data Warehouse permite obter resultados de forma mais fácil do que nas bases de dados das escolas. Para fazer análises nas bases de dados das escolas seria necessário aceder a cada uma das bases de dados e depois fazer a integração dos dados obtidos para ter resultados próximos dos pretendidos. No Data Warehouse os dados obtidos das escolas são guardados de forma integrada o que vai permitir obter os resultados pretendidos de forma mais fácil. Outra vantagem do Data Warehouse é que permite fazer análises de forma rápida combinando as dimensões que se pretende analisar.

Pelas razões apontadas o sistema de Data Warehousing é o ideal para obter os resultados pretendidos.

### 3.5.2. Validade do Projeto de Desenho e Conceptualização

Para que o projeto seja validado, em cada fase do projeto foram detalhados os procedimentos a aplicar para a conceção do sistema de Data Warehousing desde do seu planeamento até à implementação. Com base nos procedimentos detalhados é possível fazer uma estimativa dos custos associados à implementação do sistema. Apesar dos custos associados à implementação, existe a convicção que o sistema de Data Warehouse é uma mais-valia para analisar as componentes que influenciam o desempenho escolar dos alunos nas escolas, quando comparado com as metodologias atualmente utilizadas.

### 3.6. Foco do Problema

A Empresa presta serviços de consultadoria às escolas que pretenderem aderir, com base na análise das componentes que influenciam o desempenho escolar dos alunos nessas escolas. Para isso as escolas têm de fornecer dados de maneira a serem analisados no Data Warehouse. Com os dados fornecidos pretende-se fazer análises ao nível das matrículas dos alunos e ao nível das notas dos alunos. Numa análise às perspetivas definiram-se as seguintes propriedades como relevantes:

- Ano Escolar: ano escolar e ciclo escolar;
- Escola: nome da escola, agrupamento escolar, tipo de escola, freguesia, concelho e distrito;
- Contexto Familiar: habilitação do pai, habilitação da mãe, área de profissão do pai e área de profissão da mãe;
- Características do aluno: sexo, idade, escalão e se tem ou não necessidades educativas especiais;
- Disciplina: nome da disciplina;
- Estado de Avaliação (Aprovado ou Não Aprovado): nome;
- Ano Letivo: ano letivo e triénio.

Os valores a considerar para análise são as notas dos alunos e matrículas dos alunos. Em relação às notas podem ser calculadas médias ou podem ser feitas contagens. Em relação às matrículas dos alunos são feitas contagens. Esses valores estão descritos em 3.3.1.

## 4. Modelo Multidimensional

Com base nos requisitos analisados anteriormente foi feita a modelação multidimensional para responder a esses requisitos.

### 4.1. Matriz

A Tabela 1 apresenta uma matriz onde se pode observar para cada tabela de fatos as dimensões que tem [1]. As dimensões e as tabelas de fatos foram modeladas para responder aos requisitos analisados anteriormente.

Tabela 1 - Matriz

Tabela de Fatos Dimensão	Matrícula	Notas do 5º Ano ao 9º Ano	Notas do 10º Ano ao 12º Ano
Ano Letivo	X	X	X
Ano Escolar	X	X	X
Escola	X	X	X
Características do Aluno	X	X	X
Contexto Familiar	X	X	X
Estado de Avaliação	X		
Disciplina		X	X

Como se pode observar na Tabela 1, as tabelas de fatos partilham dimensões. No caso da tabela de fatos *Matrícula* não existe ligação com a dimensão *Disciplina* uma vez que nessa tabela de fatos são feitas contagens de matrículas dos alunos não sendo necessário a dimensão *Disciplina*. No caso das tabelas de fatos *Notas do 5º Ano ao 9º Ano* e *Notas do 10º Ano ao 12º Ano* não existe ligação para a dimensão *Estado de Avaliação*. A dimensão *Estado de Avaliação* permite saber os alunos que estão aprovados ou não aprovados em cada ano letivo. Como nas tabelas de fatos *Notas do 5º Ano ao 9º Ano* e *Notas do 10º Ano ao 12º Ano* são feitos tratamentos de notas de disciplinas não existe ligação com a dimensão *Estado de Avaliação*.

## 4.2. Data Mart

Neste projeto apenas se considera um Data Mart uma vez que as tabelas de fatos partilham dimensões. O Data Mart a desenvolver tem como objetivo analisar as componentes que influenciam o desempenho escolar dos alunos nas escolas. A contagem de matrículas é feita na tabela de fatos *Matrícula*. Para analisar as notas foi necessário considerar duas tabelas de fatos uma vez que do 5º Ano ao 9º Ano as notas vão de 1 a 5 e do 10º Ano ao 12º Ano as notas vão de 0 a 20. Para além disso do 5º Ano ao 9º Ano um aluno pode ter 3 notas a uma determinada disciplina (*CF, Exame e CFD*), enquanto do 10º Ano ao 12º Ano um aluno pode ter 4 notas a uma determinada disciplina (*CF, CIF, Exame e CFD*). O Data Mart é atualizado no final de cada ano letivo após a Empresa receber os dados das escolas.

## 4.3. Granularidade

A granularidade na tabela de fatos *Matrícula* corresponde a uma matrícula de um aluno num ano letivo. A granularidade das tabelas de fatos *Notas do 5º Ano ao 9º Ano* e *Notas do 10º Ano ao 12º Ano* é de um registo com notas de um aluno a uma disciplina num ano letivo.

## 4.4. Dimensões

As dimensões foram criadas segundo os requisitos analisados em 3.3.1. As dimensões criadas são as seguintes:

- Ano Letivo: permite guardar os anos letivos;
- Ano Escolar: permite guardar os anos escolares;
- Escola: permite guardar informação das escolas;
- Características do Aluno: permite guardar informação sobre as características dos alunos;
- Contexto Familiar: permite guardar habilitações e áreas de profissão dos pais;
- Estado de Avaliação: permite guardar os estados de avaliação (Aprovado ou não aprovado);
- Disciplina: permite guardar as disciplinas;

Todas as dimensões têm uma surrogate key (SK) e uma alternate key (AK). A AK permite identificar de forma única um registo na dimensão e a SK é um atributo artificial que substitui a AK [1] [22]. Todas as relações entre uma tabela de fatos e as suas dimensões é feita através da SK de

cada dimensão. As dimensões que só têm um atributo poderiam ser guardadas nas tabelas de fatos mas foi decidido criar dimensões para que o modelo seja mais legível e para dar possibilidade ao crescimento das dimensões através da inserção de novos atributos.

Para cada dimensão foi definida uma política de Slowly Changing Dimensions (SCD), o que permite para cada dimensão decidir se é guardado histórico. Os tipos de SCD são os seguintes [1] [16]:

- Tipo 1: O tipo 1 permite substituir o valor de um atributo para um novo valor. Este tipo de SCD permite corrigir erros em valores de um determinado atributo da dimensão;
- Tipo 2: O tipo 2 permite guardar versões de registos na dimensão. Uma mudança é capturada inserindo um novo registo com a atualização do valor do atributo. Quando ocorre este tipo de mudança os valores anteriores são guardados na dimensão o que permite ter um histórico nessa dimensão;
- Tipo 3: O tipo 3 permite guardar para cada atributo desse tipo os dois valores mais recentes. Para isso cada atributo desse tipo tem duas colunas uma para guardar o valor mais recente e outra para guardar o valor anterior.

#### 4.4.1. Dimensão *Ano Letivo*

A dimensão *Ano Letivo* permite guardar informação temporal dos dados. Os dados podem ser analisados de forma anual ou por triénio. Cada ano letivo é definido pelo ano e pelo triénio. Na Tabela 2 está a descrição dos atributos da dimensão *Ano Letivo* com a política SCD.

Tabela 2 - Dimensão *Ano Letivo*

Atributo	Descrição	Tipo	Chave	Domínio	Exemplo	SCD
ANL_id	Identificador único de um registo	INT	SK	-	1	-
ANL_ano	Ano letivo 2012 = 2012/2013	INT	AK	>0	2012	1
ANL_trienio	Faixa de 3 anos letivos	VARCHAR (9)	-	-	2010-2012	1

Todos os atributos têm SCD igual a 1 uma vez que uma alteração de valor nesses atributos acontece em caso de erro no valor. Foi definida a seguinte hierarquia para a dimensão *Ano Letivo*:

- H1:ANL\_ano -> ANL\_trienio

A hierarquia permite fazer análises de dados por triénio. Caso se pretenda obter maior detalhe dos dados num determinado triénio podem ser analisados os anos letivos desse triénio. Na Figura 7 está o diagrama da dimensão *Ano Letivo*.



Figura 7 – Diagrama da dimensão *Ano Letivo*

#### 4.4.2. Dimensão *Ano Escolar*

A dimensão *Ano Escolar* permite guardar os anos escolares e os ciclos escolares. Cada ano escolar é definido pelo ano e pelo ciclo escolar. Na Tabela 3 está a descrição dos atributos da dimensão *Ano Escolar* com a política SCD.

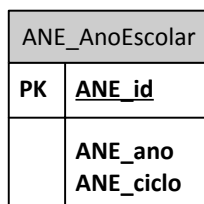
Tabela 3 – Dimensão *Ano Escolar*

Atributo	Descrição	Tipo	Chave	Domínio	Exemplo	SCD
ANE_id	Identificador único de um registo	INT	SK	-	1	-
ANE_ano	Ano escolar 5 = 5º Ano	INT	AK	>=5 <=12	5	1
ANE_ciclo	Ciclo escolar	VARCHAR (10)	-	-	Secundário	1

Todos os atributos têm SCD igual a 1 uma vez que uma alteração de valor nesses atributos acontece em caso de erro no valor. Foi definida a seguinte hierarquia para a dimensão *Ano Escolar*:

- H1: ANE\_ano -> ANE\_ciclo

A hierarquia permite fazer análises de dados por ciclos escolares. Caso se pretenda obter maior detalhe dos dados num determinado ciclo escolar podem ser analisados os anos escolares desse ciclo. Na Figura 8 está o diagrama da dimensão *Ano Escolar*.


 Figura 8 – Diagrama da dimensão *Ano Escolar*

#### 4.4.3. Dimensão *Escola*

A dimensão *Escola* permite guardar informações para cada escola. Cada escola é definida pelo nome, agrupamento escolar, tipo de escola, distrito, conselho e freguesia. Na Tabela 4 está a descrição dos atributos da dimensão *Escola* com a política SCD.

 Tabela 4 - Dimensão *Escola*

Atributo	Descrição	Tipo	Chave	Domínio	Exemplo	SCD
ESC_id	Identificador único de um registo	INT	SK	-	1	-
ESC_escola	Nome da escola	VARCHAR(200)	AK	-	Escola Básica ...	1
ESC_agrupamento	Agrupamento escolar	VARCHAR(200)	-	-	Agrupamento de Escolas ...	2
ESC_tipo	Tipo da escola	VARCHAR(7)	-	-	Público	2
ESC_freguesia	Freguesia da escola	VARCHAR(100)	-	-	Calvão	2
ESC_concelho	Concelho da escola	VARCHAR(100)	-	-	Vagos	2
ESC_distrito	Distrito da escola	VARCHAR(100)	-	-	Aveiro	1
ESC_validoDe	Início da validade do registo	INT	AK		2004	-
ESC_validoAte	Fim da validade do registo	INT			2007	-

A AK da dimensão *Escola* é formada pelos atributos *ESC\_escola* e *ESC\_validoDe* para identificar de forma única um registo na dimensão *Escola*. Os atributos *ESC\_validoDe* e *ESC\_validoAte* permitem verificar na dimensão *Escola* quando ocorreram alterações em atributos com SCD igual a 2 para uma escola [16]. Esses atributos permitem guardar entre que anos letivos um determinado registo de uma escola é válido. Quando ocorrem alterações em atributos do tipo 2 numa escola, para além de ser inserido um novo registo é necessário atualizar o atributo *ESC\_validoAte* do registo anterior dessa escola. O registo mais recente de uma escola tem sempre o valor 9999 no atributo *ESC\_validoAte*.

Os atributos com SCD igual a 1 são *ESC\_escola* e *ESC\_distrito* uma vez que uma alteração de valor nesses atributos acontece em caso de erro no valor. Os atributos com SCD igual a 2 são *ESC\_tipo*, *ESC\_agrupamento*, *ESC\_freguesia*, *ESC\_concelho*. Em relação ao atributo *ESC\_tipo* foi considerado que o tipo de uma escola pode mudar e para as análises têm de se guardar o contexto histórico da escola que muda de tipo, de maneira a obter resultados corretos quando se fazem análises pelo tipo de escola. O mesmo se passa para o atributo *ESC\_agrupamento*. Em relação aos atributos *ESC\_freguesia* e *ESC\_concelho* normalmente deveriam ser de tipo 1, mas ultimamente têm havido redistribuição de freguesias e concelhos em algumas zonas de Portugal, sendo por isso que se consideram esses atributos do tipo 2 para guardar o contexto histórico. Foram definidas as seguintes hierarquias para a dimensão *Escola*:

- H1: *ESC\_escola* -> *ESC\_tipo*;
- H2: *ESC\_escola* -> *ESC\_agrupamento*;
- H3: *ESC\_escola* -> *ESC\_freguesia* -> *ESC\_concelho* -> *ESC\_distrito*.

A hierarquia H1 permite fazer análises de dados por tipo de escola. Caso se pretenda obter maior detalhe dos dados num determinado tipo de escola podem ser analisados as escolas com esse tipo. A hierarquia H2 permite fazer análises de dados por agrupamento escolar. Caso se pretenda obter maior detalhe dos dados num determinado agrupamento escolar podem ser analisados as escolas com esse agrupamento. A hierarquia H3 permite fazer análises de dados por localização. Podem ser feitas análises de dados de escolas por distrito. Caso se pretenda obter maior detalhe dos dados num determinado distrito podem ser feitas análises de dados de escolas por cada concelho desse distrito. Para obter maior detalhe dos dados num determinado concelho podem ser feitas análises de dados de escolas por cada freguesia desse concelho. Para obter maior detalhe dos dados numa determinada freguesia podem ser feitas análises de dados das escolas nessa freguesia. Na Figura 9 está o diagrama da dimensão *Escola*.



ESC_Escola	
PK	ESC_id
	ESC_escola ESC_agrupamento ESC_tipo ESC_freguesia ESC_concelho ESC_distrito ESC_validoDe ESC_validoAte

 Figura 9 - Diagrama da dimensão *Escola*

#### 4.4.4. Dimensão *Características do Aluno*

A dimensão *Características do Aluno* permite guardar informação para cada combinação de características que um aluno pode ter. Na Tabela 5 está a descrição dos atributos da dimensão *Características do Aluno* com a política SCD.

 Tabela 5 - Dimensão *Características do Aluno*

Atributo	Descrição	Tipo	Chave	Domínio	Exemplo	SCD
CAL_id	Identificador único de um registo	INT	SK	-	1	-
CAL_sexo	Sexo do aluno	VARCHAR(12)	AK	-	Masculino	1
CAL_idade	Idade do aluno	INT		>0	15	1
CAL_escalao	Escalão social do aluno SE = Sem escalão	VARCHAR(2)		-	A	1
CAL_nee	Necessidades educativas especiais 1 = O aluno tem necessidades educativas especiais	BIT		0 ou 1	0	1

Cada combinação é definida pelo sexo do aluno, idade do aluno, escalão social do aluno e se o aluno tem ou não necessidades educativas especiais. Por exemplo para um aluno de 15 anos do

sexo masculino com escalão A e sem necessidades educativas especiais o registo na dimensão que caracteriza esse aluno guarda o sexo igual a masculino, a idade igual a 15, o escalão igual a A e necessidades educativas especiais igual a 0 porque não tem. Se houver outro aluno com as mesmas características é identificado pelo mesmo registo. A AK desta dimensão é formada por todos os atributos exceto o *CAL\_id*. Nesta dimensão não são guardados registos por cada um dos alunos como acontece por exemplo para as escolas na dimensão *Escola*. Esta dimensão foi modelada desta forma pois para as análises não interessa saber quem é o aluno mas apenas as características dele. Como esta dimensão guarda por cada registo a combinação das características que um aluno pode ter, todos os atributos têm SCD igual a 1 uma vez que uma alteração de valor nesses atributos acontece em caso de erro no valor. Esta dimensão não tem hierarquias uma vez que as análises são feitas pelos atributos. Na Figura 10 está o diagrama da dimensão *Características do Aluno*.

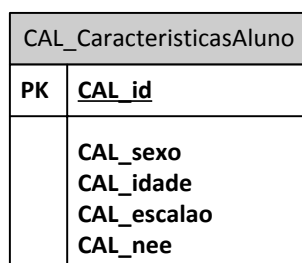


Figura 10 - Diagrama da dimensão *Características do Aluno*

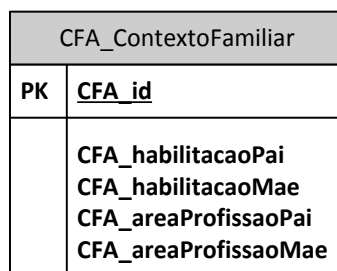
#### 4.4.5. Dimensão *Contexto Familiar*

A dimensão *Contexto Familiar* permite guardar informação para cada combinação de contextos que os pais podem ter. Cada combinação é definida pela habilitação do pai, habilitação da mãe, área de profissão do pai e área de profissão da mãe. A dimensão *Contexto Familiar* não guarda registos por cada um dos pais uma vez que para as análises não interessa saber quem é o pai e a mãe mas apenas os valores de contexto deles. Como esta dimensão guarda por cada registo a combinação de contextos dos pais, todos os atributos têm SCD igual a 1 uma vez que uma alteração de valor nesses atributos acontece em caso de erro no valor. Na Tabela 6 está a descrição dos atributos da dimensão *Contexto Familiar* com a política SCD.

**Tabela 6 - Dimensão Contexto Familiar**

Atributo	Descrição	Tipo	Chave	Domínio	Exemplo	SCD
CFA_id	Identificador único de um registo	INT	SK	-	1	-
CFA_habilitacaoPai	Habilitação do pai	VARCHAR(21)	AK	-	Ensino Superior	1
CFA_habilitacaoMae	Habilitação da mãe	VARCHAR(21)		-	Secundário e 3º Ciclo	1
CFA_areaProfissaoPai	Área de profissão do pai	VARCHAR(31)		-	Quadros Intermédios	1
CFA_areaProfissaoMae	Área de profissão da mãe	VARCHAR(31)		-	Quadros Intermédios	1

Esta dimensão não tem hierarquias uma vez que as análises são feitas pelos atributos. Na Figura 11 está o diagrama da dimensão *Contexto Familiar*.



**Figura 11 - Diagrama da dimensão Contexto Familiar**

Os atributos *CFA\_habilitacaoPai* e *CFA\_habilitacaoMae* podem ter os seguintes valores:

- Ensino Superior;
- Secundário e 3º Ciclo;
- Básico (1º e 2º ciclo);
- Sem Habilitações;
- Desconhecido.

Os atributos *CFA\_areaProfissaoPai* e *CFA\_areaProfissaoMae* podem ter os seguintes valores:

- Quadros Dirigentes e Superiores;
- Quadros Intermédios;
- Trabalhadores Qualificados;
- Trabalhadores Não Qualificados;
- Desconhecido.

#### 4.4.6. Dimensão *Estado de Avaliação*

A dimensão *Estado de Avaliação* permite guardar os estados de avaliação que um aluno pode ter. Cada estado de avaliação é definido pelo nome. Como os estados de avaliação são aprovado e não aprovado a dimensão só tem dois registos. Na Tabela 7 está a descrição dos atributos da dimensão *Estado de Avaliação* com a política SCD.

Tabela 7 - Dimensão *Estado de Avaliação*

Atributo	Descrição	Tipo	Chave	Domínio	Exemplo	SCD
ESA_id	Identificador único de um registo	INT	SK	-	1	-
ESA_nome	Nome do estado de avaliação	VARCHAR(12)	AK	-	Aprovado	1

O atributo *ESA\_nome* tem SCD igual a 1, uma vez que uma alteração de valor nesse atributo acontece em caso de erro no valor. As análises com esta dimensão são feitas utilizando o atributo *ESA\_nome*. Na Figura 12 está o diagrama da dimensão *Estado de Avaliação*.

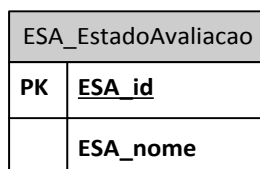


Figura 12 - Diagrama da Dimensão *Estado de Avaliação*

#### 4.4.7. Dimensão *Disciplina*

A dimensão *Disciplina* permite guardar as disciplinas. Cada disciplina é definida pelo nome. Na Tabela 8 está a descrição dos atributos da dimensão *Disciplina* com a política SCD. O atributo

*DIS\_nome* tem SCD igual a 1, uma vez que uma alteração de valor nesse atributo acontece em caso de erro no valor.

Tabela 8 - Dimensão *Disciplina*

Atributo	Descrição	Tipo	Chave	Domínio	Exemplo	SCD
DIS_id	Identificador único de um registo	INT	SK	-	1	-
DIS_nome	Nome da disciplina	VARCHAR(100)	AK	-	Matemática	1

As análises com esta dimensão são feitas utilizando o atributo *DIS\_nome*. Na Figura 13 está o diagrama da dimensão *Disciplina*.

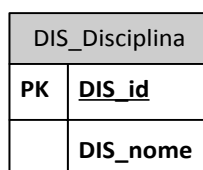


Figura 13 - Diagrama da dimensão *Disciplina*

## 4.5. Tabelas de Fatos

Foram criadas as seguintes tabelas de fatos com base nos requisitos definidos anteriormente:

- Matrícula: tabela de fatos que permite fazer contagens de matrículas de alunos;
- Notas do 5º Ano ao 9º Ano: tabela de fatos que permite guardar as notas dos alunos entre o 5º Ano e o 9º Ano de escolaridade;
- Notas do 10º Ano ao 12º Ano: tabela de fatos que permite guardar as notas dos alunos entre o 10º Ano e o 12º Ano de escolaridade.

Como já foi referido anteriormente as tabelas de fatos partilham dimensões.

### 4.5.1. Tabela de Fatos *Matrícula*

A tabela de fatos *Matrícula* tem como atributos:

- 6 Atributos (PK/FK) correspondentes a cada uma das dimensões.

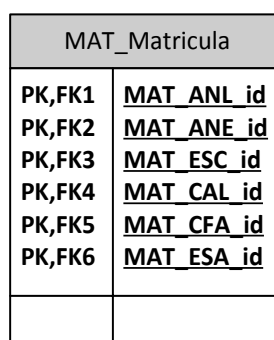
A tabela de fatos *Matrícula* não tem medidas uma vez que apenas são feitas contagens. Na

Tabela 9 está a descrição da tabela de fatos *Matrícula*.

**Tabela 9 - Tabela de Fatos *Matrícula***

Atributo	Descrição	Tipo	Chave	Domínio	Exemplo
MAT_ANL_id	Identificação de uma linha na dimensão <i>Ano Letivo</i>	INT	PK	FK1 >0	1
MAT_ANE_id	Identificação de uma linha na dimensão <i>Ano Escolar</i>	INT		FK2 >0	1
MAT_ESC_id	Identificação de uma linha na dimensão <i>Escola</i>	INT		FK3 >0	1
MAT_CAL_id	Identificação de uma linha na dimensão <i>Características do Aluno</i>	INT		FK4 >0	1
MAT_CFA_id	Identificação de uma linha na dimensão <i>Contexto Familiar</i>	INT		FK5 >0	1
MAT_ESA_id	Identificação de uma linha na dimensão <i>Estado de Avaliação</i>	INT		FK6 >0	1

Na Figura 14 está o diagrama da tabela de fatos *Matrícula*.



**Figura 14 - Diagrama da Tabela de Fatos *Matrícula***

#### 4.5.2. Tabela de Fatos *Notas do 5º Ano ao 9º Ano*

A tabela de fatos *Notas do 5º Ano ao 9º Ano* tem como atributos:

- 6 Atributos (PK/FK) correspondentes a cada uma das dimensões.
- 3 Medidas associadas aos requisitos analisados anteriormente (3.3.1):

- *CF* – classificação interna anual a uma disciplina: medida aditiva;
- *Exame* – classificação de exame a uma disciplina: medida aditiva;
- *CFD* – classificação final anual a uma disciplina: medida aditiva.

Na Tabela 10 está a descrição da tabela de fatos *Notas do 5º Ano ao 9º Ano*.

**Tabela 10 - Tabela de Fatos de *Notas do 5º Ano ao 9º Ano***

Atributo	Descrição	Tipo	Chave	Domínio	Exemplo	
NT1_ANL_id	Identificação de uma linha na dimensão <i>Ano Letivo</i>	INT	PK	FK1	>0	1
NT1_ANE_id	Identificação de uma linha na dimensão <i>Ano Escolar</i>	INT		FK2	>0	1
NT1_ESC_id	Identificação de uma linha na dimensão <i>Escola</i>	INT		FK3	>0	1
NT1_CAL_id	Identificação de uma linha na dimensão <i>Características do Aluno</i>	INT		FK4	>0	1
NT1_CFA_id	Identificação de uma linha na dimensão <i>Contexto Familiar</i>	INT		FK5	>0	1
NT1_DIS_id	Identificação de uma linha na dimensão <i>Disciplina</i>	INT		FK6	>0	1
NT1_CF	Classificação interna anual a uma disciplina	INT	-	1 a 5	3	
NT1_Exame	Classificação de exame a uma disciplina	INT	-	1 a 5	4	
NT1_CFD	Classificação final anual a uma disciplina	INT	-	1 a 5	3	

Na Figura 15 está o diagrama da tabela de fatos *Notas do 5º Ano ao 9º Ano*.

NT1_Nota5_9	
PK,FK1	<u>NT1_ANL_id</u>
PK,FK2	<u>NT1_ANE_id</u>
PK,FK3	<u>NT1_ESC_id</u>
PK,FK4	<u>NT1_CAL_id</u>
PK,FK5	<u>NT1_CFA_id</u>
PK,FK6	<u>NT1_DIS_id</u>
	NT1_CF NT1_Exame NT1_CFD

Figura 15 – Diagrama da Tabela de Fatos *Notas do 5º Ano ao 9º Ano*

#### 4.5.3. Tabela de Fatos *Notas do 10º Ano ao 12º Ano*

A tabela de fatos *Notas do 10º Ano ao 12º Ano* tem como atributos:

- 6 Atributos (PK/FK) correspondentes a cada uma das dimensões.
- 3 Medidas associadas aos requisitos analisados anteriormente (3.3.1):
  - *CF* – classificação interna anual de uma disciplina: medida aditiva;
  - *CIF* – classificação média interna a uma disciplina no ano escolar de conclusão: medida aditiva;
  - *Exame* – classificação de exame a uma disciplina: medida aditiva;
  - *CFD* – classificação média final a uma disciplina no ano escolar de conclusão: medida aditiva.

Na Tabela 11 está a descrição da tabela de fatos *Notas do 10º Ano ao 12º Ano*.

Tabela 11 - Tabela de Fatos de *Notas do 10º Ano ao 12º Ano*

Atributo	Descrição	Tipo	Chave	Domínio	Exemplo	
NT2_ANL_id	Identificação de uma linha na dimensão <i>Ano Letivo</i>	INT	PK FK1	>0	1	
NT2_ANE_id	Identificação de uma linha na dimensão <i>Ano Escolar</i>	INT		FK2	>0	1
NT2_ESC_id	Identificação de uma linha na dimensão <i>Escola</i>	INT		FK3	>0	1
NT2_CAL_id	Identificação de uma linha na dimensão <i>Características do Aluno</i>	INT		FK4	>0	1



NT2_CFA_id	Identificação de uma linha na dimensão <i>Contexto Familiar</i>	INT		FK5	>0	1
NT2_DIS_id	Identificação de uma linha na dimensão <i>Disciplina</i>	INT		FK6	>0	1
NT2_CF	Classificação interna anual de uma disciplina	INT	-	-	0 a 20	14
NT2_CIF	Classificação média interna a uma disciplina no ano escolar de conclusão	INT	-	-	0 a 20	12
NT2_Exame	Classificação de exame a uma disciplina	INT	-	-	0 a 20	12
NT2_CFD	Classificação média final a uma disciplina no ano escolar de conclusão	INT	-	-	0 a 20	14

Na Figura 16 está o diagrama da tabela de fatos *Notas do 10º Ano ao 12º Ano*.

NT2_Nota10_12	
PK,FK1	<u>NT2_ANL_id</u>
PK,FK2	<u>NT2_ANE_id</u>
PK,FK3	<u>NT2_ESC_id</u>
PK,FK4	<u>NT2_CAL_id</u>
PK,FK5	<u>NT2_CFA_id</u>
PK,FK6	<u>NT2_DIS_id</u>
	NT2_CF NT2_CIF NT2_Exame NT2_CFD

Figura 16 – Diagrama da Tabela de Fatos *Notas do 10º Ano ao 12º Ano*

#### 4.6. Revisão e Aprovação

Todos os esquemas desenvolvidos até agora devem ser revistos e aprovados pela Empresa.

### 4.7. Modelo Logico do Data Mart

Nas figuras seguintes podemos ver o modelo lógico do Data Mart. Para analisar o modelo de forma mais fácil, este está organizado pelas tabelas de fato. Na Figura 17 está a estrela com a tabela de fatos *Matrícula*, na Figura 18 está a estrela com a tabela de fatos *Notas do 5º Ano ao 9º Ano* e na Figura 19 está a estrela com as tabelas de fatos *Notas do 10º Ano ao 12º Ano*. As dimensões com o mesmo nome são partilhadas pelas tabelas de fatos.

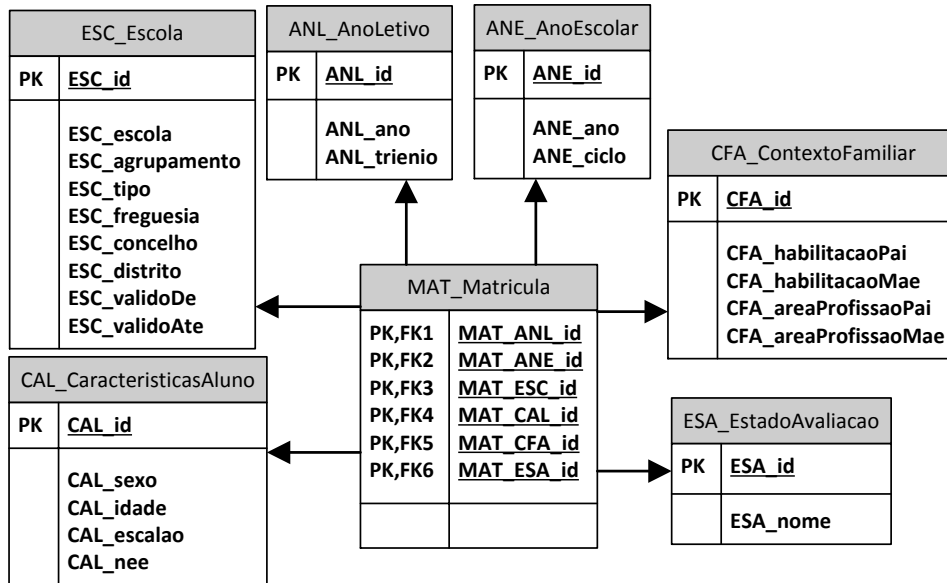


Figura 17 - Estrela *Matrícula*

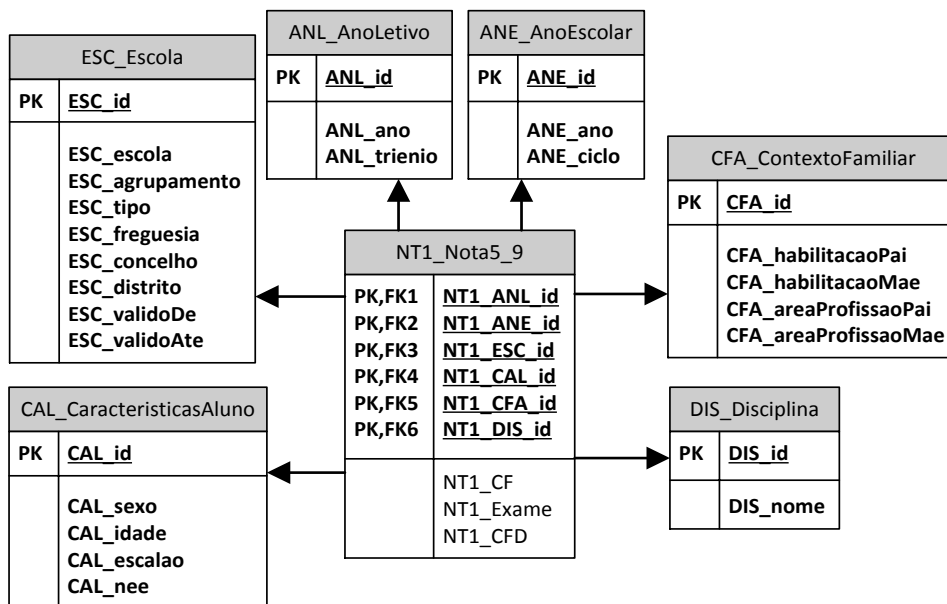


Figura 18 - Estrela *Notas do 5º Ano ao 9º Ano*

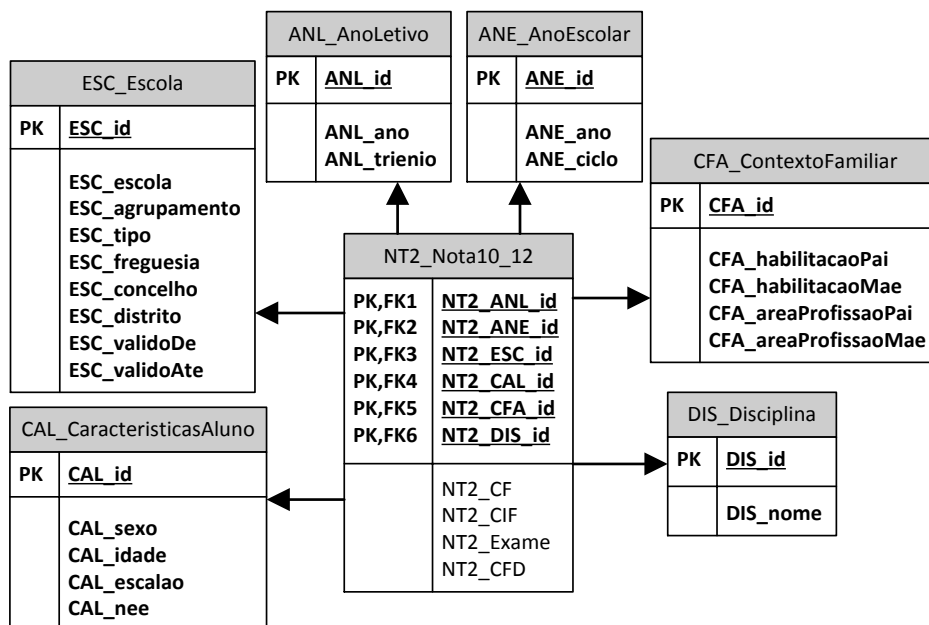


Figura 19 - Estrela Notas do 10º Ano ao 12º Ano

#### 4.8. Análise das Fontes de Dados

O Data Mart descrito anteriormente deve ser preenchido com dados provenientes das fontes. São conhecidas duas fontes de dados, uma por cada escola. Cada escola representa um agrupamento escolar por isso cada fonte tem dados das escolas do agrupamento. Ambas as fontes de dados estão em ficheiros Excel, mas esses ficheiros têm um esquema diferente em cada fonte. Para distinguir as fontes de dados das escolas vou nomeá-las *Fonte 1* e *Fonte 2*. Os ficheiros da *Fonte 1* e *Fonte 2* podem ser organizados por categorias. Na *Fonte 1* as categorias dos ficheiros que vou descrever têm o prefixo F1 e na *Fonte 2* as categorias que vou descrever têm o prefixo F2. Os ficheiros da *Fonte 1* estão organizados nas seguintes categorias:

- F1\_Alunos – Ficheiros que têm para cada aluno, o número de processo, o sexo, a data de nascimento e se o aluno tem necessidades educativas especiais. Os ficheiros têm também o agrupamento escolar. Existe um ficheiro deste tipo por cada ano letivo;
- F1\_Pais – Ficheiros que têm para cada aluno, o número de processo, a habilitação do pai, a habilitação da mãe, a profissão do pai e a profissão da mãe. Os ficheiros têm também o agrupamento escolar. Existe um ficheiro deste tipo por cada ano letivo;
- F1\_Escalões – Existem dois ficheiros desta categoria, um para o escalão A e outro para o escalão B. Cada um dos ficheiros tem o número de processo de cada aluno com esse

escalão. Os ficheiros têm também o agrupamento escolar, o escalão, o ano escolar e o ano letivo. Existe um ficheiro deste tipo por cada ano escolar de cada ano letivo;

- F1\_Turmas – Ficheiros que têm para cada aluno, o número de processo, as notas que tem a cada disciplina e o estado de avaliação. Os ficheiros têm também o agrupamento escolar, o nome da escola, o ano letivo e o ano escolar. Existe um ficheiro deste tipo por cada turma do 5º ano ao 12º ano, em cada ano letivo. Nos ficheiros do 5º ano ao 9º ano existem três notas a obter que são *CF*, *Exame* e *CFD*. As notas vão de 1 a 5. Nos ficheiros do 10º ano ao 12º ano existem quatro notas a obter que são *CF*, *CIF*, *Exame* e *CFD*. As notas vão de 0 a 20.

No Anexo I estão ficheiros exemplo para cada uma das categorias de ficheiros da *Fonte 1*.

Os ficheiros da *Fonte 2* estão organizados nas seguintes categorias:

- F2\_Alunos – Ficheiros que têm para cada aluno, o número de processo, o sexo, a data de nascimento, se o aluno tem necessidades educativas especiais, habilitação do pai, habilitação da mãe, profissão do pai e profissão da mãe. Os ficheiros têm também o agrupamento escolar, o ano escolar e o ano letivo. Existe um ficheiro deste tipo por cada turma do 5º ano ao 12º ano, em cada ano letivo;
- F2\_Escalões – Existem dois ficheiros desta categoria, um para o escalão A e outro para o escalão B. Cada um dos ficheiros tem o número de processo de cada aluno com esse escalão. Os ficheiros têm também o agrupamento escolar, o escalão, o ano escolar e o ano letivo. Existe um ficheiro deste tipo por cada ano escolar de cada ano letivo;
- F2\_Turmas – Ficheiros que têm para cada aluno, o número de processo, as notas que tem a cada disciplina e o estado de avaliação do aluno. Os ficheiros têm também o agrupamento escolar, o ano letivo e o ano escolar. O nome da escola não existe nos ficheiros de dados. Os ficheiros das turmas na Fonte 2 estão organizados por escola e com base nisso consegue-se obter o nome da escola. Existe um ficheiro deste tipo por cada turma do 5º ano ao 12º ano, em cada ano letivo. Nos ficheiros do 5º ano ao 9º ano existem três notas a obter que são *CF*, *Exame* e *CFD*. As notas vão de 1 a 5. Nos ficheiros do 10º ano ao 12º ano existem quatro notas a obter que são *CF*, *CIF*, *Exame* e *CFD*. As notas vão de 0 a 20.

No Anexo II estão ficheiros exemplo para cada uma das categorias de ficheiros da *Fonte 2*.

As fontes de dados foram analisadas tendo em consideração os seguintes pontos:

- Disponibilidade da informação para preencher o Data Mart;

- Qualidade da informação existente nas duas fontes;
- Mapeamento dos dados das fontes para o Data Mart.

#### 4.8.1. Disponibilidade da Informação

Depois de analisar as fontes de dados foi possível verificar que se pode obter a informação quase toda para preencher o Data Mart. A informação poderá ser obtida de forma direta, indireta a partir de informação existente, ou com a intervenção de um operador. A seguir estão os processos de obtenção de informação.

##### **P1: Informação sobre anos letivos e triénios**

A informação sobre os triénios não existe nas fontes de dados. Para cada ano letivo presente nas categorias de ficheiros *F1\_Turmas* e *F2\_Turmas* tem de ser definido o triénio ao qual pertence.

##### **P2: Informação sobre os anos escolares e ciclos escolares**

A informação sobre os ciclos escolares não existe de forma direta nas fontes de dados. O ciclo escolar é obtido através do ano escolar uma vez que um ano escolar pertence apenas a um ciclo escolar. Deve existir a intervenção de um operador para atribuir a cada ano escolar, o ciclo a que pertence.

##### **P3: Informação sobre as escolas**

Na *Fonte 1*, a informação do nome das escolas e do agrupamento escolar é obtida a partir dos ficheiros da categoria *F1\_Turmas*. Na *Fonte 2*, a informação do agrupamento escolar é obtida a partir dos ficheiros da categoria *F2\_Turmas*. Em relação ao nome das escolas a informação não pode ser obtida de forma direta nos ficheiros da categoria *F2\_turmas*. Como já foi referido anteriormente, os ficheiros da categoria *F2\_Turmas* estão organizados por escola e com base nisso consegue-se obter o nome da escola. A partir das fontes é possível obter o nome de cada escola e o agrupamento escolar respetivo. Os dados sobre o tipo de escola, freguesia, concelho e distrito não existe nos ficheiros das fontes. Tem de existir a intervenção de um operador que pesquise os dados em falta com base no nome de cada escola, para completar essa informação.

##### **P4: Informação sobre os alunos**

A informação de cada aluno é obtida a partir dos ficheiros da categoria *F1\_Alunos* e *F2\_Alunos*. Sempre que o sexo ou a data de nascimento forem desconhecidos, o sexo fica com o valor “Desconhecido” e a data de nascimento fica com o valor “0001-01-01”. Sempre que o valor para

as necessidades educativas especiais é desconhecido, assume-se que o aluno não tem necessidades educativas especiais e o valor para esse atributo é 0.

#### **P5: Informação da idade do aluno**

As fontes de dados não disponibilizam a idade dos alunos. No processo P4 é possível obter a data de nascimento de cada aluno. Deve ser implementado um mecanismo que calcula a idade do aluno a partir da data de nascimento.

#### **P6: Informação do escalão social dos alunos**

A informação dos alunos com escalão social é obtida a partir dos ficheiros da categoria *F1\_Escalões* e *F2\_Escalões*.

#### **P7: Informação sobre os Pais**

A informação dos pais de cada aluno é obtida a partir dos ficheiros da categoria *F1\_Pais* e *F2\_Alunos*. Se existirem valores a null para as habilitações ou profissões, esses valores ficam com o valor “Desconhecido”.

#### **P8: Informação sobre a categoria da habilitação do pai e da mãe**

As fontes de dados não disponibilizam informação sobre a categoria da habilitação do pai e da mãe. Através do processo P7 é possível obter a habilitação do pai e a habilitação da mãe. Para obter a categoria da habilitação do pai e da mãe, é necessário implementar um programa que atribua uma categoria da habilitação a cada habilitação.

#### **P9: Informação sobre a área de profissão do pai e da mãe**

As fontes de dados não disponibilizam informação sobre a área de profissão do pai e da mãe. Através do processo P7 é possível obter a profissão do pai e a profissão da mãe. Para obter a área de profissão do pai e da mãe, é necessário implementar um programa que determine a área de profissão a partir do nome da profissão. Para as profissões que não foi possível atribuir uma área de profissão, deve existir a intervenção de um operador para atribuir a área de profissão.

#### **P10: Informação sobre notas do 5º Ano ao 9º Ano**

A informação das notas de cada aluno do 5º ano ao 9º ano é obtida a partir dos ficheiros da categoria *F1\_Turmas* e *F2\_Turmas*. Do 5º ano ao 9º ano os alunos podem ter três notas a uma disciplina que são *CF*, *Exame* e *CFD*. As notas entre o 5º ano e o 9º ano vão de 1 a 5. Nos ficheiros

da categoria *F2\_Turmas* não se pode obter de forma direta notas *CFD* em disciplinas sem exame. Nessas disciplinas o *CFD* fica igual a *CF*. Se para um determinado aluno e para uma determinada disciplina todas as notas forem null, o registo deve ser descartado. Para corrigir valores null que apareçam nas notas *CF*, *Exame* e *CFD*, deve ser implementado um programa para obter a nota que está a null a partir das outras notas quando necessário.

### **P11: Informação sobre notas do 10º Ano ao 12º Ano**

A informação das notas de cada aluno do 10º ano ao 12º ano é obtida a partir dos ficheiros das mesmas categorias que no processo P10. A diferença é que as notas do 10º ano ao 12º ano vão de 0 a 20 e podem existir quatro notas por disciplina que são *CF*, *CIF*, *Exame* e *CFD*. Nos ficheiros da categoria *F2\_Turmas* não existem notas *CFD*. As notas *CFD* só têm valor para disciplinas no ano escolar de conclusão. O *CFD* fica igual a *CIF* em disciplinas no ano de conclusão sem exame. Em disciplinas no ano de conclusão com exame o *CFD* dever ser obtido através da seguinte fórmula:

- $CFD = 0,7 \times CIF + 0,3 \times Exame.$

A nota de *Exame* vai de 0 a 200 nos ficheiros e tem de ser convertida para a escala de 0 a 20. Da mesma forma que no processo P10, se para um determinado aluno e para uma determinada disciplina todas as notas forem null, o registo deve ser descartado. Para corrigir valores null que apareçam nas notas *CF*, *CIF*, *Exame* e *CFD*, deve ser implementado um programa para obter a nota que está a null a partir das outras notas quando necessário.

### **P12: Informação sobre estado de avaliação**

O estado de avaliação dos alunos aparecem nos mesmos ficheiros que estão descritos no processo P10. Em alguns casos o estado de avaliação pode aparecer com valor null. Nesses casos deve ser implementado um programa para descobrir qual é o estado de avaliação de cada aluno com base nas suas notas.

### **P13: Informação sobre disciplinas**

As disciplinas aparecem nos mesmos ficheiros que estão descritos no processo P10. A *Fonte 1* tem o nome de cada disciplina escrito por extenso, enquanto na *Fonte 2* o nome de cada disciplina está abreviado. Por exemplo para a disciplina de Matemática, na *Fonte 1* é guardado o valor *Matemática*, enquanto na *Fonte 2* é guardado o nome *MAT*. Tem de ser implementado um mecanismo de maneira que duas disciplinas iguais com nome diferente sejam identificadas de uma única maneira.

#### 4.8.2. Qualidade da Informação das Fontes

A informação existente nas fontes de dados das duas escolas é fiável uma vez que os dados estão corretos. No entanto é necessário integrar os dados das duas fontes. Por exemplo em relação às disciplinas, estas são nomeadas de forma diferente em cada uma das fontes. Esses casos têm de ser tratados atribuindo um nome padrão a cada disciplina.

#### 4.8.3. Mapeamento dos dados das fontes para o Data Mart

A Tabela 12 apresenta o mapeamento dos dados entre as fontes e o Data Mart. Na coluna *Categoria de Ficheiros* da Tabela 12 é especificada a categoria dos ficheiros de onde vem os dados quando aplicável.



Tabela 12 - Mapeamento Logico das Fontes para o Data Mart

	Destino				Origem		Transformação
	Tabela	Atributo	Tipo de Dados	SCD	Categoria de Ficheiros	Tipo de dados	
D I M E N S Õ E S	ANL_AnoLetivo	ANL_id	INT	-	-	-	Surrogate Key
		ANL_ano	INT	1	F1_Turmas, F2_Turmas	INT	Direto
		ANL_trienio	VARCHAR(9)	1	-	-	Ver processo P1
	ANE_AnoEscolar	ANE_id	INT	-	-	-	Surrogate Key
		ANE_ano	INT	1	-	-	Ver processo P2
		ANE_ciclo	VARCHAR(10)	1	-	-	Ver processo P2
	ESC_Escola	ESC_id	INT	-	-	-	Surrogate Key
		ESC_escola	VARCHAR(200)	1	F1_Turmas, F2_Turmas	Text	Direto
		ESC_agrupamento	VARCHAR(200)	2	F1_Turmas, F2_Turmas	Text	Ver processo P3
		ESC_tipo	VARCHAR(7)	2	-	Text	Ver processo P3
		ESC_freguesia	VARCHAR(100)	2	-	Text	Ver processo P3
ESC_concelho		VARCHAR(100)	2	-	Text	Ver processo P3	
ESC_distrito		VARCHAR(100)	1	-	Text	Ver Processo P3	
ESC_validoDe		INT	-	-	-	-	
ESC_validoAte	INT	-	-	-	-		

	Destino				Origem		Transformação
	Tabela	Atributo	Tipo de Dados	SCD	Categoria de Ficheiros	Tipo de dados	
D I M E N S Õ E S	CAL_Caracteristicas Aluno	CAL_id	INT	-	-	-	Surrogate Key
		CAL_sexo	VARCHAR(12)	1	F1_Alunos, F2_Alunos	Text	Ver processo P4
		CAL_nee	BIT	1	F1_Alunos, F2_Alunos	Boolean	Ver processo P4
		CAL_idade	INT	1	F1_Alunos, F2_Alunos	Text	Ver processo P5
		CAL_escalao	VARCHAR(2)	1	F1_Escalões, F2_Escalões	Text	Ver processo P6
	CFA_Contexto Familiar	CFA_id	INT	-	-	-	Surrogate Key
		CFA_habilitacaoPai	VARCHAR(21)	1	F1_Pais, F2_Alunos	Text	Ver processo P8
		CFA_habilitacaoMae	VARCHAR(21)	1	F1_Pais, F2_Alunos	Text	Ver processo P8
		CFA_areaProfissaoPai	VARCHAR(31)	1	F1_Pais, F2_Alunos	Text	Ver processo P9
		CFA_areaProfissaoMae	VARCHAR(31)	1	F1_Pais, F2_Alunos	Text	Ver processo P9
	ESA_Estado Avaliação	ESA_id	INT	-	-	-	Surrogate Key
		ESA_nome	VARCHAR(12)	1	F1_Turmas, F2_Turmas	Text	Ver Processo P12
DIS_Disciplina	DIS_id	INT	-	-	-	Surrogate Key	
	DIS_nome	VARCHAR(100)	1	F1_Turmas, F2_Turmas	Text	Ver processo P13	

	Destino				Origem		Transformação
	Tabela	Atributo	Tipo de Dados	SCD	Categoria de Ficheiros	Tipo de dados	
T A B E L A S D E	MAT_Matricula	MAT_ALN_id	INT	-	-	-	FK para ALN_AnoLetivo
		MAT_ANE_id	INT	-	-	-	FK para ANE_AnoEscolar
		MAT_ESC_id	INT	-	-	-	FK para ESC_Escola
		MAT_CAL_id	INT	-	-	-	FK para CAL_CaracteristicasAluno
		MAT_CFA_id	INT	-	-	-	FK para CFA_Contexto Familiar
		MAT_ESA_id	INT	-	-	-	FK para ESA_Estado Avaliação
F A T O S	NT1_Nota5_9	NT1_ALN_id	INT	-	-	-	FK para ALN_AnoLetivo
		NT1_ANE_id	INT	-	-	-	FK para ANE_AnoEscolar
		NT1_ESC_id	INT	-	-	-	FK para ESC_Escola
		NT1_CAL_id	INT	-	-	-	FK para CAL_CaracteristicasAluno
		NT1_CFA_id	INT	-	-	-	FK para CFA_Contexto Familiar

	Destino				Origem		Transformação
	Tabela	Atributo	Tipo de Dados	SCD	Categoria de Ficheiros	Tipo de dados	
T A B E L A S D E F A T O S	NT1_Nota5_9	NT1_DIS_id	INT	-	-	-	FK para DIS_Disciplina
		NT1_CF	INT	-	F1_Turmas, F2_Turmas	INT	Ver processo P10
		NT1_Exame	INT	-	F1_Turmas, F2_Turmas	INT	Ver processo P10
		NT1_CFD	INT	-	F1_Turmas, F2_Turmas	INT	Ver processo P10
	NT2_Nota10_12	NT2_ALN_id	INT	-	-	-	FK para ALN_AnoLetivo
		NT2_ANE_id	INT	-	-	-	FK para ANE_AnoEscolar
		NT2_ESC_id	INT	-	-	-	FK para ESC_Escola
		NT2_CAL_id	INT	-	-	-	FK para CAL_CaracteristicasAluno
		NT2_CFA_id	INT	-	-	-	FK para CFA_Contexto Familiar
		NT2_DIS_id	INT	-	-	-	FK para DIS_Disciplina
		NT2_CF	INT	-	F1_Turmas, F2_Turmas	INT	Ver processo P11
		NT2_CIF	INT	-	F1_Turmas, F2_Turmas	INT	Ver processo P11
		NT2_Exame	INT	-	F1_Turmas, F2_Turmas	INT	Ver processo P11
		NT2_CFD	INT	-	F1_Turmas, F2_Turmas	INT	Ver processo P11

## 4.9. Revisão e Aprovação do Modelo desenvolvido

Toda a modelação feita deve ser revista detalhadamente por membros nomeados pela Empresa. Só depois da aprovação do modelo desenvolvido se pode avançar para as próximas fases do projeto.

## 4.10. Estratégia de Extração, Transformação e Carregamento dos Dados

Neste capítulo é apresentado a estratégia para a extração, transformação e carregamento dos dados.

### 4.10.1. Extração

Ambas as fontes de dados são constituídas por ficheiros Excel, mas esses ficheiros têm um esquema diferente em cada fonte. Tem de ser definida uma estratégia de extração para cada uma das fontes de dados. Em ambas as fontes existem ficheiros com informação dos alunos, pais dos alunos, notas dos alunos e escalões sociais dos alunos. No ponto 4.8 foram definidas categorias de ficheiros em ambas as fontes. Na *Fonte 1*, foram definidas as categorias de ficheiros *F1\_Alunos*, *F1\_Pais*, *F1\_Escalões* e *F1\_Turmas*. Na *Fonte 2*, foram definidas as categorias de ficheiros *F2\_Alunos*, *F2\_Pais*, *F2\_Escalões* e *F2\_Turmas*. Têm de ser implementados programas para extrair os dados dos ficheiros de cada categoria.

### 4.10.2. Transformação

Os dados das fontes são carregados numa base de dados para serem tratados. A base de dados tem o nome SA (Staging Area). A base de dados SA tem duas sub-bases de dados que são *Escolas* (ver Anexo III) e *SA\_DataMart*. Na base de dados *Escolas* é feito o tratamento de dados provenientes das fontes de dados. A base de dados *SA\_DataMart* tem um esquema de base de dados semelhantes ao do Data Mart e é a partir dessa base de dados que o Data Mart é carregado. Essas bases de Dados vão ser descritas com maior detalhe no capítulo da Implementação.

No processo de transformação dos dados é necessário ter cuidado com os seguintes aspetos:

- Tentar remover alguns valores null que vêm nas fontes de dados. Estes processos estão identificados em 4.8.1;
- Transformação de alguns valores na base de dados *Escolas*, de maneira a que a informação proveniente das fontes fique coerente,

- Por exemplo, para o estado de avaliação dos alunos só se pretende guardar os valores “Aprovado” e “Não Aprovado”, mas nas fontes de dados por vezes aparecem valores “Transitou” e “Não Transitou”. Esses valores têm de ser substituídos por “Aprovado” e “Não Aprovado” respetivamente;
- Obter valores de atributos a partir de outros atributos. Por exemplo a idade do aluno é obtida através da data de nascimento;
- Criação das chaves primárias surrogate keys.

#### 4.10.3. Carregamento

Após a transformação dos dados estar concluída o Data Mart pode ser carregado com os dados. O Data Mart será atualizado com novos dados em cada ano letivo quando as escolas enviarem esses dados à Empresa.

#### 4.11. Estimativa de Dimensionamento

Para fazer uma estimativa de dimensionamento do Data Mart é necessário verificar o número de alunos no ensino básico e secundário em Portugal. A Direção-Geral de Estatísticas da Educação e Ciência fez uma estimativa do número de alunos no ensino básico e secundário para os anos letivos de 2012/2013 até 2018/2019 [23]. Para o número de alunos no ensino básico e secundário em Portugal foram consideradas as estimativas feitas para o ano letivo 2013/2014. Na Tabela 13 estão as estimativas para os anos escolares do 5º ano ao 12º ano.

**Tabela 13 - Estimativa do número de alunos no ensino básico e secundário em Portugal**

	Ensino Básico					Secundário (regular e artístico)		
	2º Ciclo		3º Ciclo					
	5º Ano	6º Ano	7º Ano	8º Ano	9º Ano	10º Ano	11º Ano	12º Ano
Nº de alunos	112380	117968	116420	114744	103603	71146	63935	62100
	230348		334767			197181		
	762296							

O número de disciplinas que um aluno pode ter varia ao longo do percurso escolar do aluno. Por isso é necessário estimar o número de disciplinas lecionadas no ensino básico e no ensino secundário em Portugal. A estimativa foi feita com base em planos curriculares do ensino básico e do ensino secundário [2] [24]. Na Tabela 14 está a estimativa para o número de disciplinas.

**Tabela 14 - Estimativa do número de disciplinas**

	Ensino Básico		Secundário		
	2º Ciclo	3º Ciclo	10º Ano	11º Ano	12º Ano
<b>Nº de Disciplinas</b>	10	12	8		6

Segundo Kimball para fazer uma estimativa de dimensionamento é necessário estimar o número de registos das tabelas de fatos do Data Mart e o número de registos em tabelas de dimensão que possam ter um número muito elevado de registos [1]. A seguir é feita a estimativa de dimensionamento para as tabelas do Data Mart. A Tabela 15 tem o número de registos necessário para cada tabela de fatos.

**Tabela 15 - Número de registos por aluno para cada tabela de fatos**

Tabelas de Fatos	Número de registos por aluno
Mat_Matricula	1 registo por ano letivo
NT1_Nota5_9	10 registos por ano letivo se o aluno estiver no 2º ciclo do ensino básico
	12 registos por ano letivo se o aluno estiver no 3º ciclo do ensino básico
NT2_Nota10_12	8 registos por ano letivo se o aluno estiver no 10º ano ou 11º ano
	6 registos por ano letivo se o aluno estiver no 12º ano

Na tabela *MAT\_Matricula* para cada aluno e para um ano letivo é necessário guardar um registo. Na tabela *NT1\_Nota5\_9* o número de registos varia consoante o número de disciplinas que o aluno tem. Na tabela *NT1\_Nota5\_9* é necessário guardar um registo por cada disciplina que um aluno do 2º ciclo e 3º ciclo do ensino básico tem. Na tabela *NT2\_Nota10\_12* o número de registos também varia consoante o número de disciplinas que um aluno tem. Na tabela *NT2\_Nota10\_12* é necessário guardar um registo por cada disciplina que um aluno do ensino secundário tem. Para

estimar o número de registos em cada tabela de fatos basta saber o número de alunos para o qual se vão guardar dados. Para estimar o número de registos para a tabela *NT1\_Nota5\_9* é necessário saber o número de alunos que estão no 2º ciclo e que estão no 3º ciclo uma vez que o número de disciplinas varia de um ciclo para o outro. Para estimar o número de registos para a tabela de fatos *NT2\_Nota10\_12* é necessário saber o número de alunos que estão no 10º ano, 11º ano e 12º ano uma vez que o número de disciplinas no 12º ano é diferente que no 10º ano e 11º ano. As tabelas que crescem mais em números de registos no Data Mart ao longo dos anos são as tabelas de fatos.

As tabelas das dimensões do Data Mart crescem pouco em termos de número de registos e algumas têm um número de registos fixo. O crescimento do número de registos nas tabelas de dimensão não depende do número de alunos. A Tabela 16 tem o número de registos para cada tabela de dimensão.

**Tabela 16 – Número de registos para cada tabela de dimensão**

<b>Dimensões</b>	<b>Número de registos</b>
ANL_AnoLetivo	1 registo por ano letivo.
ANE_AnoEscolar	8 registos, um para cada ano escolar
ESC_Escola	1 registo por escola. Se para uma escola mudarem valores de atributos de tipo 2 é necessário um registo adicional para essa escola
CAL_CaracteristicasAluno	1 registo por cada nova combinação de valores dos atributos <i>CAL_sexo</i> , <i>CAL_escala</i> , <i>CAL_idade</i> e <i>CAL_nee</i>
CFA_ContextoFamiliar	1 registo por cada nova combinação de valores dos atributos <i>CFA_habilitacaoPai</i> , <i>CFA_habilitacaoMae</i> , <i>CFA_habilitacaoPai</i> e <i>CFA_habilitacaoMae</i>
ESA_EstadoAvaliacao	2 registos. Um para o estado de avaliação “Aprovado” e outro para o estado de avaliação “Não Aprovado”
DIS_Disciplina	1 registo por disciplina



Para estimar o número total de registos de cada tabela do Data Mart com base nas estimativas feitas para cada uma das tabelas, basta saber o número de escolas que aderem ao serviço da Empresa e para cada escola saber o número de alunos em cada ano escolar e ciclo escolar.

#### **4.12. Revisão e Aprovação final da análise multidimensional**

O modelo multidimensional apresentado deve ser revisto com membros da Empresa para que este seja aprovado.



## 5. Estratégia de Implementação

Neste capítulo são apresentadas as bases de dados necessárias para o tratamento dos dados provenientes das fontes. Depois são explicados os processos envolvidos para o carregamento do Data Warehouse com os dados provenientes das duas fontes.

### 5.1. Bases de Dados

Foi necessário criar bases de dados para permitir a extração e transformação dos dados provenientes das fontes. Depois dos dados estarem tratados podem ser inseridos no Data Mart. Na Figura 20 está o diagrama com as bases de dados criadas.

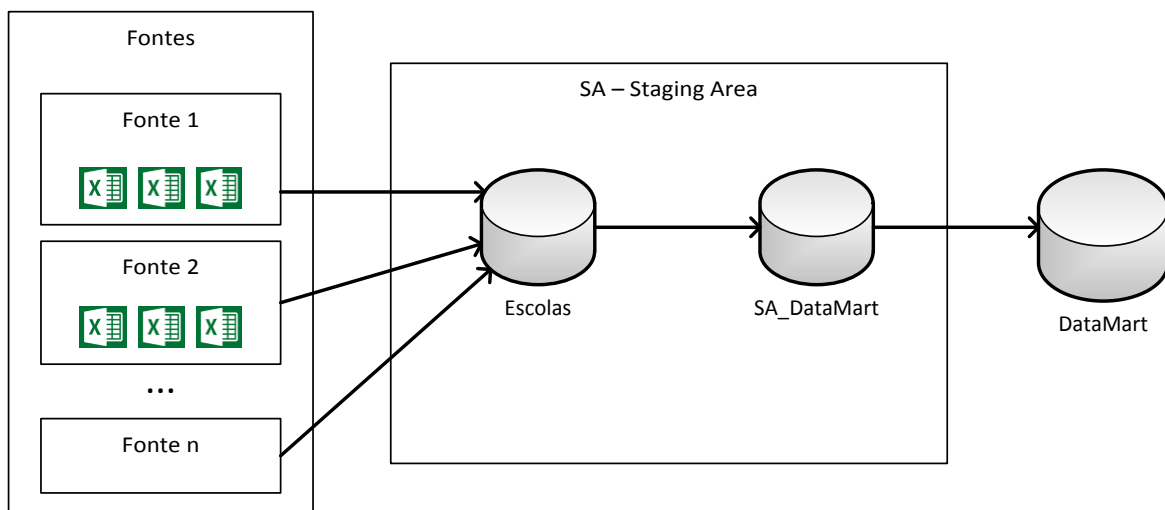


Figura 20 – Diagrama com as bases de dados criadas

No diagrama da Figura 20, as fontes conhecidas estão nomeadas *Fonte 1* e *Fonte 2*. Num caso real em que várias escolas aderem aos serviços de consultadoria da Empresa, podem existir outras fontes com diferentes esquemas e formatos. Sempre que surgir uma nova fonte de dados é necessário definir uma estratégia para a extração dos dados dessa fonte. A seguir são apresentadas as bases de dados que foram criadas para a extração e transformação dos dados.

A base de dados SA (Staging Area) permite fazer a integração e transformação dos dados provenientes das fontes. A base de dados SA está dividida em duas sub-bases de dados que são *Escolas* e *SA\_DataMart*.

## Escolas

A base de dados *Escolas* tem como objetivo juntar os dados provenientes das fontes numa única base de dados. Assim os dados provenientes das fontes ficam integrados numa única base de dados. As tabelas da base de dados foram pensadas de maneira a integrar os dados dos ficheiros das duas fontes conhecidas da forma mais direta possível. Nesta base de dados vai ser feita a transformação dos dados necessários. Os atributos *NumeroProcesso* e *AgrupamentoEscolar* aparecem em muitas tabelas da base de dados *Escolas* e permitem identificar de forma única um aluno. Só o número de processo não chega para identificar um aluno, uma vez que em agrupamentos escolares diferentes pode existir o mesmo número de processo. Um aluno é identificado de forma única através do número de processo e do agrupamento escolar da escola que frequenta. O modelo lógico da base de dados *Escolas* está no Anexo III.

## SA\_DataMart

A base de dados *SA\_DataMart* tem um modelo lógico parecido com o Data Mart, e permite guardar os dados a inserir quando se faz uma atualização do Data Mart. Existe para cada dimensão e para cada tabela de fatos do Data Mart, uma tabela semelhante na base de dados *SA\_DataMart*. A diferença entre as duas bases de dados é que em cada tabela na base de dados *SA\_DataMart* tem um atributo a mais que permite verificar se o registo está atualizado na tabela respetiva do Data Mart. Se o atributo tiver valor 0 significa que não está atualizado, se o atributo tiver o valor 1 significa que está atualizado. Para distinguir as tabelas do *SA\_DataMart* das tabelas do Data Mart, as tabelas do *SA\_DataMart* têm o prefixo SA.

## 5.2. Manipulação dos dados na base de dados SA

Na base de dados *Escolas* são integrados e transformados os dados provenientes das fontes. Depois dos dados estarem tratados são inseridos na base de dados *SA\_DataMart*.

### 5.2.1. Processos de Extração e Transformação dos dados

A seguir descrevo os processos envolvidos na inserção e transformação dos dados na base de dados *Escolas*. A estratégia de extração dos dados das fontes será definida em função das categorias de ficheiros descritas no ponto 4.8. As categorias de ficheiros da *Fonte 1* têm o prefixo F1 e as categorias de ficheiros da *Fonte 2* têm o prefixo F2. No Anexo I estão ficheiros exemplo para as categorias de ficheiros da *Fonte 1*. No Anexo II estão ficheiros exemplo para as categorias de ficheiros da *Fonte 2*.

### Inserção dos dados dos alunos e dos pais dos alunos

A *Fonte 2* tem dados dos alunos e dos pais dos alunos nos ficheiros da categoria *F2\_Alunos*. Na *Fonte 1* a mesma informação está em ficheiros de duas categorias. A categoria *F1\_Alunos* tem dados dos alunos e a categoria *F1\_Pais* tem dados dos pais dos alunos. Na Figura 21 estão as tabelas da base de dados *Escolas* para a inserção dos dados dos alunos e dos pais dos alunos.

Aluno		Pais		AreaProfissao	
PK	<u>NumeroProcesso</u>	PK	<u>NumeroProcesso</u>	PK	<u>Profissao</u>
PK	<u>AgrupamentoEscolar</u>	PK	<u>AgrupamentoEscolar</u>		AreaProfissao
PK	<u>AnoLetivo</u>	PK	<u>AnoLetivo</u>		
	Sexo		HabilitacaoPai	CategoriaHabilitacao	
	DataNascimento		HabilitacaoMae	PK	<u>Habilitacao</u>
	NEE		ProfissaoPai		Categoria
			ProfissaoMae		

Figura 21 - Tabela para a inserção de dados dos alunos e dos pais

Para cada aluno os dados do aluno existentes nos ficheiros são o número de processo, sexo, data de nascimento e se tem necessidades educativas especiais. Para cada aluno os dados dos pais existentes nos ficheiros são as habilitações e profissões do pai e da mãe. Nos ficheiros da categoria *F1\_Pais* existe também o número de processo para cada aluno. Todos os ficheiros têm o nome do agrupamento escolar. A informação relativamente às áreas de profissão do pai e da mãe não existem nas fontes de dados. Por isso foi criada a tabela *AreaProfissao* na base de dados *Escolas*, que para cada profissão tem a área de profissão respetiva. Relativamente às habilitações, estas existem nas fontes de dados mas não estão categorizadas com os valores pretendidos. Por isso foi criada uma tabela *CategoriaHabilitacao* na base de dados *Escolas* para atribuir a cada habilitação a categoria respetiva. É necessário implementar um trigger na tabela *Pais* que após a inserção de um registo, executa as seguintes operações:

- Para cada valor dos atributos *ProfissaoPai* e *ProfissaoMae*, inserir um registo na tabela *AreaProfissao* sempre que a profissão não exista nessa tabela
  - Se na tabela *AreaProfissao* não existir a profissão do pai é inserido um registo em que o valor do atributo *Profissao* dessa tabela fica igual ao valor do atributo *ProfissaoPai* da tabela *Pais*;
  - Se na tabela *AreaProfissao* não existir a profissão da mãe é inserido um registo em que o valor do atributo *Profissao* dessa tabela fica igual ao valor do atributo *ProfissaoMae* da tabela *Pais*;

- O atributo *AreaProfissao* do registo inserido fica com o valor “Desconhecido”.
- Para cada valor dos atributos *HabilitacaoPai* e *HabilitacaoMae*, inserir um registo na tabela *CategoriaHabilitacao* sempre que a habilitação não exista nessa tabela
  - Se na tabela *CategoriaHabilitacao* não existir a habilitação do pai é inserido um registo em que o valor do atributo *Habilitacao* dessa tabela fica igual ao valor do atributo *HabilitacaoPai* da tabela *Pais*;
  - Se na tabela *CategoriaHabilitacao* não existir a habilitação da mãe é inserido um registo em que o valor do atributo *Habilitacao* dessa tabela fica igual ao valor do atributo *HabilitacaoMae* da tabela *Pais*;
  - O atributo *Categoria* do registo inserido fica com o valor “Desconhecido”.

Para extrair os dados deve ser implementado um programa para cada uma das categorias de ficheiros mencionadas. Nos ficheiros da categoria *F1\_Alunos* e *F1\_Pais* existe um ficheiro por ano letivo. Nos ficheiros da categoria *F2\_Alunos* existe um ficheiro por turma com um determinado ano escolar e ano letivo. Cada programa tem de percorrer os ficheiros da categoria respetiva para extrair os dados e inseri-los nas tabelas *Aluno* e *Pais* da base de dados *Escolas*. Na Tabela 17 pode observar-se para os atributos da tabela *Aluno*, as categorias de ficheiros de onde vem os dados.

**Tabela 17 – Origem dos dados para a tabela *Aluno***

Atributos	Descrição	Categoria de Ficheiros	
		Fonte 1	Fonte 2
NumeroProcesso	Número de processo do aluno	<i>F1_Alunos</i>	<i>F2_Alunos</i>
AgrupamentoEscolar	Agrupamento escolar	<i>F1_Alunos</i>	<i>F2_Alunos</i>
AnoLetivo	Ano letivo	<i>F1_Alunos</i>	<i>F2_Alunos</i>
Sexo	Sexo do aluno	<i>F1_Alunos</i>	<i>F2_Alunos</i>
DataNascimento	Data de nascimento	<i>F1_Alunos</i>	<i>F2_Alunos</i>
NEE	Se o aluno tem ou não necessidades educativas especiais	<i>F1_Alunos</i>	<i>F2_Alunos</i>

Para evitar inserir valores null na tabela *Aluno* os programas a implementar para as categorias de ficheiros *F1\_Alunos* e *F2\_Alunos* devem executar as seguintes operações para valores desconhecidos:

- O valor do atributo *Sexo* fica com o valor “Desconhecido”;
- O valor do atributo *DataNascimento* fica com o valor por defeito “0001-01-01”;
- O valor do atributo *NEE* fica com o valor 0, assume-se que o aluno não tem necessidades educativas especiais.

Os dados relativamente às necessidades educativas especiais são guardados de forma diferente nas categorias de ficheiros *F1\_Alunos* e *F2\_Alunos*. Nos ficheiros da categoria *F1\_Alunos* para cada aluno, se este tem necessidades educativas especiais, o valor para as necessidades educativas especiais é igual a “true”, caso contrário o valor é igual “false”. Nos ficheiros da categoria *F2\_Alunos* para cada aluno, se este tem necessidades educativas especiais, o valor para as necessidades educativas especiais é igual a 1, caso contrário o valor é igual a 0. O programa a implementar para extrair os dados dos ficheiros da categoria *F1\_Alunos*, antes de inserir os registos tem que executar as seguintes operações:

- Substituir cada valor “false” pelo valor 0;
- Substituir cada valor “true” pelo valor 1.

Os dados relativamente ao sexo dos alunos são guardados de forma diferente nas categorias de ficheiros *F1\_Alunos* e *F2\_Alunos*. Nos ficheiros da categoria *F1\_Alunos* para cada aluno, se este for do sexo masculino, o valor para o sexo é igual a “Masculino”, se for do sexo feminino, o valor para o sexo é igual “Feminino”. Nos ficheiros da categoria *F2\_Alunos* para cada aluno, se este for do sexo masculino, o valor para o sexo é igual a “M”, se for do sexo feminino, o valor para o sexo é igual a “F”. O programa a implementar para extrair os dados dos ficheiros da categoria *F2\_Alunos*, antes de inserir os registos tem que executar as seguintes operações:

- Substituir cada valor “M” pelo valor “Masculino”;
- Substituir cada valor “F” pelo valor “Feminino”.

Para evitar inserir valores null na tabela *Pais* os programas a implementar devem executar as seguintes operações para valores desconhecidos:

- O valor dos atributos *HabilitacaoPai*, *HabilitacaoMae*, *ProfissaoPai* e *ProfissaoMae* ficam com o valor “Desconhecido”.

Na Tabela 18 pode observar-se para os atributos da tabela *Pais*, as categorias de ficheiros de onde vêm os dados.

Tabela 18 – Origem dos dados para a tabela *Pais*

Atributos	Descrição	Categoria de Ficheiros	
		Fonte 1	Fonte 2
NumeroProcesso	Número de processo do aluno	<i>F1_Pais</i>	<i>F2_Alunos</i>
AgrupamentoEscolar	Agrupamento escolar	<i>F1_Pais</i>	<i>F2_Alunos</i>
AnoLetivo	Ano letivo	<i>F1_Pais</i>	<i>F2_Alunos</i>
HabilitacaoPai	Habilitação do pai do aluno	<i>F1_Pais</i>	<i>F2_Alunos</i>
HabilitacaoMae	Habilitação da mãe do aluno	<i>F1_Pais</i>	<i>F2_Alunos</i>
ProfissaoPai	Profissão do pai do aluno	<i>F1_Pais</i>	<i>F2_Alunos</i>
ProfissaoMae	Profissão da mãe do aluno	<i>F1_Pais</i>	<i>F2_Alunos</i>

Para atribuir a categoria de habilitação às habilitações que estão na tabela *CategoriaHabilitacao*, deve ser implementado um programa que mapeia as habilitações para a respetiva categoria de habilitação. O programa para cada registo da tabela *CategoriaHabilitacao*, substitui o valor “Desconhecido” no atributo *Categoria* pela categoria da habilitação.

As categorias de habilitação são as seguintes:

- Ensino Superior;
- Secundário e 3º Ciclo;
- Básico (1º e 2º ciclo);
- Sem Habilitações;
- Desconhecido.

Para atribuir a área de profissão às profissões que estão na tabela *AreaProfissao* deve ser implementado um programa que mapeia as profissões para a respetiva área de profissão. O programa para cada registo da tabela *AreaProfissao*, substitui o valor “Desconhecido” no atributo *AreaProfissao* pela área de profissão. As áreas de profissão a atribuir são as seguintes:

- Quadros Dirigentes e Superiores;
- Quadros Intermédios;
- Trabalhadores Qualificados;



- Trabalhadores Não Qualificados;
- Desconhecido.

Se no fim da execução dos programas continuarem a existir valores “Desconhecido” no atributo *AreaProfissao* da tabela *AreaProfissao* ou no atributo *Categoria* na tabela *CategoriaHabilitacao*, tem de existir a intervenção de um operador para tentar substituir esses valores.

#### Inserção dos dados relativamente aos escalões sociais dos alunos

Os dados dos escalões sociais dos alunos são obtidos dos ficheiros da categoria *F1\_Escalões* e *F2\_Escalões*. Na Figura 22 está a tabela da base de dados *Escolas* para a inserção dos escalões sociais dos alunos.

Escalao	
PK	<u>NumeroProcesso</u>
PK	<u>AgrupamentoEscolar</u>
PK	<u>AnoLetivo</u>
	Escalao

Figura 22 - Tabela para a inserção de dados dos Escalões dos Alunos

Nas categorias *F1\_Escalões* e *F2\_Escalões* existem ficheiros para o escalão A e ficheiros para o escalão B. Cada um desses ficheiros tem o escalão, o agrupamento escolar, o ano letivo e o ano escolar. Para cada aluno esses ficheiros têm o número de processo do aluno. Na Tabela 19 pode observar-se para os atributos da tabela *Escalao*, as categorias de ficheiros de onde vêm os dados.

Tabela 19 – Origem dos dados para a tabela *Escalao*

Atributos	Descrição	Categoria de Ficheiros	
		Fonte 1	Fonte 2
NumeroProcesso	Número de processo do aluno	<i>F1_Escalões</i>	<i>F2_Escalões</i>
AgrupamentoEscolar	Agrupamento escolar	<i>F1_Escalões</i>	<i>F2_Escalões</i>
AnoLetivo	Ano letivo	<i>F1_Escalões</i>	<i>F2_Escalões</i>
Escalão	Escalão social do aluno	<i>F1_Escalões</i>	<i>F2_Escalões</i>

Os ficheiros das categorias *F1\_Escalões* e *F2\_Escalões* têm o mesmo esquema, por isso pode ser implementado apenas um programa para extrair os dados dos ficheiros das duas categorias. Em cada categoria existe um ficheiro do escalão A e outro do escalão B, por cada ano escolar para cada ano letivo. O programa tem de percorrer os ficheiros da categoria respetiva para extrair os dados e inseri-los na tabela *Escalao* da base de dados *Escolas*. Na tabela *Escalao* não existem valores null uma vez que são guardados apenas registos de alunos que têm escalão.

#### Inserção dos dados relativamente às notas, disciplinas e estado de avaliação dos alunos

Os dados relativamente às notas, disciplinas e estado de avaliação dos alunos são obtidos nos ficheiros das categorias *F1\_Turmas* e *F2\_Turmas*. Em cada um dos ficheiros existe para cada aluno o número de processo, as notas que teve a cada disciplina e o estado de avaliação. Todos os ficheiros têm o nome do agrupamento escolar, o ano letivo e o ano escolar. Nos ficheiros da categoria *F1\_Turmas* têm também o nome da escola. Nos ficheiros da categoria *F2\_Turmas*, não existe o nome da escola. Os ficheiros dessa categoria estão organizados por escola e com base na turma consegue-se saber a que escola pertence. Na Figura 23 estão as tabelas da base de dados *Escolas* necessárias para a inserção das disciplinas, notas e estado de avaliação dos alunos.

Nota5_9		Nota10_12		EstadoAvaliacao	
PK	<u>NumeroProcesso</u>	PK	<u>NumeroProcesso</u>	PK	<u>NumeroProcesso</u>
PK	<u>AgrupamentoEscolar</u>	PK	<u>AgrupamentoEscolar</u>	PK	<u>AgrupamentoEscolar</u>
PK	<u>AnoLetivo</u>	PK	<u>AnoLetivo</u>	PK	<u>AnoLetivo</u>
PK	<u>AnoEscolar</u>	PK	<u>AnoEscolar</u>		Escola
PK	<u>Disciplina</u>	PK	<u>Disciplina</u>		AnoEscolar
	Escola		Escola		EstadoAvaliacao
	CF		CF	DisciplinaPadrao	
	Exame		CIF	PK	<u>Nome</u>
	CFD		Exame		NomePadrao
			CFD		

Figura 23 – Tabelas para a inserção de Disciplinas, Notas e Estado de Avaliação dos Alunos

Na tabela *EstadoAvaliacao* tem de ser implementado um trigger que antes da inserção de um registo, executa as seguintes operações:

- Verificar se na tabela já existe um estado de avaliação com o mesmo número de processo, o mesmo agrupamento escolar e o mesmo ano letivo
  - Se não existir é inserido na tabela o registo com o número de processo, agrupamento escolar, ano letivo, ano escolar, escola e estado de avaliação;

- Se existir o registo não é inserido.

O trigger tem como objetivo garantir que em cada ano letivo um aluno só possa ter um estado de avaliação. Como um aluno pode em casos excecionais frequentar disciplinas de anos escolares diferentes no mesmo ano letivo, deve ser guardado o estado de avaliação do ano escolar mais elevado. Para as tabelas *Nota5\_9* e *Nota10\_12* têm de ser implementado um trigger para cada tabela. Ambos os triggers verificam depois de uma inserção se a disciplina existe na tabela *DisciplinaPadrao*. Os triggers devem executar as seguintes operações nas respetivas tabelas:

- Verificar se a disciplina inserida existe na tabela *DisciplinaPadrao*
  - Se a disciplina não existe é inserido um registo na tabela *DisciplinaPadrao*, em que o atributo *Nome* fica com o nome da disciplina a inserir e o atributo *NomePadrao* fica com o valor “Desconhecido”.

Os triggers descritos a cima têm como objetivo integrar as disciplinas existentes numa tabela e atribuir um nome padrão a cada uma das disciplinas. Nas fontes de dados foi analisado que para uma determinada disciplina, esta podia ser nomeada com nomes diferentes.

Para extrair os dados deve ser implementado um programa para cada uma das categorias de ficheiros mencionadas. Nos ficheiros da categoria *F1\_Turmas* e *F2\_Turmas* existe um ficheiro por cada turma do 5º ano ao 12º ano para cada ano letivo. Nos ficheiros da categoria *F2\_Turmas*, o nome das escolas não pode ser obtido de forma direta. Os ficheiros da categoria *F2\_Turmas* estão organizados por escola e com base na turma consegue-se saber a que escola pertence. O programa da categoria de ficheiros *F2\_Turmas* deve determinar o nome da escola a partir da turma. Nos ficheiros da categoria *F2\_Turmas*, cada disciplina de língua estrangeira tem o prefixo “L.E.” (por exemplo “L.E. 1” para a língua estrangeira 1). Para obter o nome da disciplina existe nesses ficheiros uma coluna que identifica a disciplina (“I” – disciplina de Inglês). O programa a implementar para os ficheiros desta categoria têm de definir os nomes das disciplinas de língua estrangeira.

Cada programa tem de percorrer os ficheiros da categoria respetiva para extrair os dados e inseri-los nas tabelas *EstadoAvaliacao*, *Nota5\_9* e *Nota10\_12* da base de dados *Escolas*. Os ficheiros em cada ano letivo devem ser lidos de forma decrescente por ano escolar. A leitura é feita desta forma para que o trigger da tabela *EstadoAvaliacao* descrito anteriormente funcione corretamente. Nos ficheiros da categoria *F2\_Turmas* para as turmas do 5º ano ao 9º ano de escolaridade, não se pode obter de forma direta notas *CFD* em disciplinas sem exame. O programa a implementar para essa categoria de ficheiros tem de atribuir o valor da nota *CF* à nota *CFD* em disciplinas sem exame. Se para um determinado aluno e para uma determinada disciplina todas as notas forem null, os

programas a implementar para as categorias de ficheiros *F1\_Turmas* e *F2\_Turmas* devem descartar o registo. Na Tabela 20 pode observar-se para os atributos da tabela *Nota5\_9*, as categorias de ficheiros de onde vêm os dados.

Tabela 20 – Origem dos dados para a tabela *Nota5\_9*

Atributos	Descrição	Categoria de Ficheiros	
		Fonte 1	Fonte 2
NumeroProcesso	Número de processo do aluno	<i>F1_Turmas</i>	<i>F2_Turmas</i>
AgrupamentoEscolar	Agrupamento escolar	<i>F1_Turmas</i>	<i>F2_Turmas</i>
AnoLetivo	Ano letivo	<i>F1_Turmas</i>	<i>F2_Turmas</i>
AnoEscolar	Ano Escolar	<i>F1_Turmas</i>	<i>F2_Turmas</i>
Disciplina	Nome da disciplina	<i>F1_Turmas</i>	<i>F2_Turmas</i>
Escola	Nome da escola	<i>F1_Turmas</i>	<i>F2_Turmas</i>
CF	Classificação interna anual à disciplina (1 a 5)	<i>F1_Turmas</i>	<i>F2_Turmas</i>
Exame	Classificação de exame à disciplina (1 a 5)	<i>F1_Turmas</i>	<i>F2_Turmas</i>
CFD	Classificação final anual à disciplina (1 a 5)	<i>F1_Turmas</i>	<i>F2_Turmas</i>

Na tabela *Nota5\_9* podem existir valores a null em algumas notas. Quando o registo tem apenas a nota *CF* nas disciplinas de Português e Matemática no 6º ano e 9º ano, é assumido que o aluno faltou ao exame. Quando o registo tem apenas as notas *Exame* e *CFD*, é assumido que o aluno tem estatuto autoproposto<sup>1</sup> pelo que não tem classificação interna à disciplina. Para corrigir valores null na tabela *Nota5\_9*, deve ser implementado um programa que para cada registo da tabela executa as seguintes operações:

- Em disciplinas com exame, as notas com valor null devem ser atualizadas de maneira a obter um dos seguintes resultados:
  - As notas *CF*, *Exame* e *CFD* ficam com valores diferentes de null,

<sup>1</sup> Um aluno com estatuto autoproposto tem de realizar exames às disciplinas no qual não teve positiva, de maneira a acabar o ano escolar ou ciclo escolar.

- As notas a null são obtidas com a fórmula,  $CFD = 0,7 \times CF + 0,3 \times Exame$ ;
  - A nota *CF* tem valor e as notas de *Exame* e *CFD* ficam igual a null,
    - São as notas de um aluno que faltou ao exame;
  - A nota *CF* fica igual a null e as notas *Exame* e *CFD* têm o mesmo valor,
    - São as notas de um aluno com estatuto autoproposto.
- Em disciplinas sem exame, as notas com valor null devem ser atualizadas de maneira a obter o seguinte resultado:
  - As notas *CF* e *CFD* têm o mesmo valor e a nota *Exame* fica igual a null.

Na Tabela 21 pode observar-se para os atributos da tabela *Nota10\_12*, as categorias de ficheiros de onde vêm os dados.

**Tabela 21 – Origem dos dados para a tabela *Nota10\_12***

Atributos	Descrição	Categoria de Ficheiros	
		Fonte 1	Fonte 2
NumeroProcesso	Número de processo do aluno	<i>F1_Turmas</i>	<i>F2_Turmas</i>
AgrupamentoEscolar	Agrupamento escolar	<i>F1_Turmas</i>	<i>F2_Turmas</i>
AnoLetivo	Ano letivo	<i>F1_Turmas</i>	<i>F2_Turmas</i>
AnoEscolar	Ano Escolar	<i>F1_Turmas</i>	<i>F2_Turmas</i>
Disciplina	Nome da disciplina	<i>F1_Turmas</i>	<i>F2_Turmas</i>
Escola	Nome da escola	<i>F1_Turmas</i>	<i>F2_Turmas</i>
CF	Classificação interna anual à disciplina (0 a 20)	<i>F1_Turmas</i>	<i>F2_Turmas</i>
CIF	Classificação média interna à disciplina no ano escolar de conclusão (0 a 20)	<i>F1_Turmas</i>	<i>F2_Turmas</i>
Exame	Classificação de exame à disciplina (0 a 20)	<i>F1_Turmas</i>	<i>F2_Turmas</i>
CFD	Classificação média final à disciplina no ano escolar de conclusão (0 a 20)	<i>F1_Turmas</i>	<i>F2_Turmas</i>

Nos ficheiros da categoria *F2\_Turmas* para as turmas do 10º ano ao 12º ano de escolaridade, não existem notas *CFD*. O programa a implementar para essa categoria de ficheiros tem de atribuir o valor da nota *CIF* à nota *CFD* em disciplinas em ano de conclusão sem exame. Para as disciplinas em ano de conclusão com exame, o *CFD* deve ser obtido a partir da seguinte fórmula:

- $CFD = 0,7 \times CIF + 0,3 \times Exame.$

As notas *Exame* nos ficheiros têm escala de 0 a 200, por isso é necessário converter essas notas para que fiquem numa escala de 0 a 20. Se para um determinado aluno e para uma determinada disciplina todas as notas forem null, os programas a implementar para as categorias de ficheiros *F1\_Turmas* e *F2\_Turmas* devem descartar o registo. Na tabela *Nota10\_12* podem existir valores a null em algumas notas. Assume-se que quando um registo tem apenas a nota de *Exame* e *CFD*, o aluno que fez essa disciplina é autoproposto e fez o exame para ter aprovação à disciplina. Quando um registo tem notas *CF* e *CIF* e não tem notas *Exame* e *CFD* em disciplinas com exame, assume-se que o aluno faltou ao exame. Para corrigir valores null na tabela *Nota10\_12*, deve ser implementado um programa que para cada registo da tabela executa as seguintes operações:

- Em disciplinas com exame, as notas com valor null devem ser atualizadas de maneira a obter um dos seguintes resultados:
  - As notas *CF*, *CIF*, *Exame* e *CFD* ficam com valores diferentes de null,
    - As notas a null são obtidas com a fórmula,  $CFD = 0,7 \times CIF + 0,3 \times Exame$ ;
  - As notas *CF* e *CIF* têm valor e as notas de *Exame* e *CFD* ficam igual a null,
    - São as notas de um aluno que faltou ao exame;
  - As notas *CF* e *CIF* ficam igual a null e as notas *Exame* e *CFD* têm o mesmo valor,
    - São as notas de um aluno com estatuto autoproposto.
- Em disciplinas no ano de conclusão sem exame, as notas com valor null devem ser atualizadas de maneira a obter o seguinte resultado:
  - As notas *CF*, *CIF*, *CFD* têm valor e a nota *Exame* fica igual a null,
    - As notas *CIF* e *CFD* têm o mesmo valor;
    - Quando a nota *CF* é igual a null, deve ficar com o valor de *CIF*.

Na Tabela 22 pode observar-se para os atributos da tabela *EstadoAvaliacao*, as categorias de ficheiros de onde vêm os dados.

Tabela 22 – Origem dos dados para a tabela *EstadoAvaliacao*

Atributos	Descrição	Categoria de Ficheiros	
		Fonte 1	Fonte 2
NumeroProcesso	Número de processo do aluno	<i>F1_Turmas</i>	<i>F2_Turmas</i>
AgrupamentoEscolar	Agrupamento escolar	<i>F1_Turmas</i>	<i>F2_Turmas</i>
AnoLetivo	Ano letivo	<i>F1_Turmas</i>	<i>F2_Turmas</i>
Escola	Nome da escola	<i>F1_Turmas</i>	<i>F2_Turmas</i>
AnoEscolar	Ano Escolar	<i>F1_Turmas</i>	<i>F2_Turmas</i>
EstadoAvaliacao	Estado de avaliação do aluno	<i>F1_Turmas</i>	<i>F2_Turmas</i>

Para evitar inserir valores null no atributo *EstadoAvaliacao* da tabela *EstadoAvaliacao*, os programas a implementar para as categorias de ficheiros *F1\_Turmas* e *F2\_Turmas*, devem executar as seguintes operações:

- Para cada valor desconhecido, verificar se o aluno tem notas nas tabelas *Nota5\_9* ou *Nota10\_12* dependendo do ano escolar,
  - Se o aluno tiver notas, o valor do atributo *EstadoAvaliacao* fica igual a “Desconhecido” e o registo é inserido na tabela *EstadoAvaliacao*;
  - Se o aluno não tiver notas o registo é descartado uma vez que não existe maneira de saber o estado de avaliação do aluno.

A tabela *EstadoAvaliacao* não pode ficar com valores “Desconhecido” no atributo *EstadoAvaliacao*. Deve ser implementado um programa que para cada registo com o atributo *EstadoAvaliacao* igual a “Desconhecido”, executa uma das seguintes operações:

- Se o valor do atributo *AnoEscolar* estiver compreendido entre o 5º ano e o 9º ano, pesquisar na tabela *Nota5\_9* as notas que tenham o mesmo número de processo, o mesmo agrupamento escolar, o mesmo ano letivo e o mesmo ano escolar. Com base no número de positivas e negativas do aluno é atribuído no atributo *EstadoAvaliacao* o estado de avaliação “Aprovado” ou “Não Aprovado”.
- Se o valor do atributo *AnoEscolar* estiver compreendido entre o 10º ano e o 12º ano, pesquisar na tabela *Nota10\_12* as notas que tenham o mesmo número de processo, o

mesmo agrupamento escolar, o mesmo ano letivo e o mesmo ano escolar. Com base no número de positivas e negativas do aluno é atribuído no atributo *EstadoAvaliacao* o estado de avaliação “Aprovado” ou “Não Aprovado”.

Se o programa não conseguir substituir todos os valores “Desconhecido”, deve existir a intervenção de um operador para atribuir um estado de avaliação no atributo *EstadoAvaliacao*. Na tabela *EstadoAvaliacao*, no atributo *EstadoAvaliacao* só podem existir os valores “Aprovado” ou “Não Aprovado”. Deve existir a intervenção de um operador que atualize os valores no atributo *EstadoAvaliacao* se for diferente de “Aprovado” e “Não Aprovado”. Por exemplo os registos com o valor “Transitou” no atributo *EstadoAvaliacao* devem ser atualizados para o valor “Aprovado”.

Na tabela *DisciplinaPadrao* é necessário atribuir a cada disciplina um nome padrão. Deve ser implementado um programa que para cada registo da tabela substitua o valor “Desconhecido” no atributo *NomePadrao* pelo nome padrão da disciplina. Por exemplo se para a disciplina de matemática na tabela *DisciplinaPadrao* existirem dois registos em que o atributo *Nome* é igual a “MAT” no primeiro registo e é igual a “Matemática” no segundo registo, o atributo *NomePadrao* deve ter o mesmo valor em ambos os registos, para identificar a disciplina de matemática de forma única. Se depois de executar o programa continuarem a existir registos com o atributo *NomePadrao* igual a “Desconhecido”, deve existir a intervenção de um operador para atribuir o nome padrão da disciplina em cada registo, atualizando o atributo *NomePadrao*.

#### Integração dos dados relativamente às características de cada aluno

Anteriormente descrevi a forma como eram inseridos e tratados os dados nas tabelas *Aluno*, e *Escalao*. O objetivo é juntar a informação das duas tabelas e inseri-la na tabela *AlunoCorrigido*. Na Figura 24 estão as tabelas necessárias para o processo.

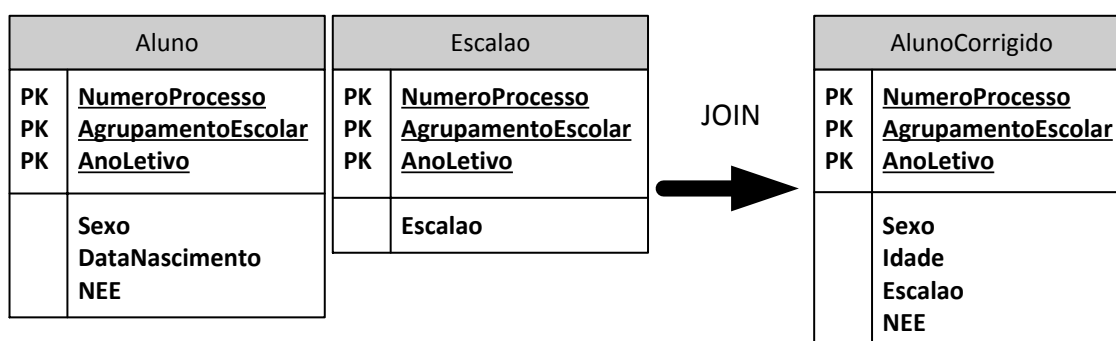


Figura 24 – Tabelas para a integração das Características dos Alunos



A tabela *AlunoCorrigido* permite ter as características de cada aluno por ano letivo. Cada registo tem o número de processo, agrupamento escolar, ano letivo, sexo, idade, escalão social e se o aluno tem necessidades educativas especiais. Deve ser implementado um programa que faça joins das tabelas *Aluno* e *Escalao* para obter os registos a inserir na tabela *AlunoCorrigido*. Na Figura 25 está o query que faz os joins das tabelas *Aluno* e *Escalao*. O resultado do query permite obter para cada registo o número de processo, agrupamento escolar, ano letivo, sexo, data de nascimento, se o aluno tem necessidades educativas especiais, e o escalão social.

```
SELECT Aluno.NumeroProcesso, Aluno.AgrupamentoEscolar, Aluno.AnoLetivo,
Aluno.Sexo, Aluno.DataNascimento, Aluno.NEE, Escalao.Escalao FROM Aluno
LEFT JOIN Escalao ON Aluno.NumeroProcesso=Escalao.NumeroProcesso
AND Aluno.AgrupamentoEscolar=Escalao.AgrupamentoEscolar
AND Aluno.AnoLetivo=Escalao.AnoLetivo
```

Figura 25 – Query para fazer joins das tabelas *Aluno* e *Escalão*

Para cada registo obtido no query, são executadas as seguintes operações:

- Se o valor do atributo *Escalao* for igual a null, deve ser substituído pelo valor “SE”, uma vez que o aluno não tem escalão.
- Calcular a idade a partir da data de nascimento
  - Se a data de nascimento for conhecida, a idade é calculada com base no ano letivo e a data de nascimento;
  - Se a data de nascimento não for conhecida (*DataNascimento* igual a “0001-1-1”), o programa tem de verificar na tabela *EstadoAvaliacao* descrita anteriormente, qual é o ano escolar que o aluno está a frequentar no ano letivo analisado. A idade é calculada pela seguinte fórmula:
    - $Idade = AnoEscolar + 6$  : Idade de um aluno com percurso escolar normal.
- O registo do aluno é inserido na tabela *AlunoCorrigido* com o número de processo, agrupamento escolar, ano letivo, sexo, idade, escalão e se tem necessidades educativas especiais.

### Integração dos dados relativamente aos pais dos alunos

Anteriormente descrevi a forma como eram inseridos e tratados os dados nas tabelas *Pais*, *AreaProfissao* e *CategoriaHabilitacao*. O objetivo é juntar a informação dessas três tabelas e inseri-la na tabela *PaisCorrigido*. Na Figura 26 estão as tabelas necessárias para o processo. A tabela *PaisCorrigido* permite ter a informação dos pais de cada aluno por ano letivo. Cada registo tem o

número de processo, agrupamento escolar, ano letivo, habilitação do pai, habilitação da mãe, área de profissão do pai e área de profissão da mãe.

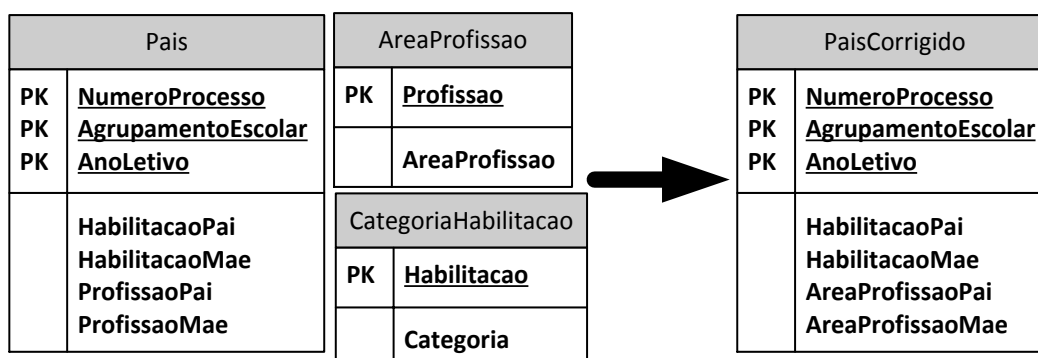


Figura 26 - Tabelas para a integração dos dados dos Pais

Na Figura 27 está o query que deve ser executado para inserir dados na tabela *PaisCorrigido*.

```
SELECT NumeroProcesso, AgrupamentoEscolar, AnoLetivo,
HabilitacaoPai=(SELECT Categoria FROM CategoriaHabilitacao WHERE Pais.HabilitacaoPai=Habilitacao),
HabilitacaoMae=(SELECT Categoria FROM CategoriaHabilitacao WHERE Pais.HabilitacaoMae=Habilitacao),
AreaProfissaoPai=(SELECT AreaProfissao FROM AreaProfissao WHERE ProfissaoPai=Profissao),
AreaProfissaoMae=(SELECT AreaProfissao FROM AreaProfissao WHERE ProfissaoMae=Profissao)
FROM PAIS
```

Figura 27 - Query para inserir dados na tabela PaisCorrigido

Na tabela *PaisCorrigidos* são inseridos os registos provenientes da tabela *Pais*. A área da profissão para o pai é obtida a partir do atributo *ProfissaoPai*, selecionando o valor do atributo *AreaProfissao* da tabela *AreaProfissao* onde *ProfissaoPai* tem o mesmo valor que *Profissao*. A área de profissão da mãe obtém-se de maneira semelhante. A categoria de habilitação para o pai é obtida a partir do atributo *HabilitacaoPai*, selecionando o valor do atributo *Categoria* da tabela *CategoriaHabilitacao* onde o atributo *HabilitacaoPai* da tabela *Pais* tem o mesmo valor que *Habilitacao*. A categoria de habilitação para a mãe obtém-se de maneira semelhante. Os registos são inseridos na tabela *PaisCorrigido* com o número de processo, agrupamento escolar, ano letivo, categoria da habilitação do pai, categoria da habilitação da mãe, área de profissão do pai e área de profissão da mãe.

### 5.2.2. Processos de Carregamento dos dados na base de dados SA\_DataMart

Depois dos dados terem sido inseridos e tratados na base de dados *Escolas*, vão ser inseridos na base de dados *SA\_DataMart*. A seguir estão descritos os processos para o carregamento das tabelas da base de dados *SA\_DataMart*.

### Carregamento da *SAANL\_AnoLetivo*

A tabela *SAANL\_AnoLetivo* permite guardar os dados a inserir no Data Mart relativamente à dimensão *Ano Letivo*. A seguir estão os passos necessários para o carregamento da tabela *SAANL\_AnoLetivo*. Na Figura 28 está a tabela *SAANL\_AnoLetivo*.

SAANL_AnoLetivo	
PK	<u>ANL_id</u>
	ANL_ano ANL_trienio ANL_atualizado

Figura 28 – Tabela para o carregamento da *SAANL\_AnoLetivo*

Para preencher a tabela *SAANL\_AnoLetivo*, deve ser implementado um programa que obtém os anos letivos existentes nas tabelas *Nota5\_9*, *Nota10\_12* e *EstadoAvaliacao*, e para cada ano letivo que não exista na tabela *SAANL\_AnoLetivo* são executadas as seguintes operações:

- Criar a surrogate key para o atributo *ANL\_id*;
- Atribuir um triénio ao ano letivo;
- Inserir o registo na tabela *SAANL\_AnoLetivo* com a surrogate key criada, ano letivo, triénio e o atributo *ANL\_atualizado* igual a 0.

Um triénio define um conjunto de três anos letivos. Por exemplo se existirem nas fontes os anos letivos 2004, 2005 e 2006 o programa a implementar pode atribuir o triénio “2004-2006” a esses anos letivos. Sempre que um registo na tabela *SAANL\_AnoLetivo* for alterado, o valor do atributo *ANL\_atualizado* passa a ser 0.

### Carregamento da *SAANE\_AnoEscolar*

A tabela *SAANE\_AnoEscolar* permite guardar os dados a inserir no Data Mart relativamente à dimensão *Ano Escolar*. A seguir estão os passos necessários para o carregamento da tabela *SAANE\_AnoEscolar*. Na Figura 29 está a tabela *SAANE\_AnoEscolar*.

SAANE_AnoEscolar	
PK	<u>ANE_id</u>
	ANE_ano ANE_ciclo ANE_atualizado

Figura 29 - Tabela para o carregamento da *SAANE\_AnoEscolar*

Os anos escolares vão do 5º ano ao 12º ano. O 2º ciclo vai do 5º ano ao 6º ano, o 3º ciclo vai do 7º ano ao 9º ano e o Secundário vai do 10º ano ao 12º ano. Como os anos escolares e ciclos escolares são conhecidos podem ser logo todos inseridos na tabela *SAANE\_AnoEscolar* por um operador. Por cada registo a inserir, o operador terá que criar a surrogate key para atributo *ANE\_id* de cada registo. O valor *ANE\_atualizado* em cada registo é igual a 0, o que permite saber que esses registos não estão atualizados no Data Mart. Sempre que um registo na tabela *SAANE\_AnoEscolar* for alterado, o valor do atributo *ANE\_atualizado* passa a ser 0.

#### Carregamento da *SAESC\_Escola*

A tabela *SAESC\_Escola* permite guardar os dados a inserir no Data Mart relativamente à dimensão *Escola*. A seguir estão os passos necessários para o carregamento da tabela *SAESC\_Escola*. Na Figura 30 está a tabela *SAESC\_Escola* com o tipo de SCD utilizado entre parênteses nos atributos.

SAESC_Escola	
PK	<u>ESC_id</u>
	ESC_escola(1) ESC_agrupamento(2) ESC_tipo(2) ESC_freguesia(2) ESC_concelho(2) ESC_distrito(1) ESC_validoDe ESC_validoAte ESC_atualizado

Figura 30 - Tabela para o carregamento da *SAESC\_Escola*

Na tabela *SAESC\_Escola* é guardado o histórico para alguns atributos, sendo por isso que existem atributos com SCD (Slowly Changing Dimensions) igual a 2. Na tabela *SAESC\_Escola* é necessário implementar um trigger que antes da inserção de um registo de uma escola, verifica se

existem registos com a mesma escola através do atributo *ESC\_escola*. São executadas as seguintes operações:

- Se existirem registos da escola,
  - Os valores dos atributos com SCD igual a 1 desses registos devem ser atualizados com os valores do novo registo, caso os valores não coincidam. O atributo *ESC\_atualizado* fica com o valor 0 nos registos atualizados;
  - Os atributos com SCD igual a 2 do registo a inserir devem ser comparados com o registo mais recente da escola. O registo mais recente tem o valor 9999 no atributo *ESC\_validoAte*. Se os atributos com SCD igual a 2 forem diferentes,
    - O registo da escola com valor 9999 no atributo *ESC\_validoAte* deve ser atualizado para o valor do ano letivo para o qual acaba a validade do registo. O atributo *ESC\_atualizado* fica com o valor 0.
- O novo registo da escola é inserido na tabela *SAESC\_Escola*.

A partir das fontes de dados só se consegue obter os nomes das escolas e os agrupamentos escolares respetivos. O nome de cada escola e do respetivo agrupamento escolar pode ser obtido nas tabelas *Nota5\_9*, *Nota10\_12* e *EstadoAvaliacao* da base de dados *Escola*. Deve existir a intervenção de um operador que faça pesquisas com base no nome de cada escola para obter o resto da informação. As pesquisas a fazer são as seguintes:

- Verificar para cada escola qual é tipo de escola, a freguesia o concelho e o distrito.

Com base nas pesquisas feitas são inseridos os registos na tabela *SAESC\_Escola*. Por cada registo a inserir, o operador terá que criar a surrogate key para o atributo *ESC\_id* de cada registo. O registo mais recente de uma escola fica com o valor 9999 no atributo *ESC\_validoAte*. Para os restantes registos da mesma escola os atributos *ESC\_validoDe* e *ESC\_validoAte* têm o intervalo de anos letivos para o qual o registo é válido. O atributo *ESC\_atualizado* em cada registo inserido fica com o valor 0. Por exemplo se uma escola foi do tipo público entre os anos letivos de 2004 e 2007 e passou a ser do tipo privado no ano letivo 2008, na tabela *SAESC\_Escola* para essa escola são inseridos dois registos. O primeiro registo tem o atributo *ESC\_tipo* com o valor “Público”, o atributo *ESC\_validoDe* com o valor 2004 e o atributo *ESC\_validoAte* com o valor 2007. O segundo registo tem o atributo *ESC\_tipo* com o valor “Privado”, o atributo *ESC\_validoDe* com o valor 2008 e o atributo *ESC\_validoAte* com o valor 9999. Sempre que um registo na tabela *SAESC\_Escola* for alterado, o valor do atributo *ESC\_atualizado* passa a ser 0.

### Carregamento da *SACAL\_CaracteristicasAluno*

A tabela *SACAL\_CaracteristicasAluno* permite guardar os dados a inserir no Data Mart relativamente à dimensão *Caracteristicas do Aluno*. A seguir estão os passos necessários para o carregamento da tabela *SACAL\_CaracteristicasAluno*. Na Figura 31 estão as tabelas necessárias para o processo.

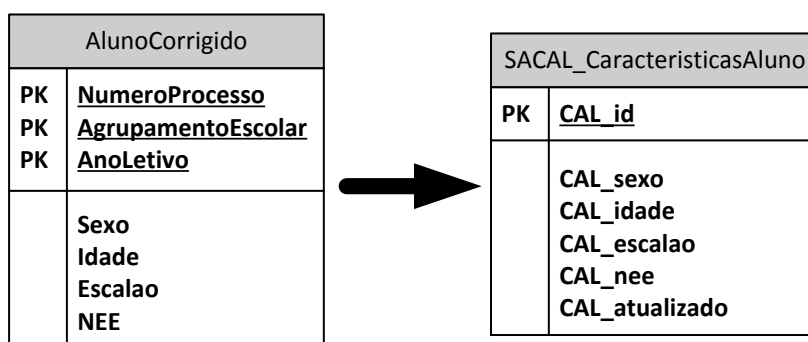


Figura 31 - Tabelas para o carregamento da *SACAL\_CaracteristicasAluno*

Os dados a carregar na tabela *SACAL\_CaracteristicasAluno* são obtidos a partir da tabela *AlunoCorrigido* da base de dados *Escolas*. A forma de essa tabela ser carregada com dados já foi descrita anteriormente. A tabela *SACAL\_CaracteristicasAluno* pode ser preenchida de forma automática a partir da tabela *AlunoCorrigido*. Para isso é necessário implementar um trigger na tabela *AlunoCorrigido*, que depois de uma inserção executa as seguintes operações:

- Verifica se existe um registo na tabela *SACAL\_CaracteristicasAluno* em que *Sexo* é igual a *CAL\_sexo*, *Idade* é igual a *CAL\_idade*, *Escalao* é igual a *CAL\_escalao* e *NEE* é igual a *CAL\_nee*,
  - Se o registo não existir, é criada uma surrogate key para o atributo *CAL\_id* e o registo é inserido na tabela *SACAL\_CaracteristicasAluno* com os valores dos atributos *Sexo*, *Idade*, *Escalao* e *NEE*. O atributo *CAL\_atualizado* tem o valor 0.

O trigger descrito acima deve ser implementado na tabela *AlunoCorrigido* antes de serem inseridos registos na tabela. Sempre que um registo na tabela *SACAL\_CaracteristicasAluno* for alterado, o valor do atributo *CAL\_atualizado* passa a ser 0.

### Carregamento da *SACFA\_ContextoFamiliar*

A tabela *SACFA\_ContextoFamiliar* permite guardar os dados a inserir no Data Mart relativamente à dimensão *Contexto Familiar*. A seguir estão os passos necessários para o

carregamento da tabela *SACFA\_ContextoFamiliar*. Na Figura 32 estão as tabelas necessárias para o processo.

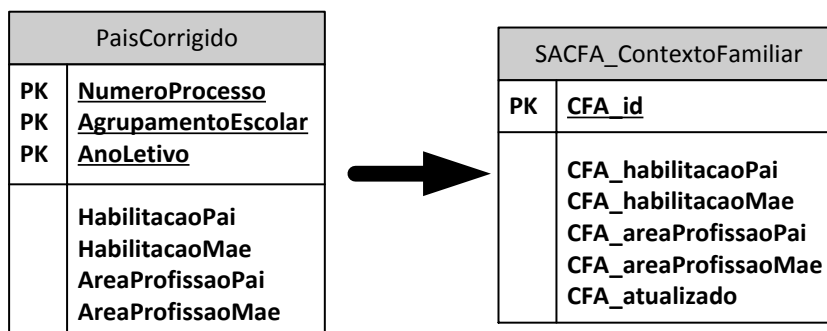


Figura 32 - Tabelas para o carregamento da *SACFA\_ContextoFamiliar*

Os dados a carregar na tabela *SACFA\_ContextoFamiliar* são obtidos a partir da tabela *PaisCorrigido* da base de dados *Escolas*. A forma de essa tabela ser carregada com dados já foi descrita anteriormente. A tabela *SACFA\_ContextoFamiliar* pode ser preenchida de forma automática a partir da tabela *PaisCorrigido*. Para isso é necessário implementar um trigger na tabela *PaisCorrigido*, que depois de uma inserção executa as seguintes operações:

- Verifica se existe um registo na tabela *SACFA\_ContextoFamiliar* em que *CFA\_habilitacaoPai* é igual a *HabilitacaoPai*, *CFA\_habilitacaoMae* é igual a *HabilitacaoMae*, *CFA\_areaProfissaoPai* é igual a *AreaProfissaoPai* e *CFA\_areaProfissaoMae* é igual a *AreaProfissaoMae*,
  - Se o registo não existir, é criado uma surrogate key para o atributo *CFA\_id* e o registo é inserido na tabela *SACFA\_ContextoFamiliar* com os valores dos atributos *HabilitacaoPai*, *HabilitacaoMae*, *AreaProfissaoPai* e *AreaProfissaoMae*. O atributo *CFA\_atualizado* tem o valor 0.

O trigger descrito acima deve ser implementado na tabela *PaisCorrigido* antes de serem inseridos registos na tabela. Sempre que um registo na tabela *SACFA\_ContextoFamiliar* for alterado, o valor do atributo *CFA\_atualizado* passa a ser 0.

#### Carregamento da *SAESA\_EstadoAvaliacao*

A tabela *SAESA\_EstadoAvaliacao* permite guardar os dados a inserir no Data Mart relativamente à dimensão *Estado de Avaliação*. A seguir estão os passos necessários para o carregamento da tabela *SAESA\_EstadoAvaliacao*. Na Figura 33 estão as tabelas necessárias para o processo.

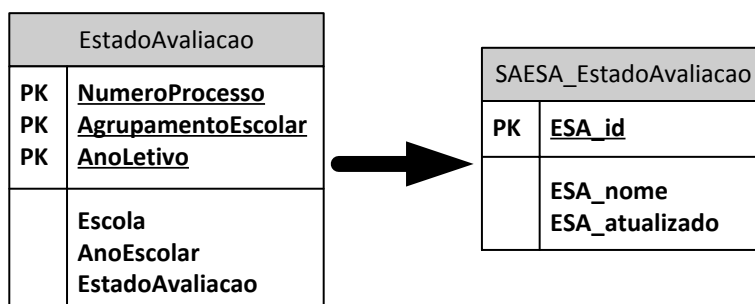


Figura 33 - Tabelas para o carregamento da *SAESA\_EstadoAvaliacao*

Os dados a carregar na tabela *SAESA\_EstadoAvaliacao* são obtidos a partir da tabela *EstadoAvaliacao* da base de dados *Escolas*. A forma de essa tabela ser carregada com dados já foi descrita anteriormente. A tabela *SAESA\_EstadoAvaliacao* vai ter apenas dois registos, uma vez que existem apenas dois estados de avaliação que são aprovado e não aprovado. Por isso só precisa de ser preenchida uma vez. Tem de existir a intervenção de um operador que selecione os valores existentes para o atributo *EstadoAvaliacao* da tabela *EstadoAvaliacao*. Dessa seleção vão resultar dois estados de avaliação. O operador tem de inserir dois registos na tabela *SAESA\_EstadoAvaliacao* um para cada estado de avaliação. Têm de ser criadas surrogates keys para cada um dos registos e o atributo *ESA\_atualizado* tem o valor 0. Sempre que um registo na tabela *SAESA\_EstadoAvaliacao* for alterado, o valor do atributo *ESA\_atualizado* passa a ser 0.

#### Carregamento da *SADIS\_Disciplina*

A tabela *SADIS\_Disciplina* permite guardar os dados a inserir no Data Mart relativamente à dimensão *Disciplina*. A seguir estão os passos necessários para o carregamento da tabela *SADIS\_Disciplina*. Na Figura 34 estão as tabelas necessárias para o processo.

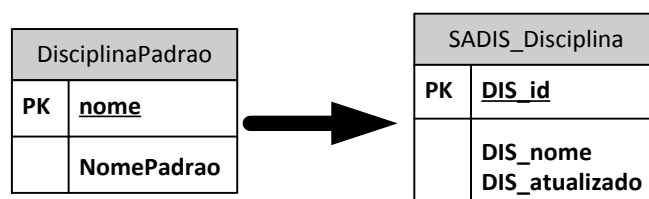


Figura 34 - Tabelas para o carregamento da *SADIS\_Disciplina*

Os dados a carregar na tabela *SADIS\_Disciplina* são obtidos a partir da tabela *DisciplinaPadrao* da base de dados *Escolas*. A forma de essa tabela ser carregada com dados já foi descrita anteriormente. Para carregar a tabela *SADIS\_Disciplina* é necessário implementar um programa que seleciona as disciplinas existentes na tabela *DisciplinaPadrao* através do atributo *NomePadrao*.



São executadas as seguintes operações:

- Verificar para cada registo obtido, se existe um registo com a mesma disciplina. Se não existir,
  - É criada uma surrogate key para o atributo *DIS\_id* e o atributo *DIS\_atualizado* fica com o valor 0;
  - O registo é inserido na tabela *SADIS\_Disciplina*.

Sempre que um registo na tabela *SADIS\_Disciplina* for alterado, o valor do atributo *DIS\_atualizado* passa a ser 0.

### Carregamento da *SAMAT\_Matricula*

A tabela *SAMAT\_Matricula* permite guardar os dados a inserir no Data Mart relativamente à tabela de factos *Matricula*. A seguir estão os passos necessários para o carregamento da tabela *SAMAT\_Matricula*. Na Figura 35 estão as tabelas necessárias para o processo.

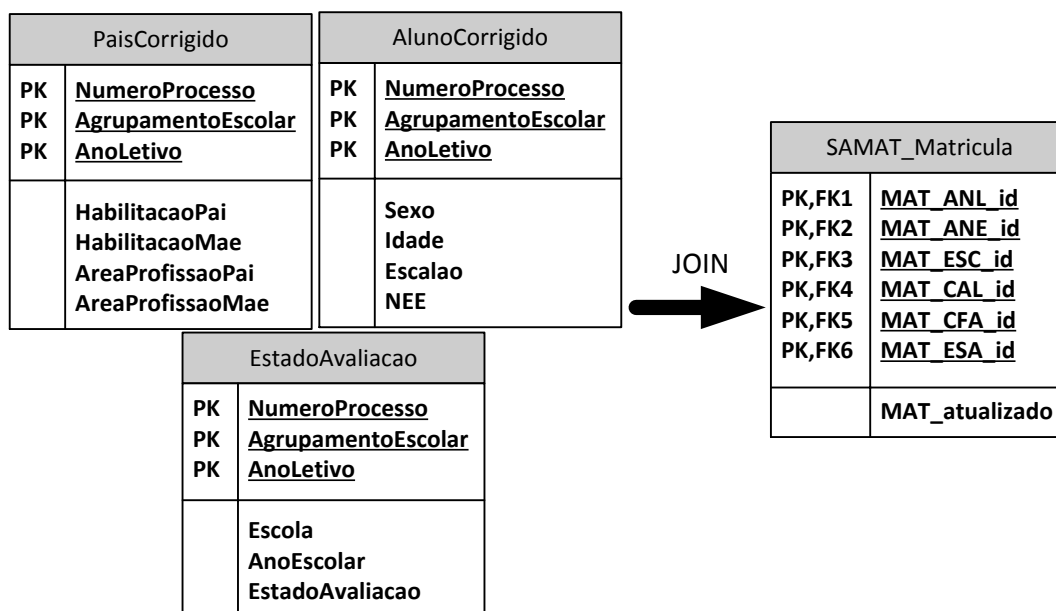


Figura 35 - Tabelas para o carregamento da *SAMAT\_Matricula*

A tabela *SAMAT\_Matricula* não tem medidas uma vez que a tabela de factos *Matricula* é utilizada para contagens. Existem 6 atributos PK/FK na tabela, um por cada dimensão. Para inserir registos na tabela *SAMAT\_Matricula* basta obter para cada registo a surrogate key para cada dimensão através da Alternate Key (AK). Na Figura 36 está o query que permite obter para cada registo as AK de cada dimensão.

```
SELECT EstadoAvaliacao.AnoLetivo, EstadoAvaliacao.AnoEscolar, EstadoAvaliacao.Escola,
AlunoCorrigido.Sexo, AlunoCorrigido.Idade, AlunoCorrigido.Escalao, AlunoCorrigido.NEE,
PaisCorrigido.HabilitacaoPai, PaisCorrigido.HabilitacaoMae, PaisCorrigido.AreaProfissaoPai,
PaisCorrigido.AreaProfissaoMae, EstadoAvaliacao.EstadoAvaliacao
FROM AlunoCorrigido INNER JOIN PaisCorrigido
ON AlunoCorrigido.NumeroProcesso=PaisCorrigido.NumeroProcesso
AND AlunoCorrigido.AgrupamentoEscolar=PaisCorrigido.AgrupamentoEscolar
AND AlunoCorrigido.AnoLetivo=PaisCorrigido.AnoLetivo
INNER JOIN EstadoAvaliacao
ON PaisCorrigido.NumeroProcesso=EstadoAvaliacao.NumeroProcesso
AND PaisCorrigido.AgrupamentoEscolar=EstadoAvaliacao.AgrupamentoEscolar
AND PaisCorrigido.AnoLetivo=EstadoAvaliacao.AnoLetivo
```

**Figura 36 - Query para obter as Alternate Key de cada dimensão da SAMAT\_Matricula**

O query da Figura 36 faz o join das tabelas *AlunoCorrigido*, *PaisCorrigido* e *EstadoAvaliacao*. Com base no resultado do query é possível obter para cada registo a surrogate key de cada dimensão através dos valores dos atributos. Em cada registo a SK é obtida da seguinte forma para cada dimensão:

- A SK da tabela *SAANL\_AnoLetivo* é obtida a partir do valor do atributo *AnoLetivo* da tabela *EstadoAvaliacao*;
- A SK da tabela *SAANE\_AnoEscolar* é obtida a partir do valor do atributo *AnoEscolar* da tabela *EstadoAvaliacao*;
- A SK da tabela *SAESC\_Escola* é obtida a partir dos valores dos atributos *Escolas* e *AnoLetivo* da tabela *EstadoAvaliacao*;
- A SK da tabela *SACAL\_CaracteristicasAluno* é obtida a partir dos valores dos atributos *Sexo*, *Idade*, *Escalao* e *NEE* da tabela *AlunoCorrigido*;
- A SK da tabela *SACFA\_ContextoFamiliar* é obtida a partir dos valores dos atributos *HabilitacaoPai*, *HabilitacaoMae*, *AreaProfissaoPai* e *AreaProfissaoMae* da tabela *PaisCorrigido*;
- A SK da tabela *SAESA\_EstadoAvaliacao* é obtida a partir do valor do atributo *EstadoAvaliacao* da tabela *EstadoAvaliacao*.

Depois das surrogates keys serem obtidas, os registos podem ser inseridos na tabela *SAMAT\_Matricula*. O atributo *MAT\_atualizado* de cada registo tem o valor 0.

#### Carregamento da *SANT1\_Nota5\_9*

A tabela *SANT1\_Nota5\_9* permite guardar os dados a inserir no Data Mart relativamente à tabela de factos *Notas do 5º Ano ao 9º Ano*. A seguir estão os passos necessários para o

carregamento da tabela *SANT1\_Nota5\_9*. Na Figura 37 estão as tabelas necessárias para o processo.

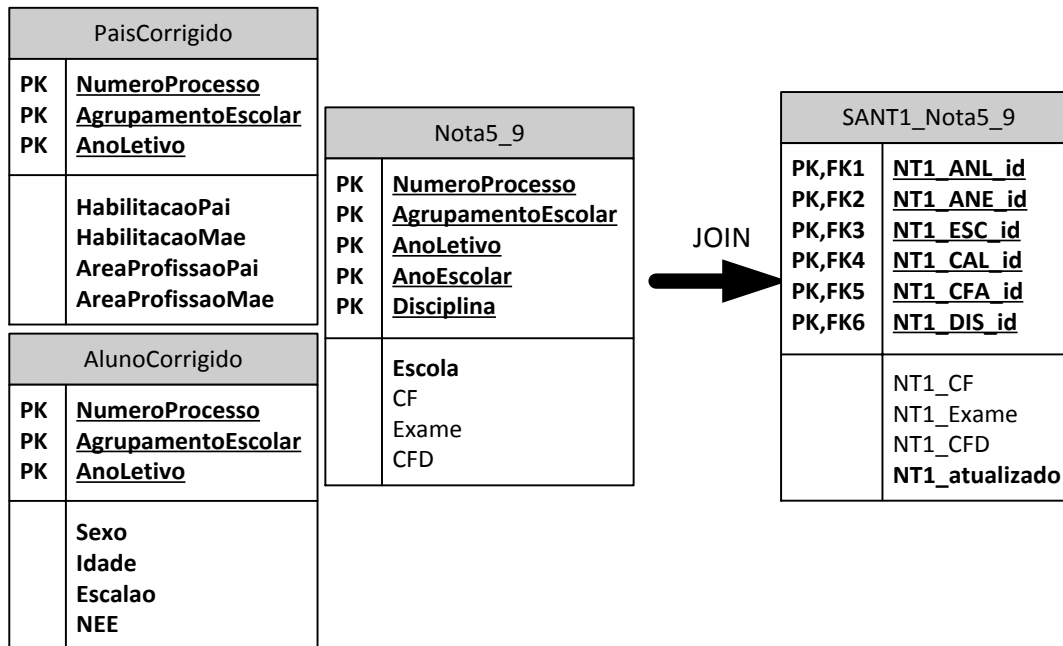


Figura 37 - Tabelas para o carregamento da *SANT1\_Nota5\_9*

Existem 6 atributos PK/FK na tabela, um por cada dimensão. A tabela de fatos *SANT1\_Nota5\_9* tem três medidas que são *NT1\_CF*, *NT1\_Exame* e *NT1\_CFD*. Na Figura 38 está o query que permite obter para cada registo as alternate keys (AK) de cada dimensão e as três medidas.

```
SELECT Nota5_9.AnoLetivo, Nota5_9.AnoEscolar, Nota5_9.Escola, AlunoCorrigido.Sexo,
AlunoCorrigido.Idade, AlunoCorrigido.Escalao, AlunoCorrigido.NEE, PaisCorrigido.HabilitacaoPai,
PaisCorrigido.HabilitacaoMae, PaisCorrigido.AreaProfissaoPai, PaisCorrigido.AreaProfissaoMae,
Disciplina=(SELECT NomePadrao FROM DisciplinaPadrao WHERE Nota5_9.Disciplina=Nome),
Nota5_9.CF, Nota5_9.Exame, Nota5_9.CFD
FROM AlunoCorrigido INNER JOIN PaisCorrigido
ON AlunoCorrigido.NumeroProcesso=PaisCorrigido.NumeroProcesso
AND AlunoCorrigido.AgrupamentoEscolar=PaisCorrigido.AgrupamentoEscolar
AND AlunoCorrigido.AnoLetivo=PaisCorrigido.AnoLetivo
INNER JOIN Nota5_9
ON PaisCorrigido.NumeroProcesso=Nota5_9.NumeroProcesso
AND PaisCorrigido.AgrupamentoEscolar=Nota5_9.AgrupamentoEscolar
AND PaisCorrigido.AnoLetivo=Nota5_9.AnoLetivo
```

Figura 38 - Query para obter as Alternate Key de cada dimensão da *SANT1\_Nota5\_9*

O query da Figura 38 faz o join das tabelas *AlunoCorrigido*, *PaisCorrigido* e *Nota5\_9*. Com base no resultado do query é possível obter para cada registo a surrogate key de cada dimensão através dos valores dos atributos.

Em cada registo do query a SK é obtida da seguinte forma para cada dimensão:

- A SK da tabela *SAANL\_AnoLetivo* é obtida a partir do valor do atributo *AnoLetivo* da tabela *Nota5\_9*;
- A SK da tabela *SAANE\_AnoEscolar* é obtida a partir do valor do atributo *AnoEscolar* da tabela *Nota5\_9*;
- A SK da tabela *SAESC\_Escola* é obtida a partir dos valores dos atributos *Escola* e *AnoLetivo* da tabela *Nota5\_9*;
- A SK da tabela *SACAL\_CaracteristicasAluno* é obtida a partir dos valores dos atributos *Sexo*, *Idade*, *Escalao* e *NEE* da tabela *AlunoCorrigido*;
- A SK da tabela *SACFA\_ContextoFamiliar* é obtida a partir dos valores dos atributos *HabilitacaoPai*, *HabilitacaoMae*, *AreaProfissaoPai* e *AreaProfissaoMae* da tabela *PaisCorrigido*;
- A SK da tabela *SADIS\_Disciplina* é obtida da seguinte forma:
  - Para o valor do atributo *Disciplina* da tabela *Nota5\_9*, é feita uma pesquisa na tabela *DisciplinaPadrao* para obter o nome padrão da disciplina através do valor do atributo *NomePadrao*.

Depois das SK serem obtidas, os registos podem ser inseridos na tabela *SANT1\_Nota5\_9* com as SK obtidas e com as medidas. As medidas são obtidas diretamente do query. O atributo *NT1\_atualizado* de cada registo tem o valor 0.

#### Carregamento da *SANT2\_Nota10\_12*

A tabela *SANT2\_Nota10\_12* permite guardar os dados a inserir no Data Mart relativamente à tabela de factos *Notas do 10º Ano ao 12º Ano*. A seguir estão os passos necessários para o carregamento da tabela *SANT2\_Nota10\_12*. Na Figura 39 estão as tabelas necessárias para o processo.

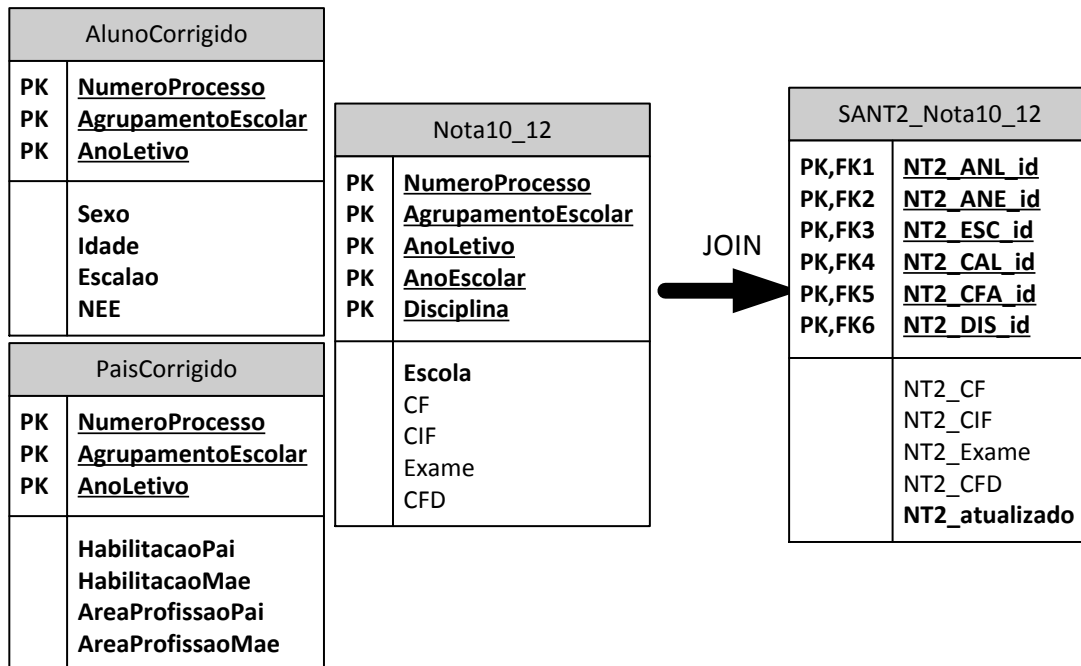


Figura 39 – Tabelas para o carregamento da *SANT2\_Nota10\_12*

Existem 6 atributos PK/FK na tabela, um por cada dimensão. A tabela de fatos *SANT2\_Nota10\_12* tem quatro medidas que são *NT2\_CF*, *NT2\_CIF*, *NT2\_Exame* e *NT2\_CFD*. Na Figura 40 está o query que permite obter para cada registo as alternate keys (AK) de cada dimensão e as quatro medidas. O query faz o join das tabelas *AlunoCorrigido*, *PaisCorrigido* e *Nota10\_12*.

```
SELECT Nota10_12.AnoLetivo, Nota10_12.AnoEscolar, Nota10_12.Escola, AlunoCorrigido.Sexo,
AlunoCorrigido.Idade, AlunoCorrigido.Escalao, AlunoCorrigido.NEE, PaisCorrigido.HabilitacaoPai,
PaisCorrigido.HabilitacaoMae, PaisCorrigido.AreaProfissaoPai, PaisCorrigido.AreaProfissaoMae,
Disciplina=(SELECT NomePadrao FROM DisciplinaPadrao WHERE Nota10_12.Disciplina=Nome),
Nota10_12.CF, Nota10_12.CIF, Nota10_12.Exame, Nota10_12.CFD
FROM AlunoCorrigido INNER JOIN PaisCorrigido
ON AlunoCorrigido.NumeroProcesso=PaisCorrigido.NumeroProcesso
AND AlunoCorrigido.AgrupamentoEscolar=PaisCorrigido.AgrupamentoEscolar
AND AlunoCorrigido.AnoLetivo=PaisCorrigido.AnoLetivo
INNER JOIN Nota10_12
ON PaisCorrigido.NumeroProcesso=Nota10_12.NumeroProcesso
AND PaisCorrigido.AgrupamentoEscolar=Nota10_12.AgrupamentoEscolar
AND PaisCorrigido.AnoLetivo=Nota10_12.AnoLetivo
```

Figura 40 – Query para obter as Alternate Key de cada dimensão da *SANT2\_Nota10\_12*

Com base no resultado do query é possível obter para cada registo a surrogate key de cada dimensão através dos valores dos atributos. Em cada registo do query a SK é obtida da seguinte forma para cada dimensão:

- A SK da tabela *SAANL\_AnoLetivo* é obtida a partir do valor do atributo *AnoLetivo* da tabela *Nota10\_12*;

- A SK da tabela *SAANE\_AnoEscolar* é obtida a partir do valor do atributo *AnoEscolar* da tabela *Nota10\_12*;
- A SK da tabela *SAESC\_Escola* é obtida a partir dos valores dos atributos *Escola* e *AnoLetivo* da tabela *Nota10\_12*;
- A SK da tabela *SACAL\_CaracteristicasAluno* é obtida a partir dos valores dos atributos *Sexo*, *Idade*, *Escalao* e *NEE* da tabela *AlunoCorrigido*;
- A SK da tabela *SACFA\_ContextoFamiliar* é obtida a partir dos valores dos atributos *HabilitacaoPai*, *HabilitacaoMae*, *AreaProfissaoPai* e *AreaProfissaoMae* da tabela *PaisCorrigido*;
- A SK da tabela *SADIS\_Disciplina* é obtida da seguinte forma:
  - Para o valor do atributo *Disciplina* da tabela *Nota10\_12*, é feita uma pesquisa na tabela *DisciplinaPadrao* para obter o nome padrão da disciplina através do valor do atributo *NomePadrao*.

Depois das SK serem obtidas, os registos podem ser inseridos na tabela *SANT2\_Nota10\_12* com as SK obtidas e com as medidas. As medidas são obtidas diretamente do query. O atributo *NT2\_atualizado* de cada registo tem o valor 0.

### 5.3. Estratégia de Carregamento do Data Warehouse

O carregamento do Data Warehouse é sempre feito de forma semelhante. Uma fonte de dados pode ter dados para uma escola ou para várias escolas de um agrupamento. No primeiro carregamento de uma nova fonte é necessário inserir no Data Mart os dados das escolas presentes nessa fonte. No segundo carregamento e seguintes de uma fonte são inseridos no Data Mart dados de novas escolas presentes nessa fonte caso existam. Na Figura 41 está o diagrama de sequência para carregar o Data Warehouse.

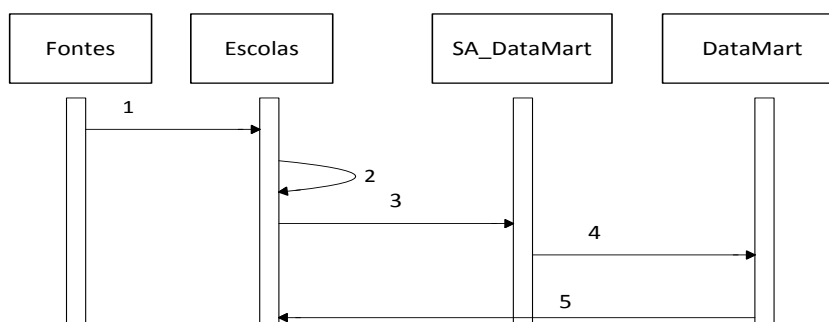


Figura 41 - Diagrama de sequência para o carregamento do Data Warehouse

A legenda das mensagens 1 a 5 da Figura 41 são as seguintes:

1. Inserção dos dados provenientes das fontes na base de dados *Escolas*;
2. Limpeza e transformação dos dados na base de dados *Escolas*;
3. Inserção dos dados da base de dados *Escolas* na base de dados *SA\_DataMart*;
4. Carregamento do Data Mart;
5. Eliminação dos registos das tabelas na base de dados *Escolas*, exceto nas tabelas *DisciplinaPadrao*, *CategoriaHabilitacao* e *AreaProfissao*.

### 5.3.1. Processos de Extração e Transformação dos dados das fontes

É necessário desenvolver programas para extrair os dados das fontes de dados para a base de dados *Escolas*. Depois dos dados estarem inseridos na base de dados *Escolas*, devem ser tratados. A extração e transformação dos dados deve respeitar os processos especificados em 5.2.1. Depois dos dados estarem tratados devem ser inseridos na base de dados *SA\_DataMart* seguindo os processos descritos em 5.2.2 pela seguinte ordem:

- Carregamento da tabela *SAANL\_AnoLetivo*;
- Carregamento da tabela *SAANE\_AnoEscolar*;
- Carregamento da tabela *SAESC\_Escola*
- Carregamento da tabela *SACAL\_CaracteristicasAluno*;
- Carregamento da tabela *SACFA\_ContextoFamiliar*;
- Carregamento da tabela *SAESA\_EstadoAvaliacao*;
- Carregamento da tabela *SADIS\_Disciplina*;
- Carregamento da tabela *SAMAT\_Matricula*;
- Carregamento da tabela *SANT1\_Nota5\_9*;
- Carregamento da tabela *SANT2\_Nota10\_12*.

As tabelas *SAANE\_AnoEscolar* e *SAESA\_EstadoAvaliacao* só são carregadas uma vez. No segundo carregamento e seguintes do Data Mart não é necessário carregar essas tabelas porque os registos dos anos escolares e estados de avaliação já se encontram nas tabelas. A tabela *SAESC\_Escola* é carregada com novos registos sempre que existam novas escolas.

### 5.3.2. Processos de Carregamento do Data Mart

A informação necessária para o carregamento do Data Mart encontra-se na base de dados *SA\_DataMart*. O carregamento das tabelas do Data Mart é feito começando pelas dimensões e terminando pelas tabelas de fatos. São inseridos os registos onde o atributo *atualizado* for igual a 0. Os atributos *atualizado* dos registos inseridos devem ficar com o valor 1 nas tabelas da base de

dados *SA\_DataMart*. Depois do Data Mart estar carregado, são eliminados os registos que estão na base de dados *Escolas*, exceto nas tabelas *DisciplinaPadrao*, *CategoriaHabilitacao* e *AreaProfissao*. Manter os registos da tabela *DisciplinaPadrao*, permite nos anos letivos seguintes atribuir um nome padrão apenas a disciplinas que não existam na tabela. Manter os registos na tabela *CategoriaHabilitacao* permite atribuir uma categoria de habilitação apenas para habilitações que não existiam na tabela. Manter os registos na tabela *AreaProfissao* permite atribuir uma área de profissão apenas para profissões que não existiam na tabela.



## 6. Discussão e Conclusão

Neste capítulo é apresentada uma discussão do trabalho e depois são apresentadas as conclusões.

### 6.1. Discussão

Com os ficheiros de dados fornecidos pelas escolas do qual obtive colaboração, foi possível verificar de que forma estavam organizados esses dados e que tipo de ficheiros existia. Com base nos esquemas dos ficheiros Excel fornecidos foi possível definir os processos de extração e transformação dos dados provenientes das fontes. Não foi possível fazer a implementação do sistema de Data Warehouse com esses dados, uma vez que para as categorias de ficheiros *F1\_Alunos*, *F1\_Pais* e *F2\_Alunos* foram obtidos apenas ficheiros exemplo o que permitiu ver o esquema desses ficheiros. Para implementar o sistema de Data Warehouse seriam necessários mais ficheiros deste tipo.

Em relação ao modelo multidimensional, este não permite verificar a evolução de um conjunto de alunos ao longo dos anos letivos, uma vez que não guarda um identificador único para o aluno. O modelo inicialmente não tinha sido pensado para fazer esse tipo de análises e quando se pensou nisso o modelo multidimensional já estava modelado. Através das reuniões feitas com membros das escolas verificou-se que em cada agrupamento escolar o aluno era identificado pelo número de processo. Se um aluno mudar de agrupamento o número de processo mudaria também. Para identificar um aluno de forma única, deveria ter sido pedido às escolas para enviar os ficheiros com dados dos alunos com um identificador único encriptado. Numa futura revisão do modelo multidimensional o aluno deve ser identificado de forma única de maneira a que o sistema de Data Warehouse permita fazer pesquisas da evolução de um conjunto de alunos ao longo dos anos letivos. Por vezes as escolas organizam as turmas tendo em conta o desempenho escolar dos alunos. Se o aluno for identificado de forma única, seria possível analisar o desempenho escolar dos melhores alunos.

### 6.2. Conclusão

Neste trabalho verifiquei que o planeamento de um sistema de Data Warehousing depende muito de uma boa caracterização dos resultados pretendidos. Se os resultados pretendidos estiverem mal especificados não será possível obter resultados de qualidade das análises feitas no Data Warehouse.

Para implementar um Sistema de Data Warehousing de qualidade a fase de planeamento é muito importante. Neste trabalho foi complicado definir sobre que perspetivas iam ser analisados os dados das escolas, pelo que foram feitas muitas reformulações do modelo multidimensional. Na fase da Implementação a base de dados *Escolas* foi modelada de maneira a inserir os dados de forma direta a partir dos ficheiros das fontes de dados. Não existem relações através de foreign keys (FK) entre essas tabelas. Existe a dúvida se esta foi a melhor forma de modelar a base de dados *Escolas*.

Nas tabelas de fatos da *SA\_DataMart* foi decidido manter os registos nas respetivas tabelas depois de estes serem inseridos no Data Mart. Em relação às dimensões esta estratégia permite inserir os registos de forma direta nas tabelas de dimensão do Data Mart. A surrogate key (SK) de um registo numa dimensão é a mesma que o registo correspondente na base de dados *SA\_DataMart*. Em relação às tabelas de fatos da base de dados *SA\_DataMart* existe a dúvida de se os registos podem ser eliminados depois de inseridos no Data Mart. Se os registos puderem ser eliminados depois de inseridos no Data Mart, o atributo que permite ver se o registo está atualizado no Data Mart já não seria necessário nas tabelas de fatos da base de dados *SA\_DataMart*.

Em relação às fontes de dados das escolas foi verificado que a mesma disciplina pode ser nomeada de forma diferente. A atribuição de uma disciplina padrão pode ser complicado caso existam muitas disciplinas. Se a Empresa tiver várias escolas aderentes e cada uma delas tiver uma forma de nomear as disciplinas, pode tornar-se demoroso atribuir um nome padrão às disciplinas. A solução seria existir uma forma padrão de nomear as disciplinas que deveria ser respeitada pelas escolas que aderem ao serviço de consultadoria privada.

Os objetivos propostos para este trabalho foram atingidos uma vez que foi feita a modelação do Data Mart com base nos resultados pretendidos e foram especificados os processos de extração, transformação e carregamento com base nos esquemas das duas fontes conhecidas.

## Referências

- [1] R. Kimball, M. Ross, W. Thornthwaite, J. Mundy e B. Becker, *The Data Warehouse Lifecycle Toolkit*, Second Edition, Wiley Publishing, Inc., 2008.
- [2] Ministério da Educação e Ciência, “Currículo e Programas Ensino Secundário, Organização Curricular,” [Online]. Available: <http://www.dgidc.min-edu.pt/ensinosecundario/index.php?s=directorio&pid=11>. [Acedido em 27 Outubro 2014].
- [3] Ginásios da Educação Da Vinci, “Exames Nacionais,” [Online]. Available: <http://www.examesnacionais.com.pt/>. [Acedido em 05 Novembro 2013].
- [4] Ministério da Educação e Ciência, “Despacho n.º 3597-A/2014,” [Online]. Available: [http://www.dgidc.min-edu.pt/jurinaconalexames/data/jurinaconalexames/legislacao/despacho\\_n\\_3597\\_a\\_2014.pdf](http://www.dgidc.min-edu.pt/jurinaconalexames/data/jurinaconalexames/legislacao/despacho_n_3597_a_2014.pdf). [Acedido em 27 Outubro 2014].
- [5] Ministério da Educação e Ciência, “Avaliação Externa das Escolas,” [Online]. Available: [http://www.ige.min-edu.pt/content\\_01.asp?BtreeID=03/01&treeID=03/01/03/00](http://www.ige.min-edu.pt/content_01.asp?BtreeID=03/01&treeID=03/01/03/00). [Acedido em 27 Outubro 2014].
- [6] Ministério da Educação e Ciência, “Relatórios do Juri Nacional de Exames,” [Online]. Available: <http://www.dge.mec.pt/jurinaconalexames/index.php?s=directorio&pid=21>. [Acedido em 27 Outubro 2014].
- [7] Público, “Ranking PÚBLICO/Universidade Católica do Porto,” [Online]. Available: <http://static.publico.pt/docs/educacao/rankings2012/>. [Acedido em 13 Novembro 2014].
- [8] Expresso, “‘Ranking’ das escolas 2013,” [Online]. Available: <http://expresso.sapo.pt/ranking-das-escolas-2013=f840093>. [Acedido em 13 11 2014].
- [9] Truncatura, “Truncatura,” [Online]. Available: <http://www.truncatura.pt/>. [Acedido em 23 Outubro 2014].
- [10] J.P.M. & Abreu, Lda, “Alunos,” [Online]. Available: <http://www.jpmapbreu.com/cgi-bin/jpmcgi.jpmpagina?id=software&num=1>. [Acedido em 20 Novembro 2014].
- [11] Truncatura, “WinGA,” [Online]. Available: <http://www.truncatura.pt/Produtos/Programas/WinGA/tabid/563/Default.aspx>. [Acedido em 23 Outubro 2014].
- [12] Truncatura, “TProfessor,” [Online]. Available: <http://www.truncatura.pt/Produtos/Programas/TProfessor/tabid/586/Default.aspx>. [Acedido em 23 Outubro 2014].

- [13] Truncatura, “Estatística e Gráficos,” [Online]. Available: <http://www.truncatura.pt/Produtos/Programas/Estat%C3%ADsticaeGr%C3%A1ficos/tabid/507/Default.aspx>. [Acedido em 23 Outubro 2014].
- [14] “Sistema MISI,” [Online]. Available: [http://www.dgeec.mec.pt/np4/179/%7B\\$clientServletPath%7D/?newsId=213&fileName=MISI\\_FAQ\\_siteDGEEC\\_20130218.pdf](http://www.dgeec.mec.pt/np4/179/%7B$clientServletPath%7D/?newsId=213&fileName=MISI_FAQ_siteDGEEC_20130218.pdf). [Acedido em 23 Outubro 2014].
- [15] Wikipedia, “Data Warehouse,” [Online]. Available: [http://en.wikipedia.org/wiki/Data\\_warehouse](http://en.wikipedia.org/wiki/Data_warehouse). [Acedido em 22 Outubro 2014].
- [16] C. S. Jensen, T. B. Pedersen e C. Thomsen, *Multidimensional Databases and DataWarehousing*, M.Tamer Özsu, University of Waterloo, 2010.
- [17] R. A. Castro, “School Performance Evaluation in Portugal: A Data Warehouse Implementation to Automate Information Analysis,” em *DSIE'11 - 6th Doctoral Symposium on Informatics Engineering*, Porto, 2011.
- [18] Y. Kurniawan e E. Halim, “Use data warehouse and data mining to predict student academic performance in schools: A case study (perspective application and benefits),” em *Teaching, Assessment and Learning for Engineering (TALE), 2013 IEEE International Conference on*, Bali, 2013.
- [19] T. Nobre, A. Trigo e P. Sanches, “SBIAES — Business intelligence system for analysis of access to higher education: The case of the Polytechnic Institute of Coimbra,” em *Information Systems and Technologies (CISTI), 2014 9th Iberian Conference on*, Barcelona, 2014.
- [20] L. Kazi, Z. Kazi e B. Radulovic, “Data warehouse based evaluation of students' achievements in information systems education,” em *MIPRO, 2012 Proceedings of the 35th International Convention*, Opatija, 2012.
- [21] Direcção-Geral do Ensino Superior, “Perguntas Frequentes. Aprovação num curso do Ensino Secundário,” [Online]. Available: <http://www.dges.mctes.pt/DGES/pt/Estudantes/Acesso/ConcursoNacionalPublico/FAQ/Exames/2.htm>. [Acedido em 04 Novembro 2014].
- [22] R. Kimbal, “Surrogate Keys,” [Online]. Available: <http://www.kimballgroup.com/1998/05/surrogate-keys/>. [Acedido em 17 Outubro 2014].
- [23] Ministério da Educação e Ciência, “Previsão do Número de Alunos no Ensino Básico e Secundário,” [Online]. Available: <http://www.dgeec.mec.pt/np4/64>. [Acedido em 21 Novembro 2014].
- [24] Ministério da Educação e Ciência, “Currículo e Programas Ensino Básico, Matrizes Curriculares,” [Online]. Available: <http://www.dgdc.min-edu.pt/ensinobasico/index.php?s=directorio&pid=6>. [Acedido em 21 Novembro 2014].

## Anexo I. Fonte 1

Fonte 1 – Ficheiro Excel da categoria *F1\_Alunos* com a lista de alunos num ano letivo

	Ministério	Educação			
	Serviço ou Organismo :		Agrupamento de Escolas XXX		
			Consulta de Alunos		
Nº de Processo		Sexo	Data de Nascimento	Nec.Educ.Especiais	
	13115	Masculino	29-09-2000	False	
	13116	Masculino	15-12-1999	False	
	13117	Feminino	13-05-2000	False	
	13118	Feminino	03-05-2000	False	
	13119	Feminino	20-10-1999	False	
	13120	Masculino	17-07-2002	False	
	13121	Masculino	19-12-1999	False	
	13122	Feminino	16-04-2000	False	
	13123	Masculino	11-05-2000	False	
	13124	Feminino	23-08-2000	False	
	13125	Feminino	20-04-2000	False	
	13126	Masculino	31-07-2000	False	
	13127	Masculino	02-11-2000	False	
	13128	Feminino	09-07-2000	False	
	13129	Feminino	04-06-1999	False	
	13130	Feminino	29-10-2000	False	
	13131	Masculino	16-09-2000	False	
	13133	Masculino	15-06-2000	False	
	13134	Feminino	30-05-2000	False	
	13135	Feminino	18-05-2000	False	
	13136	Feminino	21-06-2000	False	
	13137	Feminino	12-11-1999	False	
	13138	Feminino	31-05-2000	False	
	13139	Feminino	18-07-2000	False	
	13140	Feminino	23-06-2000	False	
	13141	Feminino	02-10-2000	False	
	13142	Masculino	31-01-2000	False	
	13143	Masculino	16-02-2000	False	
	13144	Masculino	12-10-2000	False	
	13145	Masculino	07-06-2000	False	
	13146	Feminino	18-06-2000	False	
	13147	Masculino	05-12-2001	True	
	13148	Feminino	30-05-2000	False	
	13149	Masculino	09-09-2000	False	

Figura 42 – Ficheiro Excel da *Fonte 1* com a lista de alunos no ano letivo 2013/2014



Fonte 1 – Ficheiro Excel da categoria F1\_Pais com as habilitações e profissões dos pais num ano letivo

		Ministério	Educação				
		Serviço ou Organismo :		Agrupamento de Escolas XXX			
Nº de Processo		Textohabilitacao pai	Textoprofissaopai	Textohabilitacao mae	Textoprofissaomae		
11012	Básico (2º ciclo)		Outros trabalhadores não qualificados	Básico (2º ciclo)	Outros trabalhadores não qualificados		
11013							
11017	Básico (1º ciclo)		Operador de máquinas de	Básico (1º ciclo)	Operador de máquinas de		
11018							
11019							
11020	Básico (1º ciclo)		Sem Profissão	Básico (1º ciclo)	Outros trabalhadores não qualificados		
11021	Básico (1º ciclo)		Pedreiro, calceteiro e assentador de	Básico (1º ciclo)	Outros trabalhadores não qualificados		
11022							
11023							
11025	Básico (1º ciclo)		Sem Profissão	Básico (1º ciclo)	Trabalhador não qualificado da		
11026							
11027							
11028	Formação		Sem Profissão	Sem Habilitações	Outros trabalhadores não qualificados		
11029							
11030	Básico (2º ciclo)		Operadores de caixa e venda de	Básico (1º ciclo)	Sem Profissão		
11031							
11032	Básico (2º ciclo)		Operadores de caixa e venda de	Básico (2º ciclo)	Outros trabalhadores não qualificados		
11033							
11034							
11035	Secundário		Autor e escritor				
11036	Básico (2º ciclo)		Motorista de veículos pesados de	Básico (1º ciclo)	Operador de máquinas para o fabrico		
11037							
11038							
11039	Sem Habilitações		Serralheiro de moldes, cunhos,	Básico (1º ciclo)	Outros trabalhadores não qualificados		
11040	Básico (2º ciclo)		Operadores de caixa e venda de	Básico (2º ciclo)	Operador de máquinas para preparar,		
11041							
11042	Básico (3º ciclo)		Outros trabalhadores qualificados da	Básico (1º ciclo)			
11043	Básico (1º ciclo)		Serralheiro de moldes, cunhos,	Básico (1º ciclo)	Sem Profissão		
11044	Básico (1º ciclo)		Director das indústrias transformadoras	Básico (1º ciclo)	Sem Profissão		
11045	Formação		Profissão Desconhecida	Formação	Profissão Desconhecida		
11046	Básico (1º ciclo)		Sem Profissão	Básico (2º ciclo)	Sem Profissão		
11047	Básico (3º ciclo)		Outros trabalhadores não qualificados	Básico (3º ciclo)	Outros trabalhadores não qualificados		
11048							
11053	Básico (1º ciclo)		Outros trabalhadores não qualificados	Básico (1º ciclo)	Sem Profissão		
11054							
11055							
11056							
11057							
11058							
11059							
11060	Básico (1º ciclo)		Outros operadores de instalações fixas	Básico (3º ciclo)	Lapidadores, gravadores e pintores-		
11063	Básico (1º ciclo)			Secundário			
11064							
11065							
11066							
11067							

Figura 43 - Ficheiro Excel da Fonte 1 com as habilitações e profissões dos pais no ano letivo 2013/2014

Fonte 1 – Ficheiro Excel da categoria F1\_Escalões com os alunos com um determinado escalão social

MINISTÉRIO	DGEstE - Direção de Serviços da Região : Centro		
DA			
ESCALÃO	Escola / Agrupamento :	Agrupamento de Escolas XXX	
EDUCAÇÃO E CIÊNCIA	Escola Sede :	Escola XXX	
2013 / 2014			Escalão A - 12º Ano
Listagem de Alunos Subsidiados			
Nº Proc.		Escalão AF	Observações
11470			
12040			
120018			
122873			

Figura 44 – Ficheiro Excel da Fonte 1 com os alunos que têm escalão A no 12º ano do ano letivo 2013/2014





Fonte 1 – Ficheiro Excel da categoria F1\_Turmas para uma turma do ensino básico

Agrupamento de Escolas XXX											
Escola XXX											
										3º Ciclo	
										Ano Letivo : 2013 / 2014	
										3º Período	
Registo de Avaliação											
Ano :	9	Turma :	A	Nº :	1						
Nº Processo :	11911										
Faltas Acumuladas											
Disciplina	Previstas	Dadas	Assis.		F. Inj.	F. Just.	C.F.	Prova Final	Final		
Português	33		33	33	--	--	4	4	4		
Inglês 1	20		20	20	--	--	3	--	3		
Francês 2	12		12	12	--	--	3	--	3		
História	16		16	16	--	--	3	--	3		
Geografia	17		15	15	--	--	3	--	3		
Matemática	31		31	31	--	--	4	4	4		
Ciências Naturais	19		19	19	--	--	4	--	4		
Ciências Físico-Químicas	19		19	19	--	--	4	--	4		
EDUCAÇÃO VISUAL	15		15	15	--	--	4	--	4		
Educação Física					--	--	3	--	3		
Educação Moral e Religiosa Católica		5	5	5	--	--	5	--	5		
<b>AVALIAÇÃO FINAL DE ANO/CICLO : Aprovado</b>											
Agrupamento de Escolas XXX											
Escola XXX											
										3º Ciclo	
										Ano Letivo : 2013 / 2014	
										3º Período	
Registo de Avaliação											
Ano :	9	Turma :	A	Nº :	2						
Nº Processo :	12914										
Faltas Acumuladas											
Disciplina	Previstas	Dadas	Assis.		F. Inj.	F. Just.	C.F.	Prova Final	Final		
Português	33		33	27	0	6	3	3	3		
Inglês 1	20		20	17	0	3	4	--	4		
Francês 2	12		12	11	0	1	3	--	3		
História	16		16	15	0	1	4	--	4		
Geografia	17		15	13	0	2	3	--	3		
Matemática	31		31	28	0	3	3	4	3		
Ciências Naturais	19		19	17	0	2	3	--	3		
Ciências Físico-Químicas	19		19	16	0	3	4	--	4		
EDUCAÇÃO VISUAL	15		15	13	0	2	4	--	4		
Educação Física				-2	0	2	3	--	3		
Educação Moral e Religiosa Católica		5	5	5	--	--	5	--	5		
<b>AVALIAÇÃO FINAL DE ANO/CICLO : Aprovado</b>											
Agrupamento de Escolas XXX											
Escola XXX											
										3º Ciclo	
										Ano Letivo : 2013 / 2014	
										3º Período	
Registo de Avaliação											
Ano :	9	Turma :	A	Nº :	3						
Nº Processo :	11890										
Faltas Acumuladas											
Disciplina	Previstas	Dadas	Assis.		F. Inj.	F. Just.	C.F.	Prova Final	Final		
Português	33		33	25	0	8	2	3	3		
Inglês 1	20		20	16	1	3	2	--	2		
Francês 2	12		12	12	--	--	2	--	2		
História	16		16	16	--	--	2	--	2		

Figura 45 – Ficheiro Excel da Fonte 1 do 9º A no ano letivo 2013/2014



Fonte 1 – Ficheiro Excel da categoria F1\_Turmas para uma turma do ensino secundário

Agrupamento de Escolas XXX												
Escola XXX												
Secundário												
Ano Letivo : 2013 / 2014												
3º Período												
<b>Registo de Avaliação</b>												
<b>Ano :</b>	11	<b>Turma :</b>	A	<b>Nº :</b>	01							
<b>Nº Processo :</b>	12421											
<b>Componente</b>				<b>Disciplina</b>	Prev.	Dadas	Assis.	FJ	CF/CI	CIF	Ex.	CFD
Geral	PORTUGUES				22	22	22	--	--	14	--	--
Geral	INGLES (cont)				22	22	22	--	--	14	14	--
Geral	FILOSOFIA				26	26	26	--	--	13	12	--
Geral	EDUCAÇÃO FISICA				26	26	26	--	--	15	--	--
Específica	MATEMATICA A				38	38	36	0	2	16	--	--
Específica	FISICA QUIMICA A				53	53	51	0	2	11	11	88
Específica	BIOLOGIA e GEOLOGIA				42	42	42	--	--	12	12	55
<b>Resultado Final : Transitou</b>												
Agrupamento de Escolas XXX												
Escola XXX												
Secundário												
Ano Letivo : 2013 / 2014												
3º Período												
<b>Registo de Avaliação</b>												
<b>Ano :</b>	11	<b>Turma :</b>	A	<b>Nº :</b>	02							
<b>Nº Processo :</b>	12413											
<b>Componente</b>				<b>Disciplina</b>	Prev.	Dadas	Assis.	FJ	CF/CIF	Ex.	CFD	
Geral	PORTUGUES				22	22	22	--	--	15	--	--
Geral	INGLES (cont)				22	22	22	--	--	19	19	--
Geral	FILOSOFIA				26	26	26	--	--	16	16	--
Geral	EDUCAÇÃO FISICA				26	26	26	--	--	16	--	--
Específica	MATEMATICA A				38	38	38	--	--	12	--	--
Específica	FISICA QUIMICA A				53	53	53	--	--	16	14	120
Específica	BIOLOGIA e GEOLOGIA				42	42	42	--	--	17	16	155
<b>Resultado Final : Transitou</b>												
Agrupamento de Escolas XXX												
Escola XXX												
Secundário												
Ano Letivo : 2013 / 2014												
3º Período												
<b>Registo de Avaliação</b>												
<b>Ano :</b>	11	<b>Turma :</b>	A	<b>Nº :</b>	03							
<b>Nº Processo :</b>	11962											
<b>Componente</b>				<b>Disciplina</b>	Prev.	Dadas	Assis.	FJ	CF/CI	CIF	Ex.	CFD
Geral	PORTUGUES				22	22	16	0	6	14	--	--
Geral	INGLES (cont)				22	22	16	0	6	14	15	--
Geral	FILOSOFIA				26	26	20	0	6	11	12	--
Geral	EDUCAÇÃO FISICA				26	26	24	0	2	16	--	--
Específica	MATEMATICA A				38	38	30	0	8	14	--	--
Específica	FISICA QUIMICA A				53	53	53	--	--	14	14	111
Específica	BIOLOGIA e GEOLOGIA				42	42	42	--	--	12	13	86
<b>Resultado Final : Transitou</b>												

Figura 46 – Ficheiro Excel da Fonte 1 do 11º A no ano letivo 2013/2014



Anexo II. Fonte 2

Fonte 2 – Ficheiro Excel da categoria F2\_Alunos com a lista de alunos de uma turma

AGRUPAMENTO DE ESCOLAS XXX							10º A
							ANO LETIVO :2013/2014
D.NASC.	Sexo	Necessidades Educativas	Habilitação do pai	Habilitação da mãe	Profissão do pai	Profissão da mãe	PROC.
19-04-1999	M	0	Mestrado	Licenciatura	Economista	Especialista em políticas de administração	15600
02-07-1998	F	0	Secundário	Secundário	Agente da PSP, Polícia Marítima, F	Outro pessoal de apoio de tipo administr	25008
18-05-1998	F	0	3º ciclo do Ensino Bá	3º ciclo do Ensino	Diretor e gerente de outros serviços	Operadores de caixa e vendas de bilhetes	24901
19-04-1998	F	0	2º ciclo do Ensino Bá	Secundário	Desconhecida	Comerciante de loja(estabelecimento)	15728
02-04-1998	F	0	Formação Desconhec	2º ciclo do Ensino	Desconhecida	Trabalhadora de limpeza em escritórios, I	15604
22-09-1998	F	0	Doutoramento	Doutoramento	Professor dos ensinos universitário	Professor dos ensinos universitário e sup	15606
10-09-1998	F	0	Licenciatura	Licenciatura	Professor do ensino básico(2º e 3º c	Professor do ensino básico(2º e 3º ciclos)	15607
14-04-1998	F	0	Secundário	Licenciatura	Avaliador de imóveis, seguros e out	Professor do ensino básico(2º e 3º ciclos)	24807
31-03-1998	M	0	Formação Desconhec	Formação Descon	Desconhecida	Desconhecida	24835
10-10-1998	M	0	Formação Desconhec	Formação Descon	Desconhecida	Desconhecida	15774
20-02-1998	F	0	Formação Desconhec	Formação Descon	Desconhecida	Contabilista, auditor, revisor oficial de cc	15624
06-06-1998	M	0	Mestrado	Mestrado	Outros analistas e programadores d	Professor do ensino básico(2º e 3º ciclos)	15611
24-04-1998	M	0	Licenciatura	Mestrado	Designers de produto, têxteis, moda	Diretores de outros serviços especializad	15612
11-12-1998	M	0	2º ciclo do Ensino Bá	1º ciclo do Ensino	Cozinheiro	Sem profissão	24808
18-01-1998	F	0	Mestrado	Mestrado	Diretor Financeiro	Professor do ensino básico(2º e 3º ciclos)	24811
20-02-1998	F	0	Secundário	Desconhecida	Licenciatura	Professor do ensino básico(2º e 3º ciclos)	15759
12-11-1998	F	0	Licenciatura	Licenciatura	Oficial do Exército(NE)	Professor do ensino básico(2º e 3º ciclos)	15793
26-07-1998	F	0	Formação Desconhec	Formação Descon	Desconhecida	Educadora de infância	24834
23-08-1998	F	0	Formação Desconhec	Formação Descon	Desconhecida	Arquiteta de edifícios	15751
04-09-1998	F	0	Formação Desconhec	Formação Descon	Desconhecida	Profissão Desconhecida	15615
20-04-1998	M	0	Doutoramento	Doutoramento	Professor dos ensinos universitário	Professor dos ensinos universitário e sup	15678
06-05-1998	M	0	Licenciatura	Licenciatura	Professor do ensino básico(2º e 3º c	Professor do ensino básico(2º e 3º ciclos)	15617
10-02-1998	M	0	Licenciatura	Licenciatura	Outros operadores de instalações fi	Outros agentes de nível intermédio da ad	19140
11-10-1998	F	0	Formação Desconhec	Formação Descon	Desconhecida	Profissão Desconhecida	15769
06-05-1998	F	0	Formação Desconhec	Licenciatura	Desconhecida	Especialista do trabalho social	15777
23-10-1998	F	0	Secundário	Licenciatura	Agente da administração tributária	Professor do ensino básico(2º e 3º ciclos)	15722
03-09-1998	F	0	Doutoramento	Licenciatura	Professor dos ensinos universitário	Professor do ensino básico(2º e 3º ciclos)	24810
10-06-1998	M	0	2º ciclo do Ensino Bá	3º ciclo do Ensino	Guarda dos serviços prisionais	Guarda dos serviços prisionais	15782
02-07-1998	F	0	Formação Desconhec	Formação Descon	Eletromecânico, electricista e insta	Profissão Desconhecida	15787
09-04-1998	F	0	Doutoramento	Doutoramento	Professor dos ensinos universitário	Professor dos ensinos universitário e sup	15619

Figura 47 – Ficheiro Excel da Fonte 2 com a lista de alunos do 10º A no ano letivo 2013/2014



Fonte 2 – Ficheiro Excel da categoria *F2\_Escalões* com os alunos com um determinado escalão social

<b>MINISTÉRIO</b>	DGEstE - Direção de Serviços da Região :		Centro		
<b>DA</b>					
	Escola / Agrupamento :		Agrupamento de Escolas XXX		
<b>EDUCAÇÃO E</b>					
<b>CIÊNCIA</b>	Escola Sede :		Escola XXX		
<b>2013 / 2014</b>					Escalão A - 10º Ano
			<b>Listagem de Alunos Subsidiados</b>		
<b>Nº Proc.</b>			Escalão AF		Observações
	15638				
	23972				
	24808				
	24812				
	24829				
	24915				
	24916				
	24939				
	24940				
	24941				
	24942				
	24964				
	25045				
	26016				

Figura 48 – Ficheiro Excel da Fonte 2 com os alunos que têm escalão A no 10º ano do ano letivo 2013/2014













### Anexo III. Base de dados *Escolas* – Modelo Lógico

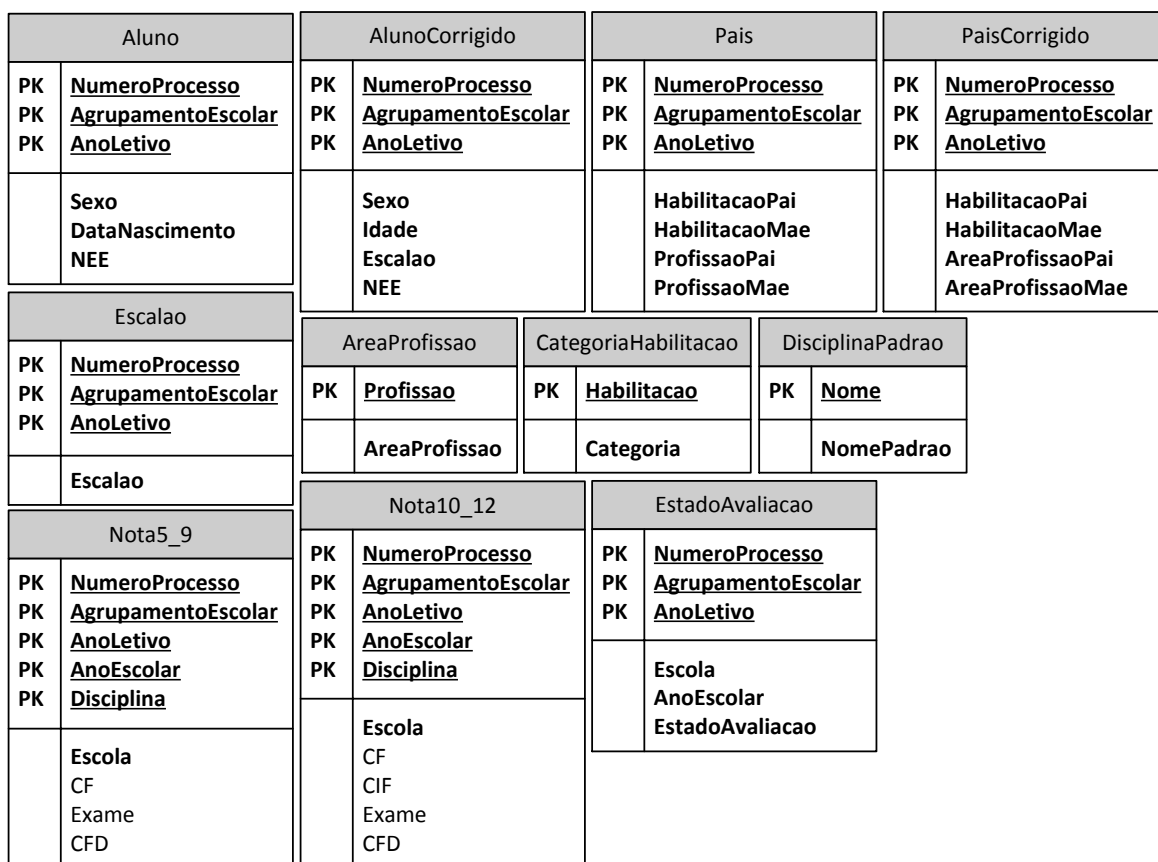


Figura 51 - Modelo lógico da base de dados *Escolas*