Universidade de Aveiro
2012

Departamento de Eletrónica,
Telecomunicações e Informática

Sónia Alexandra Resende de Pinho

APLICAÇÃO DE TÉCNICAS DE BI À INFORMAÇÃO DE IDENTIDADE

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia de Computadores e Telemática, realizada sob a orientação científica do Prof. Doutor Francisco Manuel Marques Fontes, Professor Auxiliar Convidado do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro

o júri

presidente

Prof. Doutor Joaquim João Estrela Ribeiro Silvestre Madeira
Professor auxiliar do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro

Prof. Doutor Pável Pereira Calado
Professor auxiliar do Instituto Superior Técnico de Lisboa

Prof. Doutor Francisco Manuel Marques Fontes
Professor auxiliar convidado do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro

agradecimentos

Gostaria de expressar os meus agradecimentos a todos aqueles que tornaram possível a realização desta dissertação.

Ao Prof. Doutor Francisco Fontes pelo apoio como orientador, pelos conselhos e por toda a disponibilidade que demonstrou ao longo desta dissertação.

Ao meu colega Ricardo Azevedo Pereira pelos ensinamentos que me transmitiu, pela informação disponibilizada e também pelo incentivo em todos os momentos deste trabalho.

Aos meus colegas Regina Julião, Fernando Bastos, Paula Cravo e Helena Correia pelo apoio e profissionalismo demonstrado e pelos ensinamentos na área de BI, que ajudaram a enriquecer esta dissertação e ao Filipe Rodrigues pelos conselhos na fase final desta dissertação.

À minha amiga Adriana Paula pela paciência que sempre teve comigo e por todas as palavras de incentivo que me disse quando mais precisava. De modo especial ao meu marido, aos meus filhos e aos meus pais porque sempre estiveram do meu lado e me incentivaram a concluir este trabalho.

A todos vós o meu sincero obrigado.

Sónia Pinho

palavras-chave ***business intelligence, data warehouse, personally identifiable information*, identidade, gestão de acesso, inteligência, *auditing*, *reporting*.**

resumo

Na sociedade de hoje uma pessoa tem muitas identidades, consoante o serviço ou organização onde os seus dados são disponibilizados. Gerir de forma segura e efetiva todas a identidades é uma tarefa difícil mas de extrema importância. A informação de identidade pessoal ou PII (*Personally Identifiable Information*) é o objeto deste estudo. A possibilidade de relacionar esta informação e gerar mais informação útil é uma oportunidade a não perder. Neste trabalho faz-se uma incursão pelo *Business Intelligence* e tecnologias associadas tendo em vista a sua aplicação ao tratamento seguro da informação de PII gerada nos sistemas que gerem a mesma, os sistemas de IAM. Isto é, verificar a possibilidade de dotar um sistema de IAM de inteligência e *auditing* de forma a detetar situações anómalas e diligenciar a sua resolução. Este estudo passou ainda pela investigação de algumas ferramentas de IAM já existentes e pelas suas funcionalidades de *auditing.* Os sistemas em estudo não dão resposta eficiente ao problema da resolução das falhas detetadas de forma rápida e eficaz, pelo que se propõe aqui uma arquitetura e modelo de dados que será a base de todo o sistema de *auditing* de um IAM e permitirá o processamento e relacionamento da informação pessoal gerada por este. Com a introdução de processos automáticos dessa informação é possível a resolução de alguns problemas demonstrados nos casos de uso apresentados.

keywords **business intelligence, d*ata warehouse*, personally identifiable information, identity, access management, intelligence, *auditing*, reporting.**

abstract In today's society a person has many identities, depending on the service or organization, where their data is available. To effectively and safely manage all these identities is a hard task but also an extremely important one. The personal identity information or PII (*Personally Identifiable Information*) is the subject of the presented study. The possibility to use this type of information and to link it together is an opportunity that should not be missed. This work is an incursion through the Business Intelligence and through the related technologies aiming to apply that knowledge in the secure processing of the PII generated by the systems which manage the identities, the IAM systems. In other words, this work aims to study the possibility of including intelligence and auditing, in the IAM systems, to detect abnormal behaviours and to resolve that as quickly as possible. During this study there was some investigation on the IAM tools already existents, their features and auditing. The studied tools do not respond effectively to the problem of detecting errors and failures and resolve them quickly and for good. For that reason architecture and data model which will be the basis of all the IAM auditing system, is suggested in this study and it will permit the process and link the personal data generated by the IAM system. With the introduction of automatic processing of that data it will be possible to resolve some of the problems demonstrated in the specified use cases.

Índice

Lista de figuras

1. Introdução

1.1. Motivação

Nos dias de hoje os temas da segurança e privacidade tornaram-se aspetos de extrema relevância. A implementação de soluções para abordar esta problemática é cada vez mais comum. Cada pessoa tem, perante diferentes entidades, uma identidade própria, um conjunto de atributos pessoais ou profissionais, que a identifica e caracteriza perante essas entidades. Este tipo de informação é definido como sendo ***Personally Identifiable Information*** ou PII. Para gerir toda esta panóplia de identidades, é crescente a necessidade de um conjunto de soluções ou, idealmente, a opção por uma só solução convergente de **Gestão de Identidades** e que garanta o cumprimento das normas internacionais. Uma ferramenta de Gestão de Identidades deverá preocupar-se em resolver os problemas que possam surgir relacionados com autenticação, autorização, auditoria, mobilidade e convergência físico/virtual. Associado a esta solução está sempre a questão da segurança, quer no armazenamento, quer na transferência de dados/atributos entre diferentes entidades. Assim, claramente se pode deduzir que qualquer que seja a aplicação, o aspeto segurança nunca deve ser esquecido.

O uso destas soluções gera informação, relacionada com a identidade de cada um, a que chamamos PII. Estes dados podem conter diversa informação, como o registo de entrada e saída de um utilizador num qualquer sistema, até aos registos de alteração de políticas de privilégios, entre outra informação. Contém, na maior parte das vezes, dados relacionados com a identidade pessoal ou profissional, a qual deve sempre ser salvaguardada, mas que poderá ser de grande utilidade. Tal como em outras áreas, esta informação poderá ser útil à obtenção de métricas que podem influenciar decisões e que podem ser usadas para manutenção de determinado nível de segurança. Informação pessoal que permite detetar aspetos de fraude, delinear perfis de indivíduos e de comportamentos são hoje informações muito valiosas e que já várias discussões suscitaram na legalidade do seu uso para fins comerciais, entre outros. No entanto, a sua utilização, respeitando os aspetos legais aplicáveis, é algo que suscita interesse. Toda a informação pode ser valiosa e quando se conseguem relacionar várias informações e atividades, os dados gerados por esta correlação são ainda mais interessantes do ponto de vista do negócio e da segurança.

O ***Business Intelligence*** (BI) comporta técnicas usadas já há muito tempo para análise e extração de informação a partir de outra informação armazenada. O BI permite estruturar informação e armazená-la de forma diferente, com o objetivo de extrapolar tendências e padrões. O BI é naturalmente orientado às decisão de gestão mas, o racional que suporta

as técnicas de BI, poderá também ser utilizado para inferir conhecimento a partir de outros tipos de informação como a PII.

1.2. Objetivos

O objetivo principal deste trabalho é estudar, analisar e descrever diferentes técnicas/tecnologias de BI e, associando-as à informação gerada pelas soluções de Gestão de Identidades, verificar a possibilidade de inferir outro conhecimento a partir desta informação, aplicável na área do PII. O conhecimento que se pretende inferir cobre várias áreas, como comportamentos esperados e não esperados, identificação de perfis de utilizador, deteção de fraude e inteligência associada aos *workflows*.

O trabalho passará pelo estudo do estado atual do BI e tendências observáveis. Deverá contemplar:

- análise de algumas ferramentas mais conhecidas de BI e perceber se a sua utilização é possível e viável, para o processamento de PII.
- estudo de conceitos e práticas associadas à implementação de um sistema de BI e a transposição destes métodos para os dados a analisar
- estudo das limitações legais que poderão existir ao processamento de informação do tipo PII.
- identificação de alguns casos de uso
- elaboração de uma arquitetura e modelo que suportem os casos identificados

Neste trabalho são propostos quais os dados a serem registados pela aplicação IAM (*Identity and Access Management*) da Portugal Telecom Inovação, para satisfazer os casos de uso aqui identificados.

Neste sentido, este estudo terá como principal contribuição para o tema do tratamento de PII, a apresentação de casos de estudo que devem ser desenvolvidos e implementados na aplicação de Gestão de identidades desenvolvida pela PT Inovação. Estes casos de estudo terão maior foco na introdução de inteligência associada ao processo de análise de dados pessoais. Será feita uma análise de aplicações semelhantes que já existam no mercado tendo em vista a definição de requisitos para a referida aplicação, para que esta possa trazer valor acrescentado ao mercado através da inteligência.

1.3. Estrutura da dissertação

O trabalho desenvolvido foi estruturado na presente dissertação que se inicia com uma breve introdução em que se descreve a motivação do estudo e os objetivos definidos para este.

A secção designada por estado da arte, começa por apresentar uma definição do que é *Business Intelligence*, desenvolvendo-se na descrição breve da sua história e evolução. Descrevem-se o que são os sistemas de BI e quais os processos de melhoria contínua que lhe estão associados. Apresentam-se breves descrições de algumas aplicações empresariais mais usadas nesta área.

Segue-se a descrição do que se entende por PII, e define-se genericamente o que se entende por IAM e por *Identity Intelligence*. Apresentam-se nesta fase da dissertação algumas das ferramentas comerciais que já existem nesta área. Faz-se ainda referência aos trabalhos de normalização que têm sido desenvolvidos no âmbito do TM Forum.

No capítulo seguinte, descrevem-se os aspetos legais associados ao processamento de informação de carater pessoal.

Após esta descrição, o capítulo que se segue dedica-se à apresentação de casos de uso possíveis para introdução de inteligência num sistema de IAM. Os casos identificados e apresentados são alguns dos possíveis e talvez os mais comuns, sendo que no decorrer da posterior implementação do sistema de *auditing* por parte de uma empresa/organização, com certeza irão surgir muitos mais. Não é aqui pretendida a identificação exaustiva de casos, mas antes a exemplificação de como aplicar as técnicas estudadas na melhoria de um sistema de IAM a partir do processamento e correlação da informação PII que a organização detém, tornando-o uma mais-valia para a segurança e performance dos processos da empresa ou organização.

Dedica-se ainda uma secção à apresentação de uma possível arquitetura para o sistema e de um modelo de dados que suporte o mesmo. O modelo proposto é descrito e exemplifica-se, tendo por base alguns dos casos acima descritos, a sua utilização. Aqui faz-se também uma breve distinção entre BI e *Complex Event Processing*, que também poderá ser usado como peça da arquitetura proposta, suportando alguns dos casos descritos.

A última parte desta dissertação apresenta a conclusão do estudo efetuado e a descrição do que pode ser o trabalho de continuidade deste projeto. Segue-se a lista de definições, abreviaturas e referências.

2. Conceitos e estado da arteBusiness Intelligence (BI)

Quando se fala de *Business Intelligence* (BI) normalmente referimo-nos a um conjunto de técnicas informáticas que identificam, extraem e analisam dados de negócio, com o objetivo de, normalmente, gerarem mais negócio. Permitem ter uma visão atual e/ou histórica dos dados, e muitas vezes fazer predição sobre estes.

O BI compreende tecnologias que incluem o processamento analítico de dados, o *Data Mining* e *Process Mining*, o *Complex Event Processing* (CEP), a análise da performance de gestão, o *Benchmarking* e a análise preditiva.

Todas estas tecnologias têm como principal foco o suporte à decisão ponderada e informada. É por isso que muitas vezes o BI é chamado de **Sistema de Suporte à Decisão** (DSS - *Decison Support System*), mas um sistema de BI é mais do que isso, sendo uma evolução dos *DSS's*. O BI pode muitas vezes ser referido também como *Competitive Intelligence*. No entanto podemos entendê-lo mais largamente como incluindo este tipo de *inteligência*, uma vez que o *Competitive Intelligence* se foca na análise de dados das empresas concorrentes [1].

Nos dias de hoje o BI é usado em inúmeras empresas de diferentes ramos de atividade, para melhorar a tomada de decisão, reduzir custos e identificar novas oportunidades de negócio. O BI é mais do que um conjunto de ferramentas de relatórios e um meio de relacionar dados. O BI pode ser hoje utilizado para identificar processos de negócio ineficientes e aplicar processos de reengenharia aos mesmos, com o intuito de os melhorar. Um facto a ter sempre presente é que os dados que alimentam os procedimentos de BI devem estar "limpos" e ser consistentes, para que o resultado obtido seja fiável.

Durante alguns anos, as discussões em volta do BI focavam-se fundamentalmente nas técnicas usadas para OLAP (*Online Analytical Processing*), *data mining* e *Data Warehouse*. No entanto muito pouco foco era dado à questão de como criar e implementar BI nas empresas. Num mundo cada vez mais concorrencial, é importante reagir de forma rápida e eficaz às mudanças no mercado e responder às necessidades dos clientes antecipadamente. O poder de decisão torna-se muito importante num cenário como este. O BI aplicado às empresas dos dias de hoje permite a exploração inteligente, integrada e multidimensional dos dados originados pelos diferentes sistemas de TI da empresa. Os sistemas de BI são responsáveis por transformar dados em informação e conhecimento, permitindo pensar estrategicamente. A implementação de um sistema de BI permitirá obter uma gestão efetiva dos dados e a satisfação das necessidades da empresa e dos seus clientes, a maior parte das vezes visando a redução de custos e criação de mais negócio.

2.1.1. História

O termo *Business Intelligence* foi usado pela primeira vez pelo investigador da IBM, Hans Peter Luhn. Este definiu *inteligência* como sendo: "*the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal*"[2]

O BI, como o entendemos hoje, surgiu dos Sistemas de Suporte à Decisão na década de 60 evoluindo durante a década de 80. A partir destes sistemas, no final dos anos 80, começaram a surgir *Data Warehouses*, *Executive Information Systems* (EIS), OLAP e BI.

Mais tarde em 1989, Howard Dresner, vice-presidente e investigador na Gartner Research, propôs que o termo BI se aplicasse ao conjunto de conceitos e métodos existentes para melhorar a tomada de decisão, baseada em factos. A partir da década de 90 o termo foi ganhando mais adeptos e é hoje usado amplamente.

Os sistemas de BI surgem dos Sistemas de Gestão de Informação (*Management Information Systems*) e evoluíram durante anos até às estratégias e implementações de BI de hoje em dia. Todos os seus predecessores não faziam uso da dependência de diferentes dados entre si, não havendo assim lugar à interpretação dos mesmos segundo um contexto, o que por sua vez dificulta a descoberta de novas interdependências. Estas foram as principais razões que levaram à evolução até chegarmos ao conceito de BI. Técnicas pouco apropriadas de aquisição de dados, análise, descoberta e visualização dos mesmos foram sendo "abandonadas" progressivamente. A figura seguinte ilustra a evolução dos sistemas de gestão de informação.

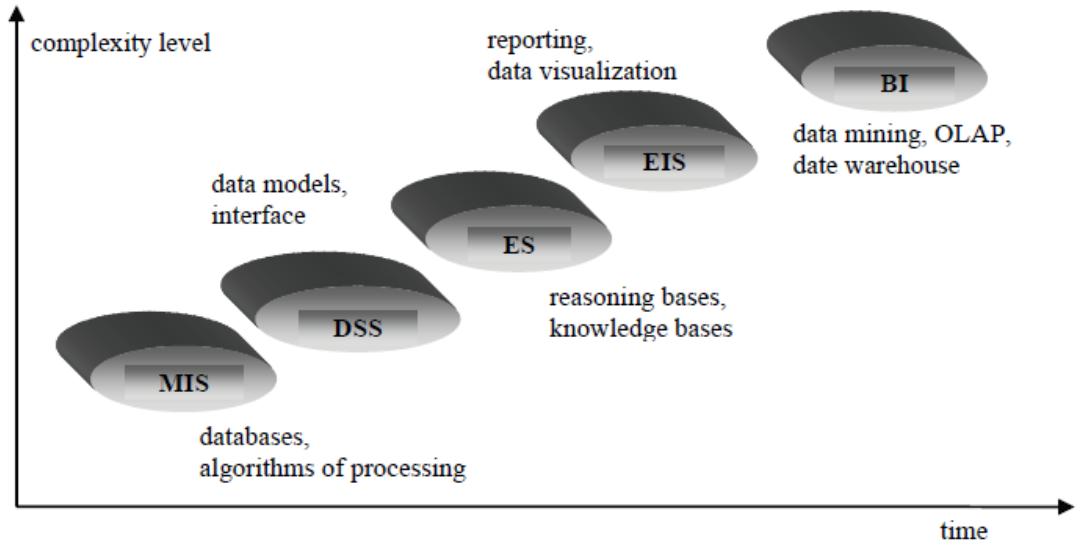


Figura 1. Evolução dos sistemas de gestão de informação [3]

2.1.1.1. BI 2.0

O BI não está parado mas continua a evoluir. Interfaces mais ricos, análise em memória (*in-memory analytics*), diferentes visualizações, tecnologias emergentes no que respeita ao armazenamento e acesso à informação, são áreas interessantes e desafiantes. É a era do BI 2.0, desenvolvida a partir de meados do ano 2000. Especialistas nestas áreas afirmam que o BI 2.0 deve incluir fatores como:

- relatórios e dados incorporando informação de contexto aos mesmos,
- dados com maior ligação às ações que deles derivam,
- acesso mais direto à informação,
- monitorização da tomada de decisão com o objetivo de associar as decisões à tática de negócio e ao contexto em que foram observadas as medidas,
- modelos mais intuitivos de apresentação de informação,
- capacidade de deteção de padrões complexos através de rotinas automáticas ou sistemas inteligentes,
- procura de informação mais facilmente e
- ligações a conteúdos não estruturados como documentos e discussões.

O BI2.0 implica a mudança das tradicionais *Data warehouses* que as ferramentas de BI utilizam, o que abrirá o caminho à introdução de contexto, contingência e a necessidade de relacionar informação proveniente de várias fontes, rapidamente.

O BI 2.0 torna-se mais popular devido à popularização das arquiteturas orientadas ao serviço (SOA - *Service-Oriented Architectures*) que permitem uma maior flexibilidade.

O crescimento das redes sociais tem criado uma grande quantidade de informação indexável e possivelmente muito útil. As redes de sensores engrossam esta informação com mais informação. Prestar mais atenção a estes dados é algo que não deve ser esquecido. Surge o BI 2.0 que deve o seu nome, em parte ao *Web* 2.0 que se foca no utilizador e colaboração em comunidade. O objetivo fundamental do BI 2.0 é a redução da latência, o tempo que decorre desde que um evento ocorre até que a reação a este seja tomada, de forma a melhorar a performance do negócio.

No BI 2.0 os dados não estão armazenados em *Data Warehouses* nem são extraídos para depois serem analisados. O resultado destes sistemas são tipicamente métricas em tempo-real e alertas que originam uma ação imediata. O BI 2.0 necessita de processos inteligentes que sejam direcionados ao evento e em tempo-real. Os sistemas devem ser capazes de processar os eventos assim que eles sucedem, compará-los com os eventos passados ou gerar previsões de eventos futuros. Acima de tudo estes sistemas devem ser escaláveis pois a quantidade de eventos não é previsível.

O BI 2.0 representa uma nova visão e estende o BI para além das *Data Warehouses* e ferramentas de *querying*. O BI 2.0 pretende criar processos mais inteligentes, de forma a procurar possíveis erros e tomar as ações necessárias para os corrigir, melhorando o desempenho [4].

2.1.2. Os sistemas de BI

A procura por sistemas de BI tem continuado a crescer apesar de a procura por *software* de IT (*Information Technology*) ter estabilizado.

Os sistemas de BI combinam a recolha de dados, o seu armazenamento e a gestão de conhecimento com ferramenta analíticas para apresentar informação complexa e competitiva aos responsáveis de decisão numa empresa. Estes sistemas permitem obter informação importante em tempo útil para gerar a "reação" aos factos representados. Um sistema de BI é um sistema proactivo que essencialmente é composto por componentes também eles proactivos, tais como:

- Dados em tempo real provenientes de *Data warehouses* e *data mining*
- Deteção automática de anomalias e exceções

- Sistemas de alertas
- Aprendizagem automática
- Visualização de dados

As fontes de informação que alimentam um sistema de BI podem ser provenientes dos mais diversos sistemas e serem apresentadas nos mais diversos formatos com o propósito de auxiliar na decisão.

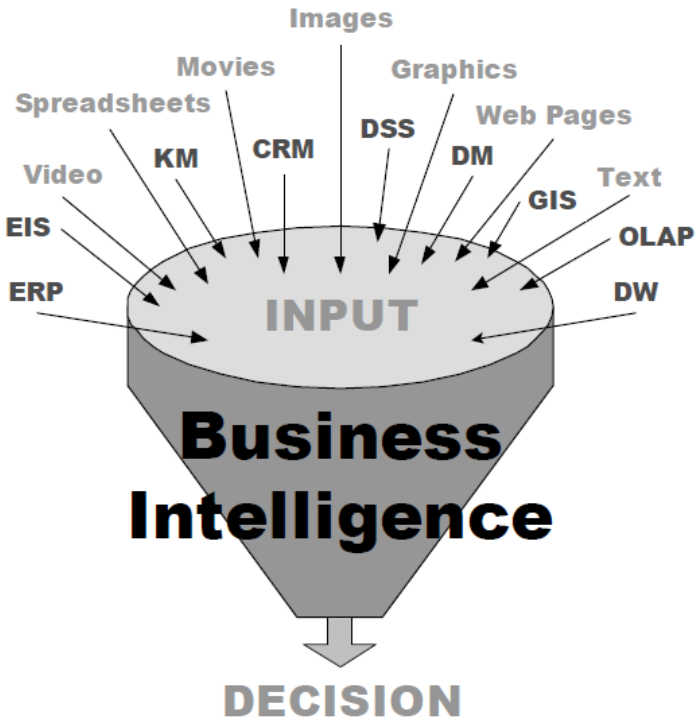


Figura 2. Inputs de um sistema de BI [5]

Várias são as definições dadas para um sistema de BI e suas aplicações, mas após um breve estudo percebe-se que as várias definições identificam essencialmente três aspetos num sistema de BI: aspetos de gestão, de tecnologia e de produto. Os aspetos relacionados com gestão ou processo, têm o seu principal foco no processo de recolha de dados de fontes internas ou externas na sua análise de forma a gerar informação relevante ao processo de decisão. O aspeto tecnológico de um sistema de BI centra a sua definição nas ferramentas e tecnologias que permitirão armazenar, recuperar, manipular e analisar a informação. Por último o aspeto relacionado com produto nas definições de BI é descrito como sendo o resultado emergente de uma análise detalhada aos dados de negócio e as práticas de análise utilizadas usando ferramentas de BI. Desta perspetiva o BI é inevitavelmente visto como um produto, resultante de um processo avançado de análise de informação e conhecimento, que suporta o desempenho e a decisão [6].

Assim, o objetivo de um sistema de BI é permitir aos gestores e analistas de todos os níveis aceder prontamente à informação e dar início à sua manipulação e análise de forma a melhorar a qualidade e o tempo de decisão. Um sistema de BI facilita a partilha de informação.

Um sistema de BI é normalmente composto por um conjunto de três tecnologias de gestão, nomeadamente, *(i) Data Warehousing*, (ii) OLAP e (iii) descoberta de conhecimento, que é predominantemente auxiliada por técnicas de *data mining*. Os componentes principais são:

- ✓ As ferramentas de ETL (*Extract*, *Transform* e *Load*), que são responsáveis pela transferência de dados dos sistemas operacionais para as *Data warehouses*.
- ✓ As *Data warehouses,* que disponibilizam o espaço para armazenamento dos dados agregados e analisados.
- ✓ As ferramentas de OLAP que analisam e modelam, permitindo o acesso e partilha da informação armazenada.
- ✓ Ferramentas de *data mining* que permitem determinar padrões, generalizações, regularidades e regras nos dados.
- ✓ Ferramentas de *reporting* e *querying* para criar diferentes tipos de relatórios.
- ✓ Ferramentas de apresentação com gráficos e interfaces multimédia para disponibilizar aos utilizadores a informação de uma forma confortável e acessível.

Para assegurar uma implementação bem-sucedida de um sistema de BI, deve-se sempre ter em conta o negócio nuclear da empresa e também os potenciais benefícios que o sistema pode trazer ao negócio. Deve-se avaliar detalhadamente todas as ferramentas existentes no mercado e decidir em conformidade com os objetivos a atingir.

2.1.3. O Ciclo de vida do BI

Desenvolver BI numa empresa, é uma estratégia de longo prazo e de continuidade. Se o negócio muda, a estratégia analítica e a tecnologia que a suportam têm de acompanhar essa evolução.

Podemos identificar quatro fases no desenvolvimento de uma estratégia de BI:

- Recolha da informação

- Reporting
- Análise
- Visualização

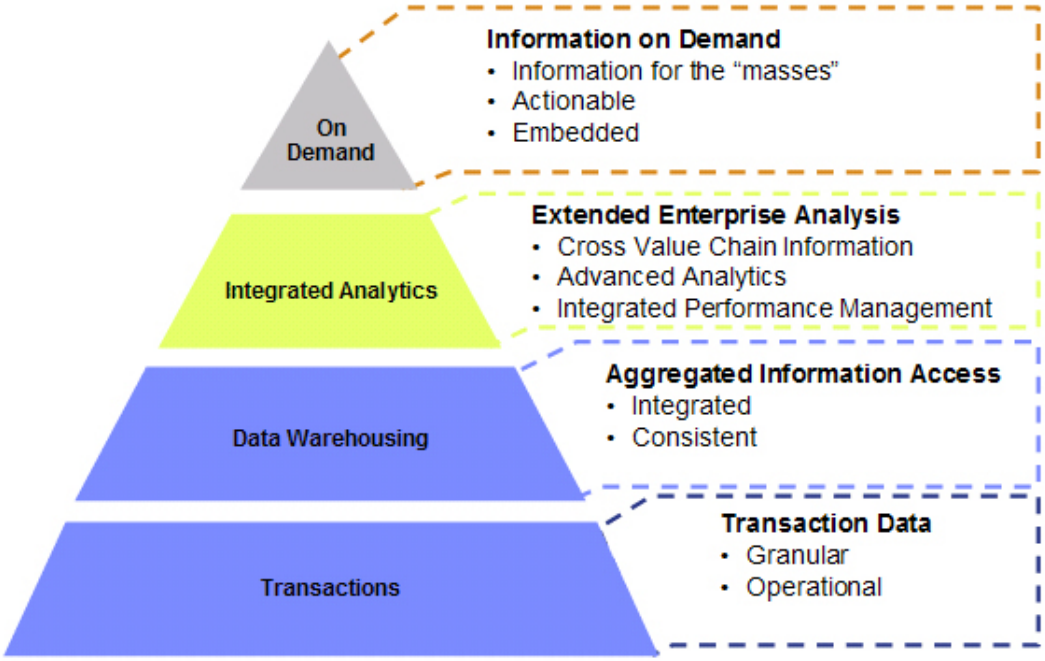


Figura 3. Dos dados à informação [7]

Na primeira fase, a de recolha de informação, os dados podem ser obtidos a partir de ferramentas de CRM (*Customer Relationship Management*), de ferramentas de desenvolvimento ou outras semelhantes como os ERP (*Enterprise Resource Planning*). Os dados podem ainda ser simplesmente registos de atividade das aplicações (*logs*). Em termos de apresentação dos dados, estes podem, nesta fase, estar guardados em bases de dados transacionais e serem alvo de operações OLTP (*Online Transaction Processing*) ou apenas em ficheiros de dados não estruturados.

Na fase seguinte, o *reporting* vai tratar os dados e criar uma forma ainda simples de os apresentar sem grandes preocupações no que respeita à sua análise. Este processamento de *reporting* pode ser direcionado ao parâmetro (tendo em conta este ou aquele parâmetro em particular) podendo ser distribuído via *e-mail* ou até por consultas *Web*. Os dados apresentados são sumarizados e derivados de um processamento básico dos mesmos. Aqui os dados apresentam-se mais uma vez na forma de bases de dados transacionais havendo preocupação pela sua replicação e utilização de modelos. É a fase de integração dos dados.

A terceira fase corresponde à fase de análise. Os dados recolhidos vão ser alvo de uma análise profunda. Nesta fase poderão ser feitas *queries* e análises às BD's (bases de dados) por parte do utilizador final. Em termos de análise dimensional, é nesta fase que se inicia o OLAP e se faz a transformação dos dados, integração dos mesmos e seu perfil histórico. Os dados são apresentados com um desenho dimensional dos mesmos, em *Data warehouses* ou usando uma semântica de cubos dimensionais. Nesta fase faz-se o armazenamento de dados em *Data warehouses*.

Por último, os dados analisados e reorganizados, serão visualizados na forma de métricas, alertas, *scoreboards* (com estratégias e objetivos) e do ponto de vista de uma análise avançada (com opção pelas operações de *data mining*). Os dados apresentam-se com um semântica complexa em *Data warehouses* avançadas ou também sob a forma de repositórios de métricas. Esta é a fase de apresentação e acesso aos dados tratados.

Sempre que uma empresa visa adotar um estratégia de BI, deverá produzir um documento em que se foque nos objetivos particulares de BI para aquela empresa. Este documento vai permitir planear de acordo com a estratégia pretendida. Este documento deve incluir:

1. a visão conceptual da estratégia (complementada com um diagrama que permita ter uma visão geral de como a tecnologia de BI se integra e forma um todo ajudando na definição e implementação da estratégia),
2. a arquitetura de dados (em que se especifica a forma como os dados vão estar organizados, guardados e como serão distribuídos),
3. a arquitetura técnica (corresponde aos componentes físicos que compõem o ambiente de BI; cada componente deve ser descrito detalhadamente) e
4. a descrição da implementação usando as diferentes tecnologias disponíveis (definindo-se tempos de implementação, processos de core e implementação das estruturas de dados).

2.1.4. Melhoria contínua em BI

Um sistema de BI não termina com a sua implementação. Um procedimento de melhoria contínua vai permitir manter o sistema atual e sincronizado com os objetivos da empresa em cada momento. Um sistema de BI saudável é a soma de quatro processos que constituem o seu ciclo: medir, analisar, planear, melhorar [8]

Figura 4. BI Improvement Cycle [9]

Medir

A fase de medição, ou observação, é a mais importante de todo o processo e a mais desenvolvida. Nesta fase as empresas reportam os dados históricos e correntes das suas métricas chave, respeitantes ao seu negócio. Estas métricas dão uma perspetiva do estado de negócio, da saúde da empresa. Embora a maior parte das empresas saiba o que medir, que indicadores usar, não será fácil obter e distribuir esta informação à sua organização e colaboradores. Implementando um sistema de BI, uma empresa/organização poderá distribuir com sucesso esta informação a todas as pessoas a que o negócio afete, seja dentro ou fora da empresa. Através do seu sistema de BI uma empresa pode divulgar indicadores e métricas que proporcionem uma vista mais aprofundada do seu negócio e que podem ser alterados para melhorar o que é atualmente medido. Hoje em dia os sistemas de BI usados permitem obter uma enorme quantidade de informação com indicadores importantes, sendo então esta a fase inicial de qualquer processo de BIIC (*Business Intelligence Improvement Cycle*). Durante esta fase, e aquando da recolha da informação, poderão ser detetadas inconsistências de dados quando estes são agregados. Esta fase é pois também importante porque permite detetar estas falhas e harmonizar o processo, permitindo a evolução do sistema. Sem esta harmonização, partir para as outras fases do processo não faz sentido, uma vez que trabalhar sobre dados e análises "suspeitas" não será a melhor aproximação.

Analisar

A segunda fase é a da análise e durante esta fase os analistas reveem e analisam os dados de várias perspetivas diferentes, tentando encontrar relações entre estes, que não estejam patentes à primeira vista. Na evolução dos sistemas de BI, várias ferramentas surgiram para simplificar este processo (*queries* à bases de dados, OLAP, visualização de

dados, etc). Estas ferramentas vieram então permitir que o trabalho dos analistas fosse simplificado.

Planear

Depois de determinar algumas das razões pelas quais as coisas ocorrem, durante a terceira fase, as empresas vão tentar determinar os efeitos nos seus resultados (*outcomes*) e se devem ou não ocorrer mudanças. Nesta fase são usadas ferramentas de simulação que permitem testar a condição "e se fizéssemos isto…", com os dados. Aqui serão desenhados vários cenários possíveis. As aplicações para esta fase de BIIC, são denominadas de 'planeamento', 'orçamentação' e 'previsão'. Usando estas ferramentas, conseguimos, por exemplo, construir um plano ou conjunto de medidas baseadas em comportamentos esperados que, combinados, com dados históricos nos permitirão prever como o lucro será afetado.

Melhorar

A fase de planeamento, naturalmente, progride para a melhoria. Nesta fase os gestores de uma empresa discutem os resultados e potenciais soluções para os problemas descobertos em fases anteriores. Tomam decisões com o intuito de melhorar esses resultados. É nesta fase que a colaboração como parte da estratégia de BI é fundamental. A colaboração dentro de BI simplifica e documenta todo o processo de comentários/votações, para que cada ponto de vista e opinião possam ser pesados e ponderados contribuindo para uma melhor decisão final. Como resultado desta fase, novas áreas ou indicadores poderão surgir e ser adicionados à nova fase de medir que se segue, de forma a acompanhar os progressos das decisões tomadas no ciclo anterior.

Desta forma o processo de BII de uma empresa é um processo perpétuo e contínuo que leva a empresa à perfeição, passo a passo. Uma vez iniciado o ciclo, será difícil pará-lo.

2.1.5. Tecnologias e ferramentas de BI

Existem atualmente no mercado bastantes soluções de BI para empresas. Algumas integram numa única ferramenta toda a panóplia de *software* e procedimentos associados às várias fases do BI. Inúmeras empresas prestam serviços nessa área, aconselhando a melhor solução após uma análise aos processos de negócio. Não é âmbito desta secção descrever todas as tecnologias e ferramentas disponíveis nesta área pelo que apenas algumas das mais usadas serão aqui referidas. As soluções apresentadas poderão não implementar todas as fazes do BI, mas apenas algumas.

2.1.5.1. *Pentaho*

O *software* Pentaho [10] é open *source* e constitui uma plataforma moderna e integrada. Está já preparada para análise de dados *big data*. O *software Pentaho Business Analytics* é uma solução bastante completa, fácil de implementar e usar, e com baixo custo. Esta suite de programas inclui acesso aos dados, visualização, integração, análise e *data mining*. No entanto o *Pentaho Data Integration* é também vendido como produto *stand-alone.*

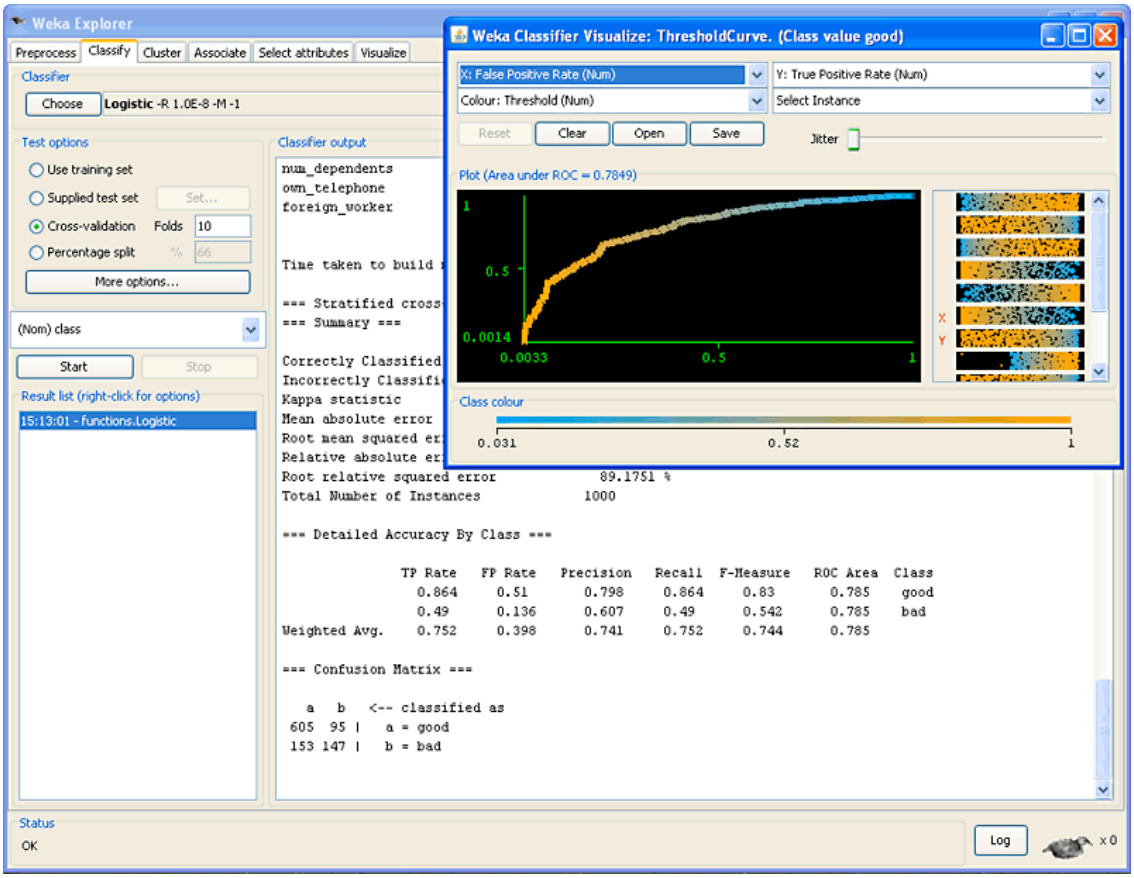


Figura 5. ***Interface do Pentaho Business Analytics- 1***

O *Business Analytics* disponibiliza uma versão bastante interativa e fácil de usar com interface *web*. Permite aceder aos dados de forma segura (com possível *single sign-on* integrado com *Active Directory*), criar relatórios e *dashboards* analisando os dados multi-dimensionalmente. Funciona com os mais variados formatos de dados e fontes, incluído as fontes de *big data* como a *Hadoop* (framework de aplicações para computação distribuída, fiável e escalável) e as bases de dados noSQL. O *software* de integração de

dados, o *Pentaho Data Integration*, permite extrair dados de fontes complexas e heterogéneas, criando estruturas de informação consistentes e de grande qualidade. Oferece uma ferramenta de ETL que cobre a maior parte das necessidades.

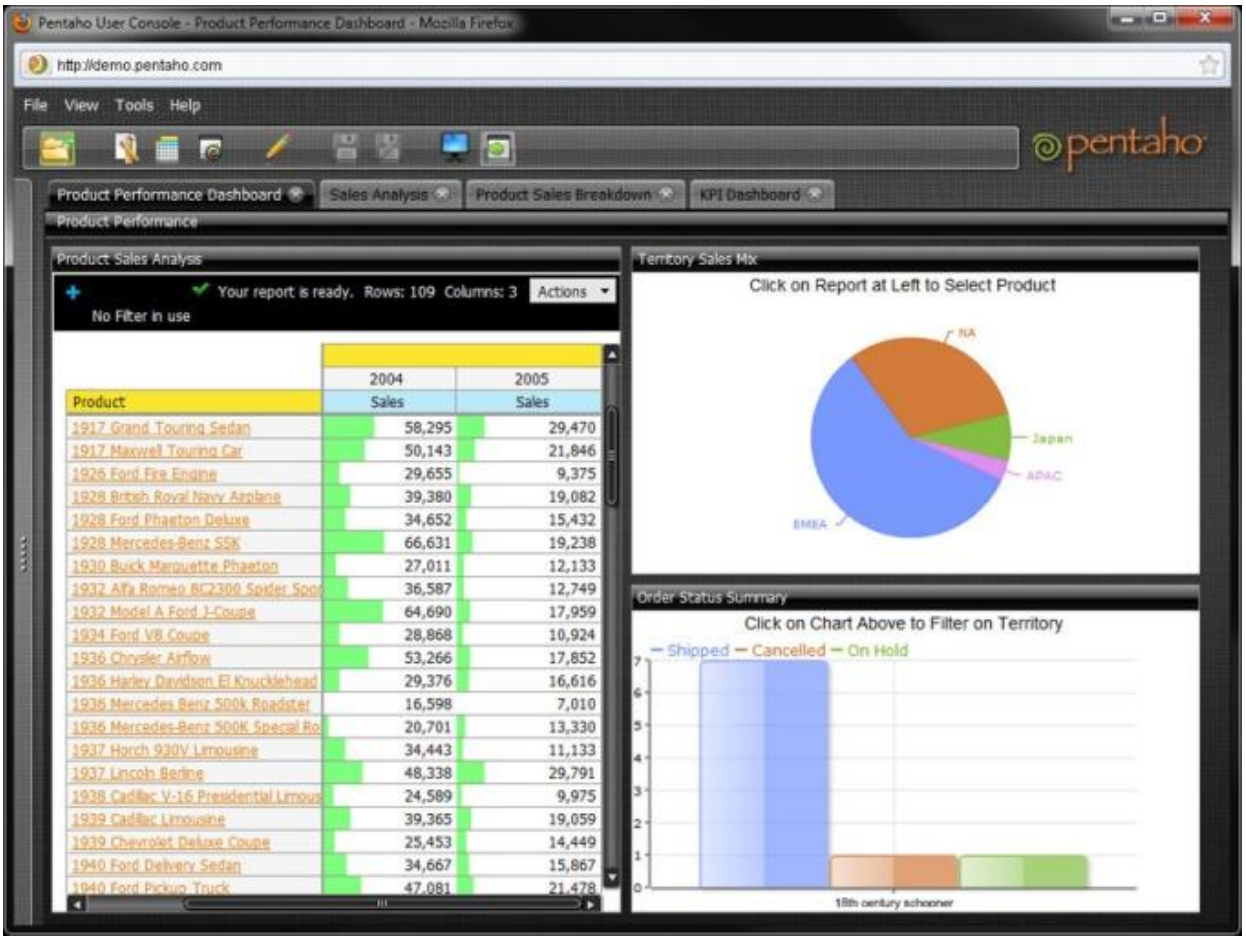


Figura 6. ***Interface do Pentaho Business Analytics-2***

2.1.5.2. MicroStrategy

A empresa MicroStrategy oferece um conjunto de produtos que visam essencialmente tornar o processo de BI mais rápido, simples e mais *user-friendly*. Segundo a própria empresa, o *software* de BI da MicroStrategy permite transformar os dados e obter uma apresentação dos mesmos de forma compreensível através de *dashboards* e relatórios que enfatizam a eficiência e produtividade, potenciam estratégias de lucro mais eficazes e preveem oportunidades. As fontes de informação destes produtos podem ser *Data warehouses*, bases de dados multidimensionais (sistemas de ERP) ou mesmo ficheiros de texto e de Excel. Pode ainda receber *inputs* de serviços *web*. A plataforma proposta na figura seguinte propõe a criação de metadados.

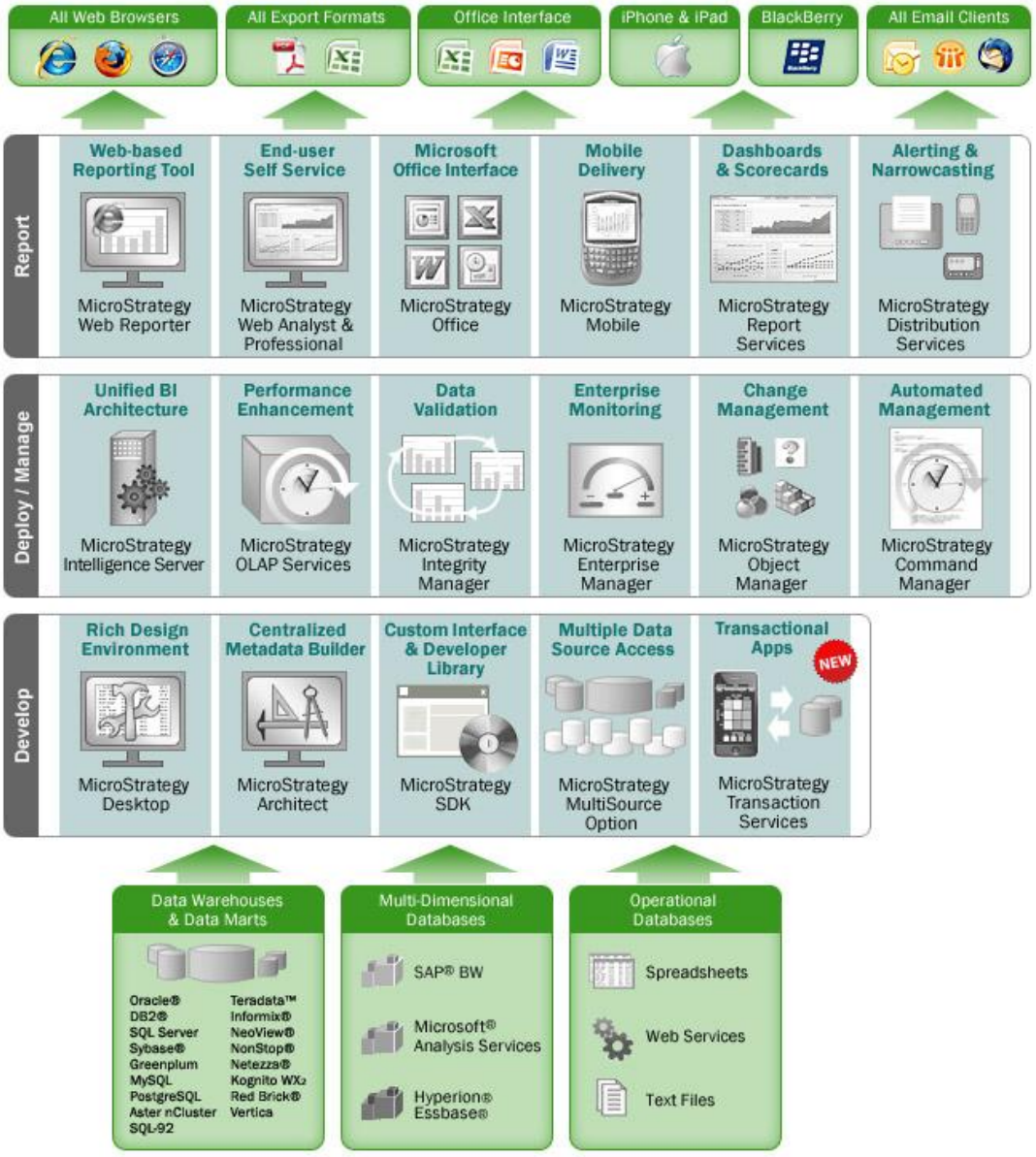


Figura 7. Mapa de produtos de BI da MicroStrategy [11]

Os metadados são informação que representa um conjunto de dados. No que diz respeito aos sistemas de BI, os metadados ligam os dados armazenados às entidades de negócio e regras definidas na aplicação de BI. No seu nível mais básico os metadados traduzem as tabelas com informação de uma *Data warehouse* em termos que façam sentido no contexto do negócio, permitindo o acesso a informação valiosa para construir mapas e relatórios, e conduzir análises. Os sistemas de BI atuais exigem metadados centralizados, reutilizáveis e dinâmicos. As plataformas da MicroStrategy disponibilizam já este tipo de metadados: centralizados e com um nível de abstração, reutilizáveis e dinâmicos, orientados aos objetos, armazenados numa base de dados relacional, aumentando a

escalabilidade e facilidade de gestão. Assim, a sua visualização será possível em inúmeros tipos de dispositivos, desde um computador a um telemóvel.

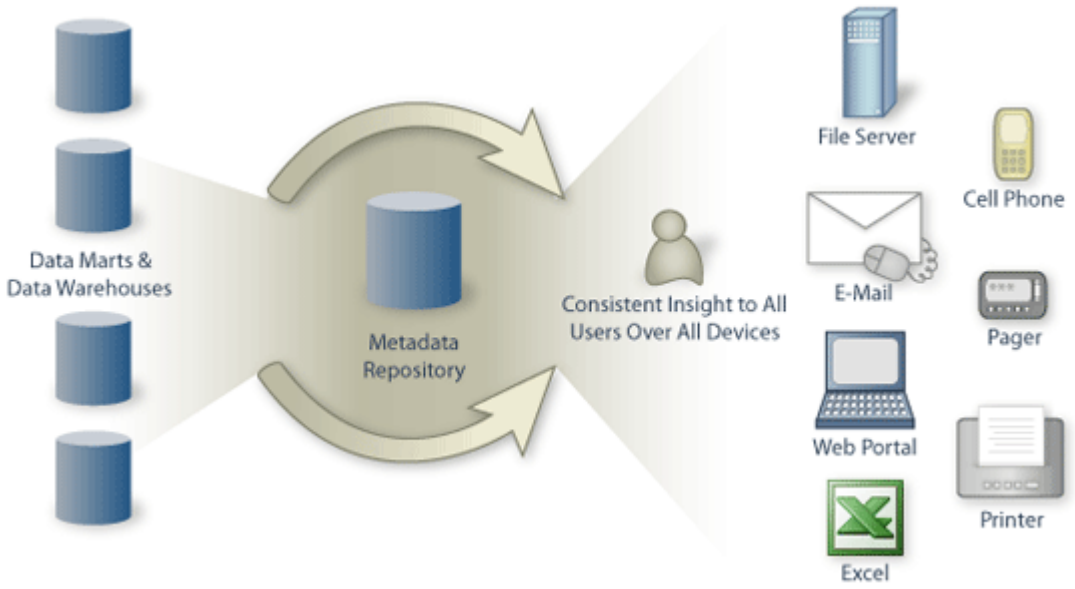


Figura 8. Produção e visualização de metadados [12]

2.1.5.3. Oracle

A Oracle é talvez a mais conhecida empresa que desenvolve e comercializa soluções de BI. A Oracle oferece um conjunto de aplicações para BI, a *Oracle Business Intelligence Applications Suite*, que suporta várias áreas funcionais. A solução é completa, *pre-built* e assegura o cumprimento das melhores práticas de análise incluindo um conjunto vasto de KPI's (*Key Performance Indicator*), métricas e *workflows*. As aplicações desta solução acedem a várias fontes de dados da organização e permitem obter uma visão geral do negócio e do seu desempenho, levando a uma tomada de decisão mais rápida e informada, ajudando a otimizar o negócio (reduzindo custos e aumentando a eficiência).

A solução da Oracle inclui ferramentas de ERP, CRM e *Industry analytics*.

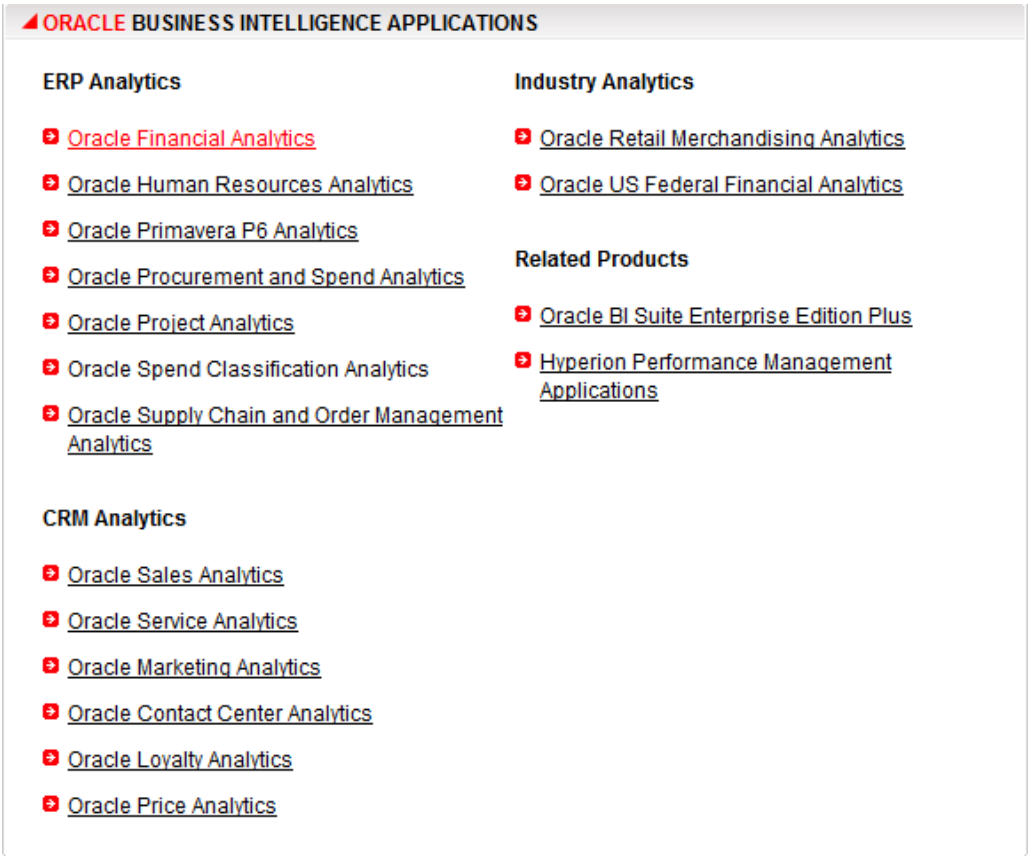


Figura 9. Aplicações Oracle BI [13]

Os produtos da Oracle são talvez dos mais conhecidos e comercializados devido, em parte, ao historial da marca no que respeita a bases de dados e *Data warehouses*. A grande desvantagem é o facto de serem aplicações com elevados custos de implementação, embora o ROI (*Return of Investment*) seja alcançado no médio prazo.

2.2. *Personally Identifiable Information* (PII)

Este é um termo usado na área da segurança de informação e define o tipo de informação que pode ser usada para identificar, contactar, caracterizar ou localizar uma única pessoa. Embora seja um termo bastante usado, ganhou popularidade com o crescimento da Internet e com a facilidade que nela existe de recolher este tipo de informação. As novas tecnologias de informação, as redes sociais e todo um vasto conjunto de serviços oferecidos na Internet, permitem a recolha e armazenamento de PII levando a um mercado cada vez mais rentável de venda de informação PII. No entanto este não é o único possível uso da PII. Esta informação pode também ser usada por criminosos com intenção de roubar a identidade de uma pessoa ou roubar bens à mesma, entre outros crimes possíveis. Como uma forma de travar estes crimes, muitos *sites*

possuem uma política de privacidade em que o tema da PII é abordado de forma específica. Para além disto os legisladores dos vários países elaboraram uma série de leis que limitam a distribuição e acesso à PII.

No entanto, a proteção da PII não é somente responsabilidade de empresas, mas também dos próprios indivíduos que podem ver a sua reputação arruinada, se as empresas que recolhem esta informação não forem fiáveis. A recolha, manutenção e disseminação deste tipo de informação deve seguir e respeitar práticas e regras bem definidas e recomendadas pelas instituições nacionais e internacionais. Algumas recomendações, como as da OCDE (Organização para a Cooperação e Desenvolvimento Económico), definem práticas de:

- Limites de recolha de informação – devem existir limites na recolha e armazenamento de dados pessoais. Os indivíduos a quem essa informação diz respeito devem dar o seu consentimento específico para cada ação que posso ser feita sobre estes.
- Especificação do propósito – o destino dos dados recolhidos deve ser definido e apresentado antes da recolha dos mesmos e o seu uso deve limitar-se a este fim.
- Qualidade dos dados – os dados pessoais recolhidos devem ser relevantes para o propósito a que foram destinados, devendo estar atualizados e corretos.
- Limitação de uso – a informação PII não deve ser revelada nem disponibilizada para outros propósitos que não o especificado, exceto se o seu "dono" ou tribunal assim o autorizarem.
- Proteção e segurança – os dados pessoais devem ser protegidos contra riscos de perda, uso, modificação ou acesso não autorizado à informação.
- Abertura – Deverá existir uma política geral de abertura acerca de desenvolvimentos, práticas e políticas que digam respeito à PII. Deve estar disponível a informação de existência, principal propósito, nome e localização desta informação.
- Participação dos indivíduos - Cada pessoa tem o direito de obter do provedor (quem recolhe) os seus dados ou a informação de que este os possui ou não. Deverá ainda poder modificá-los, apagá-los, ou corrigi-los.
- Responsabilidade – um provedor é o responsável por cumprir as regras e medidas definidas por leis e recomendações.

Para que cada regra ou recomendação seja cumprida, é também necessário que existam ações de formação para os próprios indivíduos. Estas ações de formação visam criar nos indivíduos um sentimento de alerta que pode resultar em reações preventivas da disponibilização indiscriminada de dados. É necessário perceber que apesar de ilegal, uma vez que a lei protege os menos conhecedores, não quer dizer que não aconteça. É também uma questão de criar o sentido de consciência (*awareness*). Inúmeras vezes a

falha de segurança e consequente fuga de informação, são provocadas por coisas tão simples como uma *cookie* num browser, que recolhe informação que poderá ser posteriormente usada e/ou ligada a outra levando a inferência de outra informação.

Por outro lado, esta informação poderá ser bastante útil no que diz respeito ao negócio de determinada empresa ou até na deteção de possíveis ameaças aos seus sistemas.

2.3. *Identity and Access Management* (IAM)

O IAM surgiu como forma de ajudar as organizações a acompanharem as exigências do mundo de negócios de hoje em dia e manterem a sua atividade de forma compatível com as regulamentações do seus países. O IAM faz a junção dos processos de negócio, políticas de segurança e tecnologias que estão ao dispor para ajudar as empresas na gestão das entidades digitais e controlar o acesso aos seus sistemas. De um cenário com uma visão dispersa de diferentes papéis existentes no ecossistema da empresa/organização e inúmeras responsabilidades nos sistemas da mesma, geralmente difíceis de gerir e auditar, passaremos para um cenário unificado, com uma visão global e controlo efetivos através de políticas de segurança, que permitirão reduzir custos e riscos. Um sistema de IAM deve incluir serviços de diretoria, serviços de gestão de ciclo de vida de identidades (desde o seu aprovisionamento, passando pela sua gestão e administração) e um sistema de gestão de acesso aos serviços.

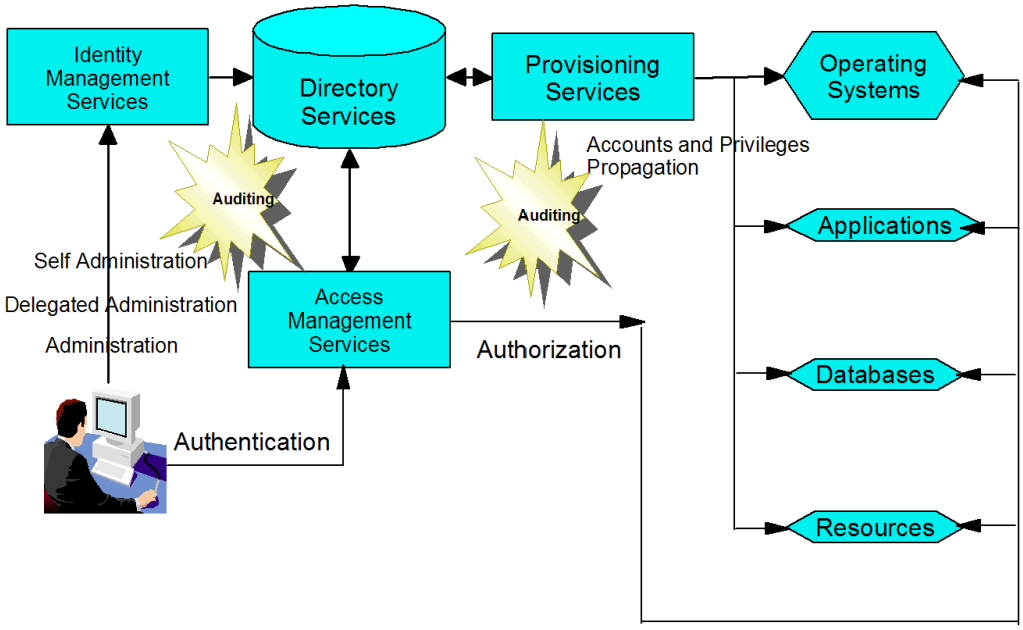


Figura 10. Descrição de um sistema IAM genérico [14]

Os serviços de diretoria são o coração do sistema de IAM pois nele estão guardados toda a informação de perfil de utilizador. São o repositório central. Poderá ser um simples sistema de ficheiros ou até uma base de dados complexa. Caso seja necessária a existência de vários repositórios, é importante manter um sistema central de entrada e acesso aos outros repositórios para que se tenha uma visão global das identidades armazenadas no nosso sistema.

Os serviços de gestão do ciclo de vida de identidades são os processos que modificam os atributos dos utilizadores assim como os seus direitos e permissões, as suas credenciais, baseando-se em políticas. Este processo inclui o aprovisionamento de identidades digitais, a sua administração e propagação das alterações a todos os sistemas de aplicação. Permite ainda delegar a administração a outros grupos de utilizadores confiáveis e a autoadministração de alguns atributos. Por último, sem esquecer de gestão de credenciais e *passwords*, as chaves da autenticação e autorização, que devem ser administradas com a máxima segurança e seguindo determinados procedimentos.

O sistema de gestão de acesso aos serviços consiste em controlar, monitorar e auditar o acesso aos recursos que existem na rede (seja esta uma rede caseira ou corporativa). Este processo baseia-se em políticas de segurança usando a autenticação e autorização em conjunto com mecanismos de confiança.

A **autenticação** é o processo que verifica a identidade do utilizador, havendo variadas formas de o fazer. A escolha da tecnologia de autenticação depende muitas vezes dos requisitos da política de segurança e deve integrar algumas recomendações como a facilidade de utilização e integração, e a possibilidade de suportar múltiplas aplicações, tudo isto com baixo custo. Podemos fazer autenticação usando: username + *password*, *Personnal Identification Numbers* (PIN), Certificados digitais X.509, *one-time passwords*, métodos biométricos (impressão digital, leitura da retina, etc.), *smart cards*, passaporte eletrónico ou *tokens* de *hardware*. Todas estas formas de autenticar são seguras havendo no entanto algumas que são mais robustas, uma vez que garantem a segurança através do uso de métodos criptográficos. Associado à autenticação e ao IAM existe o conceito do *Single Sign On* (SSO) que permite, quando possível, que os utilizadores autorizados a aceder a múltiplos recursos o façam autenticando-se uma única apenas. O SSO deve ser integrado nos sistemas da empresa passo a passo e sempre respeitando as regras e políticas existentes, permitindo a sua evolução com os sistemas.

A **autorização** é processo pelo qual se verificam as permissões do utilizador baseando-se na sua identidade. Depois de este se autenticar é feita a verificação de autorização baseando-se nas políticas de acesso e controlo, concedendo ou não o acesso. Existem vários tipos de controlo: *Discretionary Access Control* (DAC), *Mandatory Access Control* (MAC) ou *Role Based Access Control* (RBAC).

O conceito de **federação** e **confiança** é aqui muito importante pois permite a partilha de recursos de forma estruturada. A confiança é um mecanismo complexo e permite a autenticação segura e a autorização de identidades entre sistemas independentes, por exemplo de empresas parceiras, essencial para não se perderem oportunidades de negócio.

Por último falta referir o processo de **auditoria**, que é já um requisito de qualquer sistema da atualidade. A auditoria permite a obtenção de informação de quem, o quê, quando, onde e a que recursos acedeu na empresa. Devem portanto ser registados os seguintes eventos:

- Eventos de Autenticação
- Eventos de Autorização
- Modificação de objetos de diretoria

O processo de auditoria pode começar simplesmente pela filtragem de registos de atividade.

2.4. *Identity Intelligence*

O *Identity Intelligence* é a aplicação de técnicas típicas de BI e outras técnicas de análise e controlo, à informação de *Identity Management* (IdM). O termo *Identity Intelligence* tornou-se mais popular durante 2010 quando começou a ser usado por uma das grandes companhias de investigação e desenvolvimento em BI, a Gartner *[15]*. Este termo pressupõe a existência de um conjunto de características e pressupostos como:

- Existência, numa organização, de repositórios com a informação de contas de utilizador, e capacidade de recolher informação que os caracterize e aos seus direitos de acesso. Esta característica distingue-se no que respeita aos sistemas de IdM mais comuns, uma vez que normalmente os seus repositórios são mais simples e menos adequados para estes tipos de análise.
- Capacidade de correlação da informação proveniente de diferentes fontes, de forma a conseguir eficientemente popular o repositório. Para permitir uma análise rápida, completa e detalhada é fundamental ter uma ferramenta que junte a informação, relacione e a torne mais homogénea.
- Capacidade de fazer análises complexas baseadas nos princípios do BI produzindo resultados que indiquem o estado dos utilizadores na organização e a qualidade dos processos de gestão destes dados.
- Segurança e controlo avançado são outras características que devem estar presentes e ser monitoradas.

A necessidade de termos *Identity Intelligence* surge da crescente tomada de consciência de que os sistemas de IdM podem e devem fazer muito mais do que apenas gerir as contas de utilizador e automatizar a sua gestão. As ferramentas de IdM têm evoluído bastante e são hoje mais do que ferramentas de gestão. São ferramentas de *Security Governance* que permitem aumentar o nível de segurança e por conseguinte permitir que as empresas alcancem os seus objetivos no que respeita a certificações e requisitos. Esta evolução baseia-se no pressuposto de que para gerir de forma conveniente a informação de identidade deve-se primeiramente conhecê-la ao mais pequeno detalhe (estrutura, conteúdos, etc).

Também as soluções de *Cloud Computing*, cada vez mais divulgadas são fonte de processamento de informação bastante sensível e importante para os seus utilizadores. Quando falamos de domínios de *Clouds* temos a certeza de que manter o rasto a esta informação pode tornar-se um desafio. A introdução de soluções que proporcionem ao utilizador um maior controlo e gestão da sua informação é imprescindível. O simples ato de guardar acessos e tarefas realizadas num ficheiro de log é claramente ineficiente neste propósito. Ao fluxo de informação na *Cloud* é extremamente dinâmico e pode depender

de uma variedade de acontecimentos e as definições de privacidade também variam conforme o utilizador. São alvo de investigação, ferramentas que tornem esta atividade mais transparente. Algumas iniciativas como a TAMI (*Transparent Accountable Data Mining*) têm como objetivo permitir a transparência no uso desta informação tão sensível. Não é uma ferramenta mas antes uma espécie de *proxy* que permite a entidades autorizadas analisar o histórico de manipulação dos dados e efetuar previsões e inferências com base nesta análise. Alguns projetos financiados pela União Europeia já abordaram também esta problemática, como é o caso do PRIME (*Privacy and Identity Management for Europe*) que desenvolveu uma ferramenta de transparência que permite ao utilizador saber quem acedeu aos dados e monitorar estes acessos. Os *registos* são guardados e o utilizador pode consultá-los mais tarde.

Mas a tecnologia existente hoje, ainda não permite ao utilizador comum uma forma rápida e simples de monitorar esta informação. Esforços têm de ser feitos para que o utilizador possa um dia, independentemente da natureza da sua informação e das suas preferências em termos de privacidade, olhar transparentemente para tudo o que acontece relacionado com esta informação tão sensível.

Assim muitas empresas durante o ano de 2010 e 2011 adotaram mecanismos de *Identity Intelligence* mesmo durante as primeiras fases do ciclo de vida dos seus sistemas de IdM. A inovação neste contexto está em adotar este mecanismo de *Identity Intelligence* mesmo antes de implementar ou decidir por um sistema de IdM, para que o mecanismo escolhido sirva de suporte à definição de requisitos para o modelo de IdM escolhido. Este facto aumenta a probabilidade de sucesso pois permite ter noção do estado do processo de *User Management* da organização que a torna bastante útil para perceber a real necessidade da implementação do sistema de IdM.

Esta crescente tendência para adotar o *Identity Intelligence* tem vindo a fazer com que este tenha um lugar de destaque nos sistemas de BI e suporte à decisão. Os programadores e fabricantes de sistemas de IAM (*Identity and Access Management*) podem e devem evoluir, sob pena de não o fazerem e ficarem obsoletos e perderem negócio. Mas ainda nos dias de hoje os casos de aplicação são reduzidos. Existem já algumas propostas de ferramentas comerciais mas ainda há muito caminho a percorrer.

O *Identity Intelligence* é também visto como uma nova forma de segurança. O perímetro de segurança passa a ser a **Identidade** sendo que este é segurado por políticas e não pela topologia de rede. Mas se eu tenho a informação espalhada pelos mais diversos locais, como posso assegurar e gerir esta segurança? E como manter a compatibilidade com regras e políticas estabelecidas? A proposta neste sentido é de identificar, medir e gerir os riscos, transformado a informação de identidade e acesso, em informação *user-friendly* e sobre a qual se podem tomar decisões e agir. A informação pode e deve ser usada de

forma realmente inteligente e não ficando limitada apenas a medições e análises quantitativas. Tal como no BI tradicional são necessárias ações de inferência e previsão sobre este tipo de informação. A inclusão de inteligência associada ao sistema de auditoria de um IAM permitirá relacionar informação e, tal como pretendido, automatizar *workflows*, identificar perfis de utilização, perfis fraudulentos, alertar em tempo real e agir com celeridade e eficiência.

2.4.1. Soluções de mercado

Existem já em comercialização algumas soluções que são referidas como de *Identity Intelligence*. Apenas se vai aqui fazer referência a algumas delas, uma vez que o âmbito deste trabalho não é a descrição exaustiva das mesmas. A descrição baseia-se na informação disponibilizada ao público em geral. Uma vez que estamos a tratar de produtos comerciais esta informação poderá ser pouco detalhada.

2.4.1.1. Identity and Access Management da Quest Software

A Quest oferece uma panóplia de *software* desenvolvido para ir de encontro às necessidades de *Identity and Access Management* das empresas e organizações. As soluções da Quest no ramo da Identidade disponibilizam de uma forma simples a gestão de identidades e controlo de acessos, tornando o negócio mais ágil e seguro pela redução do peso na infraestrutura de IT (*Information Technologies*). Segundo a empresa, com as suas soluções conseguirá reduzir os riscos de segurança e exposição, aumentando a eficiência dos sistemas e minimizando a necessidade de realização de auditorias. As soluções da Quest passam pelo *Access Governance* que permite que cada funcionário ou utilizador tenha o acesso que necessita apenas. Permite que o administrador tenha sempre visível quem no seu ambiente acede a quê e assegurando que cada um apenas terá o acesso estrito para o desempenho da sua função e nada mais. A funcionalidade de controlo de acesso superutilizador, permite ao gestor garantir o cumprimento das políticas definidas para estes acessos, melhorando a segurança e aumentando a eficiência. No que respeita à gestão das identidades, o *software* permite consolidar identidades numa única infraestrutura, automatizando as tarefas de administração (gestão de contas e *password*s), permitindo SSO em todos os sistemas e por isso aumentando a segurança de autenticação na organização. Além destas vantagens resta falar da monitorização de utilizadores, com relatórios de atividade que permitem imediatamente responder em caso de existir um problema. Este sistema permite detetar vulnerabilidades nos sistemas e também nas políticas de acesso, prevenindo acessos não

autorizados e assegurando proactivamente a proteção das políticas de segurança contra violações à mesma.

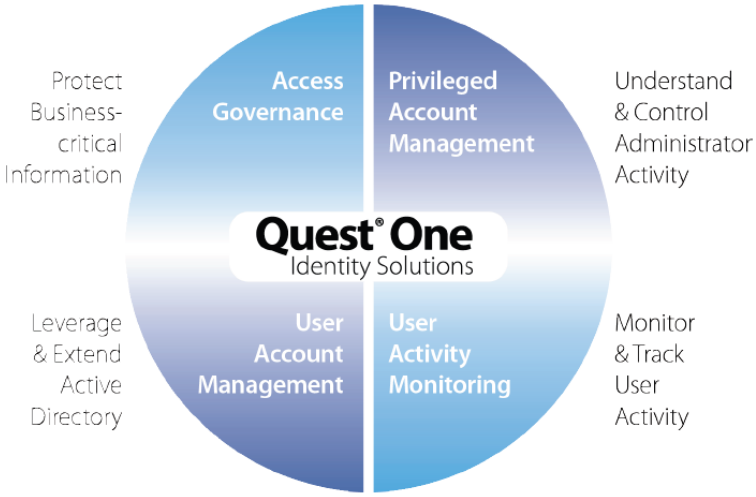


Figura 11. Soluções Quest Identity [16]

2.4.1.2. Identity Analytics *da Oracle*

O sistema Oracle *Identity Analytics* disponibiliza uma forma de as organizações gerirem e automatizarem o controlo de informação de identidade, juntando melhor do *Identity Management* e do *Business Intelligence*.

Este produto inclui uma *Identity Data warehouse* que é o repositório central onde estão armazenadas a informação de identidade, acesso e auditoria, de forma otimizada para responder a *queries* analíticas complexas e efetuar simulações. Estes dados podem ser importados de outras bases de dados de forma programada. O *Oracle Identity Analytics* permite uma forte integração das suas capacidades com outros produtos de provisionamento. Este produto permite ainda reduzir o risco de operação permitindo uma visão total dos acessos de cada utilizador. Automatiza o processo de certificação e atestação dos direitos de cada utilizador. Esta tarefa resulta na atribuição efetiva de permissões aos utilizadores baseadas apenas no que eles necessitam para o desempenho da sua função. O interface é bastante *user-friendly* e com capacidades de filtragem e ordenação bastante avançadas. No que está relacionado com as auditorias, a segregação de tarefas evita que ocorram falhas de segurança intencionais ou inadvertidas, apenas pela existência de combinações conflituosas de papéis e acessos. O *enforcement* da política de *IT Audit* permite assegurar o cumprimento de regras e regulamentações assegurando a integridade das operações realizadas. Este produto da Oracle inclui ainda a

funcionalidade de efetuar *role mining* permitindo limpar e organizar os acessos atribuídos. Os papéis definidos em cada organização, na sua estrutura, evoluem com o tempo, pelo que necessitam de ser administrados e aprovados. Qualquer mudança de papéis é analisada detalhadamente e efetuada uma análise do impacto antes de as mudanças serem aplicadas. Permite ainda simular o que poderia acontecer se esta ou aquela mudança ocorressem e efetuar *rollback* em caso de falha. A todas estas funcionalidades é adicionado um conjunto de *dashboards* bastante compreensíveis e capacidades de análise avançadas baseadas na identidade, acesso e auditoria. Todos os *dashboards* e relatórios são configuráveis de acordo com as preferências do utilizador.

Figura 12. Oracle Identity Analytics [17]

2.4.1.3. Tivoly Identity Manager *(IBM)*

Este sistema vem responder a alguns desafios propondo a implementação de uma solução simples e fácil de instalar, com capacidades *user-friendly* acrescidas, uma solução completa de *Identity and Access Management* para organizações [18]. Baseia-se no uso de papéis, contas e permissões de acesso, permitindo automatizar os procedimentos de criação, modificação e término de privilégios e contas de utilizadores. Através da implementação de uma hierarquia de papéis, o custo de administração é reduzido e esta é simplificada pelo uso de uma estrutura organizacional de *roles*.

É garantida a separação de tarefas de forma a aumentar a segurança e a conformidade, através da criação, modificação ou remoção de políticas que excluem utilizadores que pertençam a vários grupos de papéis, que possam eventualmente entrar em conflito. A possibilidade de re-certificar as permissões de vários utilizadores ao mesmo tempo, poupa em esforço ao administrador que poderá tomar decisões sobre as permissões de acesso de vários utilizadores, numa única submissão ao sistema.

O sistema de aprovisionamento é baseado em pedidos e aprovação de permissões para os vários sistemas.

O interface é configurável. Este sistema permite a criação de relatórios dos *workflows* existentes e mudanças ocorridas. Permite ainda monitorizar a conformidade com as políticas de acesso, através da correlação identificando: histórico, contas órfãs, contas não usadas e separação de tarefas. Esta ferramenta permite assim, pela automatização na criação de utilizadores, contas e grupos, a redução de erros que possam estar a ser introduzidos nesta fase do ciclo de vida de um acesso. Através do uso de um portal de funcionalidades *self-service*, adiciona ainda a redução de pedidos de ajuda aos responsáveis pela administração dos sistemas.

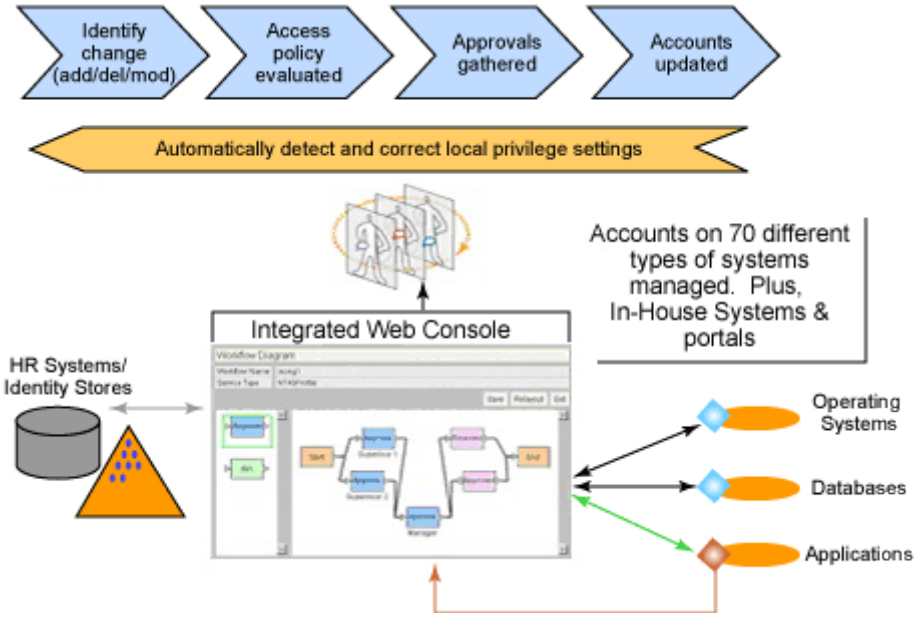


Figura 13. IBM Tivoly Identity Manager [19]

Recentemente a IBM anunciou mais uma evolução no seu sistema de *Identity Intelligence*. Esta nova funcionalidade permite às organizações uma gestão mais eficiente e sofisticada da informação dos seus colaboradores, sobre o pressuposto de que um acesso não autorizado a informação de clientes poderá colocar a organização num estado de vulnerabilidade e possibilitar fuga de informação. Esta preocupação é ainda mais flagrante no contexto da evolução do modelo de *cloud* e acesso móvel, a que hoje assistimos. Segundo o artigo publicado no site *Security Today* a 13 de janeiro do presente ano, um dos primeiros passos para a segurança de uma empresa é esta saber o que os seus colaboradores andam a fazer, e isto é possível com o novo produto IBM *Security Role and Policy Modeler.* Como é referido no artigo, este produto permite classificar os colaboradores em perfis, após uma análise aos seus dados, e gerir eficazmente os papéis e identidades dentro de uma organização através de um interface único e intuitivo [20].

2.5. **Normalização de sistemas de recolha de informação e *auditing***

Existe ainda trabalho feito no sentido no sentido da normalização. O TM Forum tem vindo a desenvolver esforços e o trabalho já feito encontra-se relatado no documento "Security Compliance Audit Automation" [21], cuja versão mais atual é a 0.7. O documento TMF_SCA_BA é um *Business Agreement* e refere-se à parte de auditoria no que respeita à conformidade dos sistemas de suporte às operações para com as diretivas de segurança, legais e corporativas.

Normalmente os dados relativos a segurança dos sistemas, nomeadamente os *logs*, dados de telemetria (atividades observadas tal como processos ativos, ficheiros abertos, processador, etc) e configurações são determinados, questionados e transferidos de forma proprietárias e específicas do fornecedor ou fabricante e em formatos estranhos, o que não os torna facilmente utilizáveis. Assim este documento procura normalizar interfaces e formatos de dados para ajudar na automatização das auditorias de conformidade em segurança. O TMF_SCA_BA ilustra os tipos de informação mais relevantes e mapeia os processos relacionados no *TMForum Business Process Framework* ou eTOM (uma espécie de ITIL, sendo que o ITL é orientado aos processos de Tecnologia da Informação e o eTOM é orientado aos processos de Operações de Telecomunicações). São definidos neste documento alguns casos de uso e deles são levantados requisitos para uma solução normalizada. Este documento é apenas o caminho inicial para mais dois que se seguem: o *Information Agreement* e o *Implementation Specification* que visam normalizarem os interfaces e especificar o formato dos dados.

Este documento define como normalizar estes processos e traz inúmeras vantagens para provedores de serviço, fornecedores de equipamentos e equipas de normalização. O princípio básico do SCA (Security Compliance Audit Automation) é que não deve existir acesso direto à infraestrutura do provedor de serviço. No sistema de gestão do SCA, uma das componentes da arquitetura, o SCA-MI (*Management Interface*) recebe os dados e permite que estes sejam acedidos a partir da sua fronteira com o provedor. Os dados são recolhidos e transmitidos pelo SCA-CT (*Collector and Transmitter*) que depois os coloca no SCA-DR (*Data Receiver*) através do interface SCA-DI (*Device Interface*). O SCA-DR guarda a informação depositando-a no repositório central SCA-CR (*Central Repository*). Este repositório disponibiliza a revisão e protege-a contra modificações. Contém ainda alguns filtros para impedir que informação critica seja armazenada (tal como passwords). Para consulta dos dados o SCA-DP (*Data Provider*) disponibiliza vistas sobre a mesma de forma segura e apenas com autorização. Poderá ainda existir um proxy, o SCA-P (*Proxy*), cuja função é fazer a adaptação, segregação ou reescrita no interface. Poderá funcionar mesmo, numa fase transitória, como interface. De seguida apresenta-se a arquitetura básica proposta no documento.

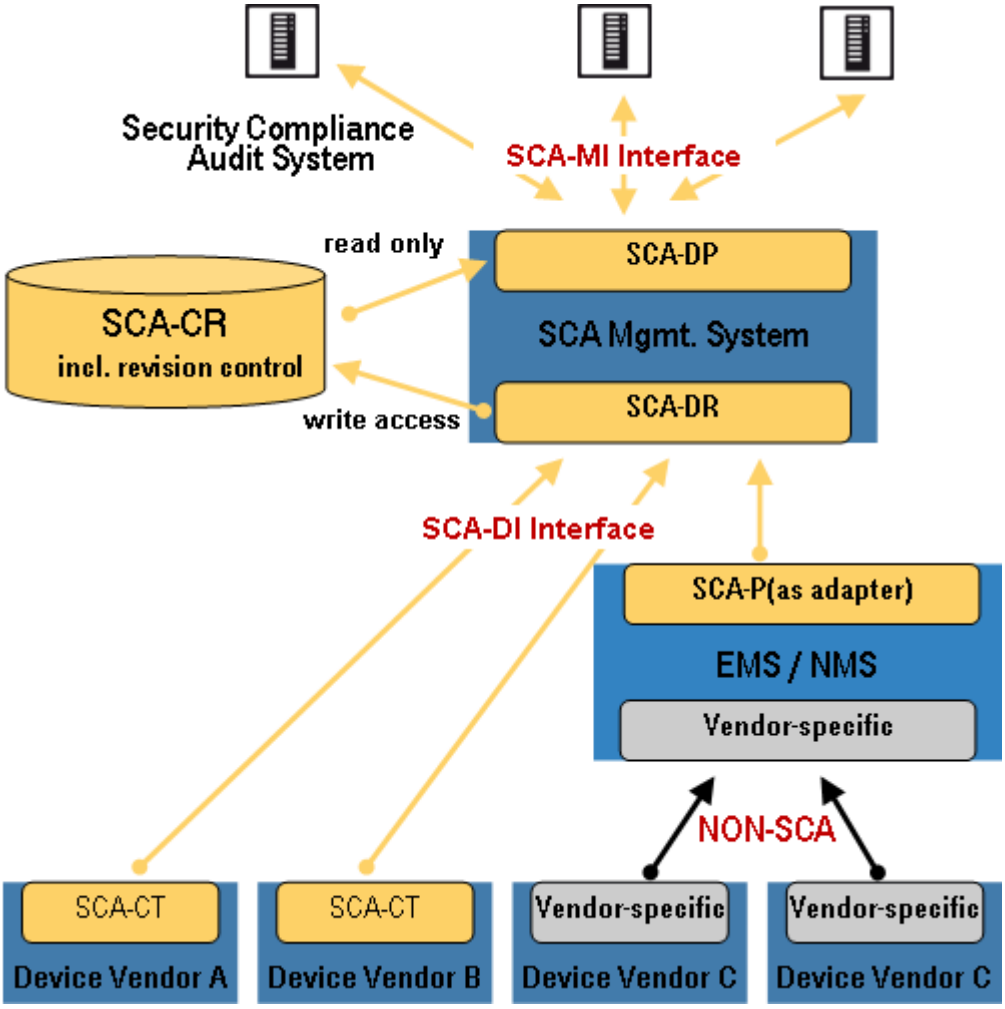


Figura 14.Arquitetura básica do SCA [21]

No que respeita aos dados, eles são divididos em três tipos, já referidos: dados telemétricos, dados de configuração e dados de *log*. Os dados de configuração serão por exemplo ficheiros com o /etc/hosts ou o /etc/ssh, são dados normalmente permanentes e que se encontram guardados em ficheiros ou bases de dados. Os dados de telemetria referem-se à observação do valor dos dados num determinado momento específico, como por exemplo a carga do processador de um sistema. São normalmente dados que não são permanentes e são recolhidos dos sistemas operativo ou das aplicações. Os dados de *log* documentam as modificações que ocorrem nos dados ou os eventos ocorridos ao longo do tempo. Os dados de configuração ou telemetria devem ser recolhidos e colocados no formato *standard* do SCA periodicamente e depois transmitidos para o SCA-DR. Para alguns dados de *log* poderá ser necessária a transmissão em tempo real ou quase real sendo que aqui se pode optar pela transferência dos dados *raw* ou pela transformação dos mesmos para um formato bem estruturado definido pelas normas.

3. Enquadramento legal

Esta análise não ficaria completa, sem referir o enquadramento legal e aspetos relacionados com as leis e diretivas existentes. Existem vários aspetos do armazenamento, processamento e comunicações eletrónicas de dados pessoais que estão protegidos pela lei. Podemos aqui destacar as leis relacionadas com a proteção de dados, transmissão e conservação de dados pessoais.

No que diz respeito a diretivas da Comissão Europeia (CE) podemos referir a diretiva nº 95/46/CE [22] relativa à proteção de dados e num âmbito mais alargado a diretiva nº 02/58/CE sobre a privacidade e comunicações eletrónicas [23]. No que diz respeito à retenção de dados, existe a diretiva nº 2006/24/CE [24] , e por último, mas não menos importante, a diretiva nº 2009/136/CE que se refere aos direitos universais do utilizador de comunicações eletrónicas e serviços [25].

Relativamente a estas diretivas, estas têm transposição em leis nacionais. No caso português podemos aqui referir a lei nº 67/98 de 26 de Outubro [26] (transposição da diretiva nº 95/46/CE), a lei nº 41/2004 de 18 de Agosto [27] (transposição da diretiva nº 2002/58/CE) e a lei nº 32/2008 de 17 de Julho [28] (transposição da diretiva nº 2006/24/CE).

Para este estudo interessa perceber as implicações legais que as aplicações de tratamento de PII devem ter em consideração.

Diretiva CE	Lei Portuguesa
Diretiva nº 95/46/CE	Decreto-lei nº 67/98
Diretiva nº 2002/58/CE	Decreto-lei nº 41/2004
Diretiva nº 2006/24/CE	Decreto-lei nº 32/2008

Tabela 1. Correspondência de Leis e Diretivas

A **lei nº 67/98** refere-se à proteção de dados pessoais e de pessoas singulares no que diz respeito ao tratamento e livre circulação desses dados. Fundamentalmente a lei refere que o tratamento de dados pessoais se deve processar de forma transparente e respeitando a privacidade, direitos, liberdades e garantias fundamentais do seu titular. Na sua redação a lei começa por definir alguns conceitos fundamentais dos quais aqui se apresentam alguns:

- Dados pessoais: qualquer informação, independentemente da sua natureza e suporte, relativa a uma pessoa singular identificada ou identificável (o titular dos dados);
- Tratamento de dados pessoais: qualquer operação ou conjunto de operações efetuadas sobre dados pessoais, de forma automática ou manual, tais como recolha, registo, organização, conservação, adaptação, alteração, recuperação, consulta, utilização, comunicação, bloqueio e destruição;
- Ficheiro de dados pessoais: conjunto estruturado de dados pessoais acessíveis segundo determinados critérios;
- Responsável pelo tratamento: pessoas singular ou coletiva que determinadas finalidades e meios de tratamento dos dados;
- Destinatário: pessoa singular ou coletiva a quem sejam comunicados dados pessoais;
- Consentimento do titular dos dados: manifestação de vontade livre, específica e informada, nos termos da qual o titular aceita que os seus dados pessoais sejam objeto de tratamento;
- Interconexão de dados: forma de tratamento que consiste na possibilidade de relacionamento dos dados de um ficheiros com dados de um ou mais ficheiros mantidos por outro responsável ou pelo mesmo mas com diferente finalidade;

Esta lei não é aplicável ao tratamento de dados efetuado por pessoa singular no âmbito pessoal ou doméstico. Os dados pessoais devem ser tratados de forma lícita e respeitando o princípio da boa-fé. Devem ser recolhidos para finalidades determinadas e não podem ser tratados de forma incompatível com as mesmas. Devem ser exatos e atualizados e, quando necessário, apagados ou retificados. Devem ser conservados de forma a permitir a identificação do seu titular apenas durante um período necessário para assegurar a prossecução das finalidades de recolha ou tratamento.

A Comissão Nacional para a Proteção de Dados (CNPD) poderá autorizar a conservação de dados para fins históricos, estatísticos ou científicos por um período superior ao inicialmente autorizado, mediante requerimento do responsável pelo tratamento. A lei refere que os dados só poderão ser tratados se o seu titular der, de forma inequívoca, o seu consentimento ou se esse tratamento for necessário para cumprimento de ação legal, de interesse público ou judicial. É proibido o tratamento de dados que se referem a convicções filosóficas, políticas, de fé religiosa, carater sindical, vida privada de origem étnica ou sexual, dados relativos a saúde ou genéticos. O tratamento deste tipo de dados só é permitido caso haja implicação legal (mediante autorização), consentimento do titular ou representante, para proteção do mesmo.

No que diz respeito à interconexão dos dados esta apenas é permitida se prevista por disposição legal, caso contrário está sujeita a autorização da CNPD. Segundo este decreto lei o titular dos dados tem o direito de saber a identidade do responsável pelo tratamento dos dados, a finalidade de recolha e outra informações tais como os destinatários dos dados e as condições de acesso e retificação. Esta obrigação só será dispensada por disposição legal ou autorização da CNPD. É da responsabilidade do responsável pelo tratamento dos dados a proteção dos mesmos, segundo técnicas adequadas. Deve proteger os dados contra destruição acidental ou ilícita, perda acidental, alteração, difusão ou acesso não autorizado, nomeadamente quando o tratamento implicar a sua transmissão numa rede. Os responsáveis pelo tratamento devem, portanto, assegurar todo o tipo de controlo sobre os dados, como o controlo do suporte dos dados, inserção, utilização, acesso, transmissão e transporte.

A CNPD pode determinar que em casos específicos a transmissão seja cifrada. Os responsáveis pelo tratamento estão obrigados ao sigilo profissional mesmo depois de cessarem as suas funções. Num outro capítulo, a lei refere as indicações a seguir no caso de transferências de dados na UE (União Europeia). É livre a circulação de dados pessoais entre estado da UE. É possível a transferência para outros estados mas apenas se estes assegurarem um nível de proteção adequado ou se o titular dos dados assim tenha dado autorização expressa para a transferência ou se esta for expressamente necessária para fins legais. A lei refere ainda que a CNPD é a entidade com responsabilidades para intervir e ser consultada sobre quaisquer assuntos que digam respeito ao tratamento de dados pessoais e quais as suas obrigações perante o Estado e a sociedade. Para além desta lei, existe mais duas que poderão ter implicações em aplicações que trabalhem PII.

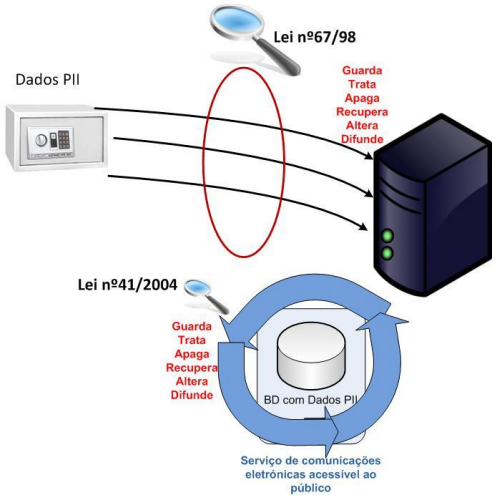


Figura 15. Proteção de dados

A **lei nº 41/2004** refere-se ao tratamento de dados pessoais e à proteção da privacidade no contexto do setor das comunicações eletrónicas. Neste sentido, as empresas que ofereçam serviços de comunicações eletrónicas ou redes de comunicações devem colaborar entre si para a adoção de medidas técnicas e organizacionais, que garantam a segurança dos seus serviços e a segurança da própria rede. No caso de risco de violação da segurança, as empresas devem comunicar aos sues clientes a existência desse risco e as possíveis soluções para o mesmo. As empresas desta área devem garantir a inviolabilidade das comunicações e dos respetivos dados de tráfego. É expressamente proibida a colocação de escutas, a instalação de dispositivos para este efeito e o armazenamento de dados de tráfegos exceto para os casos previstos pela lei.

A utilização das redes para armazenamento de informação ou para obtenção de informação armazenada no equipamento terminal de um assinante ou qualquer outro utilizador, é apenas permitida quando o assinante tem conhecimento dos objetivos desse processamento e possa recusá-lo. O armazenamento automático é permitido se necessário para efetuar ou facilitar a transmissão de uma comunicação ou para fornecer um serviço no âmbito da sociedade da informação, desde que solicitado explicitamente pelo assinante ou utilizador. Os dados de tráfego apenas devem ser mantidos até ao momento em que deixem de ser necessários para efeitos de faturação. O tratamento destes dados fica limitado aos trabalhadores e colaboradores destas empresas.

No caso de serem processados dados de localização o seu tratamento é apenas permitido se estes forem anónimos. Podem ser mantido pelo tendo estritamente necessário à prestação de serviços de valor acrescentado desde que seja obtido o consentimento prévio do assinante. Este consentimento deve poder ser retirado a qualquer momento de forma gratuita e simples. Podem ainda ser usados em caso de comunicações de emergência. Deverá ser dada a possibilidade de inibir a exibição da identificação da linha chamadora ao assinante, de forma simples e gratuita, assim como a possibilidade de rejeitar a chamada. No que respeita à lista de assinantes estes devem ser informados da inclusão da sua informação nestas listas e deve-se garantir aos mesmo a possibilidade de, sem custos adicionais, verificar, corrigir e alterar ou retirar os seus dados das mesmas.

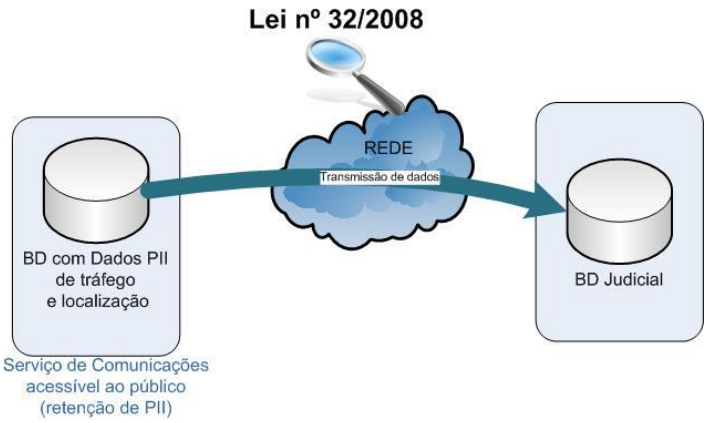


Figura 16. Proteção nas transmissões eletrónicas

Por último falta apenas referir a **lei nº 32/2008**. Esta lei regula a conservação e transmissão dos dados de tráfego e de localização, relativos a pessoas singulares ou coletivas, bem como dos dados necessários para identificar o assinante ou utilizador registado para fins de investigação por parte das autoridades competentes. A conservação de dados que revelem o conteúdo de comunicações é proibida exceto nos casos previstos nas leis referidas anteriormente. Os fornecedores de serviços de comunicações devem conservar os seguintes tipos de dados: dados que permitam identificar a fonte de comunicação e o seu destino, assim como a data, hora e duração. Devem ainda guardar dados necessários para identificar o tipo de comunicação e o equipamento usado. Estes dados devem ser conservados de forma segura para que possam ser transmitidos imediatamente, mediante despacho judicial, às autoridades competentes. No final do período de conservação estes dados devem ser destruídos a menos de indicações contrárias por ordem judicial. A CNPD deve manter um registo das pessoas autorizadas a aceder aos dados. A transmissão destes dados só poderá ser efetuada por despacho do juiz de instrução, se houver razões para crer que esta ação é indispensável para a descoberta da verdade ou para prova, na deteção e repressão de crimes graves.

3.1. Aplicação das disposições legais ao IAM e tratamento de PII

A gestão de identidades efetuada por um sistema de IAM pressupõe o tratamento de dados pessoais. É importante para uma entidade ou organização manter a sua eficiência operacional e ter a certeza de que os processos que são implementados estão de acordo com as leis locais, regulamentações e diretivas. Não basta conhecer as leis, é necessário perceber como estas se aplicam ou o que implicam nos seus processos de IAM. Muitas vezes a informação necessária para um sistema de IAM pode estar regulamentada em leis que obrigam ao cumprimento de determinados trâmites e autorizações. Informação de

caráter pessoal poderá ser necessária para dar acesso aos sistemas e serviços, pelo que os procedimentos legais devem ser tidos em conta para que situações ilegais não ocorram. Auditar um sistema que incluí um *framework* de políticas que determinam autorizações e que lidam com informação pessoal não pode ser feito sem olhar para o enquadramento legal do país.

Para efetuar o aprovisionamento dos utilizadores será necessário recolher informação pessoal. Para isso, e tendo em vista o cumprimento da lei, este deve dar o seu consentimento informado, e conhecer o responsável pelo tratamento destes dados. Deverá ser comunicada ao utilizador a finalidade desta recolha e deverão ser tomadas as medidas necessárias para assegurar a proteção dos equipamentos de armazenamento e transmissão da informação. A retenção destes dados também deve ter em conta o período estritamente necessário para tal, ou seja, quando um utilizador deixa de existir perante os sistemas de IAM, deverão ser tomadas as medidas necessárias para remover ou destruir essa informação.

Existem riscos associados à gestão inapropriada de acessos e permissões. Qualquer falha no sistema de IAM poderá conduzir a acessos não autorizados a informação sensível e a outros sistemas ou serviços. Também poderá ter impacto na performance e eficiência da entidade ou organização, uma vez que esta é responsável por dar acesso a quem o deve ter e restringir aos que dele não necessitam. Estas falhas poderão ter ainda implicação legal e comportamentos inesperados e inapropriados poderão surgir face a informação confidencial. Falhas na segurança e privacidade dos dados poderão conduzir os colaboradores a processos judiciais por denúncia junto das entidades competentes.

Assim sendo, um sistema de IAM também deverá ser alvo de auditorias regulares aos seus componentes e processos de forma a garantir a conformidade com as leis em vigor.

4. Inteligência num sistema de IAM

O IAM como produto surge da evolução de um outro produto da PT Inovação (PTIN), que permite gerir os acessos aos diferentes serviços necessários a cada utilizador para o seu trabalho diário. Como evolução da tecnologia e do paradigma de gestão de identidades e acesso faz sentido a integração de novas funcionalidades em produtos de gestão de acessos. Assim, dotar um sistema de gestão de identidades e acessos, de uma inteligência associada aos processos e *workflows*, faz sentido. A análise aqui efetuada gerará requisitos de funcionalidades e processos que qualquer sistema de IAM deve incluir, não se focando apenas no produto da PTIN. Os requisitos são genéricos e aplicáveis a qualquer sistema nesta área.

A análise e levantamento de requisitos serão baseados na identificação dos registos de atividade que a aplicação deve gerar. Assim, também serão aqui identificadas as informações que devem constar destes registos e que serão alvo de auditoria por parte do sistema. Importa distinguir entre registos de atividade e eventos, muitas vezes referidos como *logs*, dos verdadeiros ficheiros de *log*. Os *logs* são registos que servem normalmente para efetuar depuração da aplicação quando esta gera erros ou apresenta anomalias. São registos que interessam aos programadores e implementadores da solução. Nos registos de atividade e eventos estão registadas todas as interações dos utilizadores com a plataforma.

O registo de atividade de um sistema de IAM deverá conter informação sobre os acessos aos serviços por parte dos utilizadores, os eventos de SSO, as operações efetuadas nos serviços ou os pedidos de atributos, entre outra informação. Importante é o facto de todos os eventos deverem estar datados, i.e., deve ser registada a data e hora da sua ocorrência. Só com este registo de data será possível correlacionar informação do ponto de vista temporal. O tipo de registo (tipo de ficheiro, base de dados ou outro) poderá ser determinado com base na aplicação de ETL, ou seja, dependendo do tipo de ficheiro ou fonte que o sistema de ETL "eleito" aceite, poderemos ter diferentes tipos de registo. No entanto, a maioria das ferramentas de ETL aceitam formatos como o CSV (*Comma-Separted Values*) ou XLS (Excel), incluindo-se também bases de dados SQL, entre outros. Assim, a recomendação passa pela adoção de um destes tipos de ficheiro/fonte mais comuns, pois assim não ficarão restringidas as escolhas em termos de ferramenta ETL.

A introdução de inteligência no sistema de IAM é pois a implementação de um sistema paralelo de *auditing* aos registos de atividade do IAM. Podemos aqui distinguir três linhas de ação:

- *auditing* com vista à obtenção de informação estatística e histórico;
- *auditing* com vista à automatização de processos e *workflow*;

- *auditing* para deteção de anomalias e/ou fraude;

Como parte deste trabalho, identificaram-se um conjunto de *Use Cases*, que tiram partido da informação recolhida e processada. A informação que resulta do registo de atividade do IAM e dos serviços para os quais este gere o acesso e identidades, é a denominada PII. Esta informação dependendo do contexto e dos serviços pode conter dados pessoais sensíveis. Sendo PII a manipulação e salvaguarda destes dados deve ser segura e lícita. Informação de nome de utilizador, moradas, números de identificação, grupos de trabalho e atividade que este executa no âmbito dos serviços do IAM, são dados que podem ser correlacionados e levar à obtenção de novos dados confidenciais. Num contexto de uma organização estadual por exemplo, a informação manipulada pode incluir dados como identificação civil, fiscal, registo criminal, etc., que devem ser guardados de forma segura e nunca transferidos para entidades externas ao processo. A descrição e um exemplo concreto, de cada um, são apresentados de seguida, evidenciando-se a aplicação de inteligência no sistema.

4.1. *Auditing* estatístico

O sistema de *auditing* proposto, deverá fazer mais do que gerar alertas resultantes de deteção de anomalias ou sugerindo automatização de *workflows*. Como a maior parte dos sistemas de *auditing* associados aos sistemas de IAM, este deverá responder a alguns requisitos básicos como:

- Manter o registo de todos os acessos a partir de qualquer serviço/localização, aprovisionamentos, alterações e remoções
- Auditar mudanças de permissões/perfil
- Colecionar eventos e alertas gerados
- Permitir gestão de ficheiros de log
- Registar e facilitar a resolução de vulnerabilidades e falhas de segurança
- Permitir a geração de *reporting* personalizados para identificar e planear a evolução do sistema

No contexto de *reporting* poderá ser interessante retirar relatórios com:

- ✓ A hora de maior carga do sistema – tendo em vista a avaliação de performance do mesmo e melhoria da eficiência (adaptabilidade à realidade do cliente)
- ✓ O nº de acessos em simultâneo – dimensionamento do sistema

- ✓ O nº de alertas gerados durante um período de tempo – melhorar a automatização do sistema
- ✓ O nº de contas removidas por falta de uso
- ✓ O nº de aprovisionamentos
- ✓ O nº de falhas de segurança resolvidas e não resolvidas
- ✓ O tipo de acesso a determinados serviços e os acessos privilegiados, assim como ações efetuadas por estes utilizadores privilegiados
- ✓ Lista de serviços mais acedidos

O *auditing* estatístico deve permitir visualizar informação, métricas que cada grupo possa considerar mais interessante. Por exemplo, um gestor poderá querer saber os acessos a um serviço, enquanto um administrador de segurança estará mais preocupado com as falhas que poderão ter sido detetadas. O fundamental é que o sistema implemente ou possa evoluir para implementar *reporting* à medida da necessidade do cliente. Um sistema configurável à medida do cliente é mais vantajoso, do que um sistema rígido. Sem, esquecer o ponto da melhoria contínua, de corrigir erros e permitir que o sistema evolua, este tipo de relatórios podem ser muito interessantes também para apoio à decisão de gestão, sobre a continuidade do sistema.

4.2. **Casos**

4.2.1. Caso 1 – Adição de serviço disponível

Como referido anteriormente, é habitual as ferramentas de IAM disponibilizarem um portal de *Self-service*, para que cada utilizador possa selecionar novos serviços que necessite, além dos previamente configurados pelo administrador aquando do aprovisionamento deste. Mediante uma determinada característica do perfil de utilizador poder-se-á filtrar se *n* utilizadores com essa mesma característica também solicitaram acesso ao(s) referido(s) serviço(s). Se o sistema detetar que esta situação se verifica para um nº N de utilizadores, poderá ser gerado um ALERTA que notificará o administrador, sugerindo a inclusão desse(s) serviço(s) no *workflow* de criação e aprovisionamento de utilizadores. Este processo poderá ser automatizado, de acordo com as regras do sistema de IAM vigentes.

Exemplo:

Ao departamento X estão afetos um conjunto alargado de colaboradores. Aquando do aprovisionamento de utilizadores são selecionados o conjunto de serviços disponíveis para os colaboradores pertencentes ao departamento. No caso concreto os

colaboradores têm acesso aos serviços ALM e NETM. Quando o João, colaborador deste departamento, entra no sistema de IAM para aceder aos serviços que necessita para realizar o seu trabalho, verifica que necessita de mais um serviço, o GEREM, para realizar as tarefas que lhe foram atribuídas. O mesmo acontece ao António, ao Luís e ao Pedro. Os colaboradores, a partir do portal de *Self-service* do sistema, adicionam o GEREM ao seu perfil (a ação de adição de um novo serviço geralmente não é automática, gerando uma entrada no processo de autorização definido na organização). O facto de terem sido vários os colaboradores a adicionarem os mesmos serviços faz com que o sistema envie um alerta ao administrador, sugerindo a automatização deste processo. O administrador poderá então decidir incluir este serviço no *workflow* de aprovisionamento de colaboradores do departamento X, de forma automática e aquando do seu aprovisionamento. Este processo automático, retira trabalho ao *heldpdesk*, permitindo ao mesmo tempo que, de forma automática os colaboradores tenham, no momento *zero*, todos os serviços disponíveis.

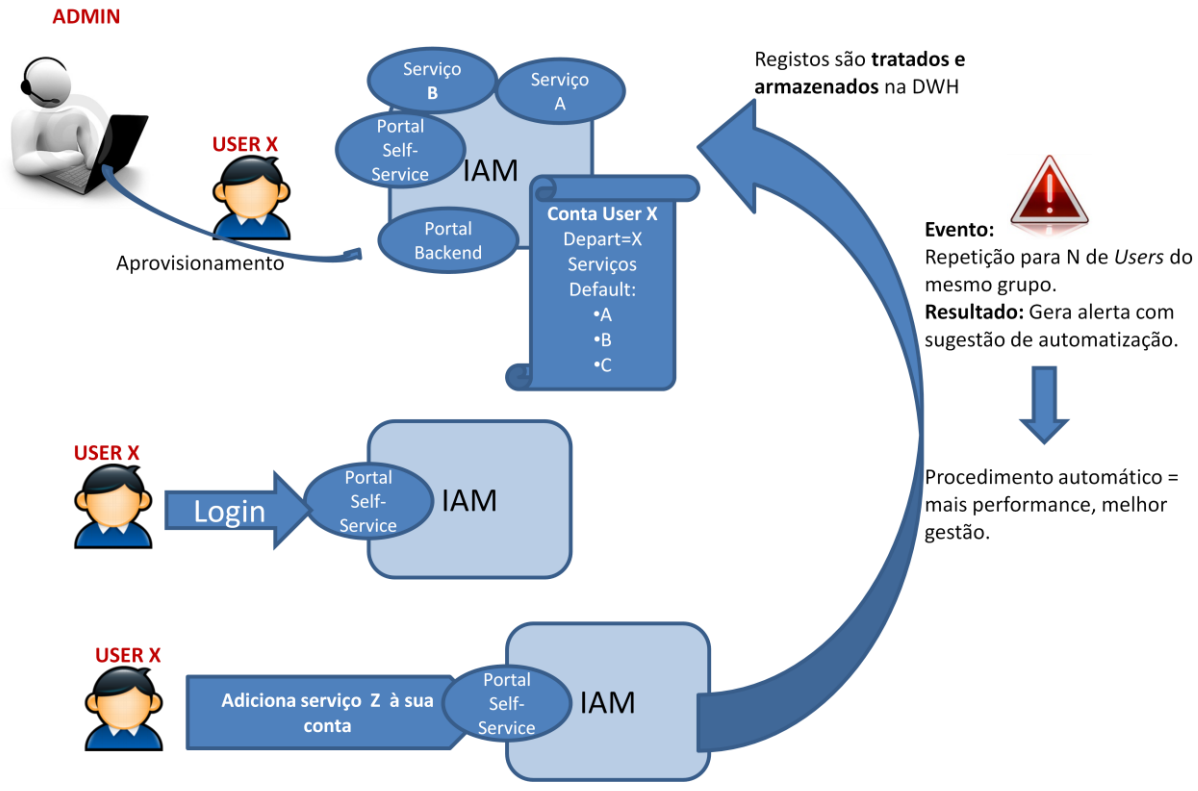


Figura 17. Ilustração do caso 1 - Adição de serviço disponível

Para implementar este alerta e a consequente automatização são necessários registos da atividade, que deverão conter a seguinte informação (todos os eventos aqui considerados deverão estar datados corretamente):

Evento	Dados a serem registados
Aprovisionamento de utilizador	• Conta de administrador utilizada • Nome de utilizador • Número identificativo (da empresa/organização/entidade) • Display Name
Atribuição de serviços por omissão a utilizador	• Conta de administrador utilizada • Nome de utilizador • Departamento/grupo/equipa atribuídos • Nome dos Serviços disponíveis por omissão
Atribuição de serviços através do portal de Self-service (on-demand)	• Nome de utilizador que fez alteração • Nome dos Serviços adicionados ao seu perfil

Os atributos necessários para implementar este caso são:

Atributo	Valor
Nome de utilizador	Username no sistema
Grupo/departamento	ID, sigla do grupo atual
Serviço adicionado	ID do serviço

Após o registo de *n* eventos de adição do serviço *x* aos colaboradores do grupo/departamento *y*, deverá gerar-se um alerta. O número *n* necessita ser definido na configuração/personalização inicial do sistema.

Este caso é do tipo de *auditing* com vista à automatização de processos e *workflow.*

4.2.2. Caso 2 – Mudança de funções

Dado o contexto de mobilidade empresarial que hoje em dia se vive, é vulgar a existência de mudanças de funções dos colaboradores (promoções/despromoções/mudança de tarefas). Existe portanto a necessidade de atualizar o perfil, a nível de permissões e/ou serviços, quando o utilizador muda de funções dentro da organização. Assim, não fará sentido que um utilizador que tinha privilégios para administrar um determinado serviço na sua região, continue a ter esse acesso após ser transferido para outro local. Este excesso de acesso poderá ser encarado como uma falha de segurança. Aqui também é útil um procedimento de alerta com automatização de *workflow*, para que estas tarefas não fiquem esquecidas. É natural pensar-se que casos deste tipo não devam existir, mas a realidade mostra-nos que é comum haver esquecimento de algumas tarefas quando se trata de mudanças de perfil. Por vezes, este facilitismo pode incorrer em falha de segurança.

Assim, o João que pertencia ao departamento X na organização onde exerce as suas funções, tinha acesso a um conjunto de serviços de gestão. O João muda de funções e vai para o departamento de testes TST onde as suas funções passam a ser a conceção e realização de testes, devendo por isso ter acesso às aplicações de teste usadas na organização. O sistema não faz automaticamente esta mudança, ou seja, o João passa a estar no departamento TST e terá acesso às aplicações de gestão (do departamento MGM) sendo adicionadas as aplicações de teste. O sistema deverá detetar esta mudança de perfil e gerar um alerta ao administrador, sugerindo que desative, para o utilizador João, os serviços no âmbito do X.

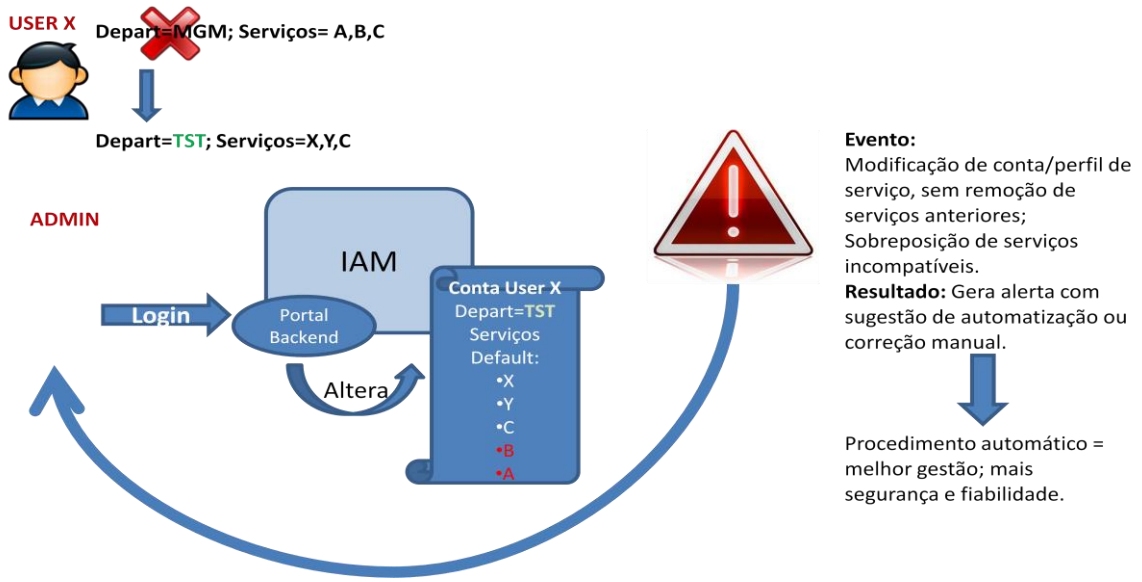


Figura 18. Ilustração do caso 2 - Mudança de funções

Para concretizar este alerta e a consequente automatização é necessário recolher os seguintes registos de atividade (todos os eventos aqui considerados deverão estar datados corretamente):

Evento	Dados a serem registados
Modificação de perfil de utilizador	• Conta de administrador utilizada • Nome de utilizador • Anterior Departamento/grupo/equipa • Novo Departamento/grupo/equipa atribuído
Atribuição de serviços a utilizador (se acontecer, para além dos atribuídos por omissão)	• Conta de administrador utilizada • Nome de utilizador • Departamento/grupo/equipa atribuídos • Nome do serviço atribuído

Os atributos necessários para implementar este caso são:

Atributo	Valor
Nome de utilizador	Username no sistema
Grupo/departamento	ID, sigla do grupo atual
Grupo/departamento	ID, sigla do grupo anterior

Detetada a mudança de perfil sem que ocorra remoção de serviços, o administrador é notificado e poderá decidir pela automatização. Este caso poderá ser incluído no grupo de auditing com vista a automatização mas também no de deteção de fraude/anomalias.

4.2.3. Caso 3 – Serviços com diferentes tipos de autenticação

Quando acedemos a um recurso é comum que o façamos usando sempre o mesmo método de autenticação. No entanto, serviços diferentes podem requerer um tipo de autenticação diferente, mais forte do que normalmente usamos, ou simplesmente diferentes serviços que implementam diferentes provedores de autenticação. Dotar o sistema de IAM de inteligência capaz de resolver esta questão é torná-lo mais eficiente e todo o sistema mais seguro. O João pertence a um grupo no IAM que necessita de acesso a dois serviços: o NETM e o WRLM. Cada um destes serviços suporta um determinado tipo de autenticação. O João costuma autenticar-se usando *username/password* da AD (Active Directory) da sua empresa – empresa X. O NETM aceita como autenticação este método, ou seja verifica as credenciais na AD da empresa X. O WRLM apenas aceita autenticação do mesmo tipo, mas na AD da empresa Z. Aqui apenas o provedor de autenticação é diferente. Assim sendo, após autenticação o João terá acesso ao NETM mas não terá acesso ao WRLM. Ou seja neste caso o sistema deverá gerar um alerta para o administrador, notificando-o da deteção desta incompatibilidade para que ele possa tomar as medidas necessárias e resolver o problema, tomando ou não a decisão de automatizar este processo, daí em diante. Ou seja, sempre que aconteça uma destas situações e não se possa retirar o acesso nem alterar a autenticação suportada por uma das partes envolvidas, o utilizador deve ser autorizado a fazer a sua autenticação por um método diferente e que seja suportado por ambos os sistemas ou ser autorizado a adicionar ao seu utilizador um novo método de autenticação ou provedor de autenticação. Assim, se por exemplo o administrador tomasse a decisão de adicionar ao utilizador João o método de autenticação *username+password* na AD da empresa Z, o problema ficaria resolvido.

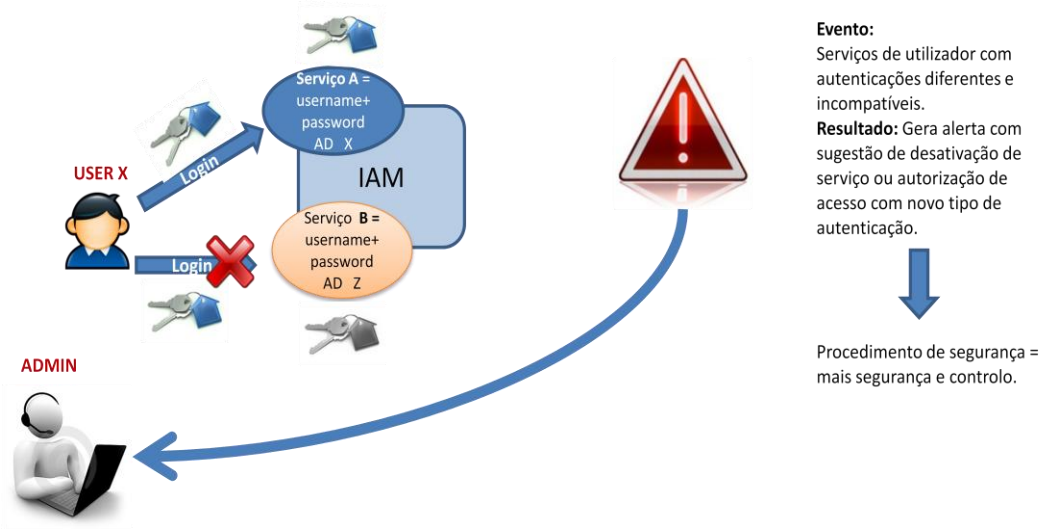


Figura 19. Ilustração do caso 3 - Serviços com diferentes tipos de autenticação

Para concretizar este alerta e a consequente automatização é necessário recolher os seguintes registos de atividade (todos os eventos aqui considerados deverão estar datados corretamente):

Evento	Dados a serem registados
Registo/modificação de tipo de autenticação de utilizador	• Conta de administrador utilizada • Nome de utilizador • Tipo de autenticação suportada
Registo/modificação de tipo de autenticação de serviço	• Conta de administrador utilizada • Nome de serviço • Tipo de autenticação suportada

Os atributos necessários para implementar este caso são:

Atributo	Valor
Nome de utilizador	Username no sistema
Autenticação de utilizador	ID de tipo de autenticação
Autenticação suportada pelo serviço	ID de tipo de autenticação

Uma vez detetada a incompatibilidade deverá existir uma notificação ao administrador que poderá tomar umas das seguintes medidas:

- Permitir que o utilizador aceda de forma diferente da suportada apenas a este serviço
- Automatizar o processo de adição de um novo método de autenticação ao referido perfil de utilizador.

Este caso também se insere no tipo de *auditing* com vista à automatização de processos e *workflow*. Este caso poderá ser concretizado através do processamento de eventos em tempo real, mas não é obrigatório que o seja. Um processo executado regularmente sobre os perfis de autenticação de utilizador e de serviços atribuídos ao mesmo, fazendo a comparação entre estes poderá ser suficiente para a deteção do problema.

4.2.4. Caso 4 – Contas não utilizadas

Um dos problemas mais comuns nas organizações, são as contas de utilizador não usadas, porque, por exemplo, o utilizador deixou a organização, e por isso a conta deixa de ser necessária. É comum manter-se a conta ativa durante um determinado período após a saída do colaborador, mas não por tempo indefinido.

Assim, sabendo que o João deixou a organização onde exercia funções, não faz sentido manter a sua conta e os seus perfis de acesso ativos durante mais tempo do que o estritamente necessário. Nesta situação, chegará o momento em que o João já não entra com as suas credenciais em qualquer dos serviços, pelo que ao fim de um determinado período de tempo o sistema deteta a conta como não usada e poderá gerar um alerta para o administrador. O administrador do IAM poderá desativar a conta de imediato ou prolongar a mesma, colocando-a num estado intermédio (*zombie*) por um período de tempo z. Ao fim de t+z a conta será desativada automaticamente. Qualquer uma das opções poderá ser tomada de forma automática, ou seja, quer desativar a conta no momento, ao fim de t ou ao fim de t+z. Deverá assim existir um procedimento associado de forma automática ao sistema de IAM que implemente este *workflow* de forma automática. Numa primeira abordagem o sistema de *auditing* poderá alertar o administrador para este acontecimento e sugerir a automatização.

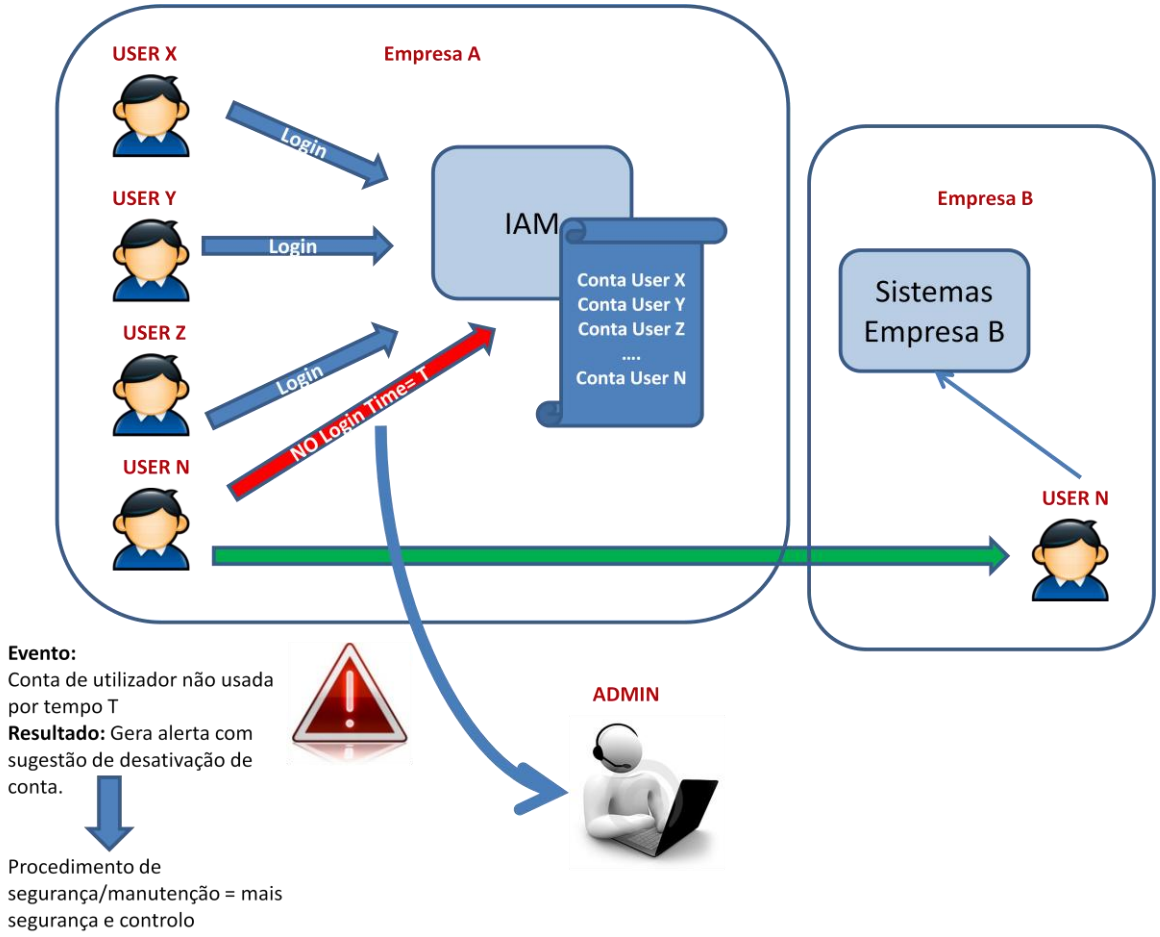


Figura 20. Ilustração do caso 4 - Contas não utilizadas

Para concretizar este alerta e a consequente automatização é necessário recolher os seguintes registos de atividade (todos os eventos aqui considerados deverão estar datados corretamente):

Evento	Dados a serem registados
Registo de entrada no sistema com conta de utilizador	• Nome de utilizador • Data de última entrada no sistema • Localização • Tipo de autenticação

Os atributos necessários para implementar este caso são:

Atributo	Valor
Nome de utilizador	Username no sistema
Data de última entrada	Data

Deverá existir uma rotina que percorra o sistema e guarde o respetivo tempo t de cada conta e compare com o valor pré- definido no sistema. Se foi previamente definido também um tempo z, a conta não deverá ser desativada até que este tempo t+z seja excedido.

4.2.5. Caso 5 – Sessões simultâneas

Um determinado utilizador acede regularmente a um dos serviços para os quais tem autorização, a partir de uma máquina localizada, por exemplo, no Porto. Quase em simultâneo é registado um novo acesso, mas desta vez a partir de uma máquina em Faro. Este acontecimento poderá considerar-se bastante estranho e poderá revelar-se uma tentativa de fraude. Perante isto deverá ser gerado um alerta para o administrador e este deve agir com celeridade para clarificar o problema. Esta ação não poderá ser automática uma vez que, imagine-se por hipótese, o sistema terminava uma ou ambas as sessões, correndo-se o risco de estar a terminar a sessão verdadeira. No entanto o administrador poderá contactar o utilizador e confirmar situação anómala. Uma outra possibilidade é a de não permitir que duas sessões existam em locais geograficamente distante e diferentes, em simultâneo.

Este é um caso que se identifica como pertencente ao tipo de *auditing* para deteção de anomalias/fraude.

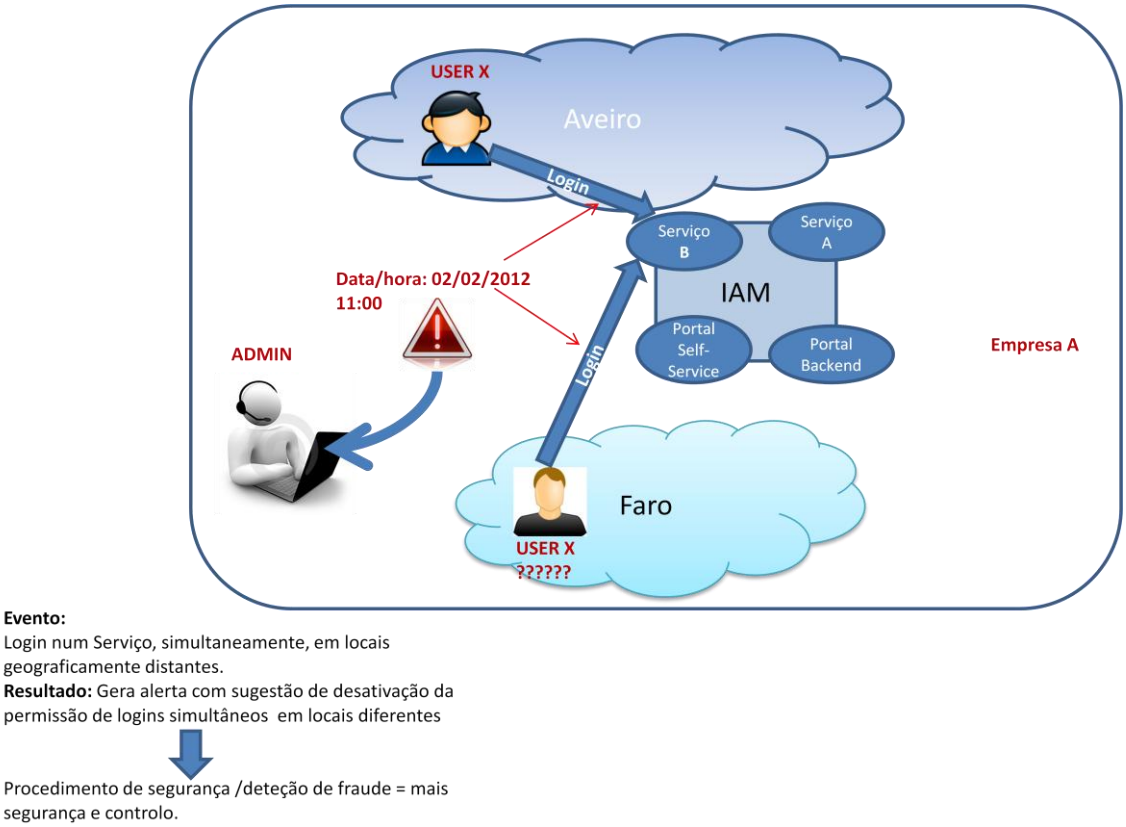


Figura 21. Ilustração do caso 5 - Sessões simultâneas

De forma a implementar este alerta é necessário recolher os seguintes registos de atividade (todos os eventos aqui considerados deverão estar datados corretamente):

Evento	Dados a serem registados
Registo de entrada no sistema com conta de utilizador	• Nome de utilizador • Data de última entrada no sistema • Localização • Tipo de autenticação

Os atributos necessários para implementar este caso são:

Atributo	Valor
Nome de utilizador	Username no sistema
Localização	ID de localização
Data de entrada	Data
Hora de entrada	Hora

Uma vez detetada a existência de sessões simultâneas para o mesmo utilizador, verifica-se a localização de ambas as sessões. Se as localizações forem diferentes e geograficamente distantes, o sistema deverá gerar um alerta para o administrador. Este após receber o mesmo alerta poderá:

- Terminar uma ou ambas as sessões e adicionar este procedimento ao *workflow* automático do sistema
- Apurar a razão destas sessões, junto da pessoa responsável pela conta de utilizador utilizada e só após esta tarefa, decidir a ação a tomar.

4.2.6. Caso 6 – Área privilegiada

O serviço NETM, é um serviço composto por diferentes serviços com níveis de acesso diferentes. O serviço tem uma área de visualização de alarmes e gestão dos mesmos, onde o João tem acesso, autorizado pelo seu perfil e onde se autentica usando *username/password* na AD da sua empresa. No entanto, existe ainda uma outra área, não visível a todos, uma vez que é uma área privilegiada de aprovisionamento de sistemas e alarmes. Para aceder a esta área os utilizadores deverão usar autenticação do tipo *fingertip.* Se o João entrou no sistema com o seu username/*password*, mesmo que tenha permissões para aceder a esta área, não irá conseguir a menos que use o tipo de autenticação solicitada. Assim, o sistema apresentar-lhe-á uma nova janela de autenticação por *fingertip* e deverá gerar um alerta para o administrador com a mensagem de existência de uma "incompatibilidade" entre a autenticação usada e a solicitada. Se a autenticação for bem-sucedida, o João conseguirá o acesso pretendido. Neste caso também é possível a automatização, ou seja, quando um utilizador precisa de acesso a uma área/sistema que exige um certo tipo de autenticação, deverá ser apenas autorizado que este aceda ao sistema, usando este método. Aquando da criação e aprovisionamento do utilizador João, este tipo de autenticação por *fingertip* deveria estar-lhe associado. É pois possível detetar estes acontecimentos e corrigi-los ou melhorá-los, automatizando as futuras decisões com base no histórico.

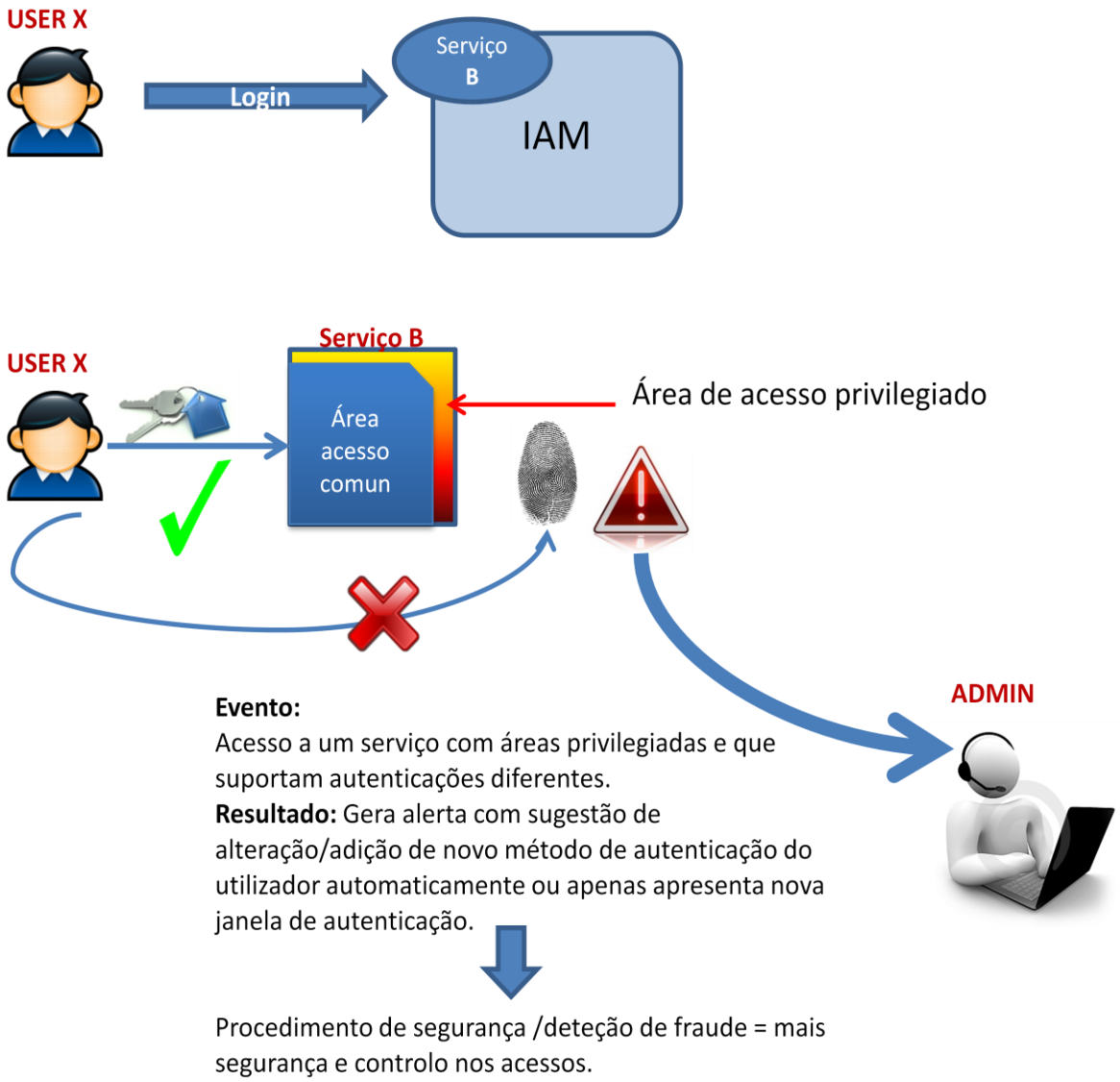


Figura 22. Ilustração do caso 6 - Área privilegiada

Para concretizar este alerta e a consequente automatização é necessário recolher os seguintes registos de atividade (todos os eventos aqui considerados deverão estar datados corretamente):

Evento	Dados a serem registados
Registo de entrada no sistema com conta de utilizador	• Nome de utilizador • Data de última entrada no sistema • Localização • Tipo de autenticação
Registo de tentativa de acesso a área privilegiada	• Nome de utilizador • ID área / sistema • Tipo de autenticação da área

Os atributos necessários para implementar este caso são:

Atributo	Valor
Nome de utilizador	Username no sistema
Autenticação de utilizador	ID de tipo de autenticação
Autenticação suportada pela área privilegiada	ID de tipo de autenticação

Uma vez detetada a incompatibilidade deverá existir uma notificação ao administrador que poderá tomar umas das seguintes medidas:

- Permitir que o utilizador aceda área privilegiada, através da inserção de uma autenticação nova do tipo suportado pela aplicação (nova janela de autenticação)
- Automatizar o processo de adição de um novo método de autenticação ao perfil do de utilizador, permitindo que este entre por este método e não tenha de se autenticar novamente.

4.2.7. Caso 7 – Mudança de comportamento de utilizador

A mudança de comportamento de um utilizador poderá, em determinadas condições ser considerada fraude.

Por exemplo, o João, no âmbito da realização das suas tarefas profissionais, autentica-se, geralmente, nos serviços a que está autorizado a aceder, através de autenticação com o seu Cartão de Cidadão. O sistema não conhece outro meio de autenticação registado pelo João. Um dia, o sistema regista uma ou mais tentativas de entrada do utilizador João, em que o tipo de autenticação usado é simplesmente *username e password*. Como este não é um método que se encontre associado ao utilizador João, este não será bem sucedido no acesso. Nesta situação o sistema deverá gerar um alerta e comunicar ao administrador

que há uma ocorrência estranha ou anomalia no comportamento esperado de utilizador. Este deverá proceder, de forma manual, à confirmação de que um novo tipo de autenticação está a tentar ser utilizado pelo utilizador em causa e não por alguém que esteja a tentar realizar algum acesso não autorizado.

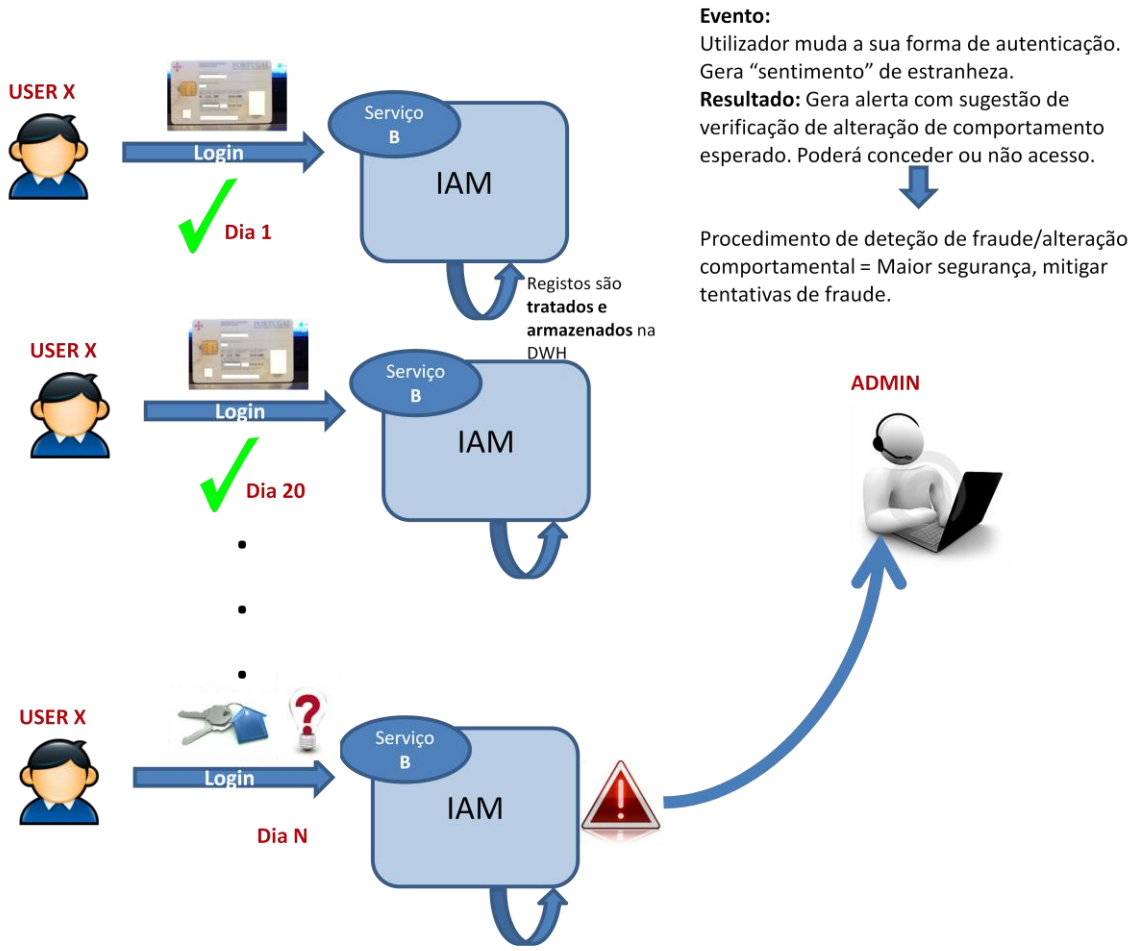


Figura 23. Ilustração do caso 7 - Mudança de comportamento de utilizador

Para concretizar este alerta é necessário recolher os seguintes registos de atividade (todos os eventos aqui considerados deverão estar datados corretamente):

Evento	Dados a serem registados
Registo de entrada no sistema com conta de utilizador	• Nome de utilizador • Data de última entrada no sistema • Localização • Tipo de autenticação

Os atributos necessários para implementar este caso são:

Atributo	Valor
Nome de utilizador	Username no sistema
Autenticação de utilizador	ID de tipo de autenticação

Uma vez detetada a tentativa de entrada do utilizador com um tipo de autenticação diferente do usual, o sistema gerará um nega o acesso e alerta para o administrador, que deve confirmar o mais breve possível o que se está a passar e uma vez confirmada a mudança de método junto do utilizador, deve proceder à adição do novo método associando-o à conta de utilizador. Para uma melhor performance, os utilizadores deverão estar informados das regras de segurança da empresas e dos sistemas, de que sempre que necessitem de adicionar um novo método à sua conta, deverão fazê-lo previamente à tentativa de acesso.

4.2.8. Caso 8 - Conta removida enquanto usada

O caso descrito de seguida é claramente um caso em que o *workflow* deve ser interrompido automaticamente e lançado um alerta. Um administrador está a efetuar operações de manutenção ao IAM e por engano apaga uma conta de utilizador. Simultaneamente existe uma sessão autenticada desse mesmo utilizador, pelo que em situação normal, este ficaria sem acesso de imediato e muitas vezes sem qualquer tipo de aviso. Numa situação normal também poderiam existir notificações para chefias etc., de que a conta tinha sido removida pela razão x ou y. Este é um caso muito desagradável e muito prejudicial ao trabalho do utilizador sendo mesmo um caso que revela uma certa falta de atenção e zelo por parte do administrador. Assim sendo, antes de uma conta ser removida devem ser verificados alguns pré-requisitos como por exemplo:

- Verificar se a data de última entrada é superior a t tempo
- Verificar que não existem de momento sessões ativas
- Confirmar que a conta deve mesmo ser removida e qual a razão

Assim, se por alguma razão o administrador tentar remover uma conta que se encontre ativa e com sessão iniciada, vai receber um alerta grave de que existe a sessão e não deverá deixar que a conta seja removida, a menos que existe razão de força maior para o fazer (por exemplo uma entrada não autorizada numa conta que deveria ter sido removida, tentativa de fraude etc.).

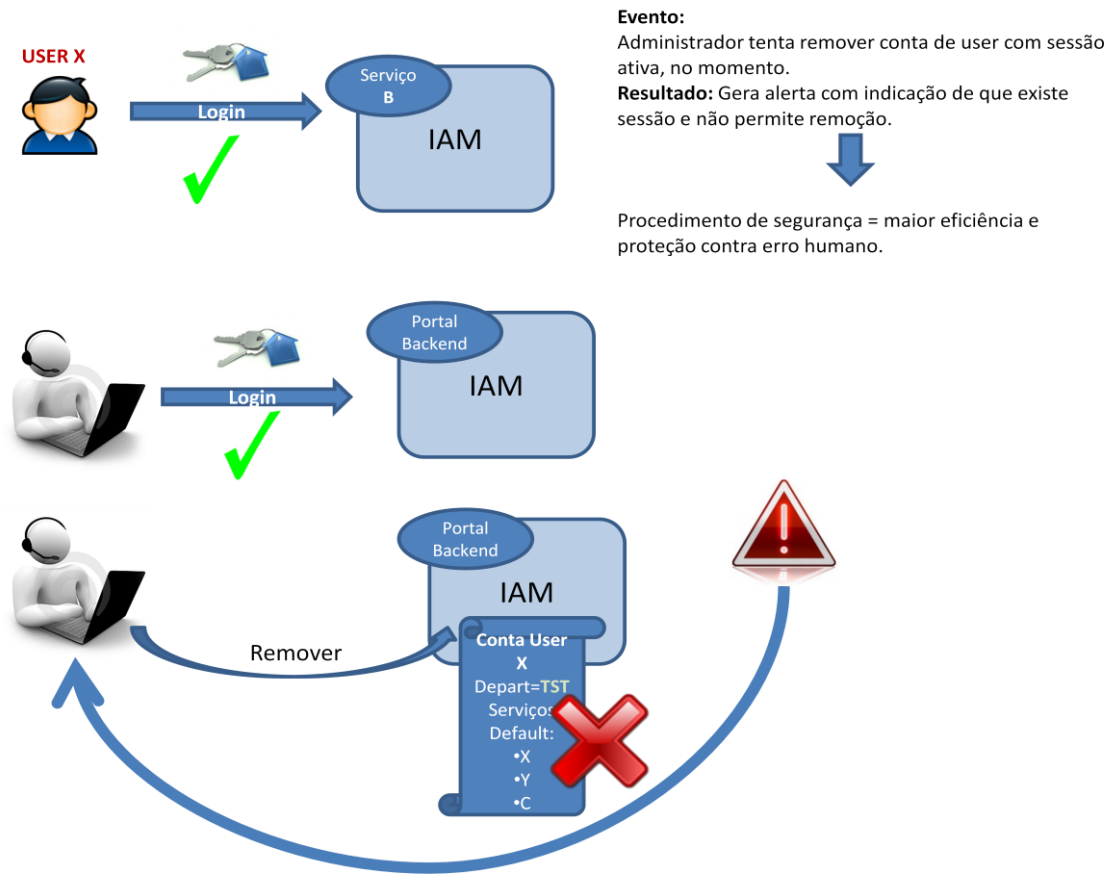


Figura 24. Ilustração do caso 8 - Conta removida enquanto usada

Para concretizar este alerta é necessário recolher os seguintes registos de atividade (todos os eventos aqui considerados deverão estar datados corretamente):

Evento	Dados a serem registados
Registo de entrada no sistema com conta de utilizador	• Nome de utilizador • Data de última entrada no sistema • Localização • Tipo de autenticação
Registo de entrada no portal *backend* pelo administrador	• Nome do administrador • Data de última entrada no sistema
Registo de modificação de conta de utilizador	• Conta de administrador utilizada • Nome de utilizador • Departamento/grupo/equipa atribuídos • ID de tarefa de alteração

Os atributos necessários para implementar este caso são:

Atributo	Valor
Nome de utilizador	Username no sistema
Nome de administrador	Username no sistema
Sigla/ID de tipo de alteração	ID de alteração

Uma vez detetada a entrada do utilizador, é registada a sessão e quando o administrador tenta efetuar uma operação de remoção sobre a conta, imediatamente lhe é apresentado na consola um alerta de sessão ativa, questionado se pretende continuar, e impedindo que a operação se concretize por exemplo, sem que seja colocada uma autenticação mais privilegiada (obrigando assim a LER o alerta).

5. Arquitetura e modelo de dados

Para a implementação de um sistema de *auditing*, paralelo ao sistema de IAM, que analise e permita inferir conhecimento a partir dos registos de atividade, tal como o BI na sua definição original o faz para dados de gestão de negócio, é necessário definir uma arquitetura que servirá de base ao sistema e ainda um modelo de dados que suporte as atividades e procedimentos a implementar. Assim, nesta secção descreve-se a arquitetura proposta e um possível modelo de dados, à semelhança dos sistemas de BI.

Não serão sugeridas ferramentas uma vez que a sua escolha requer uma análise mais aprofundada das funcionalidades de cada uma e dos vários pontos fortes e fracos e só depois poderá ser tomada uma decisão. Existem inúmeras soluções de mercado que implementam ETL, repositórios de metadados, OLAP, *Data warehouse*s, Data Marts, etc., que deverão ser comparadas e proceder-se à escolha da que mais aspetos satisfaz, nos requisitos do projeto de *auditing* IAM.

5.1. **Arquitetura**

A arquitetura proposta é genérica e poderá ser aplicada a qualquer sistema de IAM. Os registos de atividade são a base de todo o trabalho processual. Estes poderão ser simples ficheiros de texto ou bases de dados e são a fonte de toda a informação de PII a ser processada pelos sistemas de *auditing* e pelos sistemas de *reporting* do IAM.

A arquitetura inclui ainda um componente que engloba uma tecnologia emergente de processamento em tempo real ou quase real. Esta tecnologia é o CEP (*Complex Event Processing*). O CEP permite que as organizações tenham acesso ao conhecimento gerado em condições excecionais no que respeita ao tempo (tempo real) e esse fator pode ser uma mais-valia, uma vez que permite uma reação mais rápida. Todo este conhecimento permitirá um aumento da agilidade quer em termos empresariais/organizacionais quer do ponto de vista de gestão, quer na capacidade de rapidamente detetar situações anómalas e fraudulentas.

O IAM e todos os seus componentes, desde serviços, a portais *Self-Service* e *Backend*, produzem registos que contém informação considerada por definição PII, e que alimentam todo o sistema paralelo de *auditing*. Assim, todos os registos gerados pelo IAM e componentes, devem ser registados nos respetivos suportes de informação (bases de dados, ficheiros CSV ou Excel, etc.). Os dados gerados pelos registos passam ainda pelos processos do sistema de CEP onde são submetidos a várias fases de processamento,

refinamento e *assessement* até que os resultados sejam disponibilizados ao utilizador. Estes resultados são o conhecimento que permitirá ter uma reação/ação mais rápida por parte da organização, como referido anteriormente.

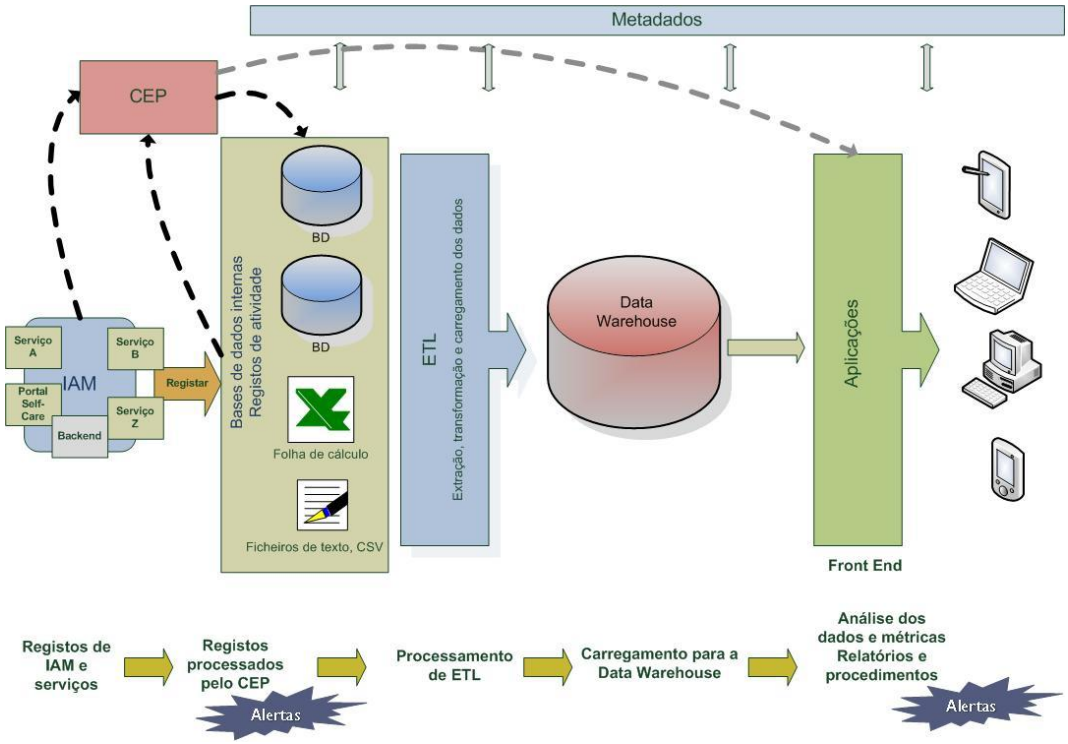


Figura 25. Arquitetura de *Auditing* do IAM

À posteriori, os dados recolhidos são alvo de um procedimento que faz a extração da informação relevante, o seu tratamento e transformação, e carregamento dos dados na respetiva *Data Warehouse*. A escolha de uma *Data Warehouse* para armazenamento dos dados é justificada pela própria definição de *Data Warehouse*, que segundo Inmon [29], "é uma coleção de dados orientada por assuntos, integrados, variáveis com o tempo e não voláteis para dar suporte ao processo de tomada de decisão." Assim, uma *Data Warehouse* orientada ao processo de *auditing* de identidades e acessos permitirá agregar de forma organizada grandes volumes de informação. A correta especificação e implementação da *Data Warehouse* torna o processo de análise e recolha de informação baseada nos dados aí armazenados muito mais rápido e eficiente, sendo este um dos mais importantes componentes dos sistemas de gestão e *reporting*.

Uma vez armazenados, os dados de PII são objeto de análise e processamento, tendo em vista a obtenção da informação de *auditing* pretendida seja estatística ou de automatização de *workflow*. Serão gerados alertas, sugestões, informação analítica e *reports* que podem ser consultados/ visualizados em qualquer tipo de equipamento fixo ou móvel através de um *front-end* específico.

Existe ainda a possibilidade de se incluírem no sistema *Data Marts*, isto é, uma camada de acesso à *Data warehouse* sobre um assunto específico, uma espécie de subconjunto de dados especializados sobre uma determinada linha de negócio ou grupo dentro da organização, por exemplo, um *Data Mart* de automatização de *workflows*, outro de dados de segurança, e outro ainda de dados analíticos.

Ao longo de todo o sistema são ainda gerados dados chamados de metadados que, por definição, são os "dados dos dados". Estes metadados são essenciais para a gestão dos sistemas de qualquer organização pois garantem a uniformização, fiabilidade e coerência dos dados. São indispensáveis para localizar os dados pretendidos e monitorar o processo de transformação dos mesmos desde a fonte até ao interface com o utilizador.

5.1.1. Complex Event Processing e Business Intelligence

Após uma análise dos possíveis registos da aplicação e tendo em vista a satisfação dos casos de uso descritos, podemos concluir que determinado tipo de reação/informação, será extremamente útil se for dada em tempo real ou quase real, tornando-se menos útil caso o alerta esteja desfasado da realidade em algumas horas. Assim, independentemente do processo de tratamento e transformação dos dados de atividade registados, que irão ser depois colocados na *Data warehouse* para produzir informação de BI, alguns dos dados devem ser tratados antes de ocorrer o processamento de extração, transformação e carregamento dos dados na *Data warehouse*. Este é o processo de CEP proposto na arquitetura que permitirá que os dados gerem alertas em tempo real ou quase real, permitindo também uma maior eficiência, eficácia e segurança na ação a tomar por parte do administrador. Este tipo de informação é relevante para a tomada de decisão, por parte da gestão, mas é ainda mais importante do ponto de vista da segurança dos sistemas.

Um motor ou uma aplicação de CEP deve ter em conta alguns aspetos fundamentais, uma vez que irá processar grandes quantidades de informação e neste caso particular, há que ter em conta a confidencialidade e segurança da mesma, por se tratar essencialmente de dados pessoais e dados de registos de atividades realizados pelos indivíduos, ou seja PII. Uma aplicação CEP processa muita informação proveniente dos eventos gerados pelo IAM e serviços associados, analisa todos eles segundo critérios de interesse pré configurados e apresenta os resultados pretendidos em tempo real. Deve integrar facilmente com os outros blocos da arquitetura e possuir um componente de depuração, caso seja necessário. Outros aspetos não menos relevantes na escolha da aplicação CEP são a resiliência, a alta disponibilidade e performance da mesma, assim como os aspetos

relacionados com a segurança da informação e da aplicação e ainda a inclusão da capacidade de monitoria às suas ações/resultados.

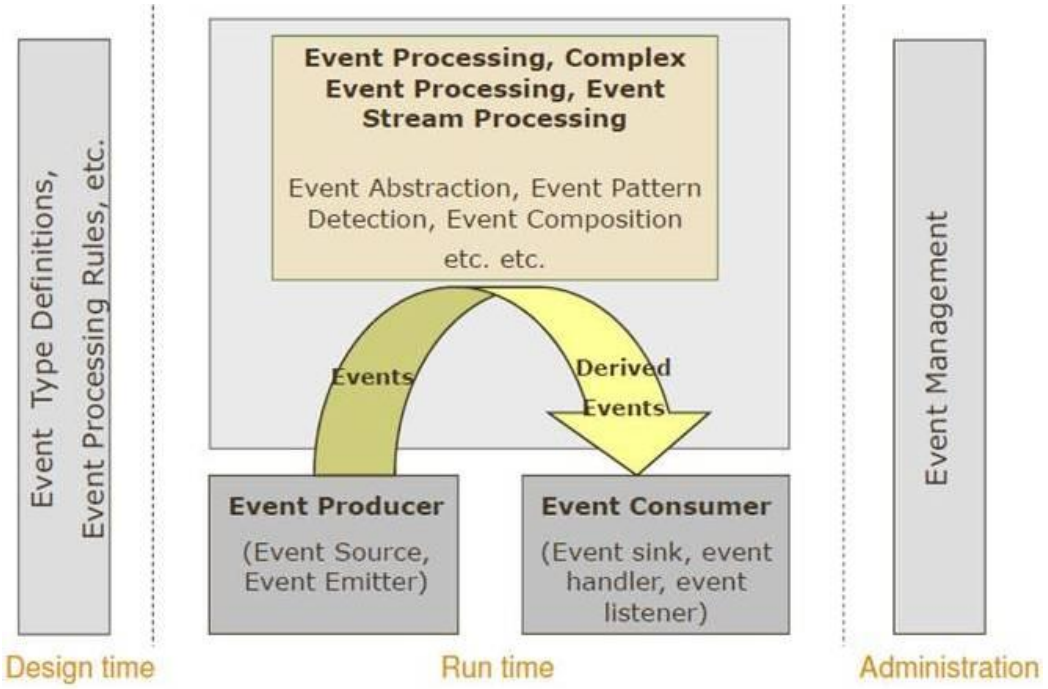


Figura 26. Visão funcional do processamento de eventos [30]

Por outro lado, um sistema de IAM deve também incluir uma *Data warehouse* e um sistema de *reporting* que permita retirar e apresentar métricas e eventos segundo um interface amigável e agradável ao utilizador, que poderá ser um técnico ou mesmo um gestor. Deverá apresentar resultados conclusivos e que apoiem a tomada de decisão. Aqui uma das caraterísticas mais importantes é a construção de um modelo de dados adequado ao fim que se pretende (análise baseada em PII).

Assim estes dois sistemas/blocos da arquitetura são em si complementares. Existem casos em que o processamento para alerta deve ser feito em tempo real e essa tarefa será levada a cabo pelo sistema de CEP. Outros casos podem ser obtidos por processamento da *Data warehouse*, segundo procedimentos regulares que ocorrem de *t* em *t* tempo. Outros ainda podem ser obtidos por meio de um processamento híbrido, isto é usando o sistema de CEP com consulta à *Data warehouse* de forma a confirmar uma ”suspeita” e gerar o alerta.

5.2. **Modelo de dados**

O modelo de dados proposto para a implementação deste sistema baseia-se no esquema em estrela (*Star Schema*). Um esquema deste tipo caracteriza-se pela existência de uma tabela central de factos e tabelas que representam as várias dimensões que constituem o sistema. Uma dimensão é uma categoria de informação e a tabela de factos inclui normalmente chaves primárias das várias dimensões e medidas numéricas que podem ser repetidas ao longo do tempo. As tabelas de dimensão contêm a informação detalhada e complementar da tabela de factos e que são de interesse para o negócio/análise. Um atributo é apenas uma parte da informação que podemos ter nas tabelas de dimensão, isto é, um nível numa dimensão. O modelo de dados vai acomodar a informação de uma forma bastante diferente da dos registos de dados, que são algo mais voláteis. Assim, ao processar os registos de dados e colocá-los num modelo dimensional é possível consolidar os mesmos e salvaguardar histórico, o que não seria possível se apenas se mantivessem os registos, sendo que esta é então a fase do ETL.

A informação contida na *Data warehouse* resulta pois do processo de ETL que deve ser executado a intervalos de tempo regulares, por exemplo, de hora a hora. O processo de ETL processa a informação dos registos regularmente e coloca-a disponível na *Data warehouse*, preenchendo a tabela de factos e as tabelas de dimensão, segundo o modelo de dados que aqui é proposto. Para melhor ilustrar o modelo de dados proposto apresenta-se o mesmo na figura seguinte:

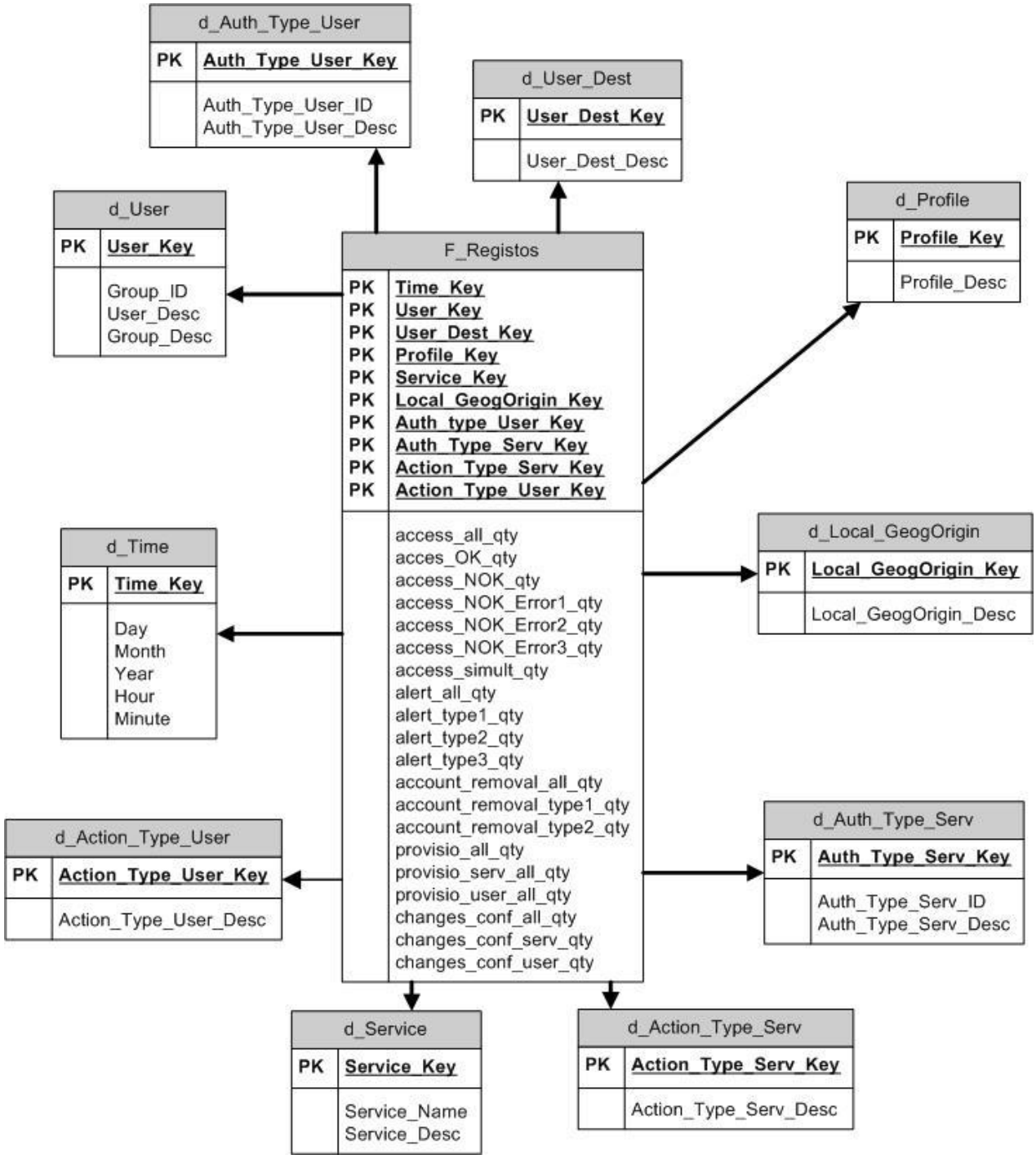


Figura 27. Modelo de Dados

Neste modelo podemos observar a tabela central de factos aqui denominada de **F_Registos**, dado que o nosso facto central e que é o alvo da análise a efetuar, são precisamente os registos de atividade. Esta tabela contém uma série de chaves primárias que permitem efetuar a ligação às tabelas de dimensão, representadas pelas tabelas que a circundam. Contém ainda várias métricas que são interessantes de analisar e de obter como resultados para análise posterior por parte da gestão ou de técnicos. As dimensões identificadas para este modelo são:

d_User - dimensão que identifica o utilizador
d_Time – dimensão tempo
d_User_Dest – dimensão que identifica o utilizador sobre o qual outro utilizador efetuou onde ocorre determinada ação realizada por outro
d_Auth_Type_Serv – dimensão que inclui o tipo de autenticação associada ao serviço
d_Auth_Type_User – dimensão que inclui o tipo de autenticação associada ao utilizador
d_Profile – dimensão que comporta os perfis de utilizador
d_Local_GeogOrigin – dimensão que contém as localizações geográficas dos utilizadores quando acedem a serviços
d_Service – dimensão que identifica os serviços
d_Action_Type_Serv – dimensão– onde se identificam os tipos de ação que se podem fazer sobre serviços
d_Action_Type_User– dimensão onde se identificam os tipos de ação que se podem fazer sobre as contas de utilizador

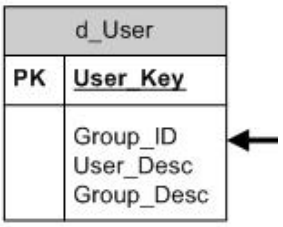
A tabela de factos não ficaria completa sem as métricas interessantes para apoiar a decisão/estratégia (métricas quantitativas como se utilizam no BI original). Foram identificadas algumas métricas que são interessantes para análises quantitativas de erros e ações. Essas métricas estão detalhadas na tabela seguinte:

Métrica	Significado
access_all_qty	Representa a contagem de todos os acessos a serviços, tendo como origem uma conta de utilizador
access_OK_qty	Representa a contagem de todos os acessos **bem sucedidos** a serviços, tendo como origem uma conta de utilizador
access_NOK_qty	Representa a contagem de todos os acessos **mal sucedidos** a serviços, tendo como origem uma conta de utilizador
access_NOK_Error1_qty **access_NOK_Error2_qty** **access_NOK_Error3_qty**	Representam a contagem de todos os acessos **mal sucedidos** a serviços, que se deveram aos erros do tipo 1, 2 ou 3 (a especificar no documento de requisitos do sistema) tendo como origem uma conta de utilizador. Os erros podem ser subdivididos em grupos por tipo, por exemplo: *password* errada, certificado expirado, *fingerprint* ilegível, etc.
access_simult_qty	Representa a contagem do nº de acessos simultâneos
alert_all_qty	Representa contagem de todos os alertas enviados pelo sistema
alert_type1_qty **alert_type2_qty** **alert_type3_qty**	Contagem de todos os alertas enviados pelo sistemas, subdivididos em grupos segundo o tipo de alerta, por exemplo alertas de erro, sugestões de automatização, fraude, etc.
account_removal_all_qty	Contagem de todas as contas de utilizador removidas
account_removal_type1_qty **account_removal_type2_qty**	Representa a contagem de contas removidas por razão para a remoção
provisio_all_qty	Número de aprovisionamentos
provisio_serv_all_qty	Número de serviços aprovisionados
provisio_user_all_qty	Número de contas de utilizador aprovisionadas
changes_conf_all_qty	Contagem de todas as alterações de configuração realizadas
changes_conf_serv_qty	Contagem de todas as alterações de configuração realizadas sobre serviços
changes_conf_user_qty	Contagem de todas as alterações de configuração realizadas sobre contas de utilizador

Apenas se identificaram algumas métricas e ao longo da implementação e configuração de cliente, podem ser necessárias e identificadas que devem por isso ser adicionadas modelo. O modelo é aberto permitindo a introdução de novas métricas.

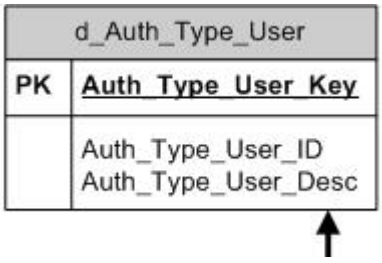
Para cada dimensão proposta pelo modelo de dados, existem atributos e o significado destes é apresentado de seguida para uma melhor compreensão do mesmo.

Dimensão d_User



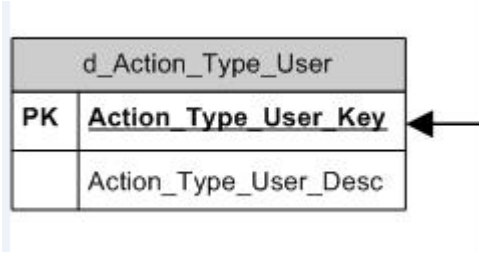
d_User	
Coluna	**Descrição**
User_Key (chave primária)	Identificador único no sistema de cada utilizador no sistema, ex. joao
Group_ID	Sigla ou identificador do departamento ou grupo de trabalho a que pertence o utilizador, ex. CTI
User_Desc	Breve descrição associada ao utilizador, podendo conter o nome, um número de identificação etc.
Group_Desc	Breve descrição associada ao grupo do utilizador.

Dimensão d_Auth_Type_User



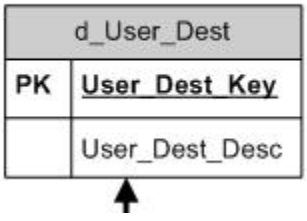
d_Auth_Type_User	
Coluna	**Descrição**
Auth_Type_User_Key (chave primária)	Identificador único no sistema de cada tipo autenticação que possa estar associada a um utilizador, ex. CERT
Auth_Type_User_ID	Pequeno identificador do tipo de autenticação, ex. “username e password “
Auth_Type_User_Desc	Breve descrição do tipo de autenticação

Dimensão d_Action_Type_User



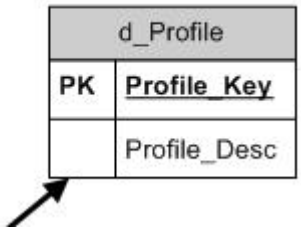
d_Action_Type_User	
Coluna	**Descrição**
Action_Type_User_Key (chave primária)	Identificador único no sistema para cada tipo de ação que possa ser realizada numa conta de utilizador,ex. adduser
Action_Type_User_Desc	Descritivo da ação realizada

Dimensão d_User_Dest



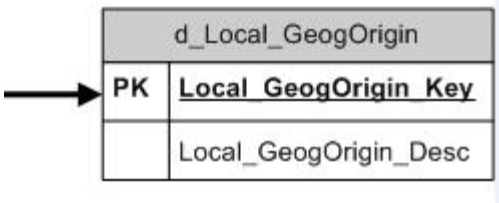
d_User_Dest	
Coluna	**Descrição**
User_Dest_Key (chave primária)	Identificador único no sistema que representa o utilizador sobre o qual foi realizada determinada ação.
User_Dest_Desc	Breve descrição do utilizador

Dimensão d_Profile



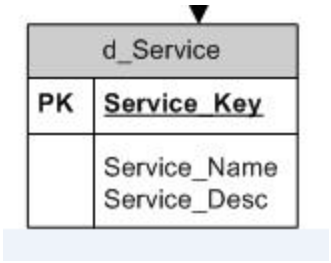
d_Profile	
Coluna	**Descrição**
Profile_Key (chave primária)	Identificador único no sistema de cada perfil que possa existir, ex. ADM
Profile_Desc	Breve descrição do perfil

Dimensão d_Local_GeogOrigin



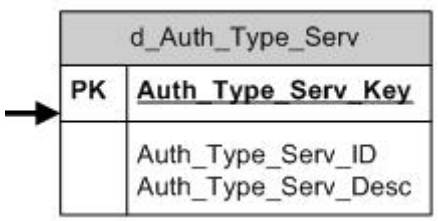
d_Local_GeogOrigin	
Coluna	**Descrição**
Local_GeogOrigin_Key (chave primária)	Identificador único no sistema de cada localização, que pode conter subdivisões por região se assim for possível de localizar, ex. AVR-01
Local_GeogOrigin_Desc	Descritivo da localização, ex. Aveiro - Cacia

Dimensão d_Service



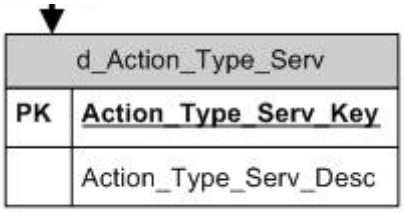
d_Service	
Service_Key (chave primária)	Identificador único no sistema de para cada serviço disponível, Ex. ALN
Service_Name	Nome do Serviço,ex. AlarmNet
Service_Desc	Breve descrição do serviço

Dimensão d_Auth_Type_Serv



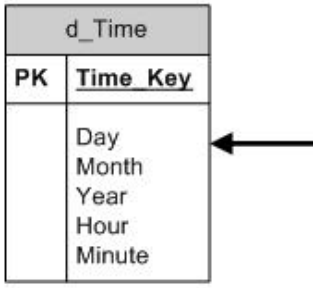
d_Auth_Type_Serv	
Coluna	**Descrição**
Auth_Type_Serv_Key (chave primária)	Identificador único no sistema de cada tipo autenticação que possa estar associada aos serviços, ex. UP
Auth_Type_Serv_ID	Pequeno identificador do tipo de autenticação, ex. “username e password “
Auth_Type_Serv_Desc	Breve descrição do tipo de autenticação

Dimensão d_Action_Type_Serv



d_Action_Type_Serv	
Coluna	**Descrição**
Action_Type_Serv_Key (chave primária)	Identificador único no sistema para cada tipo de ação que possa ser realizada sobre um serviço, ex. addserv
Action_Type_Serv_Desc	Descritivo da ação realizada

Dimensão d_Time



d_Time	
Coluna	**Descrição**
Time_Key (chave primária)	Identificador único no sistema de cada momento temporal, ex. 250620121130
Day	Dia do mês
Month	Mês
Year	Ano
Hour	Hora
Minute	Minuto

O modelo que aqui se é indicativo para implementação de um protótipo. No decorrer da implementação poderão surgir necessidades de novos atributos ou métricas que devem ser incluídos no modelo. Existem algumas práticas correntes no desenvolvimento de modelos dimensionais como a inclusão de campos genéricos que se destinam a permitir que o modelo cresça conforme a necessidade. Esta prática não é aqui observada, o que não implica que este modelo não possa ser adaptado às necessidades durante a implementação, incluindo campos adicionais que possam servir o crescimento do mesmo. Uma outra prática comum é a utilização de chaves primárias sequenciais e que não incluem dados, existindo um outro campo que mapeia essa chave no código da dimensão, por exemplo, *user_Key = 1001* e *user_map=joaosantos.* Na especificaçãpo deste modelo também não se usou esse conceito de forma a simplificar o modelo, deixando essa decisão para mais tarde, aquando da implementação.

5.2.1. O modelo e os casos

O modelo de dados que é aqui apresentado é suficiente para resolver alguns dos casos propostos, nomeadamente aqueles que não necessitam de processamento de eventos em tempo real. Assim, para alguns dos casos identificados como sendo possível o seu processamento com base na *Data warehouse*, faz-se de seguida uma breve demonstração textual de que é possível resolver o caso, com os atributos que se propõem no modelo de dados.

O caso 1- *Adição de serviço disponível* é adequado ao processamento sobre a *Data warehouse,* e algo incomportável sobre um sistema de eventos, uma vez que periodicamente os registos devem ser tratados e movidos dos seus ficheiros para a *Data warehouse*, de forma a evitar que estes cresçam até ao limite. Assim, o procedimento destinado a identificar este caso poderá ser executado periodicamente, por exemplo todas as semanas no final da semana, da seguinte forma:

Selecionar **Action_Type_User_Key = addserv** (adição de serviço a conta de utilizador) e para cada **Service_key**, contar o número de vezes, ou seja o número de utilizadores que executaram a ação addserv para um mesmo serviço, sendo que o **Group_ID** deve ser o mesmo em cada **User_Key** (por utilizador do mesmo grupo). Se se verificar esta premissa mais do que x vezes, ou seja o *count_addserv_serv1_groupy >=x*, então o sistema deve lançar um alerta (e-mail por exemplo ou SMS) ao administrador do sistema sugerindo que o serviço em questão passa a estar incluído no perfil *default* dos utilizadores do grupo y.

O caso 4 - *Contas não utilizadas* também pode ser resolvido com algum processamento sobre a *Data warehouse* e tal como os restantes casos que se inserem neste tipo de processamento, seria incomportável fazê-lo sobre os registos de dados, pela impossibilidade de manter um histórico dos mesmos a longo prazo. O procedimento destinado a verificar a ocorrência de contas não utilizadas durante um certo tempo, deverá ser efetuado de forma periódica, tal como o anterior, sendo que a base temporal deve estar ajustada ao tempo máximo de conta não utilizada que previamente se configurou. Assim:

Para cada **User_Key** obter as últimas interações com o sistema/serviços, ou seja listar todos os **Action_Type_Serv_Key** e respetiva data de acesso. Efetuar uma ordenação por data decrescente e selecionar da primeira data da lista *time_Key1***.** Comparar *time_Key1* com da data atual do sistema, *time_KeyNow*. Se o resultado da diferença de datas for superior ao tempo *t* definido como o máximo tempo de não utilização de uma conta, então o sistema deverá gerar um alerta, que tal como no caso anterior poderá ser nas mais diversas formas, alertando as contas de utilizador que se verificam esta condição e sugerindo a automatização do procedimento de encerramento de conta.

5.3. **Processamento de eventos**

Para o processamento de eventos, ou seja dos registos de atividade, deve implementar-se um motor de processamento. Os registos de atividade vão gerar acontecimentos/eventos que contêm PII e que são depois verificados contra as políticas configuradas. Segue-se o teste da condição/regra associada e se esta for verificada como verdadeira será executada uma ação, que poderá ser um envio de alerta (por mail, SMS ou outro), ou ainda a sugestão de automatização de um procedimento.

Como já foi referido, nem todos os eventos são passíveis de processamento via CEP, uma vez que a informação de PII necessária para obter resultados poderá já não estar guardada nos registos de atividade, mas sim já ter passado pelo processo de ETL e ter sido guardada na *Data warehouse* de forma segura.

Tomando como exemplo o caso de uso 5 - *Sessões simultâneas* consegue-se perceber que este é um dos casos que deve ser tratado pelo sistema de CEP e cujo resultado deve ser conhecido o mais próximo possível do tempo de acontecimento uma vez que se se tratar de uma tentativa de fraude, a ação para mitigar a ameaça deve ser tomada o mais rapidamente possível.

O fluxo de eventos que deverá resultar no alerta será:

Evento a: João acede com sucesso à aplicação ALM

data	ação	utilizador	serviço	local
02-05-2012 16:02	Login ok	joao	ALM	193.158.111.1

Evento b: João acede com sucesso à aplicação ALM

data	ação	utilizador	serviço	local
02-05-2012 16:10	Login ok	joao	ALM	194.135.88.2

Existem vários processos que são executados regularmente, por exemplo a cada 15 m, e que processam os dados de registo dos diversos serviços IAM e dos próprios portais. Estes processos contêm regras que são verificadas contra os registos. Neste caso, se o utilizador joao tem dois logins com sucesso e quase em simultâneo, então o processo vai determinar a localização dos endereços IP origem das sessões. Assim, por exemplo, **IPa** é de Aveiro e **IPb** é de Faro, logo suscita estranheza que um utilizador aceda a um determinado serviço em simultâneo mas em locais geograficamente distantes. A partir daqui o sistema gera um alerta para o administrador que indica a possibilidade de falha de segurança grave. Este deve diligenciar para que o facto seja analisado o mais rápido possível e na dúvida as sessões devem ser terminadas e averiguada a situação. Com esta reação rápida o sistema IAM ganha em segurança e eficiência na resolução dos problemas.

O caso 8 – *Conta removida enquanto usada* também poderá ser processado pelo sistema de CEP do *auditing* do IAM. Na tentativa de remover uma conta de utilizador o administrador deve depara-se com um alerta caso se encontre ativa uma sessão dessa conta. Para detetar este problema os dados necessários são:

Evento a: administrador acede com sucesso ao *backend*

data	ação	utilizador	serviço	local
02-05-2012 16:02	Login ok	adm-pedro	bckend	200.200.200.3

Evento b: administrador executa comando de remoção da conta do João

data	ação	utilizador	conta_dest	local
02-05-2012 16:05	Account_del	adm-pedro	joao	200.200.200.3

Evento c: João acede com sucesso à aplicação ALM

data	ação	utilizador	serviço	local
02-05-2012 13:00	Login ok	joao	ALM	194.135.88.2

Quando o sistema regista a operação de remoção de conta (evento a), este deve despoletar um processo de verificação de validade da operação antes de efetivar a mesma, isto é, o sistema de CEP deve procurar a existência de uma entrada bem sucedida do utilizador com *utilizador (evento c) = conta_dest (evento b)* sem que exista um evento de fim de sessão e o tempo decorrido entre o acesso bem sucedido e o momento do teste seja menor que t, sendo t o tempo máximo de sessão sem atividade a partir do qual a sessão expira. Assim, se existir sessão ativa, ou seja se se verificar a existência de evento b e c para a referida conta de utilizador, o administrador não deve conseguir concretizar a sua ação e deve ser alertado para este facto.

6. Conclusão e trabalho futuro

Este trabalho começou com um estudo sobre o estado da arte e conceitos associados ao *Business Intelligence*. Durante o estudo surgiram algumas dificuldades relacionadas em grande parte com o tipo de informação encontrada: informação sobre *software* de BI e não sobre a temática do BI propriamente dita como conceitos, evolução e novas tendências. O BI está em tudo relacionado com a gestão empresarial e económica. As ferramentas em BI estão orientadas ao processamento de métricas (quantidades de produto, quantidades monetárias, etc.) que após serem processadas e relacionadas permitem uma decisão baseada e informada por parte de administradores e gestores de organizações. O BI surge num contexto empresarial para apoiar decisões estratégicas de negócio. Apesar de ter alguns anos, o BI evoluiu e procura ser cada vez mais eficiente. Estes sistemas baseiam a sua análise nos dados recolhidos e armazenados ao longo do tempo permitindo gerar relatórios e manter o histórico. Tal como os outros sistemas também o BI deve ser alvo de processos de melhoria contínua que se destinam a afinar os processos de inferência e correlação de informação.

No decurso do trabalho foi também necessário estudar alguns conceitos sobre a temática da gestão de identidades, informação pessoal verificando-se que existem já algumas propostas de aplicação dos conceitos do BI a este tipo de informação, aparecendo estas propostas sobre o conceito de *Identity Intelligence*. A informação é escassa e mais uma vez era na sua maioria descrição de produtos nesta área, mas ajudou a perceber que é possível a adaptação destes conceitos a sua aplicação à informação do tipo PII, que se pretende tratar. Embora os processos de BI sejam mais orientados ao negócio e ao apoio à decisão nos processos com ele relacionados, que geralmente envolvem vendas, custos e proveitos, o conceito por detrás do BI pode ser aplicado ao tratamento de outro tipo de informação que por sua vez, também pode ser correlacionada, como os dados de acesso a serviços, por exemplo.

Por se pretender tratar dados que podem conter informação pessoal e privada há que tomar medidas para impedir que ocorram falhas de segurança e manipulação ilegal dos mesmos, ou seja, deve existis a preocupação com a segurança dos dados. Após uma breve análise às leis que regulam e protegem os dados e transmissões conclui-se que numa organização onde se pretenda implementar um sistema de IAM e um sistema de *auditing* dos seus dados, esta deve tomar as medidas necessárias para garantir que os dados não são manipulados de forma ilícita. Medidas como autorizações por parte dos proprietários da informação para a salvaguarda de dados e pedidos à CNPD para gerir os mesmos com um determinado fim explicitamente identificado. Medidas que impeçam a

transmissão direta dos dados entre entidades e medidas adicionais de segurança aos sistemas que manipulam os dados e os guardam.

Também no que está relacionado com a manipulação da informação pessoal, existe uma diversidade de formatos e fabricantes de ferramentas e projetos de desenvolvimento, sendo por isso necessário desenvolver esforços no sentido da normalização (formatos de registos, formatos de transmissão, etc.). Neste âmbito o TM Fórum é a entidade que está a desenvolver trabalho para garantir que mesmo que os componentes de um sistema de gestão de identidades sejam de proveniência diferente, os dados são convertidos e comunicados de forma *standard*.

Baseando-se nestes conceitos, a análise efetuada provou que também os dados considerados PII do ponto de vista legal, são passíveis de serem correlacionados, o que poderá levar à obtenção de novos dados e conhecimentos que por sua vez permitem gerar alertas e dotar um sistema de IAM de inteligência nos processos. A informação gerada por aplicações e serviços pode conter informação pessoal que identifica uma pessoa ou grupo inequivocamente. Esta informação é a PII e é nesta informação que se pretende aplicar técnicas utilizadas em BI. Por norma, esta informação está diretamente relacionada com os sistemas IAM, que gerem os acessos aos serviços e registam a atividade que ocorre nestes.

Os registos de atividade de um sistema de IAM contêm informação PII. Do estudo efetuado conclui-se que é possível tratar os dados PII e permitir que também eles apoiem a tomada de decisão numa empresa ou organização, relativamente a políticas de acesso aos seus serviços e segurança na utilização dos mesmos. No entanto e tratando-se de informação pessoal existem leis e diretiva que não podem ser esquecidas, uma vez que qualquer ação de processamento ou transferência de informação pessoal está sujeita a estas leis. Assim é possível e viável a implementação de um sistema de *auditing*, que funcione em paralelo com os sistemas de IAM e que alerte no caso de ocorrerem falhas ou anomalias. Existe já uma proposta de normalização dos sistemas de recolha de informação e *auditing* que, no geral, propõe a forma como a informação deve ser acedida e apresentada, em formatos independentes dos fabricantes.

Assim, este sistema de *auditing* paralelo permitirá inferir conhecimento a partir da informação de PII contida nos registos de atividade do IAM e para tal é necessária a implementação de uma *Data warehouse* e de políticas e processos associados a esta. Deve ainda incluir um módulo de processamento em tempo real, uma vez que existem diversas situações onde a informação rápida e reação breve podem colmatar falhas no sistema e detetar situações de fraude.

No decorrer deste trabalho foram identificados diversos casos de uso que ilustram a possibilidade de adicionar inteligência a um sistema de auditoria associado a um IAM. A inteligência não se limita ao processamento e relacionamento de quantidades e métrica, mas também à automatização e eficiência de processos e *workflows*. Os casos identificados englobam exemplos de uso que permitem detetar erros humanos, tentativas de fraude e oportunidades de melhoria dos processos. São apenas alguns dos possíveis casos e permitem demonstrar a forma como o modelo e arquitetura propostos se aplicam à sua resolução.

Para dar continuidade ao trabalho, num futuro próximo, é necessário garantir que os registos de atividade contêm toda a informação necessária à implementação do *auditing,* como se propôs anteriormente. O trabalho aqui apresentado será validado através da implementação de um protótipo baseado nos registos de atividade de todos os sistemas associados ao IAM. Este protótipo deve incluir os sistemas propostos na arquitetura. Durante a realização deste trabalho não foi possível a implementação do referido protótipo uma vez que a aplicação estava ainda em fase de desenvolvimento não existindo registos que o suportassem, e cujo conteúdo foi aqui proposto.

Lista de acrónimos

BI	Business Intelligence
BIIC	Business Intelligence Improvement Cycle
CE	Comissão Europeia
CEP	Complex Event Processing
CNPD	Comissão Nacional para a Proteção de Dados
CRM	Customer Relationship Management
DSS	Decision Support System
ES	Expert System
EIS	Executive Information System
ERP	Enterprise Resource Planning
EU	União Europeia
IAM	Identity and Access Management
IdM	Identity Management
IT	Information Technology
IT	Information Technology
ITIL	Information Technology Infrastructure Library
KPI	Key Performance Indicator
MIS	Management Information System
OCDE	Organização para a Cooperação e desenvolvimento Económico
OLAP	Online Analytical Processing
OLTP	Online Transactional Processing
PII	Personally Identifiable Information
ROI	Return of Investment
SOA	Service-oriented architectures
SSO	Single Sign on

Definições

Big data – conjunto de dados que cresce de forma desmesurada tão complexa que se torna muito difícil ou até improvável o seu tratamento com as ferramentas mais comuns de tratamento de dados.

Complex Event Processing – procedimento para análise de fluxos de informação e eventos associados, permitindo inferir conhecimento ou novos eventos. O objetivo do CEP é o processamento de eventos que permita obter padrões e responder a estes o mais rápido possível (por exemplo ameaças à segurança de informação).

Data Mart – é um subconjunto de dados de uma *Data Warehouse* que se referem, geralmente a um assunto em particular. São geralmente extrações de dados da *Data Warehouse* que se destinam a ajustar porções da mesma às necessidades de um grupo de pessoas ou departamento.

Data Mining – consiste em explorar grandes quantidades de dados tendo por objetivo encontrar padrões ou sequencias que permitam detetar novas relações e novos dados.

Online Analitical Processing (OLAP) – parte do BI e inclui o *reporting* e o *Data Mining.* Permite que os Utilizadores analisem interactivamente dados multidimensionais. Inclui a consolidação dos dados, a navegação pelos seus detalhes e a visualização de partes específicas destes.

Process Mining – é uma técnica de gestão de processos que consiste na análise do mesmo baseada nos *logs*/registos de eventos. Permite extrair informação dos registos com o objetivo de melhorar as técnicas que permitem perceber melhor o processo, os dados que gera, a informação de controlo e a estrutura dos registos que este gera.

Referências

1. Kobielus, J. *What's Not BI? Oh, Don't Get Me Started....Oops Too Late...Here Goes....* Abril 2010; Available from: http://blogs.forrester.com/james_kobielus/10-04-30-what%E2%80%99s_not_bi_oh_don%E2%80%99t_get_me_startedoops_too_latehere_goes.
2. Luhn, H.P., *A business intelligence system.* IBM J. Res. Dev., 1958. **2**(4): p. 314-319.
3. Olszak, C.M. and E. Ziemba, *Business intelligence systems as a new generation of decision support systems*, in *Proceedings of PISTA 2004 International Conference on Politics and Information Systems Technologies and Applications* 2004, The International Institute of Informatics and Systemics: Orlando: The International Institute of Informatics and Systemics.
4. Nicholls, C. *BI 2.0 The nest Generation*. novembro 2011; Available from: http://www.information-management.com/issues/20061101/1066763-1.html?zkPrintable=1&nopagination=1.
5. Negash, S. and P. Gray, *Business Intelligence* in *Americas Conference on Information Systems*2003.
6. Chee, T., et al., *Business Intelligence Systems: State of the art review and contemporary applications* in *Symposium on Progress in Information & Communication Technology*2009.
7. Software Group, S.C. *Business Intelligence solutions architecture*. novembro 2011; Available from: http://www.ibm.com/developerworks/data/library/techarticle/dm-0505cullen/.
8. Ltd, E. *The Business Intelligence Guide* dezembro 2011; Available from: http://thebusinessintelligenceguide.com/bi_strategy/The_BI_Lifecycle.php.
9. Quinn, K. *The Business Intelligence Improvement Cycle*. dezembro 2011; Available from: http://www.information-management.com/news/1045838-1.html?zkPrintable=1&nopagination=1.
10. Pentaho. *Pentaho*. dezembro 2011; Available from: http://www.pentaho.com/.
11. Microstrategy. *Business Intelligence*. dezembro 2011; Available from: http://www.microstrategy.com/Software/businessintelligence/index.asp.
12. Microstrategy. *Metadata*. dezembro 2011; Available from: http://www.microstrategy.com/Metadata/index.asp.
13. Oracle. *Oracle Business Intelligence Applications*. dezembro 2011; Available from: http://www.oracle.com/us/solutions/ent-performance-bi/bi-applications-066544.htm.
14. LINARES, M. *Identity and Access Management Solution*. janeiro 2012; Available from: http://www.sans.org/reading_room/whitepapers/services/identity-access-management-solution_1640.
15. Wikipedia. *Identity Intelligence*. janeiro 2012; Available from: http://en.wikipedia.org/wiki/Identity_intelligence.
16. Software, Q., *Identity and Access Management*.

17. Gandhi, N. *Oracle Identity Analytics 11g*. fevereiro 2012; Available from: http://identigov.wordpress.com/2010/07/22/oracle-identity-analytics-11g-all-systems-go/.
18. IBM. *IBMTivoli IdentityManager*. março 2012; Available from: http://public.dhe.ibm.com/common/ssi/ecm/en/tid10294usen/TID10294USEN.PDF.
19. Chobert, J.-P. *Develop a custom agent for IBM Tivoli Identity Manager with IBM Tivoli Directory Integrator*. março 2012; Available from: http://www.ibm.com/developerworks/library/ac-tim/index.html.
20. Today, S. *IBM Attacks the Complexity of Security with Identity Intelligence*. junho 2012; Available from: http://security-today.com/articles/2012/01/13/ibm-attacks-the-complexity-of-security-with-identity-intelligence.aspx.
21. Forum, T., *Security Compliance Audit automation, TMF_SCA_BA version 0.7*, 2012.
22. Europeia, C., *European Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data*, 24 de Outubro 1995.
23. Europeia, C., *European Directive 2002/58/EC on Privacy and Electronic Communications*, 12 de Julho 2002.
24. Europeia, C., *European Data Retention Directive 2006/24/EC, March 15, 2006.*, 15 de Março 2006.
25. Europeia, C., *European Directive 2009/136/EC amending Directive 2002/22/EC* 2009.
26. Portugal, *Leis, decretos* in *Decreto-lei nº 67/98*26-10-1998, Diário da República I Série.
27. Portugal, *Leis, decretos* in *Decreto-lei nº 41/2004*18-08-2004, Diário da República I Série.
28. Portugal, *Leis, decretos* in *Decreto-lei nº 32/2008.*17-07-2008, Diário da República I Série.
29. Wikipedia. *Bill Inmon*. abril 2012; Available from: http://en.wikipedia.org/wiki/Bill_Inmon.
30. FusionForge. *Functional View of Event Processing*. maio 2012; Available from: http://forge.fi-ware.eu/plugins/mediawiki/wiki/fiware/images/c/c1/Functional_View_of_Event_Processing.jpg.