

Mestrado em Engenharia Informática

Dissertação/Estágio

Relatório Final

Projeto DW-UC

**Desenvolvimento de uma Data Warehouse
para a Universidade de Coimbra**

Área B: Recursos Humanos

Autor:

Milton Jorge Martins Sêco

mseco@student.dei.uc.pt

Orientador:

Bruno Miguel Brás Cabral

bcabral@dei.uc.pt

Coimbra, 2 de Setembro de 2014



**FCTUC DEPARTAMENTO
DE ENGENHARIA INFORMÁTICA**
FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Resumo

Atualmente, os sistemas de suporte à decisão são peças fundamentais ao desenvolvimento e ao sucesso de inúmeras organizações por todo o mundo, tornando a procura e a necessidade destes sistemas cada vez maior.

A Universidade de Coimbra, de forma a cumprir com os seus planos estratégicos e de ação, pretende uma ferramenta que permita aos seus órgãos de gestão um fácil acesso a informação que permita observar o desempenho e evolução da universidade e assim decidir ações e rumos a tomar.

Dessa necessidade nasceu o projeto “DW-UC” que visa a construção de uma *data warehouse* onde é tratada e analisada informação de várias fontes de dados existentes dentro da UC. Por sua vez essa informação é utilizada para calcular diversos indicadores de desempenho que permitam monitorizar e avaliar o desempenho da UC em diversas vertentes.

Este relatório ilustra assim o trabalho realizado neste projeto, com especial ênfase na parte dos Recursos Humanos, área que me foi atribuída das várias divisões da *data warehouse* e análise OLAP.

Palavras-chave

Bussiness Intelligence, Data Warehouse, Key Performance Indicators, Indicadores de Desempenho, *Online Analytical Processing*, Recursos Humanos.

Agradecimentos

Em primeiro lugar, gostaria de agradecer ao meu orientador Prof. Bruno Cabral pela sua paciência e apoio prestado durante a execução deste estágio, especialmente por ter confiado e acreditado em mim apesar de todas as dificuldades que enfrentei. Um obrigado também ao Eng.º Pedro Pinto da equipa do NONIO pelas sugestões e apoio demonstrado nas reuniões semanais do projeto.

Da mesma forma, gostaria de agradecer aos meus colegas estagiários, principalmente à Beatriz Fragoso pela paciência e disponibilidade de me ajudar sempre que eu precisava, e pelo apoio e motivação que me deu em realizar este trabalho.

Um grande obrigado a todos os meus colegas da filarmónica e da tuna pelos bons momentos musicais e de grande convívio passados em Coimbra.

Agradecer também a todos os meus amigos e colegas de curso, principalmente àqueles com quem tive o privilégio de partilhar um carro de fitados e que me acompanharam durante toda a minha vida académica nesta cidade.

Um obrigado especial à minha família por todos os bons e maus momentos passados em conjunto e por me terem proporcionado uma infância e uma adolescência feliz.

Por fim, gostaria de agradecer profundamente à minha mãe a quem dedico todo o meu trabalho académico nesta universidade. Ela, que apesar do inesperado falecimento do meu pai há quase 10, e de todas as dificuldades que enfrentámos, se mostrou incansável em garantir uma boa educação e um futuro digno para mim e para o meu irmão. A ti devo o homem que sou hoje e tenho muito orgulho em ser teu filho. Obrigado mãe.

Índice

1.	Introdução	10
1.1.	Enquadramento.....	10
1.2.	Contexto Atual.....	11
1.3.	Objetivos	12
1.4.	Estrutura do Relatório	13
2.	Planeamento	14
2.1.	Plano de Ação	14
2.2.	Planeado e Realizado	16
3.	Requisitos	18
3.1.	Levantamento de Requisitos	18
3.2.	Especificação de Requisitos.....	18
3.2.1.	Requisitos funcionais gerais	19
3.2.2.	Requisitos funcionais dos Recursos Humanos	21
3.2.3.	Requisitos não funcionais.....	23
3.3.	Prototipagem Rápida.....	24
3.4.	Indicadores de Desempenho.....	26
3.4.1.	Caracterização dos Trabalhadores	26
3.4.2.	Demografia	27
3.4.3.	Outros.....	27
3.4.4.	Alterações aos Indicadores	28
4.	Arquitetura	29
4.1.	Visão Global.....	29
4.2.	Seleção de Tecnologias	31
4.2.1.	Base de Dados	31
4.2.2.	Processo ETL.....	31
4.2.3.	Análise OLAP.....	32
4.2.4.	Interface Web.....	32
4.3.	Modelo de Dados	33
4.3.1.	Modelo Área Temporária	33
4.3.1.	Modelo Multidimensional	35
5.	Implementação	39
5.1.	Processo ETL.....	39
5.1.1.	Componentes das transformações	39

5.1.2.	Carregamento da área temporária	40
5.1.3.	Carregamento das dimensões.....	41
5.1.4.	Carregamento dos factos.	43
5.2.	Cubos OLAP	44
5.3.	<i>Dashboards</i>	45
5.4.	Resultado Final	47
5.4.1.	Alterações.....	50
6.	Testes e Validação	51
6.1.	Validação dos Resultados.....	51
6.2.	Resultado dos Testes Funcionais.....	51
7.	Conclusão	55
8.	Referências.....	56

Lista de Figuras

Figura 1 - Enquadramento do projeto DW-UC.....	10
Figura 2 - Metas do planeamento, 2º semestre	16
Figura 3 - Diagrama de Gantt, 2º semestre.....	17
Figura 4 - Exemplo de protótipo dos Recursos Humanos	24
Figura 5 - Arquitetura global do sistema.....	29
Figura 6 - Tecnologias utilizadas no desenvolvimento da aplicação.....	32
Figura 7 - Modelo de dados da área temporária	34
Figura 8 - Modelo multidimensional da data warehouse	38
Figura 9 – Esquema de transformações da tabela temp_eti	40
Figura 10 - Esquema de transformações da tabela d_dem_contrato.....	41
Figura 11 - Esquema de transformações da tabela d_faixa_etaria.....	42
Figura 12 - Esquema de transformações da tabela d_trabalhador.....	42
Figura 13 - Esquema de transformações da tabela f_contrato.....	43
Figura 14 - Menu de layout do CDE.....	45
Figura 15 - Menu de componentes do CDE.....	46
Figura 16 - Menu de datasources do CDE	46
Figura 17 - Exemplo de uma querie MDX.....	47
Figura 18 - Ecrã exemplo, nível FCTUC.....	48
Figura 19 - Ecrã exemplo, nível Departamentos	49

Lista de Tabelas

Tabela 1 - Requisitos funcionais gerais	20
Tabela 2 - Requisitos funcionais dos Recursos Humanos	22
Tabela 3 - Requisitos não funcionais.....	23
Tabela 4 - Tabelas da área temporária.....	34
Tabela 5 - Tabelas de dimensão da data warehouse	37
Tabela 6 - Tabelas de facto da data warehouse.....	38
Tabela 7 - Componentes das transformações do Kettle	40
Tabela 8 - Testes dos requisitos funcionais gerais	52
Tabela 9 - Testes dos requisitos funcionais dos Recursos Humanos	54
Tabela 10 - Testes dos requisitos não funcionais.....	54

Acrónimos

BD	Base de Dados
BI	<i>Business Intelligence</i>
CDE	<i>Community Dashboard Editor</i>
DEI	Departamento de Engenharia Informática
DW	<i>Data Warehouse</i>
ETI	Equivalente a Tempo Integral
ETL	<i>Extraction, Transforming and Loading</i>
GSIC	Gestão de Sistemas e Infraestruturas de Informação e Comunicação
HTTP	<i>HyperText Transfer Protocol</i>
KPI	<i>Key Performance Indicator</i>
LDAP	<i>Lightweight Directory Access Protocol</i>
MDX	<i>MultiDimensional eXpression</i>
OLAP	<i>OnLine Analytical Processing</i>
SAMA	Sistema de Apoio à Modernização Administrativa
SAP - ERP	<i>Systems, Applications, Products in Data Processing – Enterprise Resource Planning</i>
SQL	<i>Structured Query Language</i>
TIC	Tecnologias da Informação e Comunicação
UC	Universidade de Coimbra
UO	Unidade Orgânica
XML	eXtensible Markup Language

1. Introdução

1.1. Enquadramento

A área de *Bussiness Intelligence* é certamente uma área que tem vindo a crescer ultimamente. A sua habilidade de extrair e apresentar informação que posteriormente auxilia a tomada de decisão por parte de grandes empresas é certamente a razão desse crescimento.

A Universidade de Coimbra não é exceção, e para poder cumprir com o seu Plano Estratégico e de Ação até 2015, decidiu apostar na realização de uma ferramenta que possibilite, de entre outras, monitorizar indicadores definidos no plano estratégico, fundamentar e sustentar os processo de tomada de decisão, diminuir o tempo e os custos com o tratamento, preparação e apresentação da informação.

Este estágio está integrado num projeto denominado “DW-UC” que visa desenvolver uma *data warehouse* para a Universidade de Coimbra. O projeto tem por objetivo recolher, tratar e analisar a informação de múltiplas fontes de dados existentes dentro da UC. A informação será utilizada para calcular diversos indicadores (KPI) que permitam monitorizar e avaliar o desempenho da UC nas diversas vertentes previstas no seu Plano Estratégico e de Ação. Por sua vez, este projeto está integrado num outro denominado SAMA (Sistema de Apoio à Modernização) que tem como objetivo melhor as infraestruturas e serviços TIC em várias áreas da universidade. Uma delas é a criação de instrumentos de gestão e monitorização de indicadores de desempenho no qual o projeto “DW-UC” se enquadra perfeitamente.

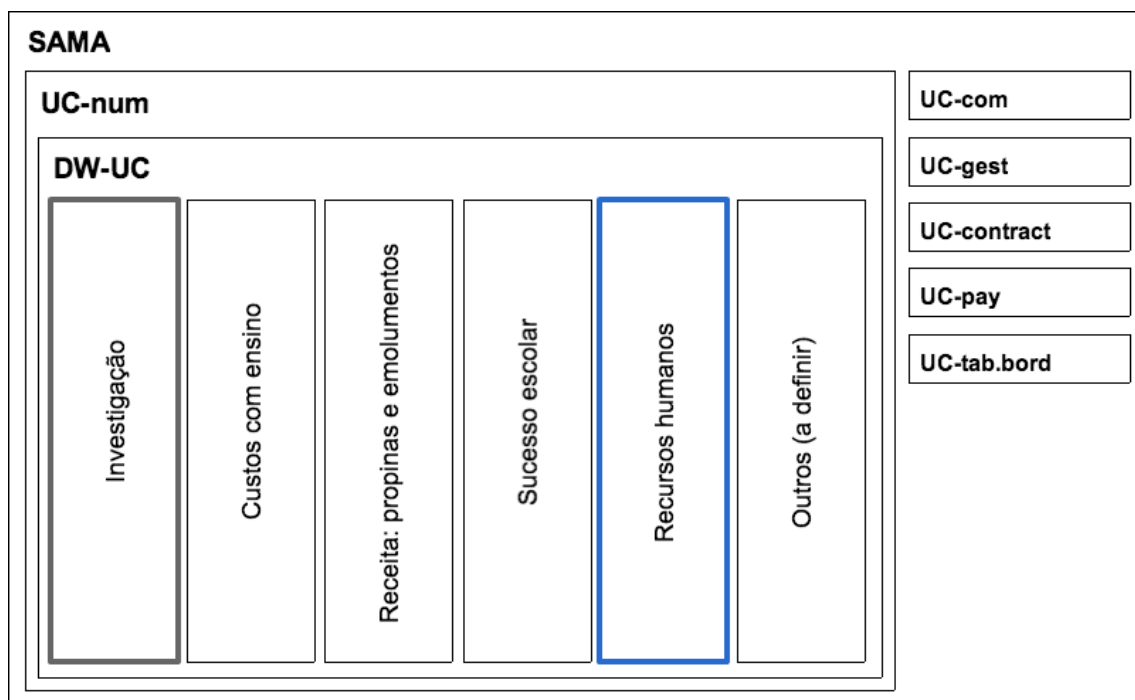


Figura 1 - Enquadramento do projeto DW-UC

O desenvolvimento da *data warehouse* e a análise OLAP estão divididos, neste momento, em 5 áreas de ação: Investigação, Custos com o Ensino, Receita, Sucesso Escolar e Recursos Humanos. A primeira área é da responsabilidade de uma equipa de três elementos que iniciou essa parte do projeto no ano letivo anterior. Cada uma das restantes quatro áreas está ao encargo de um aluno como estágio de mestrado e no meu caso coube-me a área dos **Recursos Humanos**. É nessa área que me irei focar e na qual este relatório se reflete. Para além de mim e dos meus três colegas, o resto da equipa é composta pelo nosso orientador Prof. Doutor Bruno Cabral e pelo Eng.º Pedro Pinto da equipa do NONIO.

1.2. Contexto Atual

Neste momento a UC possui essencialmente dois sistemas principais de informação onde a maior parte dos dados da universidade são armazenados: o sistema SAP ERP^[13] e o NONIO.

O NONIO é um sistema de gestão académica desenvolvido no seio da Universidade de Coimbra para uso próprio. Aqui é armazenada toda a informação académica dos alunos assim como os cursos, disciplinas e unidades orgânicas. Também é aqui que são armazenados todos os dados relativos ao serviço docente e ao pagamento de propinas e emolumentos. Apesar de toda a informação que sustenta, este sistema permite apenas uma análise estatística muito simples dos dados académicos dos alunos e de dados para a gestão dos cursos da universidade.

O SAP é um sistema integrado de gestão empresarial, produto da empresa homónima. Este contém todos os dados de gestão financeira da UC assim como toda a informação relativa aos recursos humanos, entre outros. O sistema é particularmente essencial a este módulo visto que contém toda a informação necessária relativa aos recursos humanos da universidade. No entanto, e apesar do SAP apresentar um módulo que permite gerar análises sobre os dados, este é considerado obsoleto e não oferece a flexibilidade necessária à análise de indicadores de desempenho.

O cálculo de indicadores dentro da universidade, não só para o estudo do desempenho próprio mas também pela solicitação por parte outras entidades (por exemplo, a auscultação anual do Observatório Nacional de Recursos Humanos^[14]), é um processo muito vezes manual e moroso, o que não permite uma resposta atempada e fidedigna. Nenhum dos sistemas apresenta uma forma fácil, flexível, e que permita a manipulação dos dados, como aplicar filtros ou restrições temporais, e por isso a necessidade de construir uma *data warehouse* que permita uma análise completa destes indicadores tornou-se cada vez mais uma prioridade.

No entanto, estes dois sistemas servem como fontes de dados aos vários módulos deste projeto, principalmente o SAP no que diz respeito a este módulo.

1.3. Objetivos

O objetivo principal deste projeto é então o desenvolvimento de uma *data warehouse* para a Universidade de Coimbra que é essencial um repositório central de dados onde é realizada uma análise OLAP, ou seja, uma análise sobre grande volume de dados que permite aplicar filtros, agrupamentos e outras funções sobre os dados.

Sendo assim é necessário:

1. Criar um processo automatizado para recolher, tratar e analisar a informação de múltiplas fontes de dados existentes dentro da UC, como o SAP e o NONIO.
2. Desenvolver um sistema para o cálculo de diversos indicadores de desempenho (KPI) que permitam monitorizar e avaliar o desempenho da UC como base nesses dados.
3. Desenvolver uma plataforma *web* onde os dados possam ser manipulados e visualizados através de gráficos e tabelas.

Em relação ao módulo dos Recursos Humanos, o objetivo deste é relativamente diferente aos dos restantes módulos. Enquanto nas áreas do Custo com Ensino, Receita e Sucesso Escolar pretende-se sobretudo uma análise sob a forma de valores, taxas e percentagens, nesta área o objetivo principal é analisar o número de recursos humanos e os seus custos.

Neste módulo pretende-se obter respostas a algumas questões, entre elas:

- Quantos docentes de cada categoria existem na UC?
- Qual a faixa etária dominante de cada unidade orgânica?
- Quantos funcionários existem segundo o nível de escolaridade?
- Quantos docentes existem por número de anos serviço? E qual a média de anos de serviço de cada faculdade?
- Quantos funcionários existem por tipo de contrato?
- Quais os custos com salários?

Estas questões surgem especialmente da necessidade de elaborar um documento anual sobre o Balanço Social da Universidade de Coimbra de uma forma mais fácil e percetível, e que até agora era realizado de uma forma manual e morosa. O levantamento de requisitos sobre quais os principais indicadores a calcular foi igualmente realizado junto dos órgãos de gestão dos recursos humanos da UC.

1.4. Estrutura do Relatório

O relatório está dividido em várias secções que refletem o trabalho efetuado ao longo do estágio.

No capítulo 2 é exposto o planeamento global do projeto, passando pelo plano de ação e pela comparação do que foi planeado e do que foi realizado.

O capítulo 3 apresenta o processo de levantamento de requisitos e respetivos requisitos funcionais e não funcionais. Também ilustra o processo de prototipagem rápida e os indicadores de desempenho definidos.

A arquitetura global do sistema assim como as tecnologias utilizadas estão refletidas no capítulo 4, onde se pode consultar também o modelo de dados da DW.

No capítulo 5 encontram-se os detalhes de implantação e desenvolvimento da aplicação em todas as suas fases.

No capítulo 6 são apresentados os resultados de alguns testes e validações dos dados.

Uma conclusão sobre todo o trabalho realizado é exposta no capítulo 7.

2. Planeamento

2.1. Plano de Ação

Geralmente, é bastante importante num projeto desta dimensão que sejam bem delineados quais os passos a tomar durante o desenvolvimento de uma aplicação de BI. Como em qualquer projeto deste género, este divide-se essencialmente em 4 etapas:

1. Conhecer os indicadores de desempenho (KPI) desejados para a análise OLAP.
2. Identificar os dados e as respetivas fontes necessárias.
3. Recolher, transformar e carregar os dados, isto é, todo o processo ETL.
4. Apresentar a informação aos utilizadores, neste caso, sob a forma de uma aplicação *web*.

Em relação à identificação dos indicadores de desempenho, esses foram conhecidos durante a fase de levantamento de requisitos, detalhado no capítulo seguinte. Preparar e organizar os indicadores foi igualmente importante para o desenvolvimento da aplicação pois permitiu agrupar indicadores semelhantes em categorias e estruturar a melhor forma de navegar entre eles. Essa fase do planeamento é explicada mais à frente no subcapítulo dos indicadores.

Das fontes de dados conhecidas, SAP e NONIO, foi bastante claro que a informação sobre os recursos humanos estaria alojada no sistema SAP. No entanto, o NONIO também serviu como fonte de dados relativamente às unidades orgânicas da UC. Foi-nos então disponibilizada uma cópia das tabelas do SAP ERP como forma de acesso à informação solicitada. Dada a complexidade do sistema, e do facto de a cópia conter cerca de 70 mil tabelas, foi elaborado um documento com a lista dos dados necessários ao módulo dos Recursos Humanos que foi posteriormente enviado para o GSIIC da universidade com o intuito de conhecer as tabelas que continham essa informação. De notar a colaboração por parte do técnico Nelson Costa e do Eng.º Jorge China do GSIIC na disponibilização dos dados do SAP. Para mais informação sobre os dados solicitados e quais as respetivas tabelas SAP, por favor consulte o documento em anexo intitulado “[1] doc_rh_dados-sap_2014-04-03_v2.pdf”.

Depois de conhecer os dados, o passo seguinte é recolher essa informação, trata-la e armazena-la na *data warehouse*. Claramente, deve existir uma seleção das tecnologias para a realização dessas tarefas, o que é focado mais à frente no próximo capítulo. O planeamento desta fase incidiu principalmente na construção de um modelo de dados para armazenar a informação e na fase de adaptação às tecnologias utilizadas, isto é, familiarizar e conhecer as transformações e operações de dados oferecidas pela ferramenta.

Por último, o planeamento da interface *web* é crucial para apresentar a informação ao utilizador de forma a permitir uma análise rápida e intuitiva. É importante decidir onde colocar os elementos de manipulação dos dados como as caixas de seleção, onde colocar os gráficos, como traduzi-los em tabelas e como navegar entre os vários indicadores. Os protótipos realizados na fase de levantamento de requisitos foram bastante uteis a esta parte do projeto e serviram como base ao *layout* dos *dashboards*. Como os vários módulos serão integrados numa só aplicação final, é importante definir normas de *design* de forma a manter a consistência ao nível da apresentação entre todos os módulos. Foi então elaborado um documento para essas especificações que pode ser consultado no anexo intitulado “[6] doc_todos_especificação-design_2014-01-10_v1.5.pdf”. Ambas as etapas finais estão explicadas em maior detalhe no capítulo sobre a implementação.

Ao longo de todo o segundo semestre, o desenvolvimento da aplicação foi acompanhado por reuniões semanais entre a equipa integrante. Nessas reuniões cada estagiário, responsável por cada módulo, exponha o trabalho realizado nessa semana e as dificuldades ou dúvidas encontradas. O orientador Prof. Doutor Bruno Cabral e o Eng.º Pedro Pinto do NONIO auxiliavam com sugestões aos problemas enfrentados e quais os melhores rumos a tomar.

Sendo um projeto desenvolvido paralelamente com outros alunos, eram marcadas reuniões esporádicas entres os estagiários com o objetivo de tomar decisões em prol de todo o grupo. Para uma melhor gestão da informação comum a todos, foi criado um repositório *online* partilhado entre todos os estagiários do projeto.

2.2. Planeado e Realizado

Durante o primeiro semestre, o facto de ter estado fora do país resultou num planeamento de trabalho muito irregular e que não foi cumprido na sua totalidade. No entanto, e de forma a recuperar o trabalho em falta, foi elaborado um planeamento geral para o segundo semestre com as principais metas a alcançar.

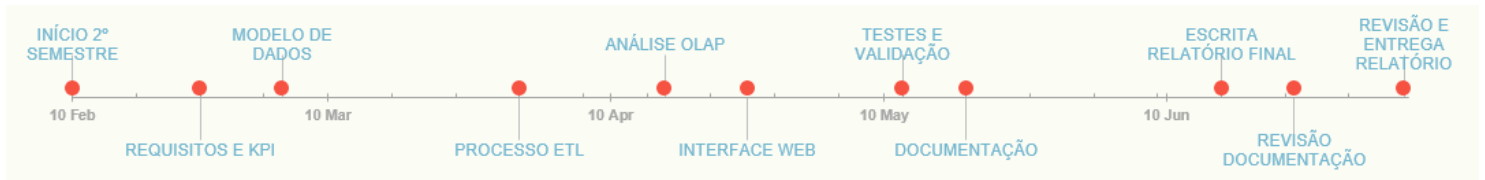


Figura 2 - Metas do planeamento, 2º semestre

Como é comum na maior parte dos projetos de engenharia de *software*, nem sempre o que é planeado é cumprido no tempo estipulado pois acontecem imprevistos ou certos eventos que alteram o planeamento. Este caso não foi exceção resultando no término do estágio mais tarde do que era suposto devido ao atraso vindo do primeiro semestre.

No diagrama de Gantt seguinte podem-se ver as tarefas realizadas ao longo do segundo semestre assim como os intervalos de tempo que elas ocuparam.

Começou-se pelo levantamento e definição dos indicadores de desempenho seguido da construção e validação dos protótipos.

O passo seguinte passou pela definição do modelo de dados assim como a recolha dos dados dos sistemas fonte.

O processo ETL foi trabalhado a seguir e foram também definidos os cubos OLAP.

A criação do *dashboard* da Caracterização de trabalhadores foi o primeiro e foi o que demorou mais tempo a ser implementado. Os restantes, como seguiam um modelo muito parecido com o *dashboard* anterior, apenas com outros *datasources* e alguns componentes diferentes, a sua implementação foi muito mais rápida. De notar que durante o desenvolvimento dos *dashboards* houveram pequenas modificações ao processo ETL e aos cubos.

Posteriormente o módulo dos Recursos Humanos foi validado e alguns melhoramentos foram feitos na aplicação.

Finalmente procedeu-se à escrita e entrega do presente relatório.

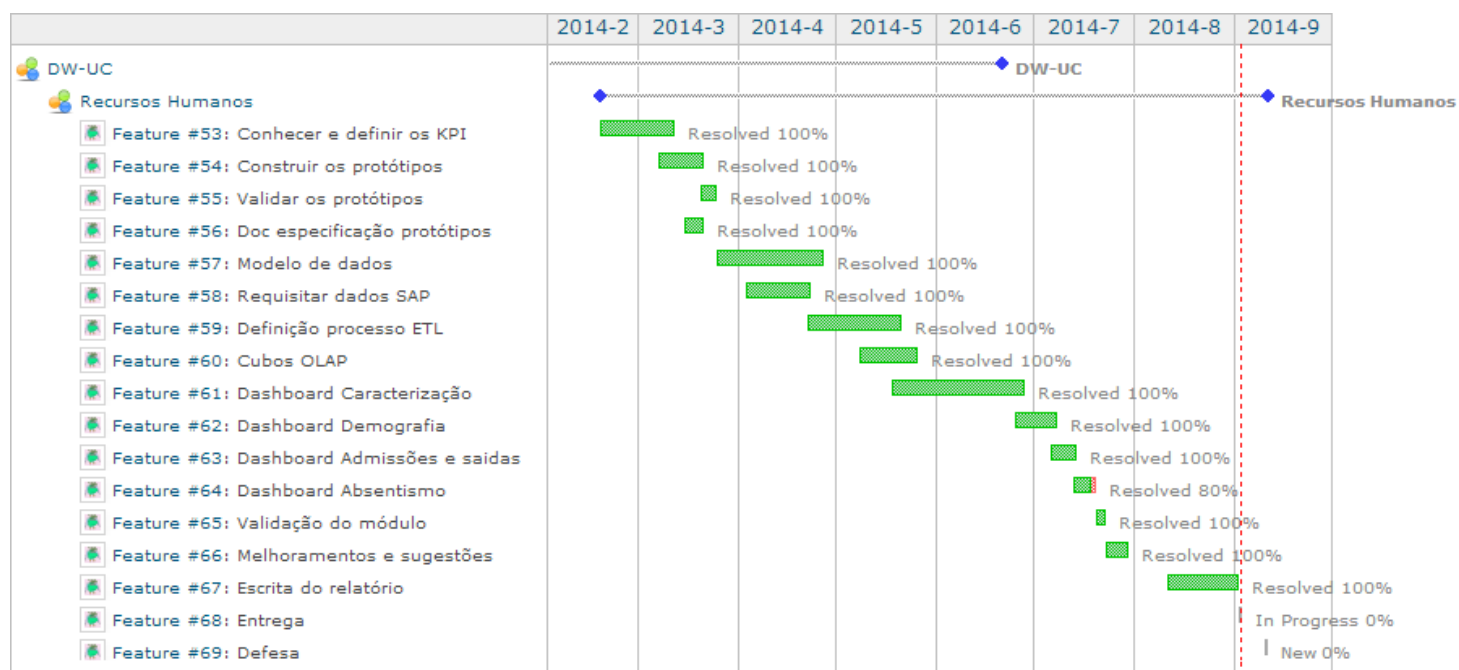


Figura 3 - Diagrama de Gantt, 2º semestre

3. Requisitos

3.1. Levantamento de Requisitos

Como em qualquer projeto na área de engenharia de *software*, o levantamento de requisitos é essencial para que se conheçam as principais funcionalidade que o sistema deve ter e assim efetuar o planeamento e a implementação destas.

Os requisitos gerais da aplicação foram levantados juntos dos principais órgãos de gestão da UC entre os quais o reitor Professor Doutor João Gabriel Silva e a vice-reitora Doutora Margarida Mano. Numa primeira fase, foram efetuadas reuniões presenciais junto destes para perceber o intuito geral de toda a aplicação.

Quanto aos requisitos específicos da área dos Recursos Humanos, estes foram conhecidos junto do chefe da Divisão do Planeamento, Gestão e Desenvolvimento, o Dr. Filipe Rocha, no início do segundo semestre. Foi elaborado por parte deste um primeiro documento com uma lista de indicadores desejados no módulo que serviram posteriormente para realizar um conjunto de protótipos rápidos. Esses protótipos foram apresentados e discutidos poucos dias mais tarde em reunião presencial com o mesmo e dessa reunião resultou um conjunto de indicadores para o módulo dos Recursos Humanos. Ao longo do desenvolvimento, houveram algumas alterações aos indicadores que discutirei no capítulo do planeamento.

3.2. Especificação de Requisitos

Relativamente à especificação dos requisitos, estes seguem o modelo FURPS+ (funcionalidade, usabilidade, disponibilidade, performance e suporte)^[3], um modelo para classificar atributos de qualidade de *software* desenvolvido na Hewlett-Packard e publicado por Grady e Caswell. Os requisitos estão assim divididos em requisitos funcionais, relacionados com as funcionalidades da aplicação, e requisitos não funcionais, que denominam a qualidade da aplicação. Dentro destes existem três níveis de prioridade quanto à sua execução: elevada, média e baixa. Têm prioridade elevada os requisitos principais que caso não sejam cumpridos comprometem a concretização da aplicação. Têm prioridade média os requisitos que quando cumpridos acrescentam valor à aplicação final e que, quando não cumpridos, não comprometem o correto funcionamento da aplicação. Têm prioridade baixa os requisitos que caso não cumpridos não comprometem o funcionamento da aplicação e que apenas contribuem pouco valor à aplicação final.

3.2.1. Requisitos funcionais gerais

Como o próprio nome indica, estes requisitos referem-se às funcionalidades principais do sistema. Os requisitos seguintes são gerais e comuns a toda a aplicação.

Código	Designação	Prioridade	Descrição
RF_GE_01	Autenticação	Elevada	A aplicação deve permitir ao utilizador a autenticação (<i>login</i>) através das credenciais utilizadas no acesso a quaisquer serviços disponibilizados à comunidade da UC (email da UC e password).
RF_GE_02	Fechar sessão	Elevada	O utilizador pode, após autenticação, efetuar o término da sua sessão (<i>logout</i>).
RF_GE_03	Término de sessão	Elevada	Para garantir a segurança da aplicação, um utilizador, depois de autenticar-se, terá associada uma sessão que deve ter um <i>timeout</i> para efetuar <i>logout</i> automaticamente.
RF_GE_04	Navegação entre módulos	Média	O utilizador deve conseguir aceder a todos os módulos a qualquer momento.
RF_GE_05	Navegação interna	Elevada	A aplicação deve permitir ao utilizador efetuar <i>drill down</i> nos dados que pretende visualizar, e a descida deve ser efetuada diretamente nos dados que vão sendo apresentados. Para que seja possível ao utilizador efetuar <i>roll up</i> a aplicação deve disponibilizar uma forma de navegação estrutural (<i>breadcrumbs</i>) que vá acrescentando o nível onde o utilizador se encontra.
RF_GE_06	Parâmetros gerais	Elevada	O utilizador deve ter disponíveis os diversos parâmetros que são permitidos aplicar sobre a análise.
RF_GE_07	Parâmetros de tempo	Elevada	Deve ser permitido ao utilizador modificar e aplicar os parâmetros temporais. Deve ser possível escolher o tipo de vista (ano letivo ou ano civil) e o período temporal (anual, semestral, trimestral ou

			mensal). Consoante o período selecionado o utilizador deve escolher o intervalo de tempo pretendido.
RF_GE_08	Esconder parâmetros	Baixa	Deve ser permitido ao utilizador esconder a barra lateral onde se encontram os parâmetros gerais e de tempo.
RF_GE_09	Secção de ajuda	Elevada	A aplicação deve disponibilizar uma secção de ajuda ao utilizador.
RF_GE_10	Informação Auxiliar	Média	Cada vista de dados disponibilizada ao utilizador deve ser acompanhada de um mecanismo que permita consultar informação referente a esses dados.
RF_GE_11	Visualização gráfico ⇔ tabela	Elevada	A aplicação deve permitir ao utilizador visualizar a informação apresentada num gráfico em formato de tabela e vice-versa.
RF_GE_12	Exportar informação da tabela	Baixa	Permitir exportar para formato Excel ou CSV a informação presente nas tabelas em análise.

Tabela 1 - Requisitos funcionais gerais

3.2.2. Requisitos funcionais dos Recursos Humanos

Em relação aos requisitos funcionais específicos do módulo dos Recursos Humanos, estes foram levantados junto do Dr. Filipe Rocha e encontram-se listados na tabela seguinte.

Código	Designação	Prioridade	Descrição
RF_RH_01	Unidades de medida	Elevada	Permitir visualizar os dados segundo diversas unidades de medida como nº de pessoas, ETIs, custo total e custo médio mensal.
RF_RH_02	Docentes por categoria profissional e regime	Elevada	Nos indicadores de caracterização dos trabalhadores, permitir agrupar os docentes por categoria profissional e por regime.
RF_RH_03	Investigadores por categoria profissional	Elevada	Nos indicadores de caracterização dos trabalhadores, permitir agrupar os investigadores por categoria profissional.
RF_RH_04	Não docentes por categoria profissional e tipo de contrato	Elevada	Nos indicadores de caracterização dos trabalhadores, permitir agrupar os não docentes por categoria profissional e por tipo de contrato.
RF_RH_05	Dirigentes por tipo de cargo	Elevada	Nos indicadores de caracterização dos trabalhadores, permitir agrupar os dirigentes por tipo de cargo.
RF_RH_06	Órgãos Sociais por tipo de cargo	Elevada	Nos indicadores de caracterização dos trabalhadores, permitir agrupar os órgãos sociais por tipo de cargo.
RF_RH_07	Agrupar por dados demográficos	Elevada	Nos indicadores de demografia, permitir agrupar os dados por habilitações literárias, faixa etária, antiguidade, nacionalidade e género.
RF_RH_08	Absentismo	Elevada	Apresentar os dados de absentismo dos docentes sobre a forma de taxa, nº de dias e duração média (em horas).
RF_RH_09	Admissões	Elevada	Calcular o número de admissões por categoria profissional.

RF_RH_10	Saídas	Elevada	Calcular o número de saídas por motivo.
RF_RH_11	Filtros gerais	Elevada	Permitir a aplicação dos filtros de habilitações literárias, faixa etária, antiguidade, nacionalidade e género a todos os indicadores.
RF_RH_12	Filtros específicos	Elevada	Permitir a aplicação dos seguintes filtros: vínculo para docentes e investigadores, categoria profissional na caracterização dos trabalhadores, tipo de contrato para não docentes, regime para docentes e investigadores, grupo de trabalhadores nos indicadores de demografia e outros.

Tabela 2 - Requisitos funcionais dos Recursos Humanos

3.2.3. Requisitos não funcionais

Nesta secção encontram-se detalhados os requisitos não funcionais deste sistema. São também conhecidos como atributos de qualidade, e tal como o nome sugere não estão ligados às funcionalidades e operações que a aplicação deve desempenhar mas sim à forma como estas se comportam. Estes estão divididos nas seguintes categorias: U – usabilidade; R – confiabilidade; P – performance; S – manutenção e suporte; O – outros.

Código	Designação	Prioridade	Descrição
RNF_S_01	Atualização de dados	Elevada	Processo ETL, carregamento e atualização da DW e cubo OLAP devem ser automáticos.
RNF_S_02	Compatibilidade (<i>browser</i>)	Elevada	Aplicação <i>web</i> é compatível com os <i>browsers</i> mais modernos, a partir das versões mencionadas: Internet Explorer 9; Firefox 20 ou superior e Safari 6 ou superior.
RNF_S_03	Compatibilidade (SO)	Média	Existem algumas restrições também quanto ao sistema operativo, são suportados oficialmente, os SO a partir de: Windows 7 e distribuições de Linux.
RNF_S_04	Licenças	Elevada	A aplicação é desenvolvida e disponibilizada através de <i>software</i> gratuito.

Tabela 3 - Requisitos não funcionais

3.3. Prototipagem Rápida

O processo de levantamento de requisitos nem sempre é claro o que poder gerar discrepâncias entre o que se pretende, e o que é entendido como requisito. Para evitar eventuais alterações durante o desenvolvimento da aplicação, foi decidido pela equipa recorrer à elaboração de protótipos, isto é, exemplos de ecrãs que se esperam visualizar no produto final. Foram então elaborados um conjunto de protótipos da aplicação para que estes passassem posteriormente por uma fase de validação, eliminando assim dúvidas de ambas as partes.

Os protótipos foram realizados com o auxílio de uma versão gratuita e limitada do *software Justinmind Prototyper*^[16]. Estes pretendem emular o que se espera ver e manipular na interface *web* da aplicação. Para além da funcionalidade, os protótipos permitem também perceber o nível de interatividade que se espera obter com os indicadores e os vários níveis de granularidade em que estes podem ser analisados.

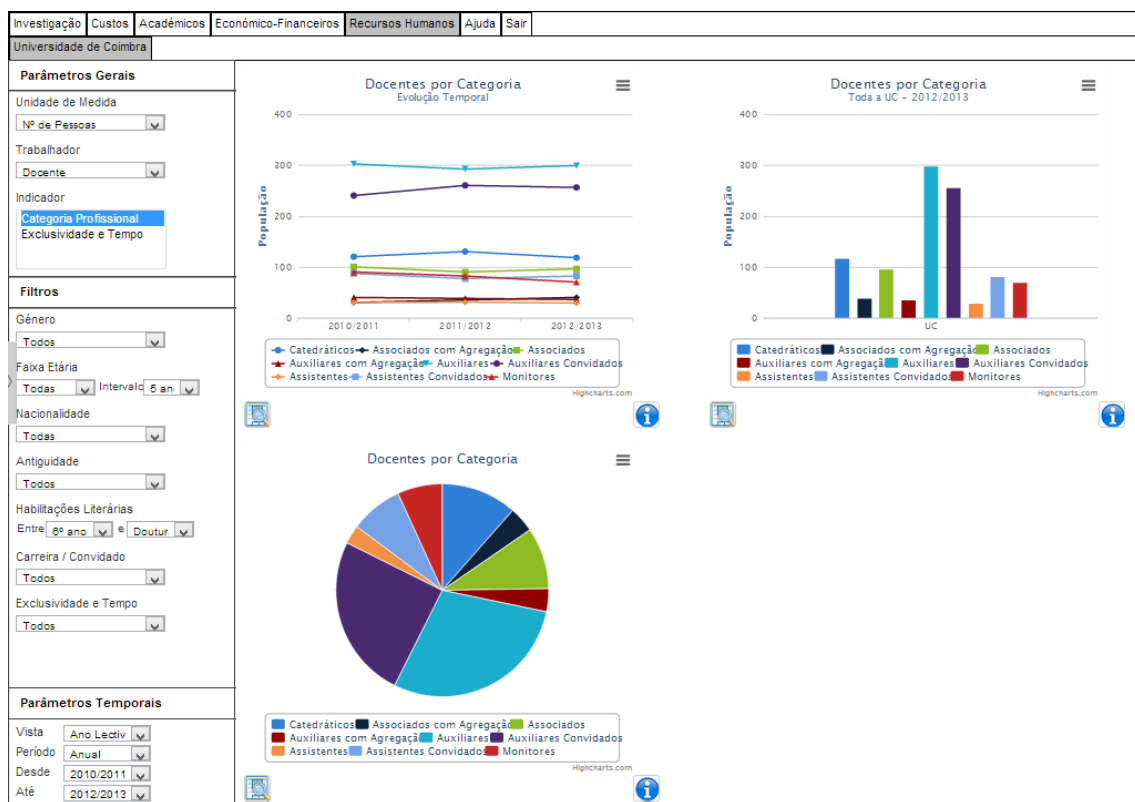


Figura 4 - Exemplo de protótipo dos Recursos Humanos

A figura anterior é apenas um exemplo dos protótipos realizados no início do segundo semestre com a ajuda da lista de indicadores compilados pelo Dr. Filipe Rocha. Estes foram posteriormente apresentados e validados pelo próprio como já referido anteriormente.

Para mais informação sobre os protótipos da área dos recursos humanos realizados nesta fase, por favor consulte o documento em anexo intitulado “[2] doc_rh_especificação-de-protótipos-recursos-humanos_2014-03-19_v1.3.pdf”. Para mais informação sobre uma primeira versão dos protótipos de todos os módulos realizados no final do primeiro semestre, por favor consulte o documento em anexo intitulado “[5] doc_todos_especificação-de-protótipos_2013-12-22_v1.4.pdf”.

3.4. Indicadores de Desempenho

Após o levantamento de requisitos e conhecidos os indicadores de desempenho a calcular na análise OLAP, é importante planejar como é que estes se organizam entre si. Isto é, é necessário identificar indicadores que se encaixem na mesma categoria, definir filtros comuns aos indicadores da mesma categoria, definir filtros específicos para cada indicador isolado e identificar as unidades de medida destes. Daqui surgiram 3 grupos principais de indicadores: **Caracterização dos trabalhadores, Demografia e Outros**. Para todos os indicadores existem 5 filtros comuns para os dados: **habilitações literárias, faixa etária, antiguidade, nacionalidade e género**.

3.4.1. Caracterização dos Trabalhadores

Para o primeiro grupo de indicadores sobre a Caracterização dos trabalhadores, existem essencialmente 4 indicadores: **Nº de pessoas, ETI, Custo total e Custo médio mensal**. Estes indicadores servem no fundo como unidades de medida dos dados que podem ser agrupados por outras propriedades dos recursos humanos. A divisão dessas propriedades para cada grupo de trabalhadores é a seguinte:

1. Trabalhadores
 - a. Docentes
 - i. Categoria Profissional (Catedrático, Auxiliar, Assistente, etc.)
 - ii. Regime (Exclusividade, Tempo Integral, Tempo Parcial)
 - b. Investigadores
 - i. Categoria Profissional (Auxiliar, Coordenador, etc.)
 - c. Não Docentes
 - i. Categoria Profissional (Técnico Superior, Assistente Técnico, etc.)
 - ii. Tipo de Contrato (Termo certo, contrato de bolsa, etc.)
 - d. Órgãos do Governo e de Gestão
 - i. Tipo de Cargo (Diretor, Vice-Reitor, etc.)
2. Contratos
 - a. Tarefa e Avença
 - b. Bolseiros e Estagiários

Isto significa que é possível calcular, por exemplo, o nº de docentes por categoria profissional, o valor de ETI dos docentes por regime, o custo total dos investigadores por categoria profissional, o custo médio mensal dos não docentes por tipo de contrato, etc. Além dos filtros gerais, este grupo de indicadores contém alguns filtros específicos: **regime** para os

docentes e investigadores por categoria profissional, **tipo de contrato** para os não docentes por categoria profissional, **vínculo** (carreira ou convidado) para os docentes e investigadores, e **categoria profissional** para os docentes por regime e para os não docentes por tipo de contrato.

3.4.2. Demografia

Em relação aos indicadores de Demografia, estes também incluem as 4 unidades de mediadas do grupo anterior, mas desta vez, o agrupamento dos dados é feito através das propriedades demográficas dos trabalhadores que, neste caso, coincidem com os filtros gerais:

1. Habilitações Literárias (12º ano, Licenciatura, Mestrado, etc.)
2. Faixa Etária (<19, 25-30, >65, etc.)
3. Antiguidade (anos de serviço, 0-4, >40, etc.)
4. Nacionalidade (país de origem)
5. Género (feminino e masculino)

Assim como no grupo de indicadores anterior, é possível calcular por exemplo, o nº de trabalhadores por habilitações literárias, o valor de ETI dos trabalhadores por faixa etária, o custo total por nacionalidade, o custo médio mensal por género, etc. Neste grupo existe um filtro específico que é o **grupo de trabalhadores** (docentes, não docentes, investigadores, etc.).

3.4.3. Outros

O último grupo contém os indicadores que não se encaixavam em nenhum dos grupos anteriores e que não têm propriamente ligações entre si. São eles:

1. Absentismo
 - a. Taxa (nº de horas não trabalhadas sobre nº de horas trabalháveis)
 - b. Nº de Dias
 - c. Duração Média (horas)
2. Formação (por Área)
 - a. Volume de Formação
 - b. Nº de Formandos
 - c. Nº de Ações de Formação
3. Admissões e Saídas
 - a. Admissões (por categoria profissional)
 - b. Saídas (por motivo)
4. Acidentes
 - a. Nº de Acidentes

- b. Índice de Frequência
- c. Índice de Incidência
- d. Índice de Gravidade

A todos os indicadores anteriores podem ser aplicados os mesmos filtros gerais, assim como o filtro de grupo de trabalhadores.

Para mais informação sobre os indicadores, como a granularidade, parâmetros temporais, tabelas fonte, etc., por favor consulte o documento em anexo intitulado “[3] doc_rh_ficha-indicadores.pdf”.

3.4.4. Alterações aos Indicadores

Durante a fase de extração dos dados e de desenvolvimento, os indicadores de desempenho iniciais acabaram por sofrer algumas alterações.

No grupo da Caracterização dos trabalhadores, os trabalhadores “Órgãos do Governo e de Gestão” foram divididos em dois grupos de trabalhadores, Dirigentes e Órgãos Sociais, para irem de encontro à categorização de trabalhadores presentes no sistema SAP. Ao não agrupar estes dois grupos distintos consegue-se uma maior discretização dos dados.

No mesmo grupo, os indicadores de caracterização de contratos, “Tarefa e Avença” e “Bolseiros e Estagiários” foram eliminados pois estão incluídos no grupo de trabalhadores dos “Não Docentes” e podem ser observados escolhendo essas categorias profissionais para a análise, ou aplicando filtros do tipo de contrato (Prest. Serviço/Avença e Contrato de Bolsa respetivamente).

No grupo Outros, os indicadores sobre as ações de formação acabaram por ser descartados pois a informação no SAP era insuficiente, apenas era guardado o título da ação de formação e não a área correspondida (gestão, informática, contabilidade, etc.). Os indicadores sobre os acidentes de trabalho acabaram igualmente por serem descartados pois essa informação só se encontra na Medicina do Trabalho da UC e em formato de papel.

4. Arquitetura

4.1. Visão Global

Uma aplicação desta dimensão é algo consideravelmente complexo e é composta por várias partes, cada uma contemplando diferentes tecnologias. Neste caso, a arquitetura global do sistema está dividida em 6 partes sendo que as 3 primeiras fazem parte do processo de ETL, ou seja, extração, transformação e carregamento de dados. Sendo assim, a divisão é a seguinte:

1. Extração das fontes de dados;
2. Transformação dos dados;
3. Carregamento dos dados na *data warehouse*;
4. Formação dos cubos OLAP;
5. Análise OLAP;
6. Interface *web* para o utilizador.

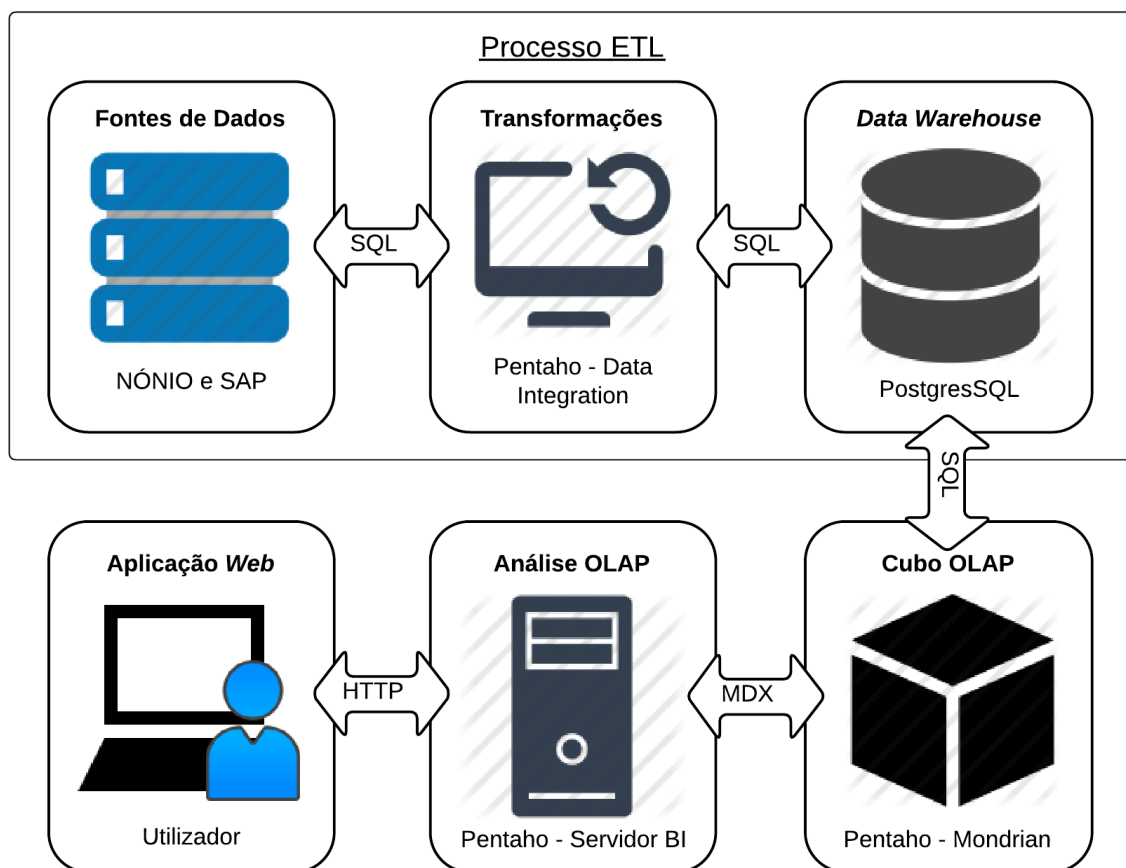


Figura 5 - Arquitetura global do sistema

Como ilustrado na figura anterior, as primeiras 3 partes do sistema estão agregadas num processo de ETL^[4] que atua sobre os dados. A primeira fase desse processo é extração dos dados das fontes de dados da UC que neste caso são o NONIO e o sistema SAP da universidade. De seguida os dados são submetidos a uma série de transformações como seleção de dados pertinentes, junção e uniformização de dados, cruzamento de informação de várias tabelas, agregação de dados, etc. Na última fase os dados remodelados são carregados na base de dados da *data warehouse* segundo um modelo multidimensional, primeiro as dimensões e de seguida os factos. O modelo de dados utilizado pode ser consultado no subcapítulo mais à frente, e os detalhes do processo ETL no capítulo da implementação.

Após os dados se encontrarem na DW, estes são mapeados para o que é conhecido como cubos OLAP. Cada um destes cubos representa um modelo “em estrela”, um tipo de modelo de dados típico das *data warehouses*. Estes são constituídos por uma tabela de factos e pelas tabelas de dimensões associadas. Geralmente existe um cubo para cada tabela de factos, que é o que acontece neste módulo. A definição destes cubos ajuda a uma melhor performance aquando das consultas dos dados pois as junções das tabelas através das chaves forasteiras são feitas automaticamente. Para além disso a informação é logo armazenada em cache, pelo que uma nova consulta semelhante será executada mais rapidamente.

A fase seguinte é onde é realizada o que se entende por análise OLAP, ou seja, é onde os dados provenientes da DW são apresentados. Aqui são efetuadas as *queries* de consulta e onde são calculados os indicadores de desempenho. Estes resultados são depois apresentados ao utilizador numa interface *web* sob a forma de gráficos e tabelas. O que faz da ilustração dos dados uma análise OLAP é facto do utilizador poder aplicar filtros, restrições temporais e descer na granularidade dos dados *on the fly* obtendo os novos resultados logo de seguida. Mais uma vez, os cubos são essenciais à manipulação destes dados, facilitando as operações de *slice and dice*.

4.2. Seleção de Tecnologias

Para um projeto desta dimensão são várias as tecnologias envolvidas no desenvolvimento de todas as componentes como o processo ETL, a base de dados, a análise OLAP e a interface *web*. Na escolha das ferramentas a utilizar, os principais aspetos tidos em conta foram: terem as funcionalidades necessárias e apropriadas ao desenvolvimento da aplicação; serem gratuitas ou *open-source* por uma questão financeira. Outro fator preponderante na decisão das ferramentas foi o facto de este projeto ter sido integrado na sequência de um módulo já desenvolvido (Projetos de Investigação) onde se teve em conta as tecnologias já utilizadas.

4.2.1. Base de Dados

Inicialmente a base de dados escolhida foi a Oracle uma vez que a UC já detém uma licença para a utilização desta BD. No entanto a licença do DEI para o Oracle não se encaixava nos termos do projeto e a base de dados escolhida para o armazenamento dos dados foi então o **PostgreSQL** uma vez que é bastante versátil e *open-source*. Esta BD permite a execução rápida de queries complexas típicas de uma análise OLAP com 5 tipos diferentes de JOINS e *queries* recursivas. Outra BD *open-source* candidata foi o MySQL já que é um das mais utilizadas em todo o mundo. No entanto o PostgreSQL têm vindo a evoluir bastante sendo utilizada por grandes empresas como a IMDB, Cisco e Skype, e é considerada das mais adequadas para *data warehouses*^{[5][6]}. Por outro lado, bases de dados não relacionais estavam fora do espectro de seleção pois, no geral, ferramentas de *Business Intelligence* não oferecem suporte a esse tipo de BDs.

4.2.2. Processo ETL

Este processo, como já referido anteriormente, está dividido na extração, transformação e carregamento dos dados. De entre estes, a fase mais importante e complexa é a transformação dos dados. Este processo deve ser automático e existem várias maneiras de o abordar. Uma delas seria programar estas transformações de raiz. No entanto, já existem ferramentas no mercado que tratam disto e a ferramenta escolhida foi o **Pentaho Data Integration (Kettle)**. Uma das razões da escolha desta ferramenta foi o facto de ela preencher certos requisitos necessários a este projeto como a possibilidade de carregar dados de diferentes fontes (xsl, csv, xml, etc.), a capacidade de elaborar diversas transformações de dados e principalmente o facto de permitir execuções agendadas e automáticas. Outra das razões deve-se ao facto de esta ferramenta ser rápida e de fácil utilização devido à sua interface gráfica. Mais detalhes sobre esta ferramenta e todo o processo ETL podem ser consultados no capítulo seguinte.

4.2.3. Análise OLAP

A análise OLAP é sem dúvida a fase fundamental deste projeto no qual são analisados e calculados dados que posteriormente serão expostos ao utilizador sob a forma de gráficos e tabelas. A ferramenta escolhida para esta parte do projeto foi o **Pentaho BI Server** que se distingue de outras ferramentas pelos seus *dashboards* de fácil construção^{[8][9]}. Com a adição do *plugin CDE*, é possível definir o *layout* da página, adicionar componentes pré-definidos como caixas de seleção, caixas de texto, gráficos, tabelas, etc., e definir os *datasources* dos dados a apresentar^[15]. Outra razão para a escolha desta ferramenta é a possibilidade de integrar o servidor de OLAP **Mondrian**^[10] que permite a construção de cubos que representam modelos de dados multidimensionais.

4.2.4. Interface Web

A apresentação dos indicadores de desempenho é feita através de uma interface *web* acessível através de autenticação dos elementos dos órgãos de gestão da UC. Esta é disponibilizada através dos *dashboards* do Pentaho BI Server e faz uso naturalmente de ferramentas típicas da *web* como o **HTML** e o **JavaScript**. Na demonstração dos gráficos é utilizado uma biblioteca de JavaScript denominada **Highcharts**^[11]. Esta foi utilizada na realização dos protótipos e é bastante versátil com diversos tipos de gráficos já definidos.

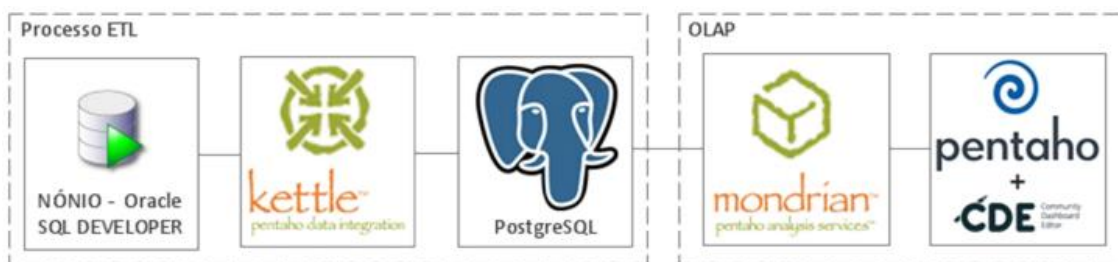


Figura 6 - Tecnologias utilizadas no desenvolvimento da aplicação

4.3. Modelo de Dados

Sem dúvida que a forma de armazenar os dados numa *data warehouse* é fulcral a uma análise OLAP fiável e responsiva. A disposição dos dados deve estar segundo um modelo multidimensional ou “em estrela” para que as *queries* à DW sejam rápidas e eficazes.

Para este módulo foram criados dois modelos de dados: uma área para dados temporários, e um modelo “em estrela” para a DW em si.

4.3.1. Modelo Área Temporária

Muitas vezes durante a extração de informação das fontes de dados, surge a necessidade de armazenar dados temporários para simplificar o processo ETL. No caso do sistema SAP, os dados dos recursos humanos estão distribuídos entre várias tabelas em que cada linha corresponde a um trabalhador, com determinadas características, de uma data inicial, a um data final. Sendo assim, é preciso intercetar os intervalos temporais das várias tabelas para cada trabalhador. Esses dados são guardados em tabelas temporais que depois servem como fontes de dados ao processo ETL.

Tabela	Descrição
temp_eti	Esta tabela armazena os valores de ETI dos trabalhadores. Esses valores podem variar durante certos intervalos de tempo por isso existe uma coluna para a data inicial e outra para a data final. A coluna do número mecanográfico é a que identifica o trabalhador em questão.
temp_rubricas	As rubricas são utilizadas para calcular a renumeração mensal dos trabalhadores e existem várias para cada um. No momento de carregamento desta tabela, é logo efetuada um soma das rubricas (indicadas pelo Dr. Filipe Rocha) de cada trabalhador para cada mês de um determinado ano.
temp_contrato	Esta é a tabela mais importante da área temporária pois armazena informação proveniente de várias tabelas do sistema SAP. Esta diz respeito a todas as especificações de contrato dos trabalhadores assim como alguns dados demográficos dos mesmos. Semelhante às outras tabelas da área temporária, ela contém uma coluna com a data inicial, outra com a data final e outra com o número mecanográfico

do trabalhador correspondente. O resto das colunas são informações sobre o contrato do trabalhador em questão.

temp_absentismo

Semelhante à tabela temporária dos contrato, esta tabela também armazena informação sobre os trabalhadores e respetivas especificações de contrato, mas neste caso só aquelas necessárias à aplicação de filtros. Respeitante ao absentismo, são guardados a duração da falta em dias e em horas, e o motivo.

Tabela 4 - Tabelas da área temporária

Visual Paradigm Standard Edition(University of Coimbra)

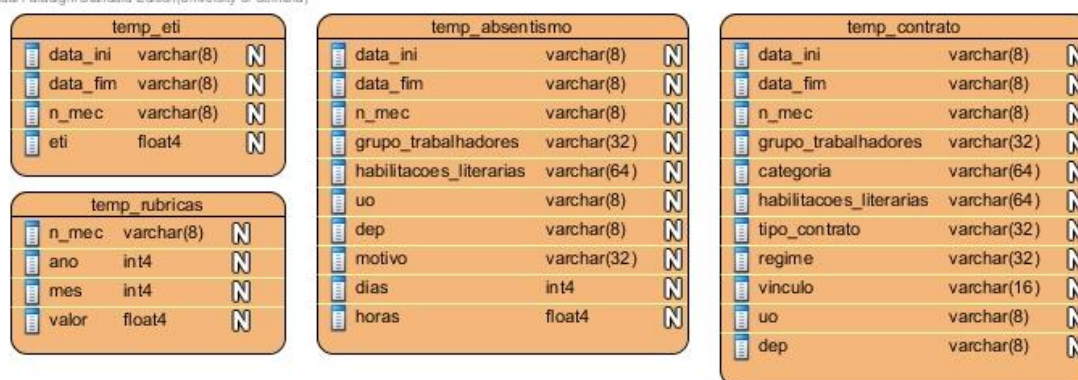


Figura 7 - Modelo de dados da área temporária

4.3.1. Modelo Multidimensional

Como já referido várias vezes ao longo do documento, o modelo multidimensional ou “em estrela” é o que mais se adequa à estrutura de uma *data warehouse*. Na verdade, o modelo usado na DW é considerado uma “constelação de factos” pois existem várias tabelas de factos partilhando dimensões entre si. Neste tipo de esquema, uma “estrela” é composta por uma tabela de factos e pelas dimensões associadas. Os factos contêm medidas ou métricas (atributos numéricos e aditivos) de determinados eventos e chaves estrangeiras para as dimensões. Estas representam as características desses eventos e complementam a informação dos factos.

A utilização deste tipo de modelo de dados trás várias vantagens à análise OLAP pois os esquemas “em estrela” são desnormalizados e isso trás algumas vantagens:

- *Queries* mais simples – a lógica de junção de tabelas é geralmente mais simples do que em esquemas relacionais normalizados.
- Melhor performance nas consultas – este tipo de esquemas oferece resultados mais rápidos em leituras à base de dados, o que é típico numa análise OLAP (muito mais leituras do que escritas).
- Agregações mais rápida – a disposição das tabelas e a simplicidade das *queries* resultam normalmente em operações de agregação mais eficazes.
- Cubos OLAP – um esquema em estrela permite a construção de cubos OLAP que aumentam consideravelmente a performance de pesquisas.

Em relação às desvantagens, a mais notória é o facto da integridade e da consistência dos dados não ser tão assegurada como num esquema altamente normalizado.^[17]

As tabelas de dimensão contêm geralmente uma chave primária que é utilizada pelas tabelas de facto para fazer a ligação entre elas. Normalmente as dimensões têm menos linhas quando comparadas com os factos mas no entanto contêm vários atributos para os descrever. As dimensões utilizadas na construção da *data warehouse* são as seguintes:

Dimensão	Descrição
d_tempo	A dimensão do tempo é geralmente considerada essencial a qualquer <i>data mart</i> e neste caso ela é associada a todas as tabelas de facto da DW. Ela representa o tempo e é utilizada para definir restrições

	temporárias nos dados. A sua granularidade mais fina é o trimestre em relação ao ano letivo e o mês em relação ao ano civil.
d_unidade_organica	Esta dimensão representa as unidades orgânicas da UC e é a única tabela que é carregada com dados do NONIO. Quando associada a uma tabela de factos significa que o trabalhador em questão pertence àquela unidade orgânica. São guardados a sigla, a designação completa da UO e se esta tem ou não departamentos associados (caso da FCTUC e da FLUC).
d_dem_contrato	Esta dimensão representa os dados demográficos dos contratos como o grupo de trabalhadores, a categoria profissional e as habilitações literárias do trabalhador associado. É utilizada no agrupamento de dados dos indicadores mas também na aplicação de filtros.
d_esp_contrato	Por sua vez, esta dimensão representa as especificações dos contratos dos trabalhadores como o tipo de contrato, o regime e o vínculo. É igualmente utilizada no agrupamento de dados e na aplicação de filtros.
d_faixa_etaria	Esta dimensão representa as várias faixas etárias em intervalos de 5 anos. É utilizada maioritariamente na aplicação de filtros, mas também como agrupamento nos indicadores da Demografia.
d_antiguidade	Como a dimensão anterior, esta representa intervalos de tempo de 5 anos mas neste caso em relação aos anos de serviço dos trabalhadores. É igualmente utilizada na aplicação de filtros, mas também como agrupamento nos indicadores da Demografia.
d_trabalhador	A dimensão do trabalhador é aquela que representa cada um dos recursos humanos. Esta identifica univocamente cada trabalhador através do seu número mecanográfico. Guarda informações como a nacionalidade, o género e data de nascimento do trabalhador, esta última utilizada no cálculo da faixa etária.
d_saida	A saída é a única dimensão que está apenas associada à tabela de factos das saídas. Simplesmente armazena os vários motivos possíveis de saída da universidade (cessação, aposentação, rescisão pelo contratado, etc.).

d_falta	Assim com a dimensão anterior, esta também só está associada a uma tabela de factos do absentismo. Contém os vários motivos possíveis de falta ao serviço (férias, licença de maternidade, etc.).
----------------	---

Tabela 5 - Tabelas de dimensão da data warehouse

Em relação às tabelas de facto, estas geralmente crescem na vertical, ou seja, contém vários registos já que representam uma granularidade no seu estado mais fino. Por exemplo, um contrato de um trabalhador durante 1 ano traduz-se em 12 registos (1 para cada mês) na tabela de factos. Os factos utilizados na construção da *data warehouse* são os seguintes:

Facto	Descrição
f_contrato	Esta tabela de factos representa os vários contratos dos trabalhadores e as suas características. É utilizada no cálculo dos indicadores de Caracterização dos trabalhadores e da Demografia. Contém métricas como o valor de ETI e a renumeração mensal, pois cada linha desta tabela está associada a um único mês da dimensão d_tempo. O cubo OLAP representante deste facto contém a medida “nº de pessoas” que é calculada através de um <i>distinct count</i> da coluna id_trabalhador. Também contém a medida “custo total” que resulta do somatório da coluna renumeracao_mensal, e a medida “custo médio mensal” que resulta da média ponderada da mesma coluna.
f_admissao	Este facto diz respeito ao número de admissões, tanto novos contratos, como mudanças de categoria profissional. É utilizado nos indicadores de Admissões e Saídas do grupo Outros. Não contém métricas e a única medida do respetivo cubo OLAP é o “nº de pessoas” que se calcula igualmente através de um <i>distinct count</i> da coluna id_trabalhador.
f_saida	Por outro lado, este facto diz respeito ao número de saídas dos indicadores Admissões e Saídas do grupo Outros. A única medida do seu cubo é também o “nº de pessoas”. No entanto, este facto tem associado a dimensão d_saida que permite um agrupamento dos dados por motivo de saída.

f_absentismo

Como o próprio nome indica, esta tabela de factos representa o absentismo, ou seja, as faltas dos trabalhadores. Contém métricas como a duração das faltas em horas e em dias. Como os outros cubos OLAP, o respetivo cubo contém a medida “nº de pessoas”. A tabela tem associada a dimensão d_falta que permite um agrupamento por motivo da falta.

Tabela 6 - Tabelas de facto da data warehouse

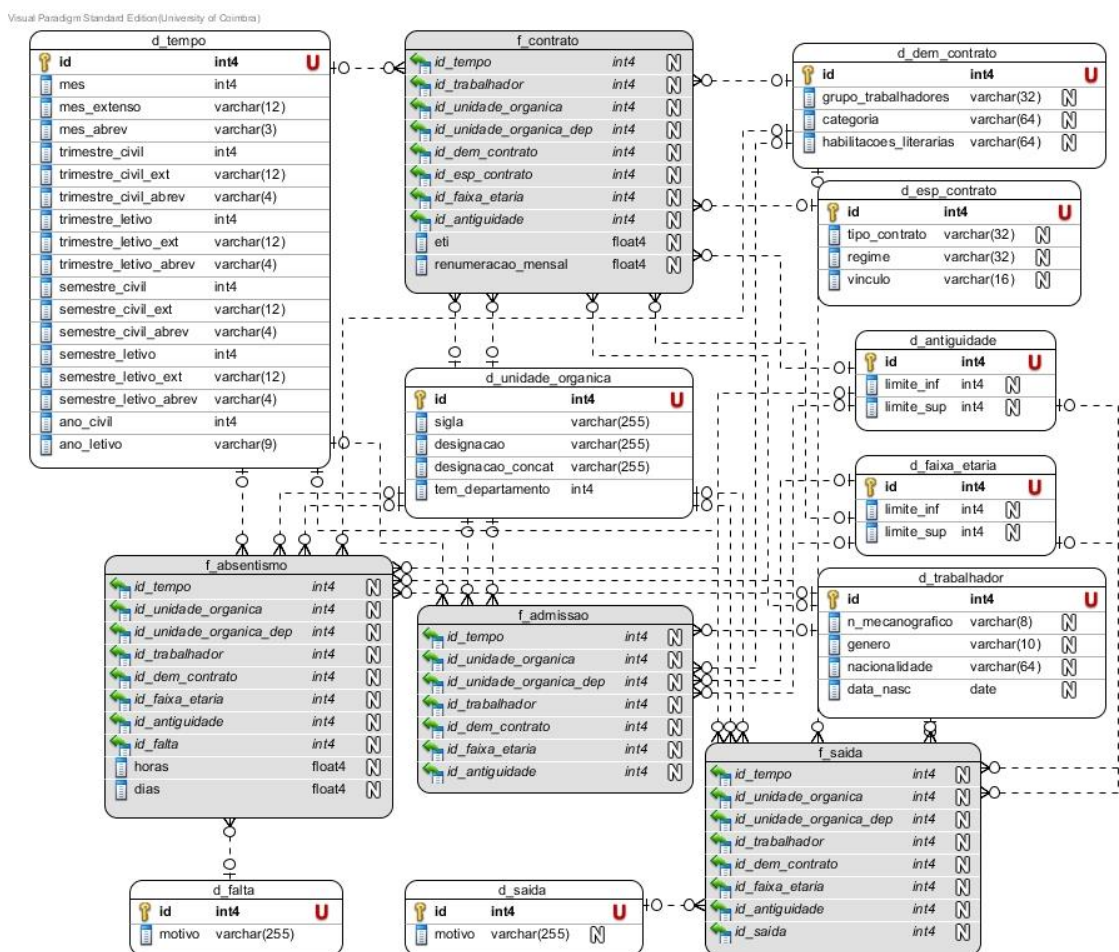


Figura 8 - Modelo multidimensional da data warehouse

5. Implementação





5.1. Processo ETL

O primeiro passo no desenvolvimento da aplicação foi a definição do processo ETL. Este foi realizado utilizando a ferramenta Kettle do Pentaho Data Integration que oferece uma interface gráfica de *drag and drop* das várias operações de extração, transformação e carregamento de dados. Para cada tabela, tanto da área temporária, como da *data warehouse* em si, foram construídas as respetivas transformações de dados através de esquemas que definem as ligações entre as operações e a sua ordem. Ao todo foram realizados 17 esquemas de transformação para todas as tabelas. Exemplos desses esquemas podem ser visualizados nas figuras dos próximos subcapítulos.

Em todo o processo de definição dos esquemas de ETL, houve a preocupação de os manter o mais simples possível, utilizando o mínimo de componentes necessários e colocando a maior parte do trabalho possível nas *queries* de obtenção dos dados. Assim, os esquemas tornaram-se não só bastante mais legíveis e de fácil interpretação, mas também ligeiramente mais rápidos na sua execução.

5.1.1. Componentes das transformações

Para perceber melhor os esquemas das transformações utilizadas no processo ETL, a tabela seguinte ilustra os componentes mais utilizados e uma breve descrição sobre as suas funcionalidades.

Ícone	Designação	Descrição
	<i>Add sequence</i>	Gera sequências numéricas a partir de um valor e com um certo incremento. Foi utilizado, por exemplo, para gerar os intervalos de 5 anos na dimensão d_faixa_etaria.
	<i>Calculator</i>	Calcula variadas operações entre valores no fluxo de dados.
	<i>Combination lookup/update</i>	Guarda um conjunto de dados numa tabela se eles já não existirem e gera as chaves primárias automaticamente. Utilizado maioritariamente no armazenamento de dados das dimensões.
	<i>Database lookup</i>	Dado um conjunto de dados, faz uma pesquisa na tabela e devolve os registos respetivos. Utilizado na pesquisa das chaves primárias das dimensões para inserir nas tabelas de facto.







	<i>Execute SQL script</i>	Executa um <i>script</i> SQL sobre a base de dados, normalmente utilizado para limpar as tabelas da área temporária.
	<i>Generate rows</i>	Gera registos com determinados dados nas colunas. Foi utilizado, por exemplo, para gerar os meses e anos da dimensão d_tempo.
	<i>Insert / Update</i>	Insere ou atualiza dados numa tabela tendo em conta um conjunto de valores a pesquisar. Utilizado maioritariamente no armazenamento de dados dos factos.
	<i>Modified Javascript Value</i>	Utilizando <i>javascript</i> , permite efetuar várias operações sobre os dados como alterar ou gerar novos valores.
	<i>Table input</i>	Dada uma <i>querie</i> SQL, devolve o resultado para o fluxo de dados da transformação.
	<i>Table output</i>	Armazena a informação num tabela dado um mapeamento entre os valores no fluxo de dados e as colunas da tabela.

Tabela 7 - Componentes das transformações do Kettle

5.1.2. Carregamento da área temporária

As tabelas da área temporária são, como o nome indica, tabelas com informação temporária e auxiliar ao preenchimento das dimensões e dos factos. Todas as transformações da área temporária são bastante similares e um exemplo disso pode ser visto na figura seguinte.

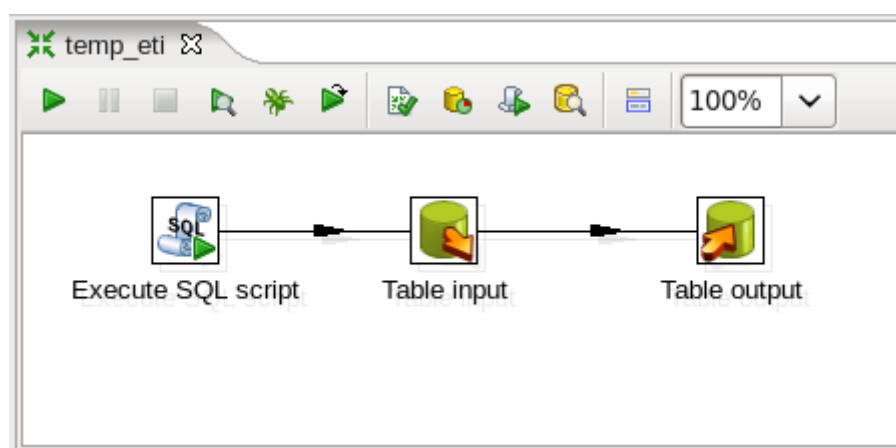


Figura 9 – Esquema de transformações da tabela temp_eti

O primeiro passo executa um *script* SQL que elimina todos os dados da tabela. Esta opção é mais simples do que atualizar os registos existente pois estes contêm intervalos temporais (colunas de data inicial e data final) o que torna a atualização dos campos mais difícil em termos de coerência de dados. Aparentemente, não existe grande problema em apagar os

registos anteriores pois são apenas dados temporários para o preenchimento das tabelas da DW. Além disso, a execução das transformações demora apenas alguns segundos. De seguida, os dados são carregados das respetivas tabelas do SAP e guardados diretamente na tabela temporária.

5.1.3. Carregamento das dimensões

Depois de carregadas as tabelas da área temporária, o passo seguinte é preencher ou atualizar as tabelas de dimensão da *data warehouse*. Como já referido, estas tabelas guardam informação que caracteriza os eventos das tabelas de facto, o que se traduz nas combinações existentes de valores das suas colunas. Por exemplo, a tabela d_dem_contrato, cujo esquema de transformações é ilustrado na figura seguinte, guarda todas as combinações de grupo de trabalhadores, categoria profissional e habilitações literárias dos recursos humanos existentes em SAP.

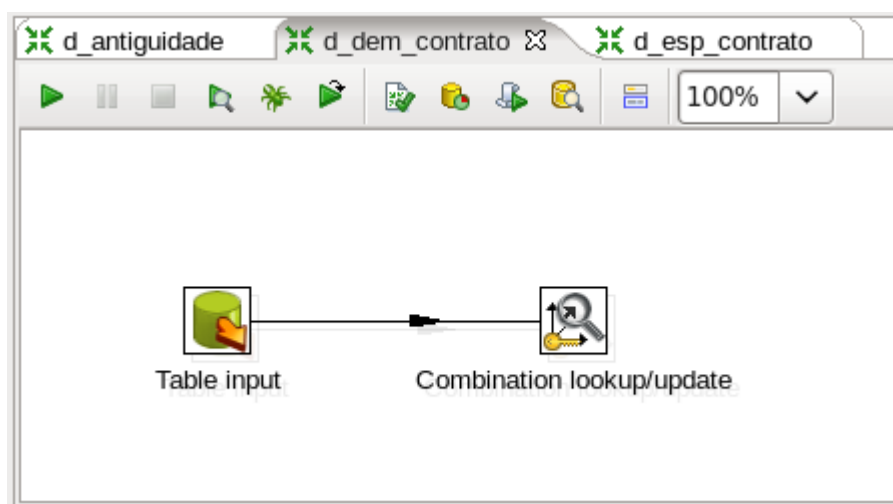


Figura 10 - Esquema de transformações da tabela d_dem_contrato

Os esquemas de transformação das dimensões são idênticos entre si, e terminam normalmente num componente de *combination lookup/update*. O que este faz é, dado um conjunto de valores a inserir, procura se esta combinação de valores já existe na tabela e só no caso de não existirem é que são adicionados. Este componente lida com as chaves primárias da tabela automaticamente.

Outros esquemas de transformação das dimensões são ligeiramente mais complexos. Alguns exigem geração automática de valores como é o exemplo das tabelas d_tempo, d_antiguidade e d_faixa_etaria.

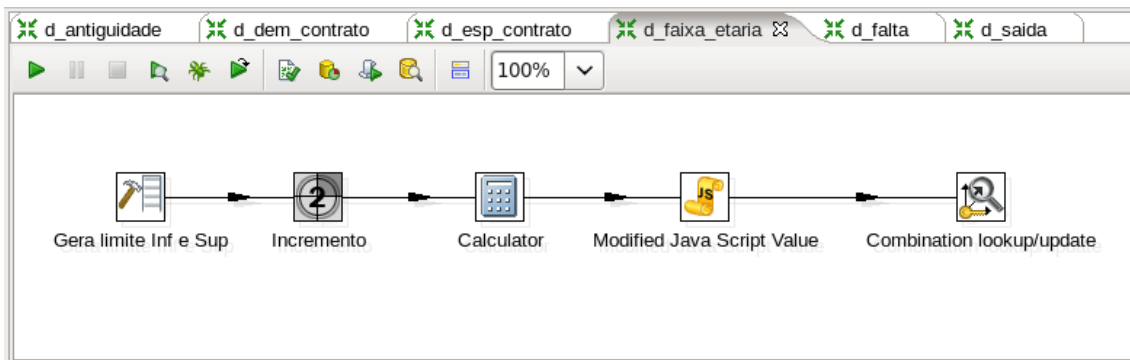


Figura 11 - Esquema de transformações da tabela d_faixa_etaria

Neste esquema, o primeiro componente gera as linhas, neste caso, duas colunas com os valores 0 e 4. O componente seguinte, a cada iteração gera uma sequência dado um valor inicial e um incremento, que neste caso era de 5. Finalmente, a calculadora é utilizada para somar estes valores, resultando na sequência de dados 0-4, 5-9, 10-14, etc., que definem os intervalos de idade da faixa etária.

Por fim, um caso especial das dimensões é a tabela d_trabalhador. Ao contrário das outras, nesta tabela os dados são armazenados utilizando um componente de *Insert / Update*. Este procura um registo na tabela dado o número mecanográfico do trabalhador, atualizando-o se ele existir ou adicionando um novo registo caso contrário.

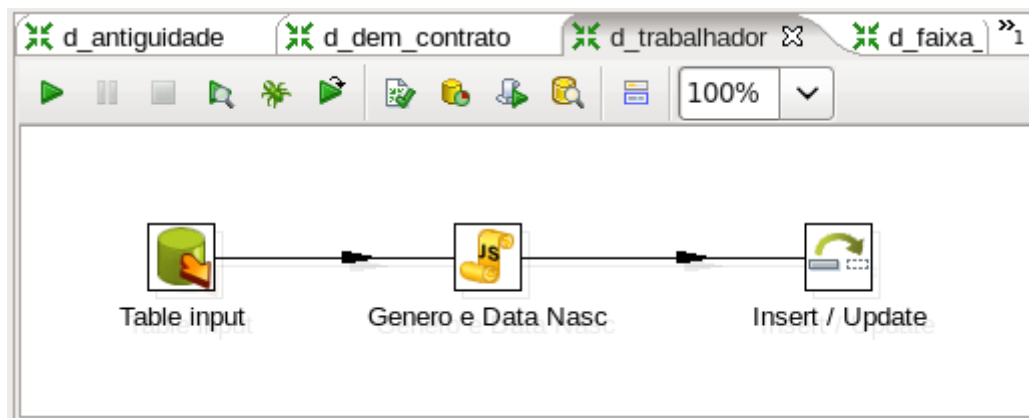


Figura 12 - Esquema de transformações da tabela d_trabalhador

Isto acontece porque com esta tabela não se pretende uma *slowly changing dimension*^[18], mas sim ter a informação pessoal mais recente de cada trabalhador. Na respetiva tabela em SAP existem alguns casos em que a informação pessoal dos trabalhadores como o género ou a nacionalidade foi alterada. Isto resulta em dois registos com o mesmo trabalhador

mas com intervalos de tempo distintos, sendo o mais recente a informação correta. Como ainda existem registos claramente errados como datas de nascimento no ano 1800 ou pessoas cujo nome é claramente feminino e no entanto estão definidas como homens, houve a necessidade de deixar espaço para a possibilidade de corrigir estes valores.

5.1.4. Carregamento dos factos.

O carregamento dos factos é a fase final do preenchimento da *data warehouse*. Estas tabelas, devido à sua granularidade mais fina ser o mês, contêm inúmeros registos pelo que as suas transformações demoram mais tempo que as outras, rondando a ordem das horas. Os factos contêm apenas métricas e chaves forasteiras para as dimensões, pelo que grande parte do processo é despendido na procura e obtenção dessas chaves. Um exemplo típico do preenchimento ou atualização de uma tabela de factos pode ser observado na figura seguinte.

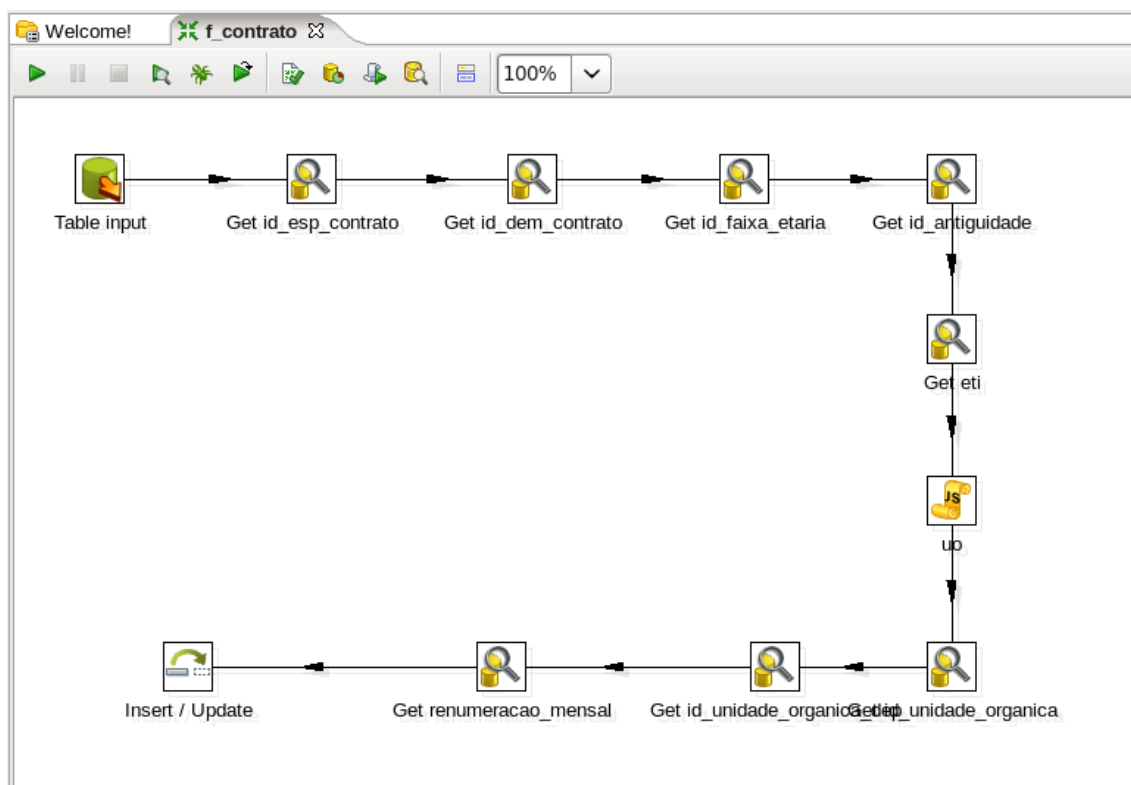


Figura 13 - Esquema de transformações da tabela *f_contrato*

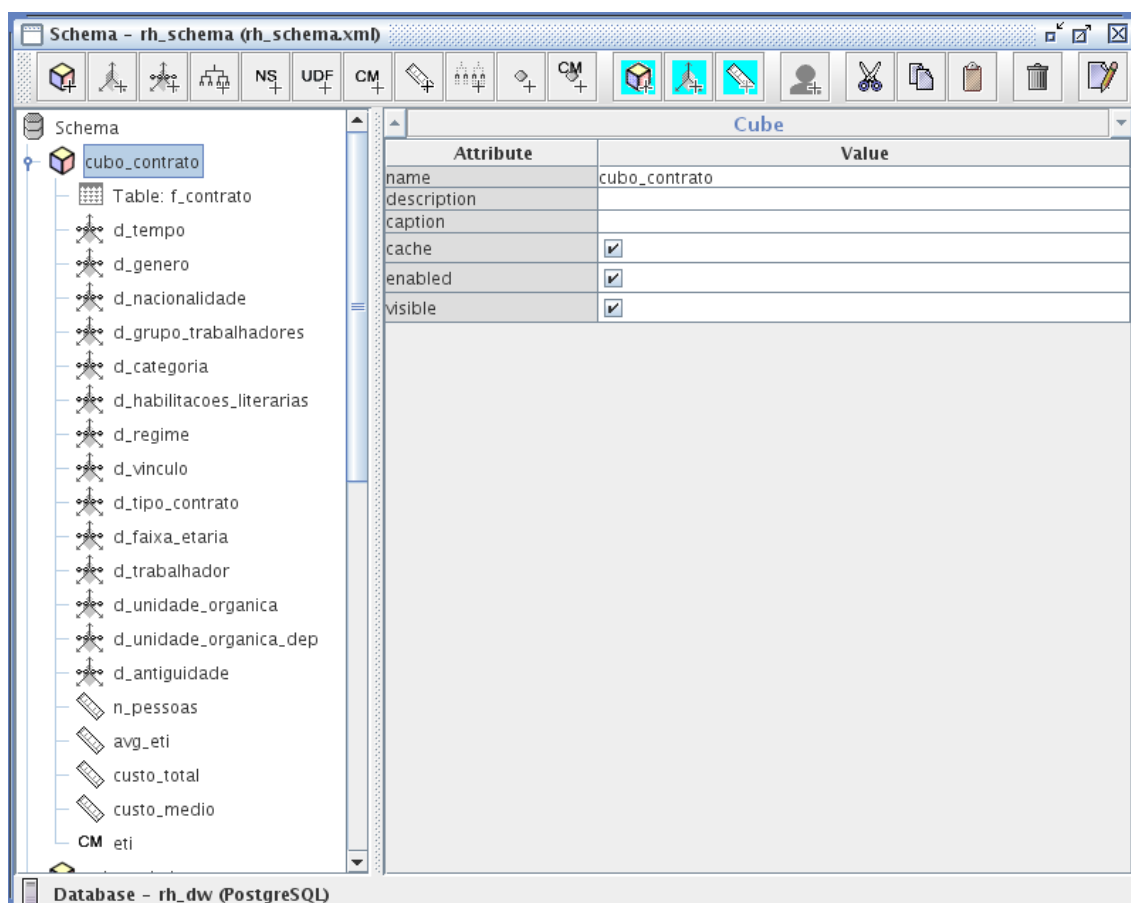
Após o carregamento dos dados, tipicamente oriundos de uma tabela da área temporária, são obtidas as chaves forasteiras para as tabelas de dimensão utilizando o componente de *database lookup*. Depois são obtidos os valores para as métricas do facto, neste caso o valor de ETI e a renumeração mensal. Finalmente os dados são armazenados utilizando um componente de *Insert / Update*, que neste caso atualiza a informação tendo em conta o

id_tempo e o id_trabalhador. Isto deve-se ao facto de, quando existe uma alteração dos dados de um contrato no sistema fonte, isso reflete-se num trabalhador num determinado período de tempo.

5.2. Cubos OLAP

A definição de cubos OLAP em aplicações deste género é muito importante pois aumenta consideravelmente a performance no acesso aos dados da *data warehouse*. Como já referido anteriormente, cada um destes cubos representa um modelo “em estrela” e são constituídos por uma tabela de factos e as dimensões que a caracterizam.

Os cubos são definidos através de um esquema em XML que posteriormente é publicado no Mondrian do Pentaho BI Server. Para a definição desse esquema foi utilizada uma ferramenta chamada **Schema Workbench**, também ela da Pentaho. Um exemplo de um cubo e da interface do programa pode ser visto na seguinte figura.



Como podemos ver, o cubo é constituído por uma tabela de factos (f_contrato), pelas suas dimensões (colunas de outras tabelas) e por medidas. Cada uma destas medidas diz

respeito a uma métrica (valor numérico ou aditivo) da tabela de factos à qual está associada uma função de agregação. Por exemplo, a medida `custo_total` é uma soma da coluna `renumeracao_mensal`. Já a medida `custo_medio` é uma média da mesma coluna. Uma dimensão pode ter várias hierarquias de valores como por exemplo, continente, país, distrito, etc.

Como já referido anteriormente, foi construído um cubo para cada uma das tabelas de facto. No entanto, na definição de um cubo, uma dimensão não é necessariamente uma tabela inteira podendo ser apenas uma coluna. Por exemplo, as dimensões `d_regime`, `d_vinculo` e `d_tipo_contrato` são todas colunas da mesma tabela. Por outro lado, a dimensão `d_tempo` corresponde à tabela homónima e contém duas hierarquias: `hie_civil` com os campos `ano_civil`, `semestre_civil`, `trimestre_civil` e `mes`; e `hie_letivo` com as colunas `ano_letivo`, `semestre_letivo` e `trimestre_letivo`.

5.3. Dashboards

Os *dashboards* da aplicação foram criados utilizando o *plugin* CDE para o Pentaho BI Server. Na construção destes existem essencialmente três menus de configuração: *layout*, *componentes* e *datasources*.

No primeiro menu, é onde é definido o HTML da página assim como os ficheiros CSS a utilizar e ficheiros externos de Javascript. Em termos de estilo da página, foi decidido entre equipa utilizar a *framework* **Bootstrap**^[18], o *plugin* **Chosen**^[19] para as caixas de seleção e a biblioteca **Font Awesome**^[20] para os ícones. Na criação dos gráficos, como já referido, foi utilizado o *plugin* de *javascript* Highcharts^[11].

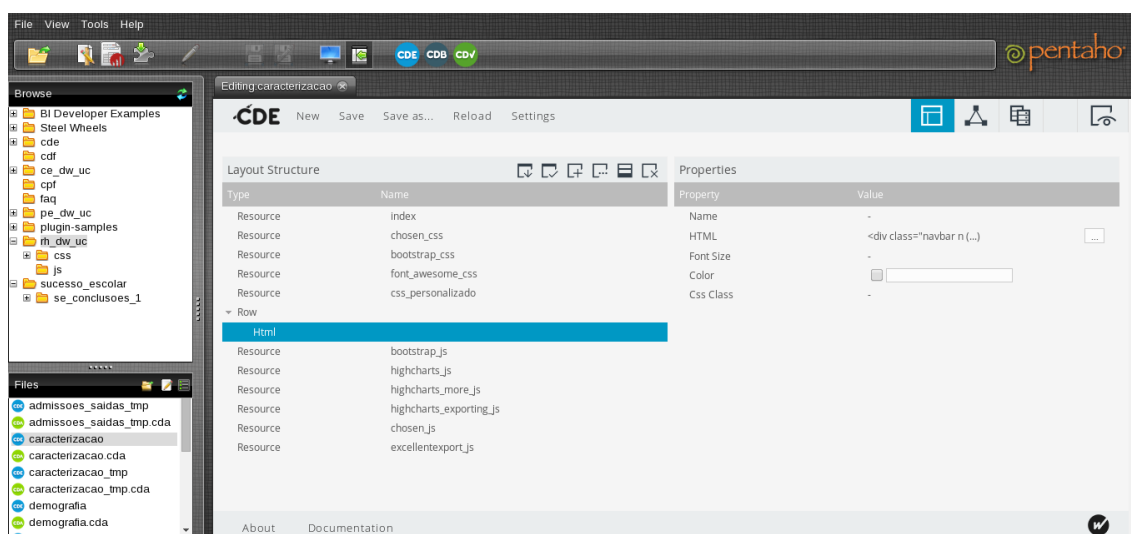


Figura 14 - Menu de layout do CDE

No segundo menu é onde é possível adicionar os componentes que são utilizados na página *web* como caixas de seleção, gráficos, tabelas, parâmetros, *scripts*, etc. Os componentes que têm um impacto visual na página como as tabelas, gráficos e caixas de seleção têm de ter um elemento HTML associado para que estes sejam visíveis e funcionais. Os parâmetros aqui definidos são o que permite a interatividade entre os componentes. Por exemplo, quando existe a alteração do valor de uma caixa de seleção, digamos do filtro género, este valor é guardado num parâmetro. Por sua vez, os gráficos devem ter esse parâmetro na sua lista de *listeners* pois uma mudança desse valor implica uma nova *querie* e a atualização dos mesmos.

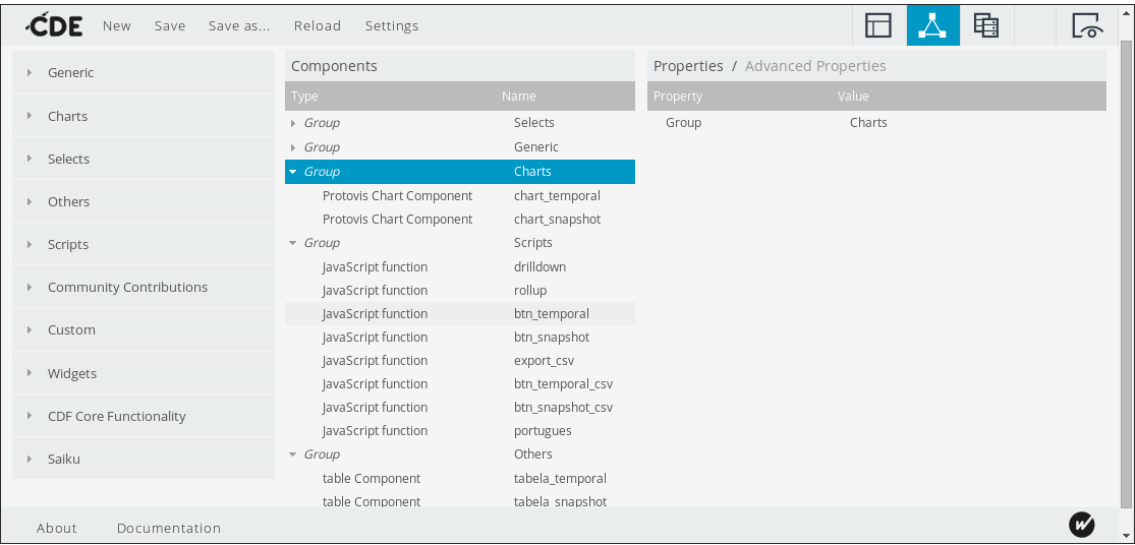


Figura 15 - Menu de componentes do CDE

Por último, o menu dos *datasources* permite, como o próprio nome indica, definir as fontes e formas de acesso aos dados.

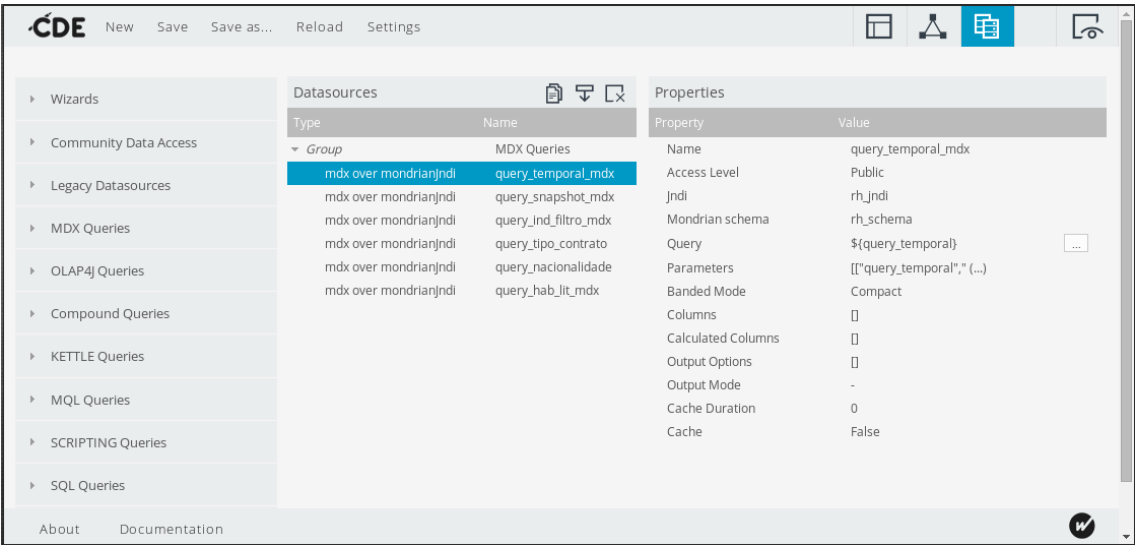


Figura 16 - Menu de datasources do CDE

Neste caso foram definidas as várias *queries* MDX que são executadas sobre os cubos OLAP. As *queries* que obtêm dados para apresentar nos gráficos são geralmente complexas e variam consoante a análise pretendida pelo que era impensável definir todas as *queries* possíveis. Sendo assim, estas utilizam uma *string* que é construída gradualmente conforme os dados pretendidos e os filtros a aplicar. Um exemplo destas *queries* pode ser visto na figura seguinte, que resulta no número de docentes por categoria para o ano letivo de 2013/2014.

```
SELECT NON EMPTY CrossJoin({[Measures].  
[n_pessoas]}, {[d_tempo.hie_letivo].  
[2013/2014]}) ON COLUMNS, NON EMPTY  
{Hierarchize({[d_categoria.hie_categoria].  
[categoria].Members})} ON ROWS FROM  
[cubo_contrato] WHERE  
{[d_grupo_trabalhadores.hie_grupo_trabalhadore  
s].[Docentes]}
```

Figura 17 - Exemplo de uma *querie* MDX

5.4. Resultado Final

O resultado final da aplicação pode ser observado através dos ecrãs disponibilizados ao utilizador. Ao todo foram construídos 4 *dashboards* para a análise de indicadores de desempenho: Caracterização; Demografia; Admissões e Saídas; e Absentismo. Um exemplo destes ecrãs pode ser visualizado nas figuras seguintes.

Na primeira figura podemos observar o resultado do *dashboard* da Caracterização dos trabalhadores. Neste exemplo é exposta a informação sobre o nº de docentes por categoria profissional, na FCTUC, entre o ano letivo de 2011/2012 e 2013/2014.

No menu de topo encontram-se as ligações para os restantes módulos e também para os outros *dashboards* dos Recursos Humanos (submenu clicando sobre o nome do módulo). Também vemos aqui os botões de ajuda (FAQ) e de terminar sessão (Sair).

De seguida temos a barra de navegação com os vários níveis de granularidade e neste caso encontramos-nos ao nível da FCTUC, isto é, ao nível da Faculdade/Unidade Orgânica selecionada. Clicando num dos níveis (Universidade de Coimbra ou Unidades Orgânicas) podemos subir na granularidade dos dados (*roll up*). Ao clicar numa das barras do gráfico de *snapshot* (à direita) podemos descer de granularidade para o nível dos departamentos (*drill down*) que pode ser visto na próxima figura.

Continuando a análise do ecrã, à esquerda podemos observar o menu lateral dividido em 3 partes: Parâmetros Gerais; Filtros; Parâmetros Temporais. Nos parâmetros gerais é possível escolher o indicador que pretendemos analisar, o grupo de trabalhadores e a opção de agrupamento. Esta última difere consoante o grupo de trabalhadores selecionado como já referido em capítulos anteriores. A caixa de seleção múltipla abaixo representa o que queremos visualizar nos gráficos em relação ao agrupamento escolhido que, neste caso, são as categorias profissionais a analisar. Estando esta caixa vazia, podemos ver todas as categorias existentes. Na secção seguinte estão todos os filtros que são possíveis aplicar sobre os dados (filtros comuns e filtros específicos). Por último, na secção dos parâmetros temporais é possível escolher qual o tipo de vista que pretendemos na análise (ano letivo ou ano civil) e o período (anual, semestral, trimestral e mensal). Dadas as opções anteriores é possível escolher o intervalo de análise (data inicial e data final).

No resto do ecrã podemos observar os gráficos de análise dos indicadores. À esquerda encontra-se o gráfico de evolução temporal e à direita o gráfico de *snapshot*. Clicando num dos períodos de tempo do gráfico temporal, neste caso um ano letivo, podemos observar os detalhes desse período no gráfico de *snapshot*.

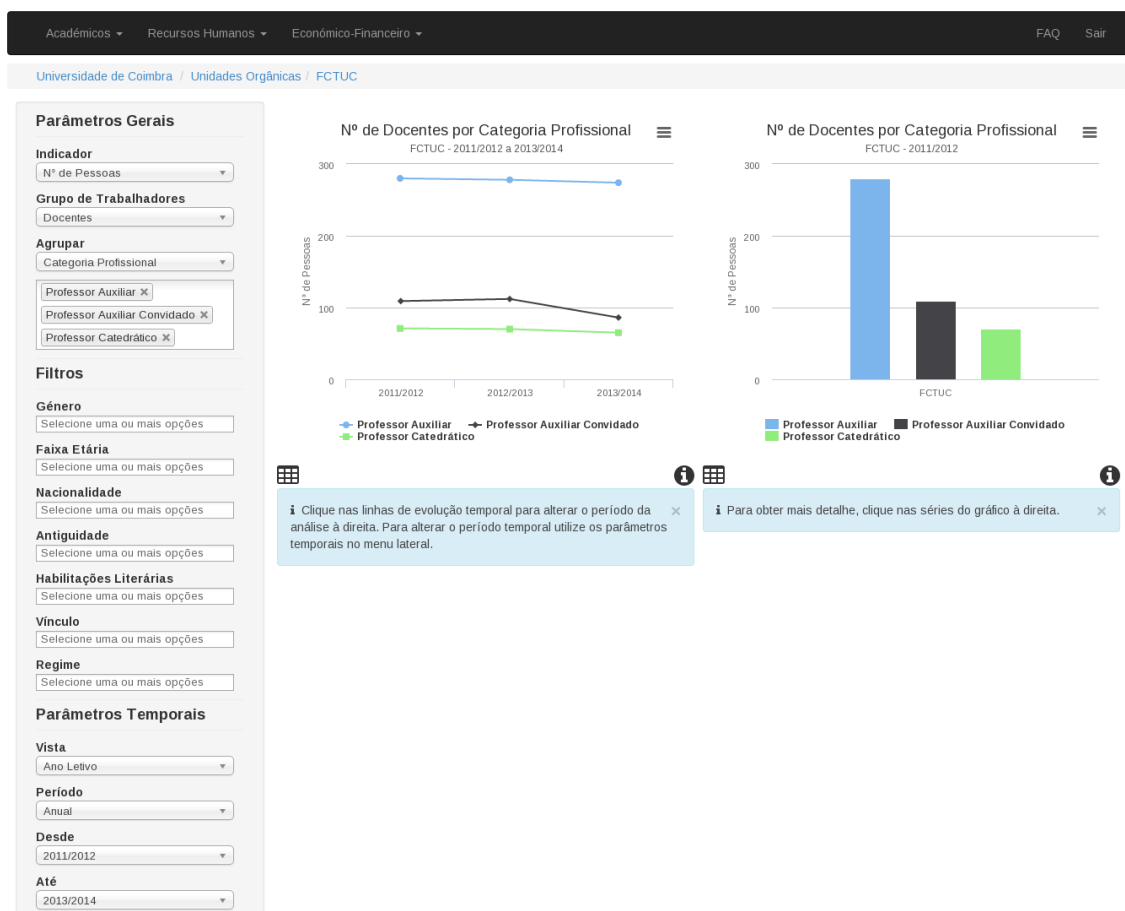


Figura 18 - Ecrã exemplo, nível FCTUC

Para cada gráfico estão disponíveis três botões com diferentes ícones. O botão no canto superior direito permite fazer *download* do gráfico sob o formato de imagem. No canto inferior esquerdo, o botão permite intercalar entre visualizar o gráfico ou a tabela respetiva. No canto inferior direito temos o botão que permite visualizar mais informação sobre o gráfico e o indicador de desempenho em análise (por exemplo, a sua formula de cálculo). Abaixo de cada gráfico encontram-se pequenas caixas de informação para ajudar o utilizador sobre como interagir com os gráficos.

Na figura seguinte podemos observar o mesmo indicador da figura anterior mas no nível de granularidade seguinte, os Departamentos. Aqui existe apenas um gráfico de *snapshot* que ocupa toda a largura do espaço de análise. A exclusão do gráfico de evolução temporal deve-se ao facto de que neste tipo de nível (Unidades Orgânicas e Departamentos) pretende-se uma comparação entre as unidades orgânicas do mesmo patamar, neste caso os departamentos da FCTUC. Para além disso, um gráfico de evolução temporal neste nível de granularidade seria praticamente impercetível pois, por cada unidade orgânica, haveria uma linha de evolução temporal para cada categoria selecionada, o que no exemplo seguinte resultaria em 36 (12x3) linhas. Assim, e dado o número de unidades orgânicas em paralelo, obtém-se mais espaço para o gráfico de *snapshot*, tornando-o mais percetível do que ocupando apenas metade do ecrã.

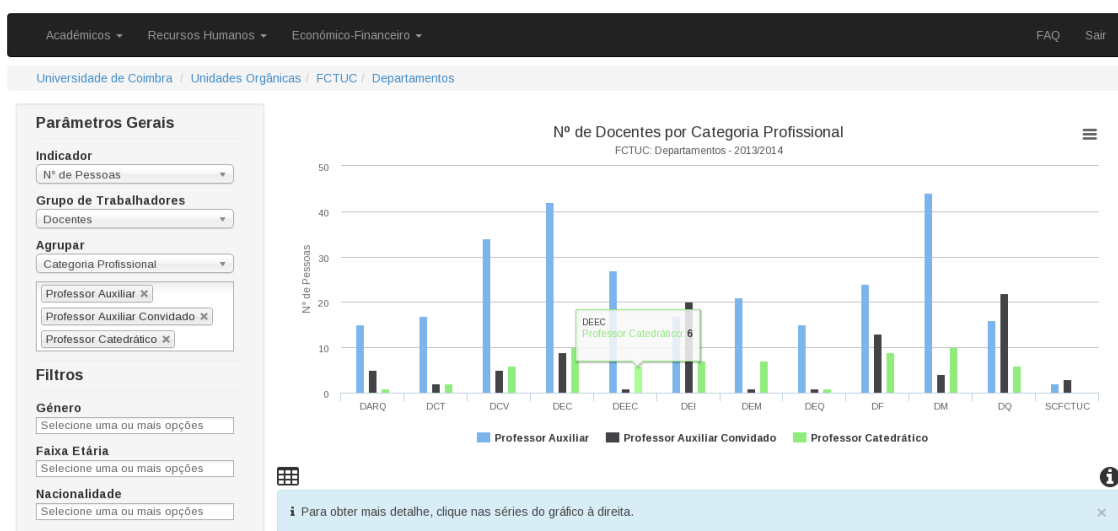


Figura 19 - Ecrã exemplo, nível Departamentos

5.4.1. Alterações

Desde a conceção dos protótipos até ao desenvolvimento dos *dashboards*, houveram algumas alterações em relação ao que estava inicialmente planeado. Algumas das diferenças podem ser observadas comparando a figura no subcapítulo da prototipagem rápida com a primeira figura do subcapítulo dos resultados finais.

A primeira diferença diz respeito às caixas de seleção dos filtros que passaram a ser caixas de seleção múltipla. Esta mudança oferece uma análise mais detalhada pois permite selecionar mais do que uma opção para cada filtro como por exemplo, filtrar os docentes com regime de exclusividade e os com regime de tempo integral. Isto trouxe outra modificação aos filtros pois deixou de ser necessário escolher um intervalo de faixa etária, isto é, escolher o intervalo de 10 anos 20-29 é o mesmo que selecionar os intervalos 20-24 e 25-29. Outra alteração parecida foi deixar de ser preciso selecionar um intervalo de habilitações literárias pelo que escolher o intervalo entre Licenciatura e Doutoramento é a mesma situação que escolher filtrar por Licenciatura, Mestrado e Doutoramento.

A última mudança foi a remoção do gráfico circular na parte inferior do ecrã. O único propósito deste gráfico era poder observar os valores percentuais relativos dos dados do gráfico de barras à direita. No entanto, este *layout* de 3 gráficos num ecrã não ia de encontro ao *design* utilizado nos outros módulos pelo que os valores percentuais podem ser antes observados passando o rato sobre as barras do gráfico de *snapshot*.

6. Testes e Validação

6.1. Validação dos Resultados

Em qualquer aplicação onde se faz um levantamento de requisitos, é claramente fundamental fazer uma validação do resultado final junto do cliente. Esta validação serve para verificar se todos os requisitos foram cumpridos e se o cliente está satisfeito com os resultados.

De forma a validar o módulo dos Recursos Humanos, foi marcada uma reunião junto do Dr. Filipe Rocha da Divisão do Planeamento, Gestão e Desenvolvimento no dia 22 de Junho no Edifício da Faculdade de Medicina, Polo I. Era importante obter a sua aprovação já que o levantamento de requisitos foi realizado junto deste.

Durante a reunião, a aplicação foi apresentada expondo todos os aspetos de análise aos indicadores de desempenho, passando pelas unidades medida, filtros e restrições temporais. O Dr. Filipe Rocha procedeu então à validação dos dados e dos *dashboards*, indicando apenas algumas sugestões para melhor aspetos de interface como adicionar *tooltips* com o nome completo de uma faculdade passando o rato sobre a sigla, ou ordenar as unidades orgânicas pela hierarquia usada na universidade. Outra sugestão dada foi aquando da visualização dos dados sob forma de tabelas, adicionar uma última linha com a soma das colunas. Estas sugestões foram posteriormente implementadas na aplicação final, pelo que se considera o módulo dos Recursos Humanos como validado.

6.2. Resultado dos Testes Funcionais

Os testes funcionais foram projetados de forma a validar os requisitos e verificar o correto funcionamento da aplicação. As seguintes tabelas apresentam os resultados dos testes realizados aos requisitos funcionais e não funcionais, indicando se eles foram cumpridos ou não. Todos os requisitos funcionais de prioridade elevada e média foram cumpridos pelo que a aplicação é considerada válida e funcional.

Código	Designação	Prioridade	Cumprido
RF_GE_01	Autenticação	Elevada	Sim. A autenticação é feita através do sistema LDAP da universidade.
RF_GE_02	Fechar sessão	Elevada	Sim. O utilizador pode terminar a sua sessão clicando no botão “Sair” no canto superior direito.
RF_GE_03	Término de sessão	Elevada	Sim. A sessão é terminada automaticamente depois de 15 minutos de inatividade.

RF_GE_04	Navegação entre módulos	Média	Sim. O acesso aos módulos é feito através da barra de navegação no topo do ecrã.
RF_GE_05	Navegação interna	Elevada	Sim. O utilizador pode descer na granularidade clicando nas barras do gráfico de <i>snapshot</i> à direita. Pode subir através da barra de <i>breadcrumbs</i> acima dos gráficos.
RF_GE_06	Parâmetros gerais	Elevada	Sim. Todos os parâmetros gerais e filtros estão disponíveis na barra lateral à esquerda do ecrã.
RF_GE_07	Parâmetros de tempo	Elevada	Sim. O utilizador pode alterar os parâmetros temporais disponíveis na barra lateral à esquerda do ecrã.
RF_GE_08	Esconder parâmetros	Baixa	Não. Requisito de prioridade baixa que apenas possibilitava mais espaço na interpretação dos gráficos. No entanto, os gráficos eram suficientemente percetíveis com a barra lateral pelo que o requisito ficou para segundo plano e acabou por não ser implementado.
RF_GE_09	Secção de ajuda	Elevada	Sim. A secção de ajuda ao utilizador “FAQ” encontra-se no canto superior direito.
RF_GE_10	Informação Auxiliar	Média	Sim. Para cada gráfico existe um botão no canto inferior direito que permite ver mais informações sobre o indicador e gráfico em causa.
RF_GE_11	Visualização gráfico ↔ tabela	Elevada	Sim. Para cada gráfico existe um botão no canto inferior esquerdo que permite visualizar a tabela correspondente.
RF_GE_12	Exportar informação da tabela	Baixa	Parcialmente. Implementada exportação para formato CSV.

Tabela 8 - Testes dos requisitos funcionais gerais

Os requisitos funcionais da área dos Recursos Humanos eram todos de prioridade elevada devido à sua importância no próprio módulo e foram cumpridos na totalidade.

Código	Designação	Prioridade	Cumprido
RF_RH_01	Unidades de medida	Elevada	Sim. O utilizador pode intercalar entre unidades de medida no topo do menu lateral.
RF_RH_02	Docentes por categoria profissional e regime	Elevada	Sim. É possível agrupar os docentes por categoria profissional e por regime.
RF_RH_03	Investigadores por categoria profissional	Elevada	Sim. É possível agrupar os investigadores por categoria profissional.
RF_RH_04	Não docentes por categoria profissional e tipo de contrato	Elevada	Sim. É possível agrupar os não docentes por categoria profissional e por tipo de contrato.
RF_RH_05	Dirigentes por tipo de cargo	Elevada	Sim. É possível agrupar os dirigentes por tipo de cargo.
RF_RH_06	Órgãos Sociais por tipo de cargo	Elevada	Sim. É possível agrupar os órgãos sociais por tipo de cargo.
RF_RH_07	Agrupar por dados demográficos	Elevada	Sim. Nos indicadores da demografia é possível agrupar os dados dos trabalhadores por habilitações literárias, faixa etária, antiguidade, nacionalidade e género.
RF_RH_08	Absentismo	Elevada	Parcialmente. Não foi possível calcular a taxa de absentismo pois não haviam dados sobre o número de horas trabalháveis, apenas a duração das faltas.
RF_RH_09	Admissões	Elevada	Sim. É calculado o número de admissões por categoria profissional.
RF_RH_10	Saídas	Elevada	Sim. É calculado o número de saídas por motivo.

RF_RH_11	Filtros gerais	Elevada	Sim. Em todos os indicadores é possível filtrar os dados por habilitações literárias, faixa etária, antiguidade, nacionalidade e género.
RF_RH_12	Filtros específicos	Elevada	Sim. Todos os filtros específicos foram implementados.

Tabela 9 - Testes dos requisitos funcionais dos Recursos Humanos

Em relações aos requisitos não funcionais foram todos cumpridos pelo que a aplicação cumpre assim com os atributos de qualidade especificados.

Código	Designação	Prioridade	Cumprido
RNF_S_01	Atualização de dados	Elevada	Sim. Na ferramenta de ETL foram agendadas execuções automáticas de mês a mês.
RNF_S_02	Compatibilidade (<i>browser</i>)	Elevada	Sim. Compatibilidade testada com os vários <i>browsers</i> .
RNF_S_03	Compatibilidade (SO)	Média	Sim. Testada e funcional em sistemas <i>Unix</i> e <i>Windows</i> .
RNF_S_04	Licenças	Elevada	Sim. Todas as ferramentas utilizadas são de licença gratuita.

Tabela 10 - Testes dos requisitos não funcionais

7. Conclusão

Neste documento foram apresentados os detalhes principais de todo o período de estágio deste ano letivo. Aqui estão explicadas todas as fases de trabalho no decorrer do projeto DW-UC, mais especificamente sobre o módulo dos Recursos Humanos. De notar que na elaboração deste documento foram tidas em conta as sugestões oferecidas pelo júri durante a defesa intermédia.

O balanço geral que retiro deste estágio é bastante positivo. Embora o facto de ter estado fora do país durante o primeiro semestre tenha dificultado a execução deste, penso que com a ajuda da equipa do projeto consegui superar as dificuldades e realizar o estágio com sucesso. Considero que cumpri os objetivos que defini e aprendi muito, não só a nível técnico como experienciar novas ferramentas e tecnologias, mas também a nível pessoal e social como trabalhar em grupo e lidar com clientes em reuniões de cariz profissional.

Em termos de perspetivas futuras, este projeto têm muito potencial para se desenvolver, seja na melhoria dos presentes indicadores como na introdução de novos. É definitivamente uma área que me desperta algum interesse e em que poderia possivelmente vir a trabalhar.

8. Referências

- [1] Kimball R. and Ross M., *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*, Second Edition, 2002.
- [2] Kimball R. and Caserta J., *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*, 2004.
- [3] Wikipedia, FURPS - <http://en.wikipedia.org/wiki/FURPS>
- [4] Wikipedia, processo ETL - http://pt.wikipedia.org/wiki/Extract,_transform,_load
- [5] Really Big Elephants: Data Warehousing with PostgreSQL - <http://www.slideshare.net/PGExperts/really-big-elephants-postgresql-dw-15833438>
- [6] 2ndQuadrant Professional PostgreSQL: Data warehousing with PostgreSQL - http://wiki.postgresql.org/images/3/38/PGDay2009-EN-Datawarehousing_with_PostgreSQL.pdf
- [7] Petaho Data Integration Steps - <http://wiki.pentaho.com/display/EAI/Pentaho+Data+Integration+Steps>
- [8] 5 Free Open Source BI Suites - <http://butleranalytics.com/5-free-open-source-bi/>
- [9] Pentaho and Jaspersoft: Good Alternates to Bigger-Name Software? - <http://slashdot.org/topic/bi/pentaho-and-jaspersoft-good-alternates-to-bigger-name-software/>
- [10] Mondrian Documentation: Mondrian and OLAP - <http://mondrian.pentaho.com/documentation/olap.php>
- [11] Highcharts JS: Interactive JavaScript charts for your web projects - <http://www.highcharts.com/>
- [12] Informação detalhada de tabelas SAP - <http://www.se80.co.uk/saptables.htm>
- [13] Wikipedia, SAP ERP - http://pt.wikipedia.org/wiki/SAP_ERP
- [14] Entidades aderentes do Observatório Nacional de Recursos Humanos - http://www.onrh.org/entidades_aderentes.php
- [15] Webdetails, CDE plugin - <http://www.webdetails.pt/ctools/cde.html>
- [16] Justinmind Prototyper - <http://www.justinmind.com/web-apps>
- [17] Wikipedia, Star Schema - http://en.wikipedia.org/wiki/Star_schema
- [18] Wikipedia, Slowly changing dimension - http://en.wikipedia.org/wiki/Slowly_changing_dimension
- [19] Bootstrap - <http://getbootstrap.com/>
- [20] Chosen - <http://harvesthq.github.io/chosen/>
- [21] Font Awesome - <http://fontawesome.github.io/Font-Awesome/>