

## **Gestão de Informação num contexto de transformação do negócio**

Rodrigo Schneeberger de Ataíde Oliveira

Relatório de Estágio apresentado como requisito parcial para  
obtenção do grau de Mestre em Gestão de Informação

*Esta página foi intencionalmente deixada em branco.*

**NOVA Information Management School**  
**Instituto Superior de Estatística e Gestão de Informação**  
Universidade Nova de Lisboa

# **GESTÃO DE INFORMAÇÃO NUM CONTEXTO DE TRANSFORMAÇÃO DO NEGÓCIO**

por

Rodrigo Schneeberger de Ataíde Oliveira

Relatório de Estágio apresentado como requisito parcial para a obtenção do grau de Mestre em Gestão de Informação.

**Coorientador:** Professor Fernando Bação

**Coorientador:** Professor Bruno Damásio

Março de 2021

## RESUMO

A indústria dos pagamentos tem vindo a evoluir a um ritmo acelerado. Isto tem sido possível com o desenvolvimento de tecnologias cada vez mais seguras e com a mudança de paradigma nos hábitos de consumo. Em Portugal, existe uma utilização generalizada de cartões bancários assente na confiança que os portugueses têm na rede multibanco. Não obstante, e talvez também por isso, ainda existe evidência de alguma resistência à utilização de métodos de pagamento digitais e na aquisição de produtos online.

Com a pandemia de COVID-19, houve uma aceleração na transição para modelos de negócio e de pagamento digitais. Com este choque surgem oportunidades para aqueles que, nesta indústria, tiverem agilidade suficiente para se adaptarem e oferecerem soluções que sirvam as necessidades de consumidores e comerciantes. É neste contexto que a Sonae Financial Services alarga a sua atividade comercial assumindo as componentes de risco, concessão, *servicing* e recuperação, aumentando a sua autonomia. Para suportar esta transformação do negócio, surge a necessidade de implementar aplicações e desenvolver novos fluxos de dados. Este estágio teve uma componente pedagógica de aprendizagem e de aplicação de conceitos teóricos e outra de apoio a uma equipa multidisciplinar responsável por garantir uma transição sem percalços.

## PALAVRAS-CHAVE

Gestão de Informação; Business Intelligence; Pagamentos; Cartões de Crédito.

# ÍNDICE

|                                                                |    |
|----------------------------------------------------------------|----|
| 1. Introdução .....                                            | 1  |
| 1.1. Pagamentos com Cartão .....                               | 1  |
| 1.2. Sonae Financial Services .....                            | 4  |
| 1.2.1. Cartão Universo .....                                   | 4  |
| 1.2.2. Modelo Comissionista .....                              | 6  |
| 1.2.3. Modelo de Concessão de crédito .....                    | 6  |
| 1.3. Motivação para a realização do estágio .....              | 7  |
| 1.4. Objetivos do estágio .....                                | 8  |
| 2. Enquadramento Teórico .....                                 | 9  |
| 2.1. Business Intelligence .....                               | 9  |
| 2.1.1. Data Warehouse .....                                    | 9  |
| 2.1.2. Modelação dimensional .....                             | 10 |
| 2.1.3. Camada Semântica .....                                  | 11 |
| 2.2. Data Mining .....                                         | 13 |
| 2.2.1. Processos Data Mining .....                             | 13 |
| 2.2.2. Data .....                                              | 14 |
| 2.3. Gestão de Conhecimento .....                              | 17 |
| 2.3.1. Metodologia de transmissão de conhecimento – SECI ..... | 18 |
| 2.3.2. <i>Agile</i> .....                                      | 19 |
| 3. Tecnologia e Aplicações .....                               | 21 |
| 3.1. Microsoft SQL Server .....                                | 21 |
| 3.1.1. SQL Server Analysis Services .....                      | 21 |
| 3.1.2. <i>Views</i> e TVF .....                                | 22 |
| 3.2. Excel .....                                               | 23 |
| 3.3. Power BI .....                                            | 23 |
| 3.4. API .....                                                 | 23 |
| 4. Projetos de Integração de Dados .....                       | 25 |
| 4.1. Integração de dados em DW .....                           | 25 |
| 4.2. Integração de dados do novo modelo de negócio .....       | 30 |
| 4.3. Novo modelo semântico – Tabular de crédito .....          | 32 |
| 5. Projetos de reporting .....                                 | 34 |
| 5.1. Central de Responsabilidades de Crédito .....             | 34 |
| 5.2. Dashboard adesões .....                                   | 39 |

|                                                            |    |
|------------------------------------------------------------|----|
| 6. Projetos de troca de informação .....                   | 41 |
| 6.1. Recuperações de crédito .....                         | 41 |
| 6.2. Imparidade e Scoring.....                             | 43 |
| 6.2.1. Ficheiros .....                                     | 43 |
| 6.2.2. Modelos.....                                        | 45 |
| 6.3. Variável Encargos financeiros e não financeiros ..... | 50 |
| 6.3.1. Encargos financeiros.....                           | 51 |
| 6.3.2. Encargos não financeiros .....                      | 54 |
| 6.4. Análise de Solvabilidade .....                        | 58 |
| 6.5. Partilha de dados com parceiro de titularização ..... | 60 |
| 7. Avaliação do estágio .....                              | 64 |
| 8. Conclusões.....                                         | 65 |
| 9. Bibliografia.....                                       | 67 |

## ÍNDICE DE FIGURAS

|                                                                                       |    |
|---------------------------------------------------------------------------------------|----|
| Figura 1 - Esquema do modelo de quatro entidades e exemplo de um Cartão Universo..... | 3  |
| Figura 2 - Pagamento com o Cartão Universo .....                                      | 5  |
| Figura 3 - Processo de transformação organizacional .....                             | 7  |
| Figura 4 - Diagrama da Metodologia CRISP-DM .....                                     | 13 |
| Figura 5 - Esquematização dos fluxos de dados.....                                    | 26 |
| Figura 6 - Esquematização dos novos fluxos de dados .....                             | 30 |
| Figura 7 - Modelo tabular de crédito .....                                            | 33 |
| Figura 8 - Arquitetura da solução de reporte à CRC .....                              | 36 |
| Figura 9 - <i>Sample</i> do <i>dashboard</i> de adesões .....                         | 39 |
| Figura 10 - Processo de geração de ficheiro com chamadas <i>call center</i> .....     | 42 |
| Figura 11 - Desenho de alto nível do processo de geração de ficheiros.....            | 43 |
| Figura 12 - Desenho dos processos de cálculo de imparidade e notação da carteira..... | 47 |
| Figura 13 - Estrutura do XML DISS do BdP .....                                        | 52 |
| Figura 14 - Output do serviço de cálculo de encargos não financeiros .....            | 56 |
| Figura 15 - Esquematização do processo de troca de ficheiros.....                     | 61 |
| Figura 16 - Pacote de geração de ficheiro.....                                        | 62 |
| Figura 17 - Pacote mestre com explicação.....                                         | 62 |

## LISTA DE SIGLAS E ABREVIATURAS

|             |                                         |
|-------------|-----------------------------------------|
| <b>AGC</b>  | Aplicação de Gestão de Cartões          |
| <b>BdP</b>  | Banco de Portugal                       |
| <b>CB</b>   | Core Banking                            |
| <b>CRC</b>  | Central de Responsabilidades de Crédito |
| <b>CU</b>   | Cartão Universo                         |
| <b>DM</b>   | Data Mining                             |
| <b>DPO</b>  | Data Protection Officer                 |
| <b>DSTI</b> | Debt Service-to-Income                  |
| <b>DW</b>   | Data Warehouse                          |
| <b>ETL</b>  | Extract, Transform, Load                |
| <b>IME</b>  | Instituição de Moeda Eletrónica         |
| <b>LGD</b>  | Loss Given Default                      |
| <b>MVP</b>  | Minimum Viable Product                  |
| <b>ODS</b>  | Operational Data Store                  |
| <b>PBI</b>  | Power BI                                |
| <b>SCD</b>  | Slowly Changing Dimensions              |
| <b>SFS</b>  | Sonae Financial Services                |

## 1. INTRODUÇÃO

A pandemia de COVID-19 está a acelerar mudanças de comportamento. Um exemplo destas mudanças é a forma como os consumidores olham para o dinheiro (notas e moedas). O seu potencial para disseminar vírus tem levado os indivíduos a procurarem formas alternativas de pagamento. Exemplo disto, são os pagamentos por *MB Way*, cartão bancário ou cartão de crédito. Desde o início da pandemia que tem sido observada uma redução significativa nos levantamentos em caixas automáticas. Isto é especialmente danoso para empresas de cartões de crédito que beneficiam das comissões aplicadas a este tipo de operação. A preocupação acentua-se, uma vez que é esperado que uma fração considerável dos efeitos deste choque seja de cariz permanente e não se retorne aos níveis de levantamentos pré pandemia. Para além disto, esta pandemia está a ser um catalisador do desenvolvimento de soluções inovadoras na indústria de pagamentos. Partindo da premissa em que "tempo é dinheiro", aplicável, em particular, a lojas movimentadas e com restrições de capacidade devido às medidas em vigor por todo o mundo, pagamentos mais rápidos, seguros e fiáveis podem dar um contributo importante na melhoria das margens de muitos negócios. É neste contexto que a Sonae Financial Services, através do alargamento do seu objeto social de IME para concedente de crédito, pretende aumentar a sua autonomia e capacidade de responder às novas necessidades dos consumidores com soluções inovadoras, como tem vindo a ser seu apanágio desde o lançamento do Cartão Universo.

### **1.1. PAGAMENTOS COM CARTÃO**

Um pagamento pode ser entendido como uma troca ou transferência de valor entre duas entidades. Os pagamentos são necessários para que uma economia funcione, já que constituem uma componente fundamental do comércio. É trivial que os pagamentos têm vindo a evoluir ao longo do tempo em paralelo com o desenvolvimento civilizacional e económico: desde as trocas, há 12 milhões de anos, até aos pagamentos com recurso a carteiras virtuais dos dias de hoje. O modelo mais simples, e com mais de dois mil anos de história, é o modelo das duas entidades, em que uma providencia um bem ou serviço, e outra líquida essa dívida, no momento, numa moeda aceite pelas duas. A partir daí, e em especial com a introdução de intermediários financeiros nos processos de pagamentos, começaram a existir entidades com o propósito de intermediar, oferecendo funcionalidades como conversões de moeda ou a possibilidade de fazer a transferência de fundos sem a troca de notas ou moedas. A forma mais comum desta funcionalidade é através da utilização de cartões bancários, de débito ou de crédito. Embora o fundamental para que este tipo de transações possa ocorrer seja que as entidades que pretendem transacionar confiem no intermediário, vamos discutir com mais detalhe como funcionam estes modelos complexos de pagamentos e o que envolvem.

Existem dois modelos distintos de transações com cartão bancário: o modelo com três entidades e o modelo com quatro entidades. O primeiro é aquele em que existem comerciantes, clientes e uma empresa intermediária que emite os cartões, gere os comerciantes e opera o sistema. Caracteriza-se por depender de um ecossistema fechado para a liquidação de transações, o que pode acabar por limitar o número de estabelecimentos em que um cartão pode transacionar. Isto porque o emissor dos cartões também tem a responsabilidade de angariar os comerciantes em que os cartões poderão transacionar. Um exemplo disto é a rede *American Express*. A SFS insere-se no modelo com quatro entidades. Este modelo é mais complexo, mas também tem mais destaque pela flexibilidade que oferece. Neste modelo estão envolvidas cinco entidades:

**Consumidor** – Aquele que pretende adquirir um bem ou serviço, utilizando um cartão ou o equivalente digital.

**Emissor** – A instituição financeira (registada na rede de pagamentos) que detém a relação contratual com os titulares dos cartões e é responsável pela sua emissão. É importante compreender que ao emitir um cartão, o emissor assume os riscos associados a esta atividade. Gerir os fundos dos clientes e saldar os pagamentos efetuados com os seus cartões é da sua responsabilidade. Logicamente, também o será cobrar aos clientes os valores das transações e receber os seus pagamentos. A sua atuação deve pautar-se pelo cumprimento da regulação em vigor. Por exemplo, em Portugal, para se ter esta atividade é preciso ter licença de Instituição de Moeda Eletrónica.

**Rede** – Disponibiliza a tecnologia, marca e regras que são responsáveis pela congregação das quatro entidades na mesma rede. No caso em estudo, é utilizada a rede *MasterCard*.

**Adquirente** – Uma instituição financeira que contratualiza com os comerciantes a aceitação e recebimento dos pagamentos. Esta entidade deve capacitar os seus clientes com tecnologia para aceitar cartões como método de pagamento, servindo assim de interface entre os comerciantes e a rede. É a entidade responsável por saldar as transações após o fecho do dia.

**Comerciante** – Um vendedor ou entidade que aceita receber pagamentos em cartão quando apresentados segundo um conjunto de regras.



Figura 1 - Esquema do modelo de quatro entidades e exemplo de um Cartão Universo

Neste modelo, a SFS IME tem o papel de emissor. Assim sendo, detém uma relação contratual com os clientes do Cartão Universo e é responsável pela sua emissão. As suas responsabilidades incluem gestão de saldos e consequentes autorizações de transações; e cobrança dos valores devidos, relativos à utilização do cartão, aos clientes. Contudo, a concessão e gestão de crédito é feita por um parceiro que também assume a responsabilidade de saldar os movimentos efetuados pelos cartões, e de assegurar a gestão comercial e de marketing dentro daquilo que está contratualizado com o parceiro.

Para além da componente operacional das transações, um cartão de crédito pressupõe a concessão de crédito. Este é um negócio relativamente simples e Consiste num acordo entre duas entidades na qual uma disponibiliza capital para aquisição de bens ou serviços, enquanto a outra se compromete a ressarcir este valor numa data futura, acrescido de juros. O juro é a remuneração pelo tempo em que a primeira entidade esteve privada do seu capital e pelo risco que existe de a segunda entidade não pagar. O cartão de crédito é a materialização deste tipo de negócio na sua forma mais simples para o portador. Este tipo de instrumento é identificado pelo BdP como um cartão associado a uma linha de crédito renovável. Ou seja, a entidade concedente de crédito garante que o portador do cartão poderá utilizar o cartão para realizar transações até ao total de um montante pré-definido, pagando este valor numa data futura e de acordo com as condições contratualizadas. O facto de a linha de crédito ser renovável significa que cada vez que o portador do cartão proceder à liquidação total ou parcial da dívida, este montante voltará a estar disponível para ser utilizado. Para além de cobrar juros, o concedente de crédito tem a responsabilidade de cobrar o Imposto de Selo Sobre a Utilização de Crédito.

Fica então claro que, para que exista um cartão de crédito, são necessárias duas atividades distintas que podem ou não ser da responsabilidade da mesma entidade, designadamente a gestão de autorizações e garantia do acesso à rede de pagamentos e à concessão de crédito como disponibilização do capital a ser adiantado ao cliente na expectativa de ser remunerado pelo risco.

## **1.2. SONAE FINANCIAL SERVICES**

A SONAE é uma das maiores empresas a operar em Portugal. No ano passado (2020), teve receitas na ordem dos 6,8 mil milhões de euros. Este estágio decorreu na Sonae Financial Services, a empresa do grupo responsável por desenvolver e implementar soluções de serviços financeiros. Alicerçada nos valores da holding - ambição e inovação - e com a responsabilidade acrescida de fazer parte de um grupo líder nos diversos setores em que opera, a SFS quer distinguir-se pela qualidade e inovação das suas soluções.

Embora o Cartão Universo seja o produto mais conhecido desta empresa, é importante referir que também fazem parte do seu portfólio os cartões *Dá*, *MoneyGram* e um conjunto de serviços relacionados.

O objetivo desta empresa, fundada em 2015, é ser líder no mercado de serviços financeiros com a oferta de produtos disruptivos desenhados à medida das necessidades das famílias portuguesas. Faz parte deste plano desenvolver parcerias estratégicas e alavancar-se nas mais de 1 500 lojas do grupo para chegar ao maior número de pessoas.

O estágio decorreu na equipa de Core & Data Solutions. Esta equipa, por sua vez, faz parte da área de Business Development, a área responsável pelo desenvolvimento de soluções tecnológicas sobre as quais os produtos são assentes. A Equipa de Core & Data Solutions é responsável por todos os fluxos de dados internos e externos. Para além disso, é responsável pelo armazenamento e disponibilização da informação aos diversos utilizadores das equipas de negócio. Junta assim fluxos operacionais e tarefas relacionadas com *business intelligence*. Esta equipa é composta por cinco pessoas de entre as quais analistas e engenheiros de dados e conta, com consultores de *business intelligence* subcontratados. Em suma, processa e integra a informação necessária à operação da SFS, disponibilizando-a depois para efeitos de monitorização, controlo e gestão.

Como o negócio da SFS é inerentemente fundado em soluções tecnológicas, as alterações aos produtos, campanhas de marketing e novas funcionalidades, são, necessariamente, avaliadas e implementadas por esta equipa. Ao juntar-me a esta equipa, a expectativa, para além do fim pedagógico do estágio, era de poder contribuir ativamente para a solidificação da relação entre as equipas de negócio e de desenvolvimento.

### **1.2.1. Cartão Universo**

O Cartão Universo foi lançado a 29 de outubro de 2015. Surgiu como a solução para juntar, num cartão apenas, os diferentes cartões de fidelidade e crédito das insígnias do grupo SONAE. Atualmente este produto conta com mais de 850 mil clientes, colocando-o em terceiro lugar na lista dos maiores emissores de cartões de crédito em Portugal. Embora este cartão seja emitido pela SFS, a linha de crédito é gerida por um parceiro reconhecido

internacionalmente na área de financiamento de particulares. A mecânica desta parceria será elaborada na próxima secção.

Uma das características que distinguiu este cartão, por ser inovadora e única no país, era a possibilidade de escolher, no terminal de pagamento, diferentes opções de parcelamento da transação com e sem juros. Atualmente esta funcionalidade já é oferecida por alguns dos concorrentes que não conseguem, no entanto, oferecer modalidades de três vezes sem juros em qualquer loja do grupo Sonae e da sua rede de parceiros. Esta funcionalidade é suportada por tecnologia *MasterCard* que também garante que este cartão seja aceite em todo o mundo. Ainda relativamente à tecnologia do cartão, está equipado com tecnologia *contactless*, permitindo a realização de pagamentos a crédito apenas com a aproximação do cartão ao terminal de pagamento. Por fim, este cartão possibilita a realização de pagamentos com o serviço *3D secure* – tecnologia que consiste no envio de um código único via SMS para que as transações sejam validadas.

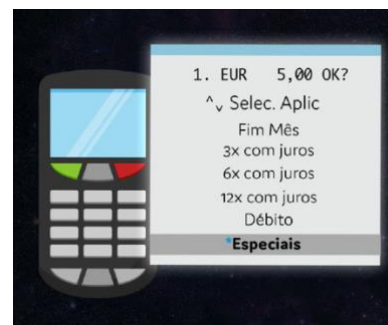


Figura 2 - Pagamento com o Cartão Universo

Outra das suas particularidades é a possibilidade de permitir a adesão de dois titulares no mesmo contrato. Tecnicamente, como veremos ao longo do relatório, esta funcionalidade oferece desafios. Não obstante, é importante, uma vez que assegura a possibilidade de partilhar a linha de crédito com um familiar que, de outra forma, não tivesse acesso a um produto deste tipo. Este cartão oferece ainda uma conta de débito até 10 mil euros. Sendo o único cartão no país, de que temos conhecimento, que tem dois números e dois códigos CVC, um para a conta de débito e outro para a conta de crédito.

Para além destas funcionalidades este cartão está associado a um programa de descontos robusto e diferenciado. Várias são as campanhas de grande relevância, destacando-se o *cashback* de 1% ilimitado e os descontos numa gasolinha até 16 cêntimos por litro. Uma particularidade é que estes descontos são carregados em Cartão Continente. Uma solução que, para a maioria dos utilizadores, é equivalente à disponibilização desses montantes no cartão e que tem como resultado o reforço da fidelização dos clientes. Este tipo de campanhas revela a importância que este cartão tem para o grupo SONAE e é a prova de que as sinergias podem ser benéficas para a melhoria da experiência do cliente numa rede de parceiros.

Este cartão oferece ainda uma aplicação móvel com funcionalidades que vão da gestão de transações a serviços adicionais como pagamentos, transferências de dinheiro entre clientes ou a subscrição de seguros. Esta aplicação já é utilizada, diariamente, por 457 mil clientes. O sucesso deve-se ao esforço de melhoria contínua que tem sido feito e que é reconhecido na avaliação de 4,5 na *App Store*.

Por fim, destaca-se uma experiência de adesão inovadora e diferenciadora. Um cliente pode escolher entre uma adesão simplificada ou completa sendo que, para a primeira, a

documentação é reduzida, com a contrapartida de poder requerer, no máximo, um *plafond* de 1 500 €. Independentemente do tipo de adesão que escolher, os novos clientes recebem, imediatamente, um cartão temporário que pode ser utilizado para transacionar nas lojas do grupo SONAE e respetiva rede de parceiros.

### 1.2.2. Modelo Comissionista

No momento em que o estágio se iniciou, o papel da SFS IME era o de emissor de cartão. Como foi mencionado na secção anterior, esta atividade era feita em conjunto com um parceiro (doravante “parceiro de crédito”). Neste modelo de negócio, descrito como modelo comissionista, existia uma relação contratual entre a SFS e o parceiro de crédito, em que o primeiro assumia todas as responsabilidades associadas com a gestão do cartão. Por seu lado, o parceiro de crédito era responsável pela disponibilização da linha de crédito e do respetivo capital. Por isto, a plataforma de adesões utilizada era a disponibilizada pelo parceiro de crédito dado que a decisão sobre a concessão ou não, e o limite a atribuir era da sua responsabilidade. Este parceiro responsabilizava-se, ainda, pela gestão de crédito. Por gestão de crédito consideram-se processos como a aplicação de juros, as responsabilidades regulamentares de reporte e os processos operacionais, como o bloqueio de cartões por risco de incumprimento. A gestão da delinquência da carteira, das recuperações de crédito devido e de provisionamento também se encontravam sob a sua alçada.

Por seu lado, a SFS era a entidade que geria a relação contratual com o cliente e emitia os cartões. Com a emissão dos cartões, surgem as obrigações da gestão dos cartões e das autorizações das transações dentro da rede *Mastercard*. A gestão das promoções comerciais, campanhas de marketing e de apoio ao cliente também é da sua competência.

Neste modelo, o parceiro de crédito tinha, como receita, o valor de juros e comissões imputadas aos clientes. Por seu lado, a SFS era remunerada pelas comissões das transações e o parceiro remunerava a empresa com base num conjunto de indicadores.

Este modelo de negócio foi uma fase importante da história da SFS, uma vez que permitiu lançar a operação com segurança e com disponibilidade de capital que suportasse um crescimento sem restrições. Contudo, como o leitor estará a adivinhar, esta parceria também coloca entraves ao crescimento do produto, no sentido em que a empresa não tem capacidade de decisão relativamente ao seu apetite por risco. Para além disso, no caso de o negócio ser bem-sucedido, a empresa fica alienada de parte significativa dos ganhos.

### 1.2.3. Modelo de Concessão de crédito

Com o aproximar da data de renegociação da parceria, a SFS ponderou várias soluções alternativas, ciente de que era importante aumentar a autonomia. Isto seria difícil no modelo

atual, uma vez que seria difícil chegar a um acordo em que o parceiro concedesse o crédito, mas não tivesse autonomia na sua gestão e concessão. A Sonae FS tomou a decisão de alargar a sua atividade à concessão de crédito. Este alargamento da atividade impõe a necessidade de garantir um conjunto de condições operacionais e tecnológicas (onde se destacam a necessidade de alterações à arquitetura, nomeadamente a implementação de um *core banking system* de crédito e respetivas integrações com o restante ecossistema), o que, naturalmente, constitui um desafio significativo.

De forma a materializar esta evolução, a Sonae FS solicitou o alargamento do seu objeto social, tomando vantagem total da sua licença de IME

Assim sendo, a Sonae FS, como entidade concessora de crédito, terá como objetivo controlar todo o ciclo de experiência dos seus Clientes e garantir total independência no desenvolvimento e manutenção dos seus produtos e serviços.



Figura 3 - Processo de transformação organizacional

### 1.3. MOTIVAÇÃO PARA A REALIZAÇÃO DO ESTÁGIO

A minha expectativa era a de que este estágio fosse uma oportunidade para pôr em prática aquilo que tenho vindo a aprender no mestrado. Estou particularmente interessado em tecer juízo crítico relativo às diferenças entre a metodologia lecionada e a utilização que é feita num contexto empresarial. A minha expectativa é a de aplicar conhecimentos ao nível do desenvolvimento de soluções em SQL ou noutra ferramenta que seja adequada. Também é meu objetivo continuar a desenvolver uma visão crítica sobre a forma como uma abordagem estratégica e unificada aos dados pode ser a raiz de uma vantagem competitiva no médio e longo prazo.

Acredito que este estágio será a melhor forma de complementar o meu mestrado, capacitando-me para discutir soluções tecnológicas e modelos conceptuais com propriedade e conhecimento de causa. A ser bem-sucedido, este estágio, contribuirá em grande medida para o meu desenvolvimento profissional.

#### **1.4. OBJETIVOS DO ESTÁGIO**

Este estágio propõe-se a colocar o aluno numa equipa com um volume de trabalho elevado, desafiando, à partida, as suas metodologias de trabalho e a sua postura enquanto profissional. A duração do estágio vai coincidir com um período de grande mudança na empresa. Uma vez que esta mudança é tanto organizacional como tecnológica, é esperado que o aluno capitalize a sua licenciatura em economia e aprofunde os seus conhecimentos em gestão de informação para dar suporte a esta transformação, fazendo a ponte entre os requisitos do negócio e os constrangimentos tecnológicos, propondo e executando soluções sempre que possível.

É importante evidenciar que, para além do propósito pedagógico, a posição de estagiário pressupõe que o aluno aumente, de forma efetiva, a capacidade da empresa de fazer face à transformação. Assim, embora seja dado espaço para desenvolver soluções e escrever código, é esperado que tenha uma abordagem mais funcional de desenho de soluções e gestão de projeto. Isto implica que, embora para fins pedagógicos exista flexibilidade nas funções desempenhadas pelo aluno, estas estarão mais próximas das de um analista de dados do que de um programador.

Alguns dos objetivos do estágio, *a priori*, são o de conhecer, compreender e gerir ativamente fluxos de dados, quer operacionais, quer de *reporting*. Para além disso, é esperado que o aluno seja responsável por acompanhar e gerir um projeto de implementação durante a totalidade do seu ciclo de vida.

No fim do estágio, este deverá ter contribuído para formar um profissional capaz de olhar para a gestão de informação com uma perspetiva de negócio. No caso particular da transformação organizacional da SFS, recai sobre o aluno a expectativa de que contribua para a implementação das diversas soluções tecnológicas que suportam o negócio e faça uso dos conteúdos lecionados durante o mestrado para disseminar na empresa as melhores práticas metodológicas e algumas das tecnologias mais avançadas existentes no mercado. Finalmente, é esperado que o pupilo seja capaz de promover a utilização do manancial de informação existente na empresa numa ótica de geração de valor comercial.

## 2. ENQUADRAMENTO TEÓRICO

Neste capítulo vão ser discutidos alguns dos tópicos abordados nas disciplinas do mestrado e que foram importantes, ou diretamente aplicados ao longo do estágio. Serão discutidos conceitos básicos e alguns modelos teóricos considerados relevantes para a construção da abordagem metodológica. É feita também uma tentativa de apontar em que situações é que cada um dos conceitos foi utilizado. Não obstante, dada a transversalidade ou o cariz conceptual de alguns dos assuntos esta ligação nem sempre é feita com detalhe.

### 2.1. BUSINESS INTELLIGENCE

Uma das formas mais simples de definir *Business Intelligence* é como sendo a evolução dos sistemas de informação executiva (*Executive information systems*). Estes consistiam na compilação de informação de negócio para a elaboração de relatórios a pedido e cálculo de indicadores. Alguns destes sistemas já incluíam a possibilidade de manter *dashboards* atualizados. Com o desenvolvimento de metodologias de armazenamento de dados, começou a ser mais comum o armazenamento de forma ordenada de grandes volumes de informação, relacionados com o negócio. À área do conhecimento que lida com a gestão da informação gerada pelo negócio com o intuito de gerar valor podemos chamar *Business Intelligence*. As definições diferem de autor para autor e, mesmo num contexto aplicado, existem departamentos de BI com propósitos completamente distintos. Podemos dizer que BI é um conceito que agrega arquiteturas, ferramentas, tecnologias e metodologias para dar suporte à tomada de decisão. No entanto, não se limita à tomada de decisão executiva, mas sim ao suporte à tomada de decisões que possam beneficiar do acesso à informação. No caso da SFS, BI está relacionado com os processos de armazenamento de informação no DW (Universo *Data Model*) e a sua disseminação quer através de *queries* para utilizadores mais avançados, como com recurso a um modelo semântico de forma mais generalizada.

#### 2.1.1. Data Warehouse

Um dos conceitos fundamentais de BI, e à volta do qual, normalmente, tudo o resto revolve, é o de *Data Warehouse*. Isto consiste num repositório físico onde os dados são organizados seguindo o modelo relacional. Segundo Inmon, existem algumas características que definem um DW. Este deve ser orientado ao assunto, ou seja, deve estar organizado de modo a que a informação útil para responder a perguntas sobre uma determinada área deva estar concentrada num único sítio. Por este motivo, o modelo do DW da SFS é um conjunto de *Data Marts*, cada um otimizado para consulta. Estes *Data Marts* são específicos a uma área de atividade ou negócio, existindo apenas algumas dimensões que são comuns entre eles. Outra característica é que um DW deve ser integrado, ou seja, deve consistir na combinação de informação de várias fontes num único repositório. Esta é uma das principais atividades da

equipa que integrei. O desafio consiste em resolver inconsistência e garantir que informação díspar é modelada de forma lógica e útil para os utilizadores. Outro requisito de um DW é o de incluir informação ao longo do tempo. Isto é essencial para tentar compreender tendências que possam existir. Finalmente, deve ser não volátil, uma vez que a informação introduzida não deve ser alterada. Isto é verdade no caso da SFS, com exceção da necessidade de corrigir integrações ou de criar campos calculados para o histórico. O motivo para este requisito é trivial. Caso a informação pudesse ser constantemente ajustada, após ter sido introduzida, os valores que seriam obtidos nas consultas iriam variar com a data da consulta o que iria gerar inconsistências e falta de confiança na informação.

Embora aquando da minha entrada na SFS não existisse um ODS e a informação dos clientes ativos se encontrasse centralizada na aplicação de gestão de cartões, existem planos, e as arquiteturas desenvolvidas já tomaram em consideração a implementação de um ODS. Isto faz parte de uma estratégia de consolidação da abordagem a BI e é esperado que traga robustez e aumento de qualidade à informação integrada. Bem como a possibilidade de vir a permitir processos de integração mais ágeis e simples.

Para concluir, importa referir que a metodologia utilizada desde 2015, e que continua a ser utilizada na SFS, é a abordagem de Kimball. Ou seja, vão sendo desenvolvidos diversos *Data Mart* à medida que o negócio evolui e se expande que são integrados na Universo DM.

### 2.1.2. Modelação dimensional

A modelação dimensional é uma das competências mais elementares de um analista de dados ou consultor de BI. Este tipo de modelação foi desenvolvido especificamente para DW's com o propósito de ultrapassar limitações de outras metodologias de modelação de bases de dados. O seu propósito é otimizar a estrutura de dados para consulta, ou seja, garantir que um utilizador consiga, de forma imediata, realizar *queries* e compreender os seus resultados. Outro dos objetivos deste desenho é o de otimizar a *performance* de leitura em detrimento das atualizações ou da escrita.

Neste tipo de modelação, a informação é dividida entre factos e dimensões. De forma simples, factos são aquilo que aconteceu, aquilo que importa. No caso do DW da SFS, factos podem ser transações ou chamadas para o *call center*. As dimensões, por seu lado, são os atributos que caracterizam os factos, por exemplo, quem realizou a transação, quem ligou para o *call center* ou o comerciante onde a transação foi feita. Neste modelo, as tabelas com factos são a peça central; depois, a estas tabelas são, através de chaves, ligadas as tabelas das dimensões que os caracterizam. Uma particularidade é a de que, por norma, cada facto, está ligado a um valor de cada dimensão. Os factos podem ser de diferentes tipos, mas no contexto do negócio da SFS, os factos correspondem a eventos. As dimensões podem ter uma hierarquia, mas no caso da SFS foi privilegiada a *schema* em estrela, quando possível, que

sugere a utilização de uma tabela por dimensão. Esta escolha baseou-se na importância que a consulta desta informação tem para os utilizadores e no volume de *queries* que são feitas.

#### 2.1.2.1. SCD

Existem diversas tipologias de dimensões. No caso do DW da SFS, a predominante é *Slowly Changing Dimensions* (SCD). Este tipo de dimensão é utilizado para capturar as alterações na informação da dimensão ao longo do tempo. Um exemplo disto é a dimensão de clientes. Alguns dos campos não devem vir a ser alterados como o NIF. No entanto, outros, como o salário, o número de dependente, ou até o nome, podem sofrer alterações, sendo de extrema importância que estas alterações sejam registadas de forma coerente para quando as tabelas forem ligadas numa *query* ou num modelo semântico, ser claro qual a informação a que estamos a aceder. Existem vários tipos de SCD. Esta tipificação pode variar de autor para autor, mas vamos focar-nos nos tipos 1 e 2 que são os mais comumente aceites.

**Tipo 1** – Não regista valores de histórico; quando existe uma alteração, é feito um *update* à linha. Esta abordagem tem a simplicidade como vantagem, mas é inconveniente a perda de informação. Não seria indicada para a nossa dimensão de clientes, uma vez que, especialmente para efeitos de concessão de crédito, é importante saber as características do cliente em cada período.

**Tipo 2** – Esta é a abordagem mais comum em modelos dimensionais e consiste na inclusão de uma nova tabela com a informação atualizada. Para além disso, inclui campos de versionamento que podem ser as datas de início e fim da validade do registo ou, de forma mais simples, uma *flag* com a indicação do estado do registo – ativo ou obsoleto. Assim, é garantido o registo de informação histórica e permite fazer a correspondência entre os factos e os valores (na dimensão) que estavam válidos quando estes ocorreram. A desvantagem desta abordagem é a capacidade de processamento que exige e o esforço acrescido no desenvolvimento da solução. Pese embora que o ganho da utilização de SCD do tipo 2, ultrapasse, normalmente, estes custos.

Existem formas diferentes de capturar informação que varia ao longo do tempo e SCD que consistem na combinação das diferentes tipologias. O mais importante, no momento do desenho de uma dimensão, é compreender a informação de modo a adotar-se uma SCD com a robustez necessária à utilização que lhe vier a ser dada.

#### 2.1.3. Camada Semântica

O conceito de camada semântica não é recente. No passado, era comum que cada ferramenta de BI dispusesse da sua própria camada semântica. Contudo, nos dias de hoje, para soluções complexas, e num ambiente de mutação constante, isto não é uma hipótese praticável. Uma das definições mais comuns de "camada semântica" é a de que esta

corresponde a uma representação da informação orientada ao negócio. O propósito é o de que os utilizadores finais consigam fazer sentido da informação que lhes é facultada.

Uma camada semântica não é mais do que um nível de abstração sobre a informação guardada numa base de dados, normalmente num DW. Esta camada garante as relações entre as diversas tabelas, apresentando apenas a informação relevante e com nomes que sejam adequados à compreensão do utilizador, ao contrário do que se encontra muitas vezes nas colunas das tabelas do DW. Esta camada também permite que a informação seja disposta nas unidades corretas e permite criar hierarquias para que se consulte a informação de forma simples. Em suma, a camada semântica agrega a lógica de negócio subjacente à informação armazenada de forma centralizada. Isto permite, não só, que as diversas aplicações de BI possam aceder à informação de forma imediata, sem necessitarem de fazer transformações, mas também que alterações à lógica de disponibilização dos dados possam ser feitas de forma centralizada. Uma vantagem evidente deste tipo de arquitetura é a de oferecer uma visão uniforme, transversal à organização. Outra forma de passar esta ideia é a de que uma camada semântica oferece uma representação comum da informação. Um fator muito importante no desenho deste tipo de soluções é garantir que estas são capazes de responder aos requisitos contemporâneos, mas que também têm a hipótese de escalar para vir a fazer face a necessidades futuras. Não ter isto em consideração pode ser fatal numa estratégia de BI e desvirtua o propósito da implementação de uma solução deste tipo.

Tecnologicamente, existem várias formas de implementar estas soluções, estando, por exemplo, no mercado, múltiplos vendedores com diversas soluções cada. A escolha recai sobre a compatibilidade com as outras instâncias da empresa, se a solução é *on premise*, *SaaS* ou *PaaS*, com a capacidade de gestão de volumes da solução e de preço.

## 2.2. DATA MINING

As definições de *Data Mining* variam de autor para autor. No entanto, num cenário empresarial é ainda menos claro aquilo que é entendido por DM. No contexto deste estágio vai ser seguida a acessão de DM que foi aplicada no curso. Desta forma, DM pode ser considerada a área que engloba o conjunto de processos utilizados para transformar *raw data* em informação útil. Isto implica, mas não é restrito, a utilização de processos automáticos e semiautomáticos para encontrar padrões em grandes volumes de dados. É importante referir que existem dois grandes catalisadores para o crescimento e desenvolvimento que esta área de conhecimento tem visto nos últimos anos, designadamente a proliferação e consequente descida de preço da capacidade de processamento, e o crescimento do manancial de dados disponíveis. Estes dois fatores permitem a crescente utilização de processos cada vez mais complexos e avançados para encontrar respostas às perguntas colocadas pelos decisores e gestores operacionais. No caso da SFS, o facto de ser um negócio inerentemente tecnológico e a sua criação em 2015, fizeram com que tenham existido as condições para, desde o início das suas operações, ser garantido o armazenamento de informação. Este cuidado com a organização sistemática da informação constitui um terreno fértil para a aplicação destas técnicas.

Neste capítulo vão ser apresentados dois processos distintos e reconhecidos de DM, e vão ser discutidos alguns conceitos relacionados com dados.

### 2.2.1. Processos Data Mining

Os processos de *Data Mining* podem ser descritos como a abordagem metodológica à procura de conhecimento num grande volume de informação. Embora estes processos não sejam mencionados nos projetos, foram uma das bases mais importantes e com mais impacto no decorrer do estágio. Permitem que, aquando da procura de significado na informação, exista uma metodologia robusta que oriente o rumo da ação. Sem estes processos a guiar a extração de conhecimento em grandes volumes de informação, este caminho poderia ser errático e pouco eficaz.

#### CRISP-DM

CRISP-DM é a sigla para *Cross-Industry Standard Process for Data Mining*. Este é conhecido como o processo *standardizado* para a realização de processos de *Data Mining*, apresentado em 1999. As vantagens da existência de um enquadramento comum são garantir uma interoperabilidade de ferramentas durante o processo de extração de conhecimento e promover as melhores práticas de uma forma generalizada. Esta metodologia tem um cariz neutro, ou seja, pode ser aplicada, virtualmente, a projetos de *Data Mining* de qualquer índole. Resume-se

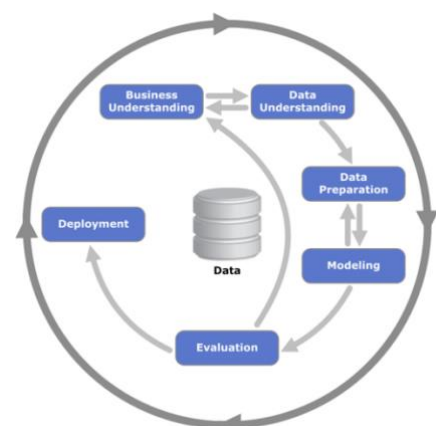


Figura 4 - Diagrama da Metodologia CRISP-DM

num conjunto de linhas orientadoras sobre as diferentes etapas que projetos nesta área devem seguir e a sua ordem.

Esta metodologia é composta por seis fases distintas:

1. **Compreensão do Negócio** – O seu principal objetivo é conseguir traduzir uma pergunta de negócio para um contexto de DM. Para isto, é preciso especificar o problema e desenhar um plano.
2. **Compreensão dos Dados** – Consiste na recolha dos dados, na sua exploração através de estatísticas descritivas, e visualizações e análise da sua qualidade.
3. **Preparação dos Dados** – Esta é a fase mais morosa, que implica o tratamento dos dados, a seleção dos dados a serem utilizados e a sua integração num formato que facilite a utilização por parte de aplicações de DM. As secções seguintes focar-se-ão em algumas das atividades inerentes a esta fase.
4. **Modelação** – Implica a escolha de uma técnica de modelação, no desenho de indicadores para a qualidade do modelo, e a construção do modelo *per se*.
5. **Avaliação** – É uma das fases mais fulcrais deste processo e onde deve ser feita uma avaliação cuidadosa do modelo, compreendendo se este é capaz de dar resposta à problematização em causa.
6. **Implementação** – É o culminar desta metodologia com a decisão da aplicação do modelo para dar resposta à pergunta inicial.

## SEMMA

*Sample, Explore, Modify, Model and Assess* é uma metodologia desenvolvida e patenteada pelo Instituto SAS. Este tem um portfólio vasto de aplicações de *Data Mining*. Estas aplicações estão, tendencialmente, organizadas seguindo as 5 fases desta metodologia. Por isto, esta forma de trabalhar tem ganho relevância. Esta metodologia deve ser interpretada como um *loop*, ou seja, quando alcançada a etapa da avaliação existem dois cenários: ou o resultado alcançado foi satisfatório; ou não se conseguiu alcançar uma resposta satisfatória para o problema em causa. Caso a solução não tenha sido satisfatória, esta metodologia sugere que se retorne à primeira etapa com a informação adicional reunida ao longo da primeira iteração do processo. Uma vez que não existe utilização de SAS no contexto da SFS, esta metodologia não foi utilizada de forma integral. Contudo, devido à ênfase que lhe foi dado durante o mestrado, foi seguida na abordagem mais imediata aos problemas de DM com que me deparei.

### 2.2.2. Data

Quando utilizamos um dos processos descritos anteriormente, de forma sintética, estamos a utilizar dados com o propósito de gerar conhecimento. Esta distinção nem sempre

é óbvia. Importa clarificar que um dado corresponde a um facto. Sem contexto, um dado pode ser considerado inútil. No contexto da SFS, dados surgem, frequentemente, relacionados com transações. Saber que o cliente 10000 gastou dez euros no comerciante 009, é uma informação. Mas o que importa, e o propósito da disciplina de DM é perceber que se este cliente realizou todas as suas transações num casino, então a sua probabilidade de incumprimento é alta e devemos reduzir-lhe o limite de crédito.

Para que seja possível dar este salto de dado para conhecimento, um aspeto fundamental prende-se com qualidade de informação.

### Qualidade dos Dados

Os processos de qualidade dos dados são importantes por, pelo menos, dois motivos, nomeadamente, garantir que a informação utilizada para monitorização da atividade económica é fidedigna e que pode ser utilizada nos processos de tomada de decisão; e permitir a construção de modelos precisos. Existem diversos motivos pelos quais os dados podem não ter qualidade, e este estágio mostrou-me que a realidade de uma empresa é a de que mais comumente se encontram falhas nos dados do que *datasets* prontos a serem utilizados. Exemplos de problemas com dados podem ter que ver com falhas na integração (uma coluna não foi integrada ou foram integrados dois campos na mesma coluna) que são simples de detetar e corrigir, mas, às vezes, existem problemas, como falha sistemática do preenchimento de um campo por não ser visível no dispositivo utilizado pelo operador. Existem também problemas de significado, por exemplo a necessidade de garantir a consistência após a atualização de um dicionário. Um problema com que nos deparámos após a transição foi como fazer sentido de informação que outrora era preenchida com recurso a um dicionário mais limitado.

Tendo sido estabelecida a importância de garantir a qualidade dos dados, importa compreender as seis dimensões fundamentais para a garantia da qualidade dos dados.

**Compleitude** – É imperativo que os dados venham preenchidos, especialmente aqueles que caracterizam de forma inequívoca o facto. Uma transação sem montante ou um cliente sem NIF não seriam úteis para qualquer tipo de aplicação.

**Consistência** – Os dados devem seguir um conjunto de normas que garanta a sua consistência de forma transversal a tabelas e aplicações. No caso da transição da SFS, foi importante garantir a consistência na marcação do sinal das transações. Se não tivesse existido esse cuidado, surgiriam muitos problemas de interpretação da informação.

**Precisão** – É crucial garantir que os dados traduzem de forma correta a realidade a que se referem. Numa realidade de transações com cartões de crédito, não faria sentido

se os montantes fossem arredondados para o milhar mais próximo ou, no caso de não serem em euros, não tivessem indicação da moeda.

**Formato** – Este ponto está relacionado com a consistência, mas importa garantir que os formatos são mantidos de forma uniforme. No caso da SFS, as datas estão todas no formato AAAAMMDD. Se surgirem situações onde este formato não é respeitado, irão surgir problemas na leitura dos dados.

**Janela Temporal** – Garantir que os dados são integrados de forma atempada enquadra-se nas características necessárias à qualidade dos dados, uma vez que, sem esta garantia, os decisores não terão confiança na informação que têm para a tomada de decisões. No projeto da CRC isto surgiu, diversas vezes, como um problema. Importa referir que dependendo da criticidade ou da latência dos processos que dão origem aos dados, aquilo que é considerado como uma janela temporal aceitável pode diferir.

**Integridade** – Existem vários tipos de integridade que necessitam de ser assegurados. Exemplos disto são a integridade relacional ou a integridade temporal. Vamos ver que nem sempre existe, o que dificulta os processos de troca de informação.

Finalmente, importa referir que a o cumprimento destes seis princípios pode ser a fundação de uma vantagem competitiva. O benefício mais evidente é a capacitação da empresa para a tomada de decisões rápidas e com confiança. Mas, como tem vindo a ser discutido, um *dataset* com qualidade também acelera o desenvolvimento de um modelo no contexto de um processo de DM.

### **2.3. GESTÃO DE CONHECIMENTO**

Gestão de conhecimento é uma disciplina que carece de aplicação prática para que se compreenda a extensão da sua importância. Pode ser vista como o conjunto de ferramentas, metodologias e sistemas envolvidos no processo de criação disseminação e utilização de conhecimento. É errado olhar para gestão do conhecimento numa ótica documentalista de gestão de ficheiros, uma vez que se dedica ao estudo de processos complexos, como a externalização de conhecimento tácito. Embora seja uma área de conhecimento, por si só, GC é um campo de estudo inerentemente multidisciplinar e que prolifera, assente em conhecimento que não lhe é exclusivo. Isto é importante porque GC é uma disciplina, por definição, que promove uma abordagem integrada às necessidades de gestão de informação de uma corporação. É imperativo enfatizar este último ponto, uma vez que o objetivo último de uma estratégia de GC é o de gerar valor a partir do capital intelectual da empresa. Exemplos de empresas com estratégias de gestão de conhecimento sofisticadas são as empresas de consultoria, dado que o seu ativo consiste no capital intelectual gerado nos diversos projetos que executam. Estas empresas acabam mesmo por atuar como uma rede de conhecimento. Um dos maiores desafios é o de transformar a experiência do próprio negócio em conhecimento palpável e passível de ser disseminado pela organização.

Para reforçar a importância desta área, importa mencionar que uma estratégia concertada de gestão de conhecimento, em muitos casos, contribui para a construção de vantagens competitivas duradouras. Estas podem ser concretizadas em aumentos de eficiência, na prevenção da perda de conhecimento e na melhoria de processos de decisão.

Um dos desafios que surgiu no estágio, e de que as organizações padecem diariamente, é o de transformar o conhecimento que os indivíduos obtêm no exercício das suas funções num recurso da empresa. Este desafio foi evidente na passagem do conhecimento que adquiri ao longo do projeto da CRC. Aquando do planeamento da minha saída, não existia uma metodologia clara para fazer transformação do conhecimento tácito em conhecimento explícito. Mesmo utilizando uma abordagem multi-suporte (apresentação, documentação escrita, e passagem *on the job*), esta mostrou-se insuficiente para, de forma efetiva, capacitar outros colaboradores com o conhecimento que interiorizei ao longo do estágio. Aspetos simples, como a quem ligar para resolver determinado problema, são exemplos deste conhecimento tácito que não é fácil de transmitir. As dificuldades desta passagem de conhecimento foram agravadas pelo contexto da pandemia de COVID-19. Foi deixado o alerta de que, pelas dificuldades acrescidas do contexto vivido (que pode ter efeitos na duração média da retenção de colaboradores), se torna, então, de suma importância o desenvolvimento de uma estratégia efetiva de gestão de conhecimento.

### 2.3.1. Metodologia de transmissão de conhecimento – SECI

Existem diferentes metodologias de transmissão de conhecimento, e várias implementações de cada uma destas com recurso a diferentes aplicações. No contexto do estágio, e sempre que existiu a necessidade de transmitir conhecimento, foi adotada a metodologia SECI, já que foi a que foi discutida com mais detalhe no contexto do mestrado.

SECI é o acrónimo para *Socialization, Externalization, Combination, Internalization*. Estas são as quatro fases deste processo, que irão ser descritas. O modelo SECI consiste num *loop* iterativo de conversão de passagem de conhecimento, não tendo um início claro nem um fim. A ideia deste modelo é que, se disseminado pela instituição e aplicado de forma consistente, permite uma partilha efetiva de conhecimento.

Para compreender melhor este modelo, importa primeiro distinguir conhecimento tácito de conhecimento explícito. Entende-se por conhecimento tácito aquele de que os indivíduos podem dispor sem que esteja necessariamente materializado. Exemplos disto são o conhecimento académico e a experiência. Conhecimento explícito, normalmente, está materializado em documentos de suporte, bases de dados ou plataformas de gestão de conhecimento. Pode também ser constituído por políticas da empresa.

**Socialização** – Consiste na transmissão de conhecimento tácito para conhecimento tácito. Ocorre quando existem interações formais ou informais entre aquele que dispõe do conhecimento e um grupo pequeno de indivíduos. Este passo do modelo está posto em causa de forma particularmente gravosa, e é o motivo por que muitas instituições da área de serviços estão reticentes a aderir ao teletrabalho. É verdade que existem formas de continuar a fazê-lo num contexto de teletrabalho, mas implica alterações profundas no modelo das interações digitais.

**Externalização** – Corresponde à transformação de conhecimento tácito em conhecimento explícito. Levei a cabo esta etapa quando fiz documentação extensiva dos projetos em que participei. Especialmente aquando da documentação de problemas encontrados e os diversos passos que levaram à sua resolução. Esta fase assume uma importância redobrada devido às limitações das interações sociais e a grande rotatividade dos colaboradores.

**Combinação** – Esta etapa corresponde à utilização de conhecimento de diferentes áreas para resolver problemas complexos. Utilizando o saber acumulado por uma área de gestão de produto e de uma área de inovação, por exemplo, pode levar a que se tomem decisões sobre o caminho a seguir, gerando mais conhecimento.

**Internalização** – Consiste na passagem de conhecimento explícito para conhecimento tácito. Uma das chaves para o sucesso desta etapa é a existência de um manancial de conhecimento explícito organizado e acessível.

Este processo foi eficaz no contexto do estágio na SFS. No entanto, verificou-se a consequência de uma das críticas que lhe é constantemente apontada, a de estar especialmente desenhado para um contexto japonês, onde a rotatividade dos colaboradores é baixa. Especialmente no processo de socialização, senti que o conhecimento ficava muito dependente das pessoas a que era passado. O que é longe de ser ideal num momento de transformação organizacional. Assim, dei particular atenção à etapa de externalização, gravando, por exemplo, sessões utilizadas com o processo de socialização.

### 2.3.2. *Agile*

*Agile* é, embora exista discussão em redor da definição, uma metodologia/abordagem ao desenvolvimento de software. Para compreender melhor *agile*, é importante contextualizar o seu surgimento, em 2001. O *agile* surge em resposta às metodologias de desenvolvimento de software vigentes (metodologia *waterfall*) que pugnavam pelo desenvolvimento de um extenso conjunto de requisitos, após o qual a solução desenvolvida era entregue, em conjunto com a documentação, e embora pudesse cumprir o caderno de encargos, muitas vezes não satisfazia o cliente. Numa época em que a tecnologia evolui cada vez mais depressa e as necessidades das empresas podem mudar numa base diária para dar resposta a esta evolução, existia a necessidade de encontrar novas formas de trabalhar. Essa é a proposta do *agile*: oferecer, como o nome indica, agilidade no desenvolvimento de software. Embora o software tenha sido o foco da introdução do *agile*, esta metodologia tem dado frutos em diversas áreas. Um exemplo disto é o *Spotify* que adotou uma estrutura organizativa alicerçada nos princípios de *agile*.

O *agile per se* é um conceito relativamente simples. Pode ser descrito como uma abordagem iterativa à gestão de projeto, visando a satisfação dos clientes. O manifesto original é composto por 12 princípios, sendo que o primeiro faz alusão à importância de satisfazer o cliente privilegiando múltiplas entregas. Isto é materializado no conceito de MVP. Ou seja, é esperado que a equipa desenvolva um protótipo funcional em cada uma das iterações que executa. Sendo que é preferível uma entrega com menos funcionalidades, mas atempada, do que uma entrega atrasada. Existem diferentes metodologias de implementação de *agile*, sendo o SCRUM e o KANBAN as mais conhecidas. Estas metodologias respeitam os mesmos princípios orientadores, propondo formas de trabalhar distintas.

O SCRUM impõe uma cadência, ou seja, é definida uma periodicidade para a reunião da equipa. Cada uma destas reuniões tem o propósito de apresentar o que foi feito e para discutir se estão a ir por um caminho que gere o valor esperado para o cliente.

O KANBAN, por seu lado foca-se em maximizar a eficiência da equipa. Isto é feito com recurso a um quadro onde é feita a gestão das tarefas realizadas, em curso, pendentes e mesmo as descartadas. A máxima desta metodologia é a de limitar o trabalho em progresso.

É importante notar que as implementações de *agile* mais bem-sucedidas correspondem a abordagens híbridas que foram cuidadosamente desenhadas para se adequarem às equipas que as irão utilizar. No caso da SFS, isto verificou-se, mas por vezes faltou um pouco de robustez metodológica que ajuda a garantir o sucesso da implementação.

### 3. TECNOLOGIA E APLICAÇÕES

Neste capítulo, vão ser discutidas algumas soluções tecnológicas utilizadas no contexto da SFS. É importante compreender que as decisões ao nível da infraestrutura e arquitetura são do âmbito da direção de sistemas de informação que é responsável pelo desenho e manutenção deste tipo de soluções. A equipa de dados adequa-se à infraestrutura disponibilizada e, embora possa discutir alguns detalhes, não tem poder de decisão sobre a estratégia de longo e médio prazo, uma vez que esta está integrada numa visão holística de desenvolvimento de médio prazo. Dentro desta estratégia, está a migração de tudo o que são soluções *on premises* para soluções com recurso a *cloud computing*. Esta estratégia vai mais longe ao sugerir que, sempre que seja possível, sejam feitas implementações em *PaaS* ou *SaaS*, ao invés de se recorrer às máquinas virtuais em *IaaS*.

Vão ser, de seguida, introduzidas algumas das aplicações utilizadas e vai ser elaborado o contexto em que surgem na SFS. Os tópicos discutidos não pretendem ser compreensivos, no sentido de especificar a totalidade dos assuntos de relevo relacionados com a aplicação, mas pretendem discutir assuntos que tenham surgido e requerido investigação no contexto do estágio.

#### 3.1. MICROSOFT SQL SERVER

O *Microsoft SQL Server* é um sistema de gestão de base de dados relacionais. Este software é desenvolvido sobre T-SQL que é uma implementação de SQL com um conjunto de capacidades desenvolvidas e detidas pela *Microsoft*. O componente principal desta aplicação é o motor de base de dados que é responsável pelo armazenamento, processamento e segurança.

No caso da SFS, esta aplicação existe em vários formatos. Aquele com que tive mais contacto foi o da instância dedicada ao DW, e que corresponde à instalação numa máquina virtual em *IaaS* da versão de 2017.

##### 3.1.1. SQL Server Analysis Services

A *Microsoft* define o SSAS como um motor de dados utilizado no suporte à decisão. Ou seja, consiste numa tecnologia, integrada no *Microsoft SQL Server*, e que permite, entre outras coisas, criar modelos semânticos. Estes modelos podem ser utilizados com recurso a duas tecnologias distintas, a dos cubos multidimensionais ou a dos cubos tabulares. No caso da SFS, a tecnologia utilizada é a do modelo tabular e, por isso mesmo, estes modelos foram designados como o Tabular Universo e Tabular de Crédito.

**Modelo Multidimensional** – Esta tecnologia data de 2005, tendo já atingido um estado de maturidade, e é conhecida como o cubo OLAP tradicional. De forma simples, esta

tecnologia organiza a informação em estruturas multidimensionais e os utilizadores acabam por aceder a células que correspondem a uma determinada coordenada desse cubo. Para além da vantagem da maturidade de tecnologia, consegue lidar com volumes muito consideráveis de informação e permite criar métricas em MDX mais complexas que as da alternativa. Contudo, é uma tecnologia que não tem tido atualizações significativas e é complexa de implementar.

**Modelo Tabular** – Os cubos tabular também são conhecidos como “Cubos em Memória”, uma vez que estes modelos se materializam em bases de dados em memória. Utiliza um motor diferente (xVelocity) e é capaz de um acesso rápido à informação devido ao armazenamento por colunas. Uma das desvantagens deste modelo são os requisitos exigentes ao nível de processamento e memória. Concluindo, é um modelo relativamente simples de desenvolver e, na maioria dos casos, com um desempenho superior ao dos modelos multidimensionais. Para além disso, é uma tecnologia em desenvolvimento com potencial para melhorias e atualizações.

A adoção do modelo tabular faz sentido num contexto em que os volumes de informação são moderados e esta solução é mais ágil devido à facilidade de implementação e com um desempenho superior.

### 3.1.2. *Views* e TVF

No projeto da CRC, existe uma fase da implementação em que são utilizadas, primeiramente, *views* e depois são substituídas por TVF. Nesta secção, vamos discutir as diferenças entre estas duas funcionalidades do SQL.

**View** – Numa base de dados uma *view* é, como o nome intui, uma perspetiva sobre a informação presente nas tabelas. De forma simples é o resultado de uma *query* guardada e que pode ser utilizada como se existisse de forma persistente na base de dados.

**TVF** – *Table Valued-Function* é uma ferramenta mais complexa. Uma função tem, por definição, a capacidade de receber argumentos e devolver um resultado. Uma TVF é uma tabela não materializada, que pode ser consultada como uma tabela normal, mas com a característica de que o resultado que devolve depende do parâmetro introduzido. Esta funcionalidade é mais complexa do que uma *View*, mas traz níveis de complexidade adicionais que podem justificar a sua utilização.

Uma nota relevante é a de que uma *View* não materializada e uma TVF sem parâmetros são muito semelhantes.

### **3.2. EXCEL**

O *Excel* é uma ferramenta trivial. A particularidade da sua aplicação no contexto da SFS é como *Interface* primário de ligação com o modelo semântico. Esta ligação permite a utilização do modelo através de *pivot tables*. Esta é uma ferramenta consolidada, e com a qual a maioria dos decisores está familiarizada. O *Excel* também pode ser utilizado para carregar informação presente nas bases de dados. Uma vantagem do *Excel* é a de estar presente em todos os computadores da empresa e permitir o armazenamento de informação e o seu manuseamento num formato que não carece de transformações para ser compreendido. Não obstante, existem duas áreas em que o *Excel* não tem capacidade para dar resposta. Uma relaciona-se com os formatos dos campos. Muitas vezes o *Excel* não é capaz de interpretar o separador das casas decimais, requerendo transformações. Isto é especialmente danoso se estivermos a realizar análise de montantes. A segunda desvantagem está relacionada com a capacidade da ferramenta. Esta está limitada, à partida, a 1 048 576 linhas. O que, muitas vezes não é suficiente para armazenar o resultado de uma *query*. Para além disso, dependendo da capacidade das máquinas e da arquitetura do sistema operativo, podem existir mais ou menos constrangimentos na execução de um ficheiro .xls.

Concluindo, o *Excel* é uma ferramenta poderosa e que tem a capacidade de facilitar muitas operações diárias, mas antes de iniciar um projeto com recurso a esta aplicação, faz sentido questionar se será a mais adequada. Atualmente, o *Power BI*, muitas vezes, apresenta-se como uma solução mais robusta do que o *Excel* para a análise de dados.

### **3.3. POWER BI**

O *Power BI* é a proposta da *Microsoft* para BI. Esta ferramenta permite a manipulação e visualização de dados com o objetivo quer de monitorizar, como de procurar informação útil para a gestão do negócio. Embora esta ferramenta esteja a ganhar relevo nos ecossistemas de BI, no caso da SFS é utilizada esporadicamente, por iniciativa dos utilizadores. Os modelos semânticos e utilizados para disseminar informação beneficiam, em grande medida, da consulta através de uma interface como o PBI. O motivo disto é o de permitir, à semelhança do *Excel*, a consulta dos dados através de listas ou *pivot tables*, mas também permitir a integração de *scripts* ou de visualizações mais complexas. Um dos problemas encontrados com esta ferramenta foi a dificuldade em integrar a informação do modelo semântico e de uma fonte externa no mesmo documento.

### **3.4. API**

API é o acrónimo para *Application Programming Interface* e consiste num conjunto de regras e protocolos que visam permitir a integração de aplicações. Como a filosofia do sistema

*core* bancário é a de oferecer uma solução ágil e integrável, esta solução está assente em comunicações via API. Neste caso, as API permitem a conexão entre a estrutura da SFS e uma aplicação *SaaS*. No contexto do estágio, contribuí para o desenvolvimento de alguns processos que recorreram a API, especialmente os de criação de contas e contrato no âmbito da plataforma de *onboarding*, mas também tive que utilizar, com frequência, API para testar integrações e para criar casos de teste.

Ficou claro que, num ambiente em que as soluções consistem, cada vez mais, na integração de várias aplicações distintas, que as API têm um papel fundamental em assegurar a integração e a qualidade do produto final. Contudo, ficou claro, também, que utilizar API nem sempre é a melhor solução, uma vez que a sua implementação não está livre de problemas. Para processos em *real-time* ou *near-real-time* são, frequentemente a melhor solução. No entanto, para processos de grande volume, as API, não conseguem oferecer a mesma garantia e robustez de um ficheiro. Assim, é evidente, que as API vão desempenhar um papel cada vez mais importante, mas não deve ser descurada a avaliação cuidada de qual a melhor forma de estabelecer comunicação entre duas aplicações.

## 4. PROJETOS DE INTEGRAÇÃO DE DADOS

Os projetos de integração de dados são a essência do propósito da equipa em que o estágio esteve inserido. As soluções tecnológicas em que o Cartão Universo está assente produzem um manancial de informação com necessidades variadas. Estas necessidades podem ser divididas em dois eixos principais, designadamente periodicidade e criticidade. Existem fluxos em tempo real e outros com periodicidades diárias ou mensais. Ao nível da criticidade, existem fluxos com criticidade muito alta, e dos quais o negócio depende para funcionar e corresponder às expectativas dos clientes. Existem outros fluxos cujo propósito é o de monitorização e *reporting* e cuja criticidade não é tão alta. Esta descrição pode intuir que os fluxos com menor latência são os mais críticos, mas uma menor velocidade de resposta não é condição suficiente para uma maior criticidade. Os projetos apresentados nesta secção são os de maior dimensão e de maior relevância para o contexto do estágio. Contudo, existiu, numa base diária, um acompanhamento dos diversos processos de integração de dados. Com o intuito de melhorar a performance dos processos, foram realizadas pequenas alterações – *debugging*, resolução de erros nas execuções e correção de problemas ou incoerências nos dados detetados pelos utilizadores.

### 4.1. INTEGRAÇÃO DE DADOS EM DW

O primeiro projeto que me foi atribuído no contexto do estágio foi o de documentar e explorar o modelo de dados do DW. As primeiras duas semanas do estágio foram inteiramente dedicadas ao desenvolvimento deste projeto, tendo continuado posteriormente a dar suporte e a acompanhar diariamente o seu progresso. O objetivo foi o de me familiarizar, quer com o modelo de dados existente, quer com os fluxos de dados que o alimentam. Uma das primeiras realidades que interiorizei, e que diferiu do ambiente de sala de aula, foi a existência de dois ambientes completamente segregados, um de desenvolvimento e outro de produção. Também ficou claro para mim que sincronizar estes ambientes e garantir que o ambiente de desenvolvimento tem dados atualizados, coerentes e com qualidade não é tão simples quanto possa parecer. Efetivamente, a falta de sincronização entre estes ambientes constituiu um obstáculo ao longo do estágio. Um exemplo simples e representativo desta situação, foi a tentativa de desenvolver um campo que utilizava outro campo transformado referente ao tipo de emprego que um cliente tinha. A dificuldade surge quando concretizo que o campo não constava no ambiente de desenvolvimento. Assim, para seguir as boas práticas de desenvolvimento, foi preciso encontrar o processo e passá-lo do ambiente produtivo para o ambiente de desenvolvimento. Foi ainda necessário atualizar as tabelas para que depois o processo fosse desenvolvido. Este exemplo pode parecer algo irrelevante, mas ilustra diferenças abissais entre o que se aprende no ambiente académico e o que é necessário colocar em prática no contexto empresarial. Outro ponto relevante relacionado com a integração de dados dentro do contexto da SFS está relacionado com a sensibilidade dos

próprios dados. Os dados de que dispúnhamos eram, na sua maioria, dados pessoais e sensíveis.

Tendo em conta que a segurança dos dados é crítica, o envolvimento do DPO da SFS é fulcral para assegurar que o acesso aos dados é garantido apenas a quem depende deste para desempenhar a sua função na empresa com sucesso.

O *Data Warehouse* da SFS está alojado numa máquina virtual *PaaS* e utiliza, como motor de base de dados, SQL Server 2017. Os fluxos de dados que vão ser descritos recorrem a *azure functions*. O estágio não incidiu sobre questões de arquitetura da solução, uma vez que essa competência é da equipa de TI.

O DW está dividido em vários *Data Marts*, e cada um destes foi desenvolvido seguindo um modelo relacional otimizado para consulta.

## Fluxos

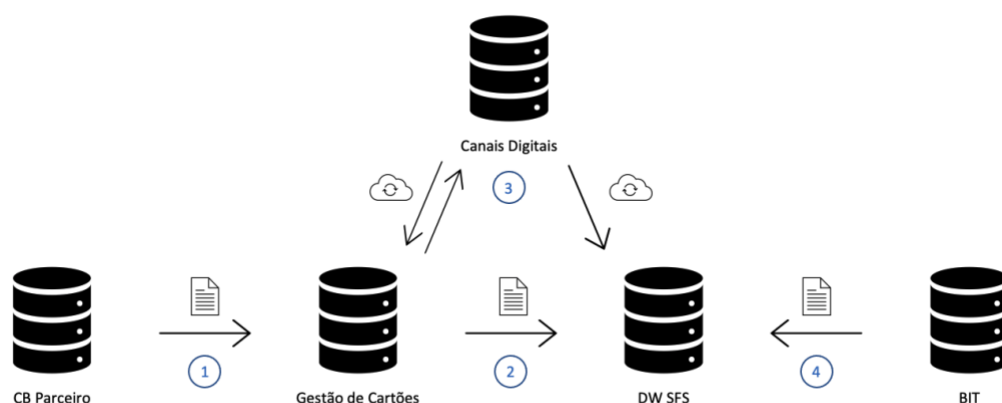


Figura 5 - Esquematização dos fluxos de dados

Como pode ser visto na figura 5, existem três fontes principais de dados: canais digitais, AGC (plataforma de gestão de cartões) e a BIT (empresa do grupo SONAE). Vamos agora explicar em que consiste cada um destes fluxos. Isto é especialmente importante para compreender o projeto seguinte, no qual foram feitas alterações significativas à origem dos dados.

1. CB Parceiro → Gestão de cartões: Este fluxo não alimenta diretamente o DW, mas contém informação essencial ao negócio. O parceiro de crédito gera ficheiros a partir do seu CB. Estes ficheiros contêm informação relativamente a juros, comissões e impostos. Estas transações são integradas e processadas pela ferramenta de gestão de cartões, acabando por serem integradas em DW como se se tratasse das transações efetuadas pelos clientes. Existe um ficheiro que sincroniza o saldo disponível no CB com o que existe na aplicação de gestão

de cartões. Este fluxo também é responsável pela comunicação de novos contratos e clientes e respetivas alterações.

Gestão de cartões → CB Parceiro: A aplicação de gestão de cartões envia, logicamente, para o parceiro de crédito, informação relativa às transações e pagamentos que os clientes efetuam com os cartões.

2. A informação acima descrita, originária do parceiro de crédito, é integrada em DW via AGC. São, numa base diária, gerados ficheiros com a informação existente na AGC. Estes ficheiros são integrados por um processo de ETL, em que a informação é primeiramente extraída para uma área de *staging*. A partir daí, a informação é integrada nas tabelas de destino, onde existem conversões de formato, criação de campos transformados e a descodificação de campos com a utilização de dicionários próprios. Um exemplo disto é a informação de género, que vem com a primeira letra do francês F – *Femme*, H – *Homme*.
3. Os canais digitais consistem na aplicação móvel, na *webapp* de gestão da conta e no portal do agente (*webapp* do *call center* para gestão das contas). O fluxo mais importante é o dos canais digitais para a AGC. Este fluxo funciona em tempo real e permite que as alterações feitas através das diferentes *interfaces* tenham efeito imediato. Exemplos disto são a alteração da percentagem de pagamento ou a alteração da modalidade de pagamento de uma transação. A alteração mais recente neste fluxo permitiu que os pedidos de novos cartões fossem processados em tempo real. Isto é especialmente importante no caso da encomenda de cartão por furto. Assim, não só o cartão é encomendado em tempo real, como o cartão antigo é bloqueado de imediato. O fluxo contrário, AGC → Canais Digitais, permite que o cliente veja o saldo disponível em tempo real. Existe, ainda, um fluxo de integração direto dos canais digitais para DW. Este fluxo é responsável por integrar informação como os cliques dos utilizadores, os logins, e o seu status digital. Esta informação é importante para otimizar os serviços disponibilizados, tanto numa perspetiva tecnológica como de negócio. Também é importante poder compreender a relação que os clientes têm com os canais digitais e a correlação desta com a utilização do cartão. Este conhecimento é utilizado, atualmente, na gestão e monitorização da carteira.

Percebemos, assim, que a aplicação de gestão de cartões se encontra no centro da operação e da integração de dados em DW. Isto faz sentido, uma vez que, nesta altura, a Sonae FS era responsável pela gestão de cartões, dependendo de um parceiro para a gestão de crédito.

4. A BIT é uma empresa de tecnologias de informação do grupo SONAE, sendo responsável por grande parte das aplicações da Sonae MC. Existem dois fluxos

entre a Sonae FS e a BIT. SONAE FS → BIT: é enviado, diariamente, um ficheiro com as transações elegíveis para *cashback* em Cartão Continente (uma das vantagens do produto).

BIT → SFS: Diariamente, a BIT, envia um ficheiro com a resposta aos pedidos de atribuição de *cashback*. Este pedido pode ser aceite ou recusado. A BIT envia, ainda, com a mesma periodicidade, um ficheiro com a informação das lojas do grupo SONAE. Este ficheiro é integrado em DW e complementa a dimensão dos comerciantes.

### Modelo Semântico – Tabular Universo

Apesar de, no DW, os dados serem integrados passando por um rigoroso processo de *data quality* e o modelo ser otimizado para consulta, a maior parte dos utilizadores consulta os dados recorrendo a um modelo semântico. Como foi discutido *supra* (secção 2), a utilização de um modelo deste género facilita o acesso à informação, uma vez que um utilizador sem proficiência em SQL consegue aceder a dados complexos e aplicar filtros compostos, sem se preocupar com *joins*. Permite, ainda, otimizar tempos de consulta e garantir o acesso a métricas complexas que, de outra forma, teriam de ser calculadas de forma esporádica, dando menos confiança nos resultados ao utilizador. Em suma, o acesso à informação foi centralizado no modelo de dados. Este modelo tenta espelhar a lógica do negócio respondendo às perguntas que os utilizadores possam colocar. Isto reduz o risco de recálculos incorretos e garante que a informação é transversal à empresa. Outra funcionalidade que este modelo oferece é a alteração dos nomes de colunas e métricas para nomes mais fáceis de compreender. No caso da SFS, é utilizado o modelo tabular, como discutido na secção da tecnologia.

Este modelo recorre a um sistema de partições para otimizar o reproprocessamento, evitando assim que, diariamente, sejam feitos cálculos redundantes. Existem partições mensais e anuais. Para as tabelas factuais, diariamente, são recalculados os últimos dois meses de calendário. Contudo, este processo é lento, pelo que estão a ser discutidas alternativas mais rápidas, como o recálculo apenas do último mês. A lógica para a existência de dois meses está relacionada com atrasos que existem na integração da informação. Mas à medida que os sistemas evoluem para integrações em tempo real, isto deixará de ser necessário. Ainda para as factuais, cada vez que um mês termina, o modelo reprocessa o ano corrente. As dimensões, por seu lado, são reproprocessadas diariamente.

Apesar da existência de partições, o reproprocessamento do modelo tabular é um dos processos mais pesados da SFS. Assim, em conjunto com a equipa de desenvolvimento, foi analisada e discutida a possibilidade de migrar este serviço de *Analysis Services* para *Microsoft Synapse*. Esta solução é cara e implica constrangimentos, para os quais, neste momento, não existe capacidade de dar resposta. Contudo, com o alargamento da atividade de crédito e com

uma utilização crescente e mais exigente do modelo semântico, isto pode vir a concretizar-se num futuro não muito distante.

Um aspeto que foi notório foi a fraca disseminação da utilização de *PowerBI* na empresa. Esta ferramenta é raramente usada para *reports* não orientados ao negócio. Ao longo do estágio, foi-me incumbida a tarefa de promover a utilização desta poderosa ferramenta que está especialmente apta para trabalhar com o modelo tabular.

Houve muita resistência para a utilização desta ferramenta, por motivos diversos – os *reports* já estavam montados e não havia capacidade para os alterar para PBI; os colaboradores sentiam-se mais confortáveis a trabalhar em *Excel*, uma ferramenta em que já têm experiência; o uso de múltiplas fontes no mesmo documento, o que não é possível com PBI quando se utiliza uma *live connection*.

Fui, ainda, encarregue de apoiar os utilizadores das diferentes equipas nas suas necessidades de informação. Esta responsabilidade incluiu diversas tarefas, das quais destaco manter-me a par da evolução do modelo de dados para poder esclarecer os utilizadores sobre onde poderiam encontrar a informação. Também surgiu a necessidade de, por vezes, esclarecer o significado dos dados e de criar novas métricas no tabular ou integrar informação adicional. É importante dar nota de que nem toda a informação que está em DW está no modelo tabular. Por vezes, o problema do utilizador ficava resolvido com uma *pivot table*, mas houve problemas que exigiram *queries ad hoc* e a entrega da tabela resultante.

Um ponto interessante é a forma como as *slow changing dimensions* são diferentes no contexto desta empresa. No contexto académico, foi reforçado que deveríamos adotar um dos tipos, assim como deveríamos evitar ter chaves redundantes. Contudo, pude verificar que, por diversos motivos, estas “boas práticas” não são seguidas, não por desleixo, mas para facilitar a utilização do DW.

## 4.2. INTEGRAÇÃO DE DADOS DO NOVO MODELO DE NEGÓCIO

Com o alargamento do objeto social, e o alargamento da atividade da SFS suportado por um CB, existiu a necessidade de atualizar os fluxos que alimentam o DW. Este processo foi, em certa medida, simplificado pelo papel central da aplicação de gestão de cartões e cuja arquitetura não sofreu alterações. A prioridade foi, aquando da transição, garantir que seria possível integrar a informação do CB em DW. Esta informação é crucial para a monitorização e acompanhamento do negócio na sua nova vertente, bem como auxiliar a tomada de decisão. Existia também a necessidade de ter esta informação disponível para a elaboração de reportes regulamentares exigidos pelo regulador e pela autoridade tributária.

### Novos fluxos de dados

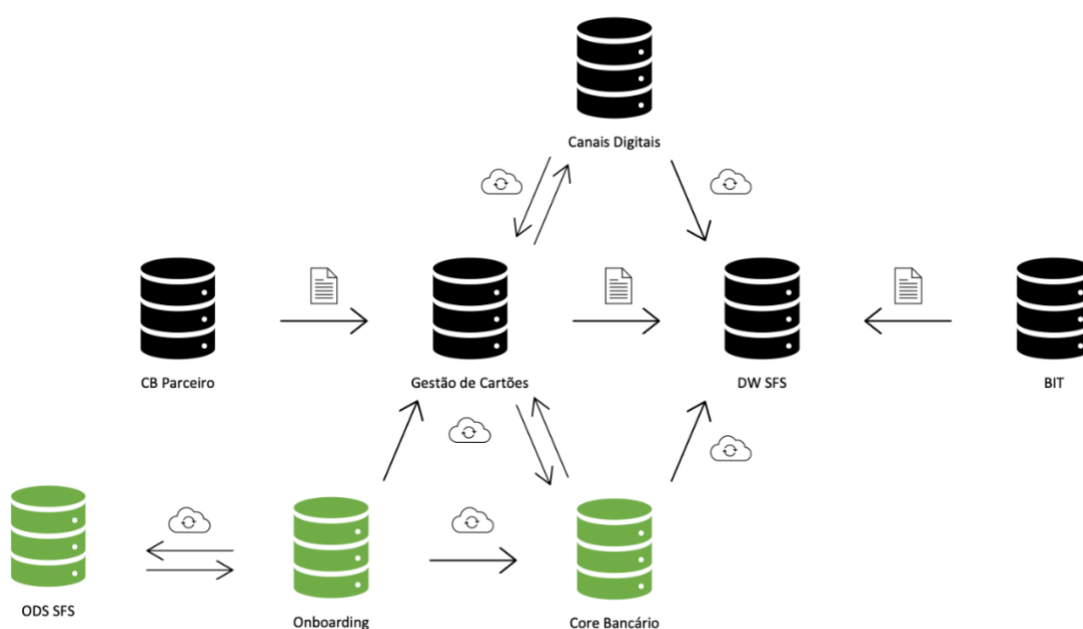


Figura 6 - Esquematisação dos novos fluxos de dados

A figura 6 tem duas grandes alterações relativamente à figura 5: a inclusão do CB e a inclusão da plataforma de *onboarding*.

Um ponto que é importante esclarecer é o da manutenção do fluxo parceiro de crédito → AGC. Aquando da transição, as linhas de crédito foram migradas, mas os clientes que tinham dívida com o parceiro mantiveram essa obrigação. Assim, para esses clientes e para efeitos legais, regulamentares e de extrato, para essas dívidas, manteve-se o fluxo. Este fluxo continua a afetar o saldo disponível; contudo, a SFS agora tem a possibilidade de intervir, o que no modelo anterior não acontecia. O ficheiro, que diariamente atualizava o balanço dos clientes para que este fizesse *match* com o do parceiro de crédito deixou de ser integrado. Agora, o valor que o cliente deve é o que está no *core* bancário.

Analisando os fluxos novos, temos:

CB → DW: Este fluxo é responsável pela integração dos dados do *core* bancário em DW. Integra os dados do *core* bancário. Não só foi a principal alteração como foi o processo que requereu mais esforço. Por ser um projeto novo, por serem novos conceitos funcionais e de negócio e por limitações impostas pela solução tecnológica. Por diversas vezes existiu a necessidade de consultar documentação e falar com as equipas do vendedor da solução para perceber o significado de um campo ou tabela. No que concerne à integração *per se*, existiam duas formas de integrar, nomeadamente a possibilidade de fazer uma cópia das tabelas transacionais e integrar esta informação em DW, ou, alternativamente, utilizar as REST API. Optámos pela segunda opção, uma vez que a aplicação está orientada para a utilização de API. A primeira opção não nos fez muito sentido, dado que as tabelas da aplicação estão otimizadas para um sistema transacional e exigiria um esforço grande de modelação e de desenvolvimento do ETL para que os dados pudessem ser facilmente consultados. Estas transformações comportariam risco tanto ao nível tecnológico como a nível funcional. Também pesou na nossa decisão o facto de esta aplicação estar em constante atualização e, caso sejam feitas alterações às tabelas exigiria um esforço maior para garantir a consistência da integração. Assim, foi construído um processo que, diariamente, após o processamento de fecho do dia da aplicação, chama as API que devolvem transações, *credit arrangements*, *groups*, e *loans*. A resposta é um ficheiro JSON que, posteriormente é integrado em DW, com recurso a um processo de ETL próprio, em tabelas que respeitam a estrutura da resposta.

Por último, importa referir que, ao longo do processo de integração dos dados de crédito, foram discutidas alternativas que garantissem segurança e controlo de acessos. A primeira solução foi segregar esta informação num *schema* de crédito, mas depressa ficou evidente que esta solução não conferia a flexibilidade no controlo e gestão de acessos e perfis que desejávamos. Assim, a solução passou pela criação de uma nova base de dados – *Credit DM* – que alberga a informação integrada pelo core bancário, da CRC e de imparidade e *scoring*.

AGC → CB: Este fluxo, funcionalmente é semelhante ao que existia da AGC para o parceiro de crédito. Embora, numa primeira fase, tenham sido feitas adaptações aos ficheiros para que possam ser consumidos pelo CB, a ideia é que num futuro próximo se evolua para o envio de transações em tempo real via API.

CB → AGC: Este fluxo vem substituir o fluxo Parceiro de Crédito → AGC e envia, para o sistema de gestão de cartões, transações relacionadas com juros, comissões e impostos.

Foi também desenvolvida uma plataforma de novas adesões (*onboarding*) para substituir a aplicação do parceiro, que era utilizada nas lojas. Esta plataforma comunica com a AGC, criando os novos contratos e associando aos cartões. Com o CB criando os contratos e os clientes. E com o ODS, registando a informação dos processos de adesão. Não comunica

diretamente com o DW, uma vez que a informação é integrada via AGC. Contudo, a perspectiva de futuro é que o ODS venha a ser o novo *pivot* para a informação de contratos e clientes.

#### Atualização do modelo semântico – Tabular Universo

O facto de existir uma camada semântica entre o DW e os utilizadores finais permitiu que estas alterações não causassem disrupção. Outra vantagem foi o facto de não ter sido necessário atualizar os *reports* e trabalho em *Excel* que os utilizadores tinham desenvolvido. Isto, de certa forma, vindicou e confirmou a importância da inclusão deste tipo de camadas numa solução de *Business Intelligence*.

Este projeto trouxe-me três aprendizagens excecionais, uma de teor tecnológico, uma a nível de conceitos de crédito, e outra ao nível de arquiteturas de BI. A utilização de API foi uma novidade. Tive a oportunidade de aprender os conceitos basilares desta tecnologia, utilizá-la para resolver problemas de negócio e integrá-la num sistema complexo com recurso ao APIM. Mas também foi importante compreender as suas limitações, ficou claro, que independentemente do quão recente uma tecnologia é ou do *hype* que existe à sua volta, é importante perceber se se adequa ao propósito daquilo que estamos a tentar alcançar. Durante o estágio, discutiu-se várias vezes se determinado fluxo deveria ser desenvolvido por ficheiro ou por API. E, existiu uma tendência, de optar por API, por ser a solução mais moderna, mesmo para processos que não o requeriam. Isto traduziu-se numa miríade de problemas e *quick fixes*. A aprendizagem ao nível dos conceitos de crédito complementou outros projetos. Não obstante, para montar esta integração foi preciso um conhecimento profundo do produto e compreender em detalhe, por exemplo, a diferença entre uma transação ou um montante retido. Por fim, estes processos ensinaram-me alguns princípios de arquitetura, especialmente por causa da ênfase que foi dada à possibilidade de utilizar o novo ODS como base de dados central da arquitetura.

#### 4.3. NOVO MODELO SEMÂNTICO – TABULAR DE CRÉDITO

Com a integração de nova informação surgiu a necessidade de a disponibilizar aos utilizadores. Foram ponderadas duas hipóteses: a inclusão desta informação no modelo semântico já em utilização ou a construção de um novo modelo. Optou-se pela utilização de um modelo especificamente dedicado a esse propósito por vários motivos, nomeadamente o facto de a informação residir numa base de dados própria; a informação ser sensível e limitada a um número restrito de utilizadores; e oferecer uma solução com menos ruído. Contudo, é expectável que esta solução venha a evoluir, passando a incluir mais informação. Neste momento está disponível a informação de contabilidade, de crédito, de imparidade e de *scoring*. O passo seguinte será incluir informação da disseminação da CRC que está a ser

integrada na mesma base de dados e é importante para fazer análises sobre a delinquência da carteira de clientes.

Este modelo foi construído em *SQL Server Analysis Services*, à semelhança daquilo que já existia e está enraizado na instituição. Optou-se por um modelo tabular uma vez que é uma tecnologia mais eficiente e com resultados comprovados dentro e fora da SFS.

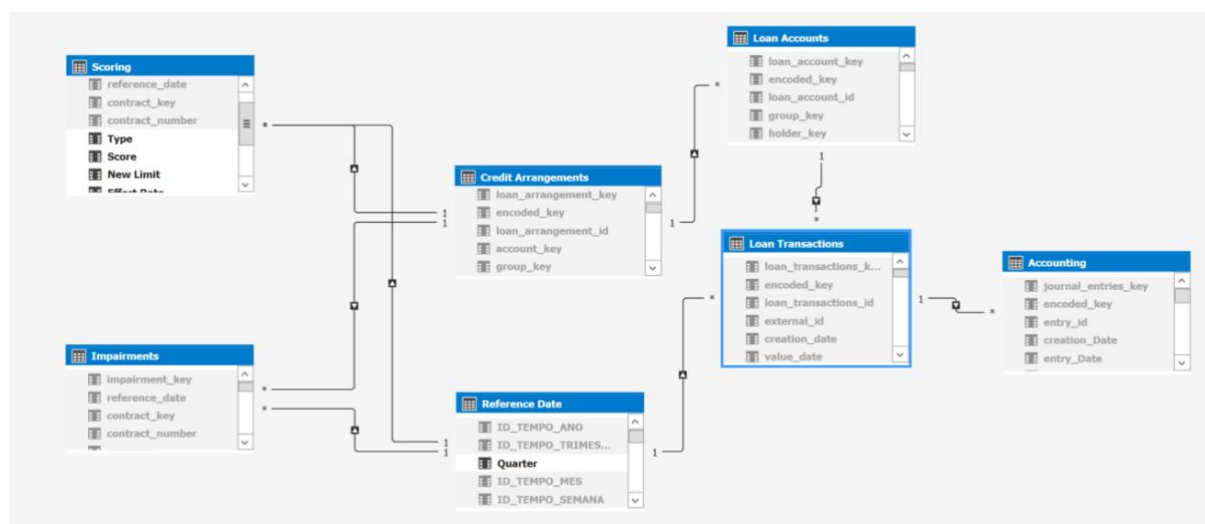


Figura 7 - Modelo tabular de crédito

Atualmente, esta solução é utilizada pelas equipas de crédito, risco e contabilidade. O seu desenho simples com um número reduzido de métricas está alinhado com a estratégia que desenvolvi para a qualidade dos dados e para a disseminação de PBI dentro da organização. O modelo tabular operacional tem muita informação de legado e até alguma redundante. Ao ter existido a hipótese de construir um modelo do zero, quisemos dar ao utilizador uma solução que disponibilizasse a informação necessária de forma palpável e simples, não colocando entraves a uma navegação fluída. Com esta solução, desenvolvi também uma metodologia para o crescimento da mesma, que pretende evitar que este modelo evolua com base em pedidos *ad hoc* e acabe por padecer dos mesmos defeitos apontados ao modelo operacional. Um passo importante definido nessa metodologia é a criação de partições que ainda não tinha sido feita.

Finalmente, importa referir que este modelo é de particular importância para garantir que a informação é consistente de forma transversal à organização. Antes da transformação do negócio, equipas diferentes utilizavam abordagens diferentes para os seus cálculos. Dava-se até o caso de diferentes equipas terem informação diferente relativa ao número de clientes. Da qualidade e relevância deste novo tabular, está dependente a unificação e simplificação do acesso à informação nesta empresa. Isto é de especial importância quando se fala de crédito, porque não é aceitável que diferentes equipas tenham informação diferente relativamente ao crédito concedido pela instituição.

## 5. PROJETOS DE REPORTING

Esta secção é dedicada em grande parte ao projeto de reporte para a CRC. Este foi o projeto principal do estágio onde estive envolvido como analista e *developer*. Existiram outras iniciativas *ad hoc* de *reporting*, mas cuja falta de robustez metodológica faz com que não sejam incluídas neste relatório. Contudo, é descrito um episódio – *dashboard* de adesões – pelas aprendizagens que trouxe para mim e para a organização e pela sua criticidade para o negócio.

### 5.1. CENTRAL DE RESPONSABILIDADES DE CRÉDITO

A CRC é uma base de dados do Banco de Portugal que agrega, num único sítio, as responsabilidades de crédito que cada indivíduo tem com instituições que concedem crédito em Portugal. O seu principal objetivo é o de auxiliar as entidades participantes na sua decisão de concessão de crédito. Isto é conseguido ao disponibilizar informação agregada sobre os requerentes de crédito. O alargamento do objeto social da SFS de IME para entidade concedente de crédito, implicou que, para cumprir com as disposições regulamentares do Sistema de Centralização de Responsabilidades de Crédito, irá interagir com o BdP, contribuindo com informação relativa ao seu universo de clientes.

Dentro da nova realidade da empresa, a sua relação com a CRC tem três vertentes, designadamente a de reporte, a de integração da disseminação, e a de recolha de dados via API para a avaliação do risco de crédito. O estágio colocou ênfase nas duas primeiras, mas também foi dado apoio funcional na implementação da terceira vertente.

Este projeto foi uma implementação *end-to-end*, desde o levantamento de requisitos à implementação da solução tecnológica. Para a implementação deste projeto, optou-se por incluir um parceiro. Este parceiro garantiu o conhecimento dos requisitos de reporte, algo que de outra forma não teria sido adquirido em tempo útil. A minha entrada veio aumentar a capacidade da SFS para dar resposta a este desafio, acabando por me tornar no ponto de contacto para o projeto e, pelo acompanhamento próximo, a pessoa com mais conhecimento do mesmo. Aquando da minha saída, isto colocou desafios ao nível da passagem de conhecimento, o que foi uma oportunidade para pôr em prática algumas das metodologias de passagem de gestão de conhecimento lecionadas no mestrado.

A criação e desenvolvimento de um modelo de reporte da CRC contempla diferentes requisitos de informação impostos pela regulação em vigor, obrigando, assim, as instituições financeiras a dispor de uma estrutura de dados orientada à especificidade deste reporte. No caso da implementação no contexto de transição, estas dificuldades são acentuadas por uma necessidade de estar preparado para vários cenários. Um exemplo deste constrangimento foi a necessidade de preparar a solução para reportar um instrumento específico que acabou por não ser disponibilizado pelo portfólio da SFS.

A primeira fase deste projeto consistiu no levantamento funcional da estrutura do reporte. Diferentes tipos de crédito estão abrangidos por diferentes requisitos regulamentares. Assim, foi importante perceber que *reports* teriam que ser enviados e os campos cujo envio é obrigatório. Depois de esta lista nos ter sido entregue pelo parceiro, coube-me avaliar de forma crítica o documento que nos havia sido entregue, desafiando alguns dos pressupostos utilizados na interpretação dos regulamentos do BdP. Tendo fechado a lista dos requisitos dos reportes no âmbito da CRC, entrámos na fase de mapeamento. Alguns dos campos puderam ser mapeados diretamente com informação disponível em DW. Estes não careceram de muito esforço. Contudo, como esperado, a grande maioria dos campos não estava disponível em DW como mapeamento direto. Para estes, deparei-me com quatro situações distintas:

1. Campos que existiam em DW, mas que não estavam no formato pedido. Não ofereceram um grande desafio, uma vez que *queries* em SQL nos permitem fazer transformações sem precisar de materializar os dados. Um exemplo disto foram as datas como é o caso da data de celebração do contrato. Em DW, esta encontra-se no formato AAAAMMDD e deveria ser reportada no formato DD-MM-AAAA.
2. Campos que necessitavam de ser transformados seguindo dicionários próprios (Tabelas de Domínio). Para estes existiu a necessidade de confrontar as tabelas de domínio com os dicionários da SFS. Existiram alguns casos simples, como o género dos clientes, em que os dicionários coincidiam. Contudo, para casos mais complexos existiu a necessidade de agregar várias entradas da SFS numa da CRC ou utilizar campos auxiliares para criar uma condicional satisfatória. Existiram também campos, que, embora seguissem as tabelas de domínio, assumiam sempre o mesmo valor, não constituindo dificuldade.
3. Campos que necessitaram de calculatória. Um exemplo disto foi o campo “Montante potencial revogável”, que consiste no valor disponível na linha de crédito. Assim teve que ser feita a diferença entre o limite de crédito do contrato e a agregação das várias *loan accounts*.
4. Campos não existiam em DW. Para esses, teve de ser feito um levantamento da existência dos campos nas aplicações e, sendo possível, integrá-los. Para outros, como o caso dos campos relacionados com imparidade ou de incumprimento. Estes dois projetos são discutidos em secções seguintes.

De seguida, passámos para a fase da implementação tecnológica. A arquitetura da solução pode ser vista no diagrama seguinte. A minha responsabilidade foi a de disponibilizar os dados de forma que pudessem ser consumidos pela aplicação. Como este projeto foi feito em *agile*, para garantir o alinhamento entre a SFS e o parceiro, foram marcadas sessões de SCRUM semanais. Na primeira sessão, foi-nos solicitado que disponibilizássemos os dados seguindo uma estrutura predeterminada e que correspondia à estrutura das tabelas de

*staging* que eram utilizadas noutros clientes. Refutei esse pedido argumentando que, numa ótica de garantir flexibilidade no futuro e facilitar o suporte a este processo, seria mais simples se a nossa disponibilização de dados fosse orientada à comunicação. Acabámos por seguir esta via, criando uma *view* por comunicação ou mais nos casos em que existiam relações de muitos para muitos.

A primeira solução encontrada para disponibilizar os dados foi a criação de *views*. Mas estas manifestaram-se insuficientes uma vez que não têm a capacidade de receber um parâmetro – devolviam apenas os dados com a *flag record active* no momento da execução. Como estes reportes são, geralmente, enviados para datas anteriores à data da execução, era imperativo que conseguíssemos disponibilizar os dados para uma data em particular. Acabámos por adaptar as *views* que tinham sido criadas para *Table Valued-Functions* que retornavam os dados ativos à data do parâmetro recebido. Esta solução serviu o propósito, contudo, de futuro, existe a possibilidade de se optar por materializar os dados a ser disponibilizados para esta aplicação, numa tabela própria. O propósito disto será o de melhorar a performance e robustez do processo.

Embora, a solução tenha sido entregue atempadamente, o processo foi turbulento, uma vez que não existia um responsável claro, nem um caderno de encargos. Como é possível prever, existiram dificuldades em atribuir responsabilidades e surgiram fricções.

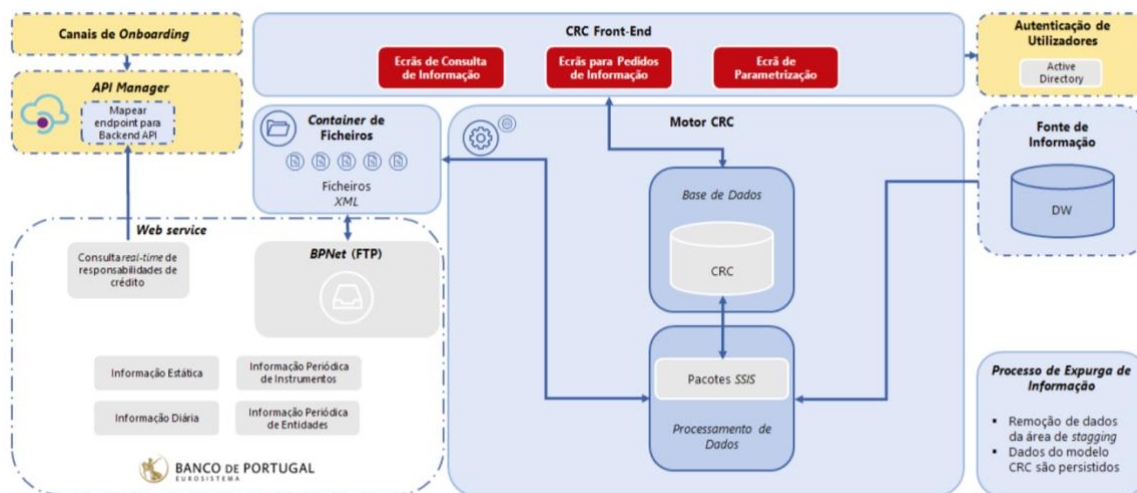


Figura 8 - Arquitetura da solução de reporte à CRC

Depois da fase de desenvolvimento, com uma duração aproximada de dois meses, entrámos na fase de testes, que se prolongou por três semanas. Na primeira semana, tentámos preparar os casos de teste ao nível do DW. O caderno de teste incluía 23 casos de teste contemplando, cada um, três execuções do motor CRC. Como foi referido anteriormente, a integração do CB ocorre diariamente, assim, para termos dados espaçados

no tempo, estes tiveram que ser criados de forma artificial, garantido todas as consistências funcionais. Foi uma das tarefas mais desafiantes do estágio, fazendo-me compreender a necessidade de antecipar os testes e ter clarividência sobre os testes a efetuar e a sua importância.

Durante a segunda semana, executámos os cenários de teste em conjunto com a equipa do parceiro. Esta semana teve uma complexidade e uma criticidade que nos levou a reunir no escritório com um horário bastante alargado, apesar das recomendações de teletrabalho.

A última semana serviu para analisar os resultados. Foram resolvidos alguns *bugs* simples e identificadas fragilidades a ter em conta aquando da passagem à produção.

A passagem à produção decorreu sem percalços e, no dia da transição, foi possível enviar os reportes necessários. A partir desse momento, começaram a surgir problemas ao nível da execução da aplicação. Como esta aplicação costumava ser implementada em máquina físicas, não estava otimizada para uma arquitetura *PaaS*. Isto levou a múltiplos *timeouts* e a inconsistências no modelo de dados da aplicação. Existiu a necessidade de fazer SCRUM diários durante cerca de dois meses para garantir que se resolviam todos os problemas identificados e aqueles que apareceram no seguimento de algumas correções. Aqui o meu papel foi o de fazer a ponte entre as equipas funcionais, as equipas de desenvolvimento e as equipas de negócio, e garantir que os conceitos e as necessidades eram compreendidos de forma transversal.

Com a minha saída a aproximar-se, surgiu a necessidade de passar o conhecimento que vim a acumular ao longo do projeto. Percebi, em primeira mão, a falta de metodologia que existe, quer na transformação de conhecimento tácito para conhecimento explícito, quer para a passagem de conhecimento em geral. A abordagem que adotámos, foi a de fazer duas sessões iniciais gravadas, em que foram passados os conhecimentos funcionais e técnicos. Foi partilhada documentação escrita. E, por fim, convidámos os recipientes a fazer parte do SCRUM diário. Depois do SCRUM, fazíamos uma sessão de esclarecimento de dúvidas para que compreendessem efetivamente o que tinha sido discutido e compilávamos as dúvidas e esclarecimentos num documento adicional.

Paralelamente, foi desenvolvido um processo relativamente simples de integração da disseminação da CRC no *Data Warehouse*. A disseminação da CRC consiste na informação que é disponibilizada pelo BdP às entidades participantes relativamente aos seus clientes. Este processo, diariamente, consultava as tabelas da disseminação e, no caso de existirem registos novos, importava-os para tabelas com a mesma estrutura na base de dados de crédito. Contudo, o BdP envia correções aos ficheiros de disseminação que envia. A partir do momento em que estes ficheiros (DISC) começaram a ser integrados pela aplicação, o processo teve que ser adaptado. O motor da CRC, ao integrar as correções, torna os registos antigos obsoletos, alterando o seu estado de 1 para 2 e integra as novas linhas. Assim, o processo que tínhamos

montado anteriormente ganhou mais um nível de complexidade. Primeiro verifica, para os registos com data de alteração diferente da data de criação e para os últimos sete dias, se estes estão diferentes daquilo que existe em DW. Se estiverem, atualiza; caso não estejam, não procede com nenhuma alteração. Depois integra as linhas novas.

Concluindo, este foi um projeto que exigiu bastante de mim. Uma dificuldade que tive foi a de ter muita responsabilidade sem muita autonomia. Por vezes, deu-se o caso de ser responsável pela resolução de incidentes, mas sem a autoridade para decidir a solução. Para além deste aspeto, este projeto aprofundou o meu conhecimento relativo ao negócio do crédito e deu-me uma base sólida sobre os requisitos de reporte ao BdP, com particular incidência na CRC. Também tornou claro para mim a importância da qualidade dos dados para produzir reportes exigidos por reguladores, e permitiu o desenvolvimento da minha proficiência em SQL e em modelação de dados. Finalmente, aprendi muito sobre o ciclo de vida de um projeto, as diferentes fases que o constituem e a importância relativa de cada uma. Não menos importante, foi a consciencialização da necessidade de alinhamento dos diferentes *stakeholders* para a conclusão virtuosa do projeto.

## 5.2. DASHBOARD ADESÕES

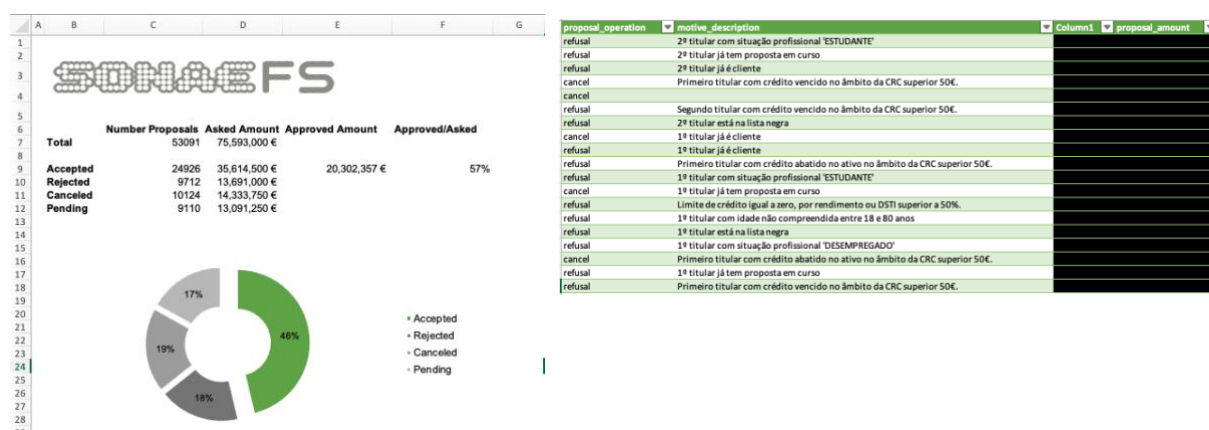


Figura 9 - Sample do dashboard de adesões

Este projeto, com a duração de apenas um dia, foi o mais curto dos projetos em que estive envolvido. Contudo, a sua inclusão neste relatório tem duas motivações. A primeira está relacionada com a multiplicidade de conceitos de Business Intelligence que foram postos em prática para a sua execução. A segunda é dar um exemplo da criticidade que o acesso a informação de forma atempada pode ter num negócio.

Na data de transição para o novo modelo de negócio, uma das aplicações que foi substituída foi o portal de adesões. Este portal é a ferramenta que permite, aos operadores nas lojas por todo o país, efetuar adesões e alterações aos contratos. Contudo, esta transição não esteve livre de percalços, uma vez que alguns dos equipamentos instalados para o efeito tiveram problemas de conectividade com a aplicação. Este problema foi identificado no início do dia, mas urgia a necessidade de, ao longo do dia e nos dias subsequentes, compreender em que estado estava a resolução do problema e monitorizar a capacidade das lojas para efetuar adesões. É importante salientar que cada adesão não feita por um problema de índole tecnológico implica, para além de uma perda de receita, dano reputacional para o Cartão Universo e para o ecossistema Sonae. Por este motivo, este incidente obteve prioridade máxima.

Saber quantas adesões estavam a ser feitas em cada loja pode parecer um processo simples, contudo, ainda não existiam fluxos de integração entre o ODS que suporta os canais digitais (incluem o portal de adesão) e o DW. Assim, quando solicitado pela direção de crédito para disponibilizar esta informação, o primeiro desafio foi localizar, no corrúpio de um dia caótico, a informação relativa às novas adesões. O segundo desafio foi conseguir criar um utilizador com acesso de leitura à tabela para poder desenvolver a integração para DW.

Esta integração foi feita com recurso a SSIS. O primeiro passo do processo é consultar uma tabela de *logs* para confirmar a data da última extração bem-sucedida e guardá-la como parâmetro. De seguida cria uma linha na tabela de *logs* com o *timestamp*. O passo seguinte é truncar as várias tabelas de *staging* e importar para as respetivas a informação com *date*

*modified* superior ao parâmetro supramencionado. O processo de importação para as tabelas de *staging* termina com a marcação do sucesso do processo na tabela de *logs*.

A partir destas tabelas de *staging* a informação é consolidada numa segunda tabela de *staging*. Neste processo são feitas várias conversões, em particular conversões de *Unicode string* to *string*. É então escrito na tabela de *logs* que o processo de transformação foi executado com sucesso. O último passo é o de integrar estas linhas na tabela de DW. Após ter sido executado, é também assinalado na tabela de *logs* e o processo é dado como concluído.

Atualmente este processo ainda está a correr num *job ad hoc*, mas na próxima revisão do ETL deverá ser incluído. O motivo por não ter sido incluído numa primeira instância prende-se com a necessidade que existiu de ter uma latência inferior à do ETL. Este *job* corre de dez em dez minutos, enquanto o ETL é executado 3 vezes por dia.

Concluída a integração dos dados para DW foi preciso desenvolver um suporte que permitisse que as equipas de negócio pudessem consumir com facilidade. Para isto foi desenvolvido um PBI, com uma versão *mobile*. Contudo, esta solução não foi bem aceite, uma vez que existia, por parte destas equipas, pouca familiaridade com a ferramenta. Foi então replicado, de forma mais simples, aquilo que tinha sido feito em PBI para *Excel*, utilizando Importação *via query*. Um dos desafios para esta implementação foi a compreensão desta informação, uma vez que existem muitas linhas para cada proposta. Dependendo da fase do processo em que está. Foi importante perceber o que era uma proposta aceite, foram encontrados casos anormais como propostas aceites com limites de crédito com casas decimais. Estes foram sintomas de problemas na aplicação que puderam ser corrigidos de forma célere. Embora tenham existido outras situações, este caso concreto mostra como o acesso a informação, através de ferramentas de visualização de BI, pode ter um impacto significativo na auscultação das soluções tecnológicas, para além da monitorização do negócio.

## 6. PROJETOS DE TROCA DE INFORMAÇÃO

Para além da gestão do DW, uma das responsabilidades principais da equipa Core & Data Solutions é a de desenvolver, implementar, monitorizar e melhorar os processos de fluxo de dados. Com a transformação, como já foi demonstrado anteriormente, existiu a necessidade de desenvolver novos fluxos, e adaptar alguns dos existentes. Nesta secção, vão ser explorados cinco projetos deste âmbito sendo que um deles, até ao momento não foi implementado. Contudo, é de extrema importância discutir também estes projetos, uma vez que são uma fonte de diversas aprendizagens.

### 6.1. RECUPERAÇÕES DE CRÉDITO

Com a inclusão da concessão de crédito nas competências da SFS, surgiu a necessidade de montar processos para as recuperações de crédito concedido a clientes incumpridores. Estes processos estarão, inevitavelmente, assentes em tecnologias de informação. Quando começámos este processo, não existia clareza, por parte do negócio, sobre como é que as recuperações iriam funcionar. Assim, pensámos a solução, numa lógica de antecipação, de modo a que se cingisse aquilo que eram factos, e não dependessem de processos de negócio. Esta plataforma iria sinalizar os clientes em incumprimento e partilhar, numa base diária, as posições da dívida em atraso.

Tecnologicamente, optámos por utilizar tabelas de base de dados num *schema* próprio da Universo DM. Foi criada uma tabela principal com o objetivo de manter o histórico, de forma acessível e fácil leitura, dos incumprimentos de cada um dos contratos/clientes. Todos os dias, corria um processo SSIS que comparava os contratos com montantes vencidos aos que se encontravam na tabela sem data de fim de incumprimento. Para os que se mantinham em incumprimento, não era feita nenhuma ação; para os que estavam em incumprimento e não tinham registo ativo (sem data de fim), era criada uma nova linha com essa data como data de início de incumprimento. Para os que estavam na tabela, mas não eram encontrados montantes vencidos, era marcada essa data como data de fim de incumprimento.

Diariamente, para os contratos em incumprimento, populávamos uma tabela de *staging* com os valores em incumprimento. O conteúdo dessa tabela era, depois, guardado num ficheiro a ser enviado para o parceiro de recuperações através de um SFTP.

No âmbito deste projeto existia também um processo que gerava um ficheiro com os resumos dos contactos recebidos pelas nossas equipas de clientes em incumprimento sobre o seu incumprimento.

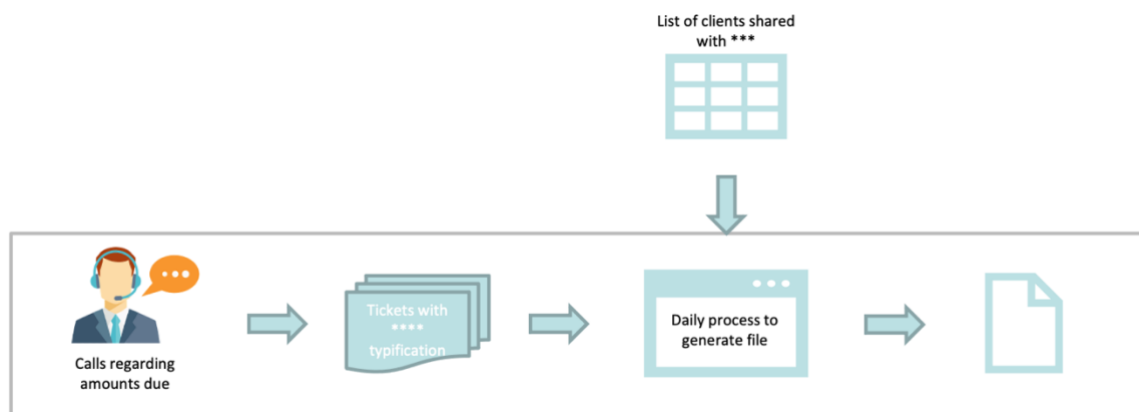


Figura 10 - Processo de geração de ficheiro com chamadas *call center*

Um dos problemas que tivemos quando desenvolvemos esta solução, foi compreender o que é um incumprimento, porque há várias definições que dependem das equipas de negócio e sobre as quais não existia clareza. Uma das complexidades adicionais que o nosso produto implica, é a de que se uma prestação não for paga, esta é transferida para a conta de *revolving* sobre a qual o cliente deve apenas 5%. Isto traduz-se num incumprimento inferior ao valor incumprido pelo cliente. Esta questão ainda não está resolvida.

Esta solução de gestão da delinquência da carteira foi desenvolvida e testada. No entanto, devido à falta de clareza da equipa de crédito, este projeto ficou arquivado até à última semana de estágio. Isto é um bom exemplo de como, mesmo trabalhando em *agile*, se não existir uma visão clara e definida do que se pretende alcançar, não é possível concluir projetos com sucesso nos prazos definidos.

Com a falta de requisitos claros e definidos, é difícil desenvolver um projeto e adiantar trabalho, mas num contexto em que o tempo escasseia, podem montar-se processos ágeis e adaptáveis.

## 6.2. IMPARIDADE E SCORING

Embora este projeto se encontre na secção de troca de dados, porque foi assim que nos chegou, acabou por ser um dos projetos mais completos e que tocou em mais pontos abordados no mestrado.

O negócio de crédito é um negócio de gestão de risco. A empresa é remunerada pelo risco que assume ao adiantar dinheiro aos clientes e pela possibilidade de estes não ressarcirem o valor utilizado. Por isto, é de extrema importância conseguir calcular o risco de um cliente entrar em incumprimento. Para isto, foi desenvolvido, em conjunto com um parceiro, uma aplicação com duas valências distintas, nomeadamente calcular o *score* do cliente e calcular a imparidade.

No início do projeto, foi-me atribuída a função de desenvolver o processo de geração de ficheiros *input* para a aplicação e o processo da integração do *output* da aplicação.

### 6.2.1. Ficheiros

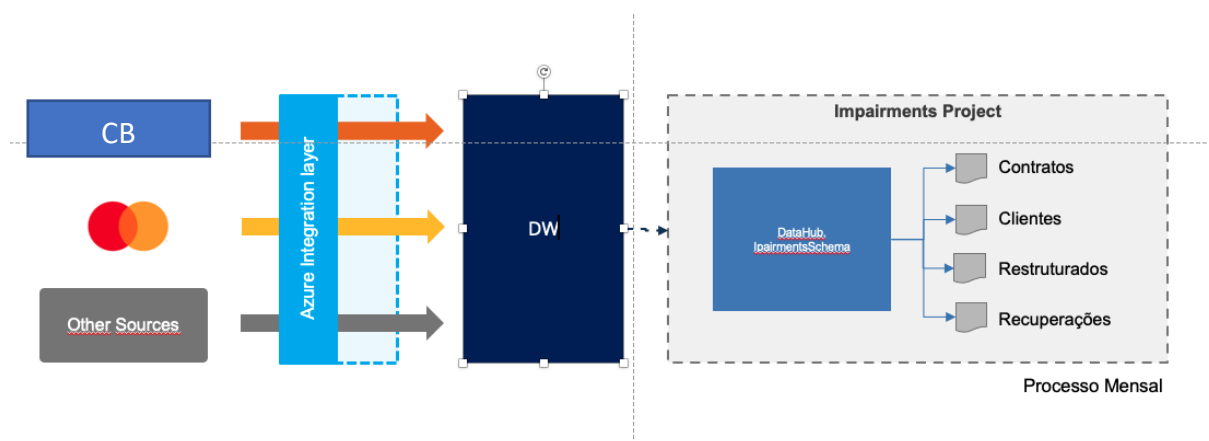


Figura 11 - Desenho de alto nível do processo de geração de ficheiros

Os dados necessários deveriam ser disponibilizados em tabelas num *schema* próprio, e a partir destas tabelas seriam gerados ficheiros .csv para serem enviados.

Contudo, à semelhança do que tem sido mencionado noutros projetos, envolvemo-nos com um grau de profundidade maior.

Como pode ser visto na figura 11 (Arquitetura inicial do projeto), foram-nos solicitados quatro ficheiros. Como a ferramenta responsável por estes cálculos é uma aplicação complexa utilizada por vários bancos em Portugal, não existiu muita margem para alterar a estrutura, podendo apenas enviar campos adicionais. Isto, para nós, constituiu um problema. O Cartão Universo tem a possibilidade de associar uma linha de crédito a dois titulares. E, embora estes casos correspondam a uma fração muito pouco significativa, não podem ser excluídos da análise. Normalmente, existem clientes solidários na dívida em linhas de utilização única, com um único pagamento, como são o crédito à habitação ou o crédito automóvel. Neste cartão, os clientes são solidários em relação à dívida de uma linha de crédito renovável. O objetivo

desta estratégia é poder incluir, como titular da linha, e portador de um Cartão Universo, familiares que possam não ter rendimento. Isto coloca problemas como o de como calcular o *scoring*: deve fazer-se o cálculo ao nível do contrato, ou deve ser feito o cálculo para cada um dos clientes e encontrar uma forma de o agregar?

Dos quatro ficheiros solicitados inicialmente, foi decidido enviar, pelo menos durante a duração do estágio, apenas os ficheiros de Clientes e de Contratos. Para lidar com a dificuldade mencionada no parágrafo anterior, foi necessário criar uma chave adicional no ficheiro de contratos que o ligasse ao segundo titular. No entanto, devido à rigidez da estrutura de envio e à sua falta de adaptabilidade às particularidades do produto, existiram campos para os quais surgiu a necessidade de fazer agregações rudimentares e que deturparam o significado da informação.

Findadas as primeiras discussões, seguiu-se uma fase exaustiva de mapeamento dos cerca de 300 campos pedidos. Este trabalho foi semelhante ao da CRC, tendo existido sinergias entre os dois projetos. Apesar disso, existiram muitos campos que tive que estudar em detalhe, perceber exatamente o seu significado para saber compreender como os poderia calcular, inferir, ou até como poderia arranjar um *proxy* que fosse bom o suficiente para ser posto no lugar de informação não disponível. Este trabalho teve um nível de complexidade adicional, uma vez que, para o envio de histórico, foram feitos mapeamentos para aquilo que era a realidade da SFS à data. Mas, para o processo mensal, tivemos que antecipar quais os campos que deveriam ir recorrendo aquilo que iria ser disponibilizado pelo CB. Embora possa parecer um tópico de somenos importância, como foi discutido, em particular nas aulas de *data minning*, se se constrói um modelo utilizando uma determinada fonte de informação e depois se utiliza outra fonte em que o dado pode assumir significados diferentes, vamos estar a pôr em causa os resultados. Foi nesta fase, e no seguimento do meu alerta para a falta de atenção a pormenores como este, que me foi atribuída a tarefa de escrutinar aquilo que estava a ser feito pelos nossos parceiros e desafiar o seu trabalho numa ótica construtiva e de melhoria contínua.

Para o envio de histórico foram gerados, para os últimos quatro anos (2016-2020), um ficheiro de clientes e contratos para cada mês. Existia informação relativa ao ano de 2015; contudo, estes dados são considerados pouco fiáveis. Deste modo, defini que a informação adicional que trariam não compensaria o risco de comprometerem o modelo com informação errada ou incoerente. Depois da geração do histórico, entrámos numa fase de *debugging*. O parceiro correu processos de *data quality* sobre os dados enviados, tendo detetado algumas inconsistências no que enviámos. Efetuei um trabalho minucioso de identificação do que eram *bugs*, *outliers* ou problemas com informação no próprio DW. Os *bugs* foram corrigidos prontamente, contudo outros casos, como problemas na própria informação (um exemplo foi o de um cliente com 90 dependentes), tiveram que ser tratados pelo parceiro. Isto mostra a importância de existir um processo frequente de *data quality*. Sugeri que fosse montado um, mas, devido à quantidade de projetos a decorrer, não foi possível. Apesar disso, acredito que,

depois de ter exposto a importância acrescida destes processos de *data quality*, agora que a SFS concede crédito necessita de informação tão fiável quanto possível para calcular o *score* dos clientes com a maior precisão, a minha chefia ficou sensível ao argumento e levará esse projeto adiante num futuro próximo. Mesmo assim, a expectativa é de que a qualidade dos dados melhore, uma vez que esta minha preocupação motivou a inclusão de regras de validação de dados na aplicação de *onboarding*.

Outro problema identificado foi o de falta de consistência nos dados. De acordo com as regras de negócio do cartão, cada cliente só pode ter um Cartão Universo e cada Cartão pode ter, no máximo, dois titulares. Contudo, a primeira premissa nem sempre se verifica. Isto pode ser por erro operacional, por se manter associado a uma conta inativa ou até por ter denunciado um contrato e começado outro e tenha existido um *delay* na integração desta informação. Não obstante, consistência é um requisito necessário para a importação dos ficheiros do lado da aplicação. Assim, montámos um processo que valida a coerência dos dados antes de gerar os ficheiros e que exclui informação que não esteja coerente. As validações feitas são as seguintes:

- Para cada contrato, os titulares (primeiro e segundo, caso exista) existem na tabela de clientes;
- Não existem contratos duplicados;
- Não existem clientes duplicados.

O leitor mais atento terá, certamente, reparado que não existe uma validação de que cada cliente está associado a um contrato. Isto acontece porque, nesta situação, os contratos são o *pivot*. Para o processo, é irrelevante serem enviados demasiados clientes.

Uma última nota a destacar, em relação a esta troca de ficheiros, é o facto de serem de grande dimensão (aprox. 1 gb) e conterem (em particular o ficheiro de clientes) informação muito sensível. Desta forma, e para garantir a segurança, foi utilizado um processo que implicava que estes chegassem ao destino com alguma latência. Isto fez com que este processo iterativo de correções ao envio levasse algum tempo. Na situação de troca mensal de ficheiros, surgiram vários problemas que fiquei encarregue de resolver. Para isto, foi necessário que articulasse as diversas partes envolvidas. Foram precisos três meses de calendário até que esta solução fosse estabilizada.

### 6.2.2. Modelos

O projeto de imparidades e *scoring* abrange duas grandes áreas, modelos de decisão de crédito e imparidade (IFRS9).

#### IFRS9

O objetivo da Norma IFRS9 é estabelecer princípios para reconhecer e mensurar ativos financeiros, nomeadamente definir uma nova metodologia de reconhecimento das perdas por imparidade de ativos financeiros com base em perdas esperadas (“expected loss model”).

Não sendo uma área sobre a qual tenha conhecimento, o meu contributo foi o de estar presente nas discussões a dar suporte relativamente às questões que foram surgindo relativamente aos dados que disponibilizámos. Uma das alternativas para o cálculo da LGD necessitava de informação de que não dispúnhamos – as habilitações literárias dos clientes – e houve a necessidade de compreender a facilidade com que essa informação podia ser obtida e a viabilidade de o fazer. Outra informação que passou pelo mesmo processo foram os encargos dos clientes que irão ser discutidos no projeto seguinte.

As discussões dentro deste fórum prenderam-se, essencialmente, com a qualidade da informação, com o seu propósito e sobre formas alternativas para a disponibilizar. O cálculo de imparidade requer a computação de *cash-flows* para os cartões. Isto implica que a informação sobre os balanços de cada contrato tem que ser precisa e, foi-nos solicitado, que fosse tão desagregada como possível. Isto constituiu uma alteração ao pedido adicional de dados, que teve que ser pensada funcionalmente e implementada. Cada uma destas alterações implicava um reprocessamento do histórico o que era um processo moroso (1 – 2 dias) e, por isso mesmo, a sua execução carecia de uma reflexão cuidada.

Atingiu-se uma fase em que, após uma análise extensa à carteira de clientes, existiram discussões relativas à metodologia a utilizar:

**Primeira Opção** – Progressão Linear Constante: “Nesta abordagem, desde a entrada em incumprimento até ao início do 4º ano em *default* (fim do “período de *workout*”) recorreu-se ao apuramento de um fator de progressão linear para aplicar aos créditos com antiguidade em *default* compreendida no período definido (1 a 4 anos em *default*).”

**Segunda Opção** – Progressão Ajustada por Coeficiente de Aceleração: “Nesta abordagem, desde a entrada em incumprimento até ao início do 4º ano em *default* (fim do “período de *workout*”) recorreu-se ao apuramento de um fator de progressão ajustado por um coeficiente de aceleração de 200% (julgamental) para aplicar aos créditos com antiguidade em *default* compreendida no período definido (1 a 4 anos em *default*).”

Outra discussão importante foi a da marcação do *stage* de um contrato. O processo de atribuição de *stages* de risco a todas as operações em carteira, para uma dada data de referência, inicia-se através da verificação da aplicação de, pelo menos, um critério de *stage* 3 ao cliente titular da exposição em análise. Se o cliente tiver alguma exposição classificada em *default* é classificada como *stage* 3.

Salienta-se que os critérios definidos para agravamento de exposições para *stage 3* são aplicados ao nível do cliente, assim, a verificação de um critério de 3 numa exposição, contagia as restantes exposições do mesmo cliente, independentemente do número de titulares associados ao contrato.

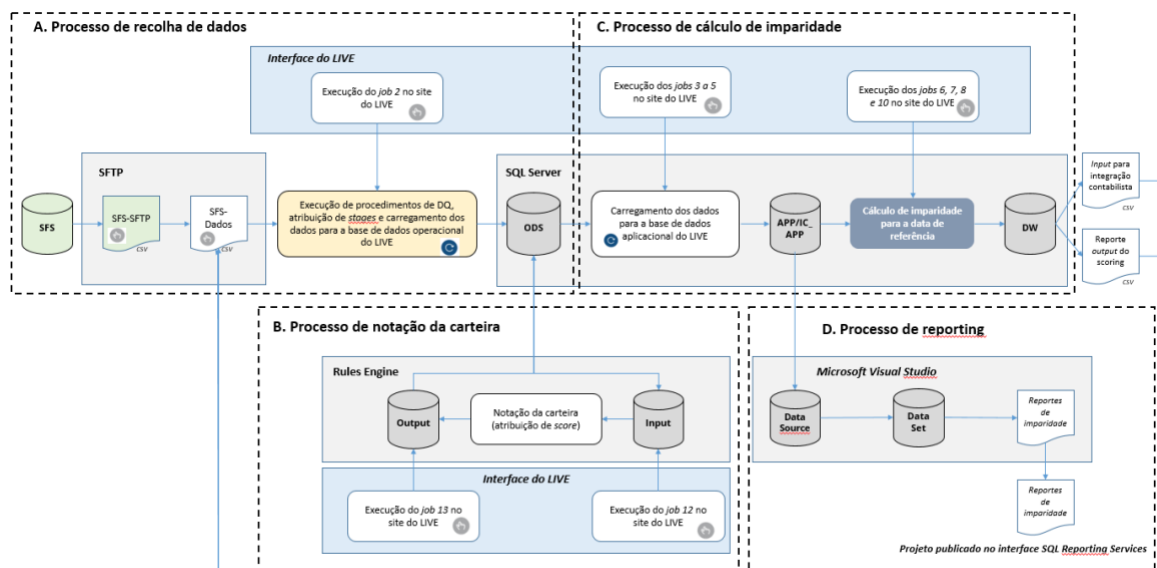


Figura 12 - Desenho dos processos de cálculo de imparidade e notação da carteira

Importa referir que, no que respeita ao cálculo de imparidade, considera-se que a classificação de um crédito em *stage 3* corresponde diretamente à definição de *default* implementada pela SFS, à qual apenas se acresce, para efeitos de *staging*, a aplicação das quarentenas para créditos em situação de incumprimento e reestruturados. Como estes critérios dependem da identificação de um determinado comportamento, existe a necessidade de perceber exatamente quais são os critérios de cada *stage* e perceber que campos podem verificar essa informação. Exemplo de um campo que é condição suficiente, é a existência de dívidas para com a autoridade tributária ou a segurança social. Esta informação, por exemplo, não existia, e teve que ser desenvolvido um projeto para a obter. Em suma, são marcados em *stage 3* clientes que estejam em incumprimento com a SFS ou qualquer outra entidade de cuja informação possamos dispor. Estarão marcados como estando em *stage 2* os clientes que tenham algum crédito em atraso. E em *stage 1* estarão todos aqueles que não padeçam deste tipo de incidências. Esta informação é importante para *reporting*, e é importante para a utilização, enquanto campo para a aplicação de regras condicionais, em vários campos de reporte relacionados com a imparidade. Com efeito, as operações devem ser classificadas num dos três diferentes *stages*, de acordo com um conjunto de regras definidas *a priori*. Esta classificação é subsequentemente utilizada no cálculo de imparidade que é integrado em DW, via ficheiro, mensalmente.

### Modelos de Decisão de Crédito

O modelo de decisão de crédito é a abordagem metodológica à atribuição de um montante a ser disponibilizado aos clientes. Este processo encontra-se dividido em quatro fases:

1. **Modelo de elegibilidade** – esta etapa é responsável pela restrição imediata aos clientes em termos de elegibilidade de crédito. Isto pode ter que ver com regras de negócio, como restrições etárias ou o requisito de empregabilidade, como com regras regulamentares – o cliente não pode ter uma DSTI acima de 50%. Este assunto vai ser discutido em mais detalhe no projeto de análise de solvabilidade.
2.
  - 2.1 **Modelo aplicativo** – atribui um *score* com base na informação sociodemográfica do cliente bem como em função do seu estado na CRC.
  - 2.2 **Modelo comportamental** – atribui um *score* a um cliente que já seja cliente na SFS. Este *score* tem em consideração as variáveis do primeiro modelo e inclui variáveis adicionais sobre o comportamento do cliente na SFS.
3. **Modelo limite de crédito** – Consome o *score* atribuído pelo modelo anterior e, recorrendo a regras regulamentares e de negócio atribui um limite de crédito.
4. **Alertas** – Esta etapa faz a validação da informação prestada pelo cliente e do resultado dos modelos anteriores, dando alertas que possam auxiliar os responsáveis pela atribuição de crédito na sua decisão.

O meu envolvimento nos modelos de decisão foi ao nível dos modelos de *score* e da abordagem holística à solução. O modelo aplicativo tem uma arquitetura em *real-time*, aqui, o meu contributo foi mais reduzindo, ficando-se pela interpretação dos campos recolhidos pelo *onboarding*, da variável de encargos que é descrita numa secção própria e da interpretação dos dados da CRC. Aqui a ideia foi a de utilizar as mesmas *flags* que estávamos a desenvolver para o modelo comportamental, mas *via web service*.

O modelo comportamental, por seu lado, é calculado mensalmente. Este modelo utiliza um conjunto muito superior de variáveis e tem dois objetivos, nomeadamente permitir, mensalmente, o acompanhamento da evolução da qualidade e da delinquência esperada da carteira e ser a base para iniciativas comerciais de aumento de *plafond*. Estas são de suma importância, uma vez que, se conseguirmos aumentar o crédito que é dado a clientes com baixa probabilidade de entrar em incumprimento, conseguimos aumentar a receita com um risco reduzido.

A arquitetura do processo é manifestamente mais simples do que a do modelo aplicativo uma vez que enviamos dois ficheiros, que são integrados e processados pela aplicação que por sua vez nos devolve um ficheiro que inclui para cada contrato, para esse mês:

**Score** – Uma escala que corresponde à segmentação das probabilidades de *default*.

**Limite Máximo** – É o resultado da aplicação do modelo de limites para cada contrato.

Embora pareça pouca informação, estes dois campos são fulcrais para o sucesso do negócio.

De um ponto de vista tecnológico este modelo não exigiu grande esforço, mas a vertente funcional, por seu lado, é mais complexa. Uma dúvida que levantei, e que constitui uma falha fundamental na abordagem metodológica deste modelo, é ter sido desenvolvido recorrendo a histórico que já havia sido selecionado. A informação de histórico de incumprimento que tínhamos era apenas dos nossos clientes, que, antes de terem sido aceites, tinham passado, *a priori*, por um modelo de *scoring* do parceiro. Isto coloca dois problemas:

- Existirem situações que o modelo desconhece e em relação às quais o comportamento pode ser errático. Para lidar com esta situação, foi sugerido que fosse dada especial atenção às regras de elegibilidade e aos alertas. O racional que suporta esta recomendação é o de que, se for possível excluir à partida estes casos que foram rejeitados pelo parceiro de crédito e para os quais o modelo não foi calibrado, então será possível reduzir de forma significativa os efeitos nefastos do enviesamento na construção do modelo.
- Ter sido utilizado uma percentagem pouco significativa de clientes predispostos à delinquência. Existiam diversas formas de lidar com isto, mas a falta de conhecimento do negócio impediu que fosse adotado, por exemplo, a abordagem de criação artificial de dados. Assim, o compromisso alcançado foi o de monitorizar frequentemente os resultados do modelo, fazendo uma calibração ao fim de seis meses e de um ano.

De facto, de um ponto de vista metodológico, a utilização de dados enviesados para treinar um modelo parece um erro crasso. No entanto, dado que a alternativa poderia ter sido a de não ter um modelo de *scoring*, ou ter um que não estivesse adaptado à realidade portuguesa, percebemos que este é um compromisso sensato.

O algoritmo utilizado foi a regressão logística. Porém, a razão da escolha deste tipo de modelo prende-se com a necessidade regulamentar de explicar as decisões de concessão de crédito. Existem modelos com resultados melhores, por exemplo redes neurais, mas cuja interpretação é complexa ou mesmo impossível.

A abordagem metodológica do modelo aplicacional e do modelo de concessão decorreram em paralelo. Como a do modelo comportamental foi mais complexa, vamos optar por descrevê-la, deixando aqui nota de que o modelo aplicacional é em tudo semelhante, mas utilizando apenas o módulo sociodemográfico.

A primeira fase foi a dos dados. Aqui foi definido e implementado um critério de *default*. De seguida, na fase de *Model Design* surgiram duas hipóteses, nomeadamente a da construção de um modelo uni modular, ou a construção de um modelo com dois módulos. Foram testadas as duas abordagens e a que fez mais sentido do ponto de vista de negócio e dos resultados obtidos foi a abordagem dos dois módulos. Para testar este modelo foi construída uma amostra de modelização que utilizou duas datas de referência com uma janela mínima de observação de doze meses. A variável *target* foi se entrou em *default* na janela de

doze meses subsequentes à data de referência. Para isto foram excluídos todos os contratos que estavam em incumprimento na data de referência.

Na fase seguinte, com o desenho do modelo concluído, foi feita uma análise univariada que consistiu em análises descritivas das variáveis, análises de valores *missing* e *outliers*. Existiram aqui diversas iterações para clarificar em que casos é que valores *null* constituíam efetivamente um *missing* ou quando tinham significado intrínseco. Também foram identificados e discutidos *outliers*. Com base nesta informação foi feita uma pré-seleção das variáveis que tinham qualidade suficiente para serem utilizadas no modelo. Fui também incumbido de discutir com vários colaboradores da SFS, o seu conhecimento sobre os incumprimentos e indicadores que lhes pudessem estar associados.

O passo seguinte foi levar a cabo uma análise bivariada, com o objetivo de analisar e selecionar as variáveis que têm uma maior capacidade discriminante e correlação esperada com a variável *target*.

Depois desta análise, algumas variáveis foram excluídas, tendo outras sido transformadas com o intuito de eliminar problemas associados à existência de *outliers* e *missing values* e, sobretudo, com o propósito de aumentar o poder discriminante. Uma das formas de o fazer, foi através da discretização de algumas variáveis.

Fechado o conjunto de variáveis passíveis de integrar o modelo, foi feita uma análise de correlações para testar e quantificar o grau de associação entre diversas variáveis explicativas. Isto permitiu a identificação de grupos de variáveis correlacionadas e a eliminação daquelas com menor poder discriminante ou significado de negócio.

O processo de seleção de variáveis culmina com a decisão de que variáveis serão incluídas no modelo final. O primeiro passo é a execução do algoritmo *stepwise*. Depois, existe uma análise com o propósito de gerir o equilíbrio entre a performance e o número de variáveis, tendo em conta que devido ao viés supramencionado, o problema de *overfitting* seria particularmente danoso.

As variáveis utilizadas são de cariz proprietário; contudo, é relevante mencionar que as variáveis com mais poder discriminante foram variáveis com transformações não lineares e rácios. Isto reforça a importância do trabalho de auscultação do negócio e de transformação de conhecimento tácito individual em dados que podem ser modelizados e contribuir para o sucesso do modelo de *score* e consequentemente para o sucesso do negócio.

### **6.3. VARIÁVEL ENCARGOS FINANCEIROS E NÃO FINANCEIROS**

Para enriquecer o modelo de *scoring* e para analisar a solvabilidade dos clientes é necessário conhecer e compreender os seus gastos e rendimentos. Os rendimentos não são um problema, uma vez que os clientes os declaram no momento da adesão. Estes podem ser

autodeclarados no caso da adesão simplificada, mas carecem de um comprovativo para a adesão a linhas de crédito superiores a 1 500€. Contudo, na altura do estágio, a SFS não dispunha de qualquer tipo de informação relativa aos encargos dos clientes. Estes encargos podem ser divididos em dois grupos: encargos financeiros e encargos não financeiros.

### **6.3.1. Encargos financeiros**

#### **Conceito**

Os encargos financeiros consistem nas obrigações recorrentes (enquanto devedores) que os indivíduos/clientes têm com instituições de crédito. Exemplos desta tipologia de encargos são o crédito à habitação, crédito automóvel ou crédito pessoal. No contexto da SFS, estes encargos são aqueles que têm prestações fixas, não incluindo, portanto, instrumentos de crédito. O racional por detrás desta posição prende-se com o facto de as prestações consistirem numa obrigação com determinada periodicidade e relativamente às quais o cliente tem uma responsabilidade que não pode ser adiada. No caso do cartão de crédito, por exemplo, o cliente pode ter o montante disponível, mas, ao não o utilizar, não está a aumentar as suas despesas recorrentes. Contudo, dentro deste âmbito, e com o propósito de monitorização da carteira, existem outras informações relevantes, como o número de cartões de crédito ou a existência de créditos vencidos.

#### **Implementação tecnológica**

A informação referente aos encargos financeiros pode ser obtida de forma direta através da CRC. Mas a sua estrutura não dá resposta às perguntas colocadas pelos processos que utilizam estes dados. Este é um exemplo claro da diferença entre dados e informação, discutido ao longo do mestrado. A informação da disseminação da CRC vem em ficheiros XML que, depois de integrado, se podem reduzir a duas tabelas principais: instrumentos e

prestações.

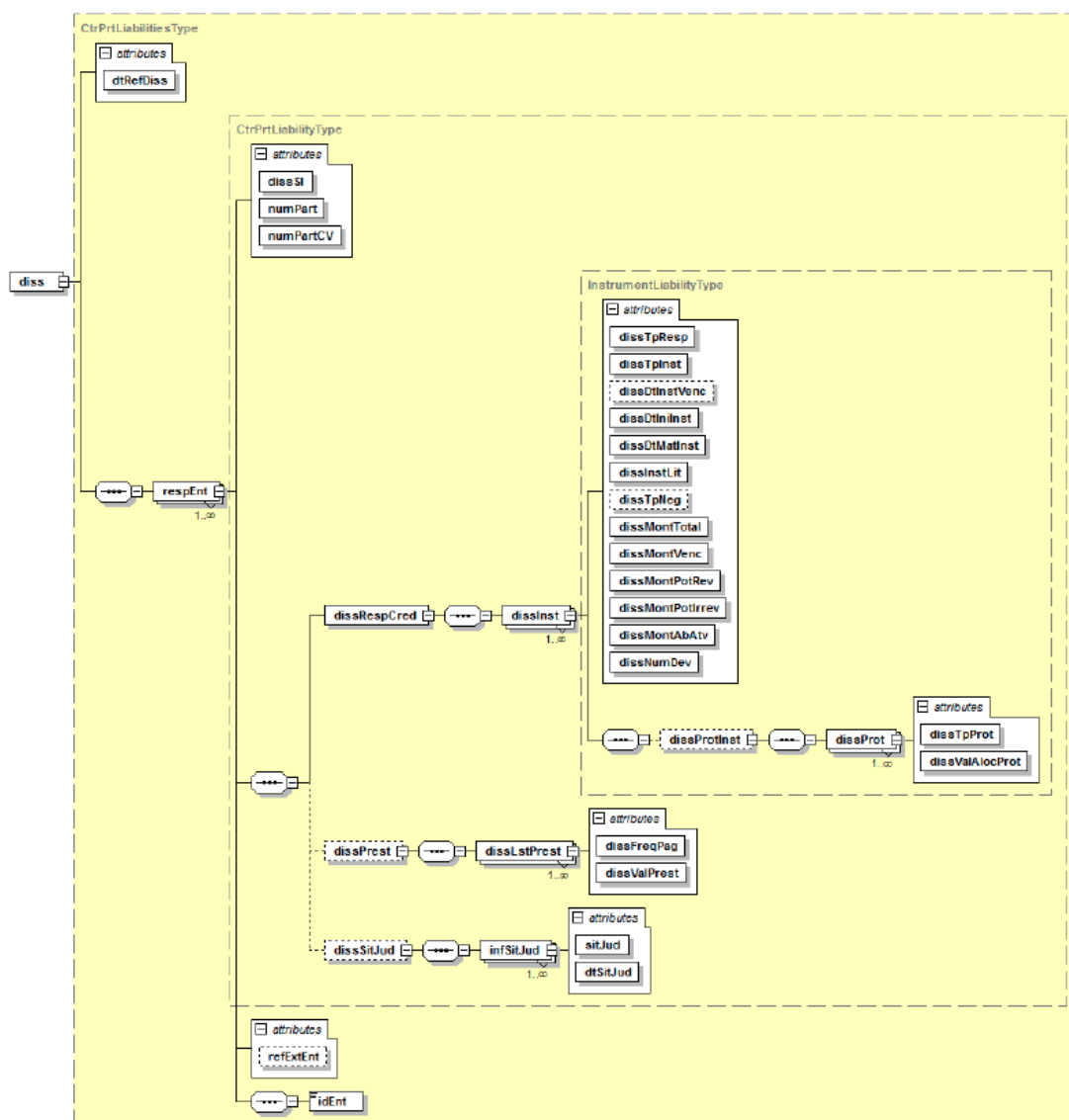


Figura 13 - Estrutura do XML DISS do BdP

### Instrumentos:

Esta tabela tem uma linha por cada instrumento de crédito sobre os quais o cliente tenha responsabilidades. Neste momento, embora sejam integradas todas as linhas, as variáveis criadas só têm em consideração instrumentos para os quais o cliente tenha responsabilidades do tipo 002 – Devedor. Contudo, de futuro, pode surgir a necessidade de iterar sobre instrumentos em que os clientes tenham a responsabilidade 003 – Avalista/Fiador.

**Flag\_Vencido:** Uma variável que itera sobre os instrumentos de cada entidade e assume o valor de 1 se existir algum com data de instrumento vencido (dissDtInstVenc) e de 0 se não existir.

**Flag\_Instrumento\_Litigio:** Uma variável que itera sobre os instrumentos de cada entidade e assume o valor de 1 se existir algum em litígio judicial (dissInstLit) e de 0 se não existir. Existiu a discussão de se faria mais sentido fazer o somatório do campo. Esta hipótese poderia trazer mais informação, contudo, podem existir vários instrumentos marcados em litígio por serem referentes à mesma concedente. Assim para o propósito da variável, optou-se por uma abordagem que traz menos informação, mas mais confiança na informação.

**Flag\_Renegociacao\_Incumprimento:** Uma variável que itera sobre os instrumentos de cada entidade e assume o valor de 1 se existir algum tipo de renegociação (dissTpNeg) = 004 – Renegociação por Incumprimento, e de 0 se não existir. O motivo para se adotar uma *flag* em vez de um somatório é semelhante ao do campo anterior.

**Montante\_Vencido:** Uma variável que itera sobre todos os instrumentos da entidade e calcula o somatório do campo Montante vencido (dissMontVenc). Esta variável parece tornar a variável Flag\_Vencido redundante, mas torna o processo mais robusto. Aconteceu existirem discrepâncias entre estas variáveis que revelaram erros na informação integrada.

**Montante\_Abatido\_Ativo:** Uma variável que itera sobre todos os instrumentos da entidade e calcula o somatório do campo Montante Abatido ao ativo (dissMontAbAtv).

Enquanto as linhas dos instrumentos constituíam apenas dados, as variáveis anteriores já comportam informação e são utilizadas ativamente em processos de negócio. Quando as *flags* Flag\_Instrumento\_Litigio e Flag\_Renegociacao\_Incumprimento assumem o valor de 1, os cartões dos clientes são automaticamente bloqueados. O mesmo é verdade quando os dois campos Montante anteriores são superiores a 50.

**Nr\_Cartoes\_Credito:** Esta variável consiste na contabilização do número de instrumentos do tipo 0041 – Cartão de crédito que cada cliente tem. Esta variável ainda não está associada a nenhum processo de negócio, mas o objetivo é estudar a sua relação com a delinquência dos clientes.

Foi discutida a possibilidade de criar outras variáveis, como o valor total de responsabilidades de crédito, mas como os instrumentos podem ter mais do que um devedor associado, existem problemas de dupla contabilização. Assim, a criação destas variáveis carece de análise futura e do detalhe da utilidade que possam ter para o negócio e gestão de risco.

## **Prestações:**

Esta tabela consiste nas próximas prestações associadas aos instrumentos reportados à CRC. Estas prestações são agregadas por frequência. O primeiro passo foi a criação de uma variável – prestação\_mensal\_total – que faz o somatório das diferentes prestações ajustadas

por periodicidade. Um cliente que tenha uma prestação mensal de 100 € e uma prestação trimestral de 300€ irá ter este campo com o valor 200€. É importante definir que na tabela de domínio está contemplada uma periodicidade irregular. Por esta se aplicar apenas a casos excecionais e não existir nenhuma agregação lógica para estes valores, foi decidido ignorar esta informação. Este campo vai ser utilizado para o cálculo da DSTI. Novamente se coloca a questão de quando existem dois titulares, uma vez que não temos informação relativamente ao número de devedores associados ao instrumento cuja prestação vem assinalada. Assim, para os casos em que existem dois titulares, existem duas abordagens, designadamente a de calcular a DSTI para os dois clientes e escolher uma das duas, ou somar os encargos e os rendimentos e dividi-los por dois.

Embora estas variáveis comportem muita informação sobre os encargos financeiros dos clientes, a sua utilização não está livre de riscos. Estas variáveis não incluem encargos financeiros com entidades não supervisionadas pelo BdP. Exemplo disso, são concedentes estrangeiras. Este risco pode ser mitigado, por exemplo, aumentando o escrutínio relativamente a clientes cuja nacionalidade é diferente do país de residência (Portugal). Também não incluem informação relativa a dívidas que os clientes possam ter com a Autoridade Tributária e Aduaneira ou com a Segurança Social.

### 6.3.2. Encargos não financeiros

#### Conceito

Para calcular a taxa de esforço de um cliente não basta ter informação sobre as suas responsabilidades financeiras. Mensalmente, existem despesas inevitáveis e que se sobreporiam à liquidação da dívida do Cartão Universo em situação de dificuldade financeira. Exemplo disto, são as despesas com bens essenciais ou com renda para indivíduos que arrendam a habitação primária. Para fazer a análise de solvabilidade, mas também para enriquecer os modelos de *scoring*, é importante reunir informação acerca das despesas que os clientes têm mensalmente. Algumas instituições de crédito solicitam esta informação no momento da adesão. A Sonae FS tem orgulho na experiência que proporciona aos seus clientes. Assim, não seria possível, durante a adesão, inquirir a um prospeto cliente quanto gasta, mensalmente em comida. Para além disso, não existe histórico de encargos declarados, o que dificultaria a utilização desta variável no modelo. Foi-me, então, pedido que desenvolvesse uma abordagem para resolver este problema.

Do senso comum vem que as despesas que os clientes têm mensalmente estarão intrinsecamente relacionadas com a informação sociodemográfica. É trivial que, *ceteris paribus*, um cliente com mais dependentes terá mais despesas mensais. O desafio foi, partindo da informação de que dispúnhamos, fazer esta estimativa com confiança e de forma que pudesse ser explicada e auditada. Foi ponderada a ideia de fazer um inquérito a uma amostra representativa dos nossos clientes e utilizar esta variável *target* para calibrar uma rede

neuronal capaz de fazer a estimativa. Apesar de razoável e de plausivelmente conduzir a melhores resultados do que a abordagem escolhida, trar-nos-ia problemas ao nível da auditabilidade do modelo.

Foi desenvolvido então um modelo que está alicerçado em informação existente sobre as despesas das famílias portuguesas e que, em função das suas características, estima os encargos anuais que têm.

$$Encargos_{l,r} = Custo Nacional_r \prod_c (\sum_{c_v} Fator_{c_v} \cdot D_{l,c_v}) \quad \forall l, r$$

Onde:

$l$  : Índice que representa cada Cliente

$D_{l,c_v}$  : Valor  $v$  para a característica  $c$  que o cliente  $l$  assume onde:

$$D_{l,c_v} = \begin{cases} 1 & \text{se o cliente possuir esse valor para essa característica} \\ 0 & \text{se o contrário} \end{cases}$$

Foi necessário um amplo trabalho de investigação para comparar os fatores que existiam disponíveis com a informação disponível em DW, e houve a preocupação de aplicar um teste à independência para minimizar a sobrevalorização de determinado atributo. A solução para este problema foi o da criação de atributos compostos. Exemplos de elevada correlação foram a situação profissional e o grupo profissional, a situação profissional e a idade, a idade e a composição do agregado familiar e a idade e o grupo profissional 'pensionistas'. Para lidar com as duas primeiras situações, removeu-se o fator relacionado com a situação profissional. Para as outras duas, condicionou-se a aplicação do fator Idade apenas a casos onde a idade não esteja implícita quer na composição do agregado como para o grupo profissional.

Resumindo, este modelo classifica cada cliente em função de uma série de atributos, para cada classificação atribui um fator dependendo da classificação que teve, e multiplica estes fatores pelo encargo de referência.

Existem quatro encargos de referência: essenciais, não essenciais, veículos e habitação. Optámos por não utilizar os encargos com veículos, uma vez que acreditamos que estes já podem estar incluídos nos outros encargos, quer nos essenciais com a rubrica transporte, quer nos encargos financeiros.

Foram utilizados então os encargos essenciais e os encargos de habitação. Os últimos foram utilizados no caso de o cliente se ter declarado como arrendatário. Caso em que a informação relativamente aos encargos à habitação não está incluída nos encargos financeiros.

Um dos problemas deste modelo foi a falta de uma variável-objetivo para validar a sua eficácia. Numa tentativa de superar este problema foi feita uma segmentação dos clientes com base nos encargos não financeiros estimados. Foram depois feitas entrevistas a alguns clientes para ganhar sensibilidade à precisão do modelo. Como consequência destas

entrevistas, foram feitos alguns ajustes menores. No fim deste projeto ficámos confortáveis com o resultado, uma vez que esta variável teve poder discriminante em ambos os modelos de *scoring*. O nosso parceiro, confirmou que esta variável seria provavelmente, o motivo pelo qual o nosso modelo aplicacional tinha um comportamento acima da média dos modelos que haviam desenvolvido até à data.

Este projeto é a prova de que, num contexto de mudança, muitas vezes existe a necessidade de saltar etapas e utilizar abordagens menos completas para resolver problemas concretos em tempo útil.

### Implementação tecnológica

Depois de desenhar o modelo, surgiu a necessidade de o implementar. Numa primeira fase, foi feita uma implementação que foi incluída no ETL. Esta solução teria que ter capacidade para reprocessar o histórico, o que implica calcular o valor de encargos financeiros para centenas de milhões de registos. Para que este processamento não demorasse um tempo excessivo a executar, optou-se por criar uma tabela com todas as combinações de fatores possíveis. O *stored procedure* cria também uma tabela temporária onde existe uma linha por registo e onde, para cada atributo, consta a classificação do cliente. Depois é feito uma correspondência exata e esse será o valor do fator para o cliente. São, depois, aplicadas as regras lógicas. Este processo, foi implementado com sucesso e faz parte do processo diário de ETL, dispondo assim o DW deste campo atualizado para cada cliente.

Como esta variável foi incluída no modelo aplicacional, surgiu a necessidade de a implementar de forma a que pudesse, recebendo parâmetros, devolver um valor.

Uma vez que este serviço vai calcular os encargos para um cliente de cada vez, não há ganhos de performance ao utilizar a metodologia anterior. Poderão existir mesmo prejuízos. Assim, a solução montada recebe sete parâmetros, como por exemplo o código postal - @postal\_code nvarchar (20) ex: '2750-781'- ou o salário - @monthly\_wage numeric (18,2) ex: 1200. Para cada parâmetro, valida qual será a classificação e o fator correspondente. Depois executa a calculatória e devolve um JSON como pode ser visto no diagrama seguinte.

```
{
  "status": {
    "code": "SUCCESS",
    "details": null
  },
  "data": {
    "monthlyWageEstimated": 1166.67,
    "monthlyExpensesEstimated": {
      "essentials": 368.93,
      "house": 0.0,
      "total": 368.93
    }
  }
}
```

Figura 14 - Output do serviço de cálculo de encargos não financeiros

Ao contrário da primeira implementação, esta não esteve livre de percalços, e, por causa de um erro num *case*, apenas devolia valores para códigos postais da Grande Lisboa.

Este problema foi resolvido e foram implementadas salvaguardas em que, no caso de não ser possível classificar um determinado parâmetro, o fator associado seria de 1, ficando este sem efeito. Isto é importante especialmente ao nível dos códigos postais, uma vez que é uma variável cujo valor que pode assumir varia com frequência e para a qual não existe processo automático de ajuste da classificação.

Embora a construção desta variável tenha sido importante para os modelos de *scoring*, este trabalho mostra que ainda existe muito potencial de capitalização do manancial de informação de que a SFS dispõe. Ao longo do estágio, acho que dei um contributo significativo para que se começasse a investir numa estratégia de valorização da informação.

#### 6.4. ANÁLISE DE SOLVABILIDADE

Antes de se decidir o modelo de transferência dos clientes, existiu a necessidade de conhecer, em detalhe, a solvabilidade da carteira. A importância desta análise advém da necessidade de mitigar assimetrias de informação nas negociações entre a SFS e o parceiro de crédito e, ao mesmo tempo, capacitar a empresa para desenhar o modelo de transição que cumprisse, da melhor forma, os seus objetivos comerciais e estratégicos.

Para levar a cabo esta avaliação, foi necessário calcular o *score* de cada um dos clientes e a sua DSTI. Segundo o BdP, a DSTI é definida como: “rácio entre o montante da prestação mensal calculada com todos os empréstimos do(s) mutuário(s) e o(s) seu(s) rendimento(s) mensal(ais), calculado nos termos do artigo 4.º”.

Para o cálculo de *scoring* são necessárias variáveis dentro das quais se incluem as *flags* discutidas na secção anterior e os encargos financeiros. Também era necessária a variável dos encargos não financeiros. Reunir estes dados pode parecer simples. Contudo, como explicado anteriormente, só têm acesso à informação de disseminação as entidades que reportaram informação financeira para determinada entidade. Como a SFS não tinha a atividade de concessão de crédito, também não tinha a responsabilidade de reporte e, consequentemente, não possuía esta informação. A solução foi fazer um pedido especial ao BdP para submeter PIRC para os clientes. Este pedido pode ser feito quer por ficheiro, quer por API. Contudo, por API tem o dobro do preço e seria moroso devido ao elevado número de clientes. Assim, a solução arranjada, foi adaptar o motor da CRC para que submetesse PIRC para uma lista de clientes, respeitando o limite de 2 500 registos por ficheiro. Para não existirem problemas de assincronismo entre os dados que estavam a ser utilizados, desenhei e implementei uma solução tecnológica que permitiu fazer calcular a informação para os clientes que existiam à data da sua implementação e, de forma automática, fazer o cálculo para novos clientes.

Funcionalmente, aquando do desenho da solução, surgiram desafios. Um ponto discutido foi o salário. Os prospetivos clientes, no momento da adesão respondem à seguinte pergunta: “Qual é o seu rendimento mensal?”. Do senso comum vem que a resposta será referente aquilo que ganham por mês e não ao seu rendimento anual dividido por 12. Assim, após uma análise mais detalhada compreendeu-se que para o cálculo da DSTI surgia a necessidade de fazer uma estimativa do salário dos clientes aplicando a seguinte regra: para clientes empregados por conta de outrem multiplica-se o salário pelo fator 14/12. Este campo passou a fazer parte do ETL e a estar disponível em DW como SFS\_adjusted\_monthly wage.

Outro ponto que foi critico, e que foi discutido na secção anterior do cálculo de *scoring* e imparidade, tem que ver com a unidade a que deve ser calculado a DSTI, nomeadamente se ao nível do cliente, ou se ao nível do contrato. A fórmula da DSTI é “rácio entre o montante total das prestações mensais associadas a todos os empréstimos detidos pelo mutuário e o seu rendimento mensal líquido de impostos e contribuições obrigatórias à Segurança Social”. Para clientes que são os únicos titulares do contrato, logicamente, não existe distinção.

Conquanto, para contratos que possam ter dois titulares existem dois cenários, designadamente calcular para cada cliente individualmente e utilizar uma regra de agregação ou calcular de forma agregada; ou calcular para cada cliente individualmente.

Uma aprendizagem importante deste projeto, é a importância em segregar os decisores e as equipas que dão suporte na utilização de dados. Neste caso em concreto, a promiscuidade entre o meu trabalho e as equipas de negócio aliada a uma falta de clareza nos requisitos levou a que alguns conceitos se fossem moldando com o objetivo de que os resultados correspondessem à expectativa. Assim, parece-me que, sempre que possível, quando é preciso informação específica para uma decisão de negócio, deve existir clareza sobre o que é que essa informação traduz de modo a que não exista margem para ajustes em função das necessidades em detrimento da precisão das soluções propostas.

## 6.5. PARTILHA DE DADOS COM PARCEIRO DE TITULARIZAÇÃO

Após a transição para o novo modelo de operação, existiu a possibilidade de se adotar um modelo de titularização. Para isso, seria necessário que o parceiro de titularização recebesse dados. Para isto foi montado um projeto com três entidades, a SFS, o parceiro e uma empresa de TI subcontratada pelo parceiro. O objetivo deste fórum foi o de implementar os fluxos de troca de informação. Um aspeto em que este projeto difere dos restantes de partilha de dados é o facto de ser um fluxo unidirecional. Os dados são enviados pela SFS para o parceiro sem que exista retorno.

O primeiro desafio foi definir os dados a enviar. Sendo um projeto novo para ambas as entidades, não existia clareza sobre aquilo que era necessário e aquilo que era possível ser trocado de um ponto de vista regulamentar (RGPD). Assim, houve a necessidade de levar a cabo múltiplas sessões, com o objetivo de chegar a uma lista consensual para a informação a partilhar. A abordagem foi a de partilhar a informação de crédito tal como é integrada em DW, o que facilita o trabalho de geração dos ficheiros a ser enviados e coloca o ónus de calcular os campos de que possa necessitar.

Para implementar este processo, a SFS adaptou-se àquilo que foram os requisitos do parceiro ao nível de estrutura e nomenclatura.

O primeiro requisito estava relacionado com o nome dos ficheiros que deviam seguir a estrutura: `yyyymmdd_hhmmss_NomeTabela`. Por exemplo, o ficheiro `20210305_154438_SFS_LkCreditArrangements.del` diz respeito à tabela `LK_CreditArrangements` e foi gerado no dia 5 de Março de 2021 às 15:44:38. Isto foi facilmente implementado no processo de SSIS utilizando um parâmetro no nome que era calculado com recurso à função *get date*.

O segundo requisito foi a inclusão de um campo `DT_Information`. Este campo consiste na data de referência dos dados e deve seguir a estrutura `AAAA-MM-DD`. Este pedido gerou alguma discussão uma vez que alguns dados são atualizados diariamente, enquanto outros são atualizados apenas quando existem alterações. Isto implica que, se este campo assumisse a data para a qual o ficheiro está a ser gerado, não corresponderia, necessariamente, à data dos dados, mas a uma data para a qual aquele dado era o mais atualizado que existia em DW. O resultado destas discussões foi manter o campo como indicado pelo parceiro, e este ficaria com o ónus de utilizar o campo de *create date* para fazer ajustes e validações. Outra dúvida que surgiu prende-se com os ficheiros de periodicidade mensal e se deveriam ter este campo preenchido apenas com o ano e mês, ou se deveriam incluir o último dia do mês a que se referiam. Como a informação é integrada em DW apenas com o ano e mês de referência, foi decidido partilhar o campo `DT_Information` neste formato e, uma vez mais, o ónus de fazer as transformações necessárias ficou do lado da equipa de desenvolvimento do parceiro.

Existiram ainda requisitos relativos à geração do ficheiro. Estes deviam ser do tipo `.csv`, utilizar o separador `§` e ter a extensão `.del`. O motivo para a utilização deste separador é evitar

que seja um caracter que possa existir, de algum modo, dentro de um campo, colocando assim em causa o processo de importação.

O último requisito de estrutura passado pelo parceiro foi a necessidade da existência de um ficheiro com o mesmo nome do ficheiro que contem informação, mas com a extensão .trg. A colocação do ficheiro no SFTP do parceiro sinaliza que o ficheiro que contém a informação já foi transferido com sucesso e está pronto para ser importado. Esta implementação foi desenvolvida pela equipa de TI da SFS que também é responsável pela infraestrutura da ligação.

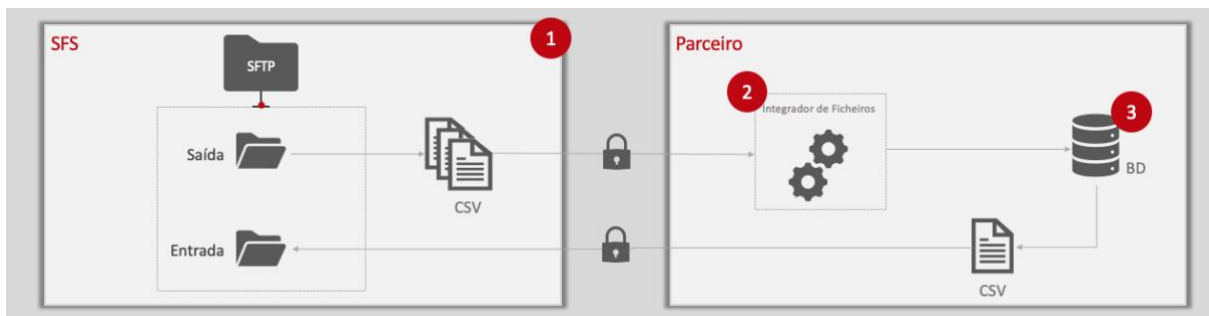


Figura 15 - Esquematisação do processo de troca de ficheiros

Definidos os requisitos funcionais e de estrutura, desenvolvemos um conjunto de processos para garantir a geração diária dos ficheiros. A solução encontrada foi um processo de SSIS que vai gerar ficheiros com a informação existente em DW. Este processo é constituído por um pacote mestre e um pacote para cada um dos ficheiros gerados. O pacote mestre tem três vertentes: orquestrador, *logging* e alarmística. Enquanto orquestrador, este pacote executa cada um dos pacotes que geram os ficheiros. Enquanto faz este processo, escreve, numa tabela de *logs* dedicada, o resultado de cada uma das execuções. Quando termina a execução dos pacotes de geração dos ficheiros, envia emails com o resultado da execução. Os pacotes de geração de ficheiros são mais simples e limitam-se a fazer uma *query* ao DW, utilizando um atributo *data* para filtrar os registos e a gravar o resultado no ficheiro. Aquando do desenvolvimento desta solução discutiu-se a possibilidade de incluir uma tabela de *staging*, entre a *query* e a geração do ficheiro. Isto traria robustez ao processo, garantiria mais flexibilidade para fazer transformações, e permitiria garantir validações para gerar o ficheiro apenas se tivessem sido devolvidas linhas. Contudo, devido à urgência, optou-se por uma abordagem mais simples e que suprime as necessidades dos intervenientes. Existe uma diferença entre o processo de geração de ficheiros para tabelas, consideradas dimensões e factuais. As tabelas de dimensão devolvem, numa base diária, as linhas ativas para esse dia. Por seu lado, os processos para as tabelas factuais geram um ficheiro com as linhas criadas no dia para o qual o ficheiro é gerado. O processo utilizado para as tabelas de dimensão comporta dois riscos, que foram sinalizados e poderão ser endereçados num futuro próximo, designadamente o tamanho do ficheiro e a capacidade necessária para o integrar, e a segurança. Ao enviar a informação completa todos os dias, implica que, no caso de ser acedida indevidamente, dará mais informação do que se fossem enviadas as alterações.

Depois de desenvolvido este processo, foi necessário testar, tendo para esse efeito sido gerados ficheiros de produção com informação mascarada. Os ficheiros foram integrados com sucesso, mas levantaram um conjunto significativo de questões funcionais, que exigiram análise e resposta. Um exemplo de uma pergunta que nos foi colocada, foi o significado de uma transação do tipo *repayment adjusted*. Outro tema que surgiu recorrentemente foram os campos a serem utilizados como chaves para ligar tabelas. Para facilitar a ligação entre as tabelas optou-se por reduzir as chaves às chaves de negócio.

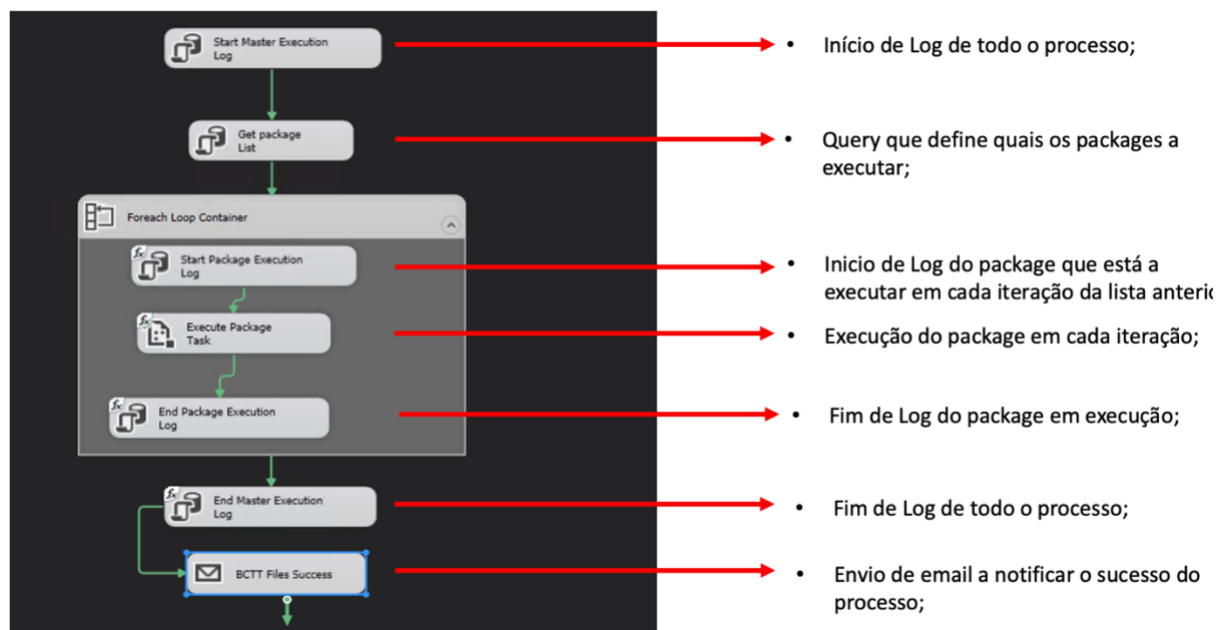


Figura 17 - Pacote mestre com explicação

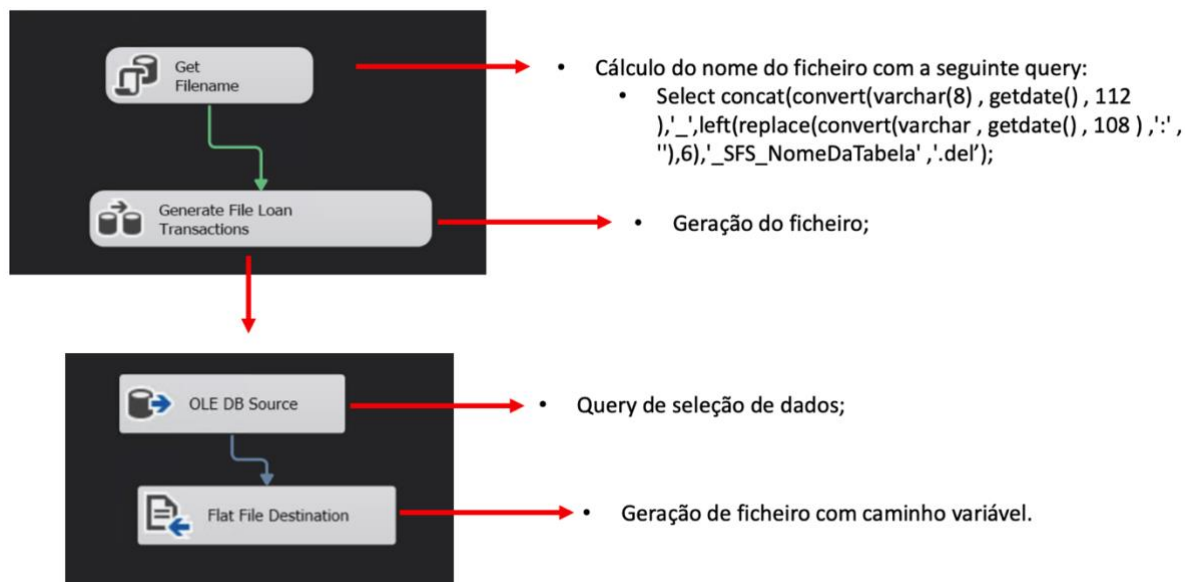


Figura 16 - Pacote de geração de ficheiro

Apesar de termos montado o processo diário de troca de informação, ficou a faltar desenvolver o processo para o envio de correções. O parceiro insistiu numa lógica de envio e

reenvio total do dia afeto. De um ponto de vista tecnológico, não existe nenhum impedimento, mas esta solução é pouco elegante e não está alinhada com as políticas de desenvolvimento internas. Assim, e como ainda não surgiu a necessidade de enviar correções, esse processo ainda não foi desenvolvido. Em relação ao envio de histórico, temos duas situações distintas, factuais e dimensões. Para as factuais, será simples enviar todas as linhas. Relativamente às dimensões, existem dúvidas se fará sentido gerar, para cada dia de histórico que desejemos integrar, a situação que vigoraria à data.

Para além das discussões relativas a pontos técnicos de estrutura e lógica dos ficheiros, existiram muitas discussões funcionais e de significado dos dados. Estas discussões tiveram uma importância transcendente às equipas que nelas tomaram parte, uma vez que serviram para que o parceiro tivesse clareza sobre o nosso produto, e em particular como este se distinguia dos restantes produtos existentes no seu portfólio. Um exemplo disto foi a clarificação de que os créditos fracionados do Cartão Universo não devem ser analisados separadamente, uma vez que o produto, por mais serviços que ofereça, é a linha de crédito renovável, e qualquer funcionalidade está assente neste serviço. Outro apoio que demos foi o de como se poderiam calcular algumas métricas. Não obstante, existiram métricas solicitadas pelo parceiro, como o somatório de eventos de liquidação antecipados que não são passíveis de ser calculados, sem um esforço desproporcional à sua utilidade.

Este foi o último projeto do estágio e colocou à prova as minhas aprendizagens relativas ao desenvolvimento de um processo de informação e de gestão de projeto. Senti que fui capaz de abordar o problema de forma mais metódica. Também desafiou o conhecimento que fui desenvolvendo relativamente aos diversos campos que integrámos e às especificidades do nosso produto. Embora pese o facto de a duração do meu estágio não ter permitido que acompanhasse a passagem a produção daquilo que desenvolvi, acho que a equipa conseguiu realizar um bom trabalho no que concerne a antecipação de pedidos por parte do parceiro. Este sucesso vem em grande medida das metodologias de trabalho *agile*.

## 7. AVALIAÇÃO DO ESTÁGIO

A avaliação do estágio é muito positiva, no sentido em que os objetivos foram cumpridos na sua totalidade. Para além disso, colocou-me em situações para lá da minha zona de conforto. Exemplos disto, foram o desenvolvimento de soluções em SQL e o desenho de modelos de dados.

Este estágio, como era esperado, complementou a componente letiva do mestrado, permitindo aplicar, num contexto real e particularmente exigente, as bases teóricas. Para além disso, foi uma oportunidade para me deparar com problemas que não existem no contexto académico como são exemplo lidar com sistemas *legacy*.

Finalmente, o balanço positivo é reforçado pela crença de que o meu desempenho contribui de forma efetiva para a resolução de problemas quotidianos e para a transição organizacional levada a cabo pela SFS. Isto foi atingido pela capacidade de trabalho, mas também pelo olhar crítico e de desafio estimulado pelo contexto pedagógico do estágio.

## 8. CONCLUSÕES

Este estágio decorreu num período crítico para o futuro da SFS, tendo-se centrado, em grande medida, nas transformações do alargamento do objeto social, e consequente aumento vertical do negócio. Ao longo desta secção, irá ser feito um resumo do âmbito do estágio, com ênfase nas aprendizagens que proporcionou.

Os projetos de integração de dados foram aqueles relacionados com a integração de informação. Esta área foi grandemente afetada pela transformação organizacional, uma vez que muitos dos processos de integração tiveram que ser transformados e existiu, também, a necessidade de desenvolver novos. Para além de me porem em contacto com um sistema de BI em contexto empresarial, estes projetos tiveram o mérito de promover a aprendizagem de novas tecnologias como API. Mais importante, permitiu-me fazer uma avaliação crítica dos vários métodos de envio e integração de informação, capacitando-me a discutir a adequação de cada um destes métodos nos seus contextos específicos. Outra capacidade que desenvolvi, foi a construção e otimização de modelos semânticos. Estes modelos constituem, no caso da SFS, uma ferramenta vital para a disseminação de informação e de cujo bom funcionamento depende a atividade do negócio. Dar este acompanhamento às equipas implicou conhecer o modelo de dados existente na empresa e desenvolver sensibilidade para os vários tipos de informação integrados em DW. Conferiu ainda a possibilidade de olhar criticamente para o manancial de informação de que dispúnhamos, sugerindo formas de o capitalizar, quer através da monetização ou do fomento da qualidade da experiência do cliente. Em suma, estes projetos deram-me uma base robusta para olhar e compreender um modelo de dados quer de uma perspetiva de negócio, quer de uma perspetiva tecnológica.

Em segundo lugar, foram apresentados os projetos de reporting. Nesta secção, explorámos o Dashboard de adesões, que foi uma experiência de um dia, mas com uma criticidade alta. Desenvolver esta solução de forma rápida, para que pudesse ser utilizada pelos diretores e administradores da empresa, foi um desafio que, embora tenha sido superado, trouxe aprendizagens. Algumas simples, relacionadas com a ligação do Excel ao modelo semântico. Outras menos palpáveis, relacionadas com a adequação da solução ao seu público alvo e a compreensão de que os requisitos normalmente não são exaustivos, pelo que é importante trabalhar de forma iterativa (se existir essa possibilidade será preferível) ou questionar o cliente da solução de forma exaustiva para que esta possa corresponder às suas expectativas. O outro projeto de reporting foi a solução de reporte à CRC. Este projeto consistia no acompanhamento da implementação da aplicação de geração de ficheiros de reporte para o BdP com foco na disponibilização de dados. Aqui, as aprendizagens podem ser divididas em três grandes áreas. As técnicas que estão na sua essência, relacionadas com a minha proficiência em SQL e DataFactgory. As de negócio, no fim de projeto fiquei a conhecer a componente funcional deste tipo de reports. E de gestão de projeto. Dentro destas, é importante reforçar a gestão de projeto per se, já que acabei por ser incumbido de

acompanhar o mesmo, desde a fase de levantamento de requisitos à passagem a produção. Para entregar este projeto, foi necessário desenvolver a capacidade de gerir os diferentes stakeholders e garantir que eram responsáveis pelos seus módulos. Isto foi, por vezes, extremamente difícil, devido à falta de definição de responsabilidades e requisitos. Após a passagem a produção, surgiram diversos problemas cuja responsabilidade pode ser imputada à arquitetura da solução, e que tiveram que ser resolvidos. Aqui, tive a oportunidade de desenvolver a capacidade de congrega os intervenientes num fórum e orientá-los para um objetivo comum. Para além disso, foi de extrema importância priorizar as diferentes ações, quer no contexto da transformação, quer no contexto de reporting (para evitar coimas).

Por fim, foram apresentados os projetos de integração de dados. Aqui, como seria de esperar, foram desenvolvidas competências na área de SQL Integration Services e na modelação da informação a ser enviada. Contudo, mais relevantes foram as aprendizagens acerca do negócio da concessão de crédito e da modelação de risco. Para poder acompanhar o desenvolvimento dos modelos de scoring e de imparidade, foi importante compreender os conceitos elementares destas áreas. Para além disso, estes projetos permitiram, de forma efetiva, integrar a informação de que dispúnhamos em modelos dos quais depende a saúde do negócio. Exemplos disto, foram a disponibilização de informação da CRC e a criação de novas variáveis a serem utilizadas pelas regressões logísticas dos modelos de scoring. Estes projetos contribuíram, assim, para o meu desenvolvimento enquanto profissional capaz de integrar as áreas técnicas de dados com as áreas de negócio, fazendo uma ponte e agilizando a comunicação.

Este estágio reforçou, em grande medida, a minha compreensão da importância de uma estratégia de gestão de conhecimento que foi reforçada pelo contexto da pandemia de COVID-19. A obrigatoriedade de teletrabalho, levou também, de forma inevitável a que desenvolvesse novas estratégias de trabalhar e estimulou a minha agilidade uma vez que tive que entrar ambiente desconhecido com recurso apenas a ferramentas digitais. Contudo, acho que o recurso a metodologias *agile* mitigou estes efeitos e facilitou a minha integração na equipa.

Foi para mim claro que é importante que exista um canal de comunicação robusto entre as equipas tecnológicas e de negócio, para que seja possível avaliar o impacto das alterações ao produto e perceber se são justificadas. Um exemplo disto, é o impacto da existência de dois titulares associados ao cartão. É ainda importante que, por muito que existam limitações tecnológicas, não sejam estas a limitar o produto nem pode recair, sobre as equipas de implementação, a responsabilidade de definir processos de negócio.

Por fim, numa nota mais filosófica, aprende-se mais facilmente com o que corre mal do que com o que corre bem, já que, quando algo corre mal, é mais fácil identificar o problema. Contudo, um espírito crítico aguçado é especialmente importante num contexto em que se utilize informação para gerar conhecimento. Nesta área, a criatividade e a discussão de ideias é de particular importância para desbloquear novas formas de gerar valor.

## 9. BIBLIOGRAFIA

- Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the operational research society*, 54(6), 627-635.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. & Wirth, R. (2000). CRISP-DM 1.0 Step-by-step data mining guide (). The CRISP-DM consortium.
- Dalkir, K. (2017). *Knowledge management in theory and practice*. MIT press.
- Darwin, C. (2017). Building a learning organization. *Knowledge Solutions*, 57.
- Dingsøyr, T., Nerur, S., Balijepally, V., & Moe, N. B. (2012). A decade of *agile* methodologies: Towards explaining *agile* software development.
- Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160(3), 523-541.
- Hann, J., & Kamber, M. (2000). Data mining: concepts and techniques.
- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.
- Linoff, G. S., & Berry, M. J. (2011). *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons.
- Liu, L., & Özsu, M. T. (Eds.). (2009). *Encyclopedia of database systems* (Vol. 6). New York, NY, USA:: Springer.
- Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization science*, 5(1), 14-37.
- Nonaka, L., Takeuchi, H., & Umemoto, K. (1996). A theory of organizational knowledge creation. *International Journal of Technology Management*, 11(7-8), 833-845.
- Omotayo, F. O. (2015). Knowledge Management as an important tool in Organisational Management: A Review of Literature. *Library Philosophy and Practice*, 1(2015), 1-23.
- Peckham, J., & Maryanski, F. (1988). Semantic data models. *ACM Computing Surveys (CSUR)*, 20(3), 153-189.
- Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, 45(4), 211-218.
- Reeve, A. (2013). *Managing data in motion: data integration best practice techniques and technologies*. Newnes.
- Schütt, P. (2003). The post-Nonaka knowledge management. *J. UCS*, 9(6), 451-462.
- Senge, P. M. (2014). *The fifth discipline fieldbook: Strategies and tools for building a learning organization*. Currency.
- Shafique, U., & Qaiser, H. (2014). A comparative study of data mining process models (KDD, CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*, 12(1), 217-222.

Sharda, R., Delen, D., & Turban, E. (2016). *Business intelligence, analytics, and data science: a managerial perspective*. Pearson.

Thomas, L., Crook, J., & Edelman, D. (2017). *Credit scoring and its applications*. Society for industrial and Applied Mathematics.

Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, 39(11), 86-95.

Wang, R. Y., Ziad, M., & Lee, Y. W. (2006). *Data quality* (Vol. 23). Springer Science & Business Media.

