

Aplicação de Técnicas de Business Intelligence a Base de Dados Prosopográficas (Análise Temporal)

Adão Baptista Pereira Lopes¹, Carlos Pampulim Caldeira²

¹University of Évora, School of Science and Technology, Master in Computer Engineering, Portugal.
E-mail: m34729@alunos.uevora.pt.

²University of Évora, School of Science and Technology, IT Department, Portugal.
E-mail: ccaldeira@di.uevora.pt

Resumo

A análise multidimensional de informação e a descoberta de padrões desconhecidos em dados são um vetor estruturante dos sistemas modernos de análise de dados. Este trabalho pretende, em termos globais, apresentar um conjunto de técnicas e métodos de Business Intelligence que aplicadas a uma base de dados prosopográfica – o sistema SPARES¹ – melhorem a forma atual de exploração de dados que é essencialmente baseada em filtros avulsos.

A análise temporal é de vital importância para Data Warehousing. A maior quantidade de problemas que surgem na resposta a dar, na análise de dados, aos utilizadores de Data Warehousing, nomeadamente aos gestores empresariais, está relacionada com a falta de uma dimensão “DATA”. Geralmente, a maioria de perguntas para tomada de decisões, tem como restrição o tempo (data), podendo-se então afirmar sem margem de dúvida que é um dos recursos mais valioso no modelo dimensional.

Feito este enquadramento, surge então a questão cerne do nosso objeto de estudo: **será possível ter uma dimensão temporal, de acordo com o ambiente de exploração de dados do sistema SPARES?** Isto é de grande importância tendo em conta a multiplicidade de vetores que definem uma data em dados históricos entre os séculos XVI e XVIII. Paralelamente ao objetivo de otimizar e operacionalizar a análise de dados de relações e eventos com o suporte em informação prosopográfica, o trabalho incide particularmente na construção de uma aplicação do tipo Cubo Multidimensional, com a finalidade de melhorar as análises e explorações de dados no sistema SPARES.

Palavras-chave: Data Warehousing, Prosopografia, Cubo Multidimensional, Business Intelligence, Análise de Dados

¹ SPARES – Sistema Prosopográfico de Análises de Relações e Eventos Sociais.

Abstract

Abstract The multi-dimensional information analysis and the discovery of unknown patterns in data are a common trend in modern data analysis systems. This work intends, overall, to present a set of business intelligence techniques and methods that applied to a prosopographic database - the SPARES² system - can improve the current form of data exploration which is essentially based on single filters.

Some of the major drawbacks in data analysis with support on data warehouses are the lack of a date dimension table designed with a proper and extensive temporal hierarchy. Generally, the majority of questions that decision makers have is strongly associated with a date (time) constraint, and it can then be argued beyond doubt that it is one of the most valuable features in the dimensional model.

Once this framework is in place, the main question of our study arises: will it be possible to have a temporal dimension, according to the data exploration environment of the SPARES system? The application of a well designed temporal framework is of the major importance in view of the multiplicity of vectors that define a date in historical data between the sixteenth and eighteenth centuries. Parallel to the objective of optimizing and operationalizing the data analysis on relations and events with the support in prosopographic information, the work focuses particularly on the construction of a cube type application, with the purpose of improving the analysis and exploitation of data in the SPARES system.

Keywords: Data Warehousing, Prosopography, Temporal Analysis, Business Intelligence, Data Analysis.

² SPARES – Sistema Prosopográfico de Análises de Relações e Eventos Sociais (Prosopographic System of Analysis of Relations and Social Events).

1. INTRODUÇÃO

Atualmente, nas empresas e organizações, estamos cientes do papel preponderante que as bases de dados assumem a nível de gestão e estratégia empresarial e organizacional, enquanto vetores essenciais para a necessidade de dotar o mundo empresarial e as organizações, independentemente da sua tipologia, de informação que habilite os distintos profissionais a tomarem decisões consentâneas com as exigências de obtenção de excelência da gestão, eficiência operacional e posicionamento estratégico. O certo é que a factualidade objetiva dessa necessidade, aliada à enorme quantidade e proliferação de dados armazenada que as organizações detêm, não possibilitam por si só, simplicidade na análise de dados, dificultando sobremaneira a obtenção da eficácia e a eficiência exigíveis na prossecução dos seus objetivos.

No processo decisório, além da experiência dos gestores, é então crucial dispor de uma enorme quantidade de dados fidedignos, informações de qualidade e conhecimento, que os auxiliem nessa árdua tarefa de tomada de decisões. *“O acesso eficiente a estas informações pode significar a sobrevivência ou falência de uma organização”*[3]. Enquadrada nesta problemática, surgiu a motivação subjacente a este trabalho e que tem como finalidade específica o estudo, aplicações de técnicas de Business Intelligence a base de dados prosopográficos, para oferecer um contributo de relevo aos investigadores de Centro Interdisciplinar de História, Culturas e Sociedades, nas análises de dados sobre os comissários do Santo Ofício na Idade Moderna em Portugal.

A organização em causa, possui um repositório de dados com suporte prosopográfico, e identificou a falta de mecanismos e ferramentas potentes que possibilitem uma maior acessibilidade e exploração dos dados. É neste contexto que se propõe uma solução, que consiste basicamente em criar um aplicativo tipo cubo para análise de dados, permitindo a obtenção de informações úteis e de melhor qualidade, levando inclusivamente à descoberta de novos padrões nos dados com a nova dimensão temporal.

Resumidamente, este projeto tem como finalidade e objetivo primordial o estudo de implementação de uma dimensão temporal, e a investigação para aplicações de técnicas de business Intelligence, sobre repositórios de dados com um suporte em informação prosopográfica, de modo a otimizar e operacionalizar de modo mais dinâmico a análise de dados de relações e eventos, entre os diversos atores intervenientes num determinado contexto social, e de modo a que informações neles contida possa ser acessível e de fácil compreensão, urge aplicar as técnicas de Business Intelligence.

2. PROSOPOGRAFIA

Não existe uma datação concreta para o aparecimento da prosopografia. Apesar da sua longa história, o primeiro uso conhecido da Prosopografia remonta à data de 1743 [7]. Sabe-se no entanto, que já no século XVI a prosopografia tinha sido muito aplicada nos períodos da história antiga e medieval, tendo nos últimos anos esse método sido

trabalhado na História Moderna e Contemporânea. “*Esse método tem uma base simples, que consiste em definir uma população a partir de um ou alguns critérios e estabelecer assim uma descrição bibliográfica cujas nuances possibilitarão traçar um perfil da sua dinâmica social, privada, pública, cultural, ideológica ou política* [1]. Para os historiadores da antiguidade, a prosopografia era considerada uma ciência prestadora de auxílio à história, tendo como objetivo primordial, estudar as biografias dos constituintes dos diferentes grupos das elites sociais, ou políticas de uma determinada sociedade.

Neste projeto, a prosopografia tem um papel decisivo, dado que é o suporte informacional dos dados do sistema SPARES. O tema fulcral para exploração e análise de dados, é a atuação dos comissários do Santo Ofício ou Inquisição do Santo Ofício em Portugal. Na Idade Moderna, surgiu o chamado Santo Ofício, instituição judicial que tinha a responsabilidade de julgar, perseguir e punir todas as pessoas consideradas como hereges ou heterodoxos, ou seja, pessoas que não estavam de acordo com as normas ou práticas impostas pela Igreja Católica.

De acordo com Olival (2012, p. 478), “*para ter segurança e uma certa honra e dignidade em termos gerais, era essencial ingressar no Santo Ofício ou nas Ordens dos Militares*”. Para esse efeito, era condicionante possuir-se determinadas habilitações que tipificavam a qualidade dos que queriam vincular-se ao Santo Ofício e que tinham uma rede de Comissários. Essa rede é que constitui a fonte informacional do sistema SPARES, sendo deste modo premissa fundamental familiarizarmo-nos com o modo de atuação dos comissários, que funcionam especificamente como auxílio na compreensão de análise de dados, e descoberta de novos padrões de dados, bem como respetivos familiares do Santo Ofício, que eram pessoas que tinham vínculo com a Inquisição, colocando os seus serviços ao dispor dos inquisidores, coadjuvando-os no exercício dos seus ofícios.

“*A historiografia tem admitido que a rede de comissários do Santo Ofício ter-se-ia iniciado no final da década de 1570 ou na seguinte*”[8]. Em Julho de 1575³ assistiu-se à nomeação do primeiro familiar para a Inquisição em Évora, mas a designação do primeiro comissário ocorreu apenas em 1586. Para o exercício desse cargo, não era exigível qualquer formação específica, pois consoante a prática da função, e a sua interação com o Tribunal do Santo Ofício, o comissário aumentaria paulatinamente o seu conhecimento e formação.

Face ao exposto, é de suma importância compreendermos todo o funcionamento da rede de comissários, as suas respetivas ocupações, sua árvore genealógica, entre outros dados, de modo a facilitar a compreensão do sistema SPARES num todo, dado que essa rede é, conforme supra referido, o suporte informacional do SPARES.

³ Tratou-se de Jerónimo de Torres, pedreiro, morador em Évora, abaixo do Chão das Covas, “mestre na obra de Sua Alteza”, o Inquisidor-Geral, nomeado “pello tempo que parecer soamente, e enquanto não mandarmos o contrajro” – ANTT, Inquisição de Évora, L^o 146, fl. 139v [8].

3. BUSINESS INTELLIGENCE

«Obter dados certos, no momento certo, para tomar a decisão certa, para a organização»

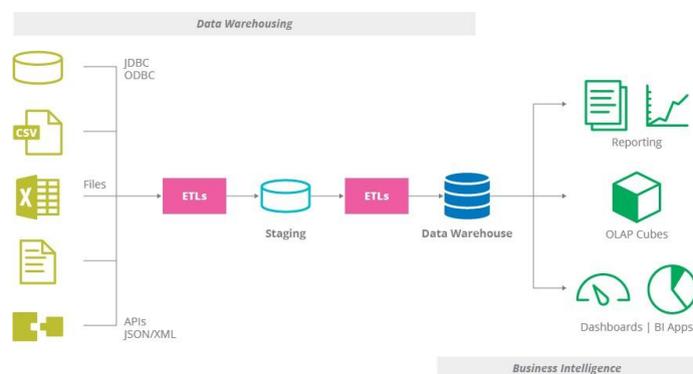
Autor Desconhecido

Em meados dos anos de 1990 surge o termo Business Intelligence⁴ por Howard Dresner do Gartner Group⁵. Somente por volta de 2005 os sistemas de Business Intelligence passaram a incluir e ser dotados de poderosos recursos e ferramentas, de tratamentos e análise de dados. São esses recursos que aplicaremos no repositório de dados prosopográfico, de modo a simplificar o tratamento, exploração e análise de dados no sistema SPARES.

Atualmente as empresas e as organizações litigam cada dia mais a rapidez e diligência, os riscos e as oportunidades do mercado, bem como o aperfeiçoamento e melhoria contínua no feedback às necessidades dos clientes. Deste modo, as ferramentas de Business Intelligence exercem um papel vital, transformando enormes quantidades de dados armazenados, em sistemas de suporte para tomadas de decisões, independentemente da sua grandeza, o que de per si constitui um inegável recurso a nível de competitividade e sustentabilidade empresarial.

A Business Intelligence deverá assim, ser considerada no âmbito deste trabalho como um conjunto de técnicas, metodologias e conceitos, implementados através de ferramentas de software, orientadas para a análise de dados. Tal permitirá a disponibilização de informações relevantes de forma compreensível e transparente, proporcionando e apoiando os gestores nas tomadas de decisões eficazes, correndo assim o menor risco possível.

Normalmente, um sistema típico de Business Intelligence é composto pelas seguintes



partes: Fontes de dados, ETL, Data warehouse e Análise de dados, como é demonstrado na Figura 1.

Figura 1. Arquitetura de apoio ao processo de BI. Fonte: [6].

Na Figura 1 é apresentado, de forma simplificada e compreensível, cada componente integrante da arquitetura de um processo de sistema de Business Intelligence. Podemos observar várias etapas imprescindíveis a seguir para implementação desse sistema.

⁴ BI é entendido como o processo de recolha, e tratamento de informações com objetivo de apoio a gestão de um negócio.

⁵ Gartner é detentora da paternidade do termo Business Intelligence. Consultora de pesquisas de mercado na área das Tecnologias da Informação.

4. CUBO MULTIDIMENSIONAL (OLAP)

A nível de contextualização histórica, foi nos finais da década de 90, que se verificou o aparecimento dos cubos analíticos (OLAP). Atualmente, a Microsoft, IBM e Oracle são as empresas mais conhecidas por produzirem ferramentas para OLAP. De acordo com Kimball e Ross (2013, p. 9) na construção de um aplicativo tipo cubo analítico, é necessário que os dados para exploração, tenham o suporte em **Esquema em Estrela**, ou seja, um esquema constituído por uma tabela de facto, que está associada a várias dimensões. Através da construção de um Cubo semelhante ao Figura 2, os utilizadores do sistema SPARES, vão poder usufruir de uma nova ferramenta de análise de dados, poderosa e dinâmica, que lhes permitirá manipular os dados, e terem à sua disposição o poder de alterarem as arestas e os segmentos do cubo de acordo com as suas necessidades.

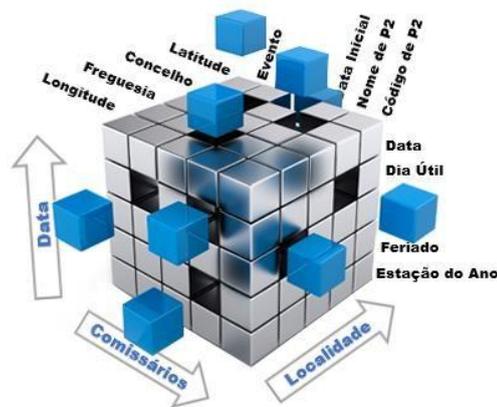


Figura 2. Aplicação do tipo Cubo para o Sistema SPARES

De modo a familiarizarmo-nos com a arquitetura do cubo multidimensional, é fundamental que se compreendam algumas expressões relacionados com a mesma, entre as quais se destacam:

- **Dimensão** - observando a Figura 2, podemos defini-la como sendo uma face do cubo. No sistema SPARES podemos mencionar «localidade», «data», «ocupação» como exemplos;
- **Medidas** - são as parametrizações usadas no processo decisório, sendo conotadas normalmente com dados numéricos que permitem realizar comparações, podendo-se mencionar a título de exemplo as «quantidades» e «médias»;
- **Hierarquia** - é a estruturação em árvore, tendo como função primordial, a organização dos membros de uma determinada dimensão;
- **Nível** - dentro de uma hierarquia, as informações podem estar estruturadas de acordo com a granularidade, ou seja, com os seus níveis de detalhe, como por exemplo na dimensão Data em que há o Ano, Semestre, Trimestre, Mês, Semana e Dia.

As principais operações que manipulam o cubo são: as operações **Roll-Up** e **Drill-Down** que aumentam ou diminuem o nível de granularidade, e as operações **Slice** e **Dice** aplicam filtros sobre o mesmo cubo. A aplicação de tipo Cubo Multidimensional, escolhido para as explorações e análise de dados no sistema SPARES, é o **Tableau**. Foi reconhecido

como líder no Quadrante Mágico Gartner para plataformas de Business Intelligence. Uma das razões mais importantes de termos escolhido o Tableau é porque oferecem a licença gratuita para estudantes e professores. Dado que todos os utilizadores do sistema SPARES, são professores ou funcionários da Universidade de Évora, têm direito a chave de ativação do Tableau, com todas as funcionalidades.

5. PROCESSO ETL



Figura 3. Modelo da área de extracção, transformação e carregamento do DW

Citando Caldeira (2012, p. 205), o processo de ETL é entendido como “*subsistema de extracção, transformação e limpeza dos dados dos sistemas operacionais no momento da sua disponibilização no data warehouse.*” O mesmo autor refere que a implementação de um sistema de BI, é um dos processos com maior complexidade, e durabilidade, podendo ocupar até 80% do trabalho. Tem como objetivo primordial, a extracção dos dados nos sistemas transacionais, transportando-os para a área de estágio, onde ocorrerá a sua transformação e posteriormente o seu carregamento no DW[2]. A Figura 3 é representativa do processo de um sistema ETL, onde são processadas enormes quantidades de dados, tornando-se por esse motivo, um processo mais crítico na construção de um sistema de Business Intelligence.

6. CASO DE ESTUDO - A DIMENSÃO TEMPORAL

O sistema prosopográfico de análise de relações e eventos desempenha duas funções: é um sistema transaccional e também é um sistema analítico. O SPARES serve para registar as transacções sobre as relações e eventos relacionados com a atuação dos comissários do Santo Ofício, contudo, tem também a serventia para análise dos respetivos dados. Dado a isso, é complexo de gerir, processar e analisar os grandes volumes de dados armazenados, este projeto está focado especificamente na aplicabilidade de técnicas de Business Intelligence sobre os dados prosopográficos.

Atualmente a forma de explorar os dados no sistema prosopográfico de análise de relações e eventos é através de filtros avulsos. É uma ferramenta com muito pouca potência para analisar os dados. Não permite realizar agregações de dados, não permite realizar consultas complexas onde constam relações entre tabelas, só permitindo realizar segmentações básicas. Os filtros não são nada mais, nada menos do que um simples «WHERE» no comando SQL. Tem como suas vantagens a possibilidade de ser considerada para uma consulta rápida, mas não como uma ferramenta para pesquisas e análises de dados, pelo facto de não ter capacidades de descobrir novos padrões de dados, e não é uma ferramenta para auxiliar os gestores nas tomadas de decisões, dado que é impossível gerar informações a partir dos dados que possam ser úteis pra tomadas de decisão.

Foi esta facticidade que fez com que nascesse esse projeto de estudo, de melhorar a forma atual de análise de dados no sistema SPARES. É então, nesse contexto, que surgem as técnicas e ferramentas de Business intelligence para resolver o grande problema do sistema SPARES, que é oferecer outra face (dimensão temporal) ao sistema no âmbito de análise e

exploração de dados. Todas as dificuldades que os investigadores tinham para analisarem os dados e descobrirem novos padrões de dados temporais, foi substancialmente melhorado com o aplicativo tipo cubo, pois torna-se tudo muito intuitivo e oferece maior qualidade de dados nas suas análises e apresentação de reporting de dados com diversos tipos de visualização.

6.1. A DIMENSÃO DATA

De acordo com Caldeira (2012, p.110), “*As datas são uma grandeza fundamental em todos os negócios e organizações. (...) A dimensão Data vai servir de exemplo para o estudo pormenorizado da estrutura hierárquica dos dados que se encontra frequentemente no interior das dimensões. O desenho desta dimensão tem que ser devidamente avaliado de acordo com as utilizações previstas*”. A dimensão Data é na prática a mais utilizada das dimensões usadas na maioria dos sistemas de Business Intelligence. Prova disso, encontramos na citação de Caldeira (2012, p.110) (...), de acordo com o qual “*Em todos os tipos de data marts é comum a presença de uma dimensão data composta por múltiplos atributos que descrevem cada uma das ocorrências da tabela*”. Apesar de ter a mesma analogia em todos os sistemas, não é composta pelos mesmos atributos, pois elas podem ser adaptadas de acordo com a necessidade do projeto envolvente.

O encaixe de uma dimensão Data no modelo dimensional é o caso principal do estudo deste projeto, dado que a sua existência é de grande relevância. O desafio centra-se em dar resposta à seguinte questão: **Será possível a concepção da dimensão Data e o seu enquadramento no sistema de Business Intelligence SPARES orientado para a análise prosopográfica do Santo Ofício entre os séculos XVI e XVIII?** No caso de ser confirmada essa possibilidade, trará grandes vantagens ao modelo de dados dimensional, fazendo com que seja extremamente simples, intuitivo e flexível analisar os dados de acordo com os segmentos temporais, o que permitirá aos utilizadores descobrirem e identificarem mais facilmente novos padrões de informação. A concepção dessa dimensão foi adaptada consoante a natureza do sistema SPARES.

Fisicamente criou-se uma tabela, denominada de “data”. A sua chave primária é o atributo “Código da Data”, que é um campo numérico e automático sem nenhum significado, a sua granularidade⁶ é o dia, ou seja, os factos são agregados por dia. Não podemos deixar de referir que é caracterizada pela sua desnormalização, possibilitando, apesar disso, um melhor desempenho na análise efetuada pelos utilizadores [5].

⁶ “Nível de detalhe com que os dados são medidos e armazenados. Quanto maior a granularidade menor o nível de detalhe e vice-versa” [2].

Tabela 1. A Dimensão Data

| Data | | Exemplo |
|------|-----------------------------|-------------------------|
| PK | Código da data | 1 |
| | Data | 21/09/2016 |
| | Descrição completa da data | 21 de Setembro de 2016 |
| | Descrição abreviada da data | 21 de Set de 2016 |
| | Ano em número | 2016 |
| | Mês em número | 9 |
| | Dia em número | 21 |
| | Dia de Semana em número | 4 |
| | Semestre em número | 2 |
| | Nome de semestre | Segundo Semestre |
| | Trimestre em número | 3 |
| | Nome de trimestre | Terceiro Trimestre |
| | Ano bissexto | É ano bissexto |
| | Nome do mês | Setembro |
| | Abreviatura do mês | Set |
| | Nome de dia | Quarta Feira |
| | Abreviatura do dia | Qua |
| | Século em Número | 21 |
| | Século em Numeração Romana | XXI |
| | Década em Número | 20 |
| | Década por Extenso | Década 20 do Século XXI |
| | Número da semana no ano | 43 |
| | Numero da semana do mês | 3 |
| | Estação do ano | Verão |
| | Final de semana | Não é final de semana |
| | Feriado | Não É Feriado |
| | Nome do Feriado | Null |
| | Dia útil | É dia útil |
| | SQL-data | 21/09/2016 |

Na Tabela 1 podemos constatar o exemplo da dimensão Data deste projeto e respetivos atributos. A dimensão Data contém diversos atributos. Podemos citar o campo de Descrição Completa da Data “Quarta-Feira, 21 de Setembro de 2016”, e também pode ser definida de forma abreviada “Qua, 21 de Set de 2016”. Segundo Caldeira (2012, p.111) “Para cada um dos dias é possível ainda acrescentar uma série de indicadores”, e de acordo com a indicação adicionamos os seguintes atributos: O Nome de Dia, na forma «Domingo, ...Sábado»; a Abreviatura do Dia na forma «Dom,...Sab»; O Dia da Semana em número começando em «1, ...7» sabendo que 1 corresponde ao domingo e 7 a sábado; o Nome do Semestre e Trimestre em número e por extenso; O Ano em Número e o Ano se é um Ano Bissexto; O Nome do Mês « Janeiro, ...Dezembro», e suas respetivas abreviaturas «Jan, ...Dez»; Se é Final de Semana; Se é Feriado e o Nome do Feriado; Se é Dia Útil. Estes últimos indicadores são exemplos de atributos a que se poderia responder com um “S”, “Sim” e “N”, “Não” . Portanto, de acordo com Caldeira (2012, p.112)” para manter o máximo de legibilidade no data warehouse, os valores nas colunas têm que ser sempre explícitos; no caso do indicador de fim de semana: «É fim de semana» ou, caso contrário, «Não é fim de semana».

Como o sistema SPARES tem como fonte os dados prosopográficos, consideramos que os casos de estudo são de séculos passados, tendo um atributo de Século e Década em número e por extenso. Existem igualmente outros indicadores particulares, como a estação do ano,

caracterizado como «verão, inverno, outono e primavera».

Nos sistemas transacionais, segundo Caldeira (2012, p.113), “as datas são comumente registradas em campos cujo domínio é um tipo especial denominado SQL-data. Essas datas são a base para todos os cálculos cujos, através do sistema ETL, (...) valores SQL-data vão passar para o data warehouse para um atributo na dimensão Data,” com o formato de «date».

APLICAÇÃO DA DIMENSÃO DATA A UM AMBIENTE PROSOPOGRÁFICO

No sistema de BI, as datas são geralmente conhecidas. Exemplo disso, é o que acontece no supermercado com a venda de um produto, processada numa data exata. Mas o caso da prosopografia é especial, dado que nem sempre conhecemos a data exata. A título de exemplo podemos referir, o caso em que não se sabe a data concreta, mas sabe-se que o evento aconteceu após essa data. Essa dimensão trará grandes vantagens ao modelo dimensional, tornando-o extremamente simples, intuitivo e flexível, uma vez que permite analisar os dados de acordo com os segmentos temporais. No caso do sistema SPARES, a dimensão Data, como de modo similar na grande maioria dos casos, não indica a data concreta do evento, tendo então sido criado um sistema de data da seguinte forma:

- Quando a data for apresentada sob a forma “2016=09=21”, sabe-se que corresponde à data exata;
- Quando a data for apresentada sob a forma “2016<09<21”, não se sabe a data concreta, mas sabe-se que o acontecimento ocorreu antes dessa data;
- Quando a data for apresentada sob a forma “2016<09<00”, não se sabe o dia concreto, nem a data concreta, mas sabe-se que o evento ocorreu antes de setembro de 2016;
- Quando a data for apresentada sob a forma “2016<00<00”, não se sabe o dia nem o mês concreto, nem a data concreta, mas sabe-se que o evento ocorreu antes do ano de 2016.
- Quando a data for apresentada sob a forma “2016>09>21”, não se sabe a data concreta, mas sabe-se que o acontecimento ocorreu depois dessa data;
- Quando a data for apresentada sob a forma “2016>09>00”, não se sabe o dia concreto, nem a data concreta, mas sabe-se que o evento ocorreu depois de setembro de 2016;
- Quando a data for apresentada sob a forma “2016>00>00”, não se sabe o dia nem o mês concreto, nem a data concreta, mas sabe-se que o evento ocorreu depois do ano de 2016.
- Quando a data for apresentada sob a forma “2016-09-21”, não se sabe a data concreta, mas sabe-se que o acontecimento ocorreu por volta dessa data;
- Quando a data for apresentada sob a forma “2016-09-00”, não se sabe o dia concreto, nem a data concreta, mas sabe-se que o evento ocorreu por volta de setembro de 2016;
- Quando a data for apresentada sob a forma “2016-00-00”, não se sabe o dia nem o mês concreto, nem a data concreta, mas sabe-se que o evento ocorreu por volta do ano de 2016.

Por esta razão, o desenho da dimensão Data foi devidamente avaliado e projetado. Após a construção física da dimensão, foi programada uma função, que recebe dois argumentos formais (data inicial e data final). Portanto, efetua o cálculo entre esse intervalo de datas, incrementando o grão da dimensão Data, e em seguida efetua a inserção na dimensão Data. A título de exemplo, utilizamos um intervalo de datas entre '1500/01/01' e '1900/12/31'.

Na Tabela 2 indicam-se alguns exemplos de dados referentes a data. Primeiro regista a data concreta preenchendo todos os campos; a seguir regista a mesma data, mas com um pormenor diferente, pois na descrição completa da data é introduzido “Sabe-se que é antes, 1 de janeiro de 1500”; a utilidade desse registo é verificada no momento em que não sabemos a data concreta, mas sabemos que o evento se realizou antes dessa data. A outra condição é verificar sempre que é início do ano, e introduz um registo “Sabe-se que é antes de 1500”, e por último, no início de cada mês é inserido um registo “Sabe-se que é antes de, Janeiro de 1500”, consequentemente, por esse motivo, nem todos os atributos da data poderão ser completados, uma vez que é desconhecido o dia e o mês, ou é desconhecido o dia.

Tabela 2. Vista parcial de dados na dimensão Data

| Código da Data | Data | Descrição da Data | Dia em Número | Mês em Número | Nome de Semestre | Ano Bissexto | Estação do Ano | SQL-Data |
|----------------|------------|--------------------------------------------------|---------------|---------------|-------------------|----------------|----------------|------------|
| 1 | 1500-01-01 | Segunda-Feira, 1 de Janeiro de 1500 | 1 | 1 | Primeiro Semestre | É Ano Bissexto | Inverno | 1500-01-01 |
| 2 | 1500-01-01 | Sabe-se que é antes de, 1 de Janeiro de 1500 | 1 | 1 | Primeiro Semestre | É Ano Bissexto | Inverno | 1500-01-01 |
| 3 | 1500-01-01 | Sabe-se que é depois de, 1 de Janeiro de 1500 | 1 | 1 | Primeiro Semestre | É Ano Bissexto | Inverno | 1500-01-01 |
| 4 | 1500-01-01 | Sabe-se que é por volta de, 1 de Janeiro de 1500 | 1 | 1 | Primeiro Semestre | É Ano Bissexto | Inverno | 1500-01-01 |
| 5 | 1500-01-00 | Sabe-se que é antes de, Janeiro de 1500 | NULL | 1 | Primeiro Semestre | É Ano Bissexto | NULL | NULL |
| 6 | 1500-01-00 | Sabe-se que é depois de, Janeiro de 1500 | NULL | 1 | Primeiro Semestre | É Ano Bissexto | NULL | NULL |
| 7 | 1500-01-00 | Sabe-se que é por volta de, Janeiro de 1500 | NULL | 1 | Primeiro Semestre | É Ano Bissexto | NULL | NULL |
| 8 | 1500-00-00 | Sabe-se que é antes do Ano, 1500 | NULL | NULL | NULL | É Ano Bissexto | NULL | NULL |
| 9 | 1500-00-00 | Sabe-se que é depois do Ano, 1500 | NULL | NULL | NULL | É Ano Bissexto | NULL | NULL |
| 10 | 1500-00-00 | Sabe-se que é por volta do Ano, 1500 | NULL | NULL | NULL | É Ano Bissexto | NULL | NULL |

METODOLOGIAS APLICADA NA INSERÇÃO DE DADOS NA DIMENSÃO DATA

Para estabelecer a ligação entre a dimensão Data e a tabela de facto comissários, que segundo Caldeira (2012, p.67), “é o elemento principal no modelo dimensional”, foi necessário criar dois atributos (Código da Data Inicial e Código da Data Final) na tabela de facto, do tipo inteiro, que são chaves estrangeiras ligadas com a dimensão Data. A existência dessas duas novas colunas na tabela de facto, deriva do facto de que cada relação ou evento tem uma data de início e uma data de terminos, mas na maioria das vezes uma ou outra é desconhecida. Por esta razão, existe uma *surrogate key* que resolve os casos em que não existe um valor para as datas, quer seja inicial ou final, como podemos constatar na *Figura 4*.

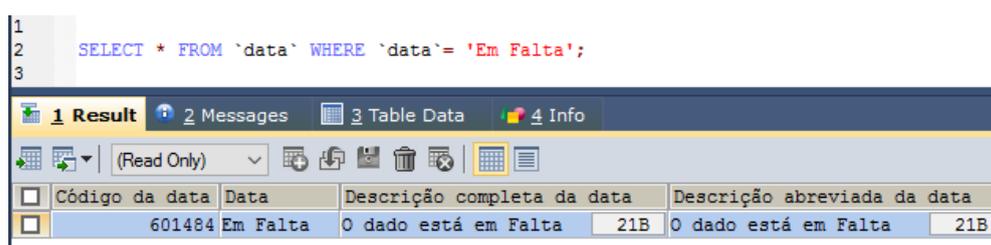


Figura 4. A utilização de surrogate key na dimensão Data

Este surrogate keys indicado na Figura 4, soluciona os eventos onde não há um valor conhecido da data. Contudo, podemos observar a chave primária da dimensão Data que é homogéneo das chaves das outras dimensões, e correspondente à utilização de chave artificial. De acordo com Caldeira (2012, p.116) “diferente das outras dimensões, a dimensão data tem uma particularidade, que é a possibilidade da chave artificial poder ter um significado concreto”, mas de acordo com o nosso caso de estudo tal facto não seria possível, dado que podemos obter conflitos com a integração dos dados, devido à necessidade de registar a mesma data quatro vezes, unicamente com a diferença que se observa Figura 5.

O primeiro registo (Código da data = 1) representa a data concreta, ou seja, o evento aconteceu nessa data; o segundo registo (Código da data = 2) corresponde a um evento que não se sabe a data exata do acontecimento, mas temos o conhecimento que tal evento aconteceu antes dessa data; o terceiro registo (Código da data = 3), da mesma forma que o segundo registo, não tem data conhecida, mas sabe-se que o tal evento aconteceu depois dessa data, e o quarto registo (Código da data = 4), sabe-se que o evento aconteceu por volta dessa data, mas não se sabe se foi antes ou depois da data.

| Código da data | Data | Descrição completa da data | Descrição abreviada da data |
|----------------|------------|--------------------------------------------------|--------------------------------------------------|
| 1 | 1500-01-01 | Segunda-Feira, 1 de Janeiro de 1500 | 35B Seg, 1 de Jan de 1500 |
| 2 | 1500-01-01 | Sabe-se que é antes de, 1 de Janeiro de 1500 | 45B Sabe-se que é antes de, 1 de Jan de 1500 |
| 3 | 1500-01-01 | Sabe-se que é depois de, 1 de Janeiro de 1500 | 46B Sabe-se que é depois de, 1 de Jan de 1500 |
| 4 | 1500-01-01 | Sabe-se que é por volta de, 1 de Janeiro de 1500 | 49B Sabe-se que é por volta de, 1 de Jan de 1500 |

Figura 5. Exemplo de registos para a dimensão Data

A dimensão data traz uma nova face ao sistema SPARES, proporcionando as análises de relações e eventos através do horizonte temporal. Segundo Caldeira (2012, p. 138), “*no desenvolvimento de um modelo dimensional surgem frequentemente múltiplas datas que funcionam como marcas temporais importantes em quase todos os processos de negócio de uma organização*”. É o caso presente no sistema prosopográfico de análise de relações e eventos, sendo a data neste sistema tratada de forma diferente de outros sistemas. As datas do acontecimento são colocadas na tabela de factos e funcionam como chaves estrangeiras da dimensão, ao contrário das demais datas que são simplesmente registadas nas dimensões como um campo.

7. CONCLUSÃO

Pode-se, em jeito de conclusão, afirmar categoricamente que o BI tem como objetivo principal transformar grandes quantidades de dados em informações de qualidade, para a tomada de decisão, de modo a possibilitar uma visão sistémica do negócio e auxiliar na distribuição uniforme dos dados entre os analistas de negócios. As aplicações de BI podem auxiliar em vários segmentos das organizações, sendo um auxílio valioso nas tendências e garantia da sustentabilidade e criação de valor das organizações. É importante ressaltar que as soluções de BI acompanham as necessidades da organização, evoluindo paulatinamente, conforme vão surgindo novos recursos e desafios. O investimento em BI por parte das organizações é incentivado pela necessidade de criação de um ambiente proativo para a tomada de decisão.

Este trabalho permitiu por um lado, apresentar efetivamente um conjunto de técnicas e métodos de Business Intelligence que aplicados a uma base de dados prosopográfica – o sistema SPARES – permite efetivamente melhorar a forma atual de exploração de dados que é essencialmente baseada em filtros avulsos, o que se traduz num recurso importante para os investigadores. Por outro lado, o seu desenvolvimento, permitiu a implementação de uma dimensão temporal no sistema SPARES, dando assim uma nova face ao aplicativo do tipo cubo.

8. REFERÊNCIAS

- [1] Almeida, C. (2011). *A prosopografia ou biografia coletiva: limites, desafios e possibilidades*. Anais do XXVI Simpósio Nacional de História – ANPUH, São Paulo. Disponível em: http://www.snh2011.anpuh.org/resources/anais/14/1300892678_ARQUIVO_anpuhsp2011.pdf
- [2] Caldeira, C. (2012). *Data Warehousing - Conceitos e Modelos*. Lisboa: Edições Sílabo, Lda.
- [3] Dalfovo, O., & Tamborlim, N. (2017). *Business Intilligence. Estudos e casos: na Gestão de Tecnologia da Informação como Inteligência nos Negócios*. 1ª edição. Blumenau: Edição Clube de Autores.
- [4] Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. 3rd Edition. Indianapolis: Wiley.
- [5] Microsoft. (2010). *A concepção da dimensão data num sistema de BI - Parte I*. Consultado em 13 de Outubro de 2017. Disponível em: <https://support.microsoft.com/pt-pt/help/2293691>.
- [6] Murillo, M. (2016). *Agile Data Warehousing and Business Intelligence in Action*. Consultado em 14 de Outubro de 2017. Disponível em: <https://www.thoughtworks.com/insights/blog/agile-data-warehousing-and-business-intelligence-action>.
- [7] Nicolet, C. (1970). *Prosopographie et histoire sociale: Rome et Italie à l'époque républicaine. Annales: économie, sociétés, civilisations*. Paris, v. 25, n. 5, p. 1209-1228. Consultado em 14 de Outubro de 2017. Disponível em: <http://www.persee.fr/doc/ah>.
- [8] Olival, F. (2012). *Comissários das Ordens Militares e comissários do Santo Ofício em Portugal: dois modelos de actuação*. As Ordens Militares: freires, guerreiros, cavaleiros: actas do VI Encontro sobre Ordens Militares, coord. Isabel Cristina Fernandes, Vol. I, Palmela: Município de Palmela / GESOS. Disponível em: <http://dspace.uevora.pt/rdpc/handle/10174/7479>.
- [9] Santos, M., & Ramos, I. (2006). *Business Intelligence : tecnologias da informação na gestão de conhecimento*. FCA - Editora de Informática.