

<https://doi.org/10.34632/gestaoedesenvolvimento.2023.11868>

Data de receção: 25/02/2022

Data de aceitação: 14/02/2023

IMPLEMENTAÇÃO DE UM DATA WAREHOUSE, UMA PERSPETIVA BIBLIOMÉTRICA

DATA WAREHOUSE IMPLEMENTATION, A BIBLIOMETRICAL PERSPECTIVE

Joana Lunet¹ orcid.org/0000-0001-8841-6905

António Ferreira² orcid.org/0000-0002-9104-7757

Filipe Ambrósio³ orcid.org/0000-0001-6048-0542

Resumo: Com o enorme fluxo de informação gerado pelas organizações, é importante que estas tenham a preocupação de armazenar os dados num sistema só, num data warehouse. Para além da extração, limpeza, transformação e entrega os dados de informação numa só base de dados, este sistema apoia as organizações na gestão e na tomada de decisões estratégicas. O objetivo deste artigo prende-se na exposição dos tópicos mais importantes a ter em conta no momento de implementar este sistema numa organização, como é o caso da escolha da arquitetura, fatores organizacionais internos que garantem o sucesso da implementação

¹ Mestranda em Gestão Aplicada no Instituto de Gestão e das Organizações de Saúde-Universidade Católica Portuguesa Viseu.

E-mail: joanalsampaio@outlook.com

² Professor Auxiliar. Universidade Católica Portuguesa – Instituto de Gestão e das Organizações da Saúde. E-mail: ajmferreira@ucp.pt

³ Professor Auxiliar Convidado do Instituto de Gestão e das Organizações de Saúde - Universidade Católica Portuguesa - Viseu. E-mail: f.ambrosio1969@gmail.com

de um data warehouse e as questões de segurança, que não devem ser esquecidas. Por fim, é referido, de uma forma breve, uma perspectiva futura, que conjuga as tecnologias 5G e data warehouse, tornando o sistema mais eficiente, seguro e mais barato.

Palavras-chave: *Data Warehouse, Sistemas de Informação, Arquitetura de Data Warehouse, 5G.*

Abstract

With the huge flow of information generated by organizations, it is important that they have the concern to store the data in a single system, a data warehouse. In addition to extracting, cleaning, transforming, and delivering the information data into a single database, this system supports organizations in the management and strategic decision-making. The objective of this article is to expose the most important topics to consider when implementing this system in an organization, such as the choice of architecture, internal organizational factors that guarantee the success of the implementation of a data warehouse and security issues, which should not be forgotten. Finally, a future perspective is briefly mentioned, combining 5G and data warehouse technologies, making the system more efficient, safer, and cheaper.

Keywords: *Data Warehouse, Information Systems, Data Warehouse Architecture, 5G.*

INTRODUÇÃO

A tecnologia *Data Warehouse (DW)* é um sistema de informação, que consiste numa “arquitetura para um sistema de informação empresarial” (Devlin e Murphy, 1988), que propõe “um armazém integrado de dados da empresa” (Devlin e Murphy, 1988), ou seja, integra uma base de dados que agrega informações sobre uma organização, que ao serem consultadas, conseguem suportar a tomada de decisões estratégicas importantes, enquanto facilita o processo de

realização de relatórios e gestão da organização. (Gonzales et al.,2015). A necessidade da adoção de um sistema de *DW* nasce do enorme fluxo de informação, espalhada por diversos bancos de dados, que existe sobre uma organização e a sua atividade. Tendo isto em conta, a gestão da informação organizacional, seria uma tarefa extremamente difícil e até, quase impossível, sem a existência de um sistema que agregasse nele toda a informação necessária (Obali et al., 2021). A adoção de um *DW* tornou-se uma prática habitual nas organizações e bastante usada em vários processos de negócio (Jukic, 2006)

O objetivo deste artigo prende-se na exploração do conceito e da aplicação do *DW*, observando alguns dos fatores essenciais a ter em conta quando uma organização escolhe implementar este tipo de sistema de armazenamento de dados.

No presente artigo, são explorados oito tipos de arquitetura de *DW*, propostos por Golfarelli & Rizzi (2009), expostos na revisão de literatura de Blažič et al. (2017), ou seja, tipos de construção de *DW* adequados a vários tipos de objetivos e necessidades das organizações aquando da implementação de um *DW*. Após a escolha da arquitetura mais adequada, as organizações devem avançar para a implementação, e é importante que as empresas em questão estejam preparadas para receber uma mudança destas. Com base na revisão da literatura de Jukic (2006), Harris (2020) e Bhansali, (2010), foi abordada a implementação do *DW* nas empresas, referindo de seguida alguns fatores a ter em conta para que a implementação seja bem-sucedida, nomeadamente a exploração dos fatores organizacionais, técnicos, de projeto, ambientais e infraestruturais, apresentados por AIMBlhouh & Ahmad (2010). No entanto, e tendo em conta que o *DW* se trata de um sistema que agrega informações confidenciais e preciosas sobre a organização, não pode ser esquecida a questão da segurança, para que seja minimizado o risco de acessos não autorizados. Desta forma, aliados aos fatores a ter em conta na implementação, as aplicações de algumas etapas de segurança devem ser seguidas, tal como certas necessidades básicas de segurança: a confidencialidade, integridade e a disponibilidade (Phoghat & Maitrey, 2015).

No final do artigo, e na tentativa de explorar as tendências futuras do tema em questão, são explicados os benefícios de aliar um *data warehouse* à tecnologia 5G, uma vez que, ao contrário das gerações anteriores de redes móveis de banda larga, na tecnologia 5G, os dados são armazenados numa base de dados separada da própria rede, permitindo a adoção de um DW.

1. METODOLOGIA DE INVESTIGAÇÃO

Em primeiro lugar, procedeu-se à recolha dos artigos através de uma fonte de pesquisa de informação (1ª etapa). Posteriormente, foram lidos (2ª etapa) e selecionados os que foram considerados relevantes para o desenvolvimento deste artigo (3ª etapa). Seguidamente, já no decorrer do trabalho, foi necessária a pesquisa de mais documentos (4ª etapa) para completar certos tópicos do artigo, em que, novamente, foram lidos (5ª etapa) e selecionados (6ª etapa) apenas os documentos com informação de valor para a finalidade deste artigo.

2. METODOLOGIA DE PESQUISA

A estratégia utilizada para realizar a pesquisa começou pela escolha de palavras-chave que conduzissem a investigação. Inicialmente, foram utilizadas as palavras-chave “data warehouse”, “data warehousing”, “warehousing”, “business data warehouse”, no entanto, ao desenvolver a presente bibliometria, foi sentida a necessidade de explorar pontos mais específicos dentro do tema, então, agregada ao termo “data warehouse”, foram pesquisados termos como: “implementation”, “adoption”, “security”, “security issues”, “architecture” e “5G”. É de salientar que a pesquisa foi efetuada na base de dados de informação científica EBSCO. É importante referir que os documentos tidos em conta para a escrita deste artigo foram apenas: artigos científicos, estudos de caso, *papers* de conferência, revistas académicas e também alguns livros, excluindo assim dissertações mestrado e teses de doutoramento.

A pesquisa iniciou-se entre dezembro de 2021 e janeiro de 2022 e, posteriormente, foi sentida a necessidade de mais recolha de informação

e, desta forma, a pesquisa foi retomada em janeiro de 2023. Em primeiro lugar, procedeu-se à recolha de vários documentos, entre os quais artigos científicos, *papers* de conferência e livros. No total, foram recolhidos cerca de 40 documentos. Após a leitura do resumo e introdução de cada um deles, foram excluídos 19, por não se enquadrarem com a finalidade do artigo: os mesmos não abordavam a forma como poderia ser implementada um *Data Warehouse*, ou não continham informação sobre o *Data Warehouse* que fosse relevante ter em conta aquando a implementação da mesma, ou por estarem diretamente ligados à investigação na área da saúde ou da engenharia, para além disso, muitas das publicações encontradas não tinham informação adaptada à realidade atual, por terem sido publicadas há algumas décadas. Algumas das publicações encontradas continham palavras chave relacionadas com o tema, mas o texto não ia de encontro à temática do *Data Warehouse*, mas abordava temas como a inteligência artificial ou outros tipos de tecnologias de informação, para além destes fatores, determinadas obras tratavam-se de estudos de caso e não abordavam o DW de um ponto de vista teórico, acabando também por não se enquadrarem no objetivo da presente bibliometria.

3. HISTÓRIA E DEFINIÇÃO

No mundo atual, altamente globalizado e competitivo, a qualidade e a precisão das informações de um determinado negócio, é um fator de elevada importância na geração de lucro e perdas, pelo que cada vez mais empresas apostam em sistemas de informação para a proteção e o armazenamento dos seus dados. Para além disso, “a necessidade de acesso rápido e fácil a informação fiável e relevante incitaram os líderes empresariais a substituir os sistemas tradicionais de computação empresarial por outros sistemas de apoio à decisão” (Dhaouadi et al., 2022)

Segundo Wiersma & Jurs (2003) sistema de informação (SI) consiste num “agregado de componentes inter-relacionados que, em conjunto, conseguem recolher, processar, armazenar e disseminar informações, para que possam apoiar as organizações nas suas tomadas de decisão,

coordenação, controlo e análise” (Vicente, 2009). As ferramentas de SI fornecem ferramentas de comunicação e análise às empresas que as ajudam a conduzir o mercado e a gerir negócios numa escala global. Desta forma, as empresas estão a tornar-se mais competitivas e eficientes, inseridas num ambiente em que todos os processos estão digitalmente habilitados, como a relação com clientes, fornecedores e funcionários.

O conceito *DW* nasce em fevereiro de 1988, com uma publicação no *IBM Systems Journal*. Este conceito foi criado pelos investigadores Barry Devlin e Paul Murphy, ao identificarem a necessidade de um sistema de apoio à decisão, estes definem o *DW* como “uma arquitetura para um sistema de informação empresarial” (Devlin e Murphy, 1988), que propõe “um armazém integrado de dados da empresa, baseado firmemente no ambiente de bases de dados relacional” (Devlin e Murphy, 1988). Segundo Devlin (2018) a definição usada mais frequentemente de um *DW* é “uma recolha de dados orientada para o assunto, não volátil, integrada e variável no tempo em apoio às decisões de gestão” (Inmon, 1996, as cited in Devlin, 2018), esta definição vai ao encontro da dos autores Gonzales et al. (2015), que afirmam que um *DW* é uma base de dados especializada, que contém “informações de um determinado período de tempo que auxiliam no processo de tomada de decisão gerência”. Este sistema extrai, limpa, transforma e entrega os dados num sistema de armazenamento de dados que apoia a consulta e a análise da informação armazenada com o propósito de auxiliar as organizações na tomada de decisão, para além disso, facilita na realização de relatórios analíticos. O armazenamento de dados tornou-se uma prática inteligente de negócio para empresas que procuram ganhar vantagem competitiva, “a maioria das empresas entende que isto é fundamental no planeamento estratégico e iniciativas, e deve ser absolutamente incluído (...) o data warehouse tornou-se numa ferramenta insubstituível para a tomada de decisões críticas e planeamento” (Castro & Ramaswamy, 2014).

O principal objetivo do *DW* é contribuir para o fácil acesso de informação. Tendo isso em conta, os dados do mesmo devem ser compreensíveis, intuitivos, a informação deve ser consistente,

organizada e de qualidade garantida. No entanto, é necessário ter em conta possíveis mudanças na informação, pelo que o DW deve ser adaptável e resiliente a mudanças. (Cigánek, 2019)

Um sistema DW tem como elementos principais: um sistema operacional que apoia diretamente a execução de um processo de negócio, ao captar detalhes sobre eventos e transações significativas e, um sistema analítico, que apoia na avaliação de um processo de negócio. Relativamente ao sistema operacional, este “captura diretamente a execução de um processo de negócio” (Cigánek, 2019), no entanto, encontra-se fora do DW e, assim, fica fora de controlo o conteúdo e o formato desses sistemas. A última ferramenta do DW é a ferramenta de acesso aos dados que, obviamente, permite a consulta dos dados armazenados no sistema.

Todavia, é de salientar que DW é comumente usado pelas organizações e denota um SI heterogéneo, isto é, a necessidade de construir um DW acontece porque os dados corporativos estão, normalmente, espalhados por diferentes bancos de dados, de diferentes formatos. Então, é necessário obter os dados e informações e colocá-los juntos para produzir um quadro geral dos dados adquiridos e, sem um DW, esta torna-se uma tarefa complicada (Obali et al., 2021).

4. ARQUITETURA EM DATA WAREHOUSE

Para que o processo de tomada de decisão seja facilitado, a organização precisa de um DW onde determinados dados sejam processados e armazenados para ações futuras. Antes da implementação de um DW, devem ser determinados critérios a serem cumpridos e, com base nisso, é feita a seleção da arquitetura de DW mais apropriada aos objetivos e necessidades da organização - “A arquitetura de DW é definida para a construção geral do inter-relacionamento dos componentes do *warehouse*, com base na análise das necessidades do utilizador” (Qian & Li-jun, 2009).

Segundo Kelly (1997) para se escolher um sistema de *data warehouse*, existem algumas propriedades indispensáveis (Golfarelli & Rizzi, 2009), nomeadamente :

- Separação: os processamentos analíticos e transacionais devem ser mantidos de parte o mais possível;
- Escalabilidade: o *hardware* e *software* da arquitetura escolhida devem ser fáceis de atualizar. O volume de dados a serem geridos e processados, o número de utilizadores e requisitos vão sempre aumentar ao longo do tempo;
- Extensibilidade: a arquitetura deve ser capaz de enquadrar em si novas aplicações e tecnologia, sem que o sistema seja comprometido;
- Segurança: a monitorização de acessos ao sistema é fundamental, dada a confidencialidade dos dados nela armazenados;
- Administrabilidade: a gestão do sistema *DW* não deve ser excessivamente difícil.

Os autores Golfarelli & Rizzi (2009) apresentaram uma explicação mais clara sobre as várias arquiteturas de design. Classificam as arquiteturas de *DW* em duas categorias. A primeira categoria inclui arquiteturas de (1) camada única, (2) duas camadas e (3) três camadas. Em segundo lugar: (4) arquitetura de *data marts* independentes, (5) arquitetura *bus*, (6) arquitetura *hub-and-spoke*, (7) arquitetura centralizada e, (8) arquitetura distribuída.

O Quadro 1 apresentado na página seguinte, representa uma comparação realizada entre todos os tipos de arquitetura apresentados pelos autores Golfarelli & Rizzi (2009), para que se possa ter um panorama geral sobre todos os tipos de arquitetura. São comparadas características como a utilização de ferramentas ETL (*Exact, Transform and Load*) no processamento de dados, o facto dos utilizadores conseguirem os dados apenas através do *DW*, se são usados *data marts* no acesso aos dados e se os *data marts* são dependentes do *DW*.

No início dos anos 2000, foram propostos diferentes métodos de modelação conceptual, lógica e física para o processo de ETL - “O processo ETL é utilizado para extrair dados de diferentes fontes; transformá-los para satisfazer necessidades analíticas específicas; e, finalmente, carregar os dados processados num sistema de

armazenamento dedicado para os apoiar, chamado *Data Warehouse*” (Dhaouadi et al., 2022).

Relativamente à primeira classificação, correspondente à (1) camada única, (2) camada dupla e (3) camada tripla, esta é orientada para a estrutura. Relativamente à segunda classificação, (4) arquitetura de *data marts* independentes, (5) arquitetura *bus*, (6) arquitetura *hub-and-spoke*, (7) arquitetura centralizada e, (8) arquitetura distribuída, esta “depende da forma como as camadas são empregadas para criar visões e *data warehouses* orientadas para as empresas ou departamentos” (Golfarelli & Rizzi, 2009).

Quadro 1

Comparação de arquiteturas de *data warehouse* e as suas características

	Arquitetura Camada única	Arquitetura Camada dupla	Arquitetura Camada tripla	Arquitetura de <i>data marts</i> independentes	Arquitetura <i>bus</i>	Arquitetura <i>Hub-and- spoke</i>	Arquitetura centralizada	Arquitetura distribuída
Ferramentas ETL usadas no processamento de dados	NÃO	SIM	SIM	SIM	SIM	SIM	SIM	SIM
Acesso direto aos dados pelos utilizadores	SIM	NÃO	NÃO	NÃO	NÃO	NÃO	NÃO	SIM
Uso de <i>data marts</i> no acesso aos dados	NÃO	SIM	SIM	SIM	SIM	SIM	SIM	NÃO
<i>Data marts</i> dependentes do DW	-	SIM	SIM	NÃO	NÃO	NÃO	SIM/NÃO	-

Fonte: (Blažii, 2017)

1. Camada única

Na prática, esta camada não costuma ser usada. Nesta camada, pretende-se remoção a redundância dos dados, minimizando a quantidade dos dados serem armazenados. A principal fraqueza da arquitetura de camada única, é falhar em separar o processamento de dados analíticos e transacionais.

2. Camada dupla

Neste tipo de arquitetura, existe uma separação em duas camadas: uma camada de fontes e dados e, a outra camada, de *DW*. No entanto,

“embora seja chamada de arquitetura de duas camadas para enfatizar a separação das duas camadas, na verdade ela consiste em quatro estágios de fluxo de dados: camada de origem, preparação de dados, camada de *DW* e análise” (Blažiū et al., 2017).

3. Camada tripla

Este tipo de arquitetura contempla três camadas fisicamente implementadas: a camada de origem, a camada reconciliada e a camada de *DW*. Nesta camada, os dados operacionais são materializados após a limpeza dos dados na fonte e a sua integração. Este processo permite que os dados sejam, integrados, consistentes, corretos, precisos e detalhados.

4. Arquitetura de *data marts* independentes

Esta prática, consiste no planeamento e implementação de *data marts* em separado, ou seja, os *data marts* não são integrados, pois os *data marts* tendem a ser constituídos por dados inconsistentes, que usam medidas e dimensões diferentes, o que irá dificultar a análise desses dados, ou seja, “Os dados são extraídos dos múltiplos ficheiros internos e externos do sistema de fonte e bases de dados em numerosos *data marts*. Os dados dos vários sistemas-fonte são transformados e integrados antes de serem carregados nos *data marts*. Os *data marts* são mini-armazéns, de âmbito limitado.” (Wang, 2022). No entanto, este tipo de arquitetura costuma ser substituída por outra para obter uma integração de dados melhor e para que o cruzamento de relatórios seja facilitado.

5. Arquitetura *bus*

Este tipo de arquitetura é muito semelhante à arquitetura de *data marts* independente. A principal diferença entre os dois tipos de arquitetura assenta no facto dos *data marts* estarem logicamente integrados. Para além disso, todas as empresas têm uma visão da informação existente. Este método é recomendado pelo autor Ralph Kimball.

6. Arquitetura *hub-and-spoke*

“A arquitetura *hub-and-spoke* consiste em fontes de dados, dados reconciliados e *data marts*. O *DW* empresarial, denominado *hub*, é criado com um conjunto de *data marts*, chamados *spokes*. Anatómicos, dados normalizados são armazenados numa camada reconciliada que alimenta um conjunto de *data marts* feitos de dados resumidos de forma multidimensional” (Blažiū et al., 2017). Este tipo de arquitetura é caracterizado pela escalabilidade e extensibilidade, assim como pela recuperação de grandes quantidades de informação.

7. Arquitetura centralizada

Esta arquitetura é vista como uma implementação específica da arquitetura *hub-and-spoke*, no entanto, esta difere da *hub-and-spoke* porque não existem marcas de dados dependentes. Consiste num *DW* que contém dados integrados e *data marts*. É um *DW* centralizado e recomendado pelo autor Bill Inmon, citado por Zhao, K. (2020).

8. Arquitetura distribuída

Este método tende a ser utilizado em contextos dinâmicos, onde os *DW* existentes devem ser integrados, com o objetivo de fornecer uma solução única. Cada *DW data mart*, integra esta arquitetura, logica ou fisicamente, ao usar chaves conjuntas, meta dados globais, consultas distribuídas, entre outros métodos.

Posto isto, é de salientar que as organizações devem analisar as arquiteturas existentes e escolher aquela que mais se adapta aos seus objetivos e necessidades, nunca esquecendo os fatores, anteriormente referidos, de separação, escalabilidade, extensibilidade, segurança e a administrabilidade. Após escolher o tipo de arquitetura mais adequada, a organização pode passar à sua implementação.

5. IMPLEMENTAÇÃO DO DATA WAREHOUSE NAS ORGANIZAÇÕES

“O DW tornou-se uma prática padrão para a maioria das grandes empresas em todo o mundo. Os dados armazenados no DW podem capturar muitos aspetos diferentes do processo empresarial, tais como o fabrico, distribuição, vendas, e marketing. Estes dados refletem explícita e implicitamente os padrões e tendências dos clientes, a eficácia das estratégias empresariais e práticas resultantes, e outras características. Tais dados são de importância vital para o sucesso do negócio (...).” (Jukic, 2006, p.83). Para que a implementação de um DW seja bem sucedida, é importante que as empresas explorem as diferenças e as soluções entre as modelos de DW disponíveis, escolhendo a mais adaptada ao ambiente da organização. Apesar da crescente abordagem ao DW por parte de investigadores e profissionais, não há um consenso entre os mesmos sobre as estratégias mais apropriadas para a implementação de um projeto de DW.

Um DW desempenha um papel importante para as organizações, uma vez que permite a redução ou até a eliminação do trabalho manual desempenhado pelos colaboradores. Este é usado para criar relatórios de dados, atuais e históricos, que são usados para criar relatórios analíticos, para uso dos colaboradores. O DW, nas empresas, é gerido por um DBA (*Database Administrator*) que monitoriza as entradas de informação na base de dados, assegurando também o funcionamento dos seus servidores online (Harris, 2020).

Bhansali (2010) propõe um modelo conceptual composto por cinco componentes que as empresas devem adotar para tirar o melhor partido da implementação de um DW, combinando o mesmo com sistemas de apoio à decisão, que consiste em:

1. Responsabilidade partilhada entre o DW e os gestores da empresa: “O nível de responsabilidade conjunta entre os gestores de empresas e de DW é fundamental para o alinhamento estratégico e a adoção bem sucedida do DW. O empenho da gestão de topo e o envolvimento das IT na

definição da estratégia empresarial é fundamental para um melhor alinhamento.” (Bhansali, 2010, p.93);

2. Alinhamento ente o plano de DW e o plano de negócios em vigor na empresa: “A estratégia de DW é afetada pela estratégia empresarial da organização e tem impacto na estratégia empresarial, em retorno. A realização do alinhamento entre os dois - DW e estratégia empresarial - é feita através da integração do armazenamento de dados e dos planos empresariais. Os planos de armazenamento de dados devem apoiar os planos de negócio.” (Bhansali, 2010, p.100);
3. Satisfação do utilizador empresarial: “No ambiente de DW, os utilizadores empresariais são os principais clientes do sistema. Um requisito básico para um DW bem sucedido é a sua capacidade de fornecer aos utilizadores empresariais informações precisas, consolidadas e atempadas. A força do DW é a sua capacidade de organizar e fornecer dados em apoio ao processo de tomada de decisões da gestão.” (Bhansali, 2010, p.104);
4. Flexibilidade no planeamento do DW: “Os gestores são continuamente confrontados com novas e sempre mutáveis pressões competitivas da desregulamentação, da globalização e da convergência de indústrias e tecnologias. A estratégia e o alinhamento estratégico precisam, portanto, de abraçar uma maior flexibilidade para estimular a criatividade e a inovação. (...) A questão da flexibilidade no planeamento é pertinente para o ambiente de DW. Nos projetos de DW, a estratégia empresarial pode mudar durante o desenvolvimento do DW. Uma mudança no plano de negócios poderia levar a uma mudança nos requisitos empresariais do DW. (...) Um modelo e uma arquitetura de DW precisa de ser flexível para refletir as mudanças no processo empresarial ao longo do tempo.” (Bhansali, 2010, p.112);

5. Integração técnica do DW: “Se for selecionado um projeto de DW com muito pouca tecnologia nova, é possível que o DW possa ser visto como não respondendo às necessidades do utilizador. Por outro lado, se for selecionado um projeto de DW que irá fornecer grandes quantidades de novas informações à empresa, poderá entrar em colapso sob o seu próprio peso. Além disso, um projeto de DW que forneça muito pouca informação nova à empresa pode ser novamente considerado como não respondendo às necessidades do utilizador. Em qualquer dos casos, fornece pouco ou nenhum valor. (...) Assim, esta discussão mostra que o grau de integração técnica é fundamental para o alinhamento estratégico e a adoção bem sucedida do DW.” (Bhansali, 2010, p.111);

Hoje em dia, o DW e os sistemas de apoio à decisão tradicionais ainda são sustentáveis para que as organizações possam gerir os seus dados, no entanto estas devem integrar as suas bases de dados tradicionais com plataformas de *front-end cloud*, principalmente num futuro pós COVID-19. A pandemia veio provar às empresas que estas precisam de confiar mais na análise de dados para operarem no dia-a-dia. (Harris, 2020).

No entanto, os autores AlMblhouh & Ahmad (2010), identificam alguns (1) fatores organizacionais, (2) técnicos, (3) de projeto, (4) ambientais e (5) infraestruturais, que afetam o sucesso da implementação do *DW* nas organizações.

1. Fatores Organizacionais

Lu & Wang (1997) identificam como fatores organizacionais o estilo de gestão. Adicionalmente, Ang et al. (2001), assume que fatores como o uso de Tecnologias de Informação (TI), a estrutura e tamanho organizacionais, o conhecimento de gestão de TI, a alta administração, recursos financeiros, o alinhamento de metas e o método orçamental como condicionantes de sucesso da implementação de *DW*. Para além destes, Hwang et al. (2004) referem o tamanho da organização, o apoio administrativo e necessidades internas como fatores. Mais tarde, Hussein

et al. (2007) afirma, à semelhança dos anteriores, que a estrutura de tomada de decisão, a administração, o alinhamento de objetivos, o conhecimento em TI, estilo de gestão e alocação de recursos, são as condicionantes de sucesso organizacionais. Por fim, Ramamurthy et al. (2008) propõe um modelo de pesquisa e identifica o impacto do compromisso organizacional, a capacidade de absorção, o tamanho da organização, a vantagem relativa e a complexidade como fatores de sucesso na adoção de *DW*.

Ao analisar os argumentos expostos pelos autores, os fatores ligados à administração são os mais mencionados, pelo que, se torna possível constatar que, para além de todos os fatores serem importantes, o bom funcionamento da administração é um ponto fulcral numa organização na hora de implementar um sistema *DW*.

2. Fatores Técnicos

Neste ponto, serão identificados os fatores que surgiram de problemas e limitações técnicas durante a implementação de projetos de *DW*, e são também aqueles que recebem mais críticas. De acordo com Mukherjee (2003) recursos como qualidade de dados, o seu acesso e disponibilidade, conhecimento técnico, disponibilidade de recursos qualificados e de equipamentos, plataformas de tecnologia e formação, são essenciais para que se evite falhas técnicas que comprometam a implementação de *DW*. Para além destes, são mencionados também como fatores, por Solomon (2005), dados limpos, meta dados, a adoção de uma metodologia padrão, conversão de dados, design de modelo de dados, design de arquitetura e ferramentas ETL.

No entanto, é necessário ter em conta que mesmo combinando todos os fatores supra mencionados, continua a existir algum risco de existência de falhas e problemas de carácter técnico, e podem surgir em várias etapas da implementação de *DW*, sendo necessário que as organizações se equipem e preparem para lidar com eventuais incidentes da forma menos prejudicial possível.

3. Fatores Projeto

Implementar um SI, envolve uma forte gestão de diversas tarefas. No que toca a projetos de *DW*, requer equipas altamente qualificadas e bem geridas, capazes de lidar com problemas que surgem durante e após a implementação do projeto. Hwang et al. (2004) afirmam que o fator dos projetos está diretamente ligado ao planeamento e resolução de todos os problemas, análise, desenvolvimento e controlo do projeto no processo de implementação. Nesse sentido, é essencial ter um plano de negócios bem estruturado que descreva todos os benefícios, recursos, custos, riscos e que apresente cronogramas estratégicos e tangíveis (Sakaguchi & Frolick, 1997). Também deve ter sido em conta um alinhamento orçamental correto do investimento no sistema *DW*, tendo em conta as mudanças nos processos de trabalho e o rumo futuro da organização em questão (Bhansali, 2007).

4. Fator Ambiental

De acordo com Hwang et al. (2004) “a empresa incorpora-se num ambiente dinâmico com grandes possibilidades de mudanças repentinas e descontroladas. A empresa deve medir e reduzir as incertezas no ambiente circundante e criar vantagens competitivas através da adoção de novas medidas de informação.” (as cited in AlMblhouh & Ahmad, 2010, p.69), pelo que o ambiente corporativo é influenciado pela utilização de novas TI, dado que a concorrência, o aumento ou diminuição de produtividade e a eficiência da implementação de novos projetos, incluindo de SI. Um bom ambiente e dinâmica corporativo, é uma garantia de sucesso na implementação de um projeto *DW*.

5. Fatores Infraestruturais

A infraestrutura de TI é definida por Applegate et al. (2008) como “um conjunto de recursos de TI tangíveis e compartilhados, que fornecem uma base para apoiar as aplicações de negócios atuais e futuros”. (as cited in AlMblhouh & Ahmad, 2010, p.69)

A gestão da infraestrutura de TI envolve a necessidade de capacidades de tomada de decisão a nível de plataformas de *software* e *hardware*, arquiteturas de rede e padrões corporativos para aquisição e

implementação de recursos, entre outros, sendo necessário adotar medidas que garantam a qualidade das fontes de dados, adquirir as ferramentas de desenvolvimento certas e dispor de bons *softwares* e *hardwares* (Weill & Broadbent, 1998). Assim, é esperado que estes fatores influenciem e apoiem o valor do negócio e os seus processos. A Infraestrutura de TI pode afetar os processos de operação e gestão, mas as melhorias traduzem-se em benefícios de negócio (informações de gestão aprimoradas e melhoria da coordenação entre partes interdependentes da organização).

De acordo com Jukic (2006, p.88), “a criação de um DW é um esforço que exige a participação de várias áreas funcionais das organizações e a integração de dados de múltiplas fontes. Conseqüentemente, as pessoas associadas a projetos de DW têm diferentes antecedentes e vários níveis técnicos de especialização”, por esta razão, a escolha do modelo de DW mais adequado à organização pode ser difícil. É importante ter em conta que a escolha da abordagem de DW a implementar pode determinar um caso de sucesso se for a mais acertada, mas também se pode traduzir num fracasso dispendioso se não for escolhida e implementada corretamente.

6. SEGURANÇA DE UM DATA WAREHOUSE

O armazenamento de dados, sejam eles antigos ou atuais, é um processo muito útil para que as organizações consigam ter uma visão única e detalhada das informações guardadas, para além de quem estas são cruciais na hora de tomar decisões de nível estratégico. O DW consiste numa coleção de dados, que devem ser protegidos por alguns sistemas de segurança, para que não se corra o risco de que alguém não autorizado aceda aos dados. No entanto, este processo é difícil e exige garantias tanto a nível tecnológico, como ao nível dos analistas de negócios da organização em questão.

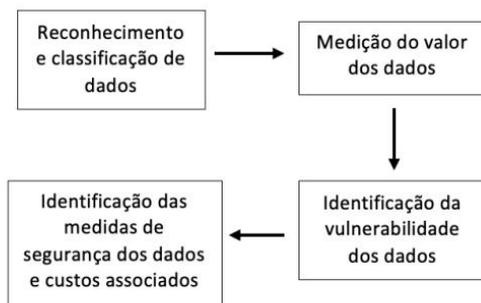
De acordo com a estratégia da Oracle, para proteger os dados do DW, é necessário ter em conta algumas considerações. Em primeiro lugar, a segurança ponta-a-ponta, ou seja, os dados integrantes de um DW são

recolhidos de várias fontes de dados e depois são colocadas no *DW*, a seguir, são distribuídos para *data marts* e outros servidores analíticos, para que posteriormente, os dados possam ser usados pelos utilizadores finais. Seguidamente, é necessário ter em conta a arquitetura do *DW* pois, “um *data warehouse* é mais fácil de proteger do que vários *data marts*, porque quando usamos dezenas de *data marts*, torna-se muito complexo e caro gerir, mas um *data warehouse* consolidado, pode ser gerido de forma simples e também mais barata” (Phoghat & Maitrey, 2015). Então, ao focar a segurança no *DW* como um todo, o nível de segurança é maior, enquanto o esforço e o investimento são reduzidos. Por último, é necessário ter sempre em conta que os dados são o elemento principal e mais importante de um *DW*, por essa razão, a adoção de um modelo de segurança multicamadas, para que haja uma segurança e proteção de dados melhor.

Segundo Phoghat e Maitrey (2015) na fase de implementação de um projeto de *DW*, existem várias etapas a serem aplicadas, para que se garanta a segurança de um sistema tão complexo como é o caso de um *DW* (Figura 1):

Figura 1

Abordagens básicas de segurança do data warehouse



Fonte: Phoghat & Maitrey (2015)

1. Reconhecimento e classificação de dados:

Deve ser feita uma lista detalhada, onde constem todas as informações disponíveis no *DW*, para que se possa fazer uma separação por tipo de dados e classificá-los. A classificação de dados é importante para o planeamento da segurança e para a análise de dados.

2. Medição do valor dos dados:

Esta etapa permite que se faça uma estimativa aproximada do custo real da segurança, para situações em que seja necessária a recuperação de mensagens de segurança, isto é, a recuperação de dados corrompidos, ou quando há perda de confidencialidade.

3. Identificação da vulnerabilidade dos dados:

Procurar conhecer a plataforma técnica, a gestão da conta e as ações do utilizador.

4. Identificação das medidas de segurança dos dados e o custos associados:

Identificar as soluções e os seus custos para as ameaças que são encontradas, e seleccionar uma medida económica para as medidas de segurança necessárias. A eficácia das medidas de segurança têm de ser analisadas.

7. PROBLEMAS ASSOCIADOS À SEGURANÇA

A implementação de um *DW* é um fator importante numa organização, que fornece à mesma informação sobre o processo empresarial como um todo. A segurança é um requisito imprescindível na implementação e manutenção de um *DW* (Kumar et al., 2016).

Grande parte das abordagens recorre a métodos de criptografia na tentativa de conseguir uma forte segurança no que toca à privacidade de dados, no entanto, este método tende a ser insuficiente para o uso de um *DW* (Phoghat & Maitrey, 2015). Por esta razão, é crucial que se recorra a uma técnica que forneça uma forte privacidade de dados, com alto desempenho.

A segurança de dados concentra três questões: a CIA (*Confidentiality, Integrity and Availability*)- as necessidades básicas da segurança:

- **Confiabilidade (*confidentiality*):** “A confidencialidade é utilizada para proteger a forma de informação de utilizadores não autorizados, utilizando métodos diretos e indiretos. Para proporcionar confidencialidade ao DW foram utilizadas diferentes abordagens no que diz respeito ao controlo de acesso” (Kumar et al., 2016, p.177);
- **Integridade (*Integrity*):** “Integridade significa proteger os dados de alterações acidentais ou maliciosas, tais como inserção, atualização ou eliminação incorreta de dados” (Kumar et al., 2016, p. 178), ou seja, é a garantia de como os dados armazenados são precisos e confiáveis. Envolve a proteção de dados contra alterações inesperadas, como é o caso da inserção de dados (Phoghat & Maitrey, 2015);
- **Disponibilidade (*Availability*):** “A disponibilidade de dados é o conceito mais importante do sistema de DW. Esta abordagem é utilizada para recuperar os dados da corrupção em tempo real. A replicação de dados é utilizada para restaurar dados de danos usando ajuda de diferentes soluções. Desta forma, podemos realizar a manutenção da base de dados de uma forma fácil.” (Kumar et al., 2016, p.179). Assim assegura-se que apenas os utilizadores autorizados têm acesso aos dados, e proíbe os utilizadores não autorizados de aceder à informação (Phoghat & Maitrey, 2015).

O foco principal, os riscos de cada um dos fatores, e o seu controlo estão expostos no Quadro 2.

Quadro 2

Comparação de problemas de *data warehouse*

Problemas (CIA)	Foco Principal	Riscos	Controlo
Confidencialidade	Segurança sensível de informação	Acesso não autorizado; Perda de dados.	Autenticação; Data masking; Encriptação.
Integridade	Controlo operacional	O dados deixam de ser exatos	Logs de auditoria
Disponibilidade	Análise de negócios e planeamento	Perturbação do negócios	Backup de armazenamento

Fonte: (Phoget & Maitrey, 2015)

A combinação das abordagens CIA são bastante úteis para se encontrar uma boa solução de segurança na adoção de um DW. É necessário que se construa um modelo que ajude a identificar as necessidades de segurança do DW, em todo o seu ciclo de vida, da organização em questão. Os autores Phoghat & Maitrey (2015) consideram que existe uma série de passos que garantem a eficácia das medidas de segurança, nomeadamente:

1. Garantir que a privacidade ao nível do utilizador é respeitada – “a privacidade de dados mantida internamente com a ajuda de diferentes tipos de restrições de integridade” (Phoghat & Maitrey, 2015, p.938);
2. A seleção dos dados corretos, em que deve ser usado um filtro para remover os dados incorretos, “a exatidão dos dados deve ser garantida antes da entrada no *data warehouse*”;
3. Depois de recolher o todos os dados (de várias fontes), deve ser feita uma limpeza dos mesmos, para que haja uma precisão maior;
4. Inserir os *data marts* num só DW, de forma a melhorar a proteção dos dados,
5. Fornecer aos clientes e/ou funcionários dados de autenticação adequados para que possam aceder ao DW, se tiverem essa necessidade;
6. Deve ser feito um registo de toda a atividade do DW, para que a verificação de erros e vulnerabilidades de segurança sejam facilitados.

8. TENDÊNCIAS FUTURAS - DATA WAREHOUSE E O 5G

“Nas telecomunicações, a norma tecnológica de quinta geração para redes móveis de banda larga é denominada Rede Móvel de 5ª Geração ou 5G.” (Keerthi & B., 2021). Nas gerações anteriores de rede, todos os dados eram armazenados nos próprios elementos de rede, no entanto, o 5G traz consigo a novidade de que os dados serão armazenados numa base de dados separada da própria rede, com o intuito de melhorar tanto a capacidade de otimização da rede como a sua fiabilidade em comparação às gerações anteriores de rede móvel de banda larga. (Keerthi & B., 2021).

“As redes 5G fornecem a velocidade e a capacidade que permitem uma infinidade de tecnologias na transformação digital (...). As redes e tecnologias 5G fornecem a alta largura de banda e a baixa latência necessárias para adquirir e analisar dados do sensor multimodal em tempo real para melhorar a eficiência do fluxo de trabalho e reduzir desperdício e custo” (Zhao et al., 2021).

Após o pré-processamento dos dados (operações de limpeza, classificação, associações, construção e armazenamento de dados), os dados agregados são direcionados para o módulo de armazenamento, onde estes são processados ao usar um método de mineração de dados para que possam ser produzidos dados organizados de acordo com os assuntos específicos estipulados. Os dados armazenados no *DW* vão variando com o tempo e são reabastecidos e atualizados pelos novos dados que são detetados no ambiente de rede, mas o *DW* é apenas estável por algum tempo, sendo que a análise de consulta é dominante e há pouca atualização e/ou remoção de dados. No entanto, o cenário é modificado se as organizações escolherem aliar o 5G a um *DW*, pois se este for “orientado para o assunto, altamente integrado, variável no tempo, e estável, pode fornecer conjuntos de dados organizados para diversas aplicações de otimização de rede inteligente que fornece um relatório só, inédito e único para os investigadores de todo o mundo” (Huang et al., 2021).

Assim sendo, é possível constatar que a utilização da tecnologia 5G nos sistemas de *DW* é altamente vantajosa, pelo que as organizações

devem ponderar a combinação de ambos, para potenciar a segurança e o funcionamento da arquitetura *DW* escolhida e, em simultâneo, reduzir os custos inerentes à sua implementação.

CONCLUSÕES

Ao longo deste artigo são expostos tópicos de grande importância para que o conceito e o funcionamento do *DW* sejam compreendidos.

Primeiramente, explorou-se conceito de *DW*, que nasce no final dos anos 80 pelos autores Devlin & Murphy (1988), que definem um *DW* como “uma arquitetura para um sistema de informação empresarial”; posteriormente, são abordados tópicos como os diferentes tipos de arquiteturas e as suas características, que as empresas devem ter em conta quando decidem adotar um *DW*, que deve ser feito de forma adaptada a cada organização. A implementação do *DW* nas organizações é, atualmente, uma prática comum no mundo empresarial, mas ainda assim há algumas empresas que falham ao adotar um sistema de *DW*, por não adotarem a arquitetura certa ou por não terem em conta alguns fatores importantes, este insucesso representa um erro dispendioso para as empresas. Nesta bibliometria também foram exploradas a questão da segurança e os problemas a ela associados, que estão concentrados em três necessidades: a confiabilidade, a integridade e a disponibilidade. Por fim, no que toca às perspetivas futuras, chega-se à conclusão de que um sistema de *DW* pode ser conjugado com a tecnologia 5G, pela forma como esta armazena os seus dados: fora da sua própria rede. Neste contexto, é possível ter uma visão geral e clara sobre os benefícios da implementação deste tipo de sistema numa organização e as implicações a ela inerentes.

Existiram algumas limitações ao desenvolver o presente estudo, nomeadamente, a falta de conhecimento prévio sobre o conceito de *DW*, que se tornou um entrave ao desenvolvimento da parte técnica e/ou prática da implementação de um *DW* numa organização. Para além desta, identifica-se também como uma limitação a escassez de estudos publicados sobre perspetivas futuras do *DW* ou da conjugação deste com tecnologias como o 5G ou 6G, pelo que dificultou o desenvolvimento

desse tópico no presente artigo. Relativamente a futuros investigadores, este artigo poderá ser um bom ponto de partida para um posterior desenvolvimento mais aprofundado do DW no geral ou dos restantes tópicos abordados em específico.

FONTES E BIBLIOGRAFIA

- AlMabhough, A. & Ahmad, A. (2010). Identifying Quality Factors Within Data Warehouse. *Second International Conference on Computer Research and Development*. 65-72. doi:10.1109/ICCRD.2010.18.
- Blažiū, G., Pošpiū P. & Jakšiū, D. (2017) Data Warehouse Architecture Classification. *0th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. 1491-1495. doi: 10.23919/MIPRO.2017.7973657.
- Cigánek, J. (2019). Design and Implementation of Open-data Data Warehouse. *6th International Conference on Advanced Control Circuits and Systems (ACCS) & 5th International Conference on New Paradigms in Electronics & information Technology (PEIT)*. doi: 10.1109/ACCS-PEIT48329.2019.9062879.
- Devlin, B. (2018). Thirty Years of Data Warehousing. *Business Intelligence Journal*, 23(1), 12–24. <https://discovery.ebsco.com/c/ljojij/details/5h2syh3dqz>
- Devlin, B. & Murphy, P. (1988). An architecture for a business and information system. *IBM Systems Journal*. 27 (1). 61-80. doi: 10.1147/sj.271.0060
- Dhaouadi, A., Gammoudi, M. M., Bousselmi, K., Monnet, S., & Hammoudi, S. (2022). Data Warehousing Process Modeling from Classical Approaches to New Trends: Main Features and Comparisons. *Data*, 7(8). 1-38. Doi: 10.3390/data7080113
- Gofarelli M. & Rizzi, S. (2009) *Data Warehouse Design: modern principles and methodologies*. Tata McGraw-Hill.
- Gonzales, R., Wareham, J. & Serida, J. (2015). Measuring the Impact of Data Warehouse and Business Intelligence on Enterprise Performance in Peru: A Developing Country. *Journal of Global Information Technology Management*. 18. 162–187.

- DOI: 10.1080/1097198X.2015.1070616
- Harris, R. M. (2020). Data Warehousing and Decision Support System Effectiveness Demonstrated in Service Recovery During COVID19 Health Pandemic. *2020 14th International Conference on Open Source Systems and Technologies (ICOSST), Open Source Systems and Technologies (ICOSST), 2020 14th International Conference On*, 1–5. Doi: 10.1109/ICOSST51357.2020.9333019
- Huang, Y., Liu, S., Zhang, C., You, X. & Wu, H. (2021). True-data testbed for 5G/B5G intelligent network. *Intelligent and Converged Networks*. 2(2). 133–149. DOI: 10.23919/ICN.2021.0002
- Jukic, N. (2006). Modeling Strategies and Alternatives for Data Warehousing Projects. *Communications of the ACM*, 49(4), 83–88. <https://doi.org/10.1145/1121949.1121952>
- Keerthi, K. S., & R., B. (2021). Unstructured Data Storage Function for 5G Core Network. *IUP Journal of Telecommunications*, 13(4), 7–19.
- Kumar, S., Singh, B., & Kaur, G. (2016). Data Warehouse Security Issue. *International Journal of Advanced Research in Computer Science*, 7(6), 177–179.
- Neera Bhansali. (2010). *Strategic Data Warehousing: Achieving Alignment with Business*. Auerbach Publications.
- Obali, M., Dursun, B., Erdem, Z. & Görür, A. (2013). A Real Time Data Warehouse Approach for Data Processing. *21st Signal Processing and Communications Applications Conference (SIU)*. 1-4. doi: 10.1109/SIU.2013.6531245.
- Ojeda-Castro, A. & Ramaswamy, M. (2014) Best practices for successful development of data warehouses for small businesses. *Issues in Informations Systems*. 15 (1). 277-284. Doi: 10.48009/1_iis_2014_277-284
- Phogat, P. & Maitrey, S. (2015) Analysis of Security Techniques and Issues in Data Warehouse. 1st International Conference on Next Generation Computing Technologies. 936-940. doi: 10.1109/NGCT.2015.7375258.
- Qian, Z. & Li-jun, S. (2009). The Architecture and Design Strategy for Data Warehouse of Highway Management. *2009 Second International Conference on Intelligent Computation Technology and Automation*. 459-462. doi: 10.1109/ICICTA.2009.826.

- Vicente, I. (2009). Factors That Affect The Successful implementation Of A Data Warehouse. *Fourth International Multi-Conference on Computing in the Global Information Technology. 1-6*. doi: 10.1109/ICCGI.2009.8.
- Wang, M. (2022). Teaching data warehousing with SAP HANA. *Issues in Information Systems. 23 (4)*. 254-264.
Doi: 10.48009/4_iis_2022_122
- Zhao, K., Zhu, M., Xiao, B., Yang, X., Gong, C. & Wu, J. (2020). Joint RFID and UWB Technologies in Intelligent Warehousing Management System. *IEEE INTERNET OF THINGS JOURNAL, 7 (12)*. 11640-11655. doi: 10.1109/JIOT.2020.2998484.

Creative Commons Attribution License | This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.