

Engenharia de Requisitos na Construção de Data Warehouses

Jorge Nery¹ *

¹Instituto de Computação – Universidade Federal da Bahia (UFBA)
Av. Milton Santos, s/nº - Ondina, Salvador – BA, 40170-110 – Brazil

²Departamento de Ciência da Computação

jorge.nery@ufba.br

Abstract. *Este artigo revisa a engenharia de requisitos atual (ER) e seu uso na construção de Data Warehouse. Ele pesquisa e identifica futuras direções de pesquisa sugeridas por sistemas emergentes nessa área específica. Em primeiro lugar, o documento faz uma visão geral do estado da arte em pesquisa de ER aplicada a Data Warehouse. A pesquisa é considerada no que diz respeito às tecnologias desenvolvidas para atender a necessidades específicas das tarefas de requisitos, como elicitação, modelagem e análise, no contexto de construção de Data Warehouse. Tal revisão nos permite identificar áreas maduras de pesquisa, bem como áreas que merecem uma investigação mais aprofundada no contexto específico de Data Warehouse. Em seguida, identificamos vários desafios na construção de Data Warehouse. Em terceiro lugar, revisamos várias estratégias de ER aplicadas à construção de Data Warehouse, para ajudar a delinear o escopo de futuras direções de pesquisa nessa área específica.*

Palavras-chaves: *Engenharia de Requisitos, Data Warehouse.*

1. Introdução

Um data warehouse é uma coleção de dados orientada por assunto, integrada, variante no tempo e não volátil para apoiar o processo de tomada de decisão da administração (W. H. Inmon)

A Engenharia de requisitos está intimamente ligada à aquisição e aplicação de conhecimento para a criação, o aperfeiçoamento e a implementação de sistemas de informação sendo assim entendendo que a Engenharia de Requisitos pode desempenhar um papel fundamental na construção dessas soluções[3]. Essas coleções de dados são um ativo valioso, mas sem uma estratégia adequada para sua organização e análise, eles permanecem como informações desconexas e de difícil acesso [13]. A Engenharia de Requisitos envolve a aplicação de práticas, técnicas e ferramentas para elucidar, identificar e analisar requisitos. Ao aplicar princípios de Engenharia de Requisitos na construção de Data warehouses, é possível garantir a qualidade dos dados, a escalabilidade do sistema, a segurança da informação, eficiência no processamento, análise dos dados, validação de algoritmos de aprendizado de máquina e a verificação de qualidade de dados ou validação de modelos preditivos. ([3], [13])

Essa pesquisa pode contribuir para o avanço da área, fornecendo diretrizes e insights para profissionais e organizações que desejam implementar soluções de Data Warehouses de forma eficiente e sustentável, aproveitando todo o potencial dos dados para impulsionar a inovação e a tomada de decisões baseadas em evidências[3].

*Professor Dr. Julio Leite e Rita Suzana

Vamos começar na seção 2 com a metodologia utilizada, seção 3 detalhando as dificuldades encontradas na construção de Data Warehouses segundo o conjunto dos artigos selecionados que podem ser mitigadas utilizando Métodos e Técnicas de Engenharia de Requisitos, na seção 4 vamos investigar como as práticas de Engenharia de Requisitos podem ser aplicadas na construção de Data Warehouses, considerando os desafios específicos relacionados à coleta, transformação e disponibilização dos dados, na seção 5 vamos estabelecer um quadro relacionando as dificuldades e as principais técnicas da engenharia de requisitos que podem ser utilizadas, na seção 6 vamos ponderar sobre as questões apresentadas e na 7 vamos a conclusão

2. Metodologia

2.1. Questão de Pesquisa

- Utilização da Engenharia de Requisitos na criação Data Warehouses
- Investigar como as praticas de Engenharia de Requisitos podem ser aplicadas na construção de Datawarehouses

2.2. Questão Estruturada (facetas)

A questão da pesquisa foi enquadrada no seguinte critérios:

- População: Artigos publicados sobre a utilização de técnicas de Engenharia de Requisitos na formação de Datalakes e Datawarehouses
- Estratégias adotadas na elucidação de requisitos
- Desafios enfrentados na definição de requisitos
- Abordagens de Engenharia de Requisitos

2.3. Palavras-chave de Pesquisa

Consideramos a possibilidade de usar as seguintes palavras-chave: População:

- (("requisitos") OR ("requirements") OR ("engenharia de requisitos") OR ("requirements engineering")) AND (("hipotese") OR ("hypoyhesis")) AND ("data warehouse")
- (("requisitos") OR ("requirements") OR ("requirements engineering")) AND ("data warehouse") AND (("Métodos Ágeis") OR ("Startup") OR ("agile methodologies"))
- (("requisitos") OR ("requirements") OR ("engenharia de requisitos") OR ("requirements engineering")) AND ("data warehouse")

2.4. Seleção de artigos para estudos

2.4.1. Com relação aos critérios de inclusão, este mapeamento sistemático considerou:

- Artigos publicados descrevendo como utilizar a Engenharia de Requisitos na formação de Data Warehouses;
- Quando vários estudos foram relatados no mesmo artigo, cada estudo relevante será tratado separadamente.
- Artigos mais relevantes e com mais citações
- Artigos relevantes citados nos artigos selecionados

2.4.2. Com relação aos critérios de exclusão, este mapeamento sistemático considerou:

- Quando vários artigos relatam o mesmo estudo, apenas o mais recente será incluído;
- foram removidos artigos menos relevantes e de literatura cinza

2.4.3. Processo de seleção e análise

A seleção dos artigos e o processo de mapeamento foi utilizando planilha eletrônica do google e algumas automações no google forms para facilitar a tabulação das informações dos artigos e foram seguidas as questões para extração dos dados (segue link: <https://github.com/jorgenery/artigo-er-building-dw>) :

- Quais as respostas para as questões?
- Lista de abordagens de ER (Práticas, técnicas, Metodos e Processos) identificadas e sua respectiva definição.
- Quantos artigos citam cada uma das abordagens?
- A proposta foi avaliada?
- E qual tipo de avaliação foi realizada (experimento, estudo de caso,...)?
- Quantidade de artigos totais, artigos duplicados, taxa de inclusão e exclusão.
 - Título e autores
 - Fonte workshop/conferência/jornal
 - Ano de publicação
- Respostas para as questões abordadas pelo estudo

3. Dificuldades encontradas na construção de Data Warehouses

A complexidade no design de um data warehouse decorre da falta de padronização nas abordagens de análise de requisitos do usuário. No entanto, ao aplicar as práticas de Engenharia de Requisitos, é possível mitigar essa complexidade. Através da identificação clara dos objetivos organizacionais, decisões críticas e informações necessárias, a equipe de engenharia pode estruturar o processo de análise de requisitos de forma eficiente [1]. Segue as dificuldades encontradas na construção de Data Warehouses que segundo o conjunto dos artigos selecionados, podem ser mitigadas utilizando Abordagens, Metodos e Técnicas de Engenharia de Requisitos([1],[4],[10],[14])

- Projetos de armazenamento são projetos de longo prazo, e é muito difícil antecipar requisitos futuros para o processo de tomada de decisão; conseqüentemente, apenas alguns requisitos podem ser declarados do começo.
- Os requisitos de informação para aplicativos DW são difícil de especificar porque os processos de decisão são flexivelmente estruturado, mal compartilhado entre grandes organizações, zelosamente guardado pelos gerentes e instável ao longo do tempo para acompanhar a evolução dos negócios processos.
- Os requisitos para tomada de decisão geralmente se referem a informações que não existem na forma exigida, e deve ser derivado de fontes de dados integrando, transformando-os e limpando-os
- A informação em tais sistemas é uma questão de exploração e integração de dados sob um paradigma orientado a assuntos

4. Melhores Práticas ER em DW

Os últimos anos têm visto propostas e esforços no campo da engenharia de requisitos de sistemas de Data Warehouse (DW). Diferentes abordagens têm sido adotadas, com algumas enfatizando a melhoria de processos e outras focadas na extração de dados ou na modelagem de objetivos. Essas abordagens geralmente consideram o diagrama ER como entrada para a fase de engenharia de requisitos do DW. A análise detalhada de processos de negócios nem sempre é considerada um bom ponto de partida, e há propostas para desenvolver sistemas DW que exclusivamente suportem processos de decisão. As metas e objetivos da organização são fundamentais na definição das necessidades do DW, identificando as capacidades de tomada de decisão a serem suportadas([9],[7],[2]). A comunicação eficaz entre as partes interessadas e os usuários é valorizada, e a captura das dependências dos stakeholders é considerada importante para alcançar seus objetivos [8].

Vamos explorar como as práticas de Engenharia de Requisitos podem ser aplicadas na construção de Data Warehouses, considerando os desafios específicos relacionados à coleta, transformação e disponibilização dos dados. A análise dos requisitos do usuário desempenha um papel crucial nesse processo, permitindo a compreensão do universo de objetivos, decisões, informações e suas inter-relações dentro de uma organização. Ela ajuda a identificar as necessidades dos usuários finais e as informações relevantes que devem ser armazenadas e disponibilizadas pelo sistema. Ao considerar as diferentes abordagens de análise de requisitos do usuário, como a abordagem orientada por dados, orientada pelo usuário, orientada por metas e abordagem mista, os projetistas de data warehouse podem avaliar suas forças e fraquezas em relação ao contexto específico de seu projeto([8],[3],[7],[12]).

Ao considerar os desafios relacionados à coleta, transformação e disponibilização dos dados em um data warehouse, a Engenharia de Requisitos oferece técnicas e metodologias que ajudam a superar esses obstáculos. Por exemplo, a elicitação de requisitos permite obter uma compreensão profunda das necessidades dos usuários e definir requisitos de dados precisos[7]. Além disso, a modelagem de requisitos auxilia na representação clara e concisa dos requisitos, facilitando a comunicação entre as partes interessadas.

Em suma, ao integrar as práticas de Engenharia de Requisitos no processo de construção de Data Warehouses, é possível enfrentar os desafios relacionados à descoberta e compreensão dos requisitos do usuário. Essa abordagem possibilita a construção de um data warehouse que atenda às necessidades dos usuários finais, fornecendo as informações necessárias para apoiar as decisões organizacionais de maneira eficiente e eficaz.

Segue as Abordagens de ER para DW

- Supply-driven (also called data-driven) - Abordagens orientadas por dados, projetar o DW a partir de uma análise detalhada das fontes de dados. Os requisitos do usuário afetam design, permitindo que o designer selecione quais blocos de dados são relevantes para a tomada de decisão e determinando sua estruturação de acordo com o modelo multidimensional[15].
- Demand-driven (or requirement-driven) - Abordagens orientadas por demanda (ou orientadas por requisitos) é uma abordagem oposta à supply-driven, onde a tomada de decisão é impulsionada pelas necessidades e demandas dos clientes ou usuários. Em vez de se basear nos dados disponíveis, a abordagem demand-driven

busca compreender as necessidades dos clientes e adaptar as estratégias de acordo. Começa determinando os requisitos de informação de usuários empresariais. O problema do mapeamento desses requisitos nas fontes de dados disponíveis é enfrentado apenas a posteriori, projetando ETL e adequado as rotinas[14].

- GRAnD (Goal-oriented Requirement Analysis for Data warehouses) - É uma análise focada no que é chamado de requisitos iniciais em engenharia de software. Em outras palavras, a análise diz respeito aos objetivos de alto nível dos stakeholders e tomadores de decisão, em vez das funcionalidades específicas do sistema a ser (neste caso, o DW) e do ambiente organizacional onde ele irá operar. Isso permite que o analista compreenda e explore as verdadeiras motivações (o porquê) por trás das solicitações dos usuários antes de explorar as possíveis soluções para sua implementação. Como comumente reconhecido na literatura de engenharia de software, analisar os requisitos iniciais diminuirá significativamente a possibilidade de mal-entendidos nas solicitações dos usuários e, consequentemente, reduzirá o risco de falha do projeto DW.[4]
- AGDI (Agent-Goal-Decision-Information) - Engenharia de Requisitos iniciais e tardias de DWs, estendendo o modelo GDI proposto em (Prakash, N., Gosain, A., Requirements driven data warehouse development, proceedings of CAiSE Short Paper, 13-17, 2003), é um modelo conceitual que descreve os componentes e interações envolvidos em sistemas de apoio à decisão. Ele é utilizado para analisar e entender o processo de tomada de decisão em organizações e como as informações são usadas nesse processo[11].
- Triple-Driven propõe uma metodologia de modelagem de dados que integra três abordagens existentes normalmente usadas isoladamente: orientada por objetivos, orientada por dados e orientada pelo usuário. É composto por quatro etapas. O estágio orientado por metas produz assuntos e KPIs (Key Performance Indicators) das principais áreas de negócios. O estágio orientado a dados produz um esquema de dados corporativos orientado ao assunto. O estágio orientado pelo usuário produz requisitos analíticos representados por medidas e dimensões de cada assunto. O estágio de combinação combina os resultados[5].
- DWARF (Data Warehouse Architectural Framework) propõe em um conjunto de fases. Cada fase segue os níveis de abstração do aplicativo em profundidade, à medida que os requisitos do projeto são reunidos para formar uma linha de base de requisitos. Em torno desse ciclo encontra-se uma fase de backbone denominada Controle de Gerenciamento de Requisitos, com o objetivo de realizar avaliação permanente da qualidade da evolução dos requisitos. Juntamente com o processo acima, os modelos de documentos propõem uma estrutura pré-definida para a documentação de requisitos de data warehouse, enquanto funcionam como instrumentos de registro de fatos e pontos de ponta para o desenvolvimento do sistema.[9]

5. Quadro Abordagem ER para DW

É fundamental escolher a abordagem correta de requisitos do usuário que possa ser adequada ao ambiente da organização. Reconhecer as forças e fraquezas de cada abordagem pode auxiliar o processo de seleção da análise de requisitos adequada (Tabela 1).

Abordagem	Características	Vantagens	Desvantagens
Supply-driven (also called data-driven)	<ul style="list-style-type: none"> -Ênfase nos dados -Decisões fundamentadas em dados -Análise de dados em larga escala 	<ul style="list-style-type: none"> -Personalização e segmentação com base nos dados -Iteração e aprendizado contínuo -Acompanhamento de métricas: 	<ul style="list-style-type: none"> -Viés nos dados -Compreensão limitada do contexto -Custos e complexidade -Falta de contexto humano -Risco de subestimar aspectos qualitativos
Demand-driven (or requirement-driven)	<ul style="list-style-type: none"> -Abordagens orientadas por demanda (ou orientadas por requisitos) começa determinando os requisitos de informação de usuários empresariais. -Destaque das forças e fraquezas em diferentes contextos. -Adaptação e flexibilidade 	<ul style="list-style-type: none"> -Maior relevância das soluções -Engajamento -Iteração e aprendizado contínuo 	<ul style="list-style-type: none"> -Dificuldade em equilibrar diferentes requisitos -Complexidade no design do Data Warehouse. -Necessidade de equilibrar com outras considerações estratégicas
GRAnD (Goal-oriented Requirement Analysis for Data warehouses)	<ul style="list-style-type: none"> -Orientação por metas e decomposição de metas -Alinhamento com os objetivos de negócio -Focado em Análise de requisitos - Importância das metas e objetivos da organização. - Ênfase na comunicação eficaz e captura das dependências dos stakeholders. 	<ul style="list-style-type: none"> -Melhoria de processos e suporte à tomada de decisão. -Consideração dos objetivos e metas da organização. -Identificação abrangente de requisitos 	<ul style="list-style-type: none"> -Complexidade do processo -Rigidez do processo -Falta de consideração detalhada dos processos de negócio. -Necessidade de comunicação eficaz com os stakeholders. -Uso de técnicas específicas
AGDI (Agent-Goal-Decision-Information) -	<ul style="list-style-type: none"> -A abordagem AGDI se concentra em agentes e Tomada de decisão -Design do Data Warehouse com base nos dados das fontes operacionais. -Seleção e estruturação dos dados relevantes para a tomada de decisão. 	<ul style="list-style-type: none"> -Otimização de metas e decisões -Análise do impacto das decisões -Flexibilidade na modelagem -Interdependência entre elementos -Iteração e aprendizado contínuo 	<ul style="list-style-type: none"> -Complexidade da modelagem -Dificuldade na obtenção de informações precisas -Papel secundário dos requisitos do usuário na determinação do conteúdo da informação. -Complexidade na definição de metas e objetivos

Abordagem	Características	Vantagens	Desvantagens
Triple-Driven	<ul style="list-style-type: none"> -Equilibra as estratégias Supply-driven, Demand-driven e Technology-driven -Consideração das capacidades tecnológicas. -Foco na criação de valor. 	<ul style="list-style-type: none"> -Alinhamento estratégico -Utilização eficiente de recursos -Flexibilidade na modelagem -Criação de valor sustentável -Iteração e aprendizado contínuo 	<ul style="list-style-type: none"> -Complexidade de desenvolvimento -Dificuldade na sincronização -Requisitos de expertise. -Desafios de coordenação -Conflitos de prioridades
DWARF (Data Warehouse Architectural Framework)	<ul style="list-style-type: none"> -Orientação à arquitetura -Modularidade e escalabilidade. -Abstração de camadas. -Ênfase na qualidade dos dados. -Orientação a metadados 	<ul style="list-style-type: none"> -Flexibilidade na modelagem de dados -Qualidade dos dados -Uso de metadados -Modularidade e escalabilidade 	<ul style="list-style-type: none"> -Complexidade inicial -Restrições de Flexibilidade -Requisitos de expertise. -Tempo e recursos de implementação -Dependência de tecnologias e ferramentas específicas -Complexidade de manutenção

Tabela 1. Quadro comparativo

6. Ponderações

Todas as abordagens relatadas têm méritos e deméritos próprios. Todas se aplicam de acordo com os requisitos do domínio e se adequam conforme a situação, mas ainda são necessárias muitas pesquisas e desenvolvimento na área para alcançar os melhores resultados[8]. As abordagens discutidas acima estão principalmente preocupadas com dados, usuários e metas, na qual há um alto envolvimento das partes interessadas, somente os usuários finais são capazes de definir corretamente os objetivos de negócio dos sistemas de data warehouse, de modo que os usuários finais devem ser capacitados a especificar os requisitos de informação por si mesmos. No entanto, os usuários finais não são capazes de especificar seus requisitos de informação porque sua visão é subjetiva por definição, porque eles não podem ter conhecimento suficiente de todas as fontes de informação disponíveis e porque eles usam apenas uma interpretação específica da unidade de negócios dos dados[15]. Além disso, os usuários finais muitas vezes não conseguem imaginar quais informações inovadoras o sistema de data warehouse poderia fornecer. Devido a isso o risco de falhas é muito alto.

Podemos reduzir esse risco de falha incluindo técnicas de elicitação de requisitos mais otimizadas e eficientes, no entanto, temos a sensação de que isso está fora do escopo do nosso trabalho. Outros itens a se discutir são: No desenvolvimento de data warehouse, o processo de requisitos sustenta uma abordagem cíclica de aquisição, representação e avaliação de requisitos para produzir gradualmente uma especificação de sistema. Assim, um processo iterativo parece ser mais apropriado para suportar esse fluxo de trabalho e a complexidade das estruturas organizacionais;

7. Conclusão

A escolha da abordagem mais adequada para utilização da Engenharia de Requisitos na Criação de Data Warehouses é de extrema importância para o sucesso de um projeto, especialmente ao considerar o ambiente específico da organização. Reconhecer as forças e fraquezas de cada abordagem pode auxiliar na seleção adequada e evitar problemas ou ineficiências durante o processo. Ao selecionar a abordagem de ER, é fundamental considerar os seguintes pontos:

- Objetivos do projeto: Compreender claramente os objetivos do projeto de ER é essencial para escolher a abordagem mais apropriada.
- Natureza dos sistemas e tecnologias envolvidas: Cada abordagem de ER tem suas peculiaridades em relação aos tipos de sistemas e tecnologias que podem ser abordados com eficiência.
- Complexidade do sistema: A complexidade do DW também desempenha um papel na seleção da abordagem.
- Analisar as forças e fraquezas de cada abordagem, é possível fazer uma seleção informada e adequada às necessidades da organização.

Portanto, ao escolher a abordagem mais adequada, é importante considerar cuidadosamente o contexto, os objetivos e as limitações da organização, garantindo uma seleção que ofereça resultados eficientes e úteis para o projeto em questão.

Projetos de Data Warehouse só podem ser bem-sucedidos se as peças de definição estiverem sólidas em duas dimensões entrelaçadas. Qualquer tentativa de definir peças

isoladas do data warehouse que, posteriormente, não possam ser úteis para serem integradas, causará o fracasso do projeto. De acordo com [6], se alguém espera construir um data warehouse robusto e resiliente diante de requisitos em constante evolução, deve aderir a uma definição de DW na qual dimensões e fatos comuns estejam conformes. Uma dimensão é considerada "conforme" quando significa a mesma coisa para todas as tabelas de fatos às quais está ligada. Da mesma forma, um fato está conforme ao modelo geral se for utilizado o mesmo vocabulário em todos os DW que representarem seu conteúdo.

Referências

- [1] Nur Hani Zulkifli Abai, Jamaiah H Yahaya, and Aziz Deraman. User requirement analysis in data warehouse design: a review. *Procedia Technology*, 11:801–806, 2013.
- [2] Mohamed Lamine Chouder, Rachid Chalal, and Waffa Setra. Requirement analysis in data warehouses to support external information. In *Computer Science and Its Applications: 5th IFIP TC 5 International Conference, CIA 2015, Saida, Algeria, May 20-21, 2015, Proceedings 5*, pages 243–253. Springer, 2015.
- [3] Alexandre Manuel Pereira Mendes da Costa. *A gestão da qualidade dos dados em ambientes de data warehousing na prossecução da excelência da informação*. PhD thesis, 2006.
- [4] Paolo Giorgini, Stefano Rizzi, and Maddalena Garzetti. Grand: A goal-oriented approach to requirement analysis in data warehouses. *Decision Support Systems*, 45(1):4–21, 2008.
- [5] Yuhong Guo, Shiwei Tang, Yunhai Tong, and Dongqing Yang. Triple-driven data modeling methodology in data warehousing: a case study. In *Proceedings of the 9th ACM international workshop on Data warehousing and OLAP*, pages 59–66, 2006.
- [6] Ralph Kimball, Margy Ross, Warren Thorthwaite, Bob Becker, and Joy Mundy. *The data warehouse lifecycle toolkit*. John Wiley & Sons, 2008.
- [7] Manoj Kumar, Anjana Gosain, and Yogesh Singh. Stakeholders driven requirements engineering approach for data warehouse development. *Journal of information processing systems*, 6(3):385–402, 2010.
- [8] Sandeep Mathur, Girish Sharma, and AK Soni. Requirement elicitation techniques for data warehouse review paper. *International Journal of Emerging Technology and Advanced Engineering*, 2(11):456–459, 2012.
- [9] Fábio Rilston Silva Paim and Jaelson Freire Brelaz de Castro. Dwarf: An approach for requirements definition and management of data warehouse systems. In *Proceedings. 11th IEEE International Requirements Engineering Conference, 2003.*, pages 75–84. IEEE, 2003.
- [10] Deepika Prakash and Naveen Prakash. *Data Warehouse Requirements Engineering*. Springer, 2018.
- [11] Naveen Prakash and Anjana Gosain. Requirements driven data warehouse development. In *CAiSE short paper proceedings*, volume 252, 2003.
- [12] Naveen Prakash and Anjana Gosain. An approach to engineering the requirements of data warehouses. *Requirements Engineering*, 13(1):49–72, 2008.
- [13] Carlos Eduardo Vazquez and Guilherme Siqueira Simões. *Engenharia de Requisitos: software orientado ao negócio*. Brasport, 2016.
- [14] Robert Winter and Bernhard Strauch. Demand-driven information requirements analysis in data warehousing. *Journal of Data Warehousing*, 8(1):38–47, 2003.

- [15] Robert Winter and Bernhard Strauch. Information requirements engineering for data warehouse systems. In *Proceedings of the 2004 ACM symposium on Applied computing*, pages 1359–1365, 2004.