

## OS 5 V'S DO BIG DATA, IMPACTOS E APLICAÇÕES NA ERA DIGITAL: IMPLEMENTAÇÃO DE PROJETO DE BIG DATA

**Andrew de Sousa Oliveira**

Pós-graduação em Engenharia de Dados, Big Data e Computação em Nuvem  
andrew.oliveira01@aluno.unifametro.edu.br

**Ricardo Evangelista de Araújo Sousa**

Pós-graduação em Engenharia de Dados, Big Data e Computação em Nuvem  
ricardo.sousa@aluno.unifametro.edu.br

**Kaio Gefferson de Almeida Mesquita**

Orientador - Centro Universitário Fametro  
kaio.mesquita@professor.unifametro.edu.br

**Área Temática:** Engenharia de Software e Computação em Nuvem

**Área de Conhecimento:** Ciências Tecnológicas

**Encontro Científico:** XV Encontro de Pós-graduação

### RESUMO

**Introdução:** O Big Data representa um dos principais motores da transformação digital, impulsionado pelo aumento da produção de dados em escala global. Sua relevância está diretamente associada à possibilidade de transformar grandes volumes de informações em vantagem competitiva. **Objetivo:** Este trabalho tem como objetivo apresentar os conceitos e aplicações do Big Data, destacando os 5V's (volume, velocidade, variedade, veracidade e valor), além de discutir seus impactos na era digital. **Métodos:** O estudo baseou-se em revisão de literatura nacional e internacional sobre Big Data, complementada por um estudo de caso prático desenvolvido em ambiente acadêmico, no qual se estruturou um processo de *Extract, Load and Transform* (ELT) para integração de dados referentes ao transporte público de Fortaleza. **Resultados:** A pesquisa evidenciou a ampla utilização do Big Data em setores como saúde, educação, finanças e mercado corporativo, bem como a importância de ferramentas como Hadoop, Airflow, Power BI e Snowflake na manipulação e análise de dados. O estudo de caso demonstrou a aplicabilidade do ELT no tratamento eficiente de dados massivos, confirmando o potencial da tecnologia na geração de valor estratégico. **Considerações finais:** Conclui-se que o Big Data possibilita avanços significativos, mas demanda investimentos em infraestrutura, segurança da informação e capacitação técnica para sua implementação eficaz.

**Palavras-chave:** Big Data; 5V's; Ciência de Dados; ETL/ELT; Transformação Digital.

## INTRODUÇÃO

O crescimento exponencial da produção de dados digitais nas últimas décadas resultou no surgimento do conceito de Big Data, definido inicialmente pela Gartner (2001) a partir dos três pilares: volume, velocidade e variedade. Posteriormente, Kitchin (2014) e outros autores ampliaram o modelo para incluir a veracidade e o valor, consolidando os 5V's como estrutura de referência.

O Big Data representa a capacidade de processar, armazenar e analisar grandes volumes de dados em tempo real, provenientes de fontes variadas como redes sociais, sensores, transações financeiras e sistemas corporativos. Sua adoção está presente em setores estratégicos, possibilitando personalização de serviços, aumento da eficiência operacional e inovação em modelos de negócio (MARR, 2015).

Entretanto, ao mesmo tempo em que promove oportunidades, o Big Data impõe desafios relacionados à qualidade da informação, custos de implementação e riscos de segurança, especialmente diante da Lei Geral de Proteção de Dados (LGPD, 2012). Vazamentos de informações sensíveis podem acarretar perdas financeiras e comprometer a imagem institucional.

Diante desse cenário, este estudo tem como objetivo analisar os impactos e as aplicações do Big Data, discutindo seus 5V's, metodologias e ferramentas, e apresentando um estudo de caso prático de implementação em ambiente acadêmico.

## METODOLOGIA

A pesquisa desenvolvida adota caráter descritivo e qualitativo, fundamentada em revisão de literatura sobre Big Data em bases acadêmicas, livros e relatórios técnicos de organizações de referência, como Gartner e IEEE. Foram considerados artigos publicados entre 2001 e 2024, permitindo identificar a evolução do conceito e suas aplicações.

Além da revisão bibliográfica, foi realizado um estudo de caso prático durante a disciplina de Engenharia de Dados da pós-graduação. Esse estudo consistiu na implementação de um processo de *Extract, Load and Transform* (ELT) utilizando o ecossistema Microsoft como base tecnológica, aplicado a dados referentes ao transporte público de Fortaleza. O objetivo foi testar a aplicabilidade de metodologias e ferramentas discutidas teoricamente.

Como não houve coleta de dados sensíveis ou participação direta de seres humanos, não se aplicou submissão a Comitê de Ética.

## RESULTADOS E DISCUSSÃO

### Os 5V's do Big Data

A análise revelou que os 5V's constituem a base conceitual do Big Data:

- **Volume:** representa a quantidade massiva de dados gerados continuamente.
- **Velocidade:** refere-se à rapidez com que os dados são produzidos e processados.
- **Variedade:** corresponde à diversidade de formatos e origens das informações.
- **Veracidade:** diz respeito à confiabilidade e qualidade dos dados coletados.
- **Valor:** está relacionado à capacidade de transformar dados em informações úteis para a tomada de decisão.

Esses elementos, em conjunto, permitem compreender o potencial estratégico do Big Data, mas também evidenciam os riscos de decisões baseadas em dados imprecisos.

### Processos de manipulação de dados

Foram identificados dois modelos principais: ETL (*Extract, Transform, Load*) e ELT (*Extract, Load, Transform*). O ETL segue o fluxo tradicional de extração, transformação e carregamento, enquanto o ELT prioriza o carregamento dos dados antes da transformação, otimizando o processamento em ambientes de alta escala.



### 01 Extract

Utiliza o **PySpark** para extrair os dados de diferentes fontes.

### 02 Transform

Utiliza o **PySpark** para aplicar uma lógica de padronização para o mesmo tipo, como por exemplo parquet, e também para limpeza de dados, retirando os valores nulos e os metadados.

### 03 Load

Nessa etapa uma das ferramentas que são mais utilizadas é o **Snowflake** que funciona como **Data Warehouse** para armazenar esses dados tratados para serem analisados posteriormente.

**Figura 1** – Metodologia para aplicação de ETL (Extract, Transform and Load)



### 01 Extract

Utiliza o **PySpark e APIs** para extrair os dados de diferentes fontes.

### 02 Load

Utiliza o **Airbyte** para estabelecer uma conexão com o **Databricks**, local onde serão salvos os dados, para depois serem transformados.

### 03 Transform

Nessa etapa utilizamos o **Databricks + Spark** para fazer a transformação dos dados.

**Figura 2** – Metodologia para aplicação de ELT (Extract, Load and Transform)

## Metodologias e técnicas

Entre as metodologias mais recorrentes estão:

- *Machine Learning* (identificação de padrões);
- Análise preditiva (projeções futuras);
- Análise de sentimentos (interpretação de emoções em textos);
- *Stream processing* (processamento de dados em tempo real).

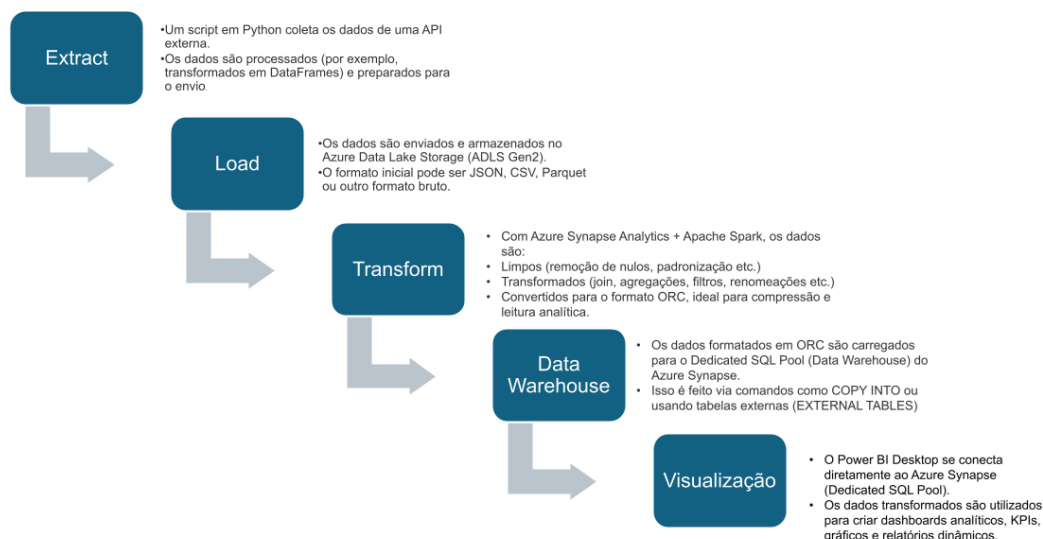
Essas abordagens reforçam o papel do Big Data na inovação tecnológica e no suporte à decisão.

## Ferramentas

Foram destacadas ferramentas amplamente utilizadas:

- **Airflow**: orquestração de fluxos de trabalho;
- **Snowflake**: armazenamento e integração de dados em nuvem;
- **Power BI**: visualização e relatórios gerenciais;
- **Hadoop**: processamento distribuído de grandes volumes de dados.

## Proposta Técnicas de Implementação de ELT



**Figura 3** – Proposta técnica de aplicação de ELT

## Estudo de caso

O estudo de caso aplicado ao transporte público de Fortaleza demonstrou a viabilidade da utilização do ELT no ecossistema Microsoft, permitindo estruturar dados de

forma escalável e eficiente. Essa experiência prática reforçou a importância do domínio de ferramentas e técnicas para a implementação bem-sucedida de projetos de Big Data.

### Modelo Relacional

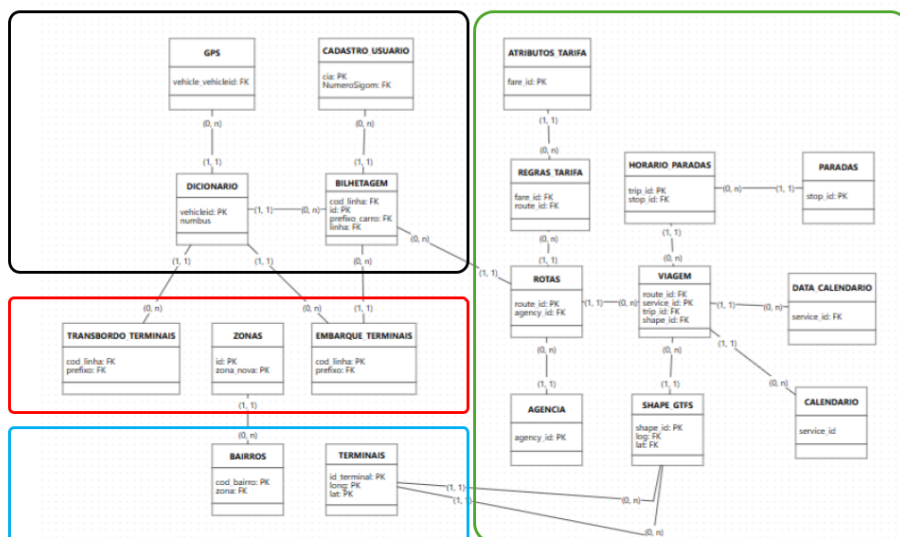


Figura 4 – Estrutura final modelagem

Os resultados obtidos confirmam que o Big Data contribui significativamente para a inovação em diferentes setores. No entanto, sua implementação requer investimentos em infraestrutura tecnológica, políticas de governança de dados e equipes qualificadas. A ausência desses elementos pode comprometer tanto a eficácia das análises quanto a segurança das informações.

### CONSIDERAÇÕES FINAIS

O estudo permitiu compreender que o Big Data se consolidou como ferramenta essencial na era digital, proporcionando às organizações capacidade de extrair valor estratégico de grandes volumes de informações. Os 5V's representam um guia conceitual fundamental para lidar com os desafios e oportunidades do tema.

O estudo de caso realizado demonstrou, em ambiente acadêmico, como metodologias e ferramentas podem ser aplicadas para integrar e analisar dados reais, confirmando a aplicabilidade do Big Data no suporte a decisões estratégicas.

Conclui-se que, apesar de seu grande potencial, a adoção do Big Data exige cautela quanto a custos de implementação, complexidade operacional e segurança da informação. Como perspectiva futura, recomenda-se ampliar pesquisas práticas em diferentes áreas, além de investir em políticas de proteção de dados e formação profissional especializada.

## REFERÊNCIAS

GARTNER. *3D Data Management: Controlling Data Volume, Velocity, and Variety*. Stamford: Gartner, 2001. Disponível em: <https://www.gartner.com/doc/491061>. Acesso em: 20 maio 2025.

KATAL, A.; WAZID, M.; GOUDAR, R. H. Big Data: Issues, challenges, tools and good practices. In: *Sixth International Conference on Contemporary Computing (IC3)*. IEEE, 2013. p. 404-409.

KITCHIN, R. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Londres: SAGE, 2014.

MARR, B. *Big Data: Using smart big data, analytics and metrics to make better decisions and improve performance*. Chichester: Wiley, 2015.

CHEN, M.; MAO, S.; LIU, Y. Big Data: A survey. *Mobile Networks and Applications*, v. 19, n. 2, p. 171–209, 2014.

APOLLO SOLUTIONS. *Big Data: o que é e quais as técnicas utilizadas*. 2023. Disponível em: <https://apollosolutionsdev.com/big-data-o-que-e-e-quais-as-sao-as-tecnicas-utilizadas%EF%BF%BC/>. Acesso em: 20 maio 2025.

CIÊNCIA E DADOS. *Análise de sentimentos e machine learning*. 2023. Disponível em: <https://www.cienciaedados.com/analise-de-sentimentos-e-machine-learning/>. Acesso em: 20 maio 2025.

MULTIVIX. *Os 5 V's do Big Data: Impactos e aplicações na era digital*. *Revista Dimensão Acadêmica*, v. 4, n. 2, 2020. Disponível em: <https://multivix.edu.br/wp-content/uploads/2020/07/revista-dimensao-academica-v04-n02-artigo05.pdf>. Acesso em: 20 maio 2025.

TOTVS. *Os 5 V's do Big Data: entenda os conceitos*. 2023. Disponível em: <https://www.totvs.com/blog/inovacoes/big-data/>. Acesso em: 20 maio 2025.

XP EDUCAÇÃO. *O que é Big Data?*. 2023. Disponível em: <https://blog.xpeducacao.com.br/o-que-e-big-data/>. Acesso em: 20 maio 2025.