



Data Warehouse e Business Intelligence

Metodologias, Técnicas e Ferramentas para Tomada de Decisão Estratégica

By Prof. Cloves Alves da Rocha (<https://linktr.ee/clovesrocha>). Graduado em Gestão de Tecnologia da Informação (2014/GTI-FG) Mestre em Ciência da Computação (2017/CIn-UFPE), Professor Universitário (Graduação e Pós), Cientista de Dados (2022/Edx **Harvard**) certificado na escola de mestres (Seminário Teológico Carisma), **Florida University of Science and Theology (FUST)** **Florida University of Science and Theology (FUST)**, PhD Candidate in Theology, professor mestre responsável da Escola Bíblica Dominical (E.B.D. EFATÁ) da Comunidade Batista EFATÁ (@c.b.efata).

Visão Geral da Disciplina

Ementa e Objetivos

Esta disciplina explora os processos de negócio e tomada de decisão através do Business Intelligence, abordando desde os fundamentos teóricos até a implementação prática de soluções de Data Warehouse. Os alunos desenvolverão competências críticas para avaliar e utilizar metodologias, técnicas e ferramentas empregadas na área de DW e BI, capacitando-se para construir bases de dados robustas voltadas à extração de conhecimento estratégico.

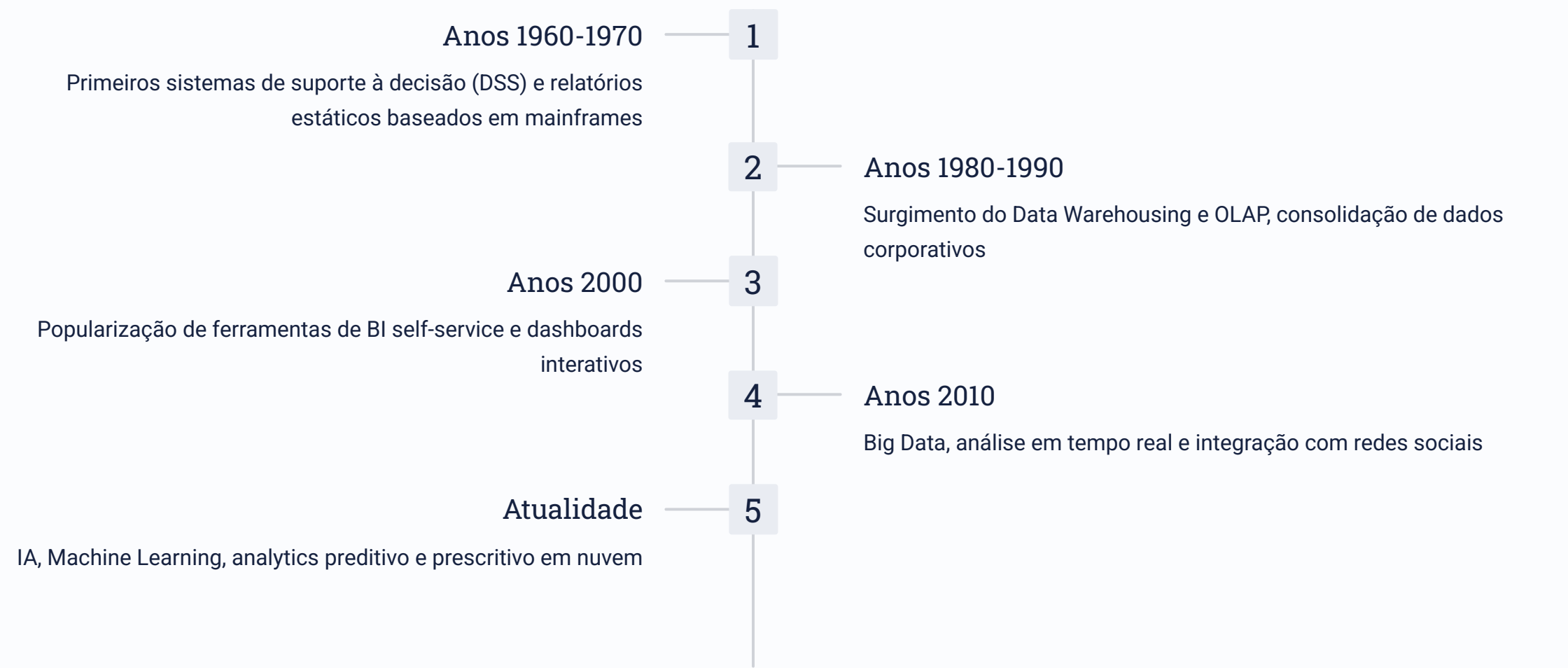
O curso integra teoria e prática, enfatizando o desenvolvimento de aplicações reais que preparam os estudantes para os desafios do mercado de trabalho em análise de dados e inteligência de negócios.

Competências Desenvolvidas

- Avaliação crítica de metodologias de DW e BI
- Domínio de técnicas e ferramentas especializadas
- Construção de bases de dados para extração de conhecimento
- Implementação de soluções OLAP multidimensionais
- Desenvolvimento de aplicações práticas de BI

Evolução do Business Intelligence

A evolução do Business Intelligence representa uma jornada fascinante desde os primeiros sistemas de relatórios até as plataformas avançadas de análise preditiva e prescritiva que conhecemos hoje.



Esta evolução reflete a crescente necessidade das organizações por insights acionáveis derivados de volumes cada vez maiores de dados, impulsionando a transformação digital dos negócios.

Unidade I: OLAP e Modelos Multidimensionais

Introdução ao Apoio à Decisão

Sistemas de apoio à decisão (SAD) fornecem informações estruturadas e análises que auxiliam gestores na tomada de decisões estratégicas. Diferentemente dos sistemas transacionais (OLTP), que focam em operações do dia a dia, os sistemas OLAP são otimizados para consultas complexas e análises multidimensionais.

Consultas de Agregação

Operações como drill-down, roll-up, slice, dice e pivot permitem navegar pelos dados em diferentes níveis de granularidade. Essas consultas agregam informações de vendas, financeiras ou operacionais, revelando tendências e padrões ocultos nos dados corporativos.

Modelos de Dados Multidimensionais

Os modelos multidimensionais organizam dados em cubos, permitindo análises por múltiplas perspectivas (dimensões). Incluem esquemas estrela (star schema) e floco de neve (snowflake), que estruturam fatos e dimensões para consultas eficientes e análises rápidas de grandes volumes de dados.

Técnicas de Implementação OLAP

ROLAP (Relational OLAP) utiliza bancos relacionais, MOLAP (Multidimensional OLAP) usa cubos pré-calculados, e HOLAP combina ambas as abordagens. A escolha depende do volume de dados, requisitos de performance e complexidade das análises necessárias.

Unidade II: Data Warehousing

Fundamentos de Data Warehousing

Um Data Warehouse (DW) é um repositório centralizado que integra dados de múltiplas fontes heterogêneas, organizando-os de forma otimizada para análise e geração de relatórios. Diferentemente dos bancos de dados operacionais, o DW é orientado por assunto, integrado, variante no tempo e não-volátil.

A arquitetura típica inclui camadas de staging (área de preparação), integração e apresentação. Os processos ETL (Extract, Transform, Load) são fundamentais para garantir a qualidade, consistência e integração dos dados provenientes de sistemas transacionais diversos.

Visões e Materialização

Visões materializadas armazenam resultados de consultas complexas, acelerando significativamente o tempo de resposta. A escolha de quais visões materializar envolve análise de custo-benefício entre espaço de armazenamento e ganho de performance.

Data Marts e Especialização

Data Marts são subconjuntos do DW focados em áreas específicas de negócio (vendas, marketing, finanças). Oferecem maior agilidade e performance para departamentos específicos, com modelagem otimizada para necessidades particulares de análise.

A implementação pode seguir abordagens top-down (do DW para os Data Marts) ou bottom-up (dos Data Marts para o DW), cada uma com vantagens específicas em termos de custo, tempo e complexidade de implementação.

Estratégias de Atualização

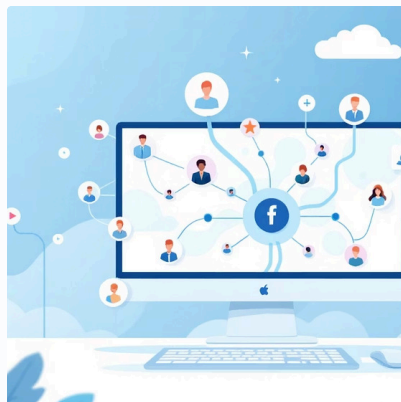
A manutenção de visões materializadas pode ser incremental ou completa, síncrona ou assíncrona. A escolha depende da frequência de atualização dos dados fonte e dos requisitos de latência da informação.

Unidade III: Mineração de Dados da Web Social



Extração de Dados do Twitter/X

A API do Twitter permite coletar tweets, hashtags, menções e métricas de engajamento. Técnicas de processamento de linguagem natural (NLP) identificam sentimentos, tendências e tópicos emergentes. Análise de grafos revela redes de influência e propagação de informação.



Análise de Dados do Facebook

Graph API possibilita extrair dados de páginas, grupos e interações. Métricas de alcance, engajamento e demografia fornecem insights sobre audiências. Análise de conteúdo identifica posts de alto desempenho e padrões de interação.



Mineração no Instagram

Análise de conteúdo visual usando computer vision identifica objetos, cenários e emoções em imagens. Hashtags e geolocalização revelam tendências regionais. Métricas de engajamento indicam preferências e comportamentos de consumo visual.

A mineração de redes sociais oferece insights valiosos sobre comportamento do consumidor, tendências de mercado, gestão de reputação de marca e análise de sentimento em tempo real. Ferramentas como Python (Tweepy, BeautifulSoup), R e plataformas especializadas facilitam a coleta, processamento e análise desses dados.

Aplicações Práticas: Análise de Sentimento

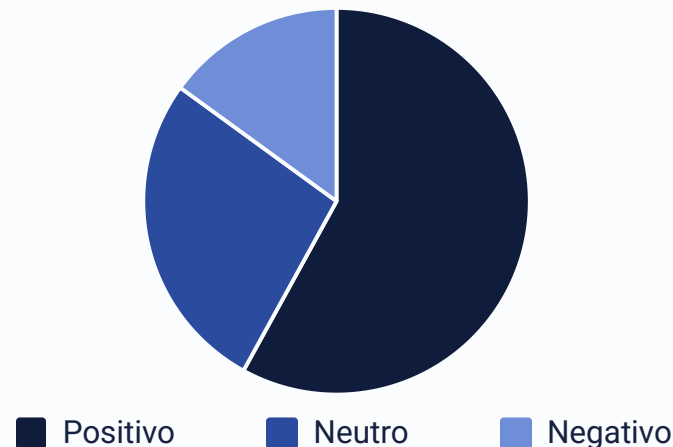
Exemplo Prático: Monitoramento de Marca

Considere uma empresa que deseja monitorar a percepção de seu produto nas redes sociais. O processo envolve:

1. Coleta de dados via APIs (Twitter, Facebook, Instagram)
2. Pré-processamento: limpeza, normalização, remoção de stopwords
3. Análise de sentimento usando modelos de NLP (VADER, BERT)
4. Classificação em positivo, negativo ou neutro
5. Visualização em dashboards interativos
6. Geração de alertas para sentimentos negativos

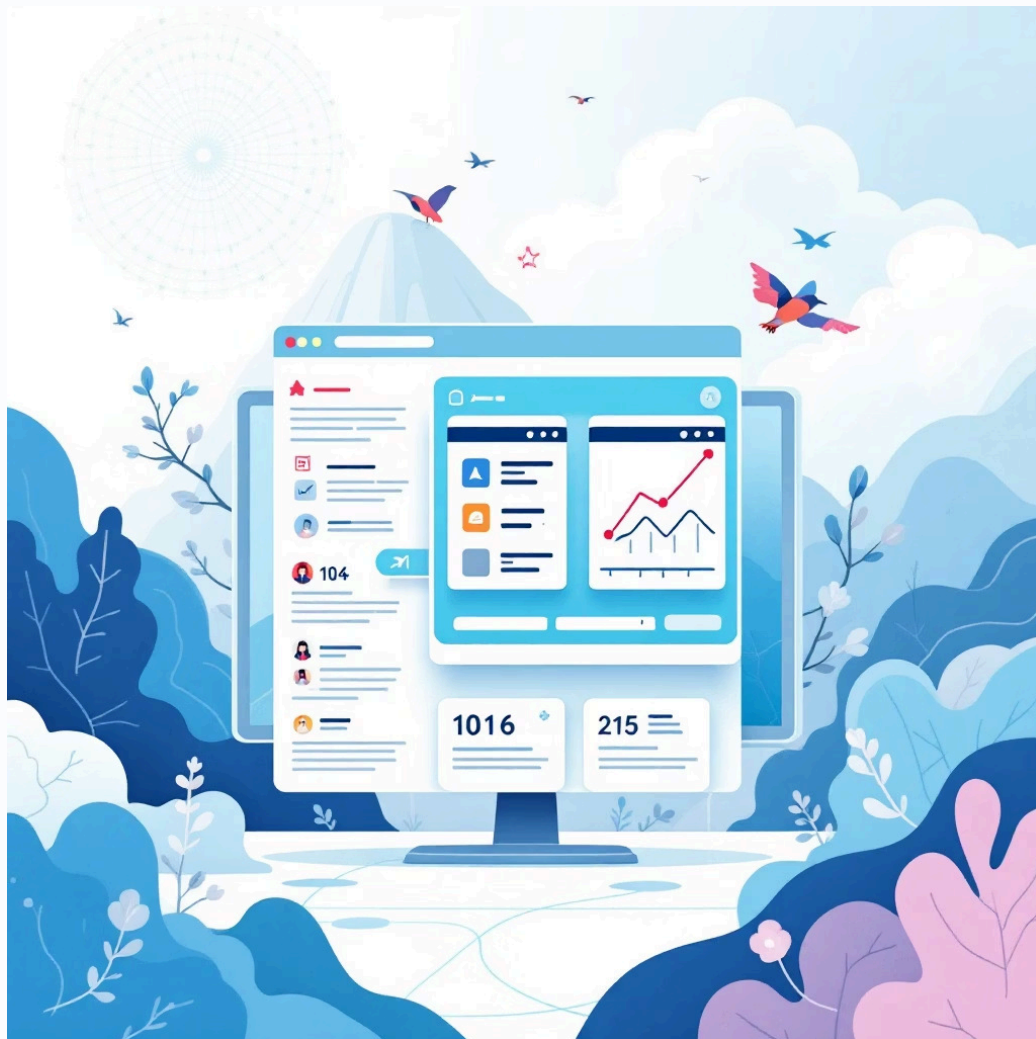
Bibliotecas Python como NLTK, TextBlob, spaCy e Transformers facilitam a implementação. O resultado permite respostas rápidas a crises de reputação e identificação de oportunidades de melhoria.

Métricas de Avaliação



A distribuição de sentimentos permite avaliar a percepção geral da marca e identificar áreas que requerem atenção imediata.

Unidade IV: Mineração de Dados Abertos



O Poder dos Dados Abertos

Dados abertos governamentais e públicos representam uma fonte rica de informações para análise e desenvolvimento de aplicações. Portais como dados.gov.br disponibilizam datasets sobre saúde, educação, segurança, economia e infraestrutura.

A mineração desses dados permite:

- Análise de tendências socioeconômicas
- Desenvolvimento de aplicações de interesse público
- Fiscalização e transparência governamental
- Pesquisa acadêmica e inovação social
- Tomada de decisões baseada em evidências

Formatos comuns incluem CSV, JSON, XML e APIs RESTful. Ferramentas como Pandas, R e Tableau facilitam a exploração e visualização desses dados.

Ferramentas para DW e Business Intelligence



Power BI

Plataforma da Microsoft para visualização e análise de dados. Oferece integração nativa com Excel, Azure e outras fontes. Interface intuitiva de drag-and-drop para criação de dashboards interativos e relatórios dinâmicos.



Tableau

Líder em visualização de dados com capacidades avançadas de storytelling visual. Suporta múltiplas fontes de dados e permite criar visualizações complexas sem programação. Excelente para análises exploratórias.



Pentaho

Suite open-source completa para ETL, análise e mineração de dados. Pentaho Data Integration (Kettle) é amplamente usado para processos de transformação de dados. Oferece recursos de Big Data e integração com Hadoop.



QlikView/Qlik Sense

Plataforma associativa que permite explorar dados de forma livre sem limitações de hierarquias pré-definidas. Motor de análise in-memory proporciona respostas rápidas em grandes volumes de dados.



Python (Pandas, Scikit-learn)

Ambiente programático para análise de dados e machine learning. Bibliotecas como Pandas, NumPy, Matplotlib e Scikit-learn oferecem flexibilidade total para manipulação, análise e modelagem de dados.



Apache Spark

Framework de processamento distribuído para Big Data. Oferece APIs para análise em larga escala, machine learning (MLlib) e processamento de streams. Ideal para ambientes com volumes massivos de dados.

Metodologia de Ensino e Avaliação

Abordagem Andragógica

A disciplina adota metodologia de ensino e aprendizagem centrada no estudante, combinando atividades síncronas e mediadas. A organização da mediação entre o graduando e os conteúdos programáticos utiliza diversas ferramentas e estratégias para desenvolver competências teóricas e práticas.

01

Exposições Teóricas

Apresentação de conceitos fundamentais de DW e BI

02

Estudos de Caso

Análise de implementações reais no mercado

03

Laboratórios Práticos

Desenvolvimento hands-on com ferramentas

04

Projetos Aplicados

Solução de problemas reais de negócio

Sistema de Avaliação

A avaliação do processo de ensino e aprendizagem incide sobre frequência e aproveitamento:

- **Frequência mínima obrigatória:** 75% da carga horária.
- **Aproveitamento escolar:** Verificações parciais durante o período letivo.
- **Sistema de notas:** Escala de 0 a 10
- **Avaliação Final:** Quando necessário, conforme regulamento.
- **Prova Colegiada:** Modalidade opcional conforme regulamento específico.

As verificações parciais incluem provas teóricas, projetos práticos, apresentações e atividades em laboratório, garantindo avaliação abrangente das competências desenvolvidas.