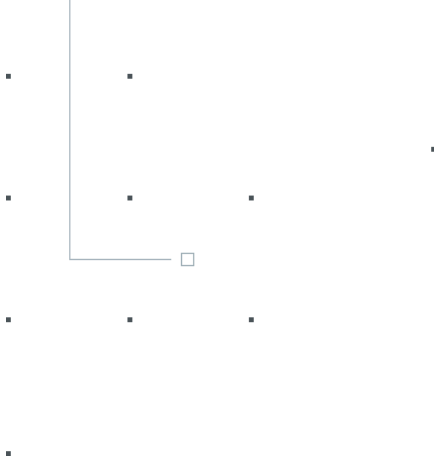
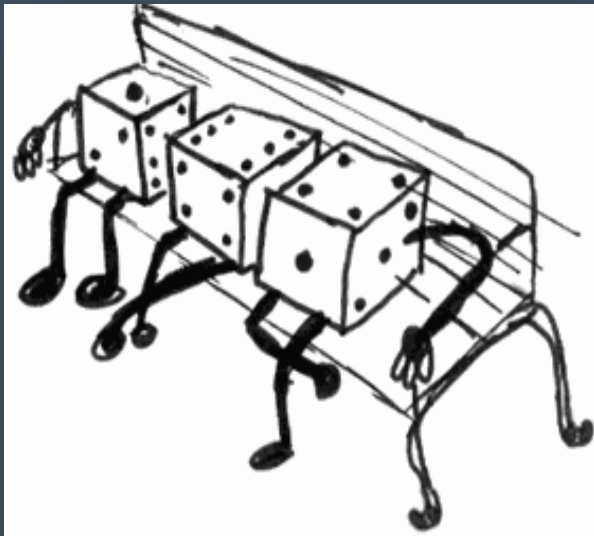


FIAP



Data Engineering



APRESENTAÇÃO

- Apresentação da professora;
- Objetivo da Disciplina;
- Conteúdo Programático;
- Metodologia Aplicada;
- Referências Bibliográficas.

Tassiana Rugoni de Campos

Short Bio

Big Data | Machine Learning | Business Intelligence SME | MBA Coordinator | MBA Professor | Startup Mentor

- **Consultora** na área de dados e **professora** há cerca de **23 anos**;
- Atuou como **Head of Data & Analytics** na **triggo.ai**;
- Coordenadora de **MBA** na **FIAP**;
- **Mentora** do programa de aceleração de **Startups**, o Amplifier da Rock New Ventures;

Autora de livros didáticos sobre:

- Big Data Development;
 - Arquitetura de Data Lake e Streaming de Dados;
 - Tecnologias Disruptivas (Blockchain, Metaverso, IAs Generativas, IoT, Cloud);
 - Think Deep (machine learning);
 - Data Modeling;
 - Data Integration & Ingestion;
 - Qualidade de Software – CMMI;
 - Desenvoltimentos: Mobile Android (Java), Python, Web (PHP/SQL) e Jogos (Unity – C#).
-
- **Professora** nos MBAs de Arquitetura de Soluções, BI, Big Data, IA, Eng. Software, Eng. De Dados, Gestão de TI, Cloud, BPM e Gestão de Negócios;
 - Professora responsável por curso de graduação em Análise e Desenvolvimento de Sistemas;
 - Experiência profissional na **gestão** de equipe de infraestrutura;
 - E gestão de times de Dados e Geotecnologia.



www.linkedin.com/in/tassianarugoni
proftassiana.campos@fiap.com.br

*Tassiana é apaixonada por saber e fazer saber,
investindo em pessoas e sendo agente de
transformação através do conhecimento!*

OBJETIVOS DA DISCIPLINA

- Aperfeiçoar os conhecimentos envolvidos em um típico ambiente analítico.
- Criar situações práticas baseadas em estudos de casos para capacitar o aluno a definir uma Arquitetura de Dados que suporte à tomada de decisão.
- Apresentar os principais conceitos envolvendo Repositórios para Armazenamento de Indicadores de Desempenho, sua arquitetura e utilização visando aumentar o nível de maturidade no tratamento de informações nas empresas.
- Capacitar o aluno a tomar decisões, de maneira isolada ou em grupo na fase de concepção da Arquitetura de uma solução analítica.



RESUMO DO CONTEÚDO PROGRAMÁTICO

- Contextualização do cenário atual de dados (compreensão do mercado);
- Conceitos sobre BI e BA (DW, Data Marts, Modelagem Dimensional, Flatten);
- Conceitos sobre Big Data (Data Lake, Hadoop, Map Reduce, HDFS, Data Lakehouse, Spark, Delta);
- Bancos SQL versus NoSQL versus NewSQL;
- ACID x BASE;
- Teorema CAP
- Ingestão e Streaming de Dados.

METODOLOGIA APLICADA

- Aula expositiva para apresentação do conteúdo através da exposição de slides.
- Aula expositiva para apresentação do conteúdo através da exposição de slides.
- Debates em sala de aula visando troca de experiências.
- Hands On com Aplicação de Estudos de caso.



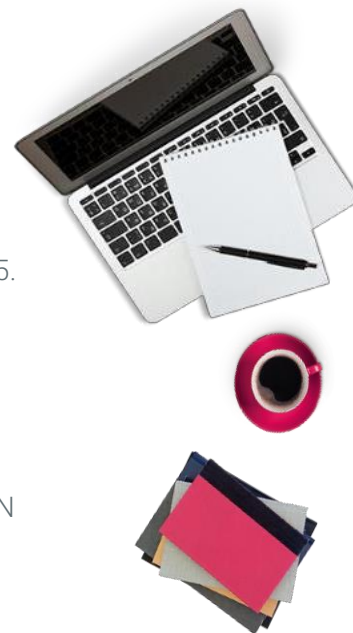
REFERÊNCIAS BIBLIOGRÁFICAS

- FOREMAN, J. W. **Data smart: usando data science para transformar informação em insight**. Rio de Janeiro: Alta Books, 2016.
- KNAFLIC, C. N. **Storytelling com dados: um guia sobre visualização de dados para profissionais de negócios**. Rio de Janeiro: Alta Books, 2019.
- BARBIERI, Carlos. **Governança de dados: práticas, conceitos e novos caminhos**. Rio de Janeiro: Alta Books, 2019.
- CLARKE, E.. **Everything Data Analytics-A Beginner's Guide to Data Literacy**. Kenneth Michael Fornari, 2022.
- DEVLIN, B. **Business unIntelligence: Insight and Innovation beyond Analytics and Big Data**. USA: Technics Publications, LLC, 2014.



REFERÊNCIAS BIBLIOGRÁFICAS

- MACHADO, Felipe N. R. **Big Data – O Futuro dos dados e aplicações**. São Paulo: Erica, 2018.
- CASTRO, L., FERRARI, D. **Introdução à mineração de dados**. São Paulo: Saraiva, 2016.
- SIMON, P. **Too Big to Ignore: The Business Case for Big Data**. USA: Wiley, 2015.
- HILLS, T. **NoSQL and SQL Data Modeling: Bringing Together Data, Semantics, and Software**. EUA, Technics Publica, 2016.
- RUSSELL, S; NORVIG, P. **Inteligência Artificial - Uma Abordagem Moderna**. GEN LTC, 2022.



DICA PARA CONSUMO BIBLIOGRÁFICO


← → ↻ oreilly.com

O'REILLY® TEAMS ▾ INDIVIDUALS FEATURES ▾ BLOG CONTENT SPONSORSHIP 🔍 SIGN IN Try Now >

The **learning** platform your tech teams need to stay ahead

More than 60% of all Fortune 100 companies trust O'Reilly to help their teams learn the technologies of today—and be ready for what's next. We can help yours too.

Request a demo > Try it free >



<https://learning.oreilly.com/>



AULA 1 - OBJETIVOS

- Contextualizar o Cenário de Dados;
- Compreender o papel do Engenheiro em Soluções de Dados;
- Exercitar conceitos de integração de dados e modelagem Flatten utilizando Alteryx.

Uma Cultura Data Driven

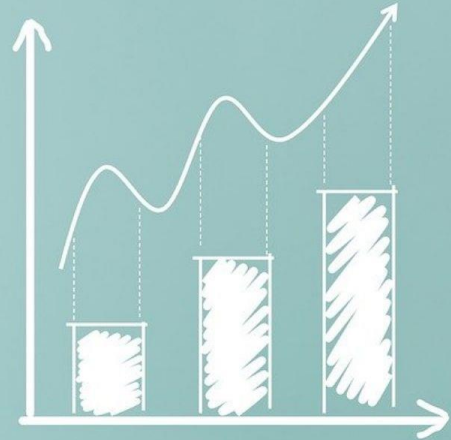


Como a evolução tecnológica impacta no modo com o qual lidamos com os dados?



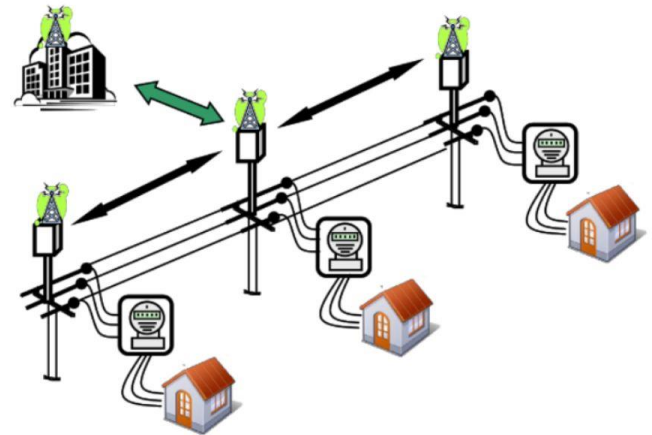
Crescimento Exponencial da Captura de Dados

Novas Fontes e Tipos de Dados



1 - Sensoriamento de Objetos e Captura de Eventos

Smart Grid



Smart City

Centrica implanta cluster Hortonworks em grande escala para impulsionar negócios

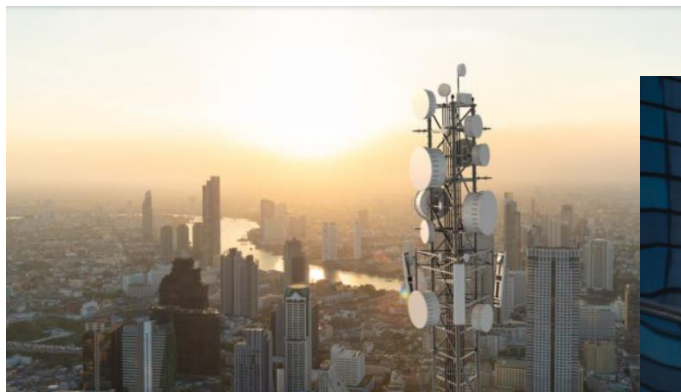
A empresa começou a trabalhar com a fabricante de caldeiras de aquecimento central Worcester Bosch [para coletar diagnósticos de caldeiras remotamente](#). Com essa caldeira conectada, o sistema de termostato de internet Hive da empresa começa a se tornar um serviço, disse Hussain.

Como a Computer Weekly informou anteriormente, a [divisão britânica de gás da empresa usou painéis do QlikView](#) no data lake do Hadoop para suportar seu programa de [medidores inteligentes](#).

O QlikView fornece análise de autoatendimento sobre um lago de dados do Hadoop. O data lake contém mais de [nove bilhões de registros](#) e coleta dados de fontes, incluindo mais de [150 tabelas SAP](#).



Os benefícios do 5G e da IA nos Smart Grids



e-commerce

Streaming de Dados - Real time

Tipo de Produto

- ☐ Calças
- ☐ Camisetas
- ☐ Shorts
- ☐ Tênis

Tamanho

EP	P	M	G
GG	34	35	3

Preço

- ☐ R\$ 90 - R\$ 180
- ☐ R\$ 180 - R\$ 250



Tênis Nike Air Zoom Structure 24 Masculino - Cinza

Ref.: 21C-7449-010-42

R\$ 899,99
até 10x de R\$ 90,00 sem juros

 R\$ 899,99 com N Card,
ou em 15x de R\$ 60,00 [Peça o seu](#)

Vendido e Enviado por
Nike

Cor: Cinza



Tamanho:

37	38	39	40	41	42	43	44
45	46	47	48	49	50	51	

ADICIONAR AO CARRINHO

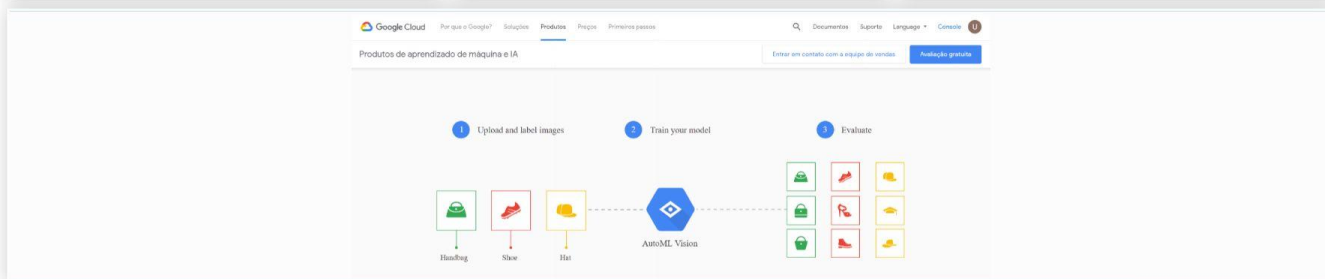
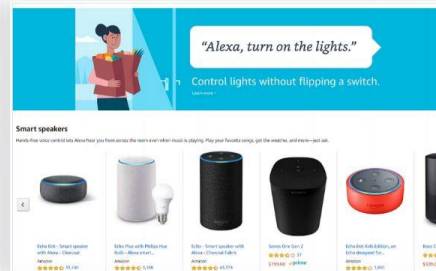
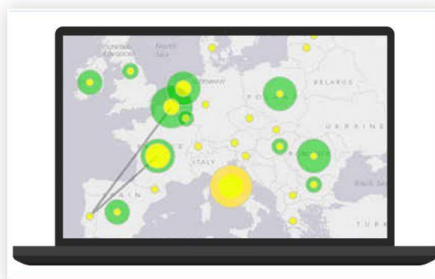


-14%

m Terra Kiger 8

MADE WITH
beautiful.ai

Novas Fontes e Tipos de Dados



Modernas arquiteturas de telecom integradas com Hadoop

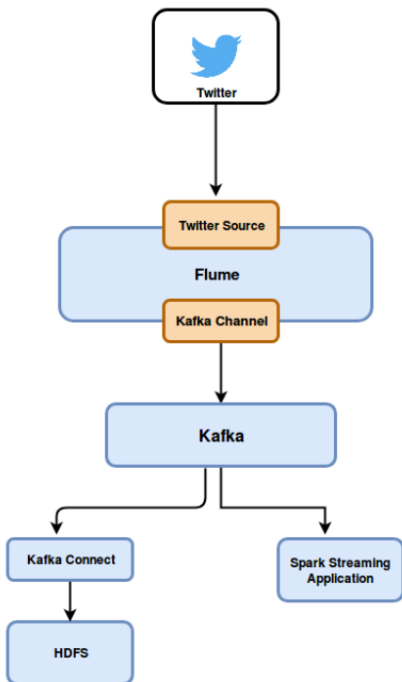
Muitas das maiores empresas de telecomunicações do mundo usam o [Hortonworks Data Platform](#) (HDP) para gerenciar seus dados. Por meio da parceria com essas empresas, aprendemos como nossos clientes usam o HDP para melhorar a satisfação do cliente, fazer melhores investimentos em infraestrutura e desenvolver novos produtos.

A parceira da Hortonworks, [Teradata](#), recentemente deu alguns [exemplos de casos de uso neste vídeo](#) sobre como a Verizon Wireless usa o Teradata em combinação com a Hortonworks Data Platform para manter a queda de seus clientes abaixo de 1%.

Rob Smith, diretor executivo de TI da Verizon Wireless, descreve como sua equipe usa sua plataforma de descoberta para melhorar as interações com os clientes:

- Identificando maneiras melhores de se comunicar com os clientes sobre seus pagamentos
- [Analisando as mídias sociais para entender melhor os sentimentos dos clientes sobre as mudanças na política da Verizon](#)
- Adaptar as comunicações de marketing às necessidades individuais de cada cliente

Análise de Sentimento no Twitter|X



A captura de tela mostra uma interface de usuário com uma barra superior contendo três botões: APIX, API de anúncios X e X para sites. O título principal é "Acesso em tempo real à conversa global". Abaixo do título, há um texto explicativo: "Use a API X para ouvir e analisar conversas públicas, interagir com pessoas no X e inovar. Aqui você pode ver como o endpoint de pesquisa recente foi usado para encontrar postagens com palavras-chave específicas."

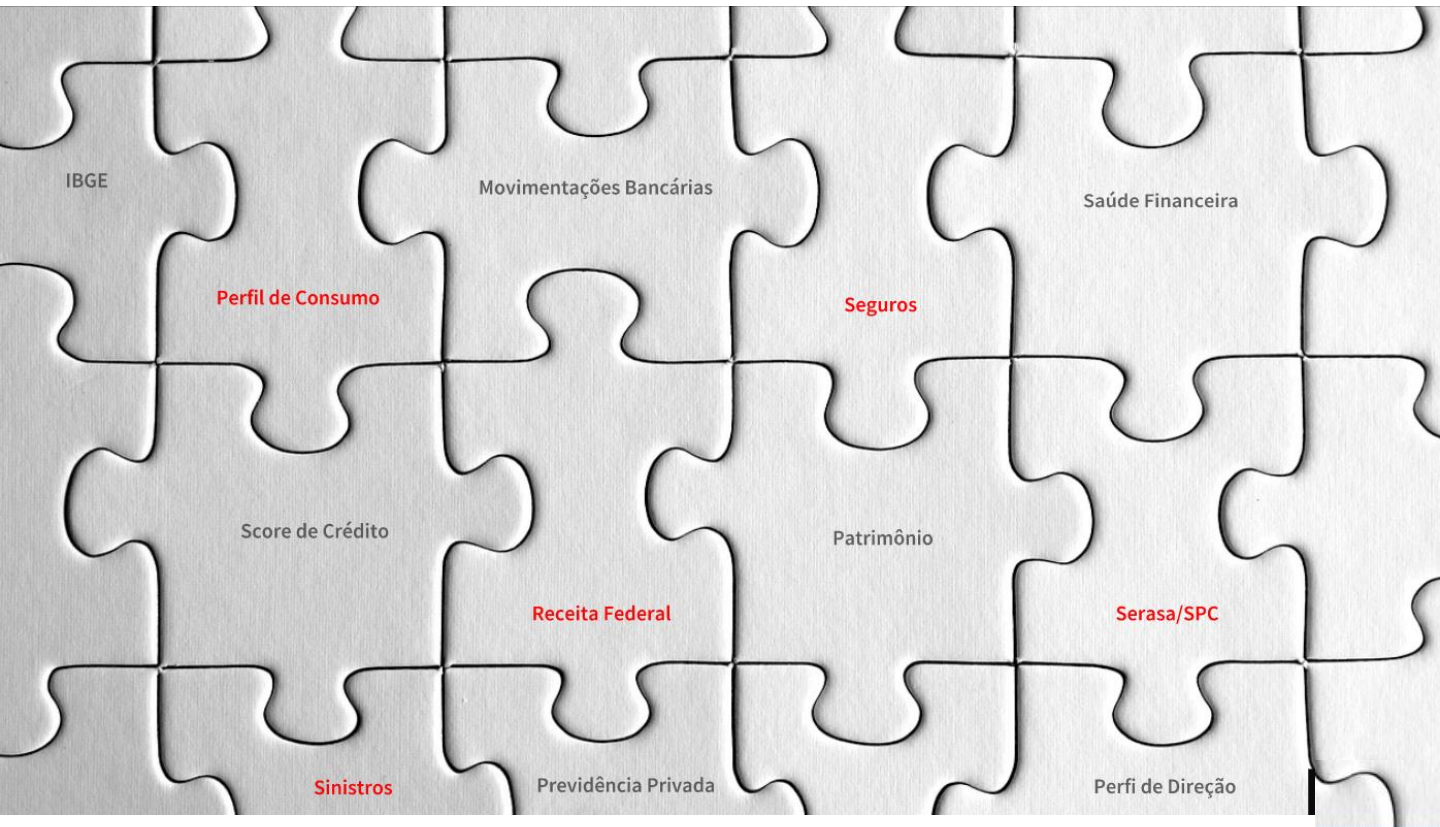
Exemplo de solicitação Exemplo de resposta

```
curl --request GET --location
'https://api.twitter.com/2/tweets/search/recent?
tweet.fields=context_annotations&max_results=100&query=camping(nature%20
--header 'Authorization: Bearer $BEARER_TOKEN'
```

An aerial photograph of a winding asphalt road that curves through rolling green hills. The hills are covered in dense vegetation, and the road is flanked by trees and shrubs. The lighting suggests a late afternoon or early morning scene, with long shadows and warm tones. The text "Open Everything" is overlaid in large white letters on the right side of the image.

Open Everything





Insurtechs

Sinistros por Tipo e Localidade

Índices geográficos de criminalidade

Clima

+

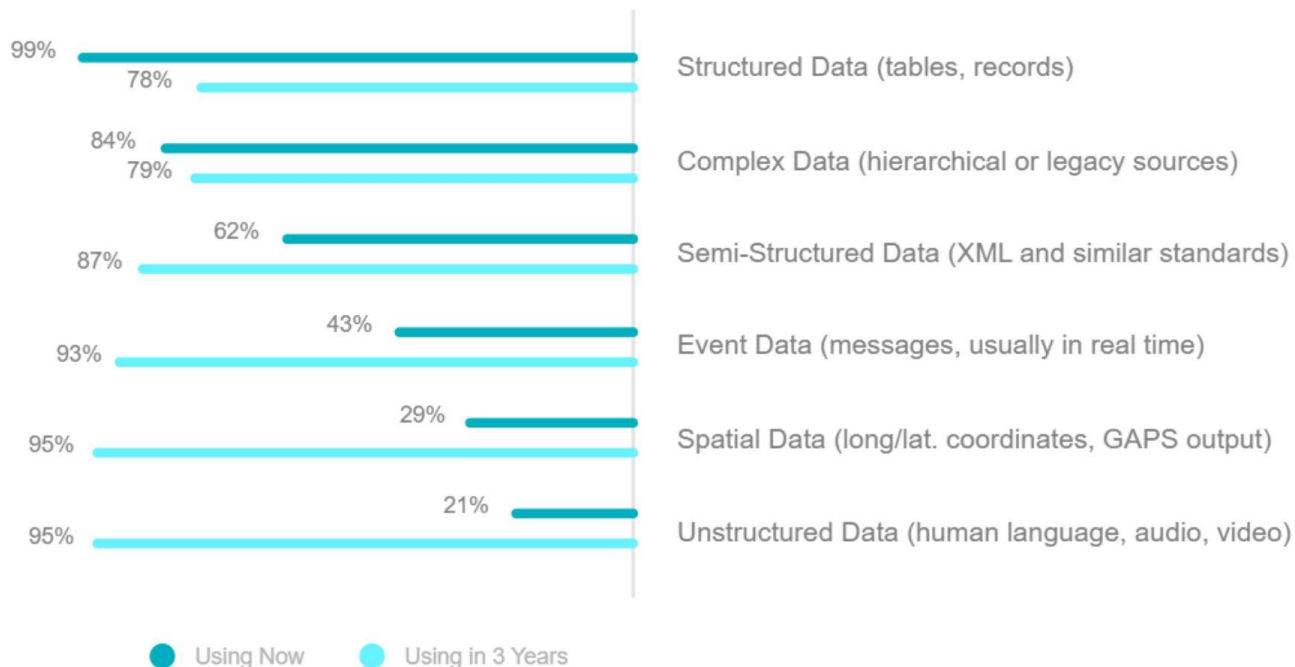
GPS e Sensores no Veículo



**IoT: Até 2025 cada
cidadão brasileiro
terá em torno de
si 7 equipamentos
máquina-máquina**



Participação por Tipos de Dados nos Projetos



O que um engenheiro precisa saber para projetar soluções cuja análise e mineração dos dados sejam vitais?

+

Liderança



■ **Hard Skills**

+

Python

👍 0 💬 1

Enter your comment...

ADD

R também



■ Percentual Negócio/Técnico

+

50%/50%



<https://easyretro.io/publicboard/RkZMDraAlgPWRSSaMmSYqmFZrHJ3/b3610eb2-8f1c-491c-8fb0-35d76522c3aa>

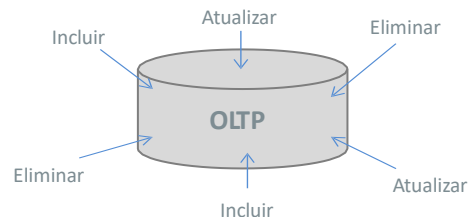
OLTP x OLAP x OLAP com Big Data



AMBIENTES TRANSACIONAIS

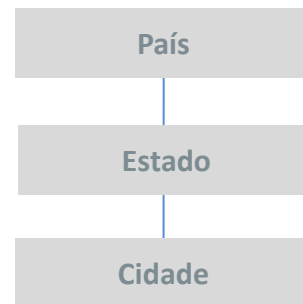
OLTP (*On-Line Transaction Processing*)

- Sistemas que tratam o negócio: CRMs, ERPs, entre outros.
- Ênfase nas operações diárias.
- Bases orientadas à escrita.



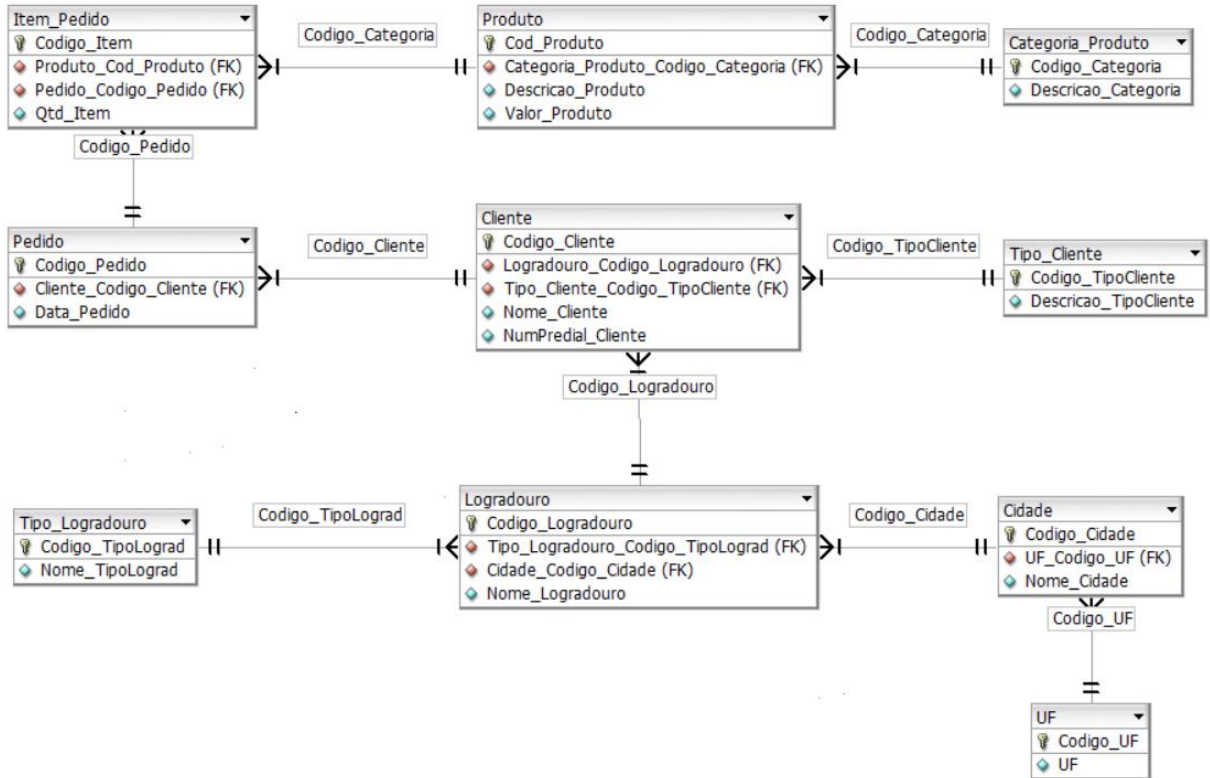
Modelagem de Dados – OLTP

- Sem repetições ou redundâncias.
- Terceira forma normal.
- Chaves de Negócio.
- Tabelas de Lookup.



AMBIENTES TRANSACIONAIS

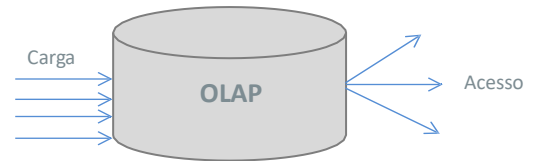
Modelagem de Dados OLTP



AMBIENTES ANALÍTICOS

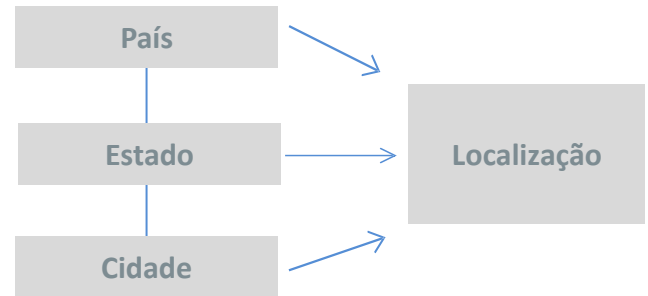
OLAP (On-Line Analytical Processing)

- Sistemas que analisam o negócio: relatórios, painéis, indicadores de desempenho.
- Dados Históricos.
- Bases orientadas à consulta.



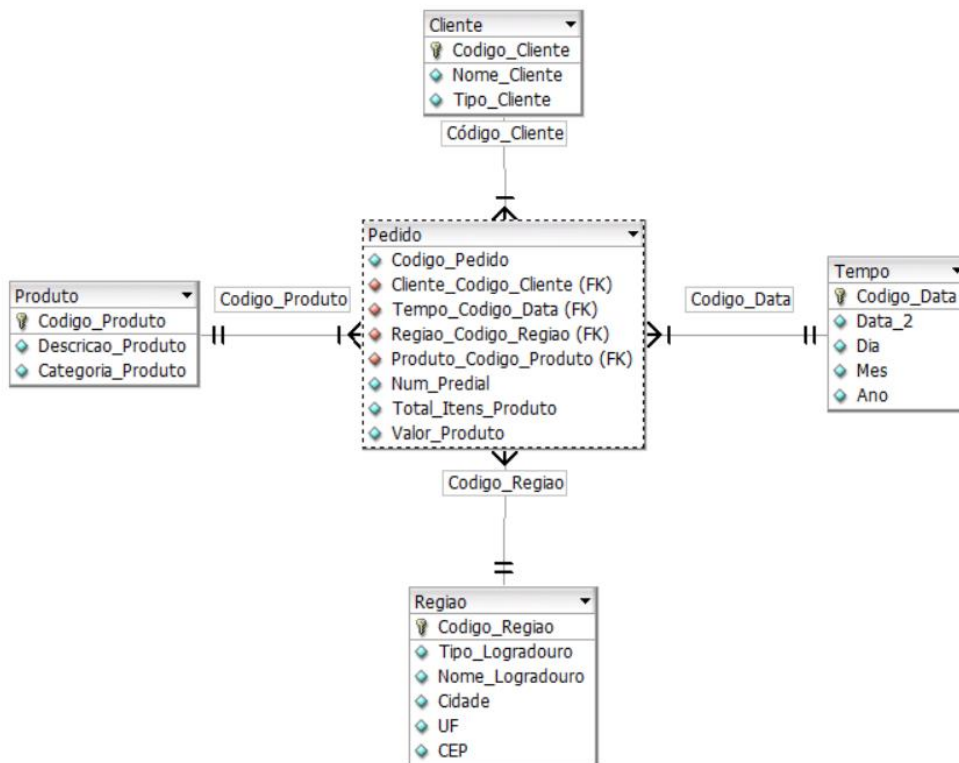
Modelagem Dimensional

- Permite redundâncias.
- Desnormalização.
- Chaves Substitutas.
- Modelo orientado aos processos de negócio.



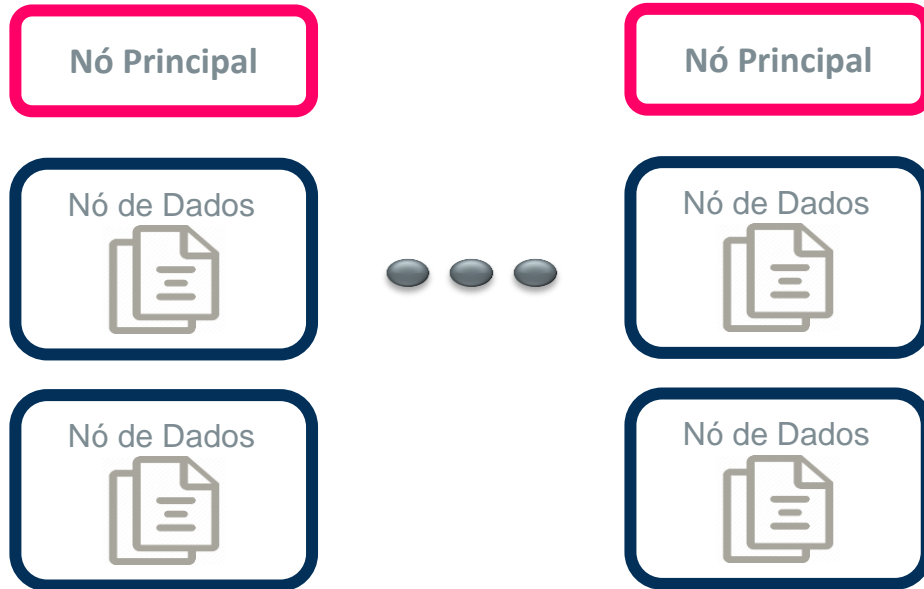
AMBIENTES ANALÍTICOS

Modelagem Multidimensional



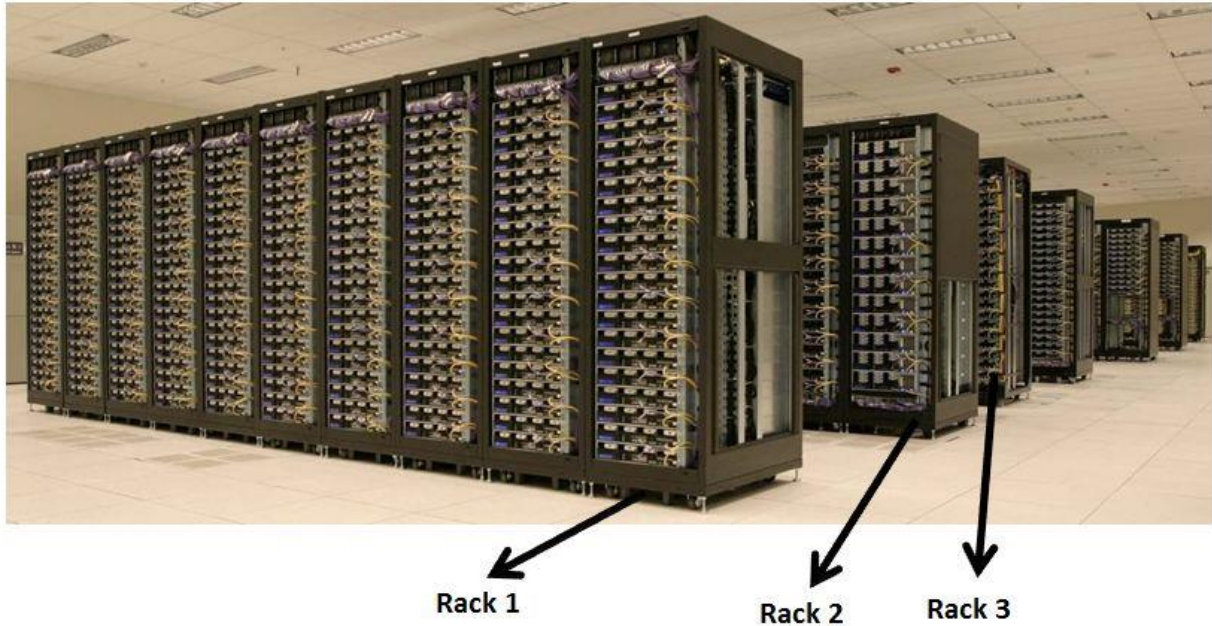
AMBIENTES ANALÍTICOS COM GRANDE VOLUME DE DADOS

HDFS - Hadoop Distributed File System



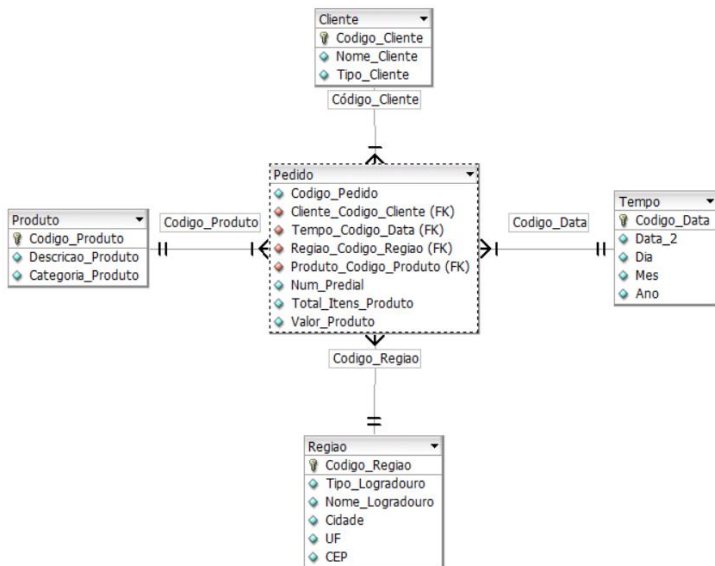
ENTENDENDO A ARQUITETURA HADOOP

Cluster Hadoop



AMBIENTES ANALÍTICOS COM GRANDE VOLUME DE DADOS

Modelagem Multidimensional



Modelagem Flatten





Hands On

VAMOS APLICAR?

Transforme o log de acesso (dados semiestruturados) em uma estrutura flatten (arquivo csv).



FlattenViews

DataVisualizacao

HoraVisualizacao

Produto

Sistema Operacional

TotalViews

Como foi a sua experiência com a aula de hoje?



<https://fiap.me/AutoglassArquiteturaSoftware>

Hoje foi o início da nossa jornada;)



<https://www.linkedin.com/in/tassianarugoni>



profassiana.campos@fiap.com.br

FIAP

Copyright © 2024 | Professora Tassiana Rugoni de Campos
Todos os direitos reservados. Reprodução ou divulgação total ou parcial deste documento, é expressamente proibido sem consentimento formal, por escrito, do professor/autor.

