Foto de Evelyn Paris no Unsplash

Uma Breve História da Arquitetura de Dados Lakehouse

Lourenço Taborda

FIAP Meetup #5
06.11.2020 19h00

Uma Breve História da Arquitetura de Dados Lakehouse

Lourenço Taborda

FIAP Meetup #5
06.11.2020 19h00

LOURENÇO TABORDA

ORACLE Certified Professional

ORACLE Certified Specialist

cbip Certified Business Intelligence Professional

Inovação com dados em nuvem COORDENADOR

TDC ONLINE

Sala ORACLE

Inovação com dados em nuvem

TRILHA

#TheDevConf Oracle

Foto de Edgar Chaparro no Unsplash

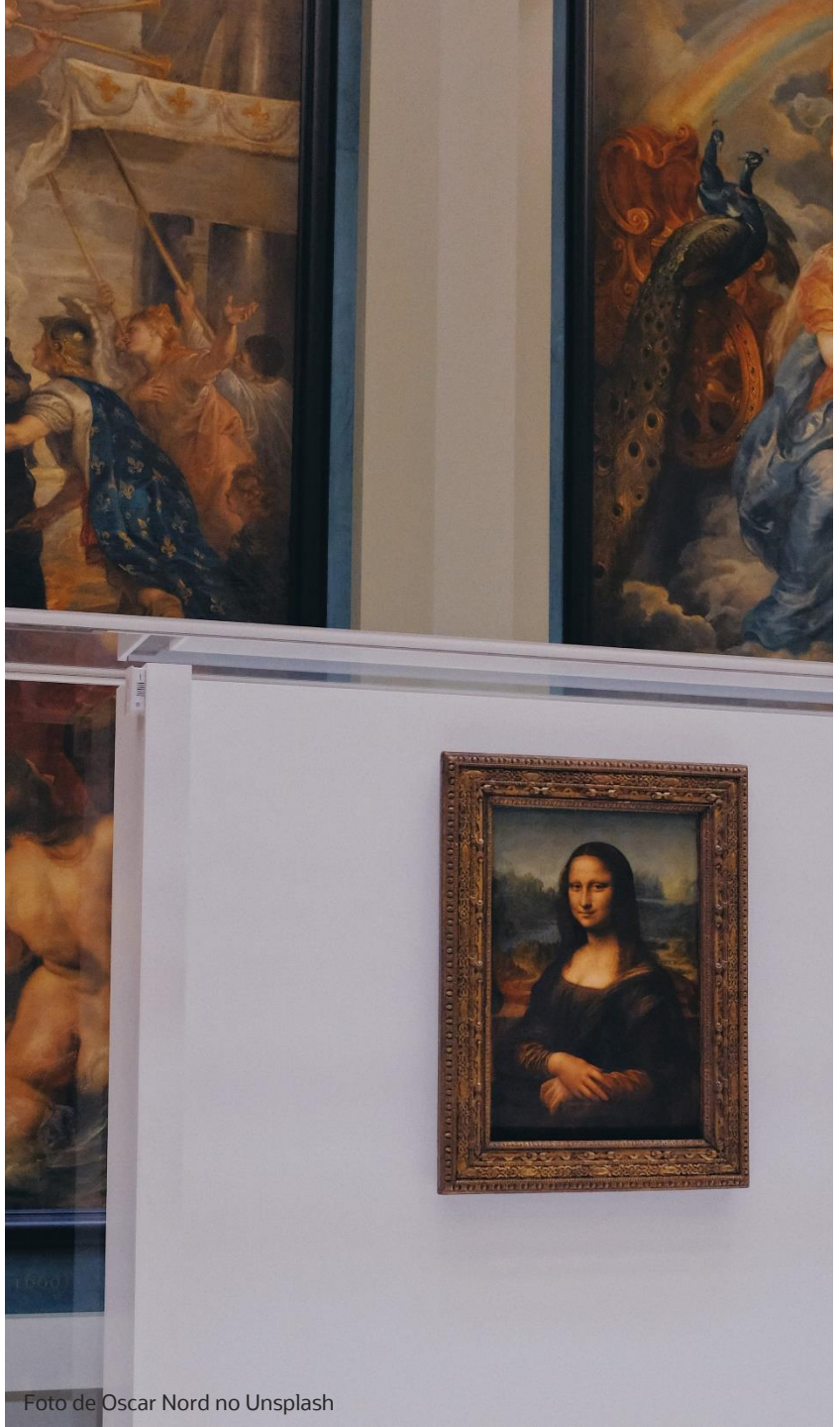
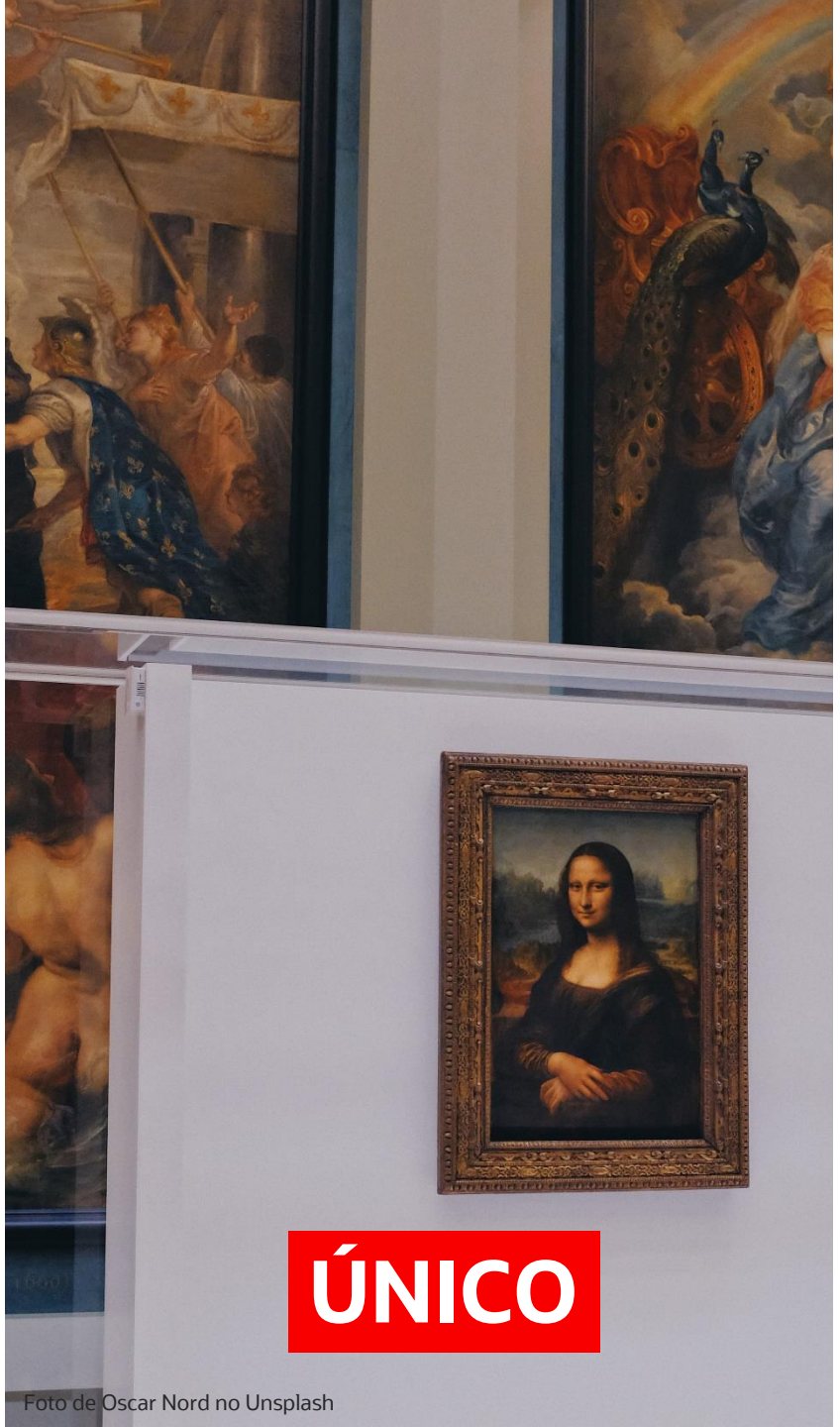


Foto de Oscar Nord no Unsplash

Foto de Sebastian Bjune no Unsplash

EXPERIÊNCIA

Foto de Edgar Chaparro no Unsplash



ÚNICO

Foto de Oscar Nord no Unsplash

NÃO-RIVAL

Foto de Sebastian Bjune no Unsplash

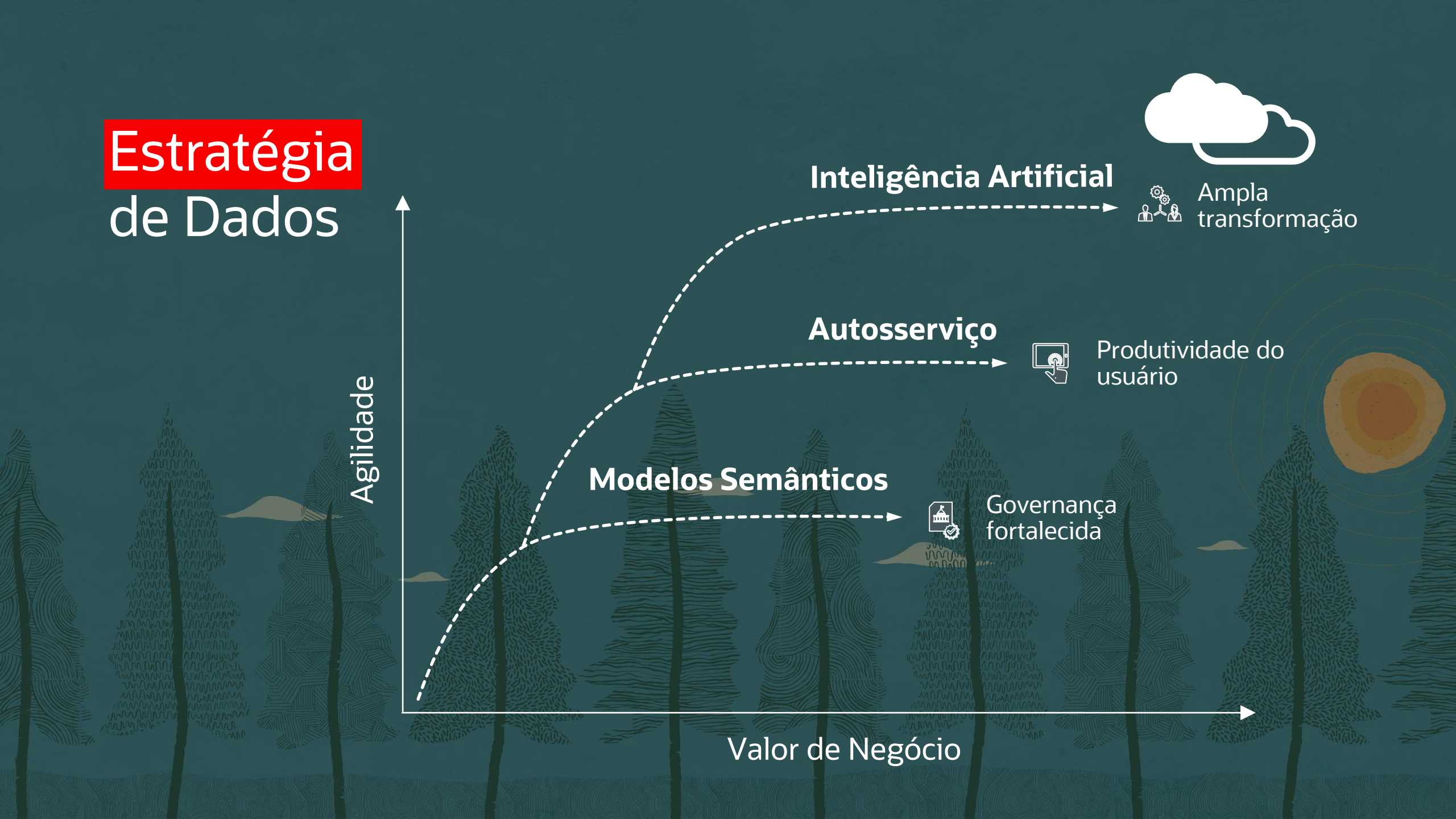
Por que construímos Data Warehouses, Data Lakes e agora estou nesta palestra sobre Lakehouse?

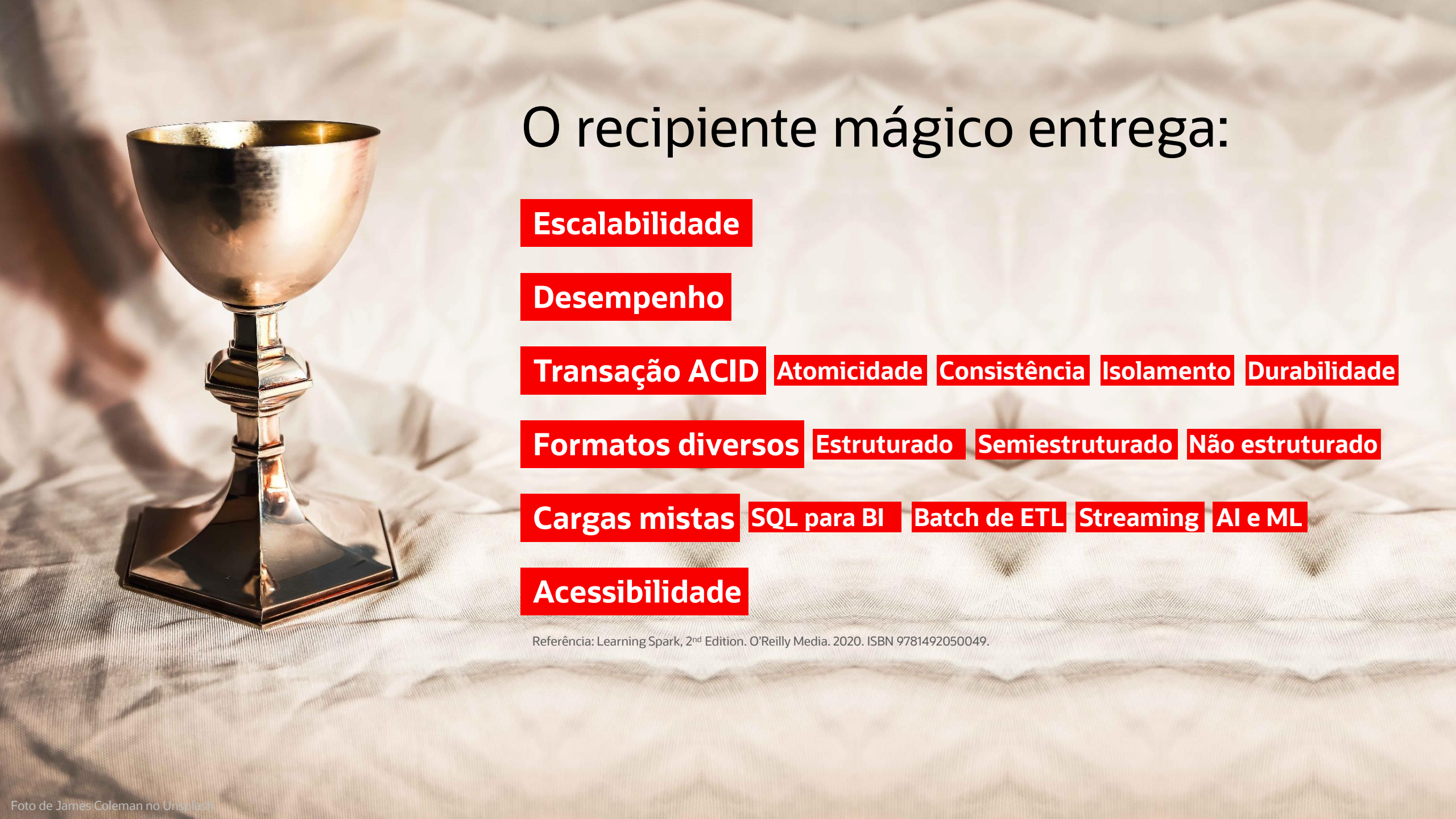
PORQUE DADOS GERAM VANTAGEM COMPETITIVA PARA OS NEGÓCIOS.

DADOS SÃO FATORES DE PRODUÇÃO DE BENS E SERVIÇOS DIGITAIS.

DADOS SÃO UM BEM DE EXPERIÊNCIA, INFUNGÍVEL E NÃO-RIVAL

Estratégia de Dados





O recipiente mágico entrega:

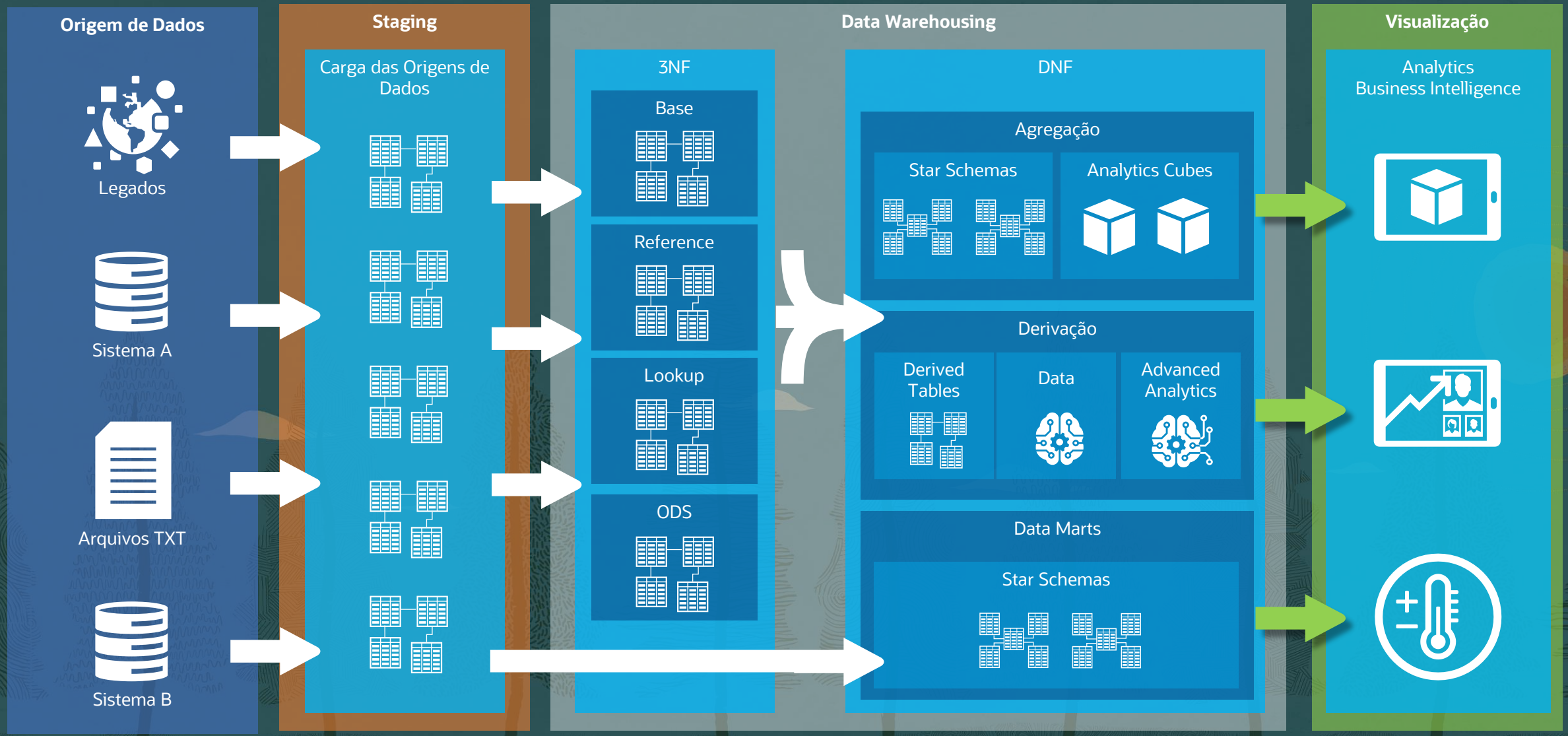
- **Escalabilidade**
- **Desempenho**
- **Transação ACID** — **Atomicidade** **Consistência** **Isolamento** **Durabilidade**
- **Formatos diversos** — **Estruturado** **Semiestruturado** **Não estruturado**
- **Cargas mistas** — **SQL para BI** **Batch de ETL** **Streaming** **AI e ML**
- **Acessibilidade**

Referência: Learning Spark, 2nd Edition. O'Reilly Media. 2020. ISBN 9781492050049.

Foto de James Coleman no Unsplash

Foto de Ming Han Low no Unsplash

Arquitetura Clássica do Data Warehouse

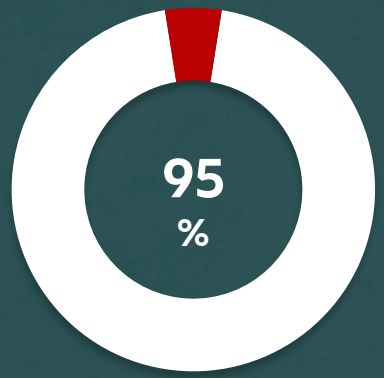


A fama do Data Warehouse

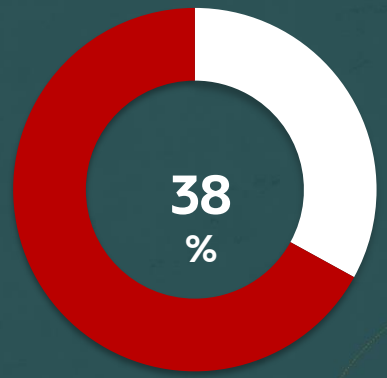
Manual

Complexo e caro

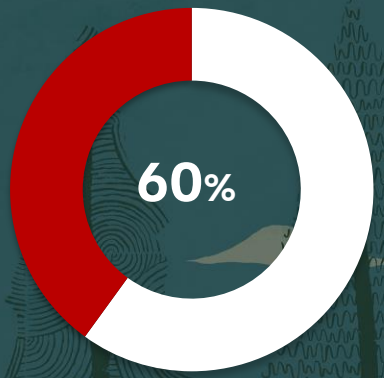
Lento



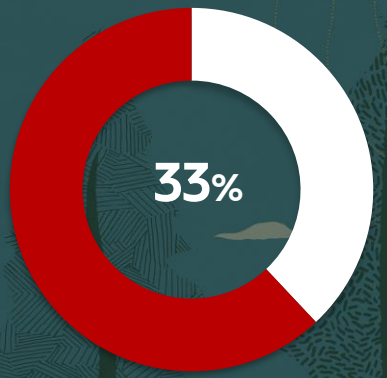
Requer extenso envolvimento manual



Aquisição inicial e custos contínuos com manutenção



Muito complexo de gerenciar

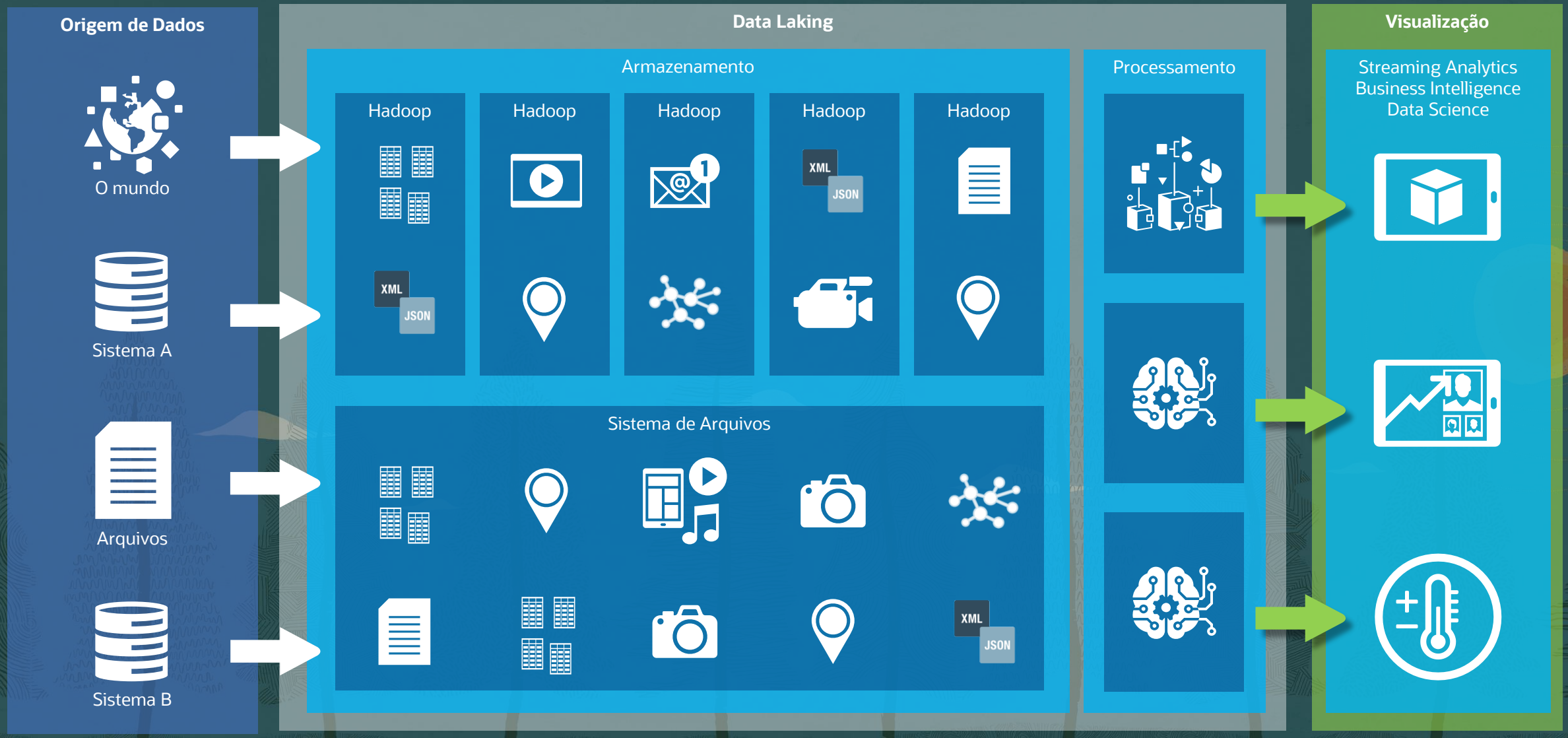


Lento e demorado de implementar

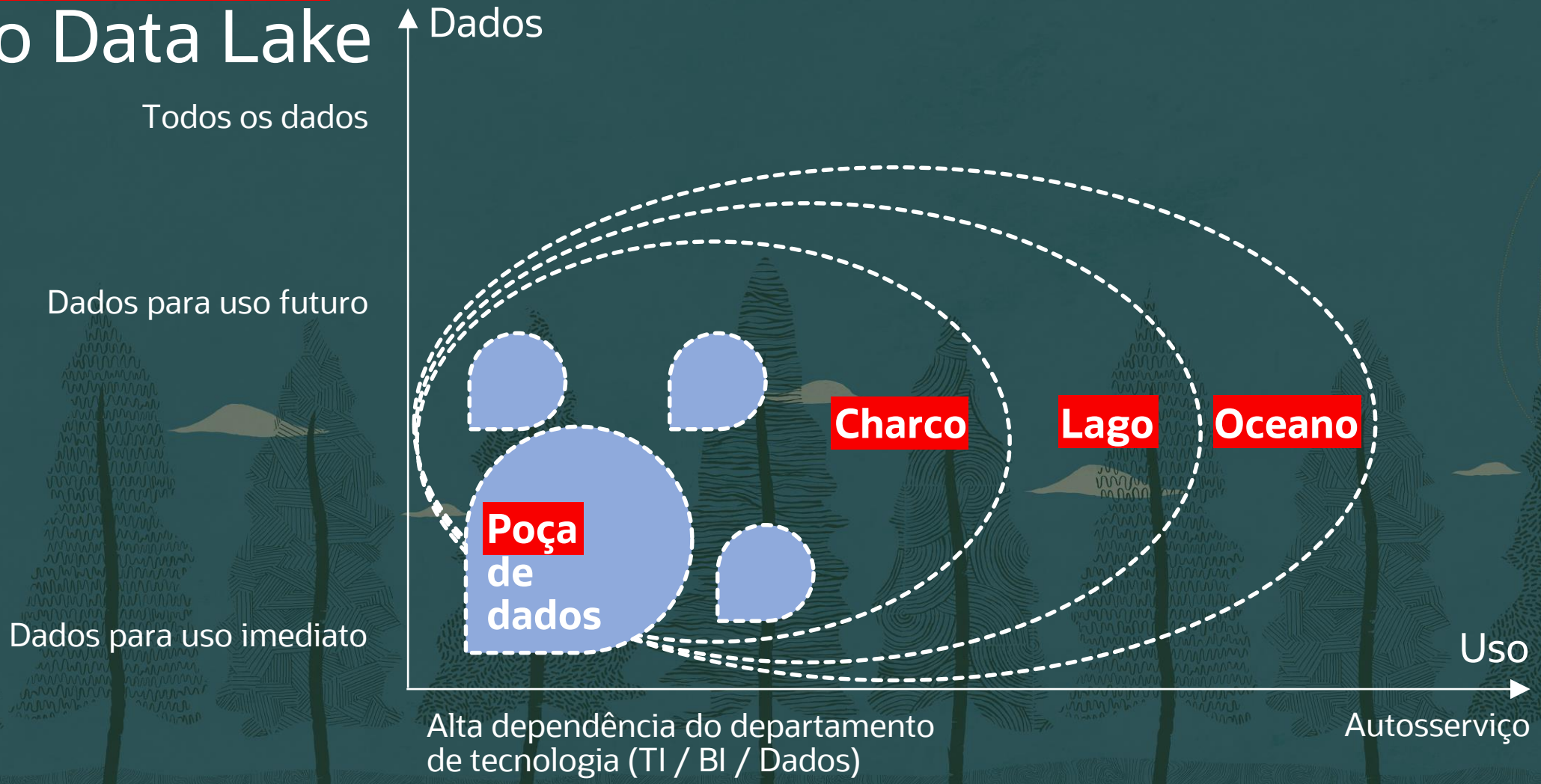
Fonte: Dimensional Research – The State of the Data Warehouse

Foto de Felix M. Dorn no Unsplash

Arquitetura Clássica do Data Lake



Maturidade do Data Lake



Referência: The Enterprise Big Data Lake. O'Reilly Media. 2019. ISBN 9781491931554.

A fama do Data Lake

Falta de atomicidade e isolamento transacional

Inconsistência de dados e qualidade de dados reduzida

Caótico e complexo

Referência: Learning Spark, 2nd Edition. O'Reilly Media. 2020. ISBN 9781492050049.

Foto de Henry Dick no Unsplash

Arquitetura de Solução Cloud Data Warehouse & Lake

ORIGEM DE DADOS

INTEGRAÇÃO DE DADOS

DATA WAREHOUSING

APRESENTAÇÃO

ORACLE SERVICE CLOUD
TERADATA
ORACLE CUSTOMER EXPERIENCE CLOUD
mongoDB
MySQL
Java JDBC
presto
PostgreSQL
ORACLE SALES CLOUD
amazon REDSHIFT
ORACLE Essbase
IBM DB2
salesforce
Microsoft SQL Server
Dropbox
Amazon Aurora
ORACLE DATABASE
ORACLE MARKETING CLOUD
Spark
hadoop

ORACLE DATA INTEGRATOR

ORACLE AUTONOMOUS DATA WAREHOUSE

ORACLE ANALYTICS CLOUD

DATA LAKING

ORACLE OBJECT STORE

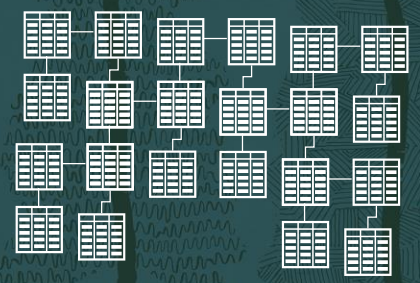
DATA WAREHOUSE MODERNO

DATA LAKE MODERNO

Abordagem para Data Warehouse Moderno

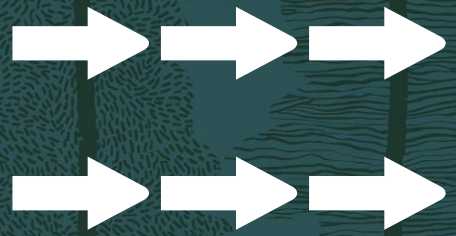
ORIGEM DE DADOS

ORACLE SERVICE CLOUD · TERADATA · ORACLE CUSTOMER EXPERIENCE CLOUD · mongoDB · MySQL · Java JDBC · presto · X · PostgreSQL · ORACLE SALES CLOUD · amazon REDSHIFT · ORACLE Essbase · IBM DB2 · Spark · Microsoft SQL Server · ORACLE DATABASE · hadoop · ORACLE MARKETING CLOUD · Dropbox



INTEGRAÇÃO DE DADOS

ORACLE DATA INTEGRATOR



DATAWAREHOUSING

ORACLE AUTONOMOUS DATA WAREHOUSE

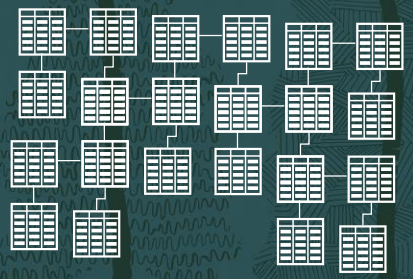
APRESENTAÇÃO

ORACLE ANALYTICS CLOUD

Abordagem para Data Lake Moderno

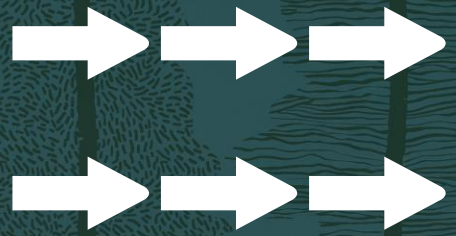
ORIGEM DE DADOS

ORACLE SERVICE CLOUD
TERADATA
ORACLE CUSTOMER EXPERIENCE CLOUD
mongoDB
MySQL
Java JDBC
presto
PostgreSQL
ORACLE SALES CLOUD
amazon REDSHIFT
ORACLE Essbase
IBM DB2
Spark
Microsoft SQL Server
ORACLE DATABASE
hadoop
ORACLE MARKETING CLOUD
Dropbox



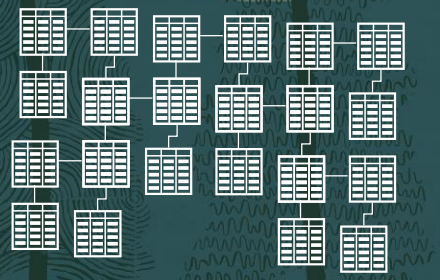
INTEGRAÇÃO DE DADOS

ORACLE DATA INTEGRATOR



DATA LAKING

ORACLE AUTONOMOUS DATA WAREHOUSE



APRESENTAÇÃO

ORACLE ANALYTICS CLOUD

E se... eu quiser um Data Lake “mesmo”?

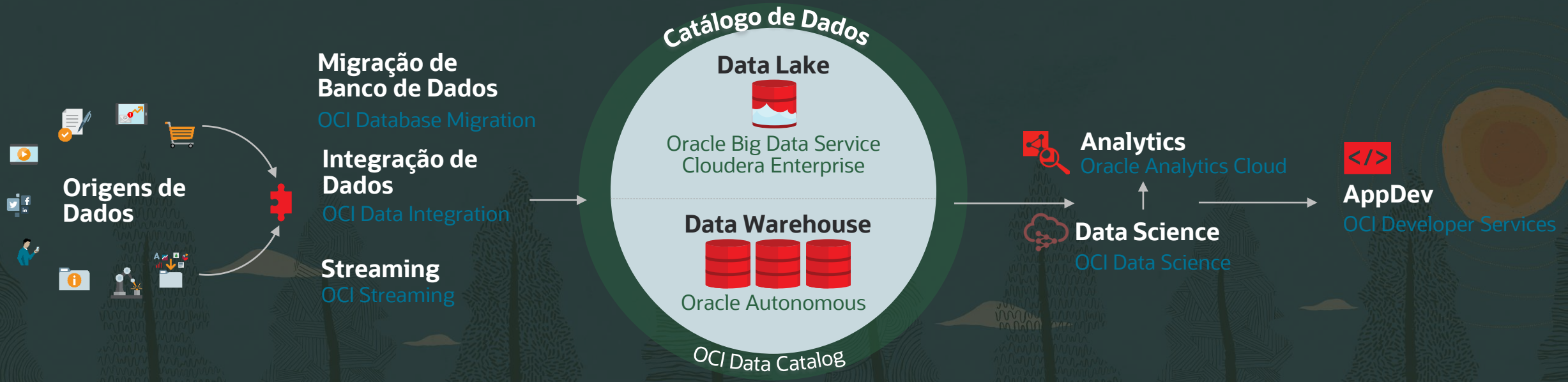


Foto de Evelyn Paris no Unsplash

Lakehouse: um novo paradigma que combina elementos de Data Lake e Data Warehouse.

TRANSAÇÃO ACID

CONFORMIDADE DE ESQUEMA

FORMATOS DIVERSOS E ABERTOS

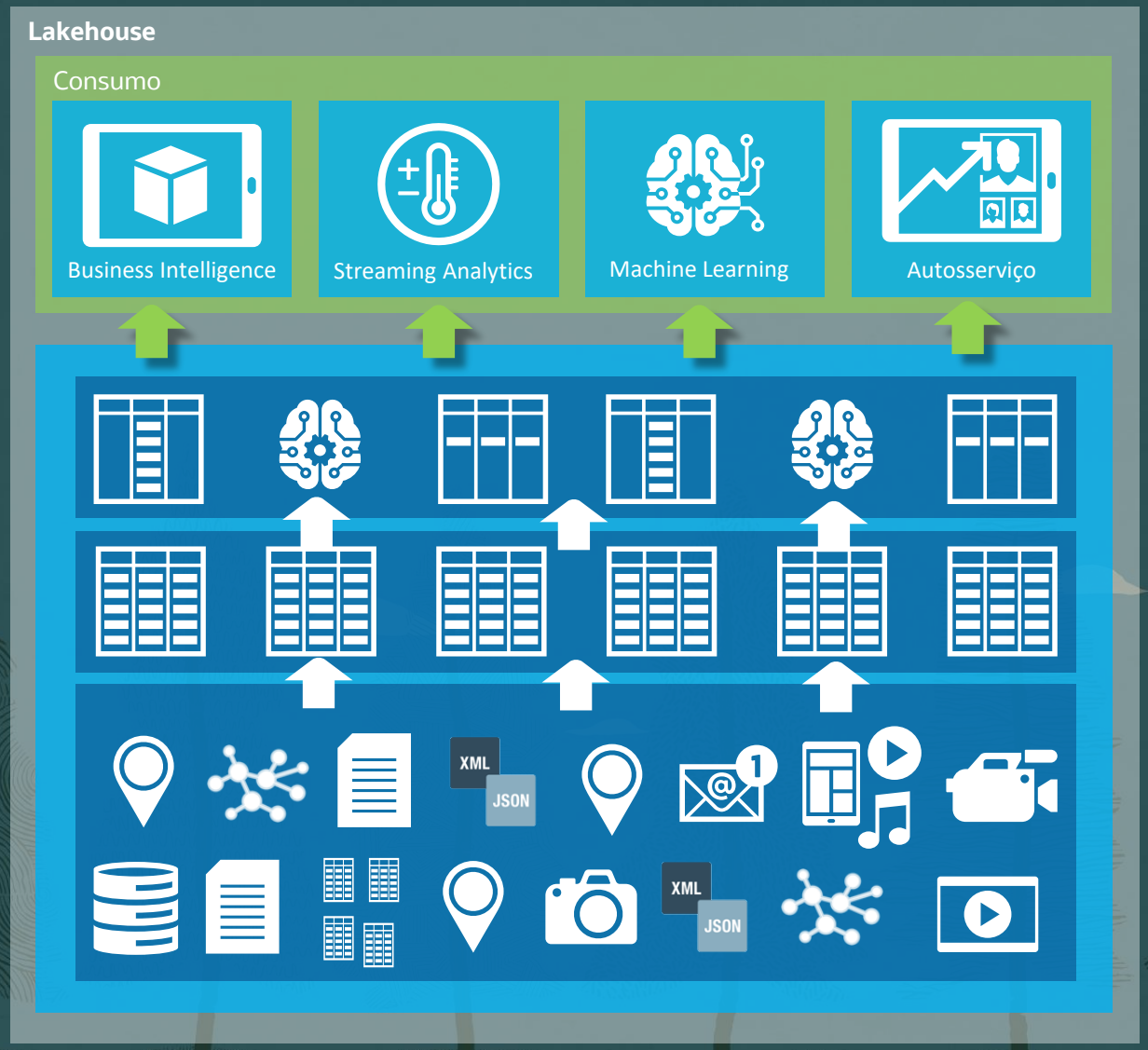
CARGAS MISTAS

UPSERT & DELETES PARALELOS

GOVERNANÇA DE DADOS

Referência: Learning Spark, 2nd Edition. O'Reilly Media. 2020. ISBN 9781492050049.

Arquitetura de Dados Lakehouse



Plataforma única para consumo

Motor de alto desempenho para consultas

Camada transacional estruturada

Data Lake para todos os dados

Referências: https://www.youtube.com/watch?v=yumysN3XwbQ
https://databricks.com/wp-content/uploads/2020/08/p975-armbrust.pdf

Projetos Lakehouse

APACHE HUDI

Hadoop Update Delete and Incremental

Focado em upserts e deletes em Chave-Valor

Combina formatos colunares e lineares

APACHE ICEBERG

Focado em propósito geral de armazenamento em tabelas únicas de grande tamanho

Evolução de esquema e particionamento, versionamento e isolamento serializado

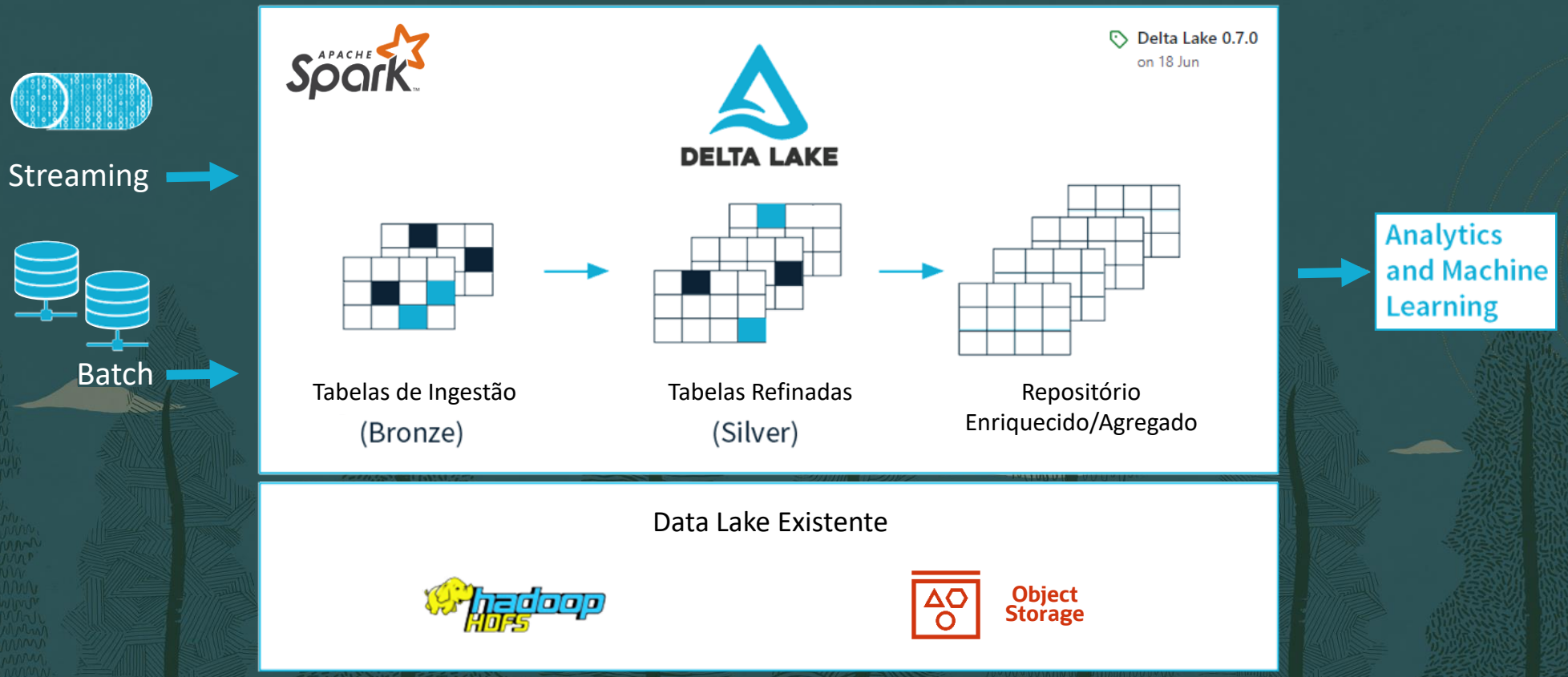
DELTA LAKE

Mantido pela Linux Foundation e construído pelos criadores do Apache Spark

Formato aberto de armazenamento com suporte transacional e evolução de esquema

Referência: Learning Spark, 2nd Edition. O'Reilly Media. 2020. ISBN 9781492050049.

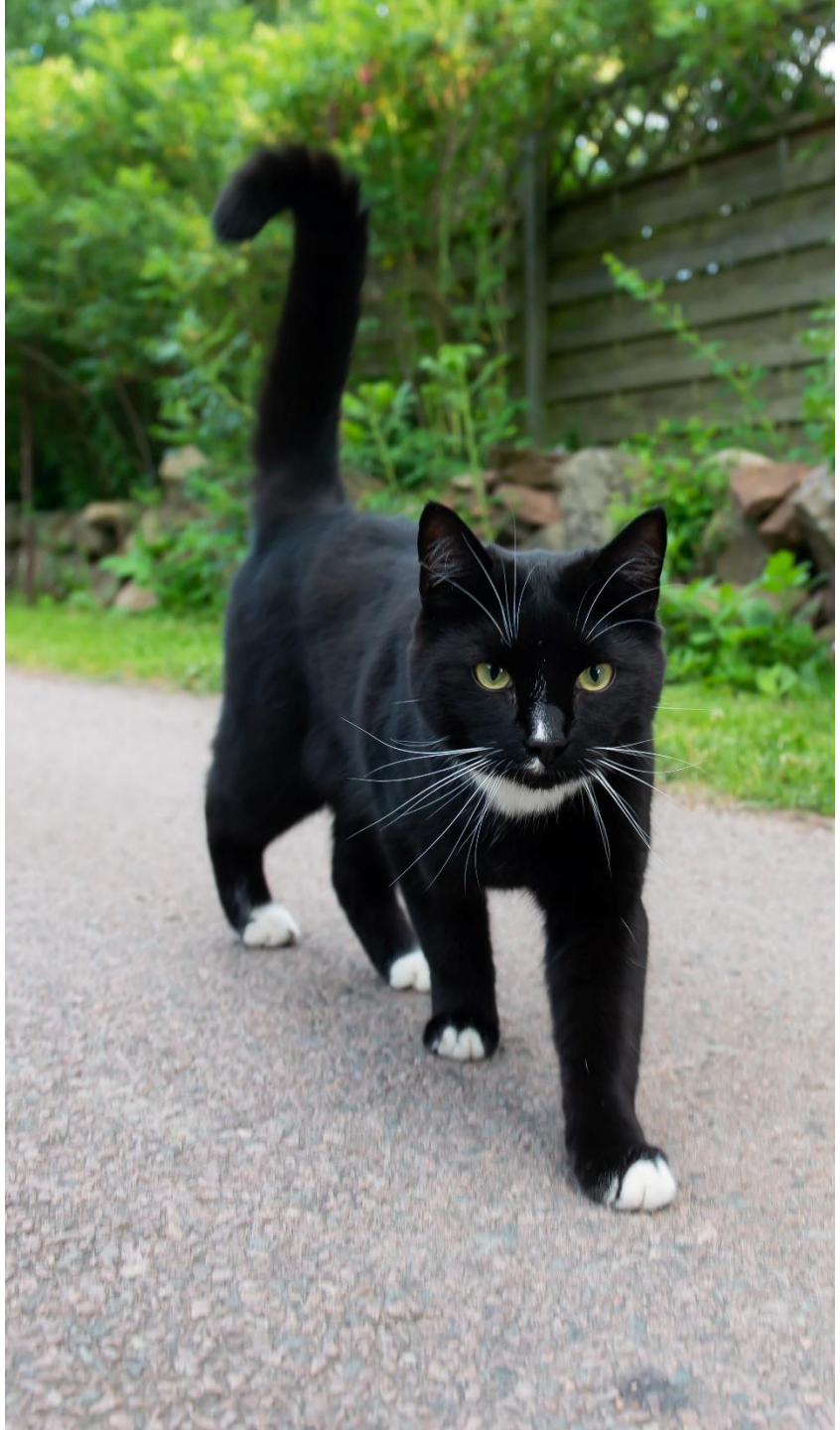
Arquitetura de Referência Delta Lake



Referência: https://delta.io/

Foto de Davidson Luna no Unsplash

Foto de Anders Nord no Unsplash



Componentes do Data Warehouse moderno

INTEGRAÇÃO
Streaming, batch data, on-premises e cloud

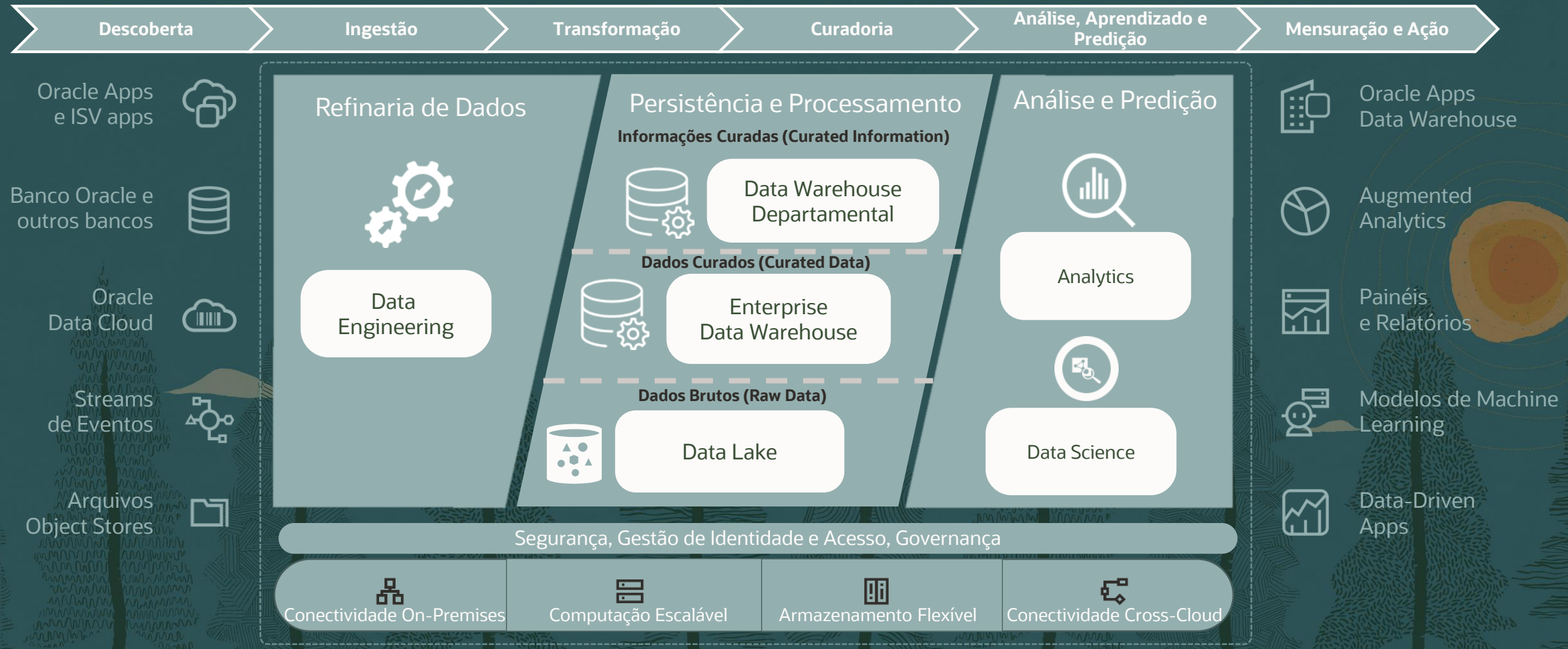
DATA WAREHOUSE
Autonomous, self-driving, self-securing, self-repairing

DATA LAKE
Baseado em Object storage e integrado com o Data Warehouse

ANALYTICS
Visualização e inteligência analítica baseadas em Machine Learning

DATA SCIENCE
Machine learning de propósito geral e in-database

Arquitetura de Solução do Data Warehouse moderno



Gravidade de dados...

Qual é a sua?

Oracle: Base de Dados Convergente

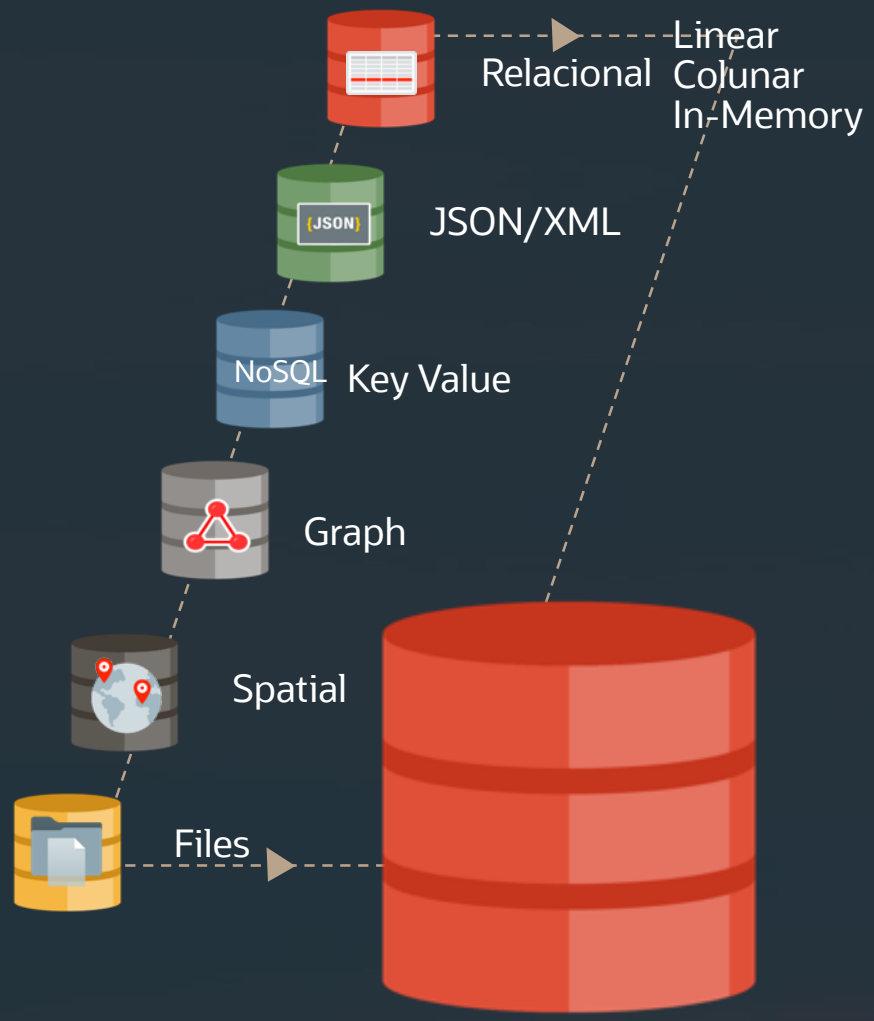


Foto de Davidson Luna no Unsplash

Resumindo...

Conheça a gravidade de dados do seu negócio e use todos os seus ativos de dados com as ferramentas maduras que você domina.

Data Warehouse e Data Lake são paradigmas complementares, relevantes ao seu negócio e permanecem com alta demanda pelo mercado.

O paradigma Lakehouse é o conjunto de práticas de sucesso conhecidas no manejo de dados e sua implementação pode ser aderente a diferentes tecnologias.

Data Warehouse e Modelagem Dimensional de Dados em Tempos Modernos de Nuvem

TDC ONLINE

Sala ORACLE

Inovação com dados em nuvem

TRILHA

#TheDevConf Oracle

09.11.20
19h30