



Foto de Evelyn Paris no Unsplash

Uma Breve História da Arquitetura de Dados Lakehouse

Lourenço Taborda

Trilha Inovação com dados em nuvem
29.10.2020 13h30

Uma Breve História da Arquitetura de Dados Lakehouse

Lourenço Taborda

Trilha Inovação com dados em nuvem
29.10.2020 13h30



Este trabalho está licenciado sob uma Licença Creative Commons Atribuição-Compartilhamento 4.0 Internacional. Para ver uma cópia desta licença, visite <http://creativecommons.org/licenses/by-sa/4.0/>.



LOURENÇO TABORDA



Certified Business
Intelligence Professional



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.

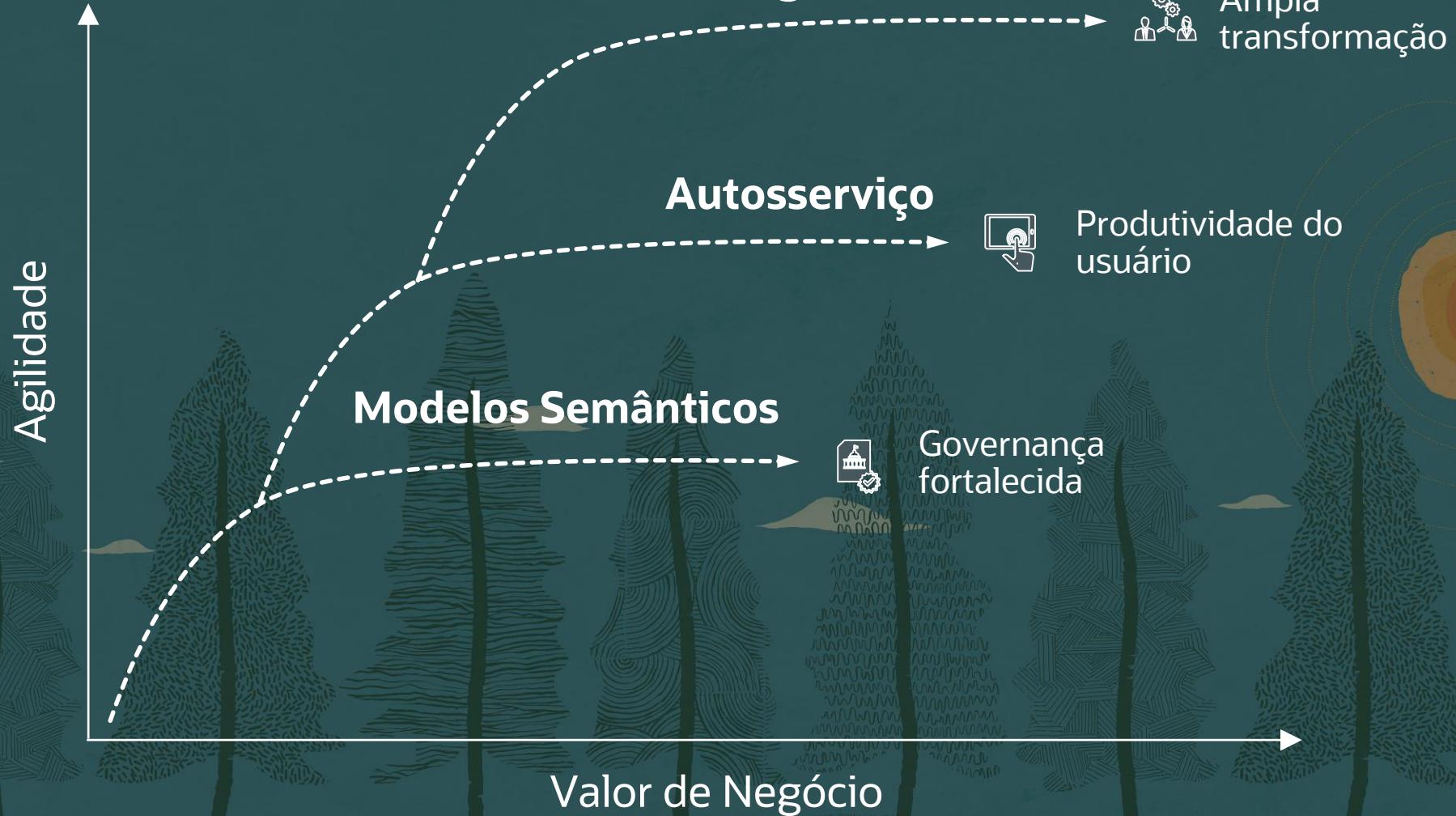


DADOS SÃO FATORES
DE PRODUÇÃO DE
BENS E SERVIÇOS
DIGITAIS.



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

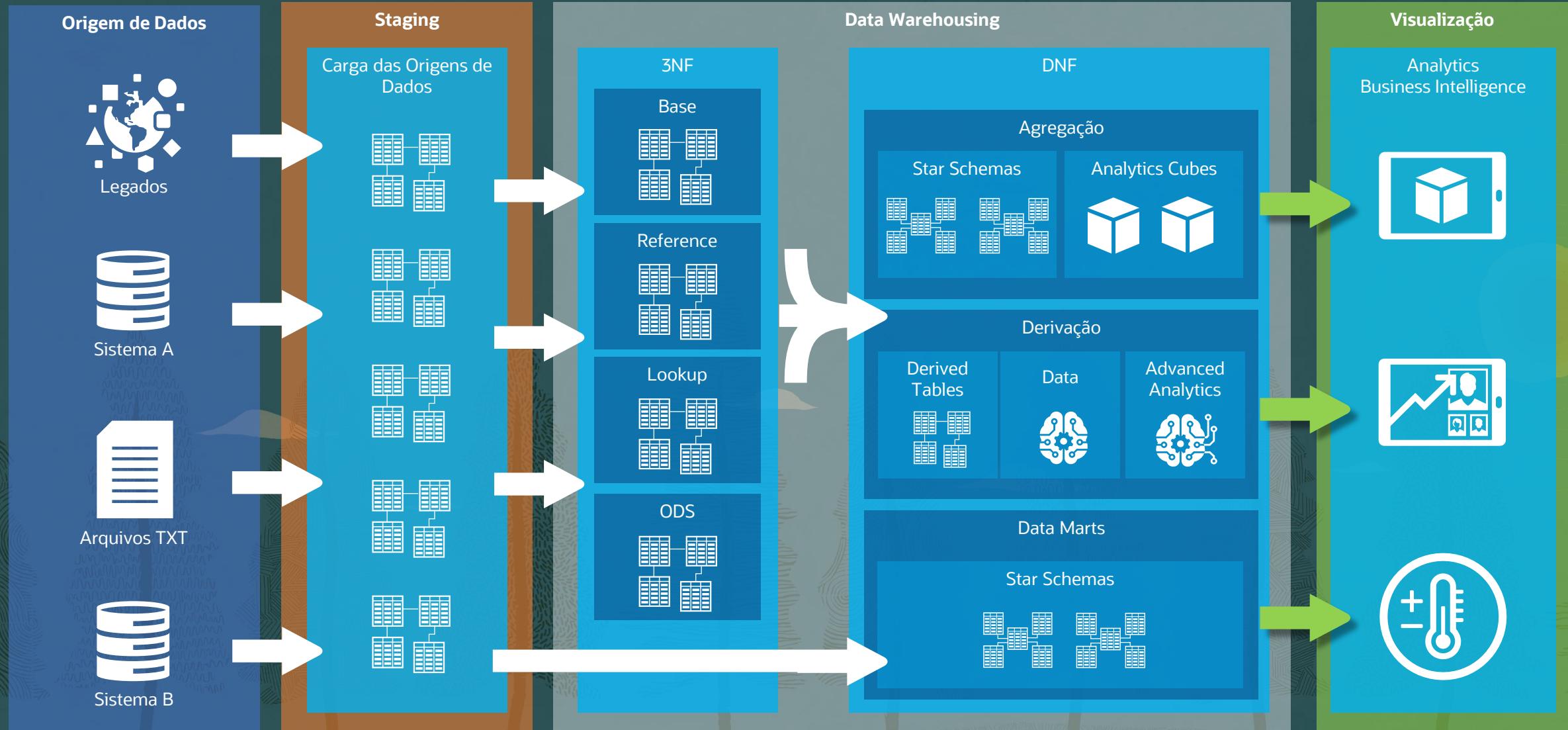
Cargas mistas | SQL para BI | Batch de ETL | Streaming | AI e ML

Acessibilidade

Referência: Learning Spark, 2nd Edition. 2020.



Arquitetura Clássica do Data Warehouse



A fama do Data Warehouse



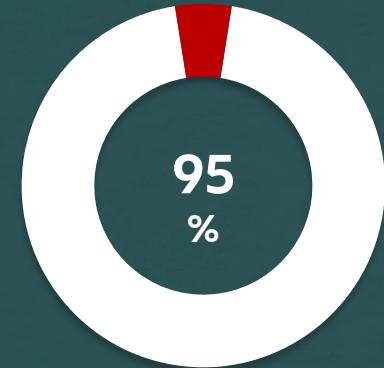
Manual



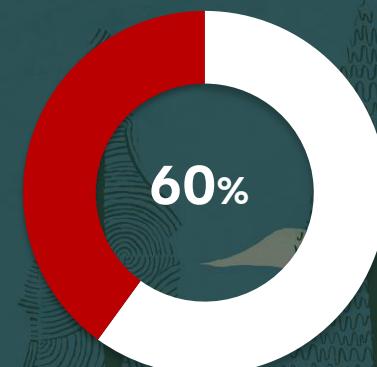
Complexo e caro



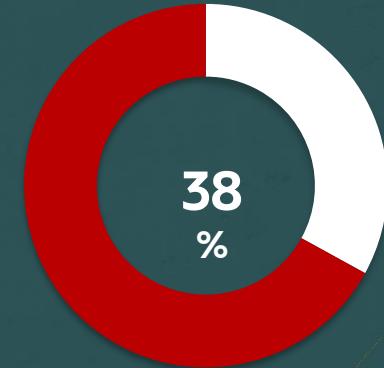
Lento



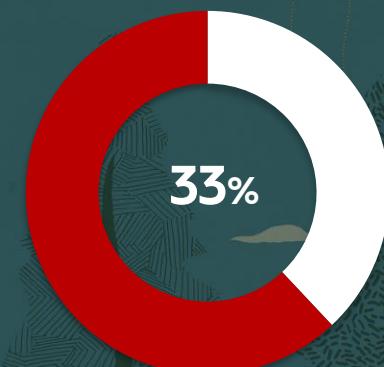
Requer extenso envolvimento manual



Muito complexo de gerenciar



Aquisição inicial e custos contínuos com manutenção



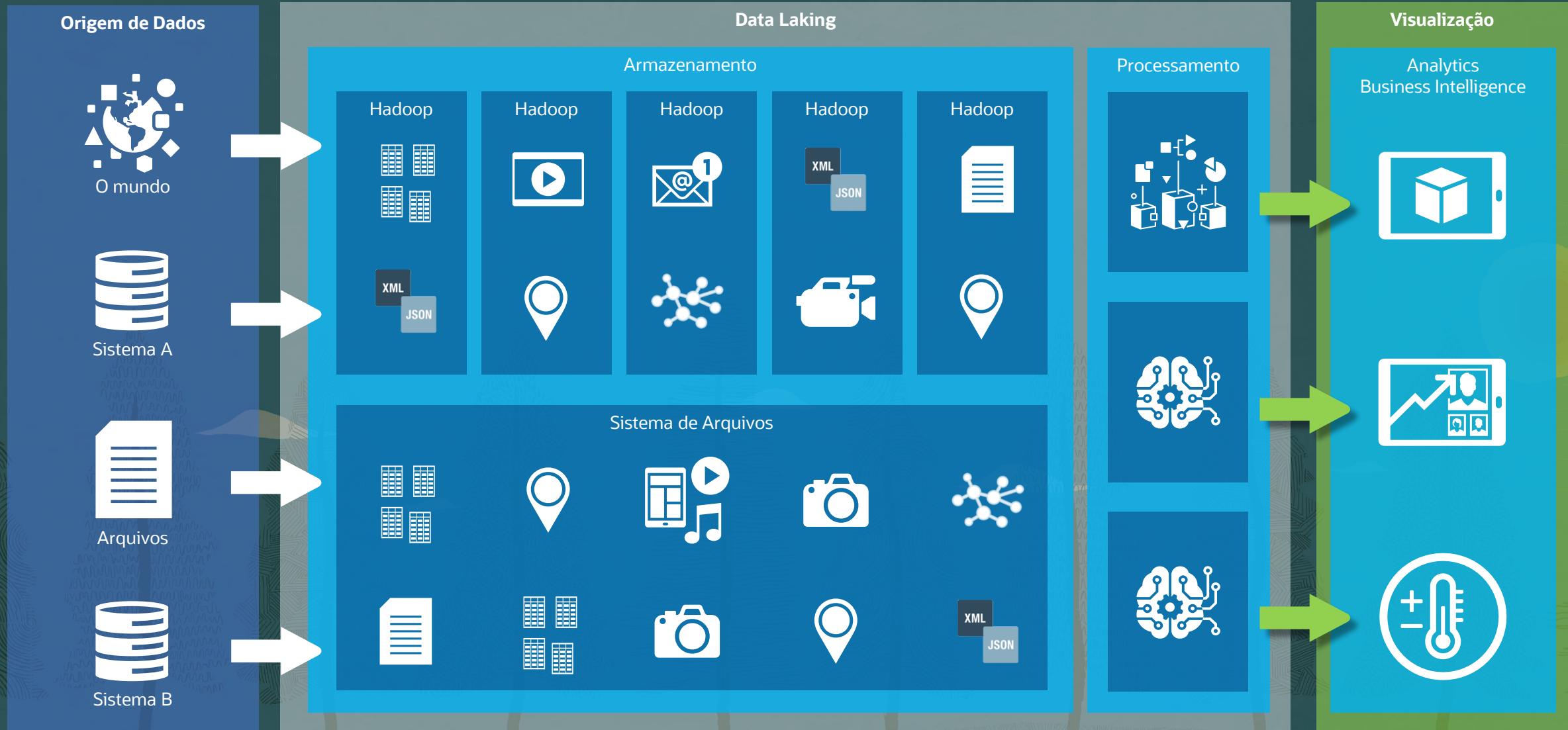
Lento e demorado de implementar

Fonte: Dimensional Research – The State of the Data Warehouse



Foto de Felix M. Dorn no Unsplash

Arquitetura Clássica do Data Lake



A fama do Data Lake



Falta de atomicidade e isolamento transacional



Inconsistência de dados e qualidade de dados reduzida

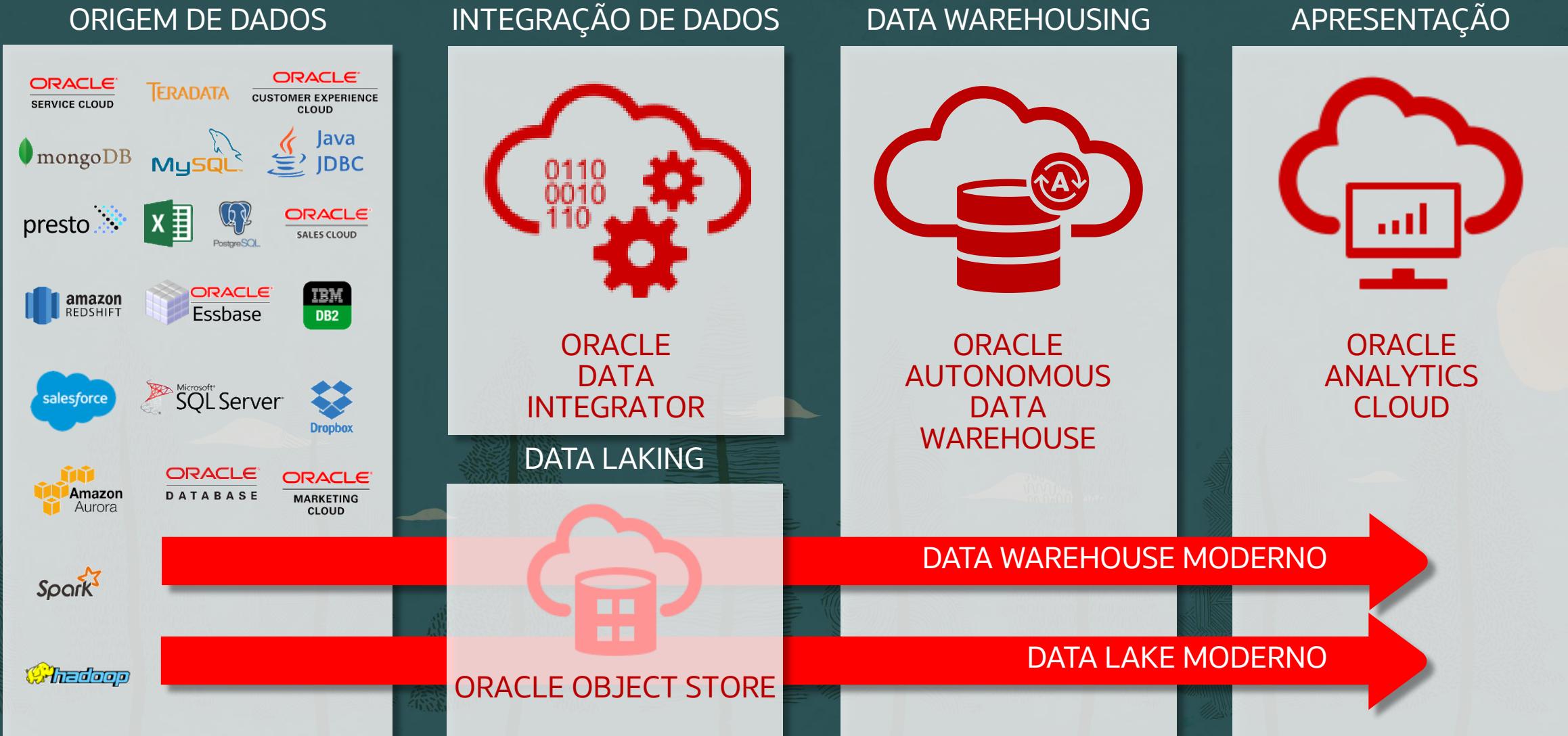


Caótico e complexo

Referência: Learning Spark, 2nd Edition, 2020.



Arquitetura de Solução Cloud Data Warehouse & Lake



Abordagem para Data Warehouse Moderno

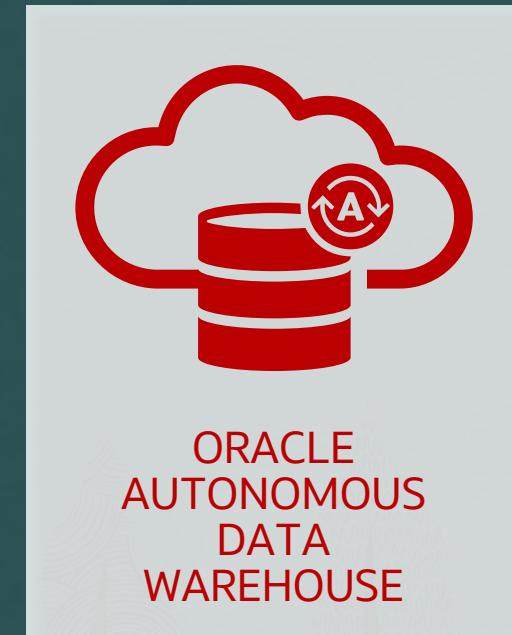
ORIGEM DE DADOS



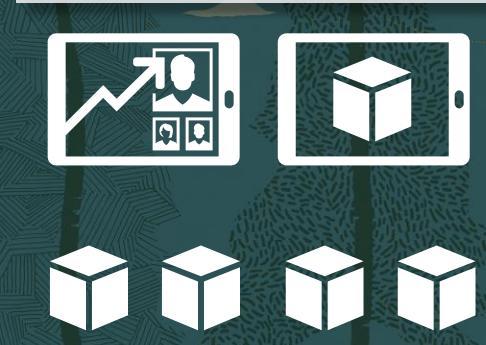
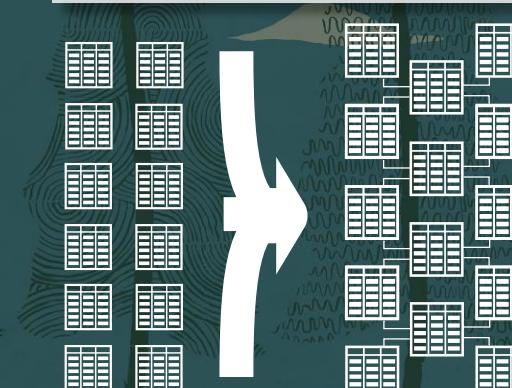
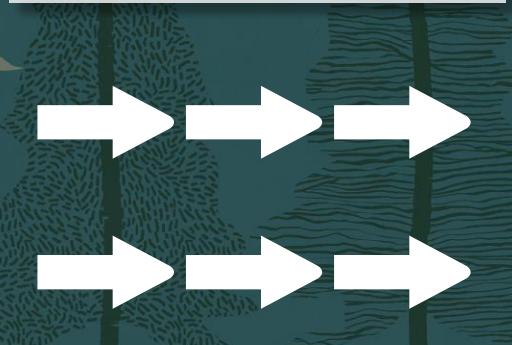
INTEGRAÇÃO DE DADOS



DATAWAREHOUSING



APRESENTAÇÃO



Abordagem para Data Lake Moderno

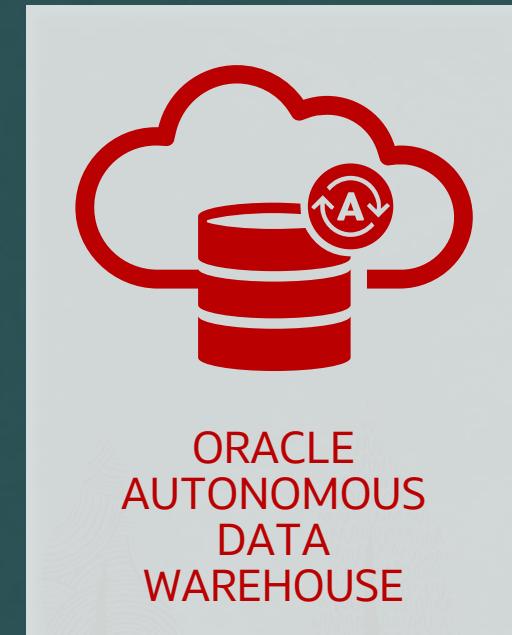
ORIGEM DE DADOS



INTEGRAÇÃO DE DADOS



DATA LAKING



APRESENTAÇÃO



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

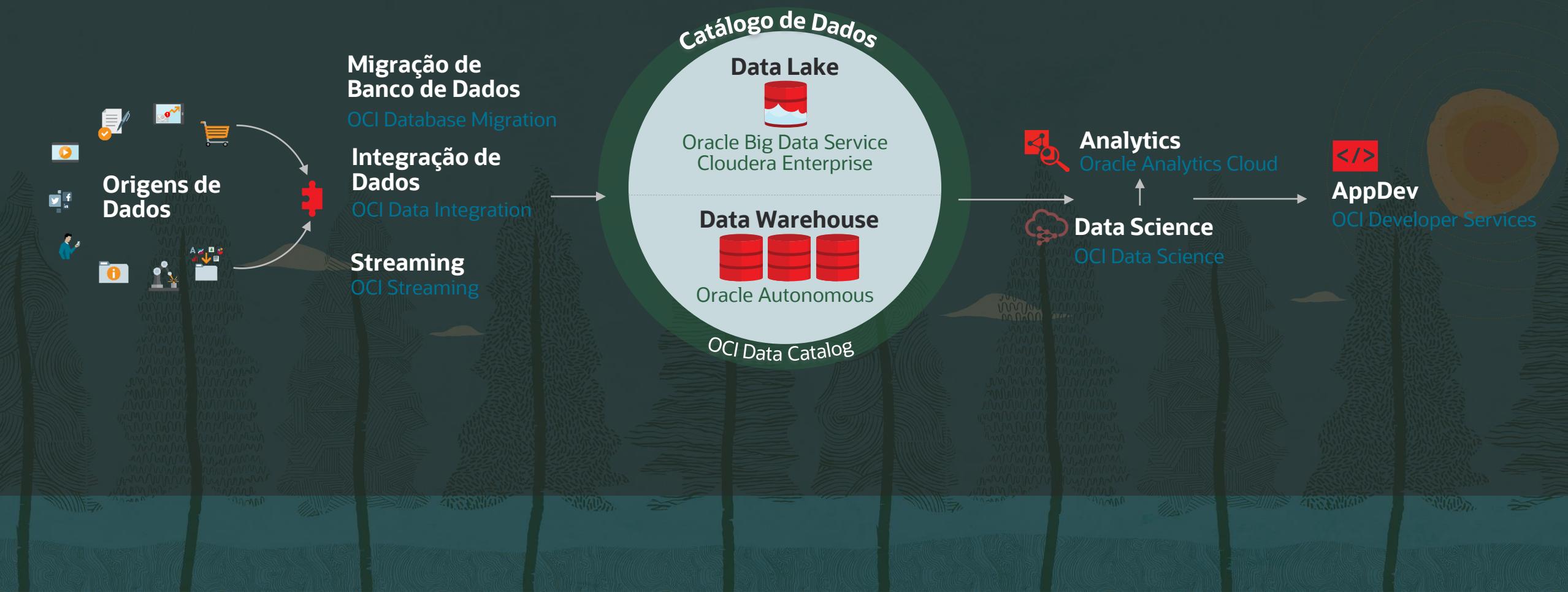




Foto de Evelyn Paris no Unsplash

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



TRANSAÇÃO
ACID



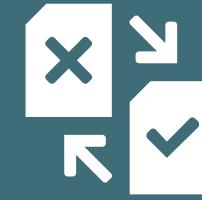
CONFORMIDADE
DE ESQUEMA



FORMATOS
DIVERSOS E
ABERTOS



CARGAS
MISTAS

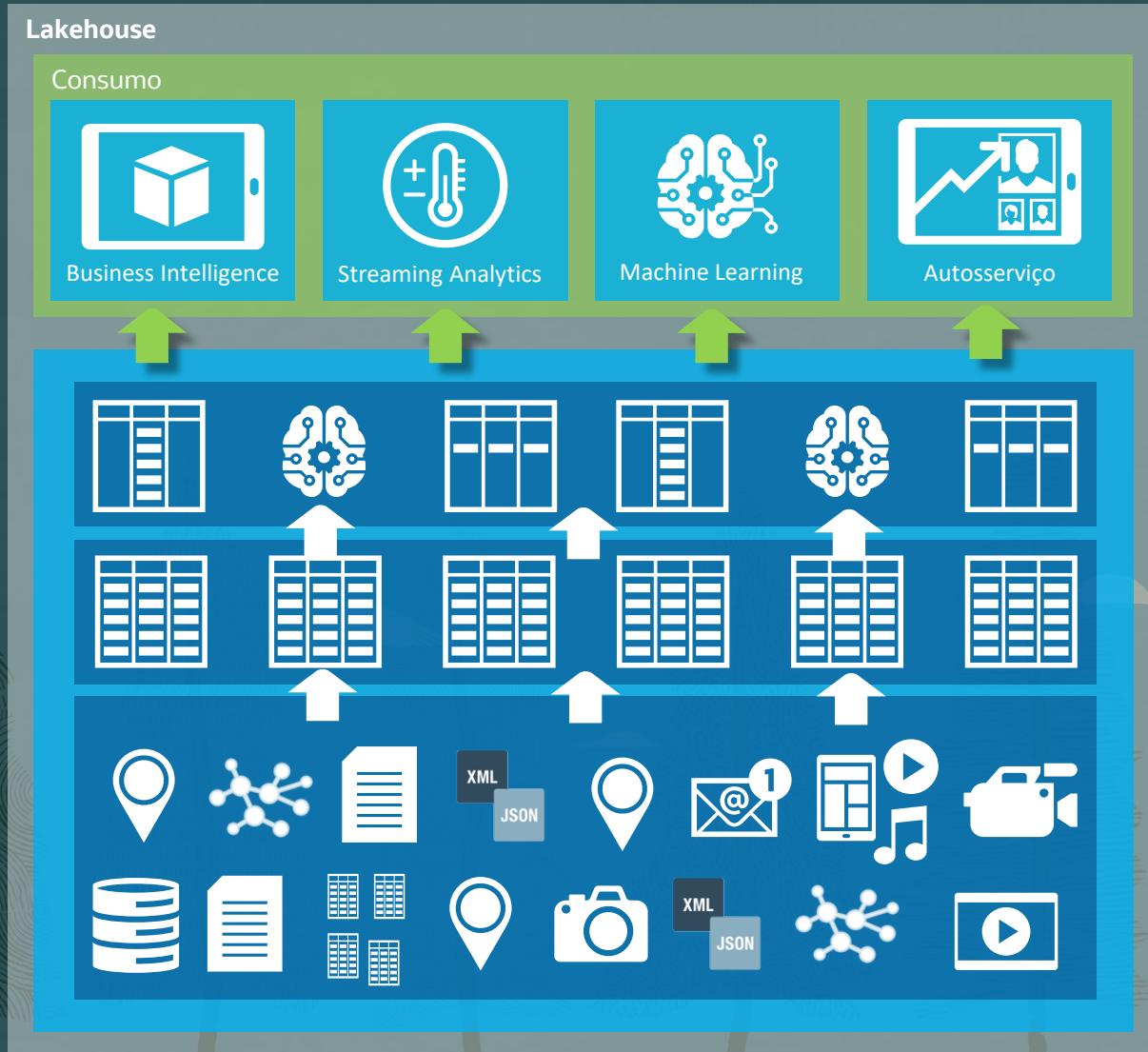


UPSERT &
DELETE
PARALELOS



GOVERNANÇA
DE DADOS

Arquitetura de Dados Lakehouse



Plataforma única para consumo

Motor de alto desempenho para consultas

Camada transacional estruturada

Data Lake para todos os dados

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

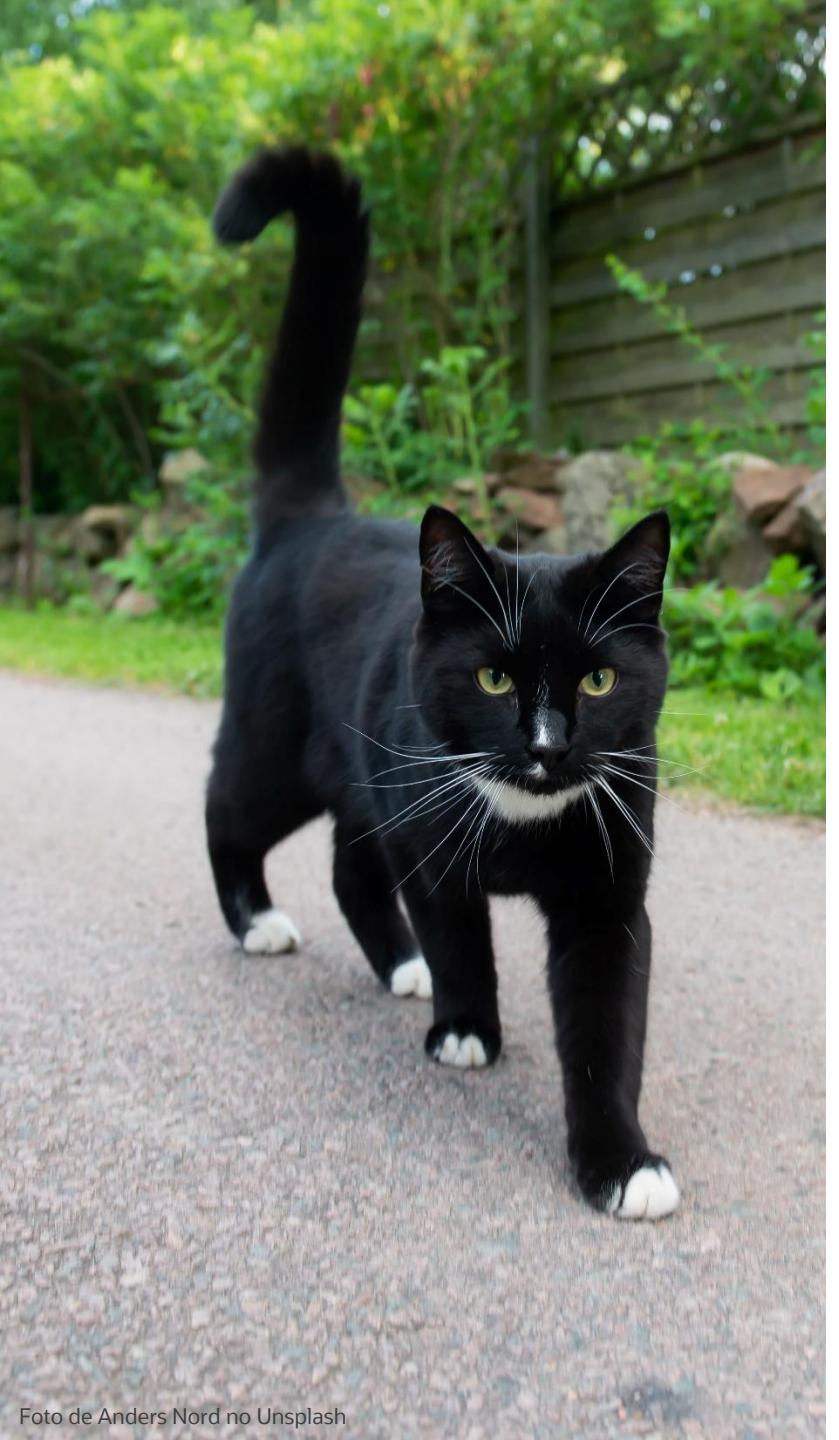
Evolução de esquema e particionamento, versionamento e isolamento serializado

DELTA LAKE

Mantido pela Linux Foundation e construído pelos criadores do Apache Spark

Formato aberto de armazenamento com suporte transacional e evolução de esquema





Gravidade
de dados...

Qual é a sua?



Ok, entendi a gravidade de dados do meu negócio!

Componentes do Data Warehouse moderno



INTEGRAÇÃO

Streaming,
batch data, on-premises e cloud



DATA WAREHOUSE

Autonomous,
self-driving,
self-securig,
self-repairing



DATA LAKE

Baseado em
Object storage e
integrado com o
Data Warehouse



ANALYTICS

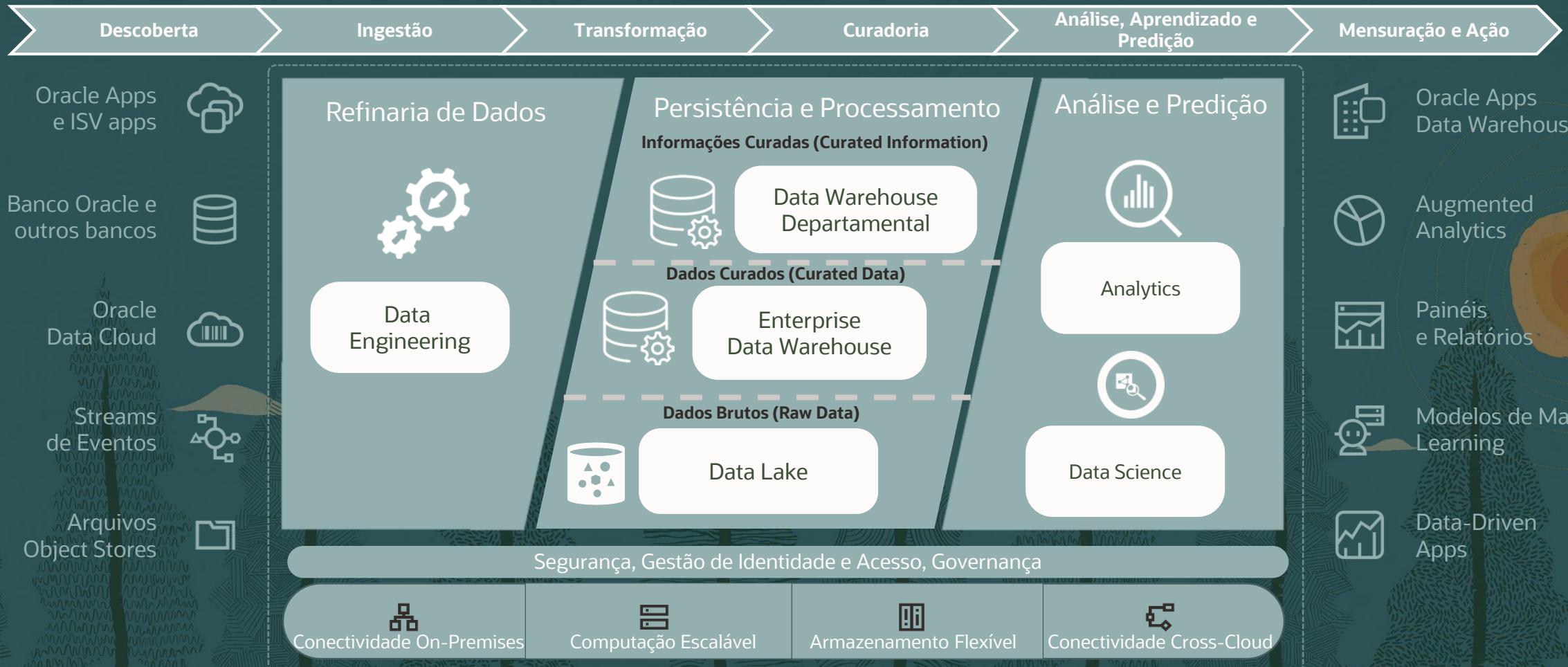
Visualização e
inteligênciia
analíticai baseadas
em Machine
Learning



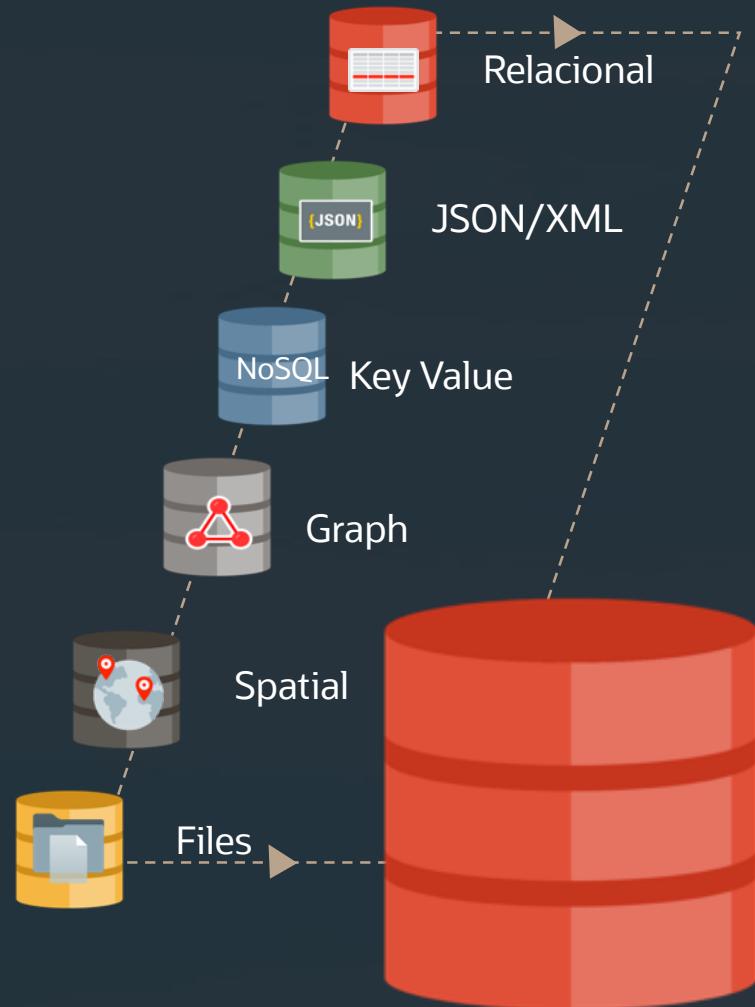
DATA SCIENCE

Machine learning
de propósito geral
e in-database

Arquitetura de Solução do Data Warehouse moderno



Oracle: Base de Dados Convergente



Enfim...

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.

Oracle Cloud Modo Gratuito

Uso Livre

Serviços que você pode usar por tempo ilimitado



Avaliação Gratuita de 30 dias

US\$ 500 em créditos gratuitos



Inovação com dados em nuvem

UMA BREVE HISTÓRIA DA
ARQUITETURA DE DADOS
LAKEHOUSE
29.10.20

