



# O Data Warehouse Moderno

Américo Silva Junior  
Abril/14





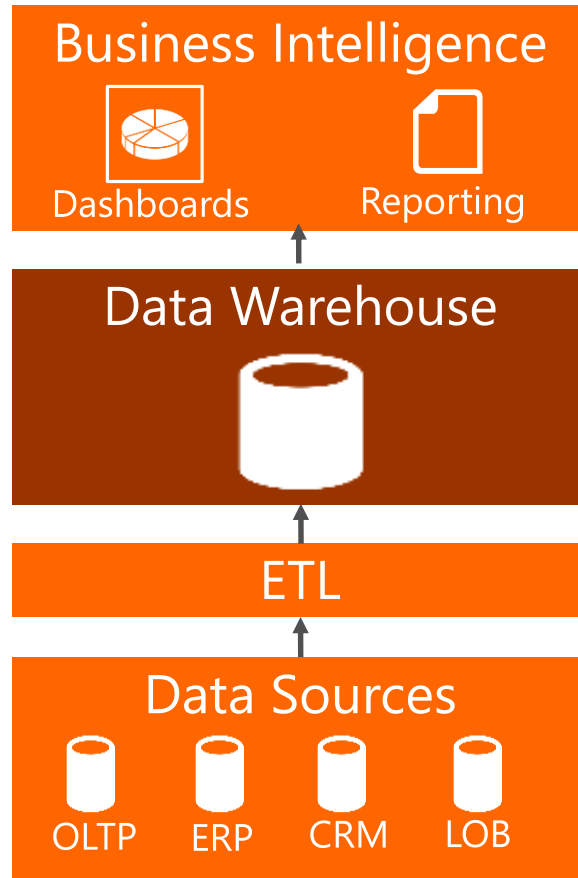
Unlocking  
Insights on  
Any Data

Breakthrough  
Data Platform  
Performance with  
SQL Server 2014

Enabling Familiar,  
Powerful Business  
Intelligence

O Data Warehouse  
Moderno

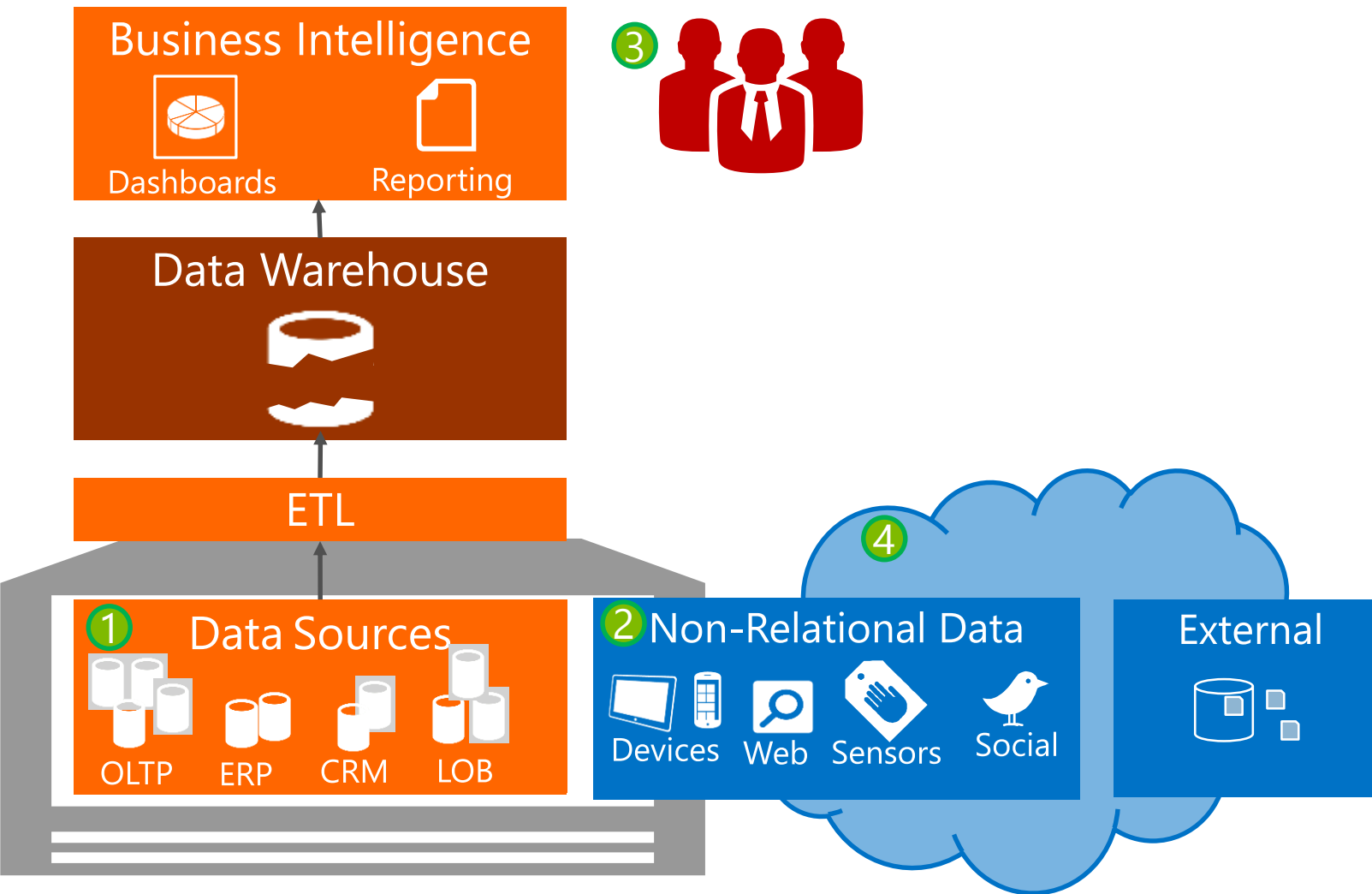
# ○ Data Warehouse Tradicional



“ ... data warehousing has reached the most significant tipping point since its inception. The biggest, possibly most elaborate data management system in IT is changing. ”

– Gartner, “The State of Data Warehousing in 2012”

# Fatores Críticos em um Data Warehouse Tradicional



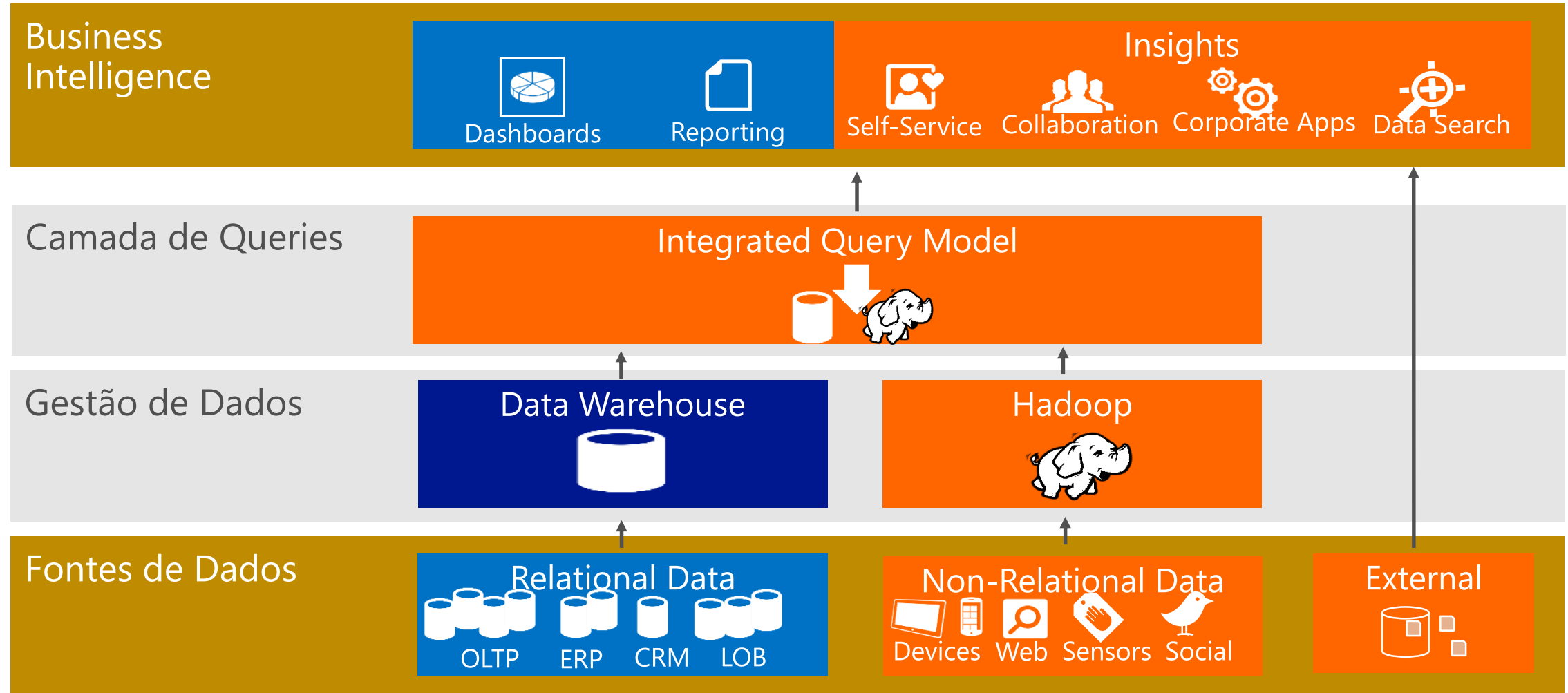
1. Maiores Volumes de Dados

2. Novos Tipos de Dados

3. Expectativas dos Usuários

4. Dados Externos

# Um Data Warehouse Moderno tem tudo isso...



# “The Modern Data Warehouse”

What Enterprises Must Have Today and What They'll Need in the Future\*

Deploy a modern data warehouse as a foundation for leveraging big data

Make sense of multi-structured data for new and unique business insights

Implement advanced forms of analytics to enable discovery analytics for big data

Empower the business to operate in near real time by delivering data faster

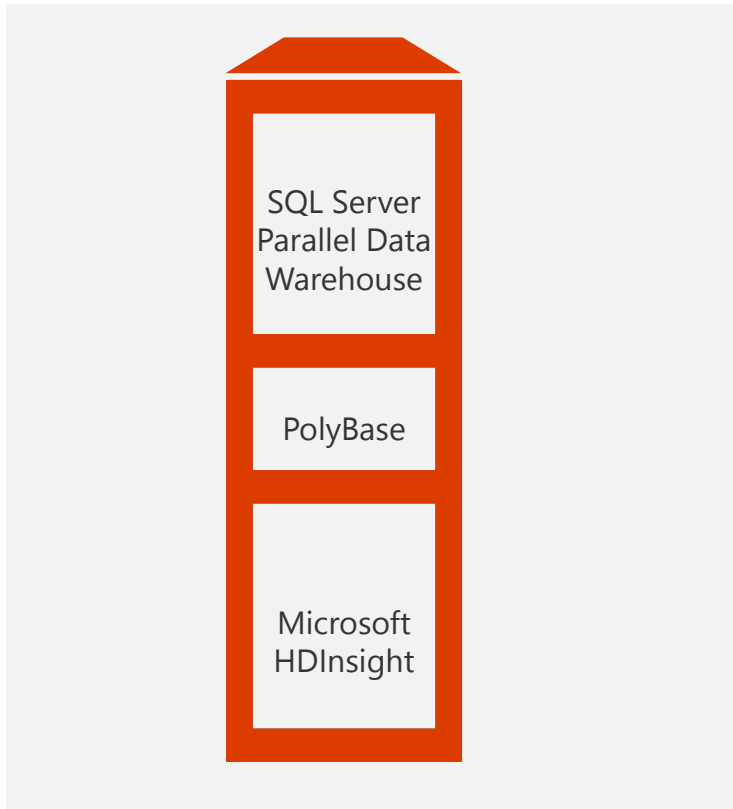
Integrate multiple platforms into a unified data warehouse architecture

Demand high performance and scalability of all components of a data warehouse

\* The Data Warehousing Institute (TDWI), is the premier provider of in-depth, high-quality education and research in the business intelligence and data warehousing industry.

# Analytics Platform System

## Appliance, On-premise; "Big Data in a Box"



- Appliance pré-montado com software instalado na fábrica
- Processamento Paralelo de Dados (MPP - Massively Parallel Processing) com escalabilidade para até 6 Pb de dados
- +100x de melhora em performance com uso de In-memory columnstore
- Região dedicada para Hadoop
- Consultas a dados relacionais e dados Hadoop em query única, usando linguagem SQL
- Disponível pela HP e Dell



Qualquer  
Volume



Performance  
Tempo Real



Qualquer Dado





Qualquer  
Volume



Performance  
Tempo Real



Qualquer Dado







Usando o Parallel Data Warehouse, a Hy-Vee conseguiu usar **7 anos de dados** aos invés de apenas **2 anos de dados**, além de obter ganhos de mais de **100x faster** nas consultas em relação a seu data warehouse anterior.

“Agora conseguimos entender porque e quando nossos clientes compram, e com isso tomamos melhores decisões sobre os precificação e publicidade.”

**TOM SETTLE**  
ASSISTANT VICE PRESIDENT  
HY-VEE

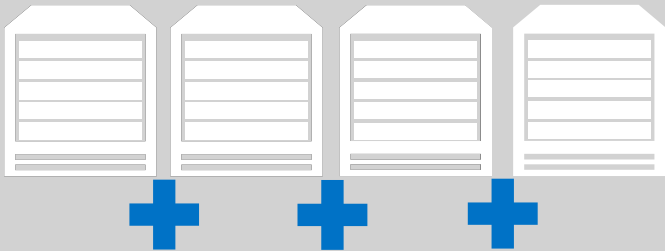
**HyVee**®



# Escalabilidade em dados relacionais

## Tecnologias no Analytics Platform System

### Scale OUT



- Processamento Massively Parallel Processing (MPP) paraleliza as queries
- Múltiplos nós com CPU, memória e storage dedicados
- Incremento linear de HW para escalar a múltiplos Pb
- Lida com complexidade de queries e concorrência de usuários
- Preserva investimento inicial ao incrementar capacidade

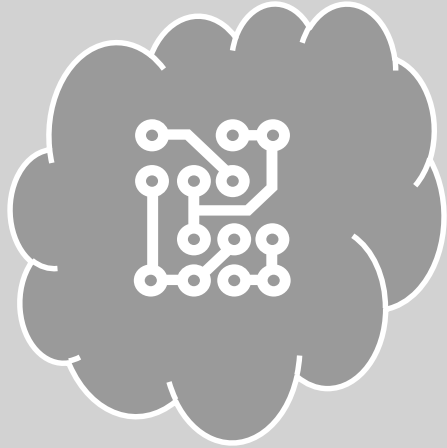
De Terabytes para Multi-Petabytes



# Escalabilidade em dados não-relacionais

HDInsight (em Azure ou APS)

Scale OUT



- Crie clusters Hadoop para suas necessidades
- Adicione mais poder de processamento de acordo com a demanda
- Cresça linearmente
- Desligue os clusters quando queira

Scale Out "Big Data"

Qualquer  
Volume



Performance  
Tempo Real



Qualquer Dado







O Bank of Nagoya ganhou **600x de performance em queries** ao usar o SQL Server, permitindo às agências ter dados de seus clientes instantaneamente.

“Ao usar o “in-memory columnstore”, conseguimos extrair dados em 2 ou 3 segundos versus os 30 minutos necessários previamente.”

**ATSUO NAKAJIMA**

ASST. DIRECTOR

SYSTEMS DEVELOPMENT GROUP



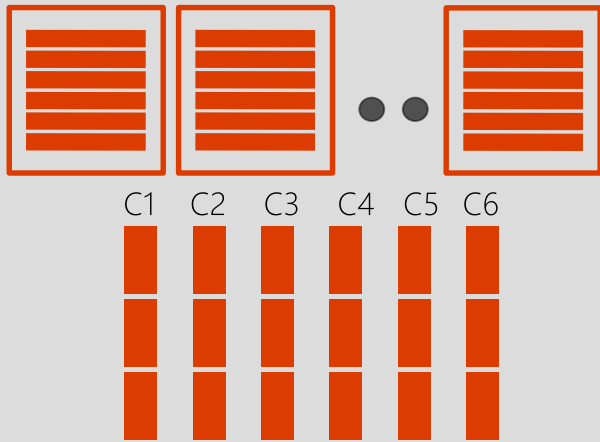




# Performance In-memory

In-memory columnstore para alta performance

Columnstore index representation



Armazene dados em format columnar para compressão massiva

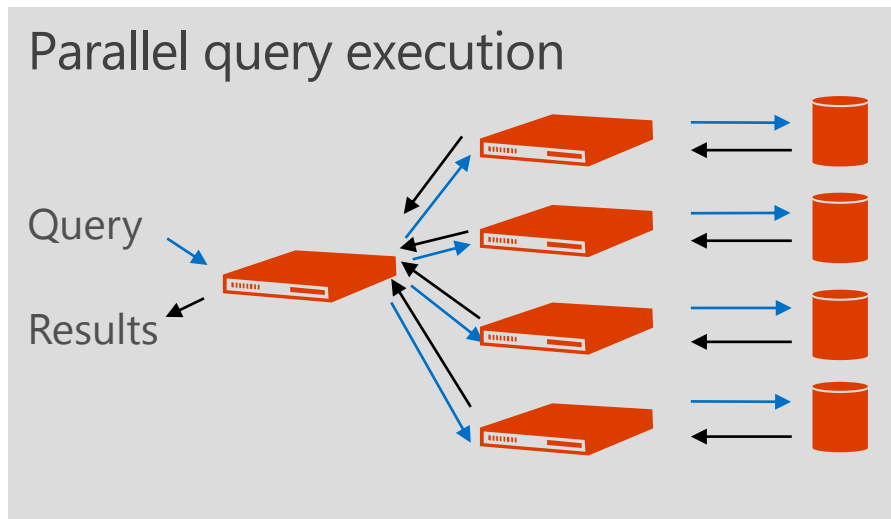
Carregue dados de/para memory para alta performance

Carga "real time" possível e em clusters

100x mais rápido e compressão 10x

# Concorrência e cargas de trabalho mistas

## Grande performance para workloads variados



- Massively Parallel Processing (MPP) paraleliza consultas
- Lida com complexidade de consultas e concorrência
- Geração de insights juntamente com cargas de dados
- Vários tipos de workloads ao mesmo tempo

Performance em queries com escala



# Insights “Near real-time”

Processamento complexo de eventos em tempo real



- Baixa latência e baixo processamento de grandes *streams* de eventos
- Insight contínuo através de data mining histórico
- Simplicidade de gestão e flexibilidade na implantação

Insights *Real-time*



Qualquer  
Volume




Performance  
Tempo Real



Qualquer Dado







Graças à fácil integração do PDW com Hadoop, a Direct Edge pode usar dados não estruturados para analyses Big Data.

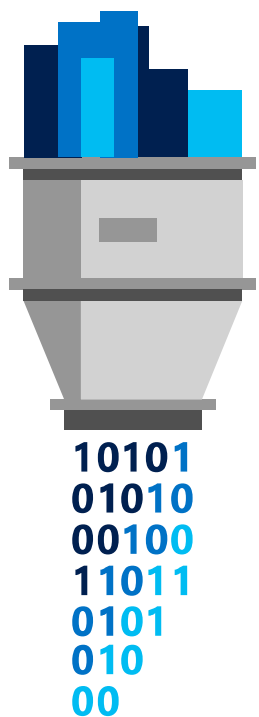
“Nossos analistas agora possuem um entendimento muito mais profundo sobre dados de trading. Conseguem, por exemplo, ter uma melhor visibilidade sobre flutuações mensais de taxas sobre transações.”

**RICHARD HORCHRON**  
CHIEF TECHNOLOGY OFFICER  
DIRECT EDGE

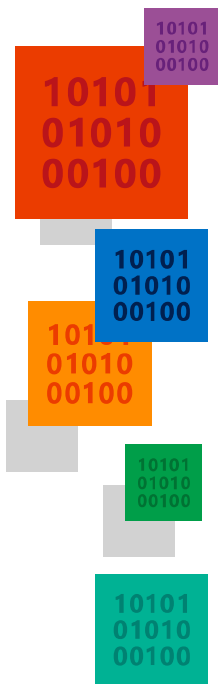


# O acúmulo de informação está nos levando ao limite

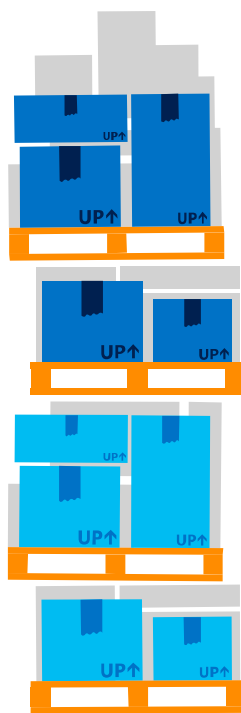
Criação de dados em tempo real



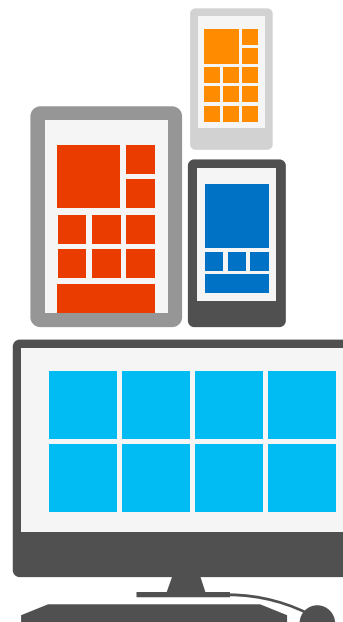
Múltiplos tipos de dados



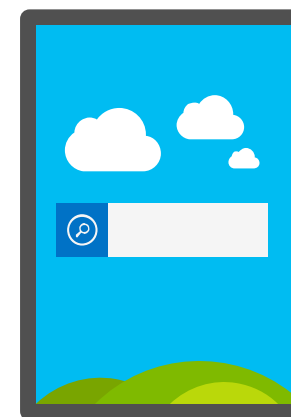
Grandes volumes de dados



Mobilidade



Expectativas dos usuários



Custos de hardware e armazenamento

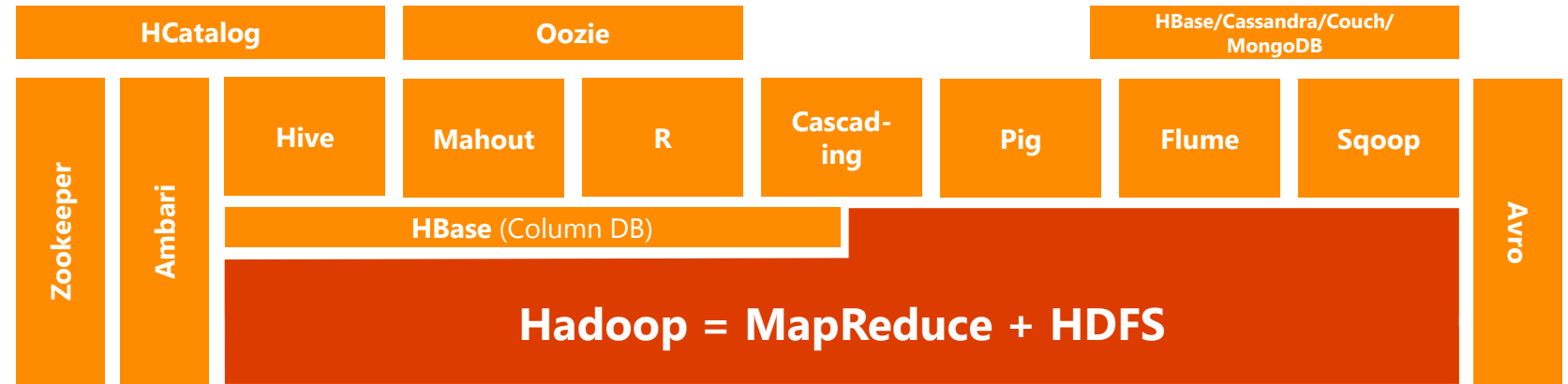
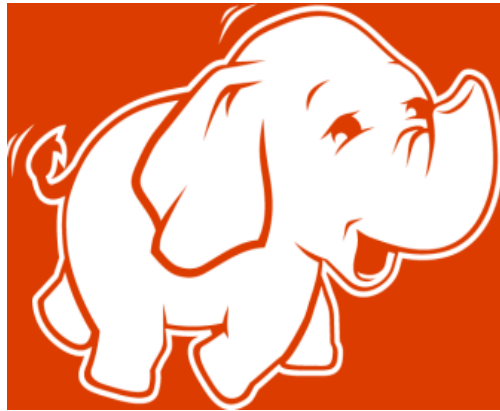




# O que é Big Data?



# O que é Hadoop?



- Sistemas distribuídos, escaláveis em Hardware commodity
- Composto de algumas partes:
  - HDFS – File system distribuído
  - MapReduce – Modelo de programação
  - Outros: HBase, R, Pig, Hive, Flume, Mahout, Avro, Zookeeper

# Cenários Típicos de Big Data



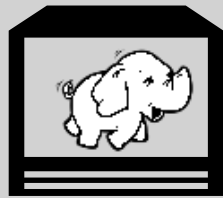
# Explorando dados não-relacionais

Cluster Hadoop em HDP for Windows, HDInsight (Azure, APS)

Windows Azure



Analytics  
Platform System



Hortonworks  
Data Platform



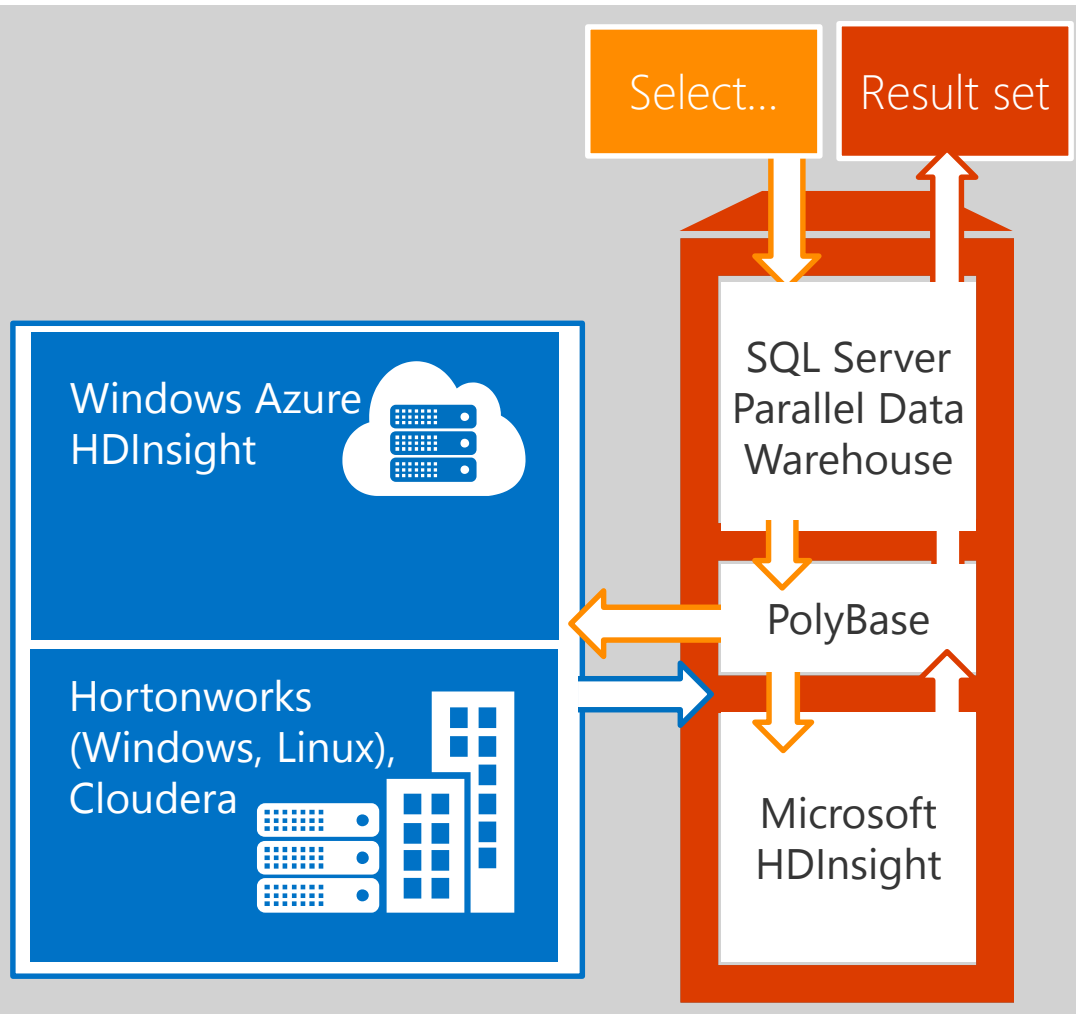
- Gerencie dados não-relacionais
- 100% baseado em Apache
- Simplicidade de gerenciamento do Windows
- Hadoop em software, appliance, nuvem

"Big Data" com simplicidade



# Integre dados relacionais e Hadoop

Consultas integradas usando PolyBase no Analytics Platform System



- Query Hadoop com skills T-SQL existentes
- Query relacional e Hadoop em paralelo
- Query simples
- Híbrido: Hadoop on-premise ou na nuvem
- Sem necessidade de ETL de dados Hadoop para o DW

Query relacional + não relacional

# Plataforma de Dados Microsoft



Microsoft SQL Server

Analytics Platform System

## BI & ANALYTICS



Self-service



Corporate



Collaboration



Mobile



Predictive

## DATA ENRICHMENT & FEDERATED QUERY



Single query model



Extract, transform, load  
& Replication Services



Data quality



Master data management

## DATA MANAGEMENT & PROCESSING



Relational



Non-relational



Analytical



Streaming



Internal & External

## INFRASTRUCTURE



Scalability



Virtualization



Security & Identity



Quality of service



# Get Started Today!

- Sign up for a free architectural design session with your Microsoft representative
- Upgrade your SQL Server DW to PDW – learn more at [www.upgradetopdw.com](http://www.upgradetopdw.com)
- Try HDInsight at [www.microsoft.com/bigdata](http://www.microsoft.com/bigdata)
- Try SQL Server for data warehousing in Windows Azure VMs at [www.windowsazure.com](http://www.windowsazure.com)
- Try SQL Server 2014 CTP1 at [www.microsoft.com/en-us/sqlserver/sql-server-2014.aspx](http://www.microsoft.com/en-us/sqlserver/sql-server-2014.aspx)





# PDW Customers



# HDInsight Customers

