

Building the Enterprise Data lake

Michael Heuer
Senior Specialist Systems Engineer

GLOBAL SPONSORS



Microsoft

DELL EMC / Forum

DELL EMC Data lake momentum

#1

Scale-Out NAS
Market Leader



>8000

Customers

#1

Hadoop Shared Storage

>2000

Big Data Analytics
Customers

#1

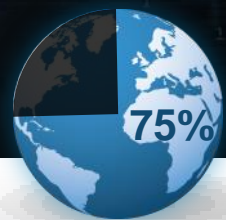
Data Lake

>20%

YoY Growth



UNSTRUCTURED DATA GROWTH



2015



71 EB



2016



106 EB



2017



133 EB

Source: IDC



Total Capacity Shipped, Worldwide



% of Unstructured Data

Digital disruption

By 2018, **33%** of all industry leaders will be disrupted by digitally enabled competitors¹

Source: IDC

45%

believe there is a possibility their business will become obsolete within the next 3 – 5 years

Source: Dell Technologies

*“Digital is the main reason just over half of the companies on the Fortune 500 have **disappeared** since the year 2000”*

Pierre Nanterme
CEO of Accenture

50%

of the G2000 by 2020 will see the majority of their business depend on digitally-enhanced products, services and experiences

Source: IDC

78%

see digital start-ups as a threat to their business, now or in the future

Source: Dell Technologies

What is digital transformation?

Digital Transformation is the application of digital capabilities to products, services, and processes in order to increase customer value, optimize operational efficiency, and develop new monetization opportunities.



Digital displacement

Disruptive business models disintermediate the status quo

sam goody®

Vs.
Music



AOL

Vs.
Online

Google



BORDERS®

Vs.
Books

amazon.com



Vs.
Electronics

overstock.com
amazon.com



Vs.
Movies

NETFLIX



Vs.
Photography



Vs.
Transportation

UBER



Vs.
Hospitality

airbnb



Transforming data into business value



Intelligent Apps



Insights



Analytics



Data Aggregation



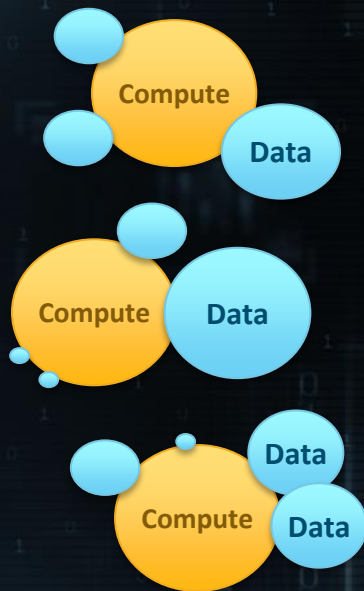
Data



Expanding Big Data Requires - A New Approach

1980s

Bring Data to Compute



**Process-centric
businesses use:**

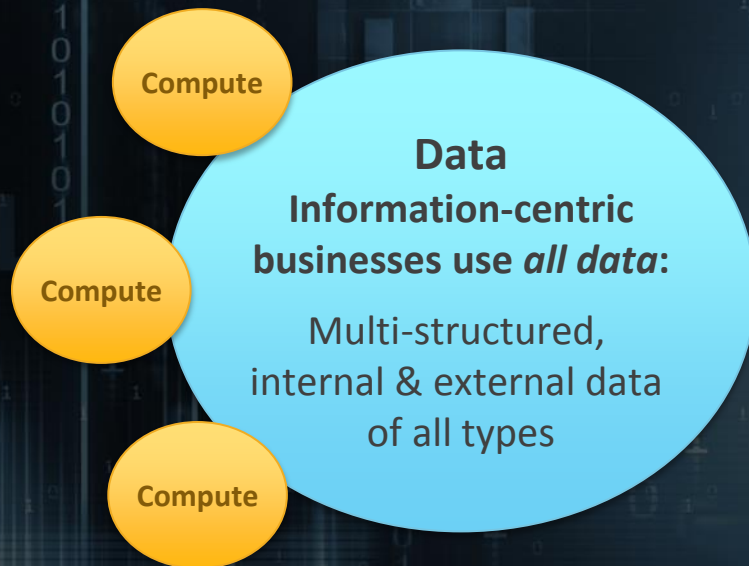
- Structured data mainly
- Internal data only
- “Important” data only

Relative size & complexity



Now

Bring Compute to Data



**Data
Information-centric
businesses use *all data*:**

Multi-structured,
internal & external data
of all types

The OLD Way: Bringing Data to Compute

Complex Architecture

- Many special-purpose systems
- Moving data around
- **No complete views**

Cost of Analytics

- Existing systems strained
- No agility
- **BI backlog**

Time to Data

- Up-front modeling
- Transforms slow
- **Transforms lose data**

Visibility

- Leaving data behind
- Risk and compliance
- **High cost of storage**



The NEW way: Bring Compute to the Data

Business needs
for insights

Operations

Marketing
campaign

Finance

Distill data



Keep the history

The “Business Datalake”

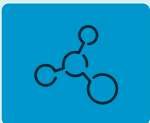
Load data from
all sources



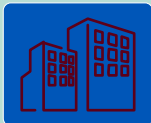
Web



Sensor



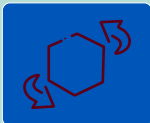
Network



ERP



CRM



PLM



Supplier



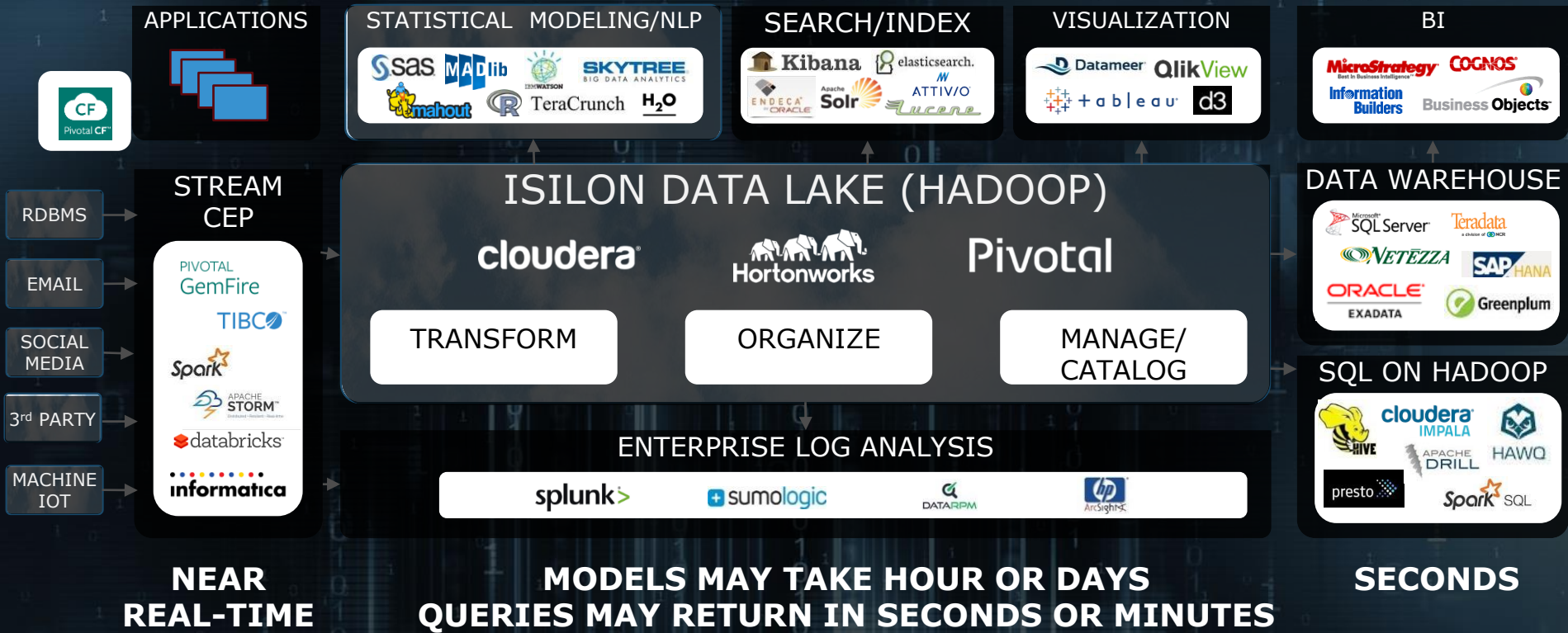
Social Media



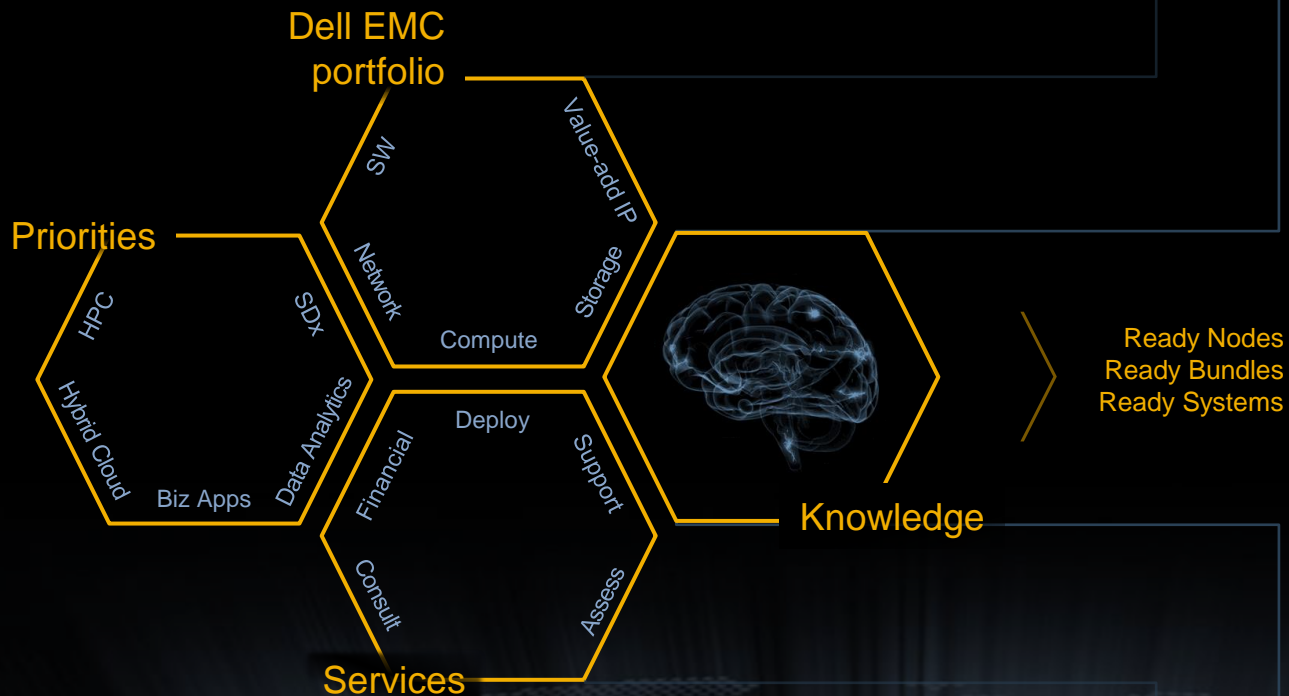
Market

Transactional systems

OUR BIG DATA LANDSCAPE



The Ready Solutions formula



Dell EMC Ready Bundles for Hadoop



Dell EMC Ready Bundle for Cloudera Hadoop

End-to-end data management, processing and analytics with 'no-code-needed' deployment plus high double-digit performance gains

Dell EMC Ready Bundle for Hortonworks with Isilon

Capacity optimized and efficient data processing and data lake platform leveraging Isilon for shared storage for Hadoop with HDFS as a protocol.

Dell EMC Ready Bundle for Hortonworks Hadoop

Open Source data management, processing and analytics solution that efficiently process multi-structured data volumes using existing tools and resources



Dell EMC Ready Solutions for Big Data

Unlocking data insights for up to 50% better revenue growth

Dell EMC differentiation

Fully validated and certified configurations

Rapid, worry-free Hadoop deployments with exclusive 'no code needed' tools

Reduced Splunk OpEx via HCI options that eliminate clustered management overhead and administration

Full-lifecycle, outcome-focused services

Big Data Vision • Accelerator • Health Check • Deployment • Support • Financial

#1 TPC-x
Big Bench

Scale Factor 10000
Price/Performance
May 2017

60% 
ETL time reduction

\$1.97 
for every \$1.00
ROI in 6 months

76% 
Lower data
analytics costs

Cloudera Hadoop • Splunk • Hortonworks Hadoop



Over 350 customers



AUDIENCE SCIENCE®

ebay

ExxonMobil

GROUPON®



MERKLE

Praedicat

StateFarm

SONY



verizon✓

WELLS FARGO

"We've completely redesigned how we do analytics. We went from 10,000 transactions per second to 150,000 transactions per second. After we implemented the new Dell PowerEdge Servers, our analysts and business leaders can mine that data in ways they couldn't before."

Jim Bacheza, VP • Audience Science

Tony Giordano, EVP • Merkle

DELLEMC