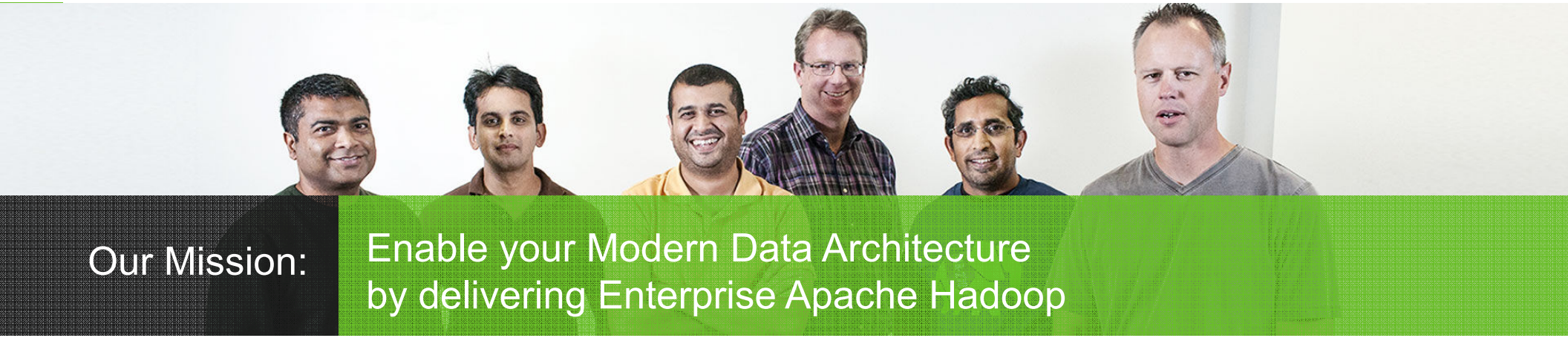
Modern Data Architecture with Enterprise Apache Hadoop

Hortonworks. We do Hadoop.

Jeff Markham
Technical Director, APAC
jmarkham@hortonworks.com



Our Mission: Enable your Modern Data Architecture by delivering Enterprise Apache Hadoop

Open Leadership

Drive innovation in the open exclusively via the Apache community-driven open source process

Enterprise Rigor

Engineer, test and certify Apache Hadoop with the enterprise in mind

Ecosystem Endorsement

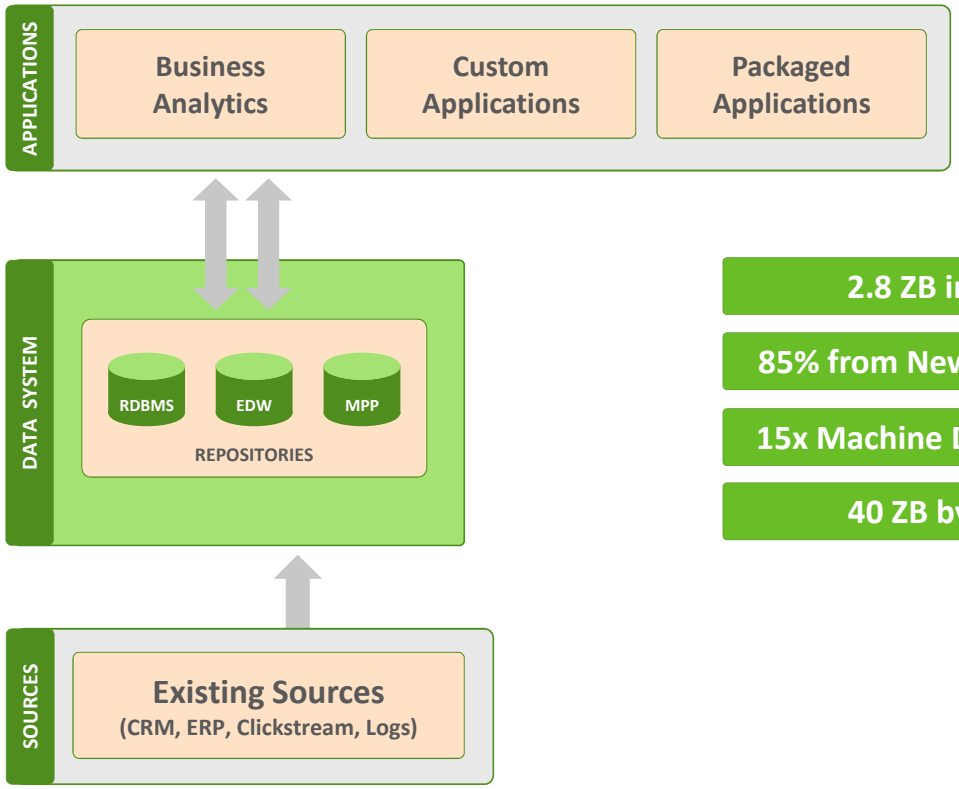
Focus on deep integration with existing data center technologies and skills

Headquartered in Palo Alto, CA; 300+ employees and growing

Reseller Partners: Microsoft TERADATA SAP hp

Hortonworks

A data architecture under pressure from new data



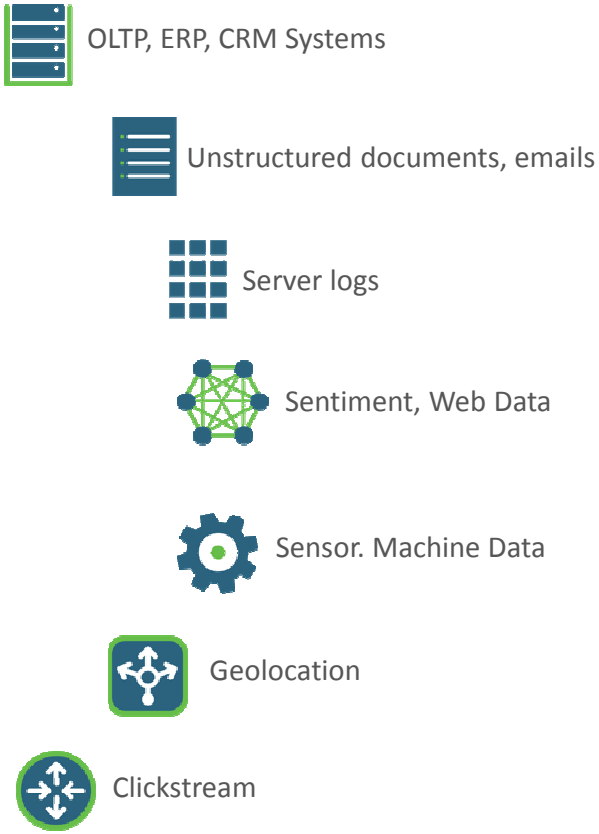
2.8 ZB in 2012

85% from New Data Types

15x Machine Data by 2020

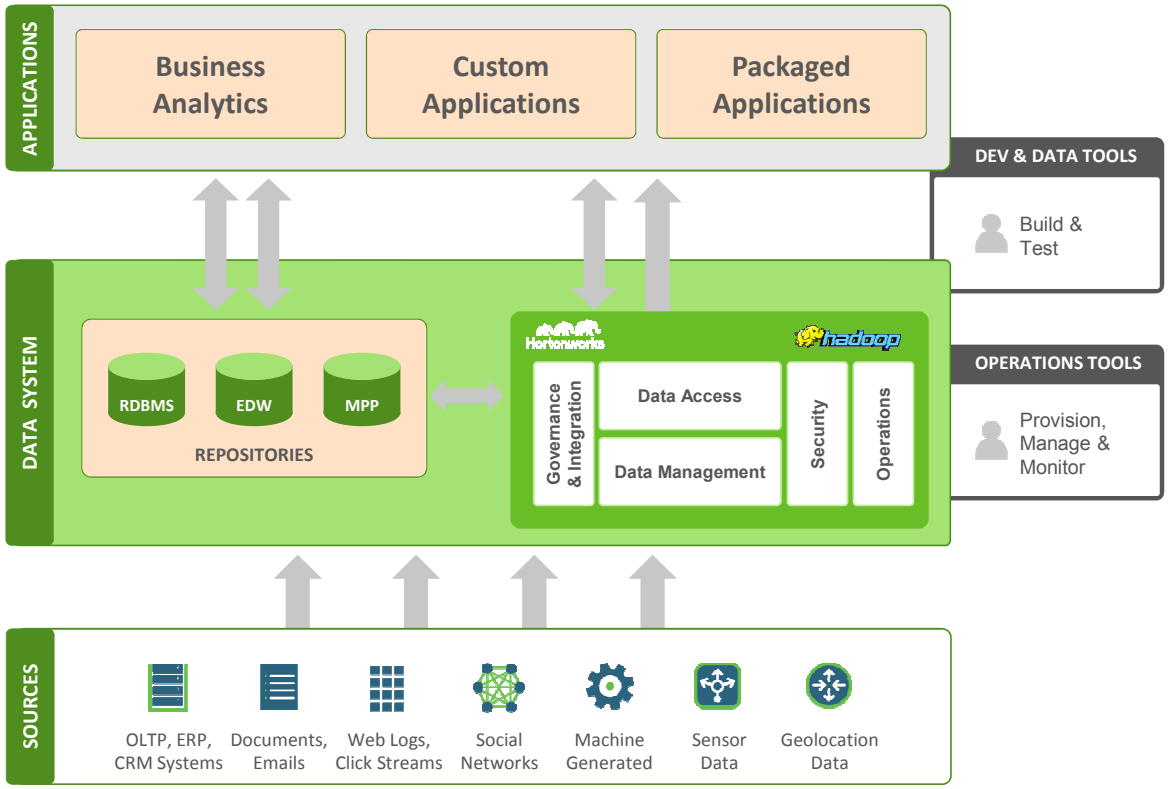
40 ZB by 2020

Source: IDC



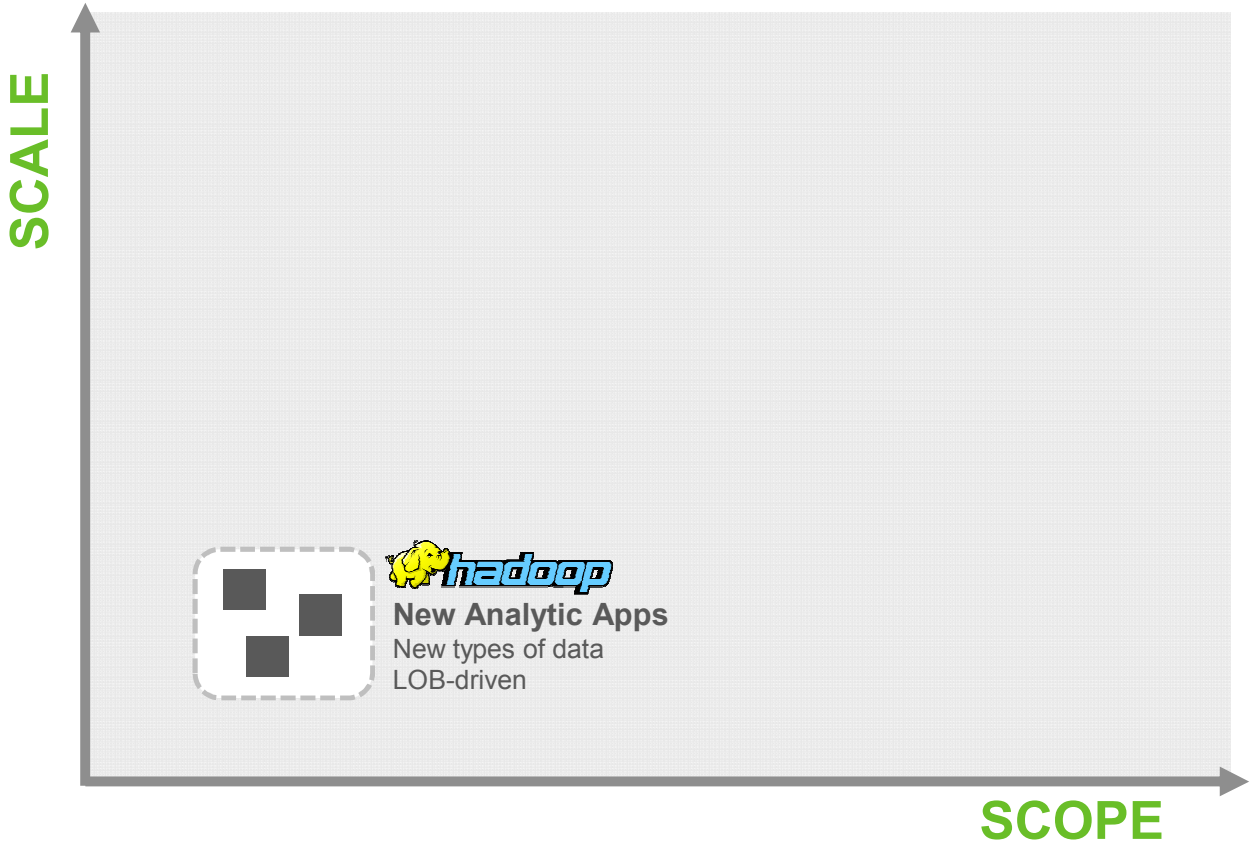
Hortonworks

Hadoop within an emerging Modern Data Architecture



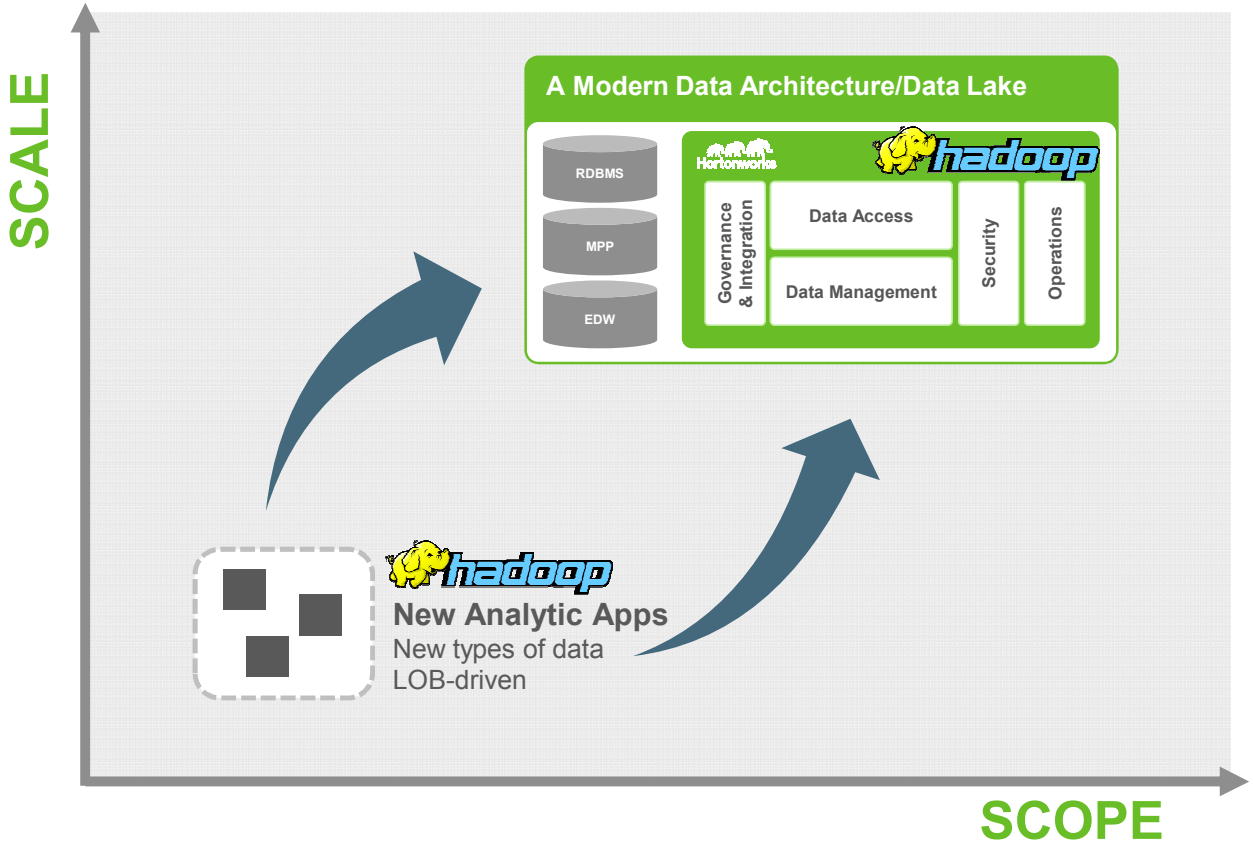
Hortonworks

Hadoop: typically used for new analytic applications…



Hortonworks

… and incrementally delivers a ‘Data Lake’



Data Lake

An architectural shift in the data center that uses Hadoop to deliver deeper insight across a large, broad, diverse set of data at efficient scale

Hortonworks

Hadoop Value: New types of data

Sentiment
Understand how your customers feel about your brand and products – right now

Clickstream
Capture and analyze website visitors' data trails and optimize your website

Sensors
Discover patterns in data streaming automatically from remote sensors and machines

Geographic
Analyze location-based data to manage operations where they occur

Server Logs
Research logs to diagnose process failures and prevent security breaches

Unstructured
Understand patterns in files across millions of web pages, emails, and documents

Hadoop unlocks a new approach: Iterative Analytics

Current Reality

Apply schema on write

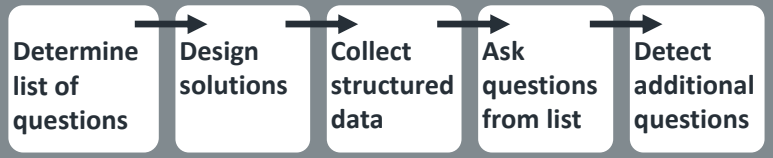
Dependent on IT

Augment w/ Hadoop

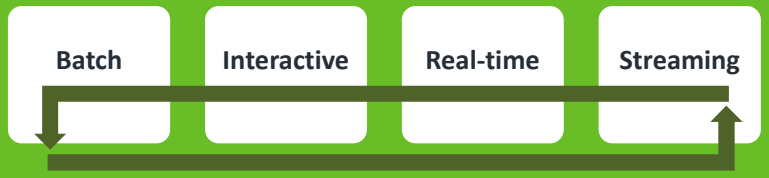
Apply schema on read

Support range of access patterns to data stored in HDFS: polymorphic access

SQL
Single Query Engine
Repeatable Linear Process



Hadoop
Multiple Query Engines
Iterative Process: Explore, Transform, Analyze



New analytic applications for new types of data

Financial Services

- New Account Risk Screens
- Fraud Prevention
- Trading Risk
- Maximize Deposit Spread
- Insurance Underwriting
- Accelerate Loan Processing

Retail

- 360° View of the Customer
- Analyze Brand Sentiment
- Localized, Personalized Promotions
- Website Optimization
- Optimal Store Layout

Telecom

- Call Detail Records (CDRs)
- Infrastructure Investment
- Next Product to Buy (NPTB)
- Real-time Bandwidth Allocation
- New Product Development

Manufacturing

- Supplier Consolidation
- Supply Chain and Logistics
- Assembly Line Quality Assurance
- Proactive Maintenance
- Crowdsourced Quality Assurance

Healthcare

- Genomic data for medical trials
- Monitor patient vitals
- Reduce re-admittance rates
- Store medical research data
- Recruit cohorts for pharmaceutical trials

Utilities, Oil & Gas

- Smart meter stream analysis
- Slow oil well decline curves
- Optimize lease bidding
- Compliance reporting
- Proactive equipment repair
- Seismic image processing

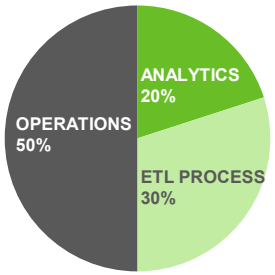
Public Sector

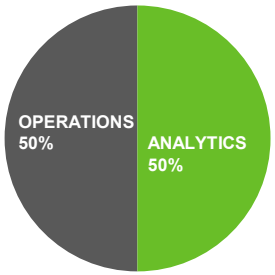
- Analyze public sentiment
- Protect critical networks
- Prevent fraud and waste
- Crowdsource reporting for repairs to infrastructure
- Fulfill open records requests

Hortonworks

Hadoop delivers compelling economics

EDW Optimization





Current Reality

EDW at capacity: some usage from low value workloads

Older data archived, unavailable for ongoing exploration

Source data often discarded

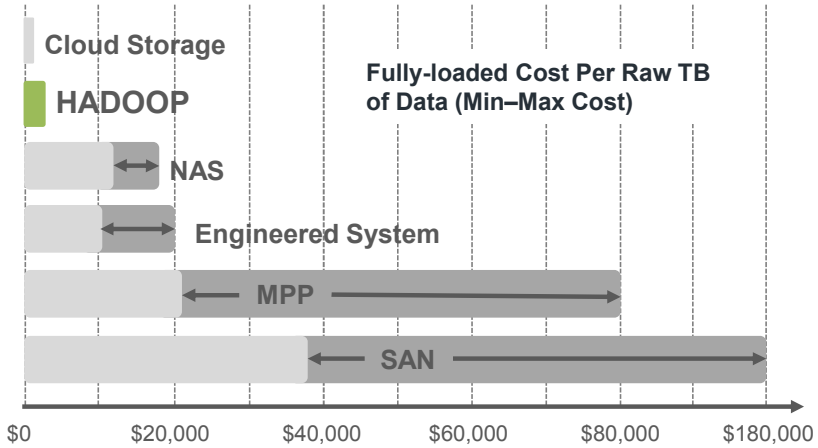
Augment w/ Hadoop

Free up EDW resources from low value tasks

Keep 100% of source data and historical data for ongoing exploration

Mine data for value after loading it because of schema-on-read

Commodity Compute & Storage



Hadoop Enables Scalable Compute & Storage at a Compelling Cost Structure

Hortonworks

Delivering the Core Capabilities of Enterprise Hadoop

Hortonworks

Core Capabilities of Enterprise Hadoop

PRESENTATION & APPLICATION

Enable both existing and new application to provide value to the organization

ENTERPRISE MGMT & SECURITY

Empower existing operations and security tools to manage Hadoop

GOVERNANCE & INTEGRATION

Load data and manage according to policy

DATA ACCESS

Access your data simultaneously in multiple ways (batch, interactive, real-time)

Store and process all of your Corporate Data Assets

DATA MANAGEMENT

SECURITY

Provide layered approach to security through Authentication, Authorization, Accounting, and Data Protection

OPERATIONS

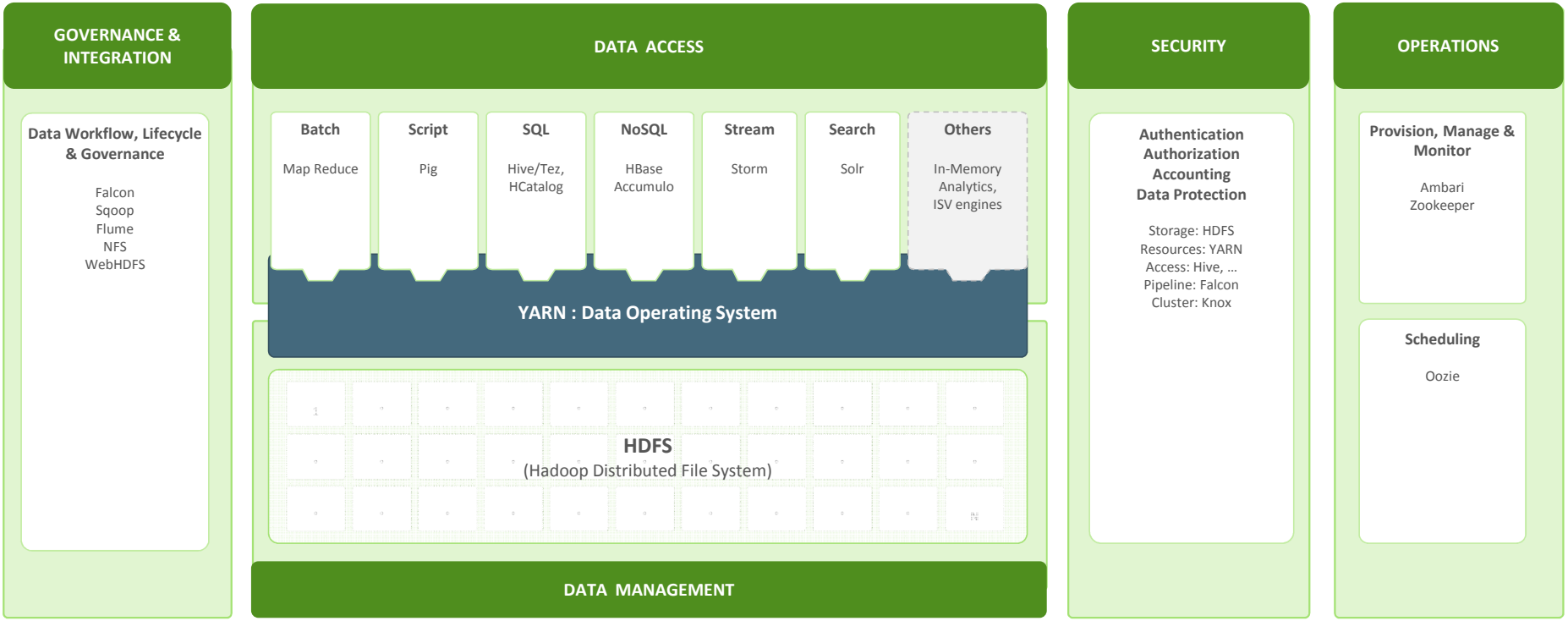
Deploy and effectively manage the platform

Provide deployment choice across physical, virtual, cloud

DEPLOYMENT OPTIONS

Hortonworks

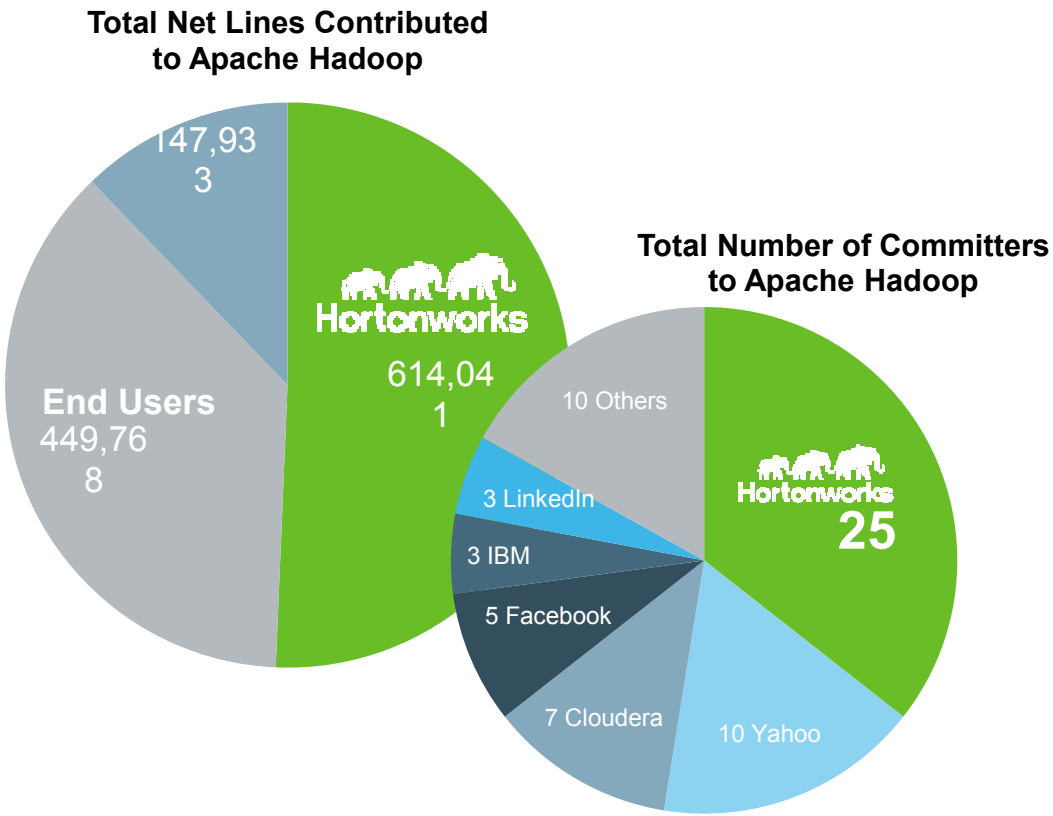
Core Technologies of Enterprise Hadoop



Hortonworks

Hortonworks provides leadership to Hadoop

Apache Project	Committers	PMC Members
Hadoop	21	13
Tez	10	4
Hive	15	3
HBase	8	3
Pig	6	5
Sqoop	1	0
Ambari	21	12
Knox	6	2
Falcon	4	2
Oozie	2	2
Zookeeper	2	1
Flume	1	0
Accumulo	2	2
Storm	1	0
TOTAL	**103**	**49**

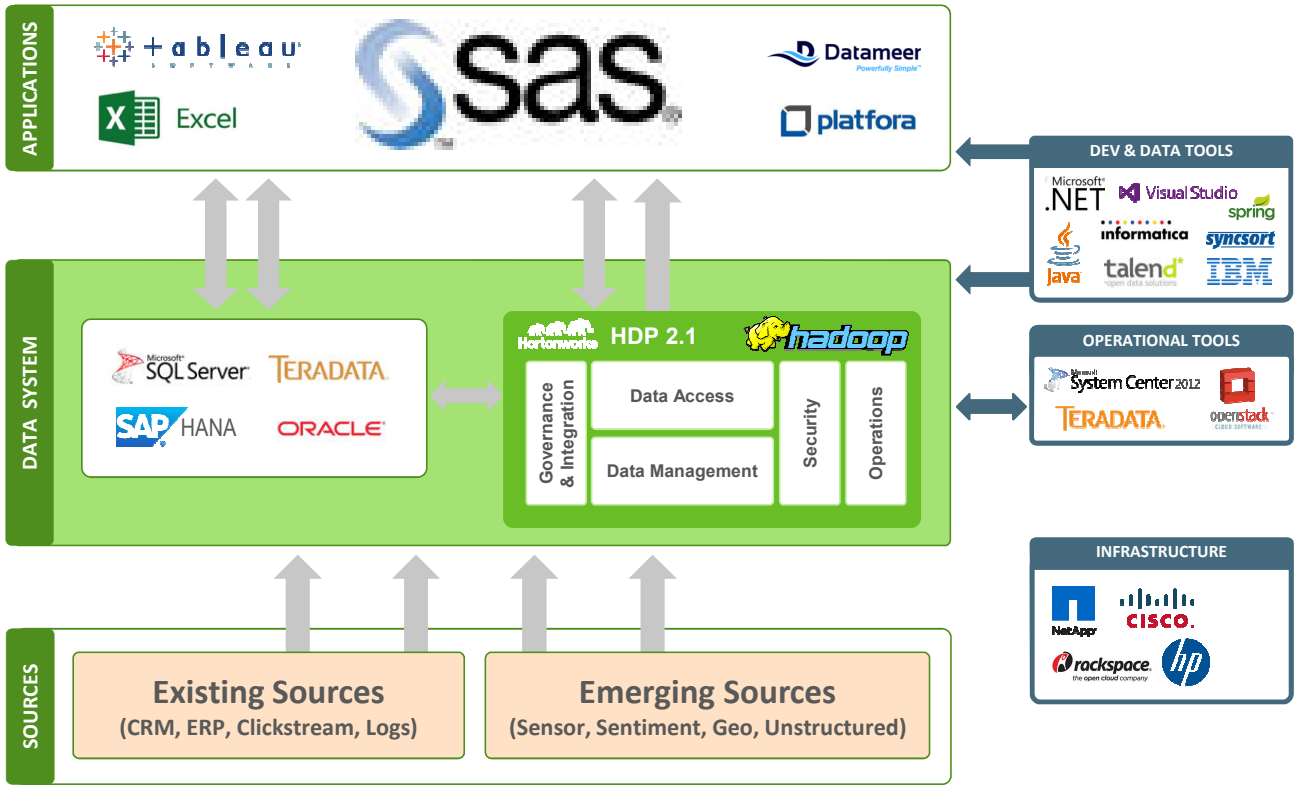


Hortonworks

The Partners You Rely On,
Rely On Hortonworks for Hadoop

Hortonworks

Hadoop is wholly integrated into the data center



DEPTH

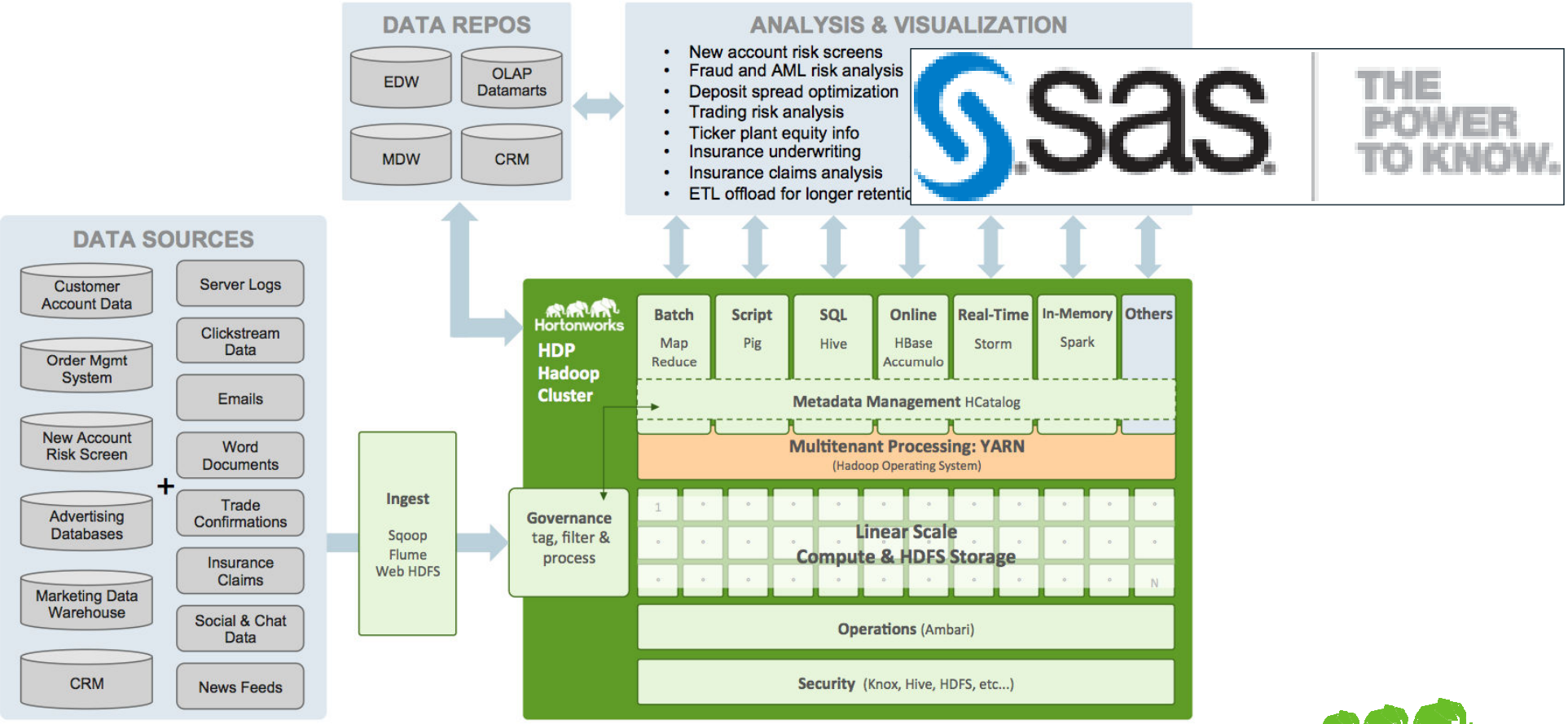
Hortonworks engages in deep engineered relationships with the leaders in the data center, such as Microsoft, Teradata, Redhat & SAP

BREADTH

Hundreds of partners work with us to certify their applications to work with Hadoop so they can extend big data to their users

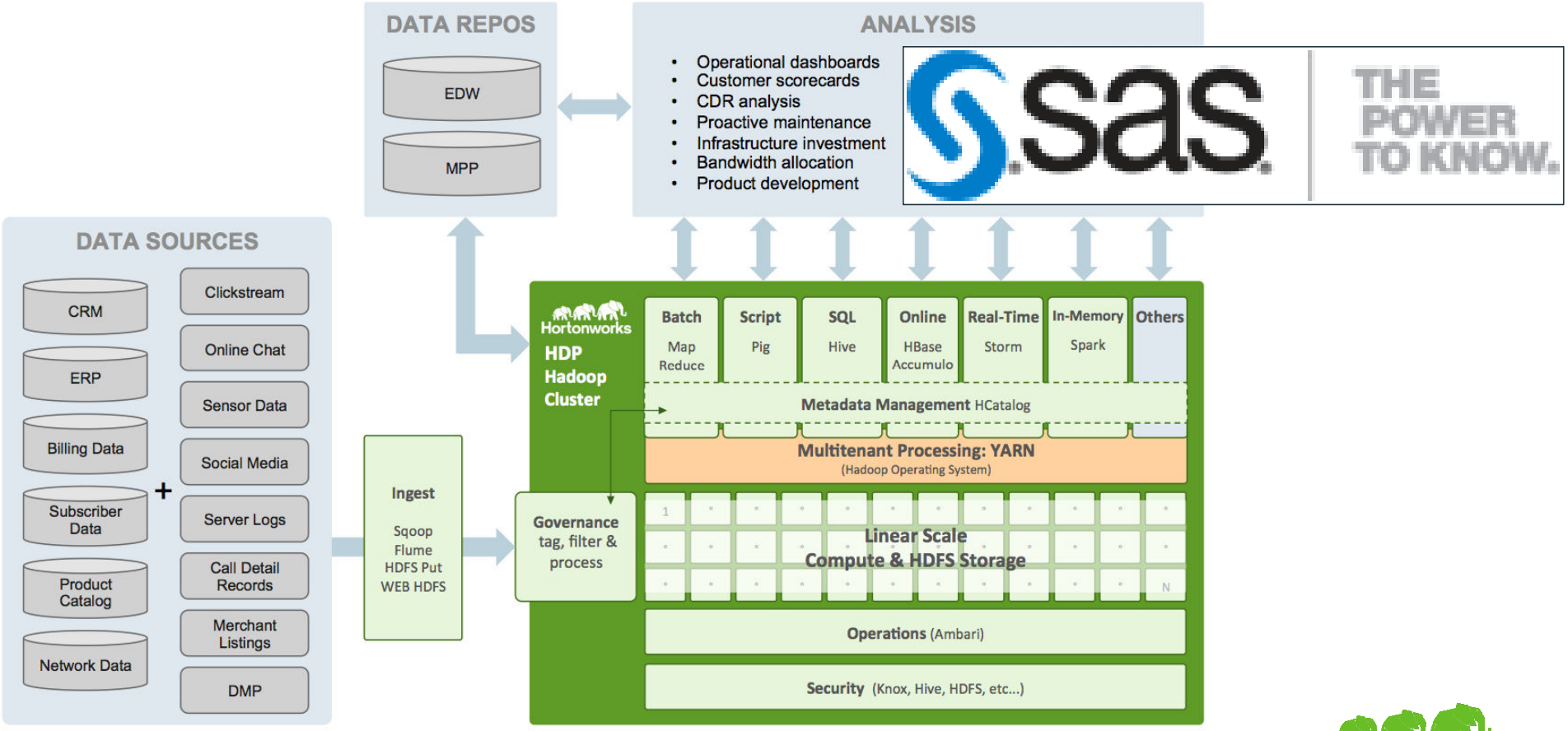
Hortonworks

Financial Services High Level Use Cases



Hortonworks

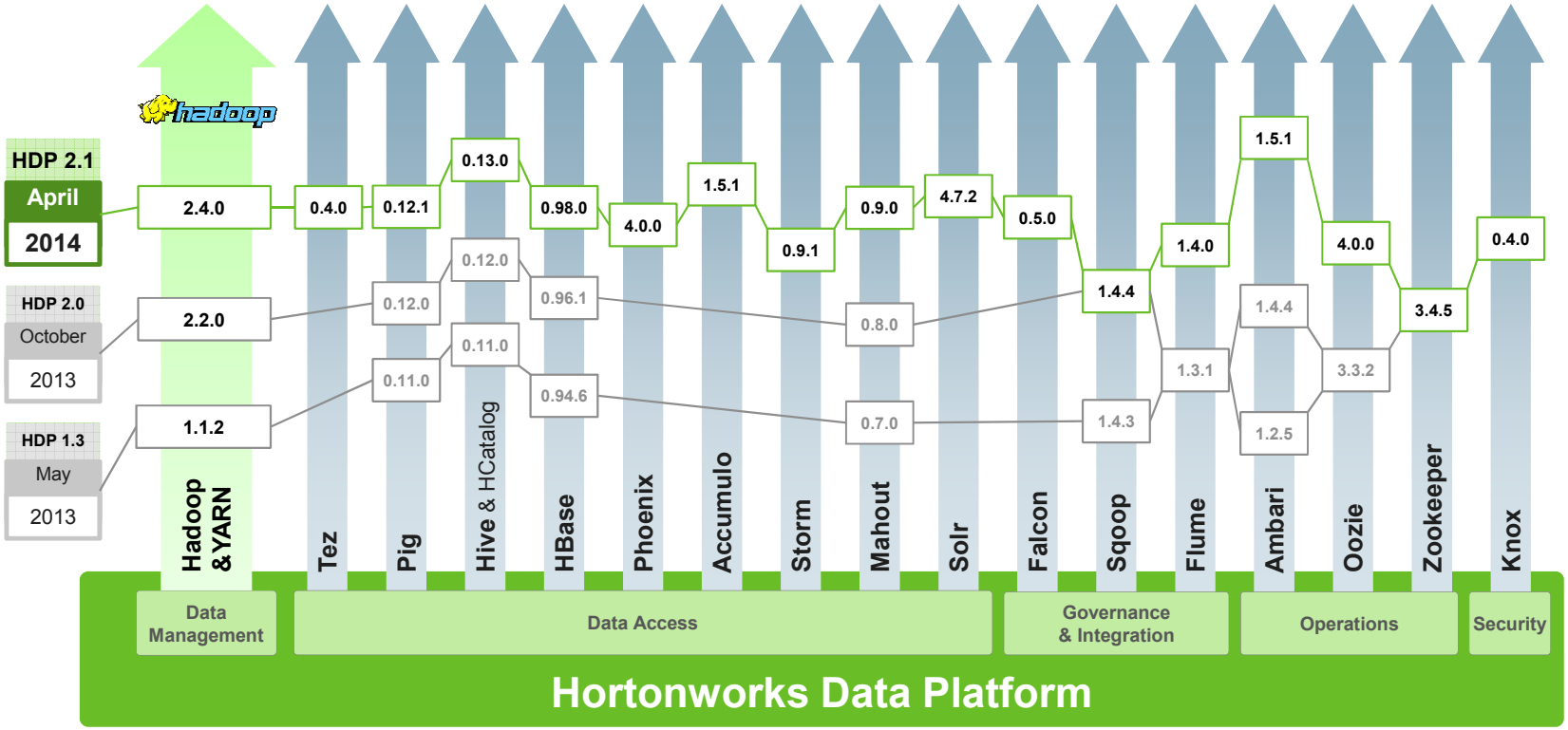
Telco High Level Use Cases



Hortonworks

HDP 2.1: Reliable, Consistent & Current

HDP certifies most recent & stable community innovation



Hortonworks

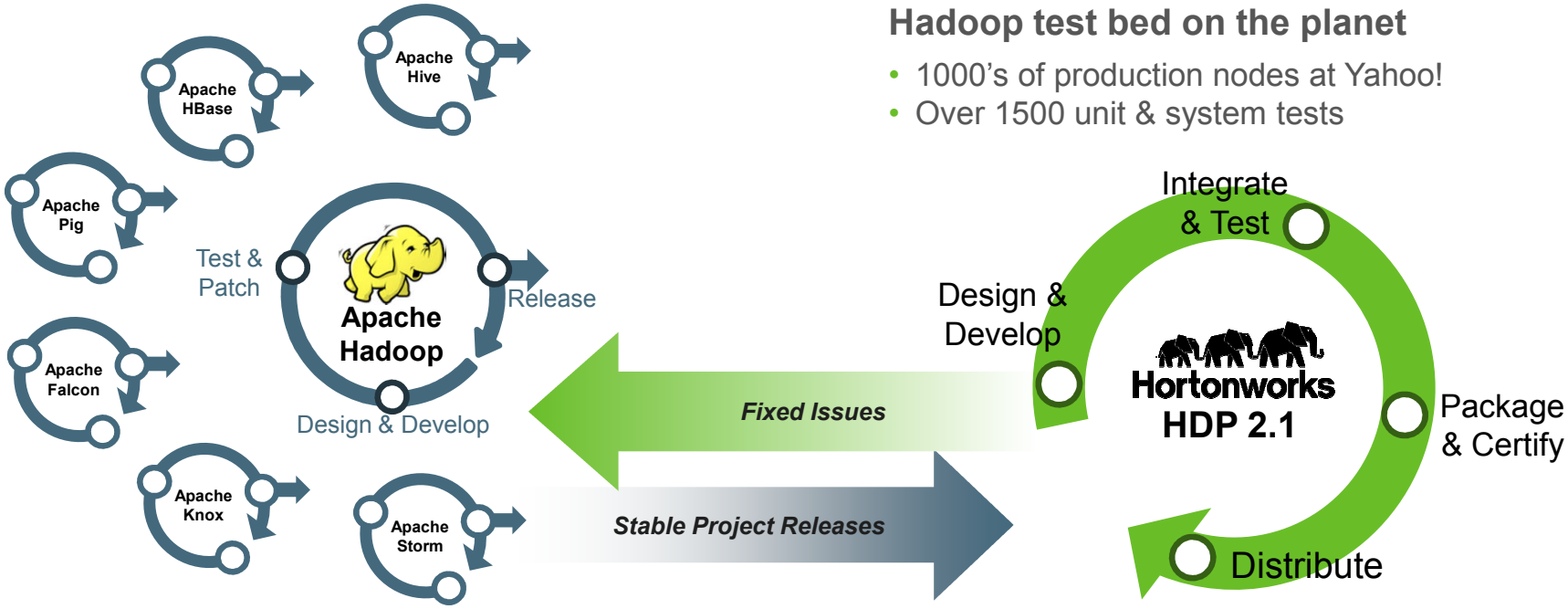
Hortonworks process for Enterprise Hadoop

Upstream Community Projects

Downstream Enterprise Product

Certified at scale using the most advanced Hadoop test bed on the planet

- 1000's of production nodes at Yahoo!
- Over 1500 unit & system tests



***Virtuous cycle** when development & fixed issues done upstream & stable project releases flow downstream*

Hortonworks

Working in the community for the enterprise

Hortonworks de-risks your investment…

All support and implementation backed by the largest and most experienced Hadoop team on the planet

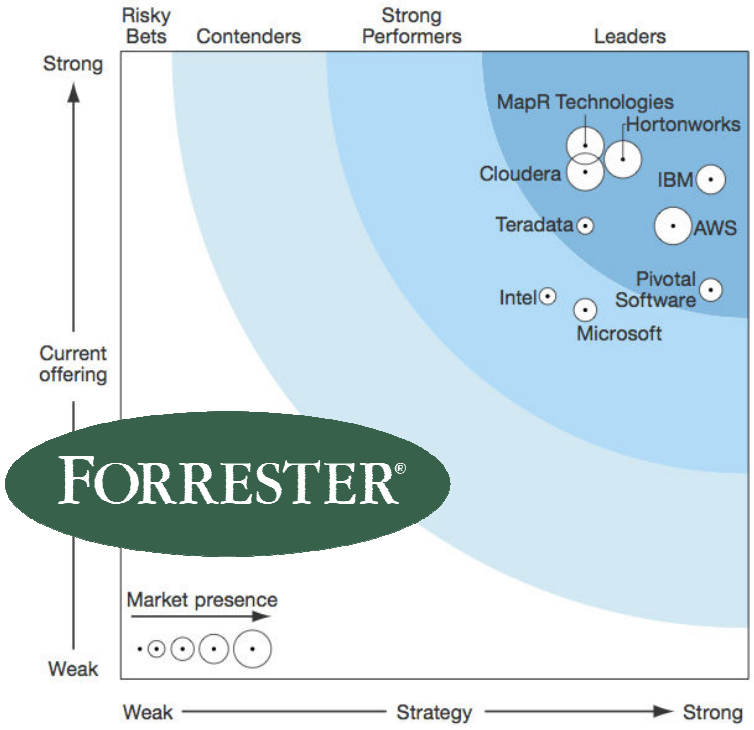
A Two Way Street

- Gain access to the expertise and knowledge of the Hadoop community
- All issues or feature requests represented in the community

Hortonworks

Hortonworks: A Leader In Hadoop

The Forrester Wave™: Big Data Hadoop Solutions, Q1 2014



"Hortonworks loves and lives open source innovation"

Vision & Execution for Enterprise Hadoop.
Hortonworks leads with a strong strategy and roadmap for open source innovation with Hadoop and a strong delivery of that innovation in Hortonworks Data Platform.

World Class Support and Services.
Hortonworks' Customer Support received a maximum score and was significantly higher than both Cloudera and MapR.

Key Strategic Partnerships.
Hortonworks unique strategic partnerships with Microsoft, SAP, Teradata and others are a key strength as part of its overall strategy of ecosystem partnership to accelerate Hadoop adoption in the enterprise.

Hortonworks

The value of Open…

Connect With the Community

We employ a large number of Apache project committers & innovators so that **you** are represented in the open source community

Avoid Vendor Lock In

Hortonworks Data Platform is close to the open source trunk as possible and is developed 100% in the open so **you** are never locked in

Partnerships that Matter

We work with partners to deeply integrate Hadoop with data center technologies so **you** can leverage existing skills and investments

Certified for the Enterprise

We engineer, test and certify the Hortonworks Data Platform at scale to ensure reliability and stability **you** require for enterprise use

Support from the Experts

We provide the highest quality of support for deploying at scale. **You** are supported by hundreds of years of Hadoop experience

Thank You

Jeff Markham
Technical Director, APAC
jmarkham@hortonworks.com

Hortonworks