Hortonworks
CISC Innovation day

Simon gregory
sgregory@hortonworks.com

Here was the ask…

Hortonworks' data reposition - how this works and the types of data you work with. 1: Data Types & Value.

What have been your latest challenges and how have you overcome them? 2: Challenges.

In a fast-moving space, what directions do you see things moving in future? 3: Future.

Are there things in the 'standards' arena that are particularly trying? 4: Standards (or lack of).

Data types and Value…

New data types:

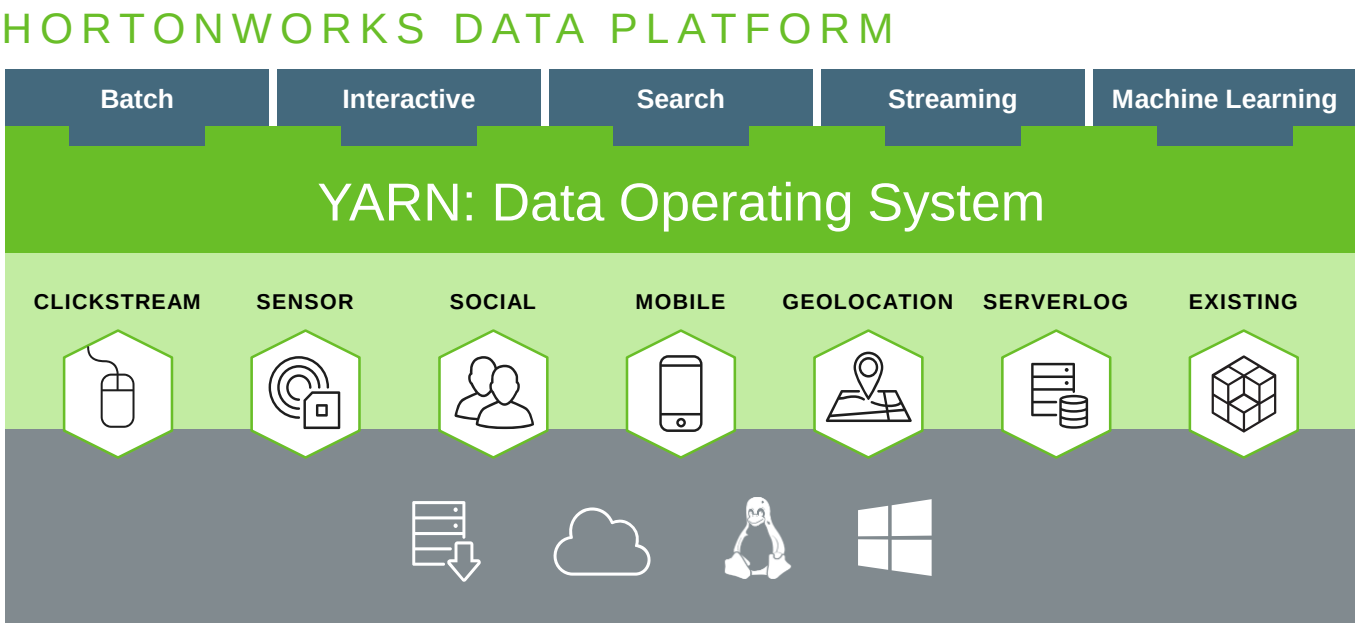
Most new data types don't typically fit neatly into rows and columns making them unsuitable for existing data structures without considerable transformation.

Hadoop supports countless data types with no transformation required on ingest.

There's a significant shift from batch orientated use cases to real time streaming and interaction. Hadoop allows for both by enabling multiple ingest and execution engines within a single cluster. Many customers now combine these approaches.

Value:

The ability to "late bind" with no existing or predefined schema provides the ability to model mixed data types rapidly.
The ability to combine multiple data types provides new insights whilst maintaining the source data for remodeling purposes or other use cases.



An example of mixed and new data types.

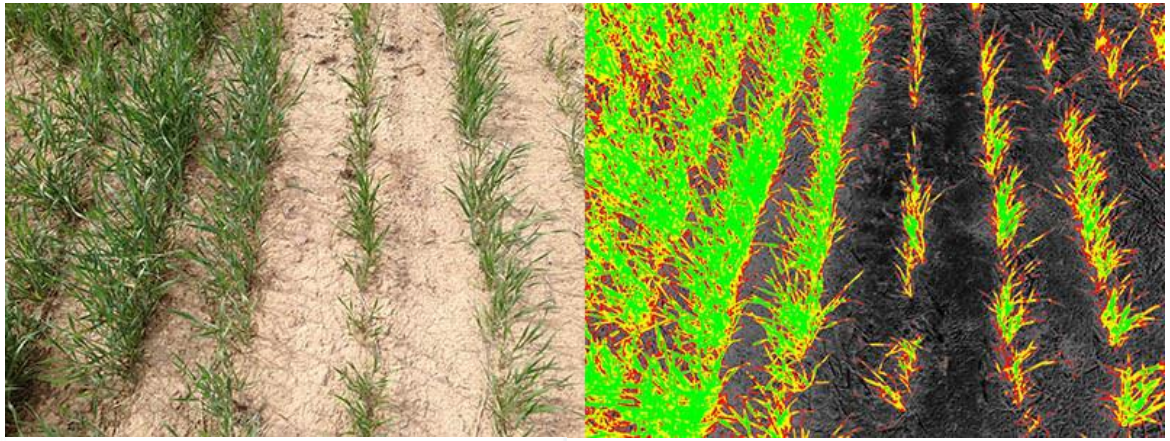
Precision Agriculture:

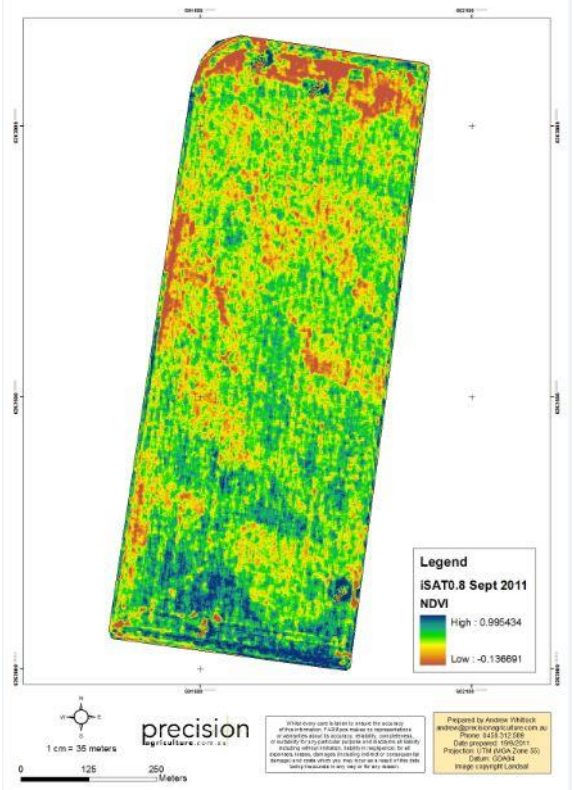
Soil sensor data

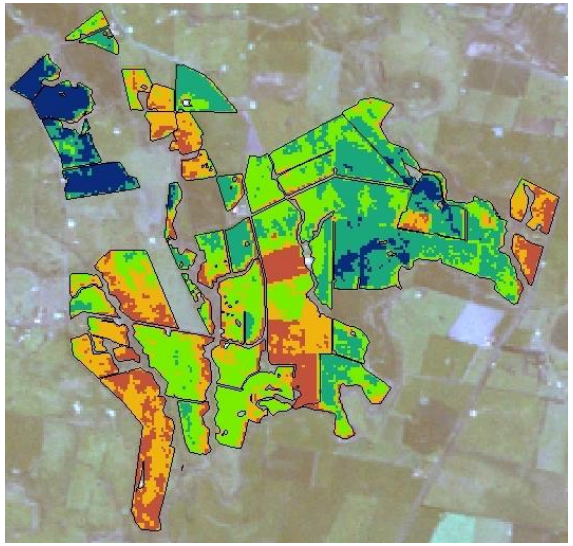
Historical yield results

Historical weather information

Satellite imagery







http://video.esri.com/watch/4656/becks-hybrids

http://video.esri.com/watch/4657/big-data-_and_-geoanalytics

The net result is that the correct seed can be identified for any specific soil / environmental conditions and the exact date for sewing can be specified to produce the highest yield based on the crop location.

Hortonworks®

Challenges…Hadoop is a platform not a product.

Therefore we must deliver enterprise software capabilities across the platform.

Consistency:

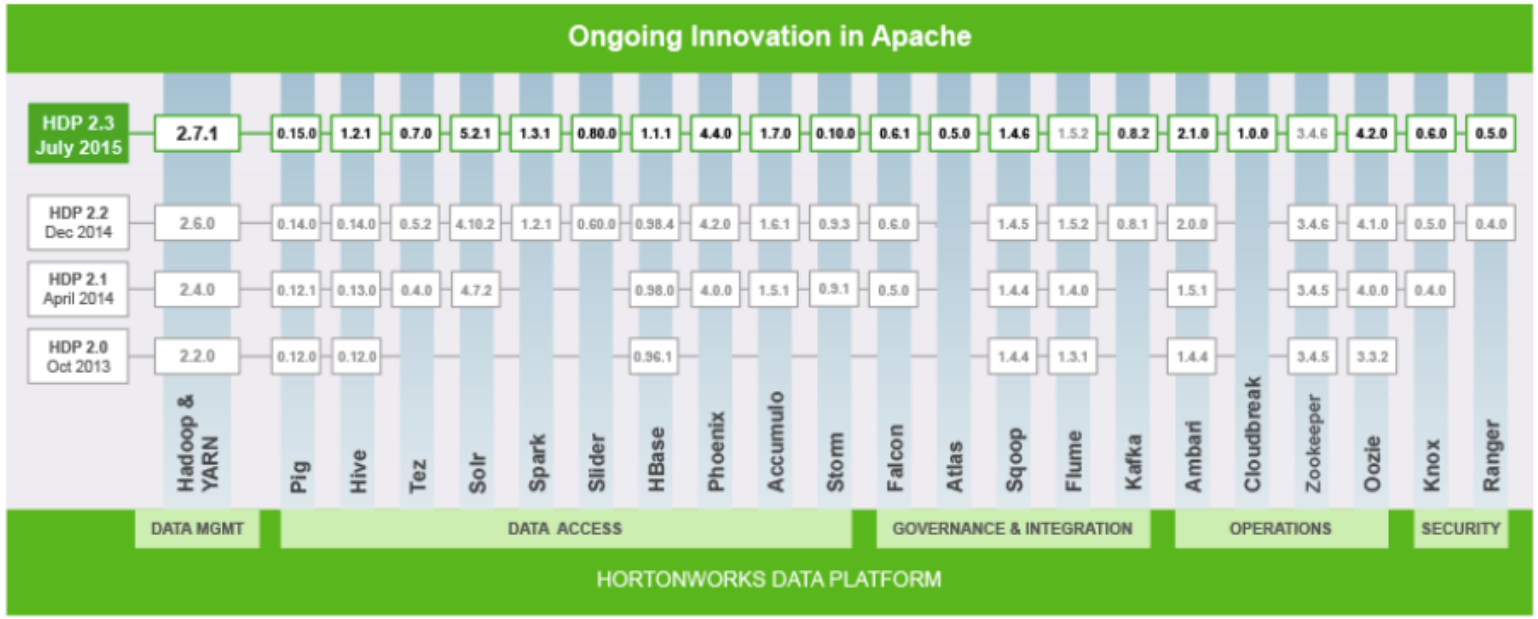
We at Hortonworks not only commit 100% of our code to the Apache Software Foundation (ASF) but focus on applying consistency across multiple Apache projects in 3 key areas: Security; Operations, Governance and integration.

Simplification:

Deliver ease of use for the deployment, Management and ongoing operations and support of the platform.

Innovation:

Develop existing and and contribute new projects to the ASF community based on our customers requirements.



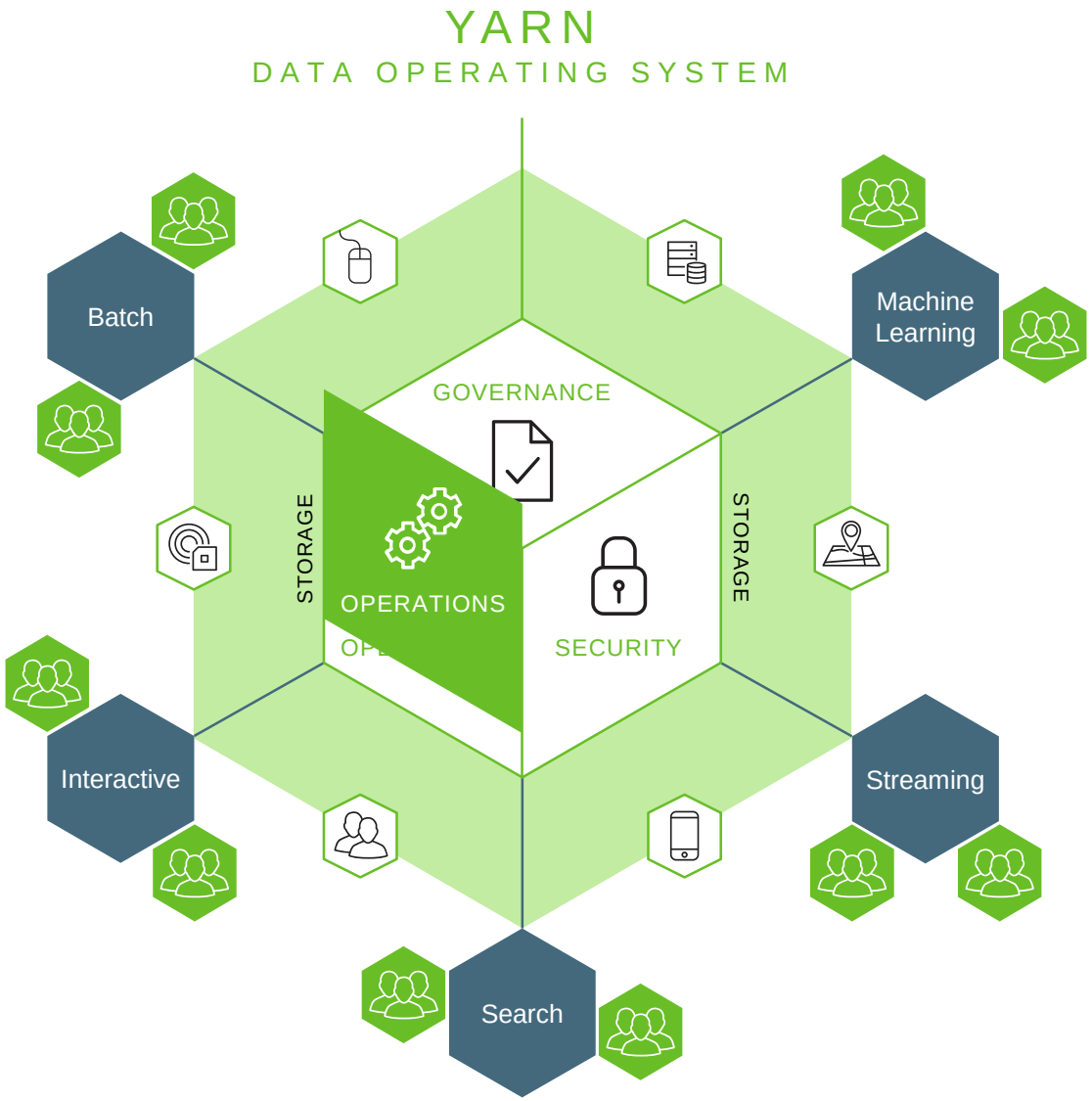
Predictability:

Co-founders of the ODP initiative driving testing and compatibility standards across the Hadoop and partner eco-system. http://opendataplatform.org

Hortonworks®

Challenges:

Provide Consistent Operations



Centralized

management and monitoring of Hadoop clusters and workloads must be centralized across all projects in the platform: **Apache Ambari**

Automated Provisioning

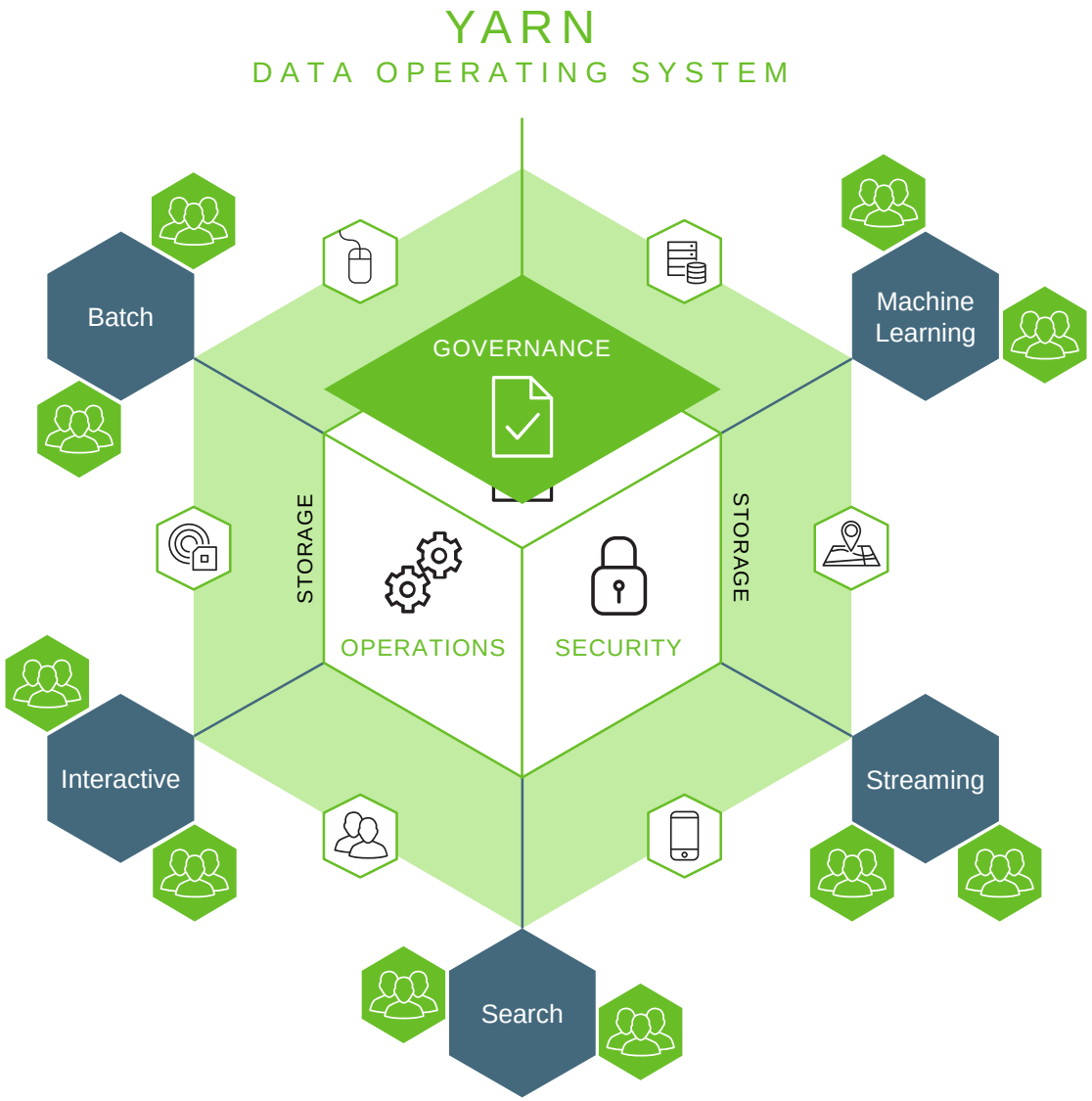
Customers require combinations of on-premis, cloud and hybrid services: **Cloudbreak**

http://publicrepo-1.hortonworks.com/HDP/cloudbreak/cloudbreakdeplo yer_1.0.0_Linux_x86_64.tgz

Hortonworks®

Challenges:

Governance, Provenance and Workflow



Data Management: **Apache Falcon & Apache Atlas**

along the entire data lifecycle

Data Governance

enables comprehensive data lineage through a hybrid approach

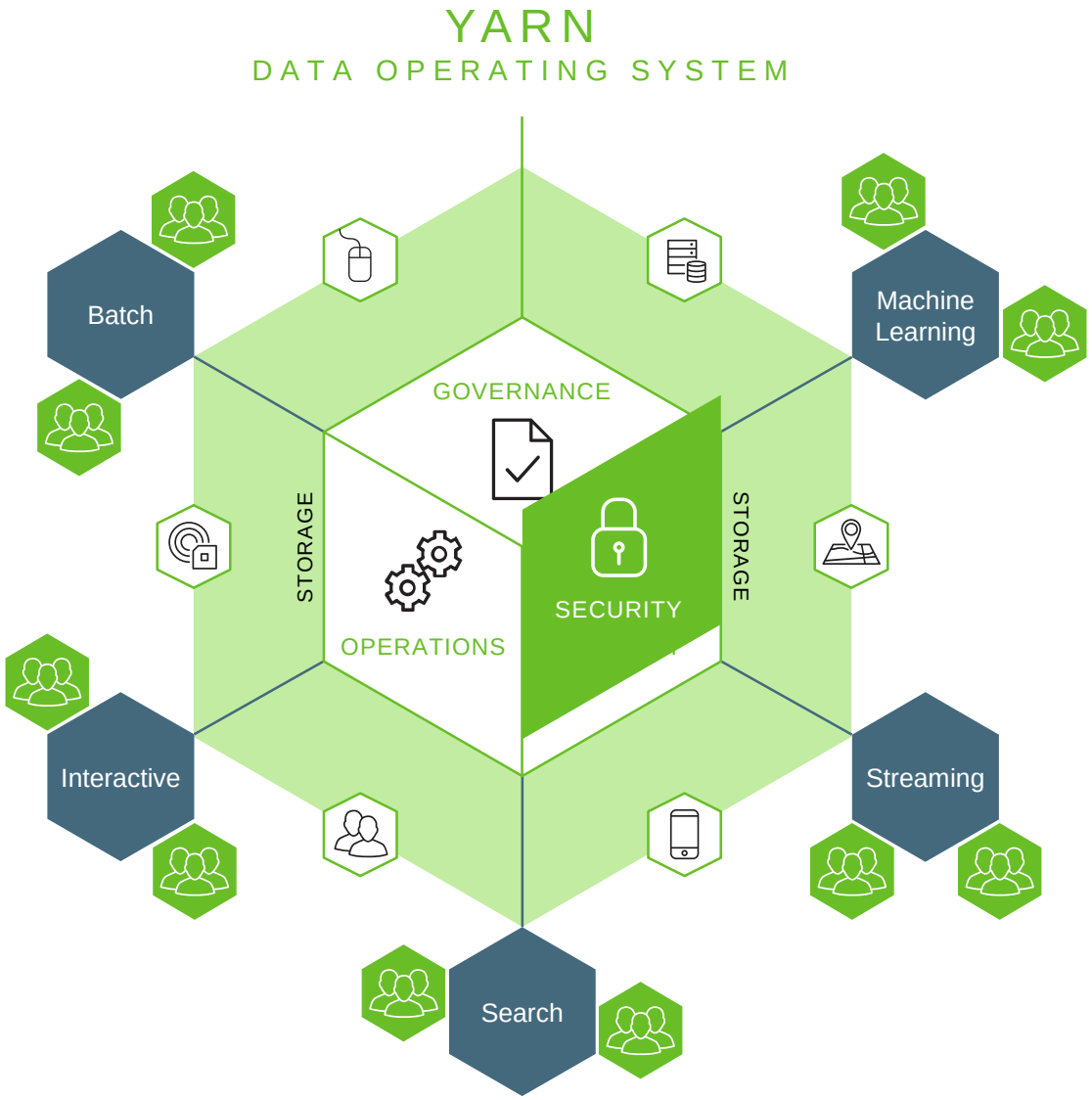
Interoperability of data

across the Hadoop ecosystem, through a common metadata store

Hortonworks®

Challenges:

Provide Comprehensive Security



Comprehensive Security:

Apache Ranger & Apache Knox

Encrypt Data

at rest and in motion.

Centralized Administration

of security policies and user authentication

Fine-Grain Authorization

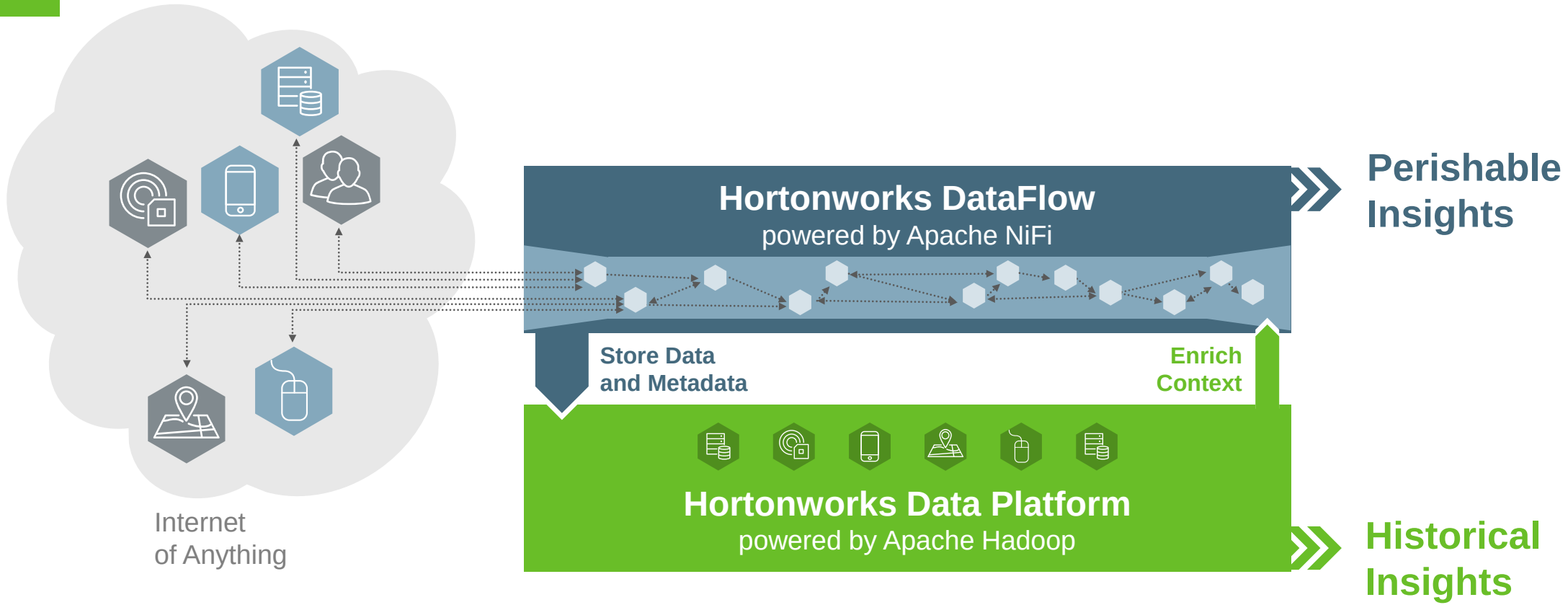
for data access control

Centralized Audit capability

Who, what and when.

Hortonworks®

3: The Future is everything….literally.

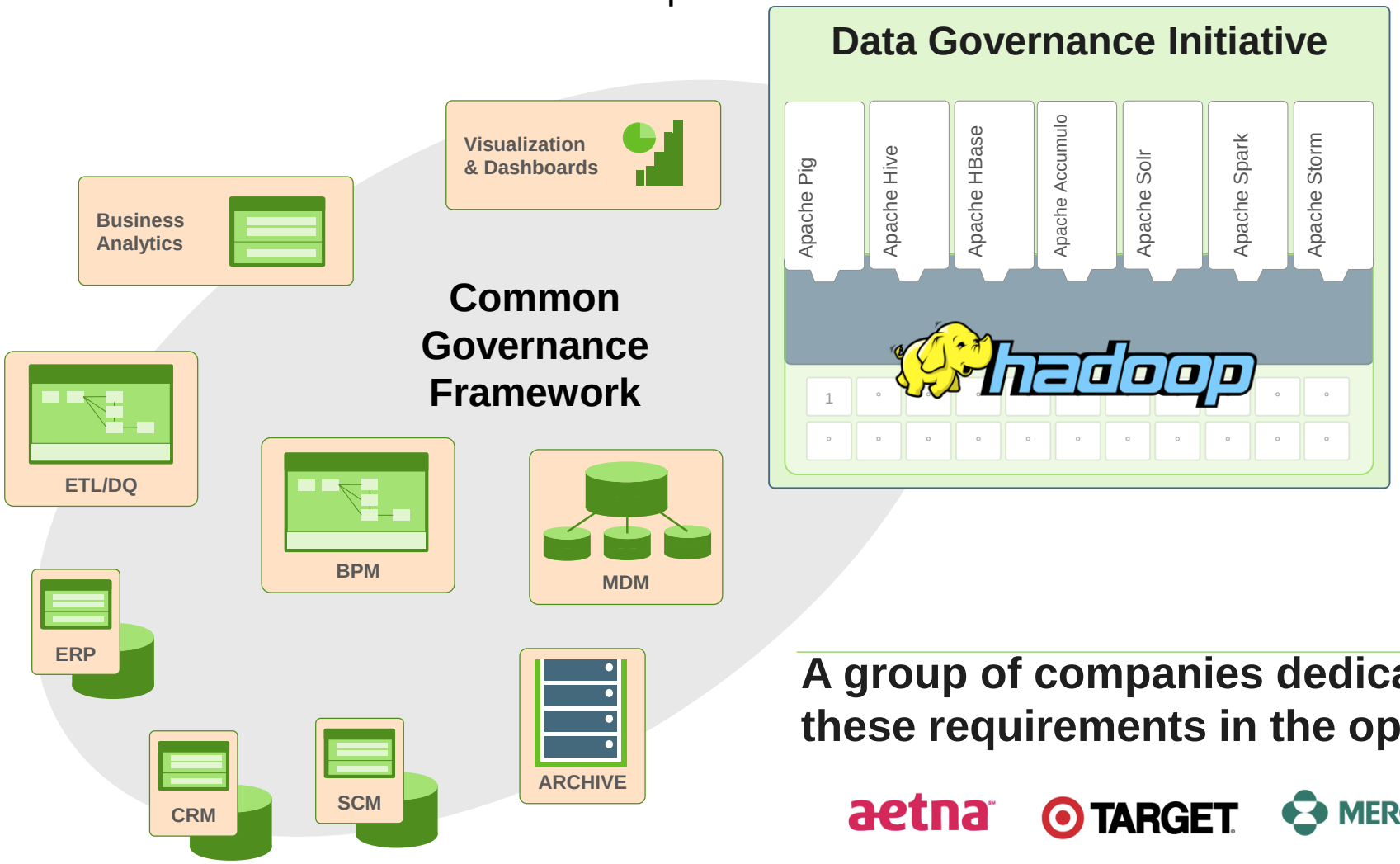


The ability to combine both traditional and core data with any and all edge data.
Secure workflow of any data type.
Bi-directional.
Actionable at a point in time – over a time series etc..
Matched against a long term Hadoop data store for ML and more accurate predictions.

Hortonworks®

4: Standards – Apache Atlas:

Data Governance Initiative for Hadoop



Requirements

1. Hadoop must snap in to the existing frameworks and be a good citizen
2. Hadoop must also provide governance within its own stack of technologies
3. **Metadata Services**
4. Deep Audit Store
5. Advanced Policy rule engine

A group of companies dedicated to meeting these requirements in the open

aetna TARGET MERCK Be well Hortonworks

SAP J.P.Morgan Schlumberger sas

About Hortonworks

Customer Momentum

- 550 customers (as of August 5, 2015)
- 119 customers added in Q2 2015

Hortonworks Data Platform

- Completely open multi-tenant platform for any app and any data
- Consistent enterprise services for security, operations, and governance

Partner for Customer Success

- Leader in open-source community, focused on innovation to meet enterprise needs

Hortonworks®

Getting started with Hadoop

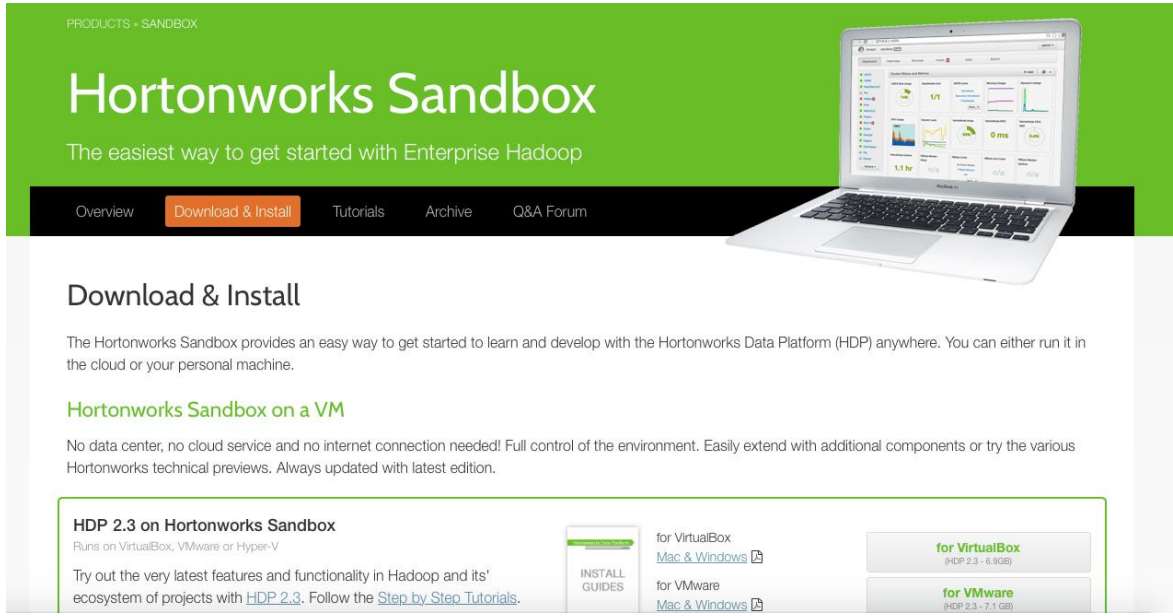
The HDP Sandbox 2.3

http://hortonworks.com/products/hortonworks-sandbox/#install

The full Hadoop distribution:
On your laptop in a virtual machine (no cost)
On Azure (no cost)
Tutorials and real world examples provided from Hortonworks, the Open source community and our partners.

The full HDP distribution (no cost):

http://hortonworks.com/hdp/downloads/



Thank You