
ORACLE®

Stay Informed During and After OpenWorld

Twitter: @OracleBigData, @OracleExadata, @Infrastructure
Follow #CloudReady

LinkedIn: Oracle IT Infrastructure– Oracle Showcase Page
Oracle Big Data – Oracle Showcase Page

October 1–5, 2017
SAN FRANCISCO, CA

Roadmap

Big Data: Cloud Service, Cloud Machine and Big Data Appliance

Jean-Pierre Dijcks
Product Management
Big Data
October, 2017


@jpdijcks

Safe Harbor Statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

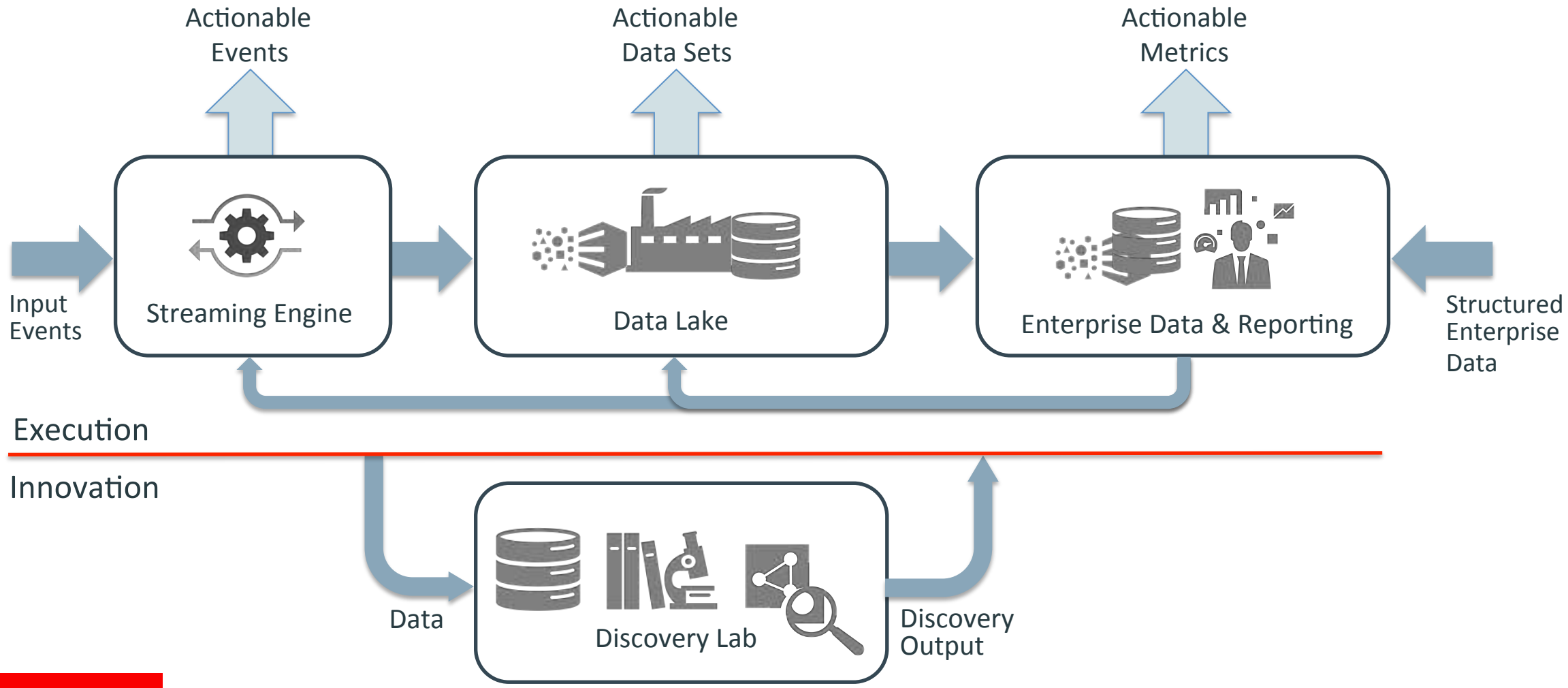
Program Agenda

1. Today
2. Roadmap Near Term
3. Roadmap Long Term

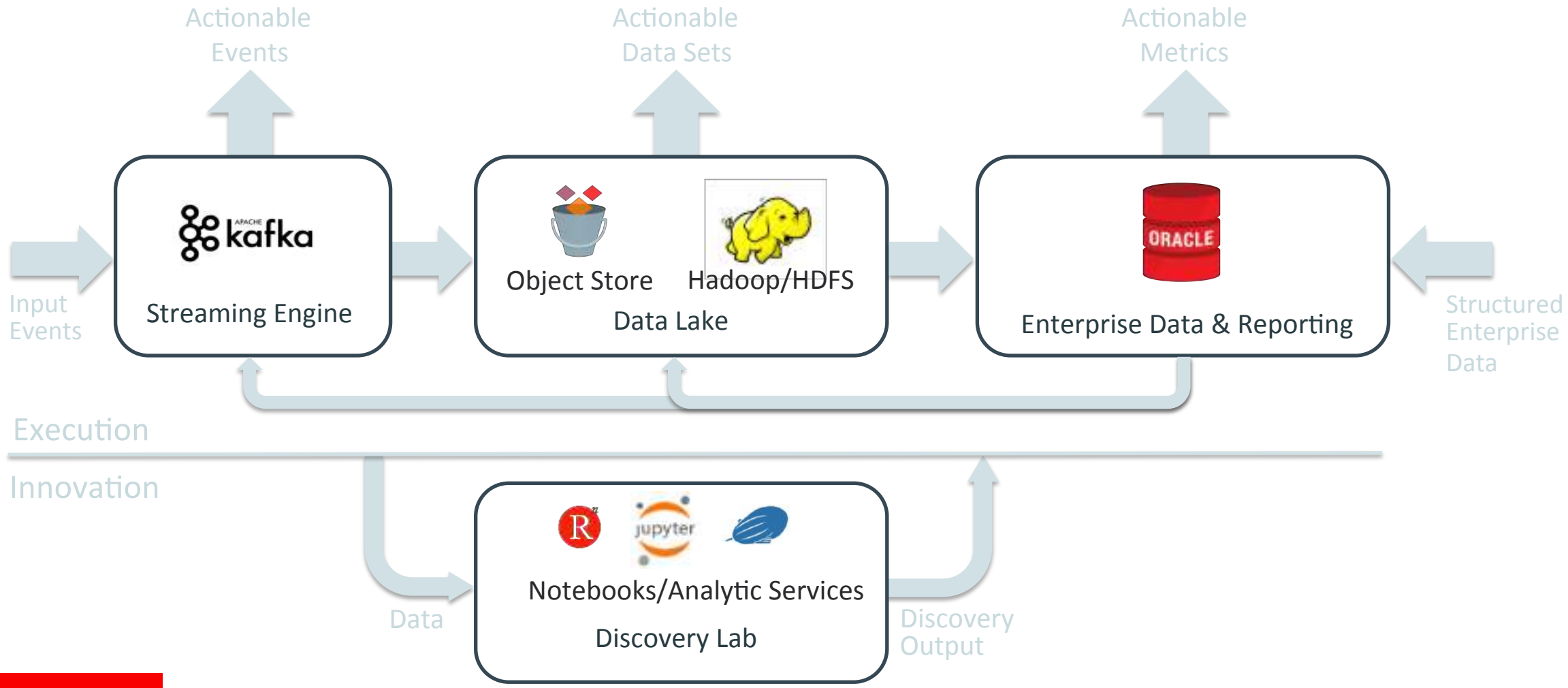
Program Agenda

1. Today
2. Roadmap Near Term
3. Roadmap Long Term

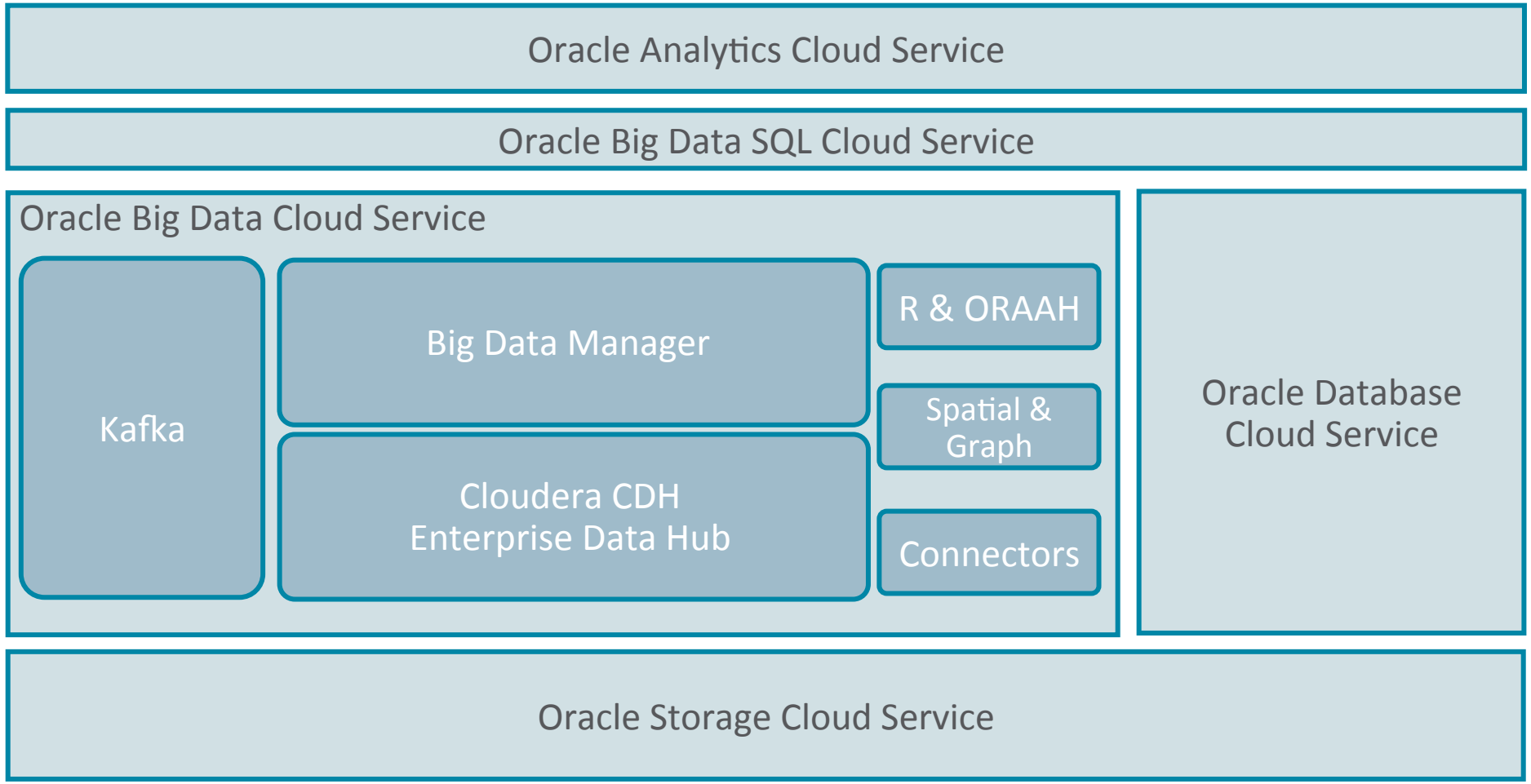
Information Management Reference Architecture



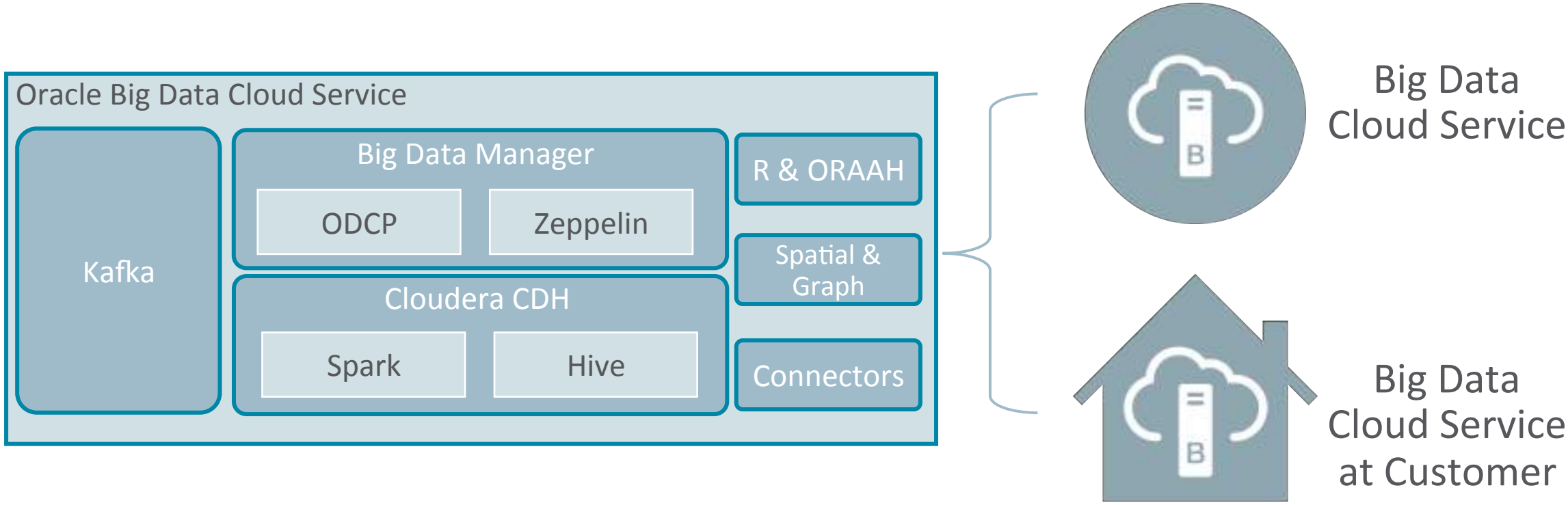
Implementation: Typical Components



Implementation: Oracle Cloud Services



Focus: Oracle Big Data Cloud Service

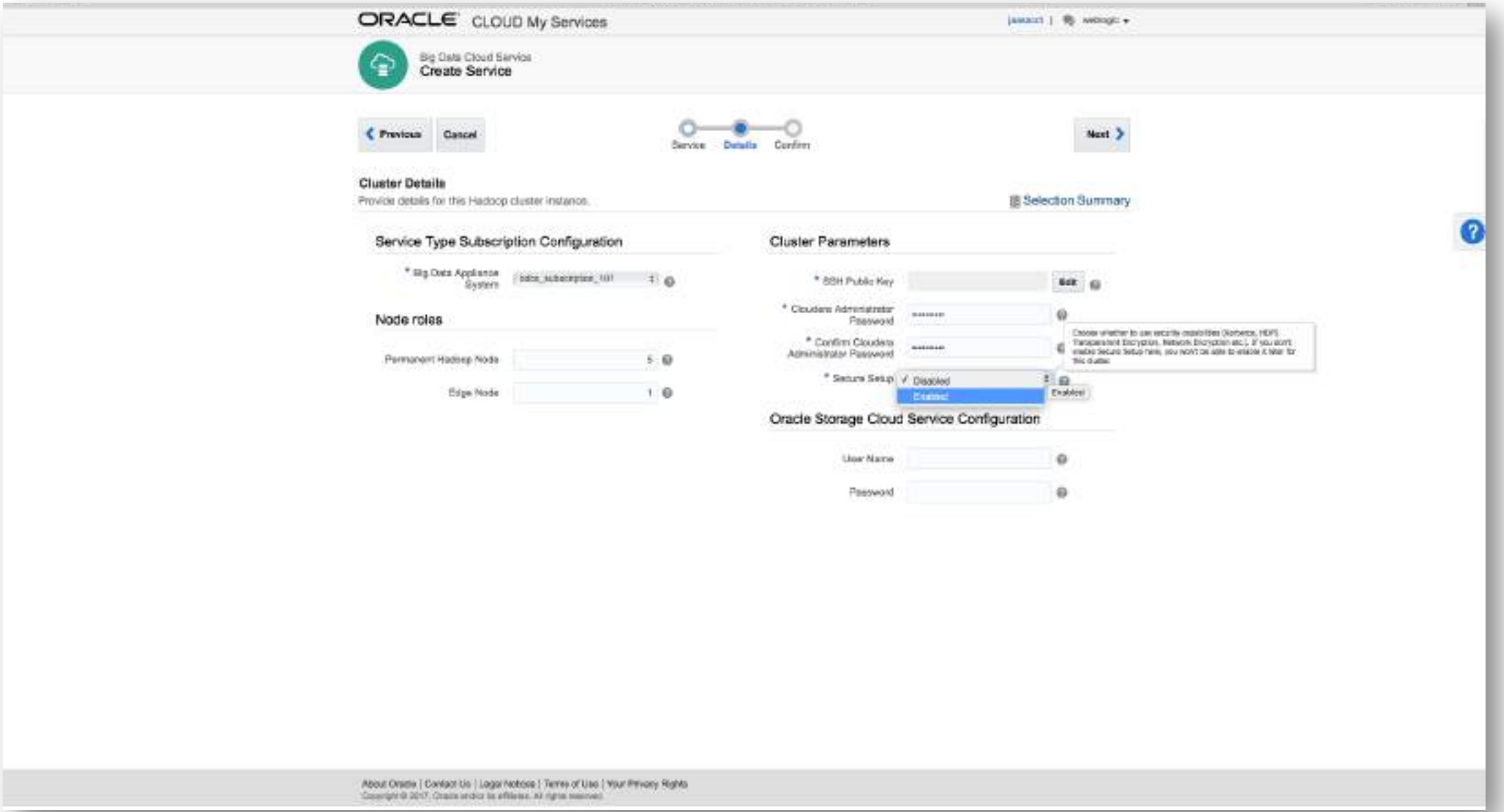


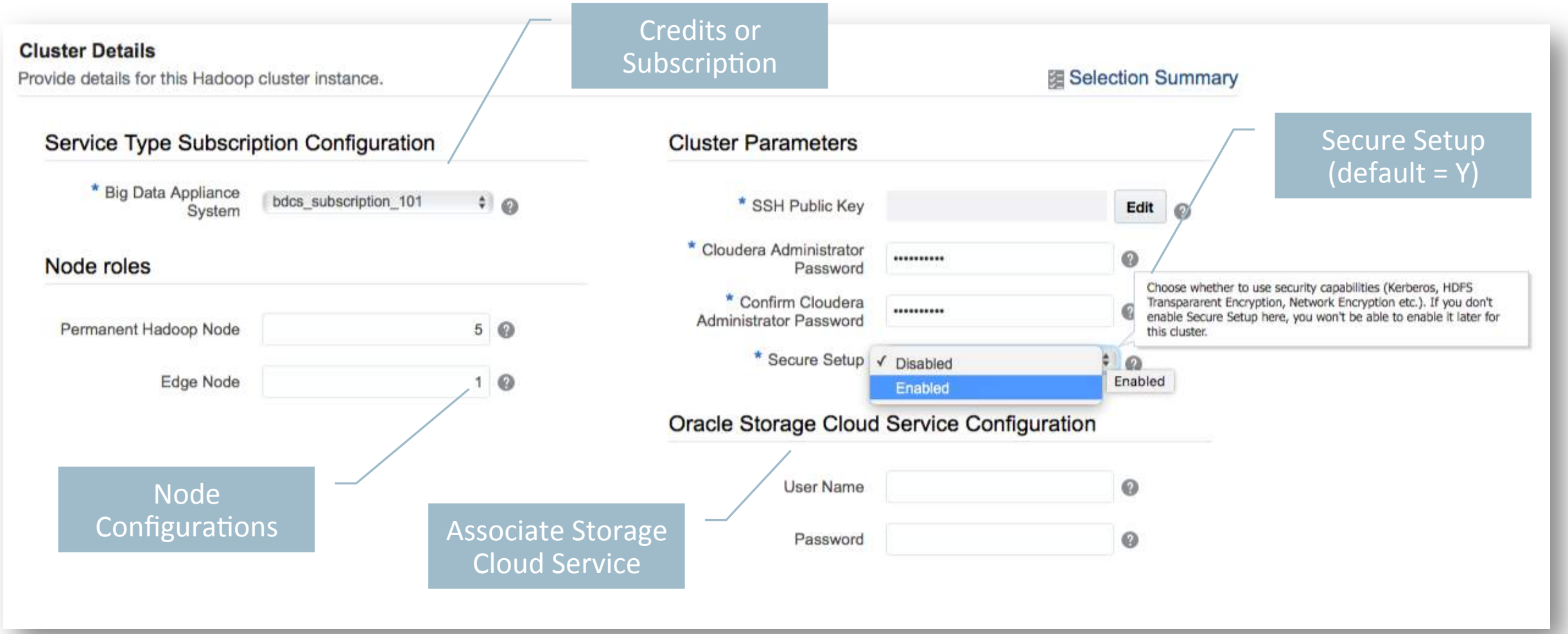
Definitions

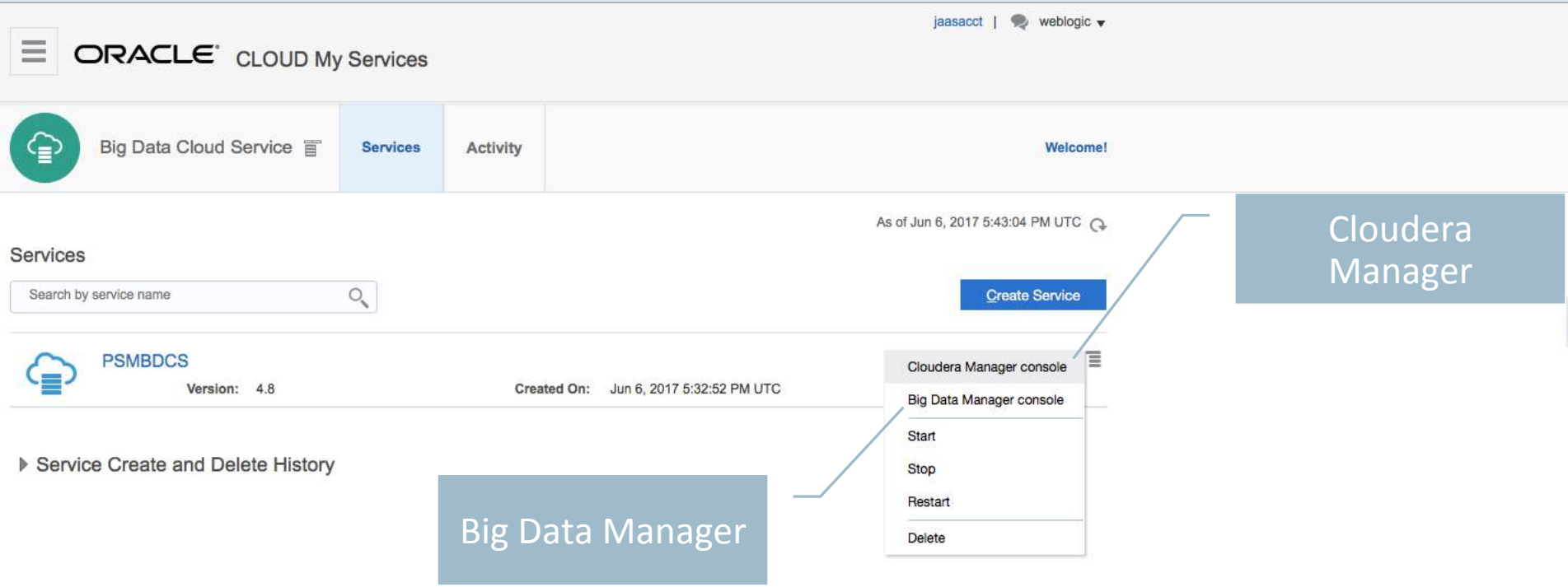
Big Data Cloud Service

Big Data Cloud Service at Customer

- Hadoop, Spark, Kafka and more, delivered as an integrated Cloud Service
  - Cloudera Enterprise – Data Hub Edition 5.x
  - Oracle Big Data Connectors
  - Oracle Big Data Spatial and Graph
  - Oracle Data Integrator Enterprise Edition
  - Notebook Integration
- Dedicated Compute Shapes with Direct Attached Storage and networking
  - Clusters start as small as 3 nodes and grow seamlessly
  - Burst/Shrink Compute when Required
  - Embedded Edge Nodes
  - Secure by Default







hdfs (BDCSDevTest) Actions

Today, 2:40 AM UTC

Status | Instances | Configuration | Commands | File Browser | Charts Library | Cache Statistics | Audits | Web UI | Quick Links

HDFS Summary

Configured Capacity	324.9 GiB/148.1 TiB

Health Tests

Create Trigger

Show 8 Good

Status Summary

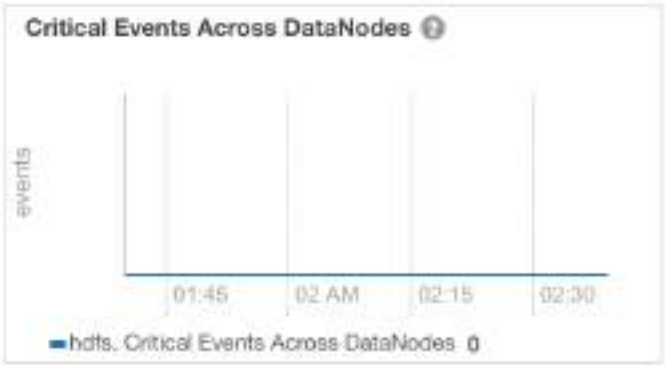
Balancer	1 None
DataNode	4 Good Health
Failover Controller	2 Good Health
Gateway	1 None
HttpFS	1 Good Health
JournalNode	3 Good Health
NameNode	2 Good Health
Hosts	4 Good Health

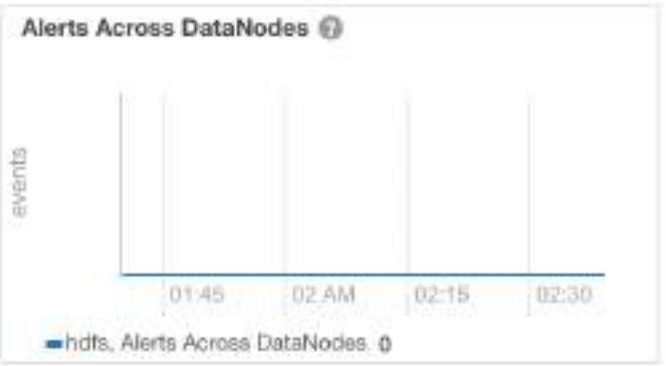
Health History

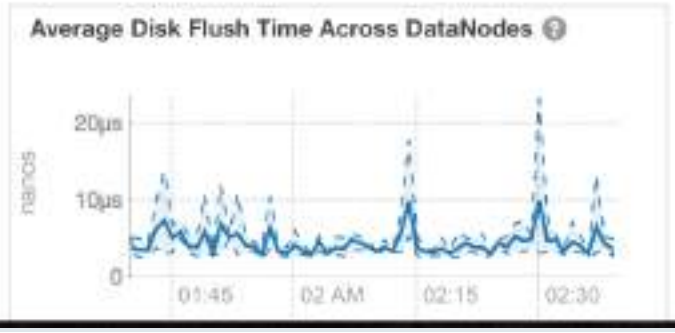
Display a menu

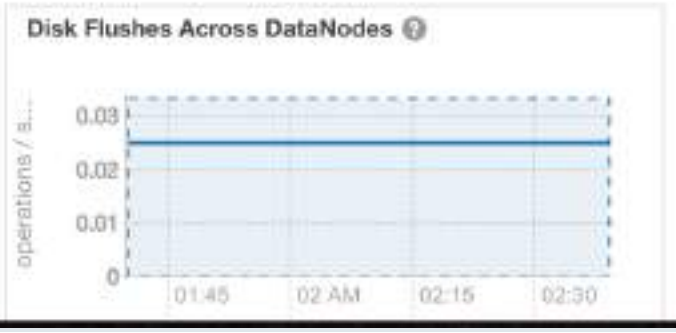
Charts

30m 1h 2h 6h 12h 1d 7d 30d









Oracle | Integrated Cloud Ap | Big Data Cloud Service Details | Big Data Manager

Search

ORACLE

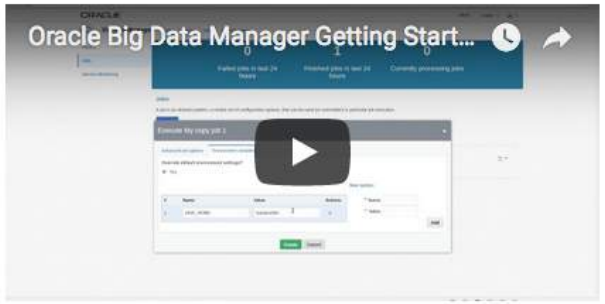
a422306 | bigdatamg

Big Data Manager

Home | Data | Jobs | Monitoring | Administration | Dashboard | Notebook

New to Big Data Manager?

Go to the Getting Started Guide to learn more



Getting Started Guide

Jobs

View your jobs

Jobs

Data explorer

Explorer and manage data in your storage

Data Explorer

Monitoring

Status of HDFS, Yarn, cluster etc.

Cluster Monitoring

- Healthy nodes: 5 of 5

Service Monitoring

- Yarn
  OK
- WebHDFS
  OK
- Notebook
  OK

Monitoring

About Oracle | Contact Us | Legal Notices | Terms of Use | Your Privacy Rights

What did you just do?

- Allocated a set of nodes from the pool
- Configured the IB network to ensure secure tenancy
- Created VMs to host the Hadoop cluster
  - Enabled 32 OCPUs
  - Delivery 48TB of Disk
  - And 256GB of memory
- Created a CDH cluster in a standard and certified manner
- Created the HDFS file system, configured Yarn, Spark
- Enabled HA for CDH components like NameNodes, Zookeeper, etc.
- Applied latest Best Practices for Cloudera Set up
- Enabled Kerberos on the cluster for all services and integrated with IDCS*
- Set up MySQL in Active-Passive mode for Hive metastore
- Configured Key Trustee Server in HA mode for HDFS encryption
- Set up Sentry for SQL access security
- Configured the client network for client access
  - Enabled Hue, Cloudera Manager
- Set up the edge nodes if selected in cluster creation
- Connected to Storage Cloud to enable ODCP for data copy back and forth
- Configured Big Data Manager
  - Zeppelin Notebooks and Interpreters
- Configured R and ORAAH*
- And a bunch of stuff we casually forgot about…

A few other things to mention

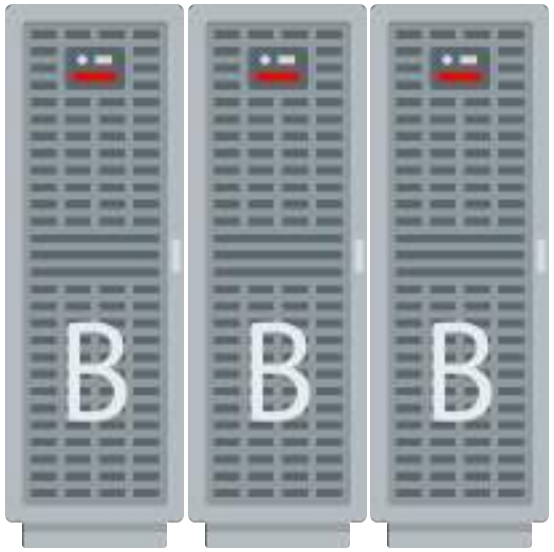
- BDCS is running in Production for a large set of customers
  - Now also available under the new Universal Credits Model
- BDCS at Customer (BDCM) is Generally Available
  - Certification with the latest Control Plane software is ongoing

The following roadmap items apply to both offerings!

Program Agenda

1. Today
2. Roadmap Near Term
3. Roadmap Long Term

Big Data Appliance



- Edge Nodes
  - Linux 7 Edge Nodes for CDSW ✓
  - Fully Managed Edge Nodes
- Full Clusters on Linux 7
- Pre-configured Kafka Clusters
- Big Data Manager on BDA
  - Full UI Capabilities
  - Integration with Notebooks etc.
- Non-stop Clusters
  - HA for all components

Big Data Cloud Service Design Goals

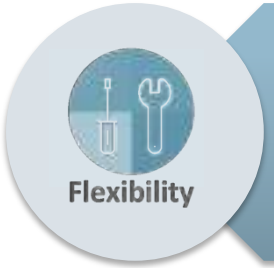
Performance

Security

Flexibility

Data Science

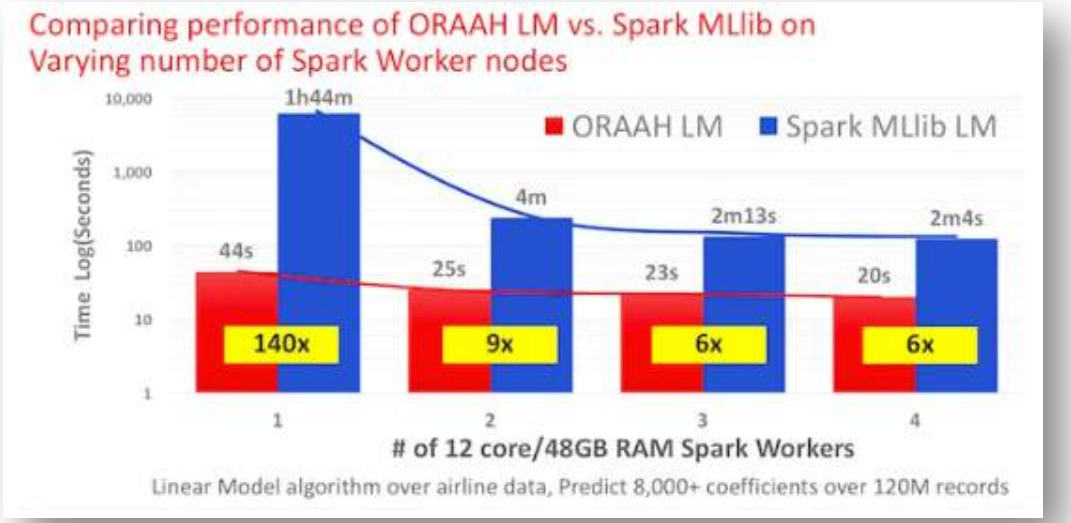
Configure R & ORAAH



Integrate Apache Zeppelin

Enable Cloudera Data Science Workbench

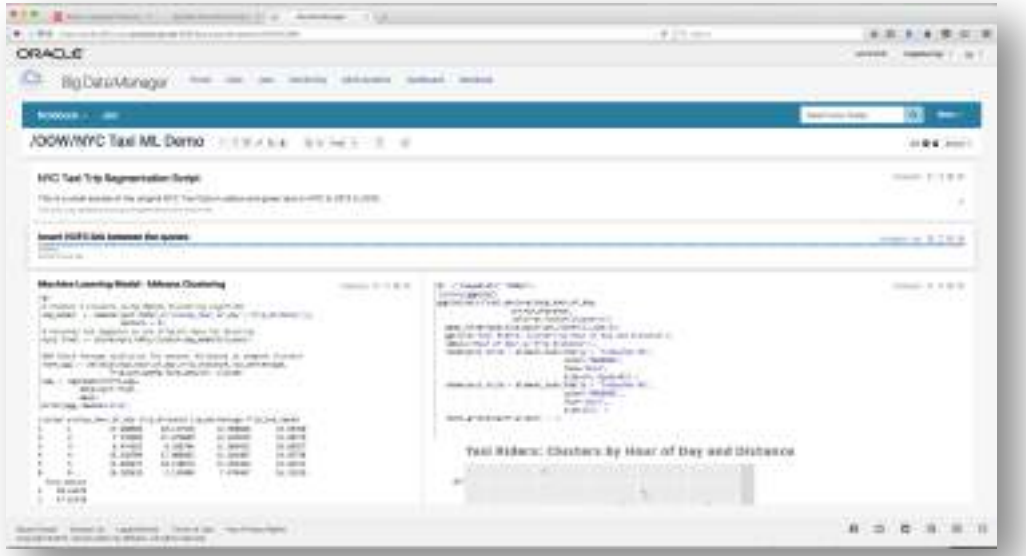
Data Science: Configure and Enable ORAAH



- Configure BDCS and BDCM clusters to default to using ORAAH for R Analytics
- Connect up to Zeppelin and other environments
- Enable dramatic performance improvements for Analytics out of box

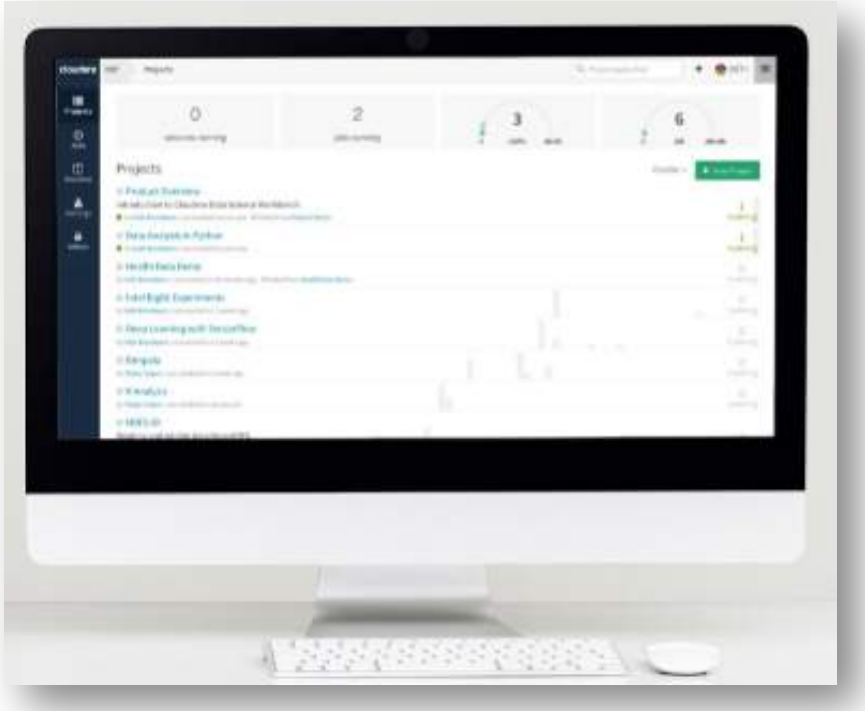
http://www.oracle.com/technetwork/database/database-technologies/bdc/r-advanalytics-for-hadoop/overview/index.html

Data Science: Integrate Apache Zeppelin



- Embed Apache Zeppelin into Big Data Manager ✓
- Enable simple mechanics to use interpreters, notebooks and files ✓
- Configure R for easy usage in notebooks
- Initially:
  - Non-secure clusters ✓
  - Secure clusters with IDCS and User Management

Data Science: Enable Cloudera CDSW on BDCS



- Cloudera Data Science Workbench requires very specific OS versions for its nodes
- Enable configurable Edge Nodes in BDCS to accommodate this image
- Enables use of CDSW within BDCS and BDCM

Note: Separately licensed component

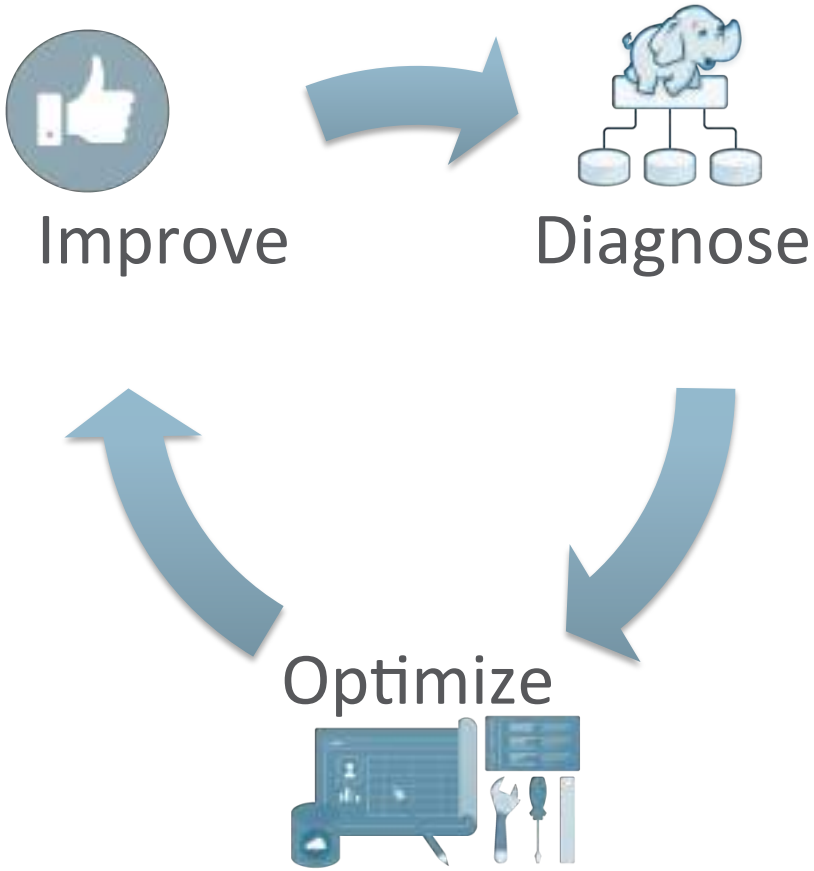
Service Management

Pro-Active Health Checks

Auto Node Failover

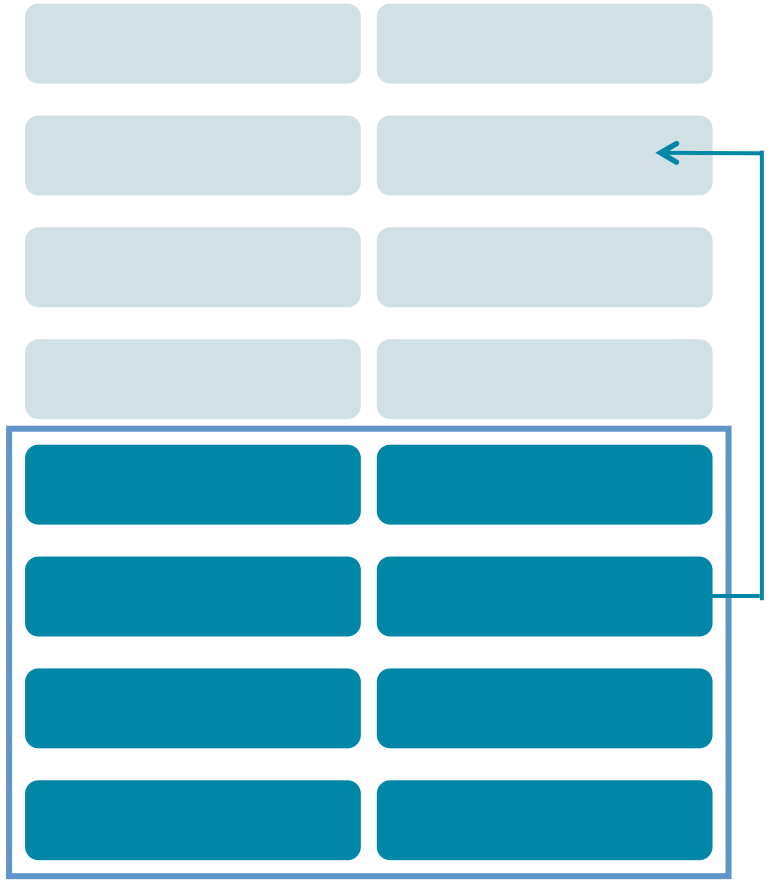
IDCS Integration

Pro-Active Service Health Checks



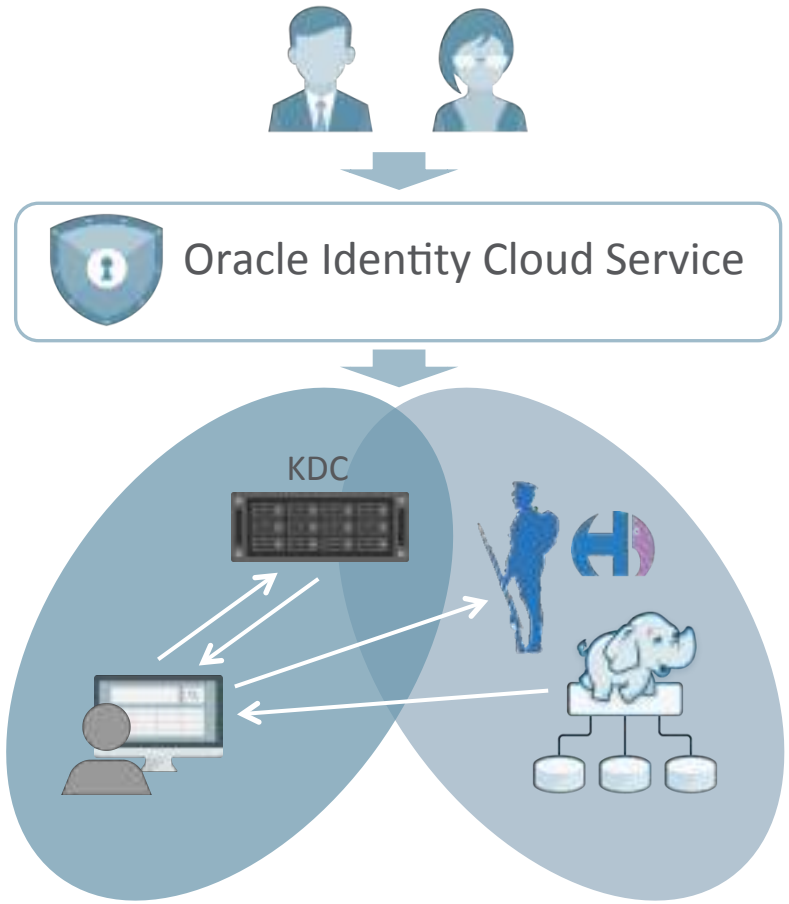
- BDCS encapsulates a fast moving ecosystem
- Pro-Active Cluster Health Checks ensure:
  - Optimal configurations of clusters to underlying compute shapes
  - Better security by pro-active patch recommendations
  - Specific settings for your workloads
  - Updates for the latest Cloudera settings and best practices

Auto Node Failover



- Because of Performance and Security, BDCS runs direct attached storage
- Auto Node Failover ensures:
  - Always the same compute capacity dedicated to jobs
  - Lights-out management
  - Much higher cluster resiliency

ICDS Integration & Firewall Setup

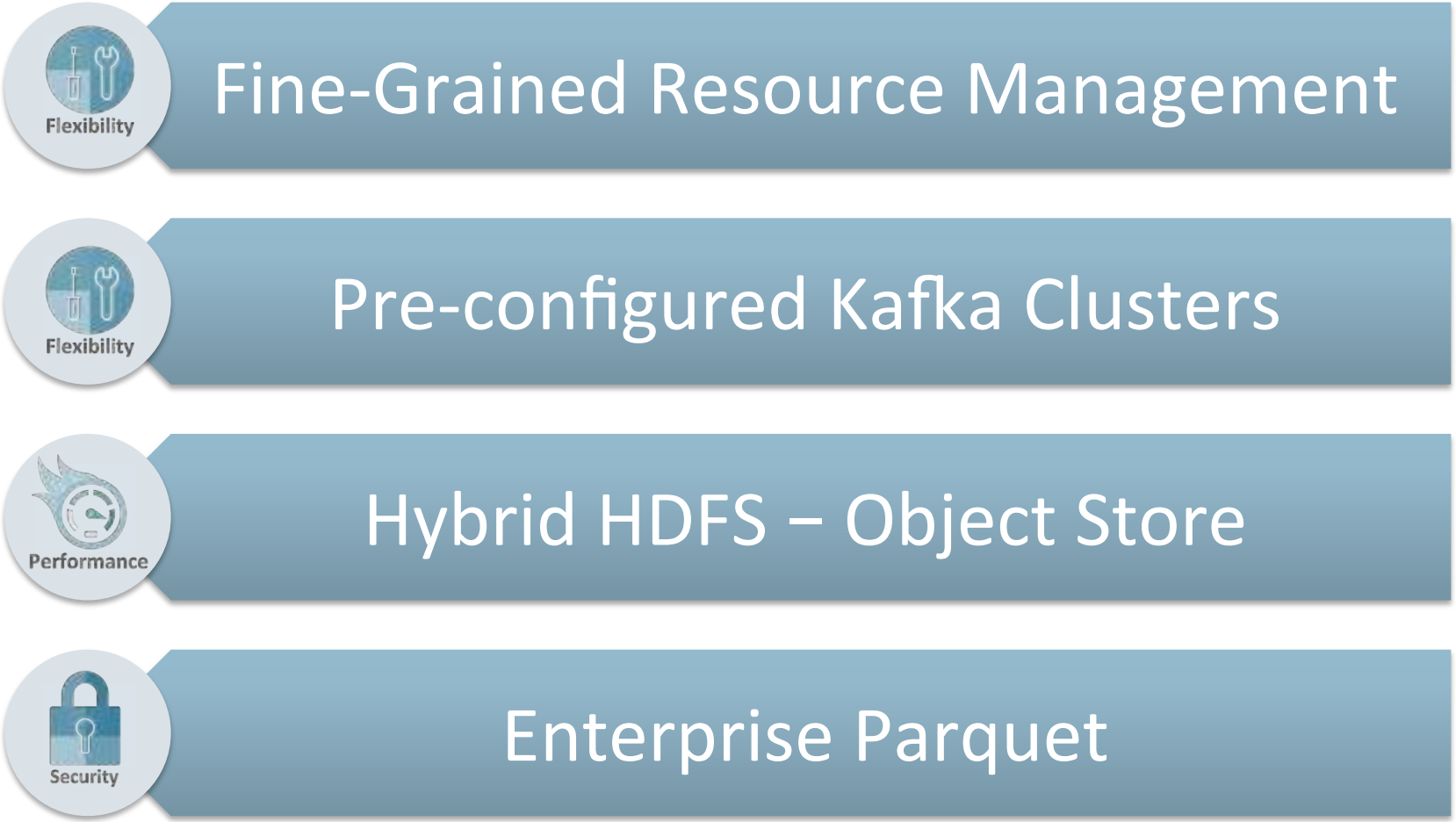


- Secure Hadoop is comprised of many components:
  - This leads to many manual connections
- IDSC integration delivers:
  - Single Oracle Cloud Identity
  - Integrated with Kerberos and other AAA security features
  - User portability across services
- Firewall Configuration enables self-service port management

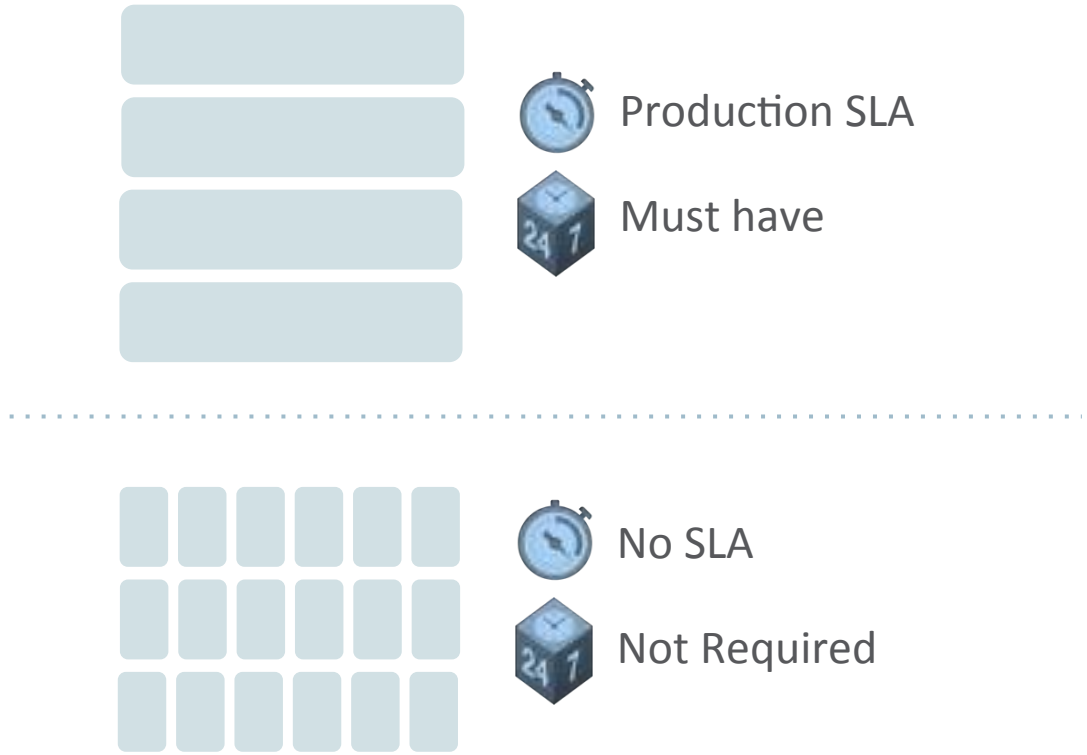
Program Agenda

1. Today
2. Roadmap Near Term
3. Roadmap Long Term

Long Term Roadmap Work

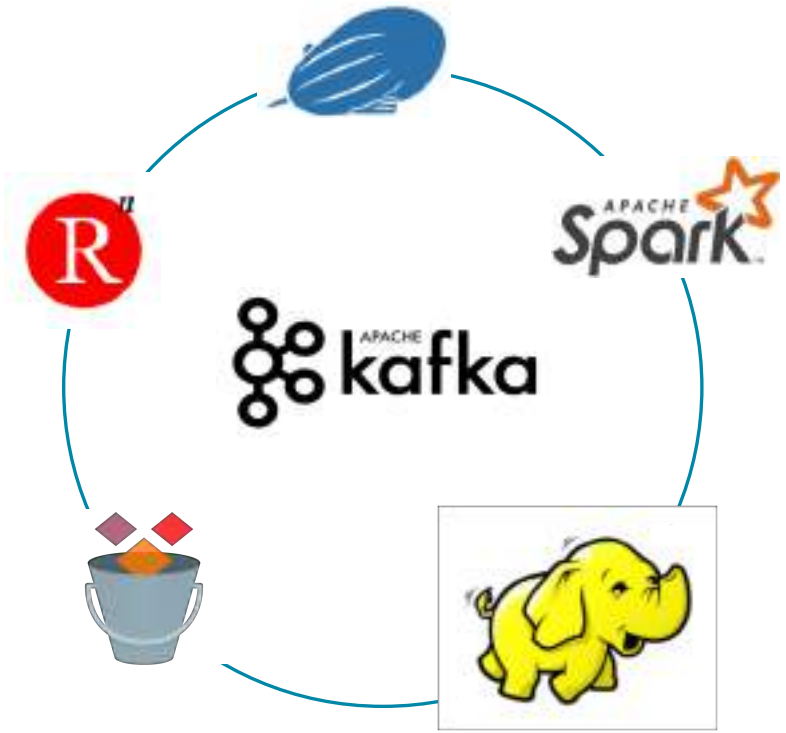


Fine-Grained Resource Management



- Enable fine-grained resources and billing
- Choices will range from:
  - Small VM based systems with multiple nodes
  - Dedicated Cluster Nodes for performance sensitive workloads
  - And anything in between
- Clusters will consist of multiple (small) nodes
- Customers can run Cloudera CDH as Dev/Test/ Prod systems in the same infrastructure with the same version
- Customers can grow or shrink systems to fit their need

Pre-Configured Kafka Clusters

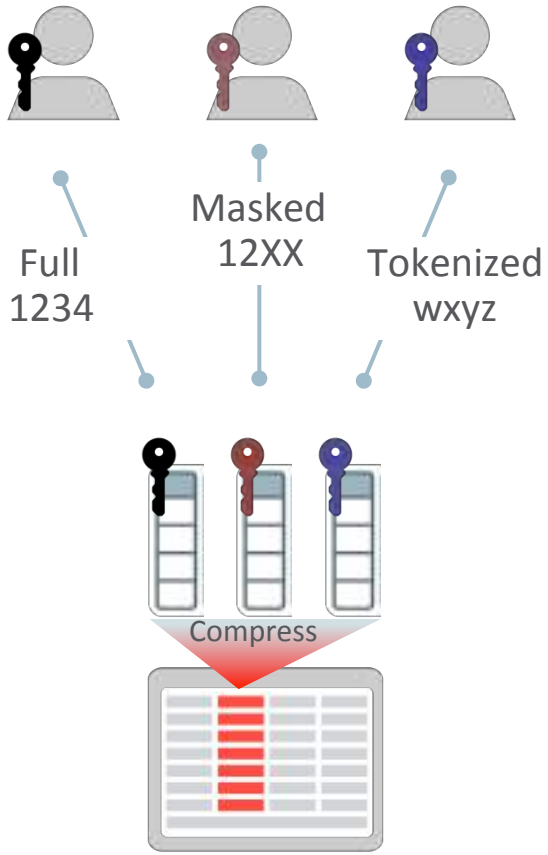


- Fully Supported: Create and configure Kafka on BDCS today
- Automate Kafka on BDCS as a configuration choice, delivering:
  - Either a small cluster sharing hardware with HDFS
  - A full blown multi-node cluster for scale out workloads
  - A multi-VM system with isolation
- All pre-configured for the platform
- All licensed and supported in the cluster

Hybrid HDFS – Object Store

- Object Stores increasingly replace HDFS
- But, not all workloads will run well on that **today**
- BDCS will enable Hybrid HDFS – Object Store for data in the service:
  - Services like Big Data SQL will query data either in Object Stores, or in HDFS
  - System will lazily load for better performance
  - Customer can pin for better security
- Get best of both worlds, integrated

“Enterprise Parquet”



- As file stores become increasingly important for analytics, security poses an especially important topic
- Enterprise Parquet solves:
  - The issues around data security by enabling multiple column versions (redacted, tokenized) while not increasing file sizes
  - Performance issues around Schema on Read
  - File proliferation through dynamic query matching
- Decreases security risks, cost and maintenance

$ sqlplus scott/tiger

SQL> SELECT ssn, fname FROM person_ep;

SSN	FNAME_VAL
'***-**-****‘	'Sheldon'
'***-**-****‘	'Cedrick'
'***-**-****‘	'Carolyne‘
'***-**-****‘	'Santiago‘
'***-**-****‘	'Lora‘
'***-**-****‘	'Kaden‘
'***-**-****‘	'Imelda‘
'***-**-****‘	'Maximillia‘
'***-**-****‘	'Dorothea‘

$ sqlplus epdemoUser/privilegedPassword

SQL> SELECT ssn, fname FROM person_ep;

SSN	FNAME_VAL
'574-81-3280'	'Sheldon'
'496-03-4793'	'Cedrick'
'365-75-6602‘	'Carolyne'
'560-88-9346‘	'Santiago'
'596-90-4843‘	'Lora'
'754-88-2912‘	'Kaden'
'572-06-4933‘	'Imelda‘
'623-95-0295‘	'Maximillia'
'439-27-0802‘	'Dorothea’

And one more thing…

Autonomous or managed “Hadoop” is being planned…

Follow this Space!!

Integrated Cloud

Applications & Platform Services


ORACLE®