Oracle Big Data SQL: Roadmap to High-Performance Data Virtualization

Marty Gubar
Director - Oracle Big Data PM

Gaurav Singh
Energy Australia
Big Data & Data Warehouse Solution Architect

October 3, 2017

ORACLE

Copyright © 2017, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

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.

Copyright © 2017, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

Agenda

1. Introduction
2. High Value Transformation at Energy Australia
3. Big Data SQL Today and Development Themes

Copyright © 2017, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

Big Data SQL Goals

Easily access any data across big data stores

Provides a unified security model across the sources

Analyze all data using Oracle's rich SQL dialect

Fast performance using Big Data SQL Smart Scan

Copyright © 2017, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

Big Data SQL

Gaurav Singh
Big Data & Data Warehouse Solution Architect
Information & Analytics

EnergyAustralia
LIGHT THE WAY

Introducing EnergyAustralia

Energy Generation

Owns and operates a multibillion dollar portfolio of Energy Generation Assets.

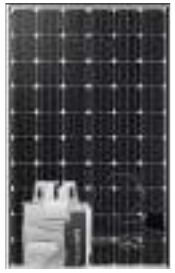
Has made an investment of $1.5 billion in Renewables.

- Yallourn Power Station
- Hallett Power Station
- Mt Piper Power Stations
- Tallawarra Power Station
- Pine Dale Mine

- Cathedral Rocks Wind Farm
- Waterloo Wind Farm
- Stony Gap Wind Farm Project
- Mallee Solar Park Project

Energy Retailing

Is one of Australia's largest and most experienced energy retailers, servicing over 1.7 million households and businesses across Australia.



EnergyAustralia
LIGHT THE WAY

Business Capabilities

…utilising Oracle Big Data Solution

High-Value capabilities delivered

A query that took **23 seconds** on a legacy platform takes **0.08** seconds on the new platform

Increased speed of query performance

New Smart Meter Analytics Capabilities

Oracle R Advanced Analytics for Hadoop

Revenue assurance improvements

Bringing data to life

Strategic BI and Analytical tools will translate data into visual dashboards that tell a compelling story

Golden Record of Customer & Premise

ORACLE FUSION MIDDLEWARE GOLDENGATE

ORACLE Enterprise Data Quality

Single View of Agent

ORACLE DATA INTEGRATOR

ORACLE DATABASE 12c

550+

Enterprise definitions endorsed

ORACLE FUSION MIDDLEWARE METADATA MANAGEMENT 12c

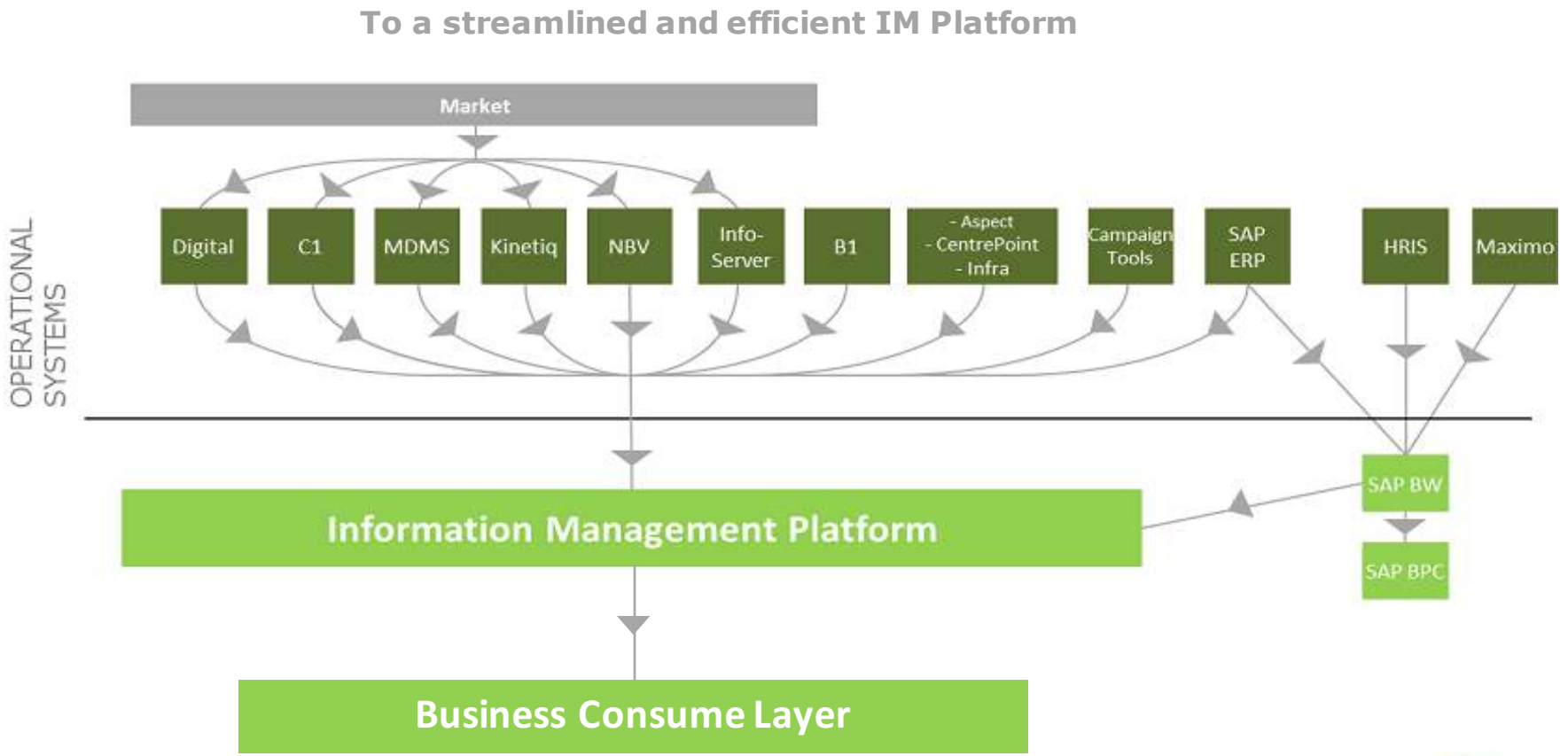
Across these domains…

- Customer
- Customer Service & Agent
- Financial Performance
- Meter
- Onboarding & Billing
- Payments, Credit & Collections
- Product & Pricing
- Sales & Offer

Past Methodology

…utilising complex interesting custom built solutions

Legacy analytics eco-system

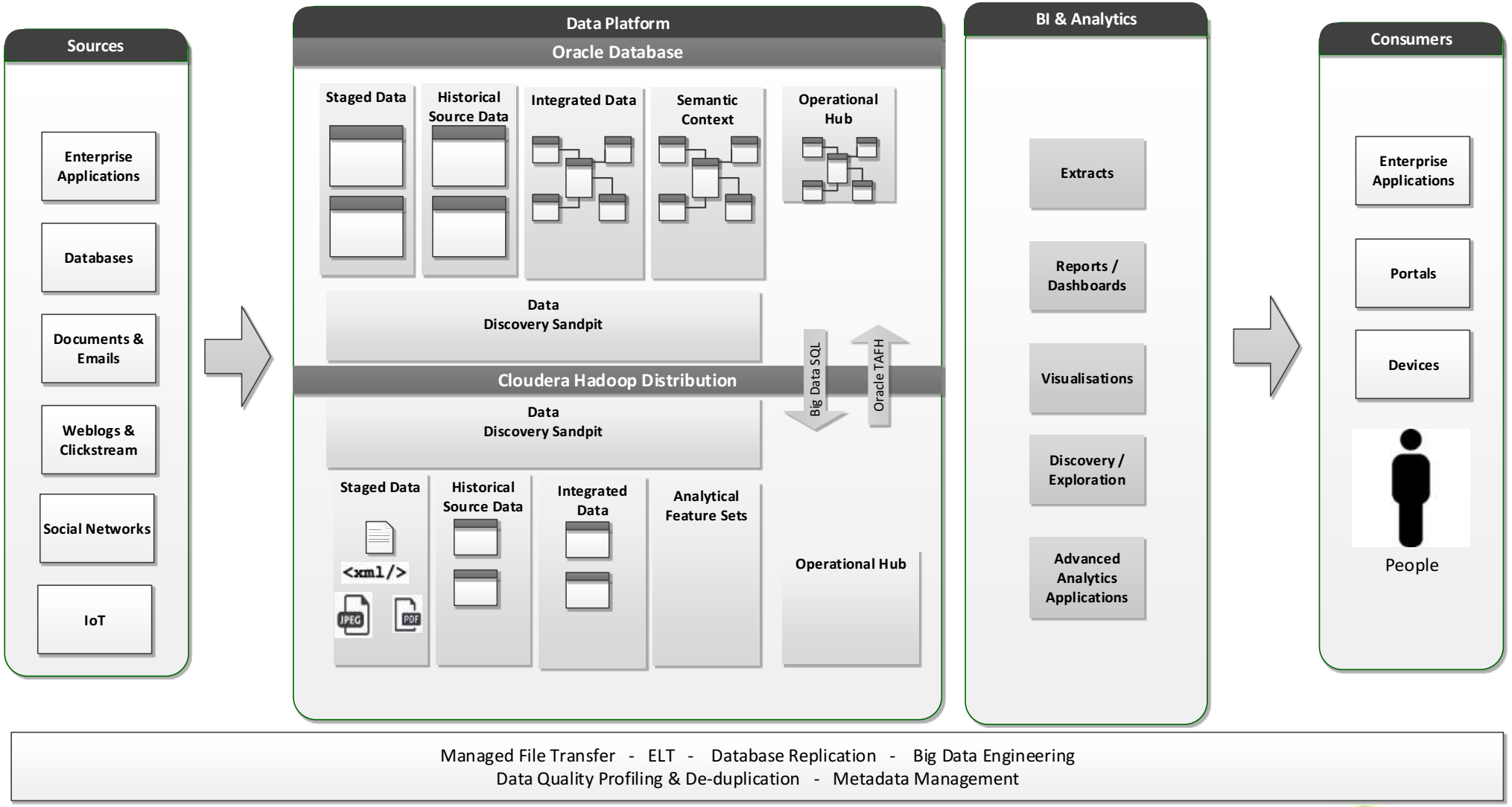


EnergyAustralia
LIGHT THE WAY

New Methodology

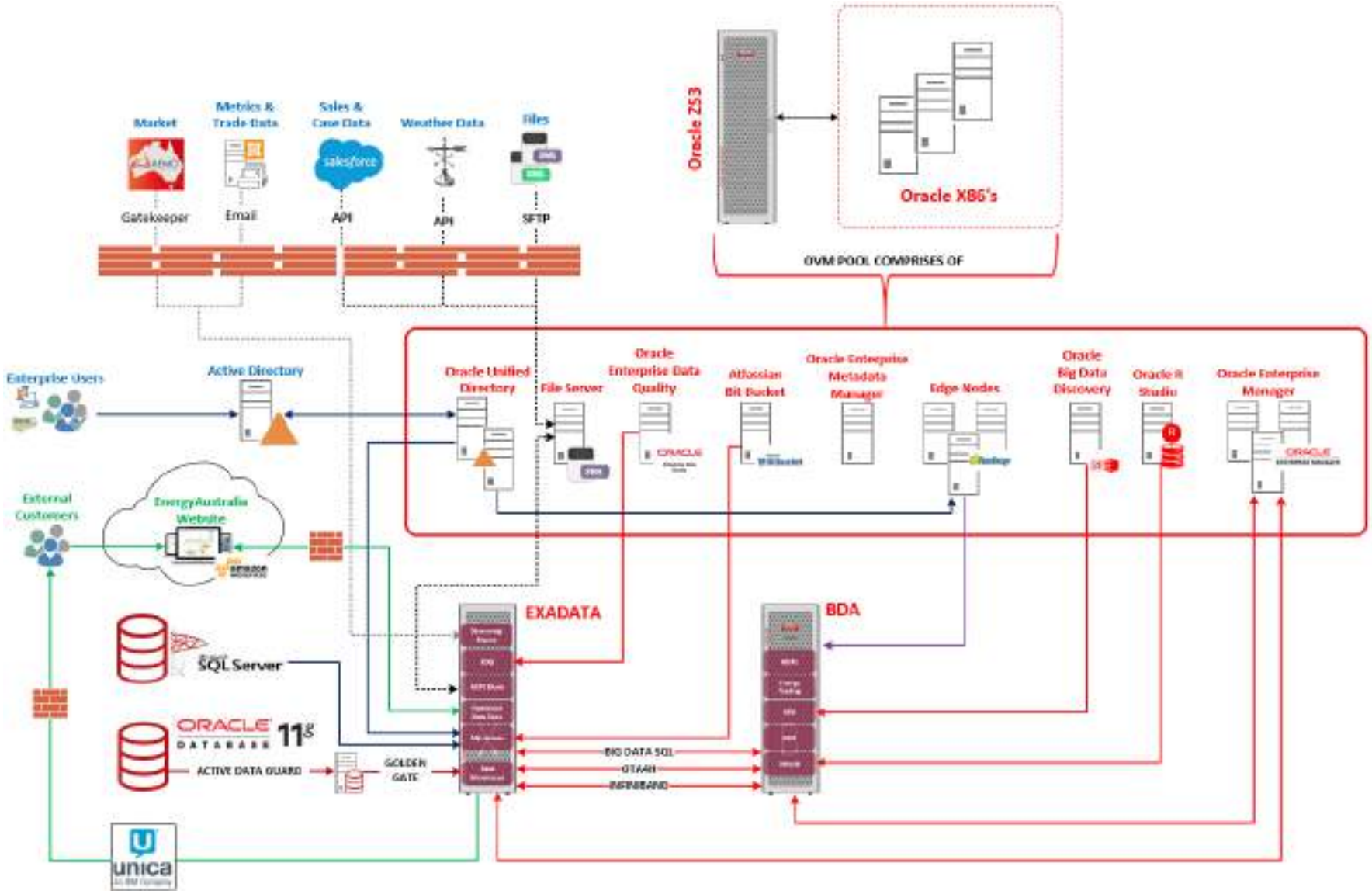
…utilising Oracle Big Data Solution

Solution Architecture



EnergyAustralia
LIGHT THE WAY

Technical Architecture

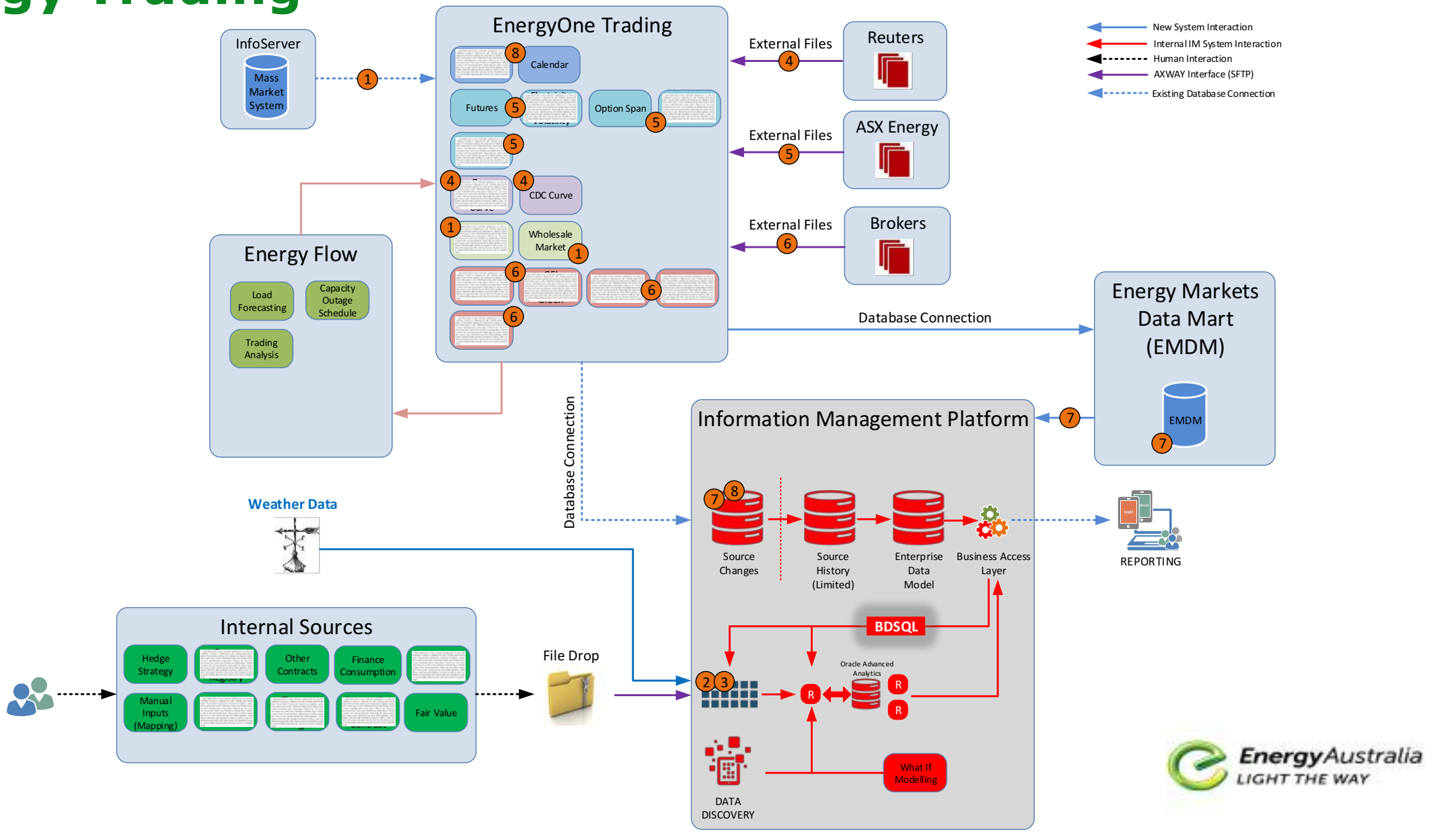


EnergyAustralia
LIGHT THE WAY

Business Enablement

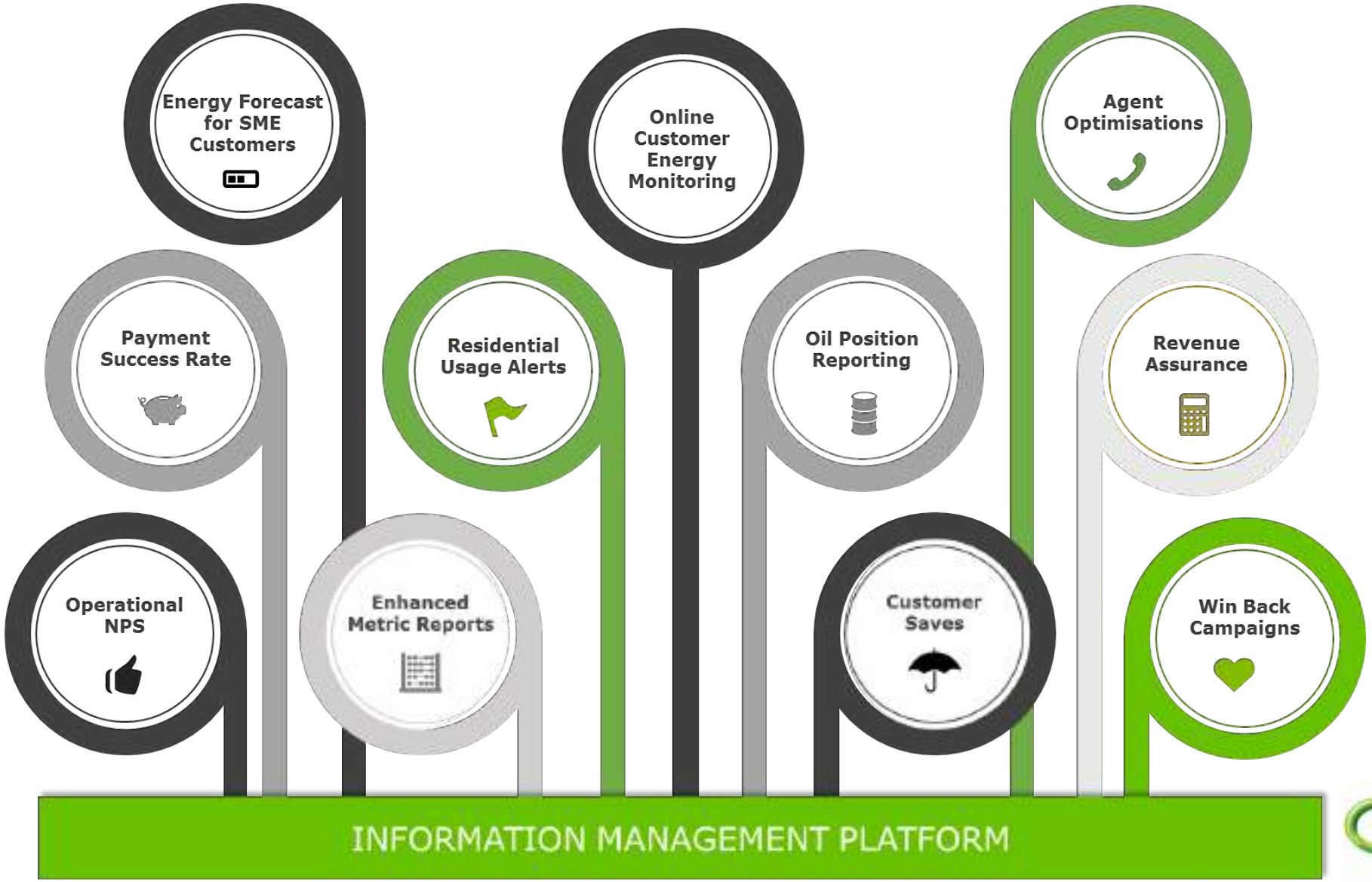
…through Big Data SQL

Energy Trading



EnergyAustralia
LIGHT THE WAY

Major business initiatives delivered



EnergyAustralia
LIGHT THE WAY

Learnings

- **Big Data SQL** allows for seamless access to underlying data sets, **regardless** of source
- A **heterogenous BI ecosystem** provides the **flexibility** to complement an integrated data platform utilising **BDSQL**
- **Connect and Analyse** data and make it available to a growing pool of **business users** who do not need to learn how to Hadoop, but still **utilise the power** of Hadoop through **BDSQL**
- Running **advanced analytics in-place** helps further leverage the investment in engineered systems and makes the use case of utilising a Big Data solution **viable and beneficial** to the organisation
- **Return on investment** by a factor of **15 times**

EnergyAustralia
LIGHT THE WAY

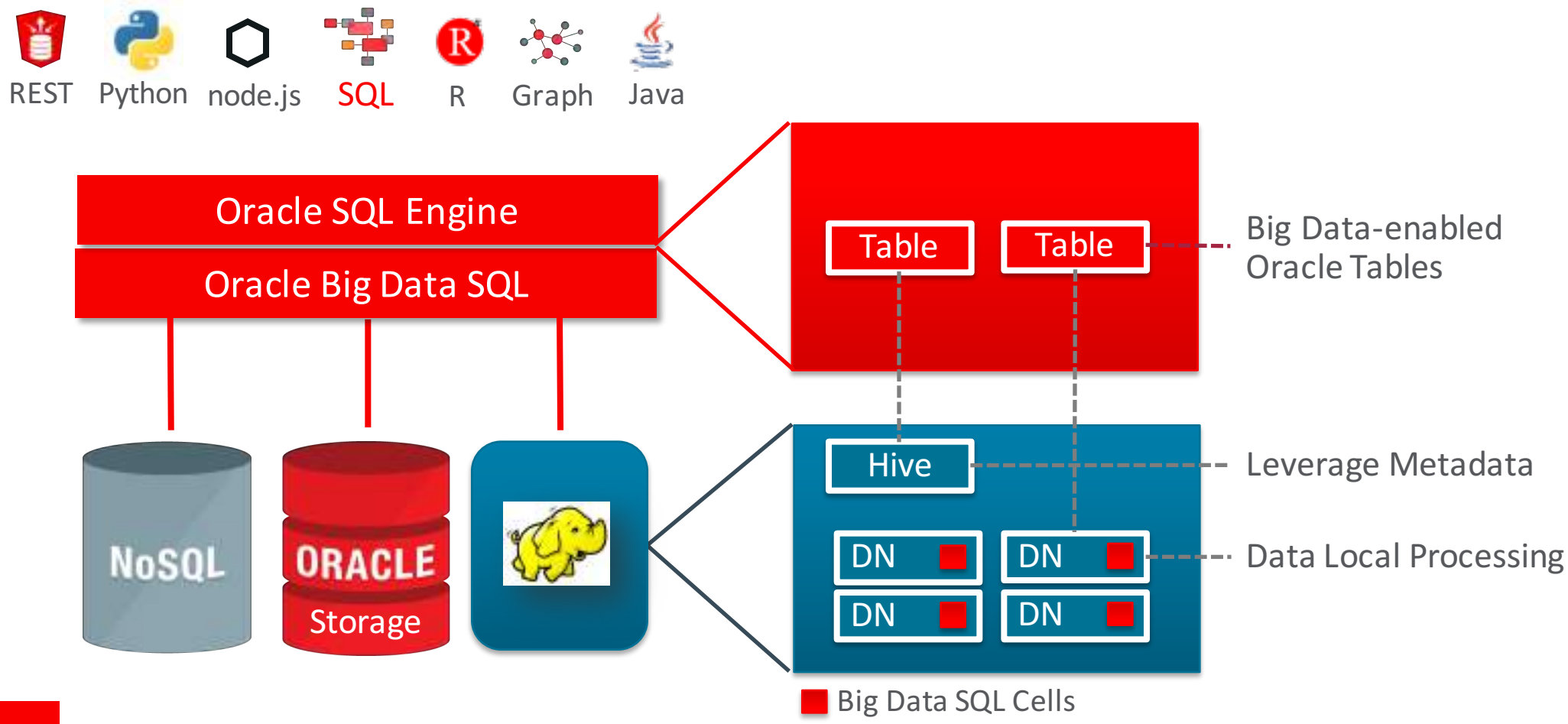
Big Data SQL Today

Architecture and Key Features

ORACLE

Copyright © 2017, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

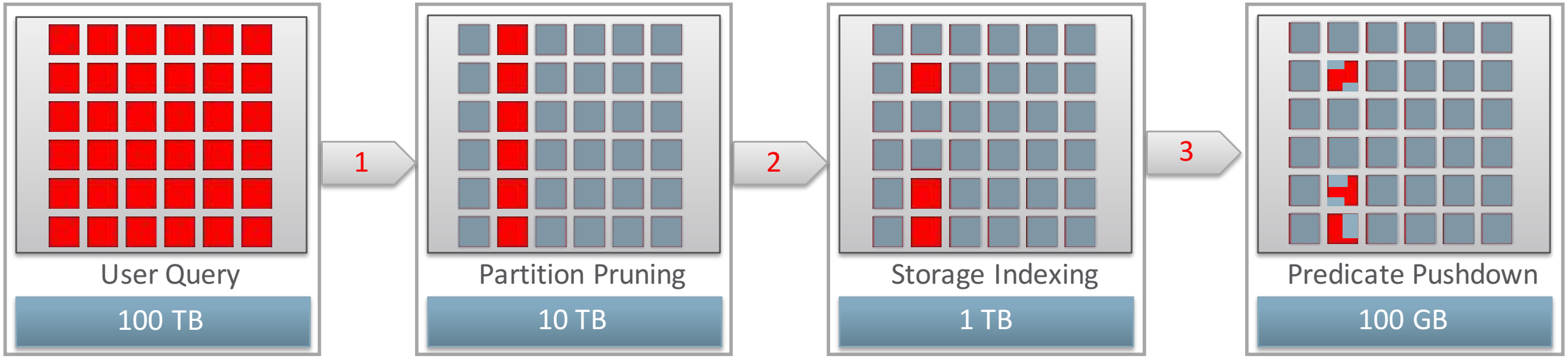
Big Data SQL Architecture



Copyright © 2017, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

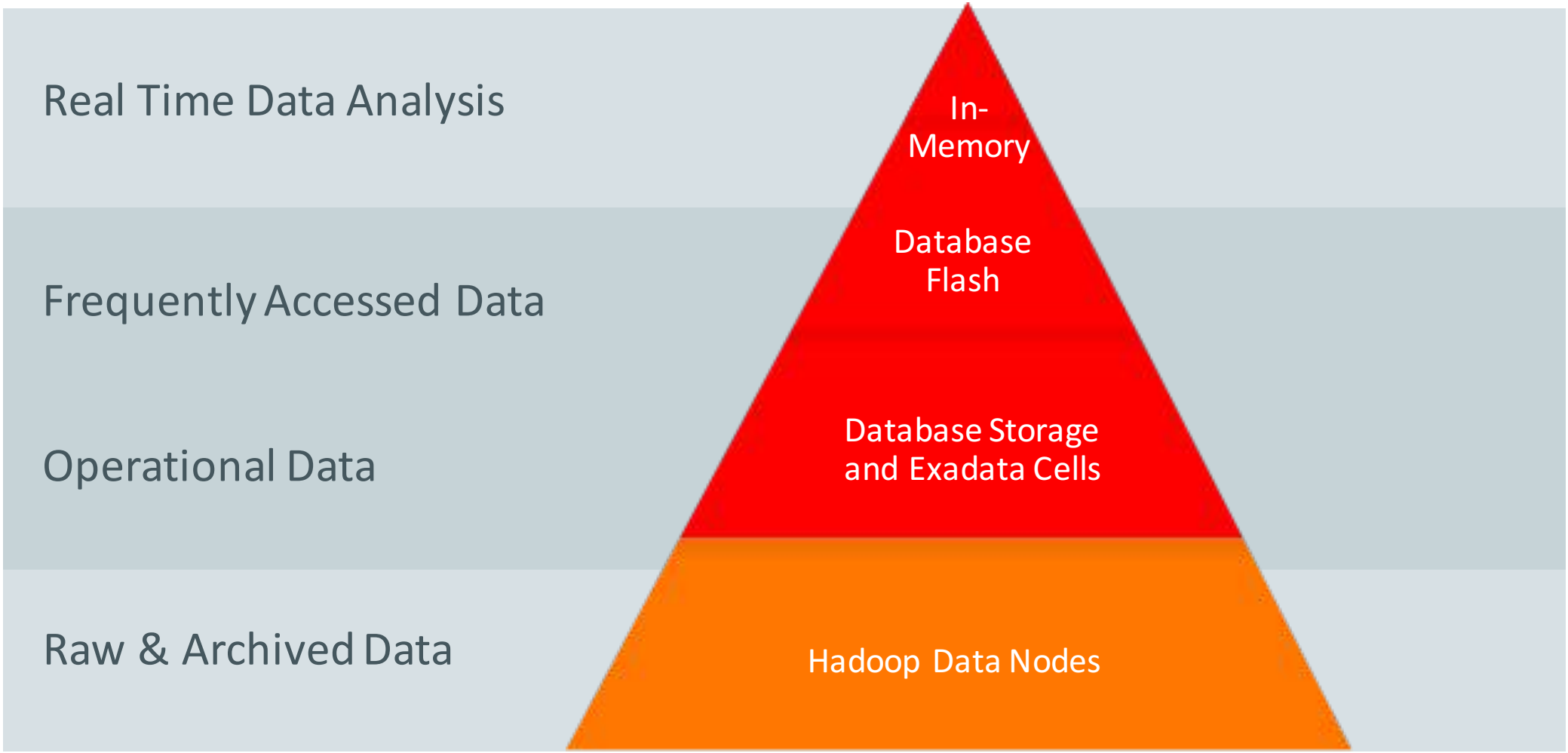
Big Data SQL Performance Features

Compound IO Reduction thru Smart Scans



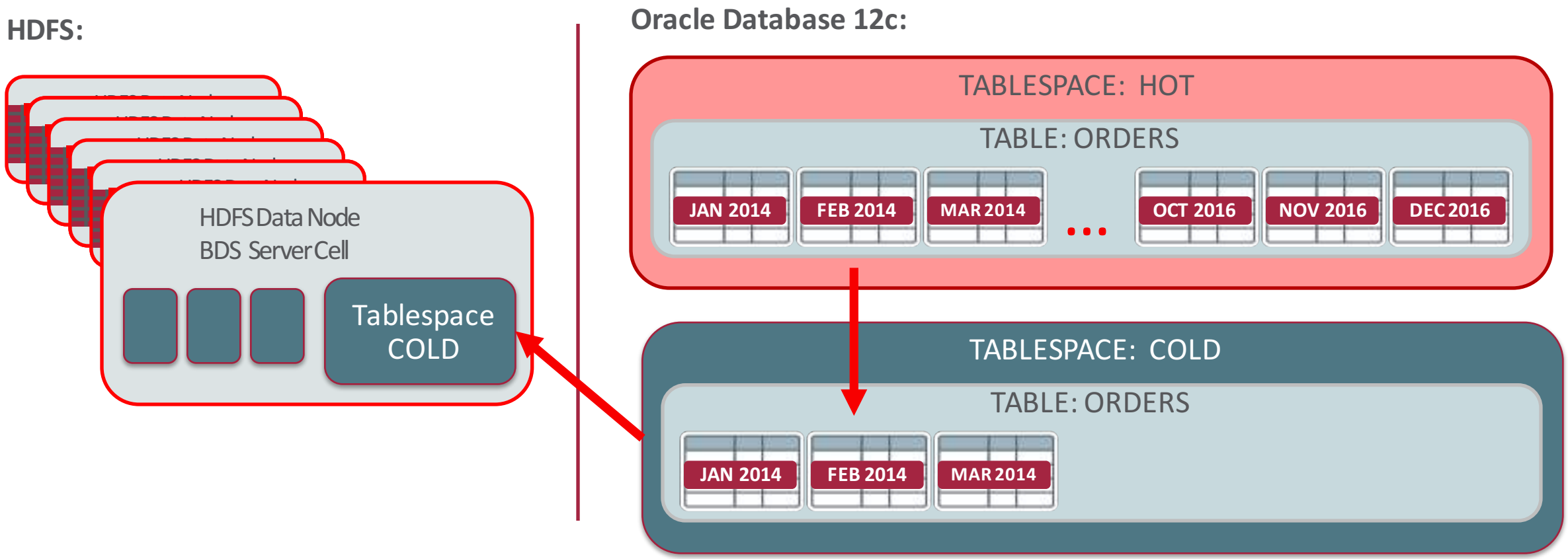
Copyright © 2017, Oracle and/or its affiliates. All rights reserved. |

Big Data Use Case: Information Lifecycle Management



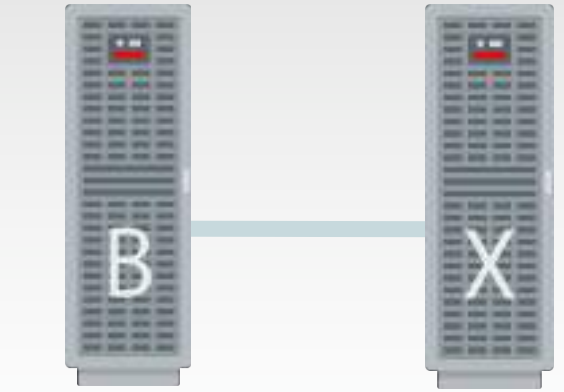
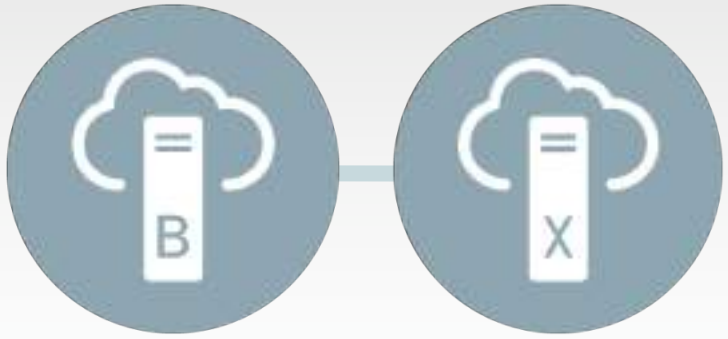
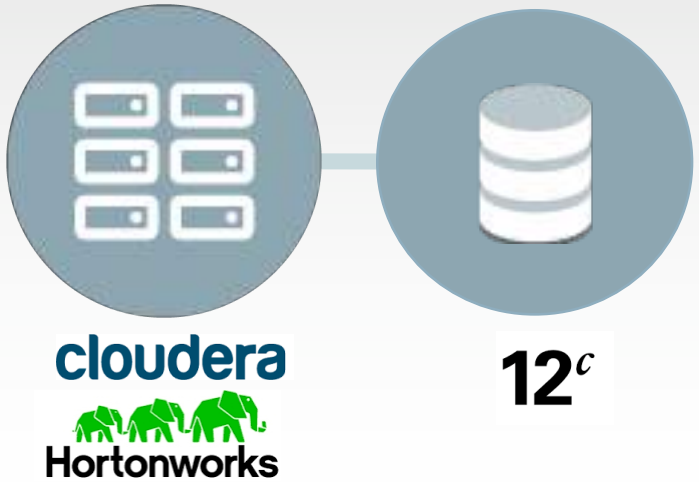
Copyright © 2017, Oracle and/or its affiliates. All rights reserved. |

Archive Data to “Cold” Partitions



Copyright © 2017, Oracle and/or its affiliates. All rights reserved. |

Flexible Deployment Options

Engineered Systems	Oracle Cloud	Commodity
		

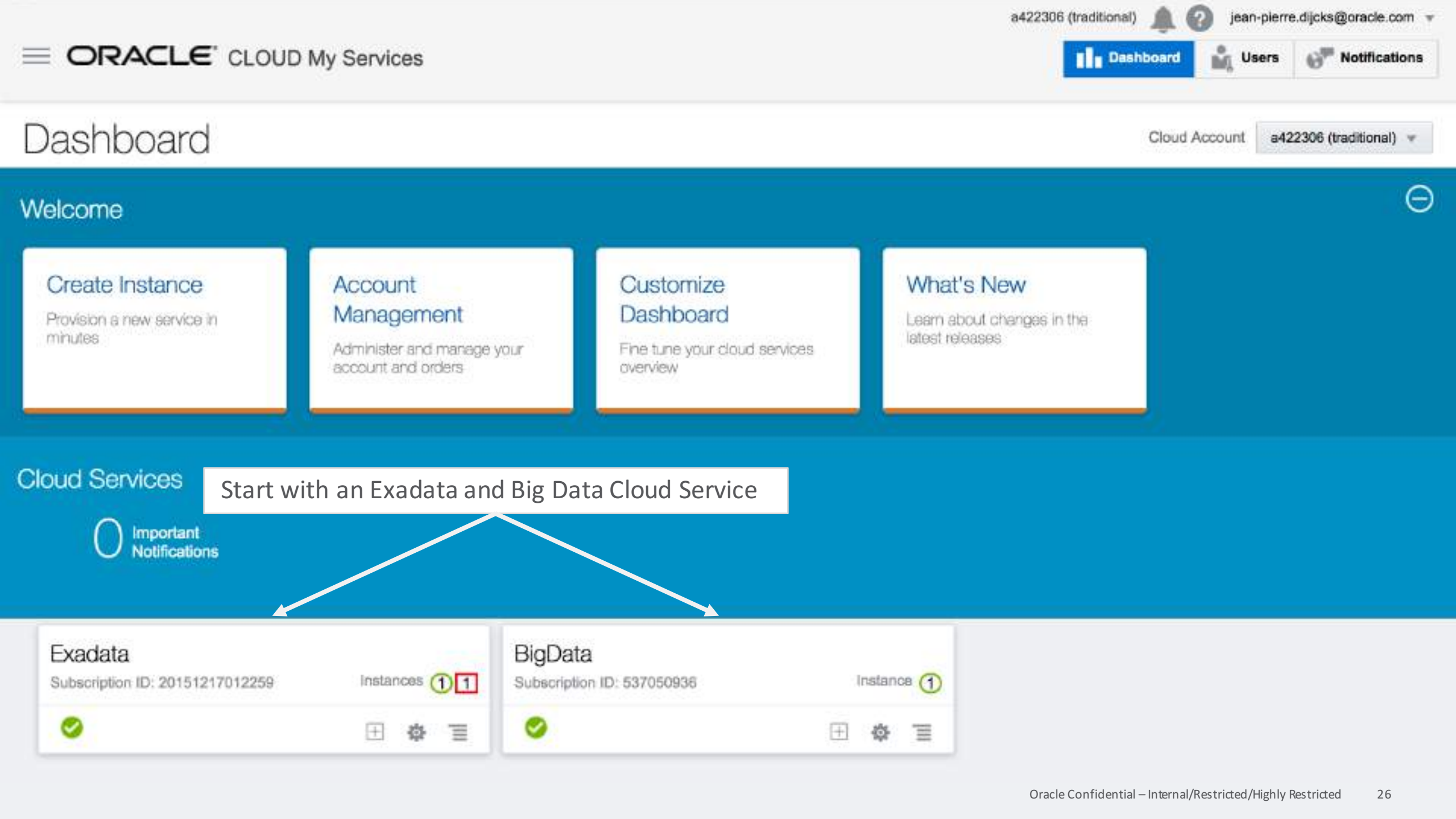
Copyright © 2017, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

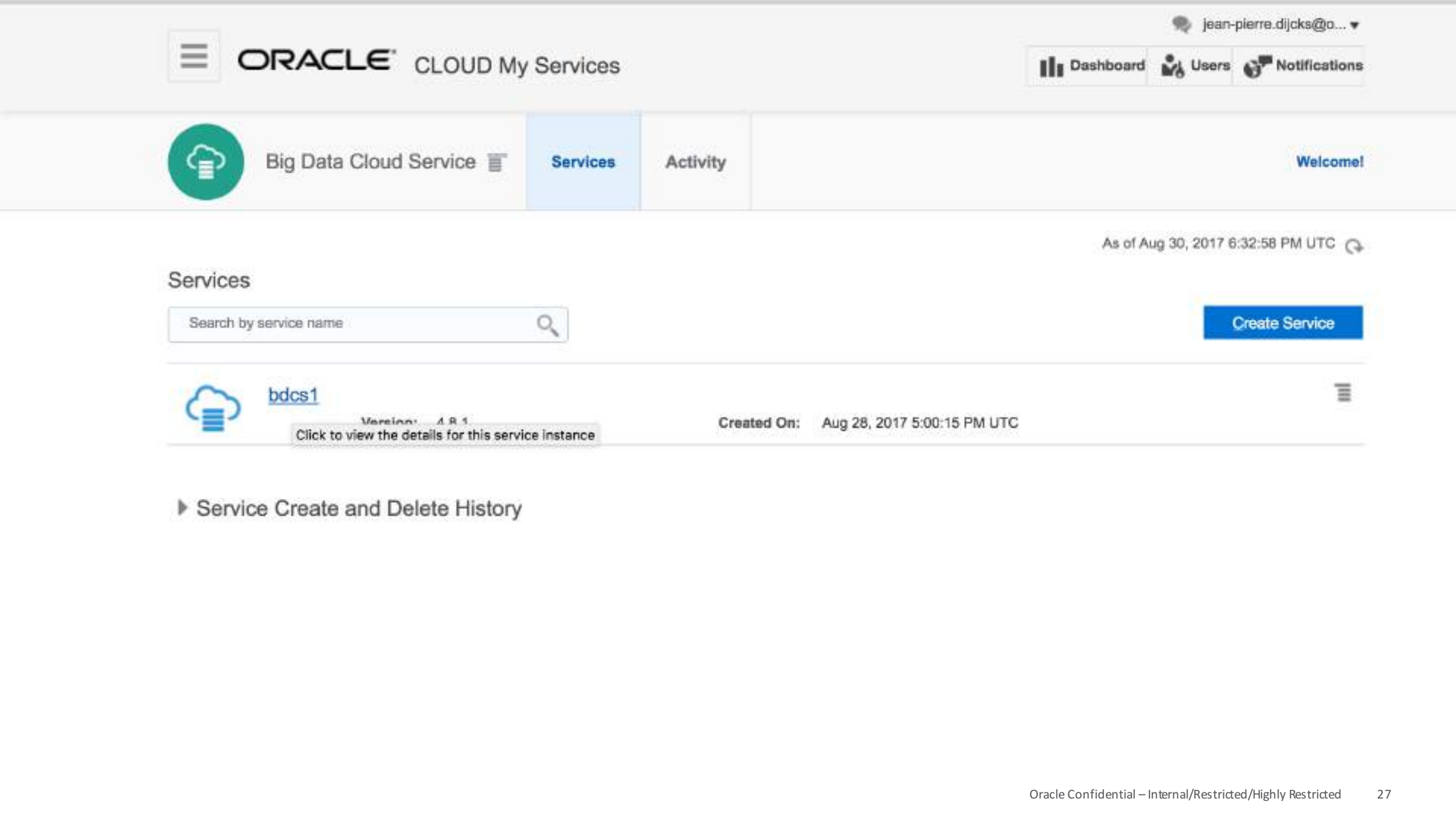
Big Data SQL Cloud Service

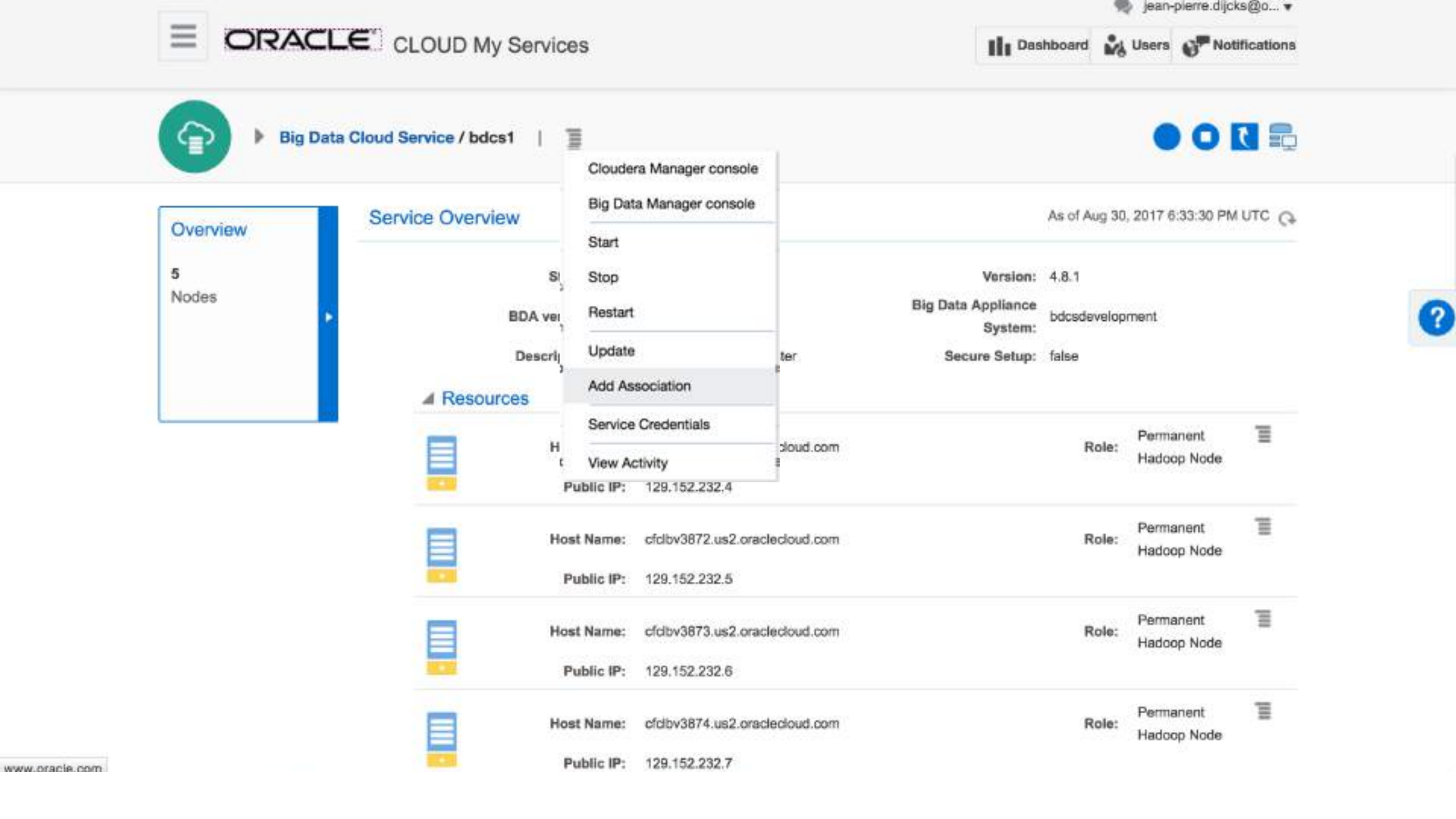
Cloud Simplicity

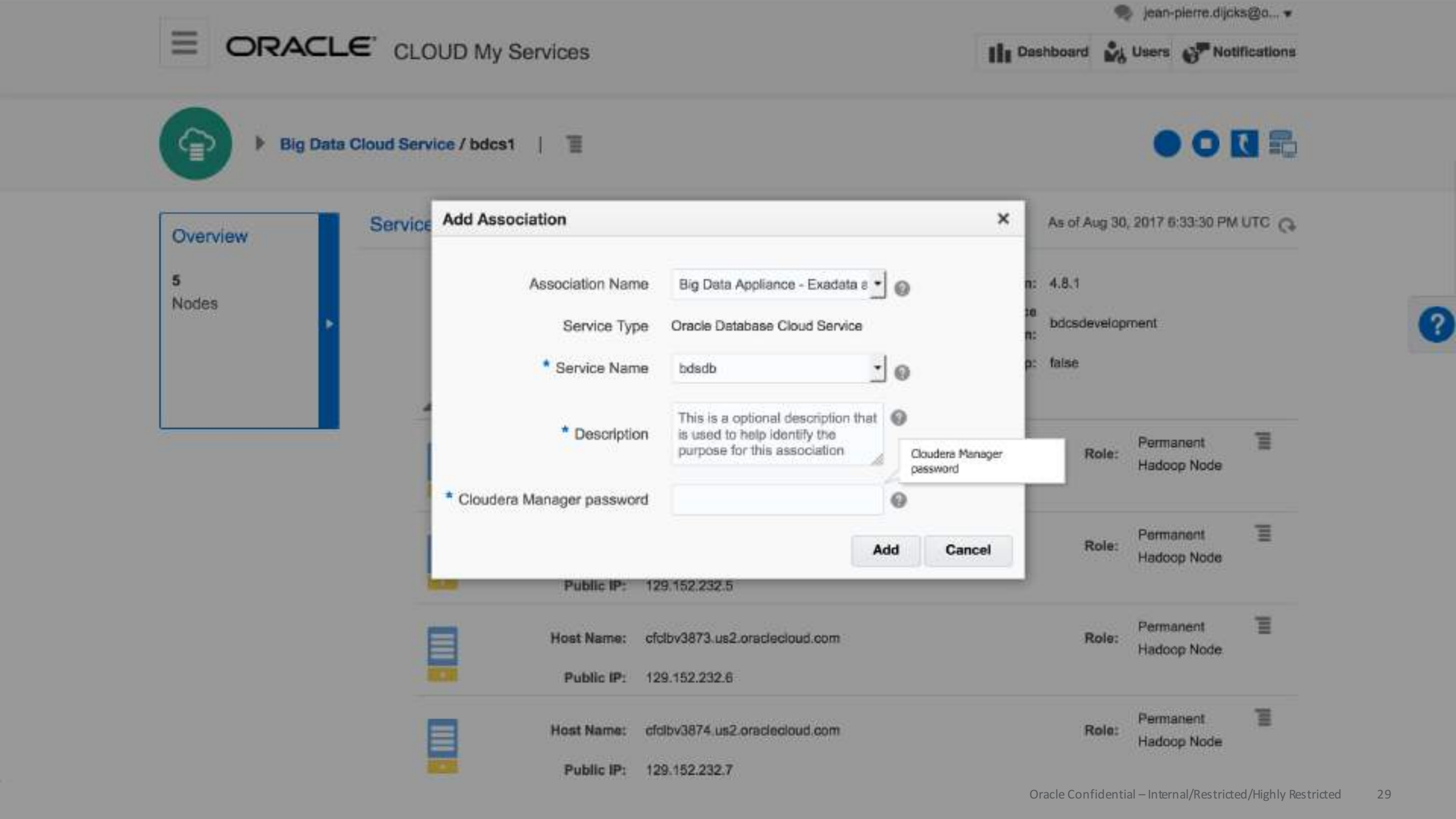
ORACLE

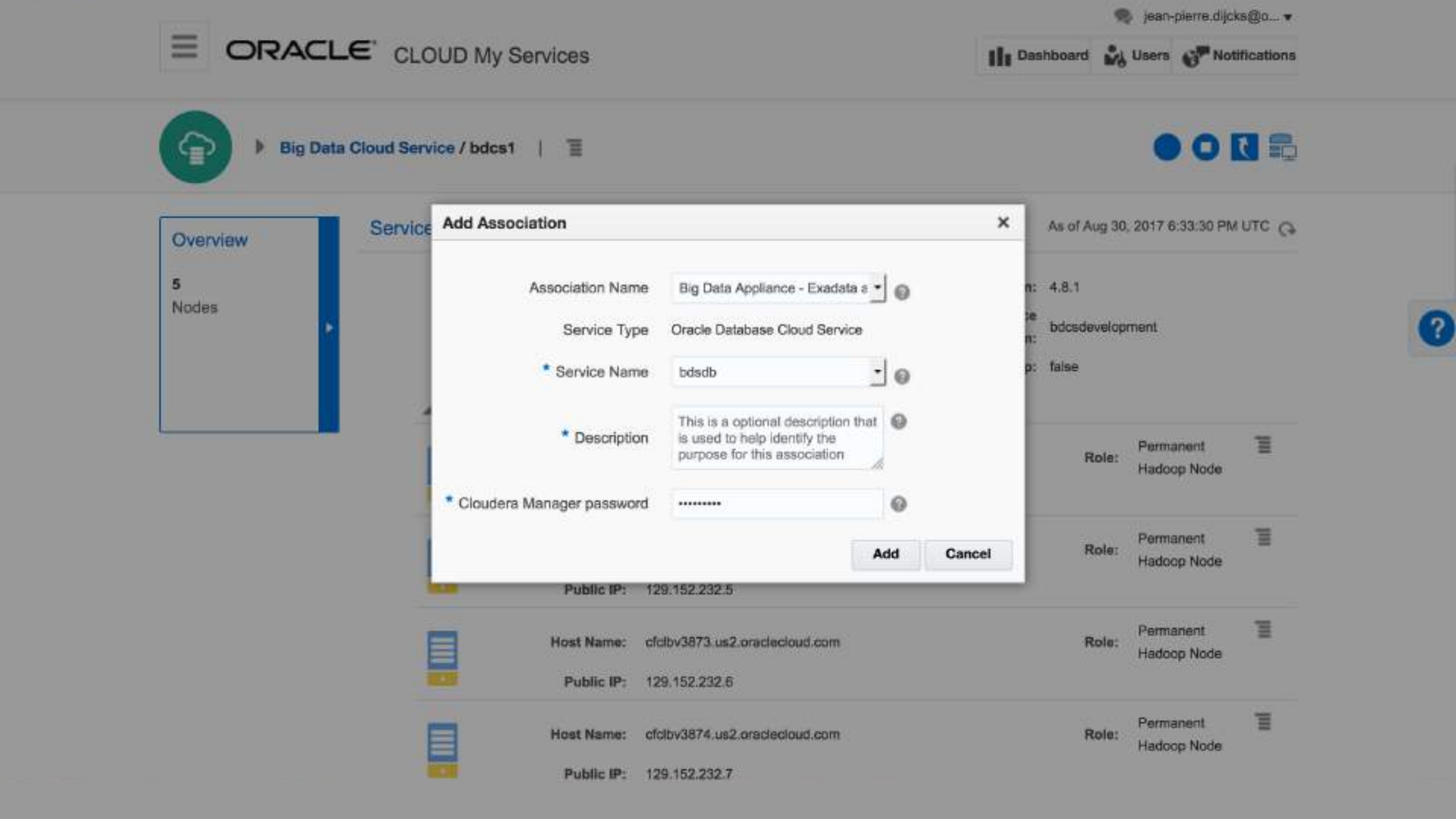
Copyright © 2017, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

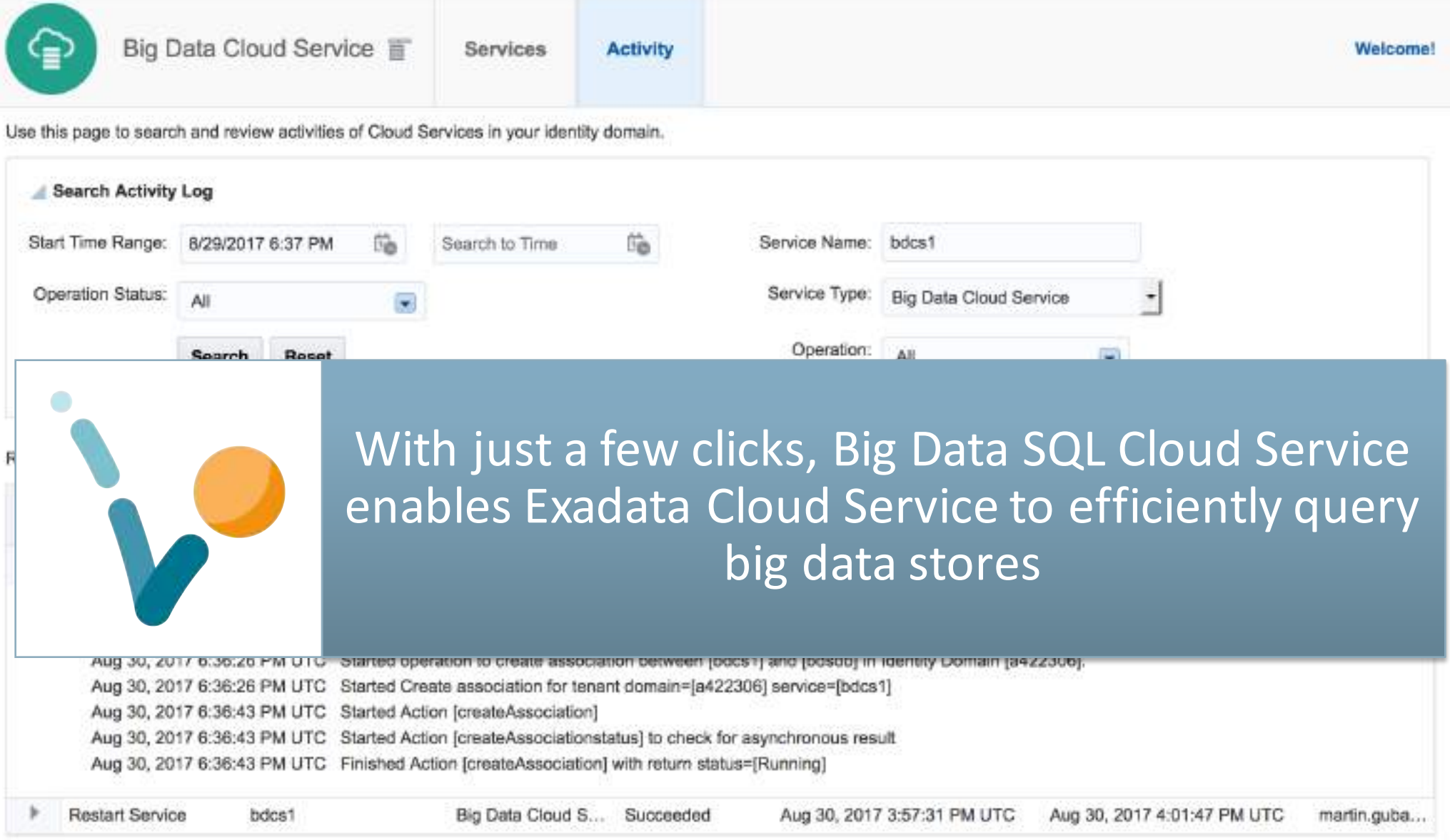

a422306 (traditional)
jean-pierre.dijcks@oracle.com
ORACLE CLOUD My Services
Dashboard
Users
Notifications
Dashboard
Cloud Account
a422306 (traditional)
Welcome
Create Instance
Provision a new service in minutes
Account Management
Administer and manage your account and orders
Customize Dashboard
Fine tune your cloud services overview
What's New
Learn about changes in the latest releases
Cloud Services
Start with an Exadata and Big Data Cloud Service
0 Important Notifications
Exadata
Subscription ID: 20151217012259
Instances 1 1
BigData
Subscription ID: 537050936
Instance 1


jean-pierre.dijcks@o...
ORACLE CLOUD My Services
Dashboard
Users
Notifications
Big Data Cloud Service
Services
Activity
Welcome!
As of Aug 30, 2017 6:32:58 PM UTC
Services
Search by service name
Create Service
bdcs1
Created On: Aug 28, 2017 5:00:15 PM UTC
Click to view the details for this service instance
Service Create and Delete History


jean-pierre.dijcks@o...
ORACLE CLOUD My Services
Dashboard
Users
Notifications
Big Data Cloud Service / bdcs1
Cloudera Manager console
Big Data Manager console
Start
Stop
Restart
Update
Add Association
Service Credentials
View Activity
Overview
5
Nodes
Service Overview
As of Aug 30, 2017 6:33:30 PM UTC
Version: 4.8.1
Big Data Appliance System: bdcsdevelopment
Secure Setup: false
Resources
Public IP: 129.152.232.4
Role: Permanent Hadoop Node
Host Name: cfclbv3872.us2.oraclecloud.com
Public IP: 129.152.232.5
Role: Permanent Hadoop Node
Host Name: cfclbv3873.us2.oraclecloud.com
Public IP: 129.152.232.6
Role: Permanent Hadoop Node
Host Name: cfclbv3874.us2.oraclecloud.com
Public IP: 129.152.232.7
Role: Permanent Hadoop Node


jean-pierre.dijcks@o...
ORACLE CLOUD My Services
Dashboard
Users
Notifications
Big Data Cloud Service / bdcs1
Overview
5
Nodes
Service
As of Aug 30, 2017 6:33:30 PM UTC
Add Association
Association Name
Big Data Appliance - Exadata a
Service Type
Oracle Database Cloud Service
* Service Name
bdsdb
* Description
This is a optional description that is used to help identify the purpose for this association
* Cloudera Manager password
Cloudera Manager password
Add
Cancel
n: 4.8.1
bdcsdevelopment
p: false
Role: Permanent Hadoop Node
Public IP: 129.152.232.5
Host Name: cfclbv3873.us2.oraclecloud.com
Public IP: 129.152.232.6
Host Name: cfclbv3874.us2.oraclecloud.com
Public IP: 129.152.232.7


jean-pierre.dijcks@o...
ORACLE CLOUD My Services
Dashboard
Users
Notifications
Big Data Cloud Service / bdcs1
Overview
5
Nodes
Service
As of Aug 30, 2017 6:33:30 PM UTC
Add Association
Association Name
Big Data Appliance - Exadata s
Service Type
Oracle Database Cloud Service
* Service Name
bdsdb
* Description
This is a optional description that is used to help identify the purpose for this association
* Cloudera Manager password
Add
Cancel
4.8.1
bdcsdevelopment
false
Role:
Permanent Hadoop Node
Public IP:
129.152.232.5
Host Name:
cfclbv3873.us2.oraclecloud.com
Public IP:
129.152.232.6
Host Name:
cfclbv3874.us2.oraclecloud.com
Public IP:
129.152.232.7


Big Data Cloud Service
Services
Activity
Welcome!
Use this page to search and review activities of Cloud Services in your identity domain.
Search Activity Log
Start Time Range: 8/29/2017 6:37 PM
Search to Time
Service Name: bdcs1
Operation Status: All
Service Type: Big Data Cloud Service
Search
Reset
Operation: All
With just a few clicks, Big Data SQL Cloud Service enables Exadata Cloud Service to efficiently query big data stores
Aug 30, 2017 6:36:26 PM UTC Started Create association for tenant domain=[a422306] service=[bdcs1]
Aug 30, 2017 6:36:43 PM UTC Started Action [createAssociation]
Aug 30, 2017 6:36:43 PM UTC Started Action [createAssociationstatus] to check for asynchronous result
Aug 30, 2017 6:36:43 PM UTC Finished Action [createAssociation] with return status=[Running]
Restart Service bdcs1 Big Data Cloud S... Succeeded Aug 30, 2017 3:57:31 PM UTC Aug 30, 2017 4:01:47 PM UTC martin.guba...

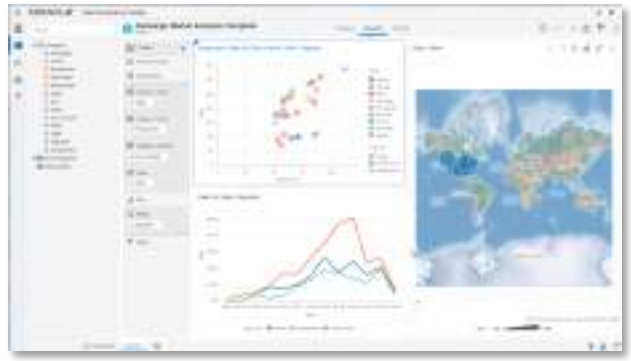
Securely Access All Data

Define

- Create big data-enabled tables
- Leverage underlying metadata
- Use tooling to simplify creation

Secure

- Grant access to users/groups
- Apply Oracle advanced security policies to authorize access

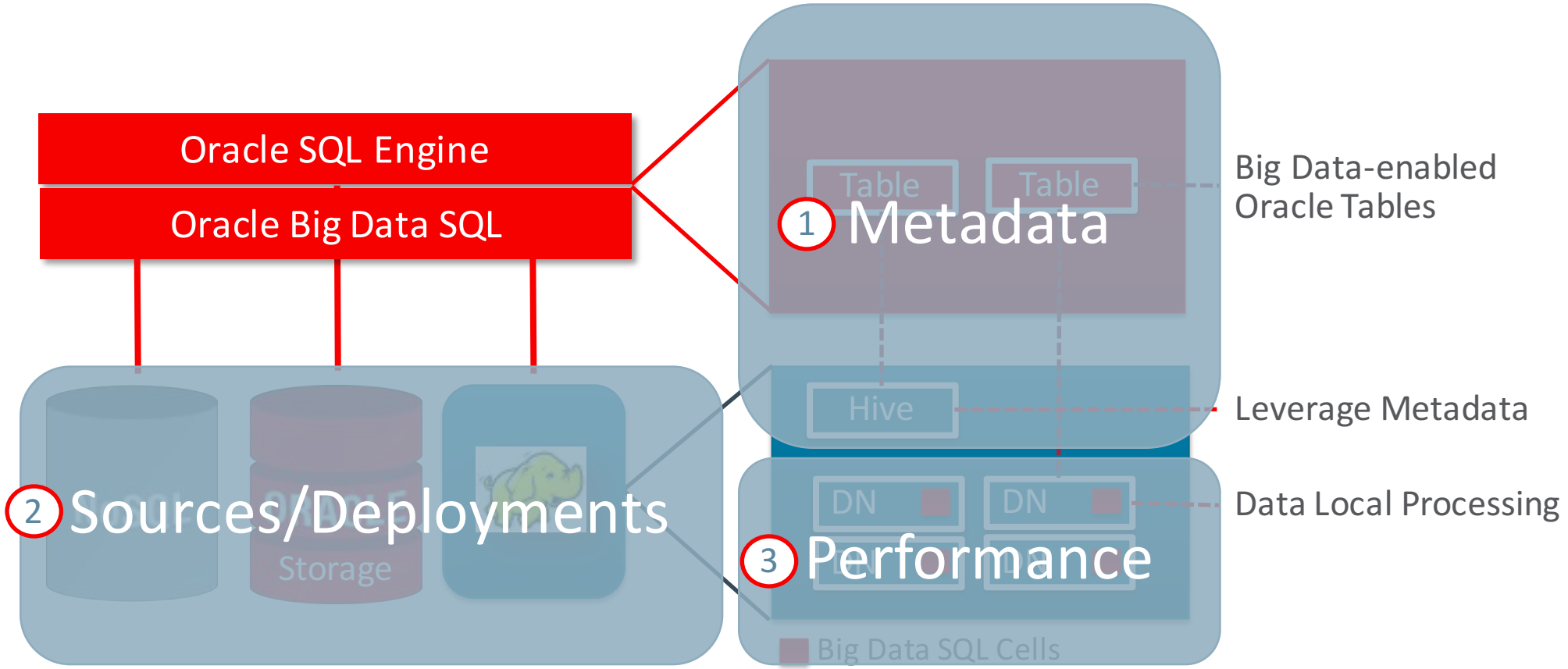


Query

- Use Oracle SQL to analyze data
- Seamlessly combine big data sources
- Use favorite tools and applications

Copyright © 2017, Oracle and/or its affiliates. All rights reserved. |

Big Data SQL Focus Areas



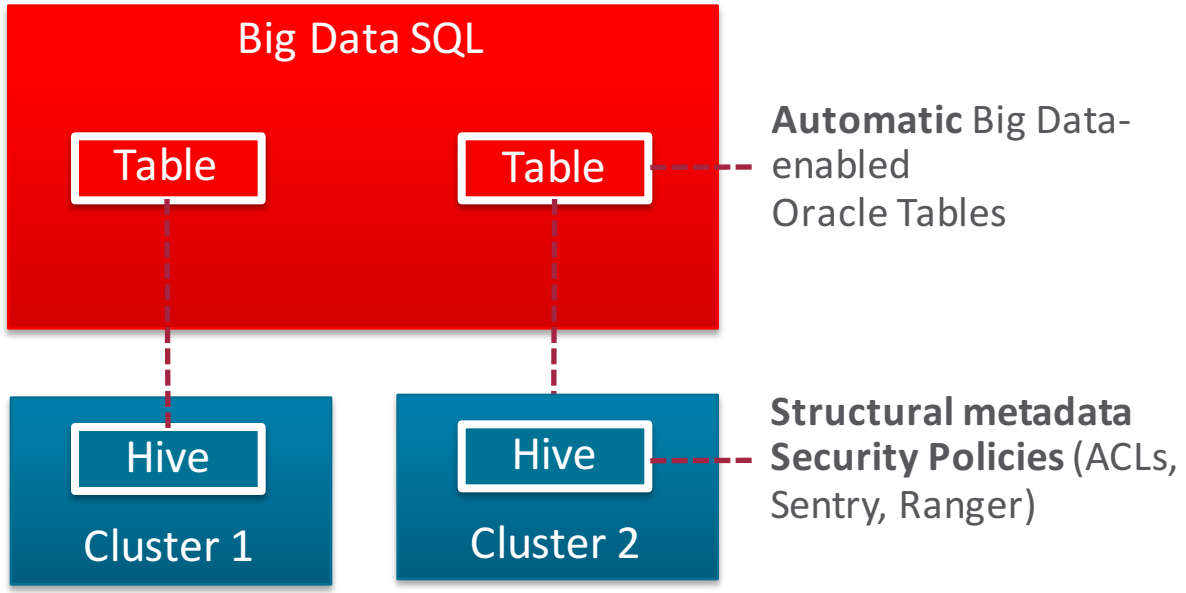
Confidential – Oracle Internal/Restricted/Highly Restricted

Copyright © 2017, Oracle and/or its affiliates. All rights reserved. |

Metadata

Big Data SQL Aggregates Metadata from Underlying Sources

- **Simplify Access**
  - Automatic table definition when metadata is available
  - Discover metadata from structured files
- **Support Source Security Policies**
  - Use security policies defined on source
- **Extend with Unique Capabilities**
  - Advanced Oracle Security policies



Copyright © 2017, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

New Sources

Apache Kafka

- Query Apache Topics for real time analysis
- Blend real time streams with other data sources

Object Stores

Oracle Object Store

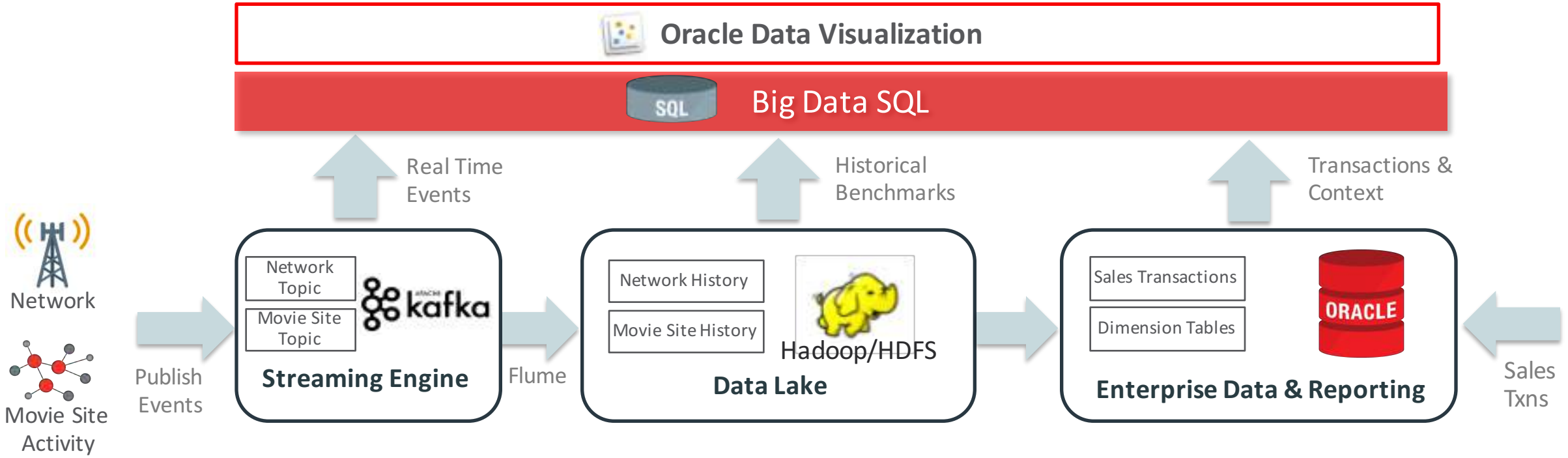
Amazon S3

- Efficiently query object store data
- Support any file type

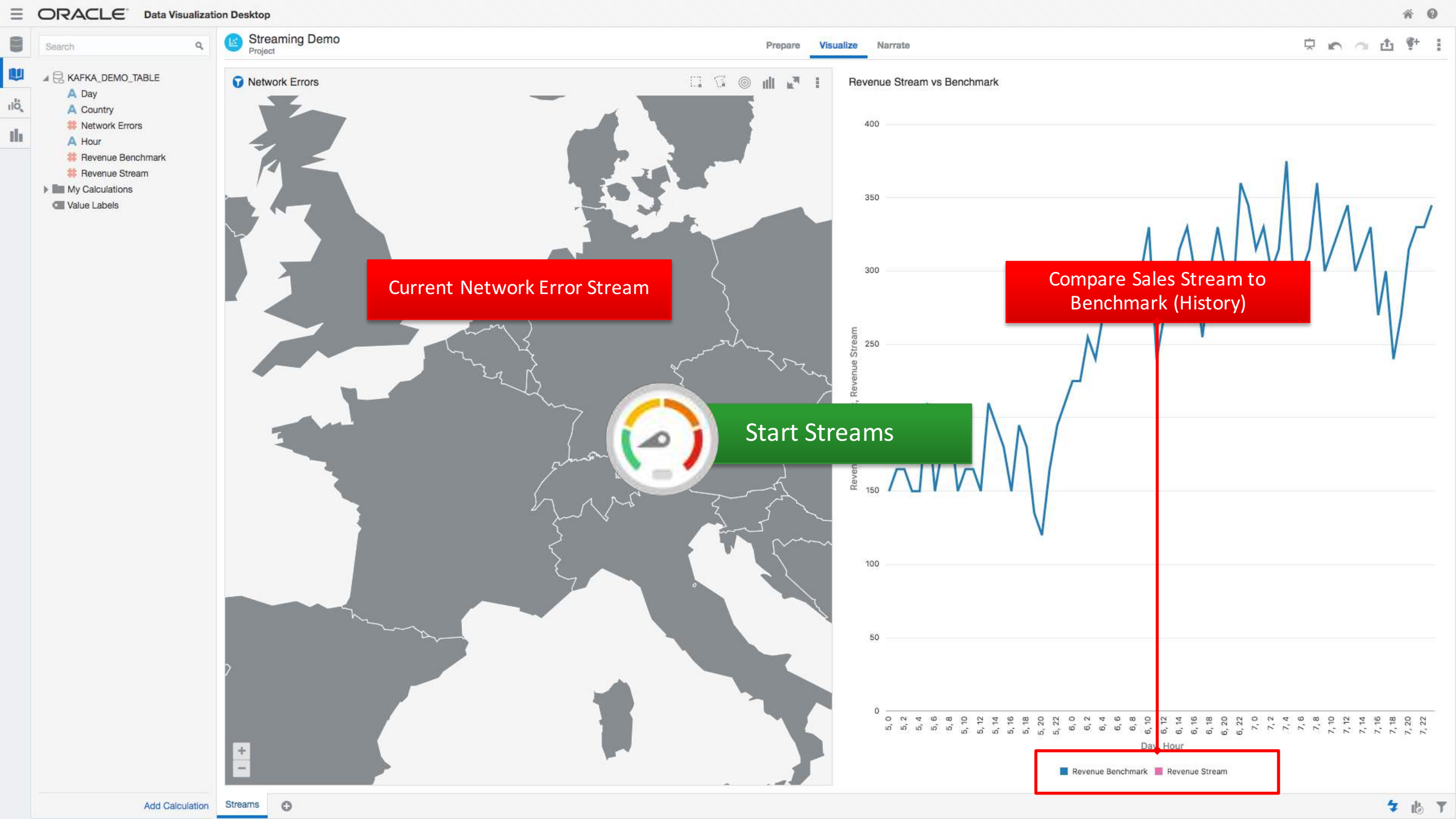
Copyright © 2017, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

Demonstration Scenario

Analyzing Real-time Streams



Copyright © 2017, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted


ORACLE Data Visualization Desktop
Search
KAFKA_DEMO_TABLE
Day
Country
Network Errors
Hour
Revenue Benchmark
Revenue Stream
My Calculations
Value Labels
Add Calculation
Streaming Demo
Project
Prepare
Visualize
Narrate
Network Errors
Current Network Error Stream
Start Streams
Revenue Stream vs Benchmark
Compare Sales Stream to Benchmark (History)
400
350
300
250
150
100
50
0
Revenue Benchmark
Revenue Stream
Streams

Search

- KAFKA_DEMO_TABLE
  - Day
  - Country
  - Network Errors
  - Hour
  - Revenue Benchmark
  - Revenue Stream
- My Calculations
- Value Labels

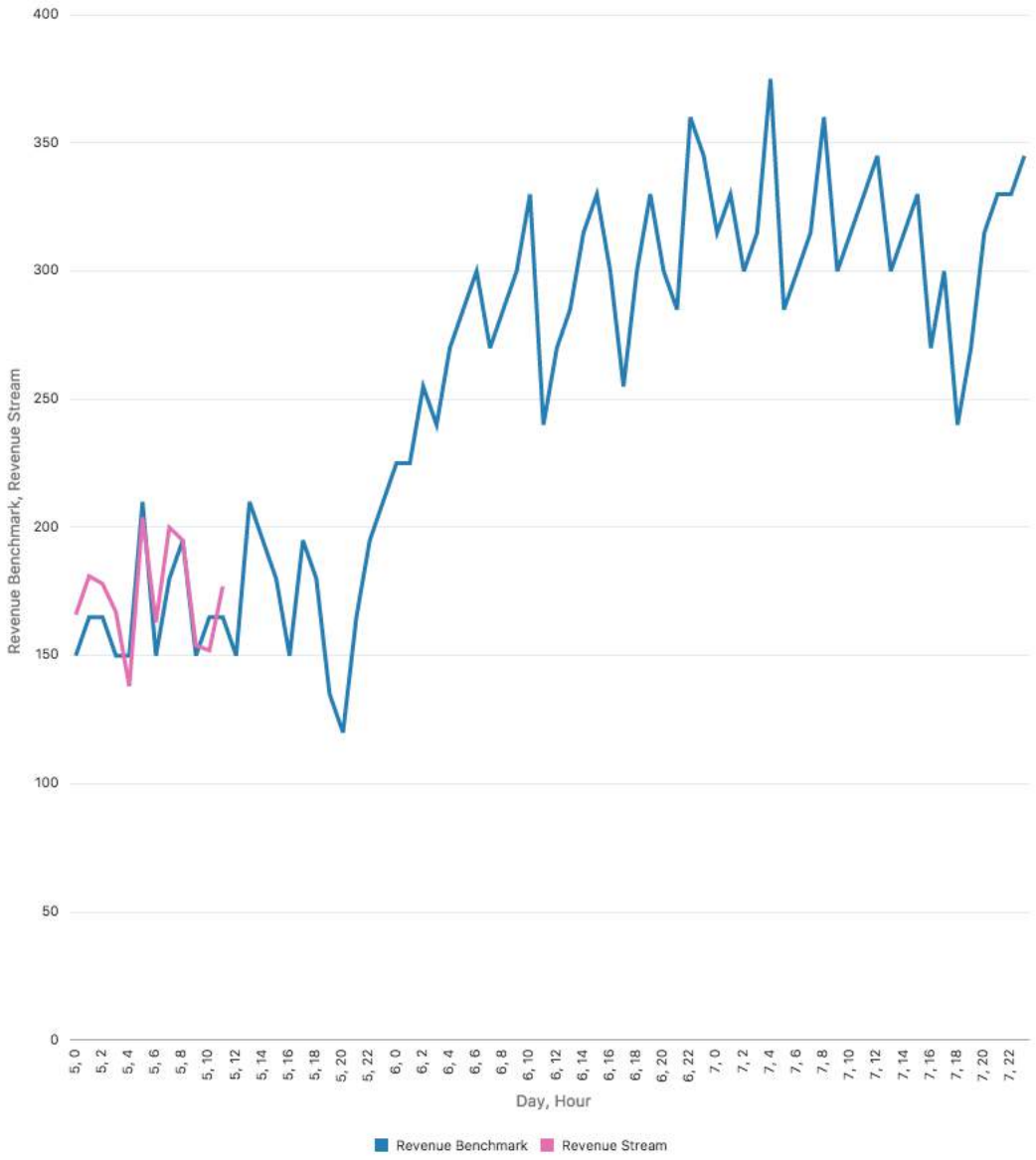
Add Calculation

Streaming Demo
Project

Prepare | Visualize | Narrate

Network Errors

Revenue Stream vs Benchmark



Streams

Search

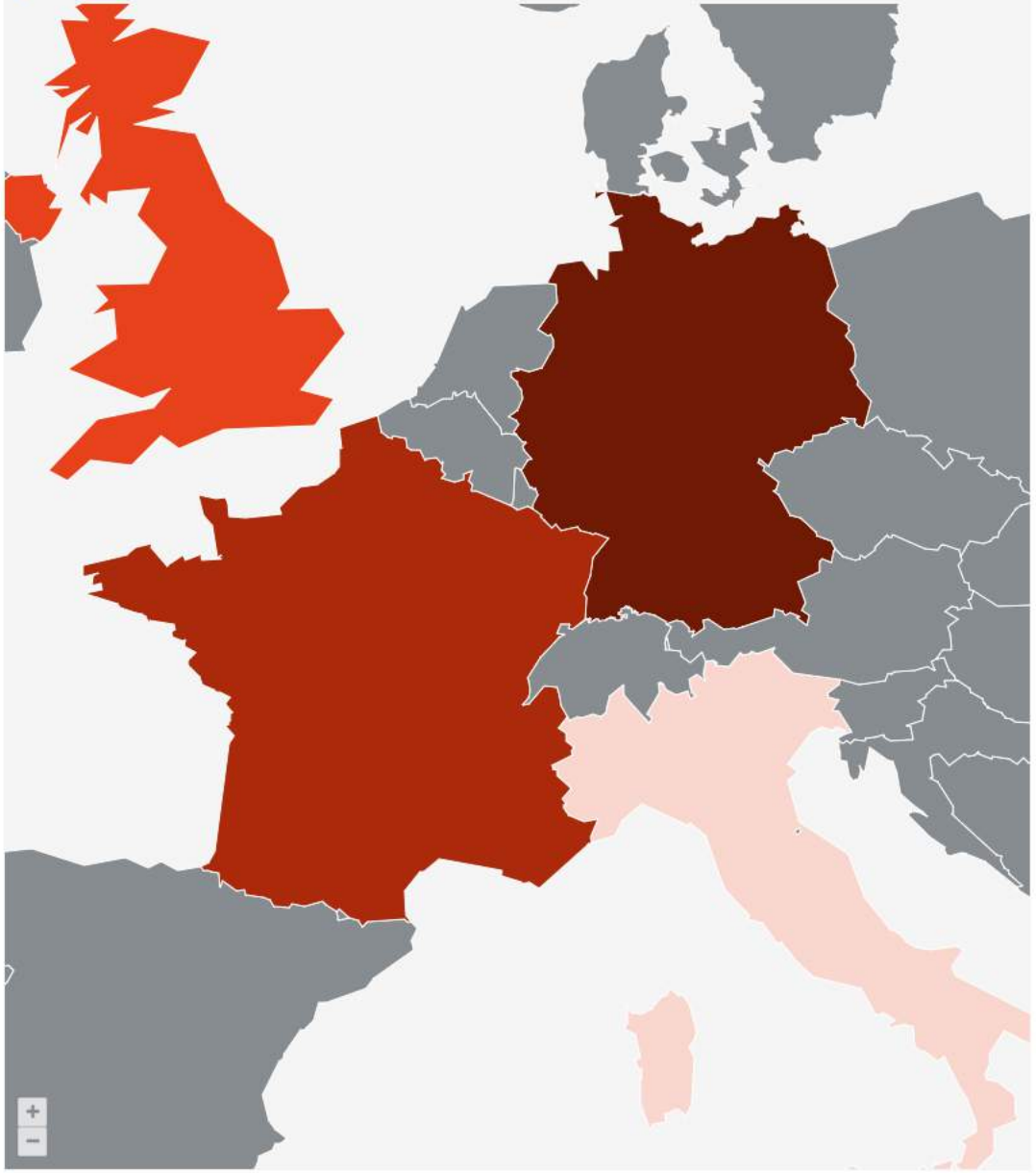
- KAFKA_DEMO_TABLE
  - Day
  - Country
  - Network Errors
  - Hour
  - Revenue Benchmark
  - Revenue Stream
- My Calculations
- Value Labels

Add Calculation

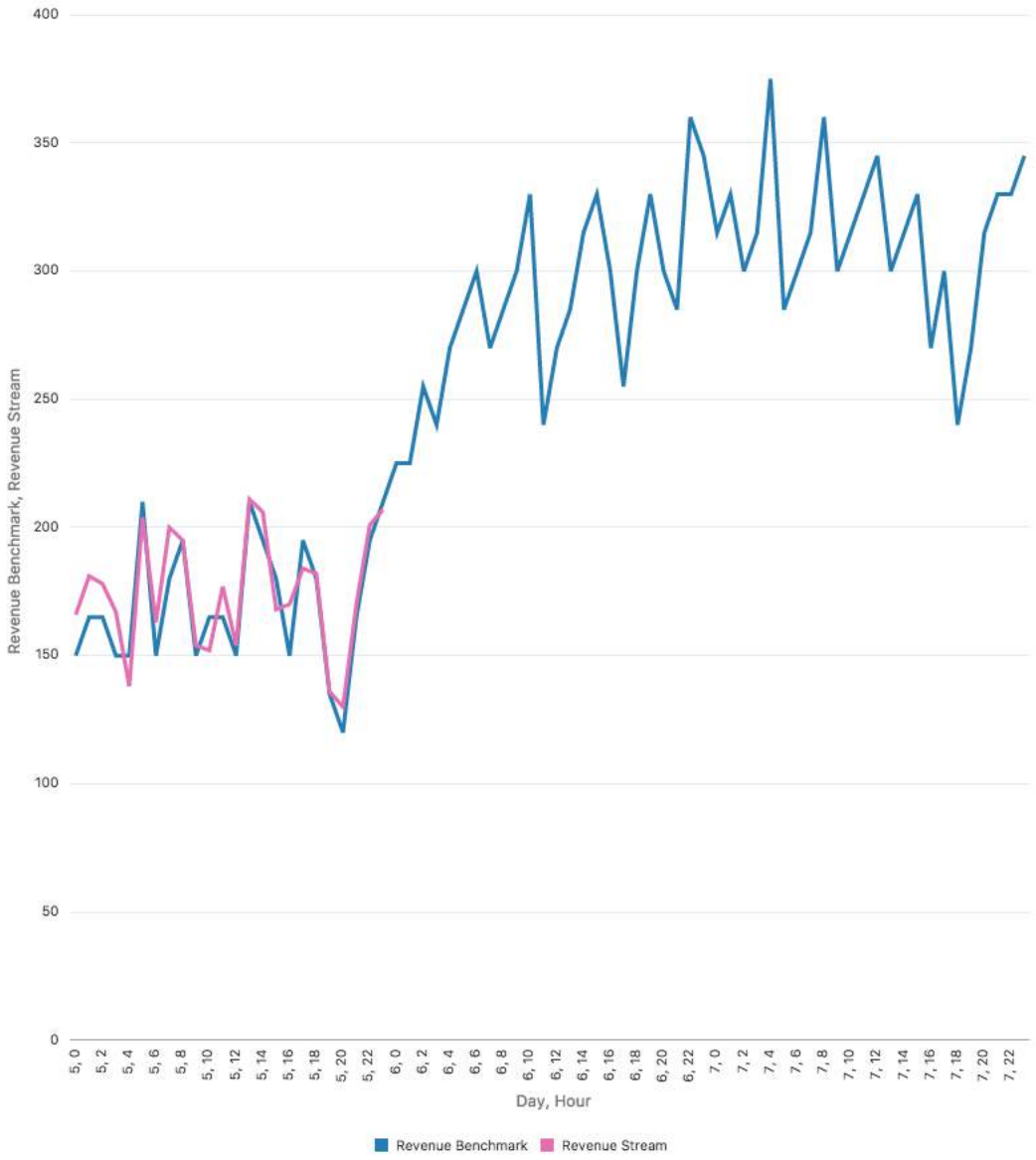
Streaming Demo
Project

Prepare | Visualize | Narrate

Network Errors



Revenue Stream vs Benchmark



Streams

Search

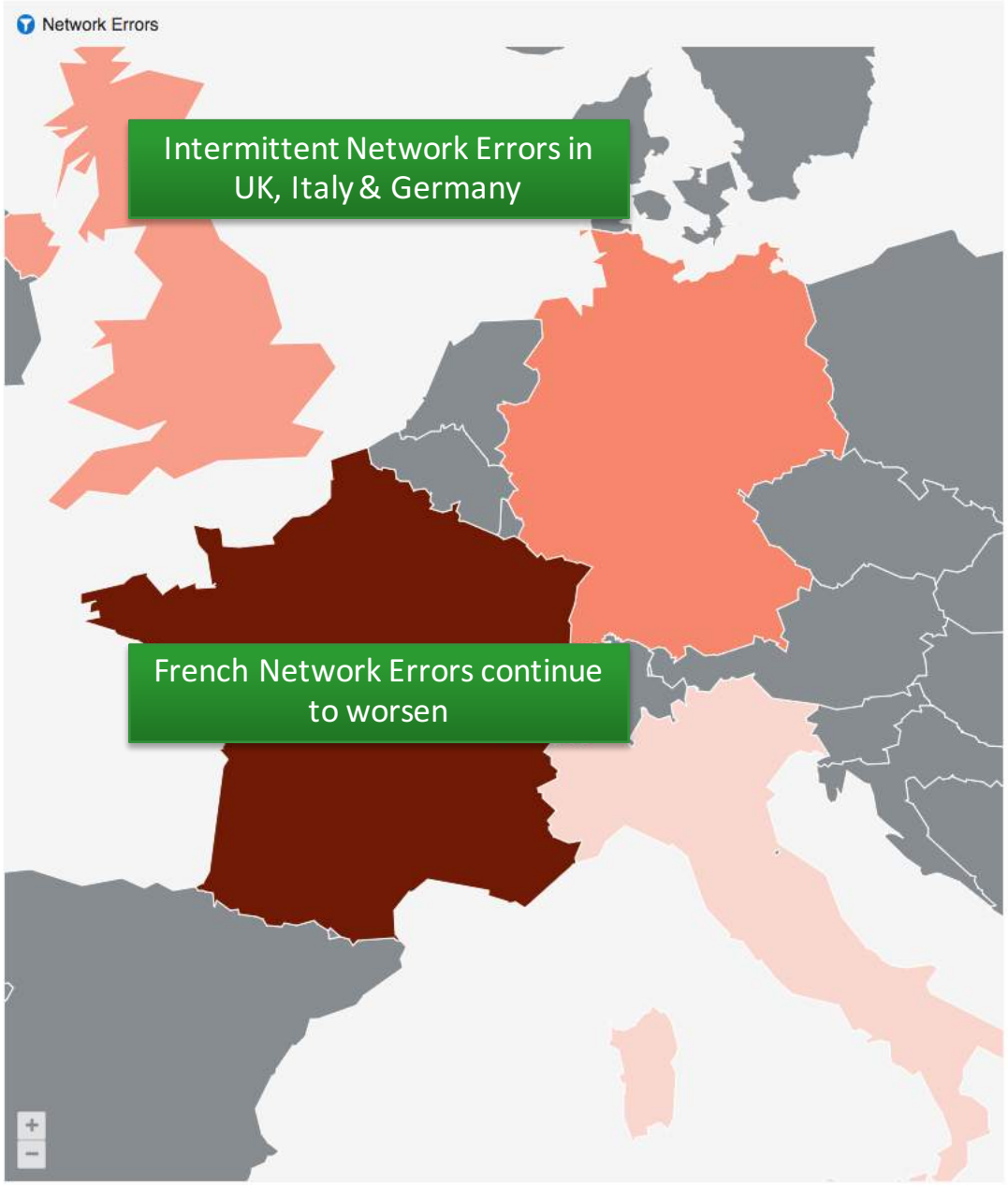
- KAFKA_DEMO_TABLE
  - Day
  - Country
  - Network Errors
  - Hour
  - Revenue Benchmark
  - Revenue Stream
- My Calculations
- Value Labels

Add Calculation

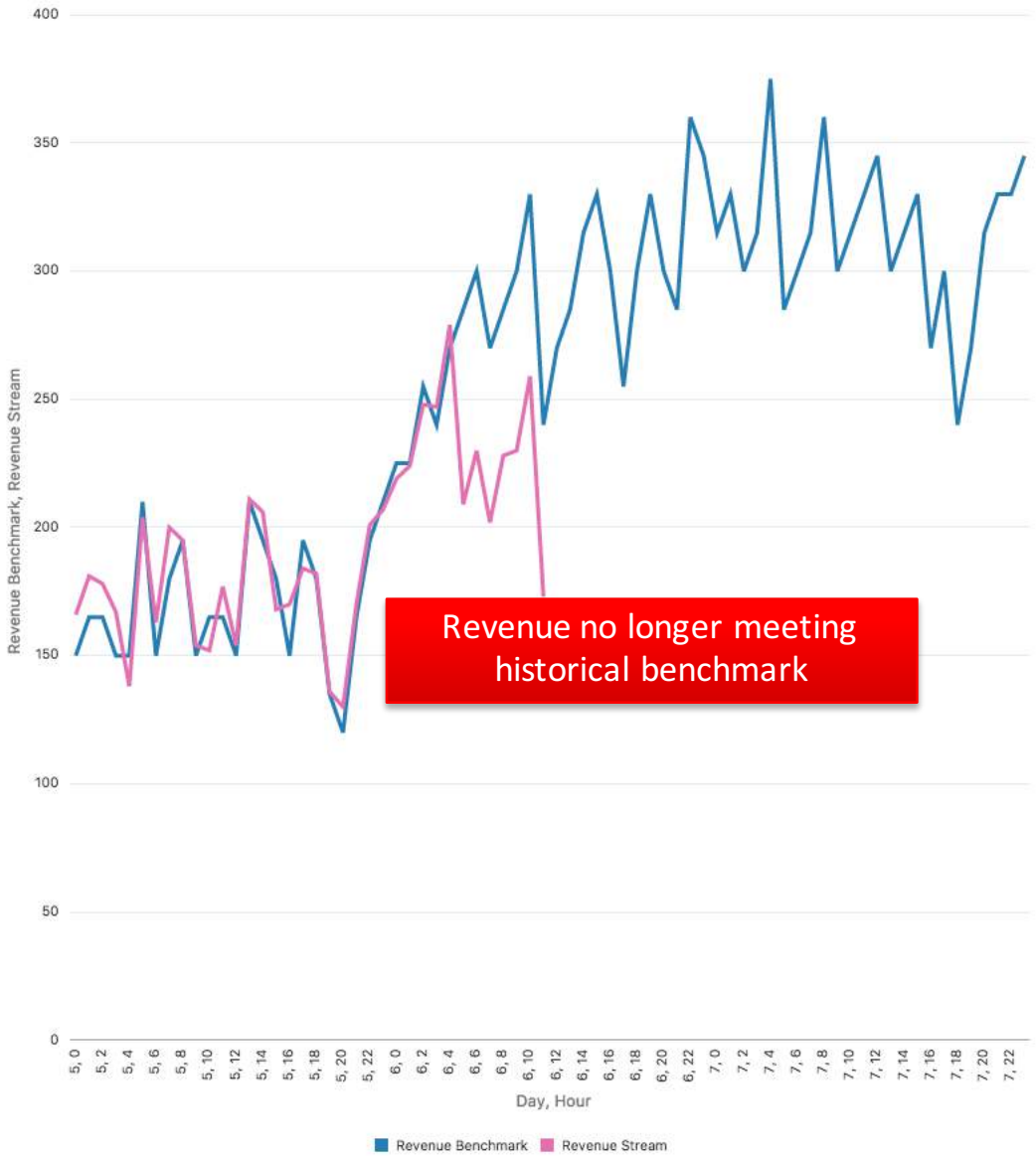
Streaming Demo
Project

Prepare | Visualize | Narrate

Network Errors



Revenue Stream vs Benchmark



Streams

Search

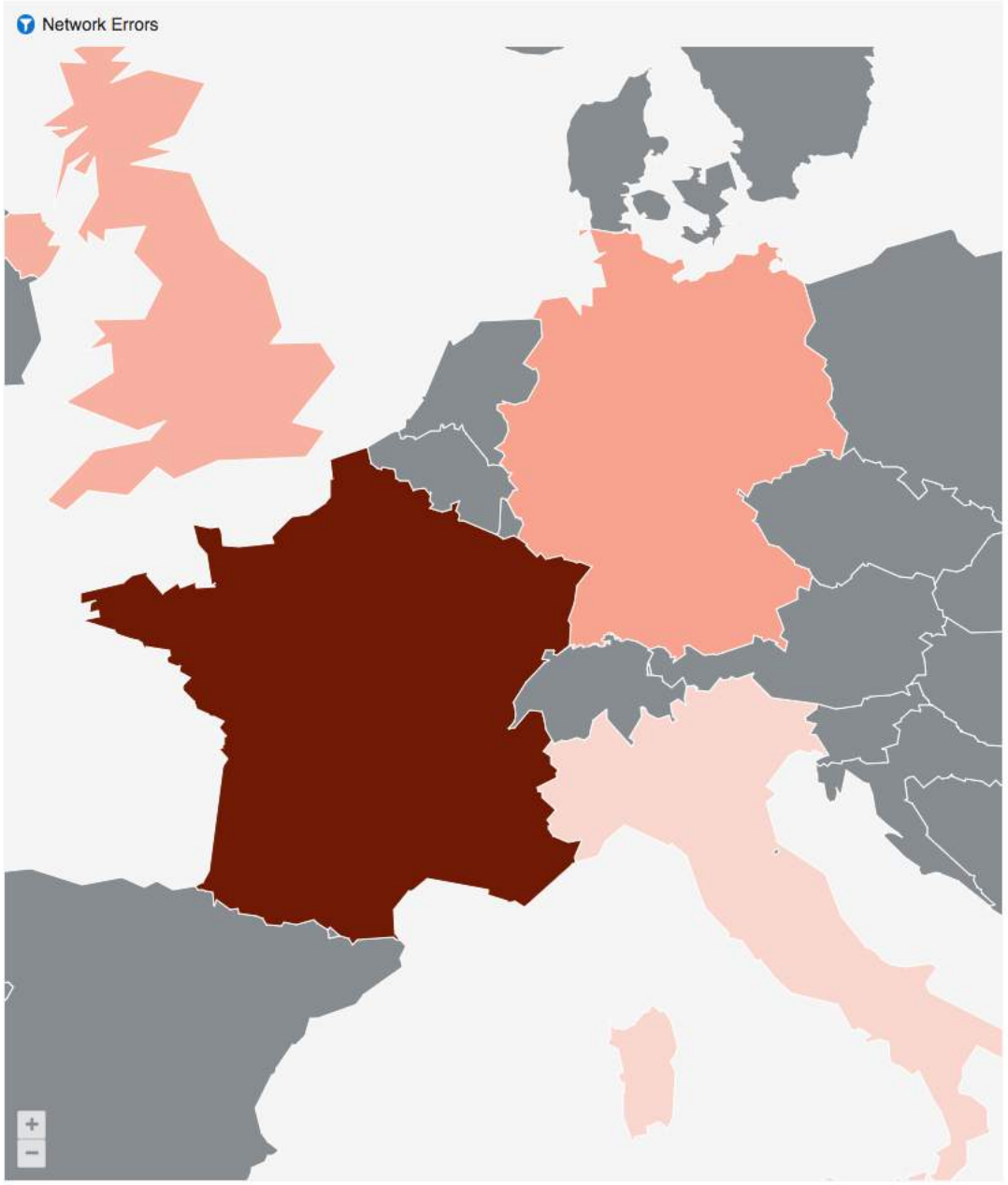
- KAFKA_DEMO_TABLE
  - Day
  - Country
  - Network Errors
  - Hour
  - Revenue Benchmark
  - Revenue Stream
- My Calculations
- Value Labels

Add Calculation

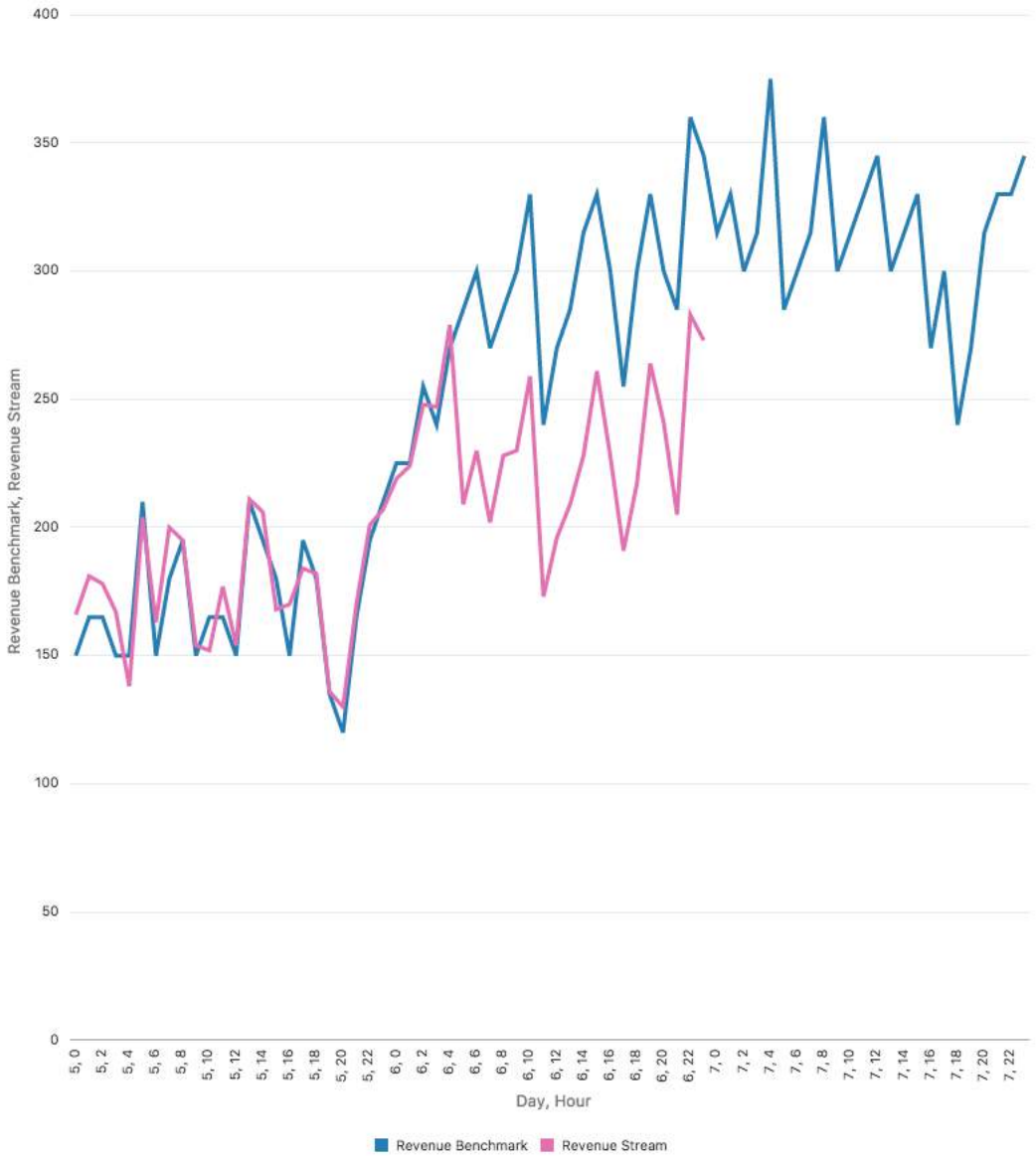
Streaming Demo
Project

Prepare | Visualize | Narrate

Network Errors



Revenue Stream vs Benchmark



Streams

Search

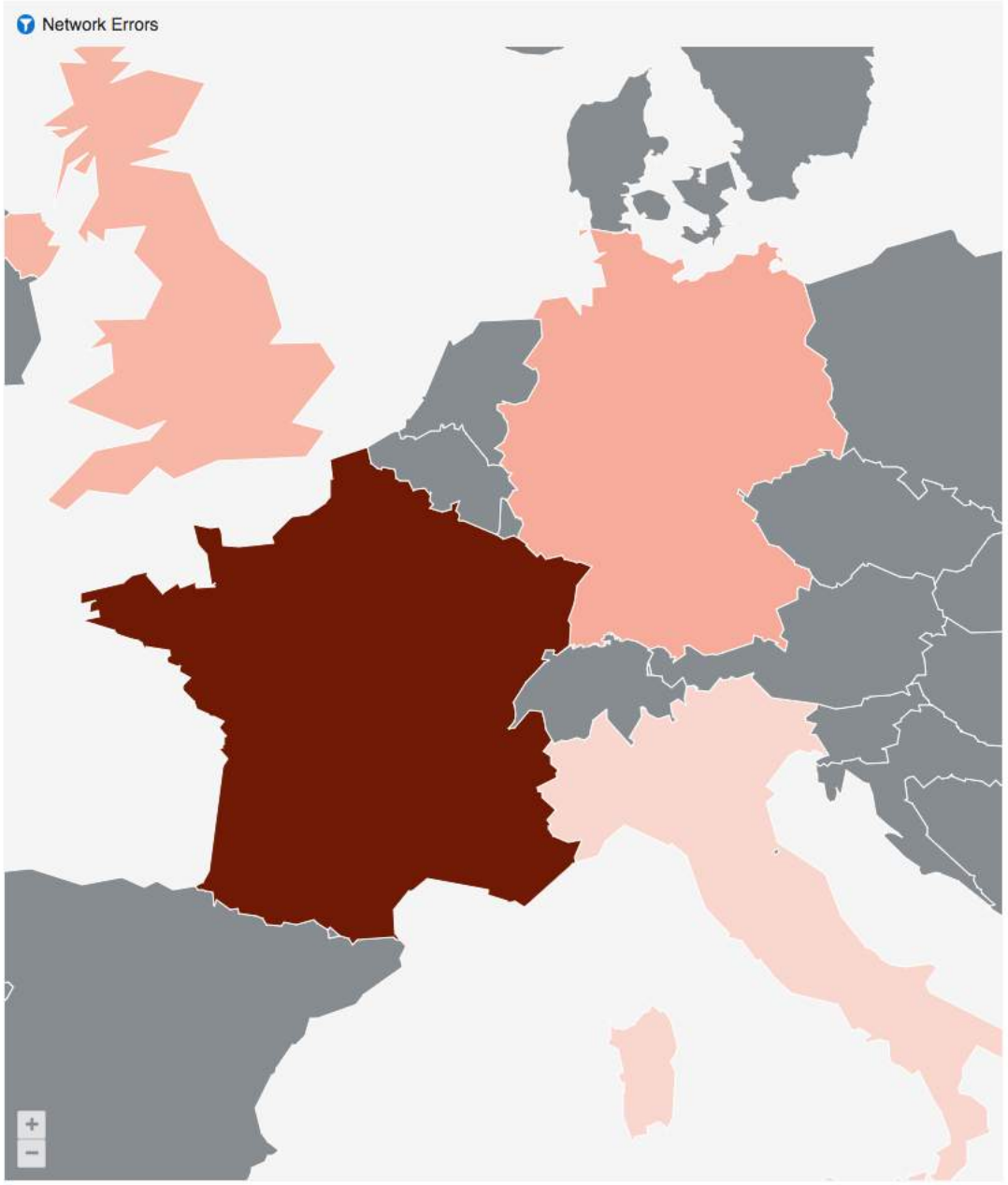
- KAFKA_DEMO_TABLE
  - Day
  - Country
  - Network Errors
  - Hour
  - Revenue Benchmark
  - Revenue Stream
- My Calculations
- Value Labels

Add Calculation

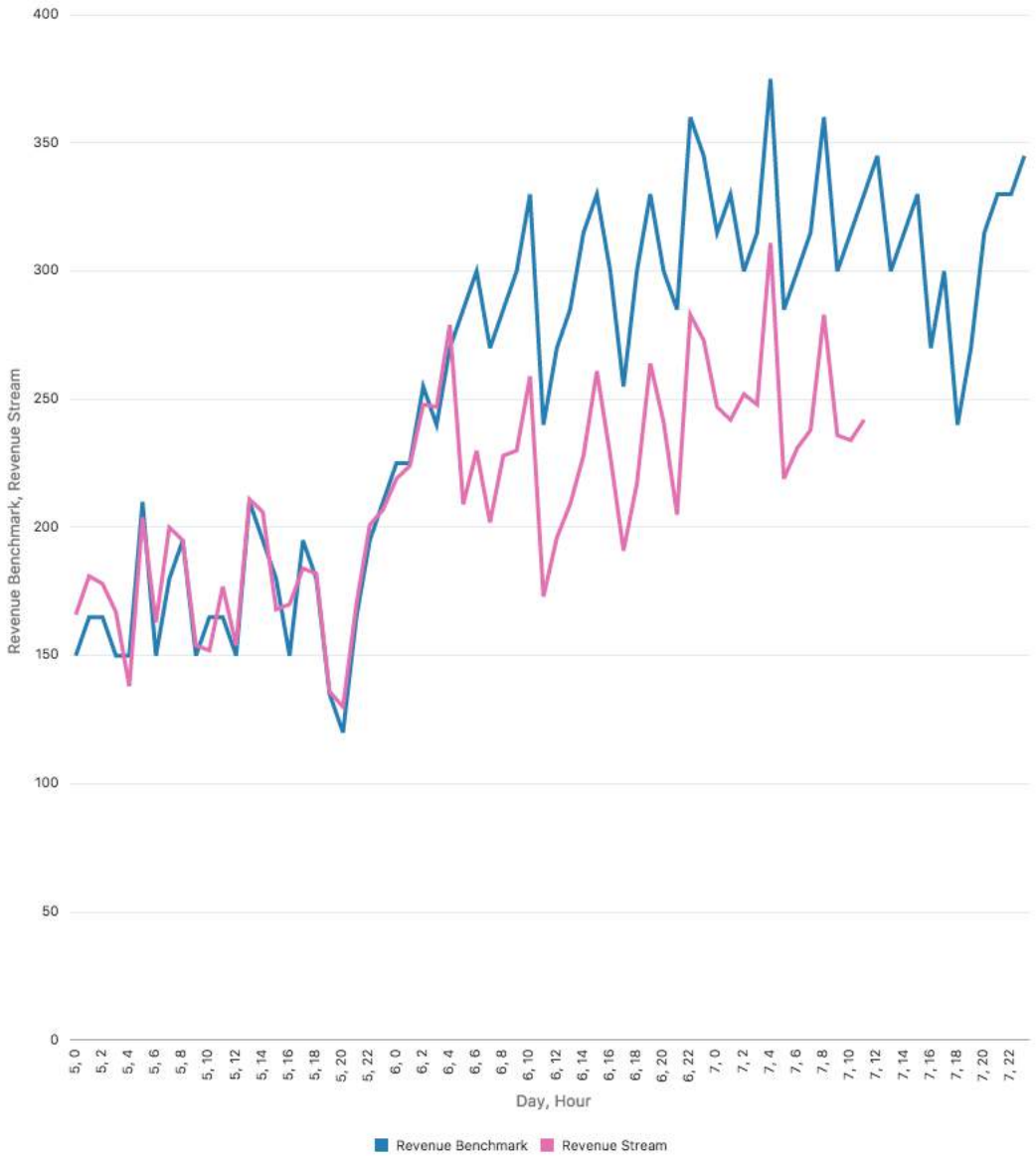
Streaming Demo
Project

Prepare | Visualize | Narrate

Network Errors



Revenue Stream vs Benchmark



Streams

Search

- KAFKA_DEMO_TABLE
  - Day
  - Country
  - Network Errors
  - Hour
  - Revenue Benchmark
  - Revenue Stream
- My Calculations
- Value Labels

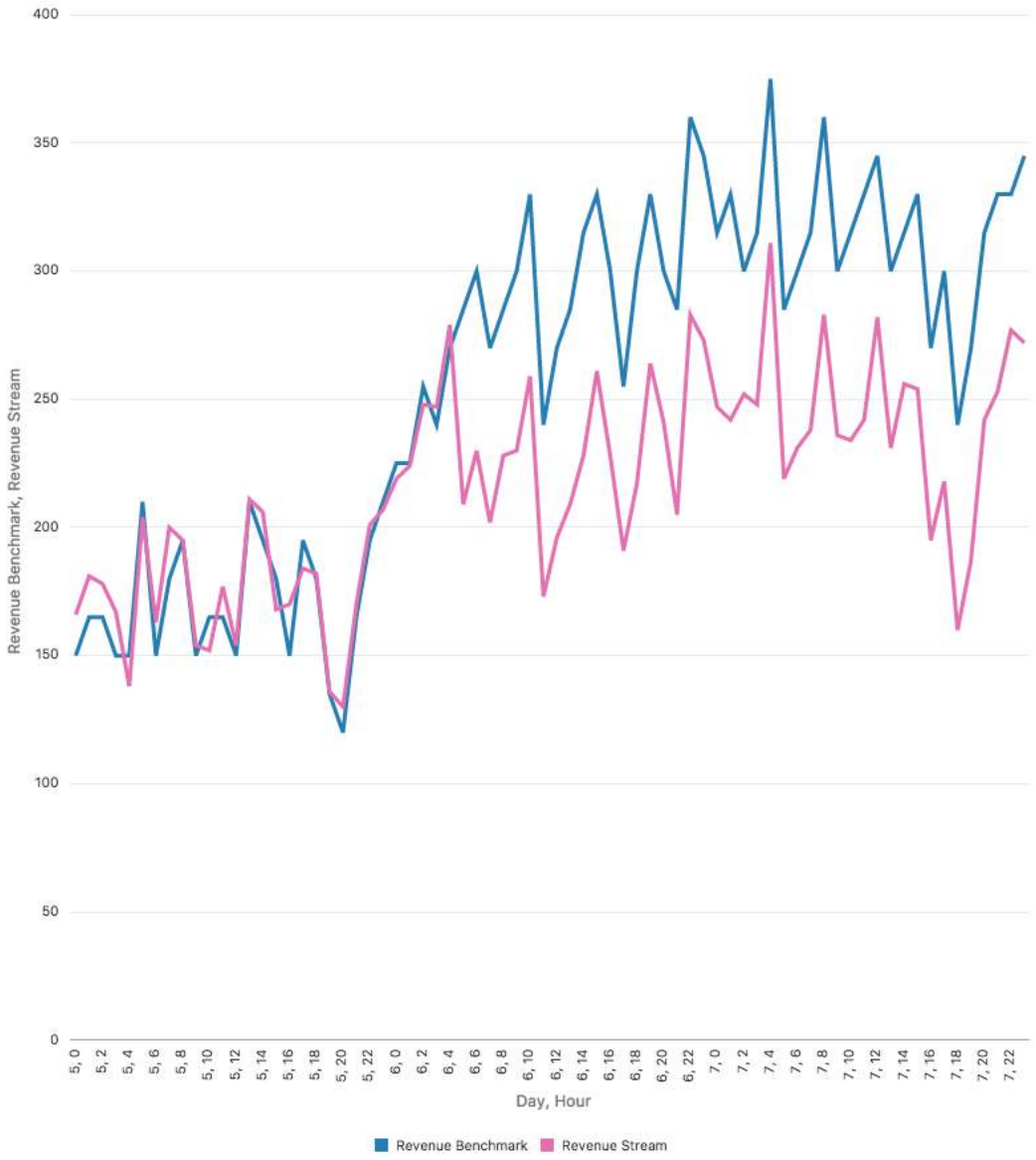
Add Calculation

Streaming Demo
Project

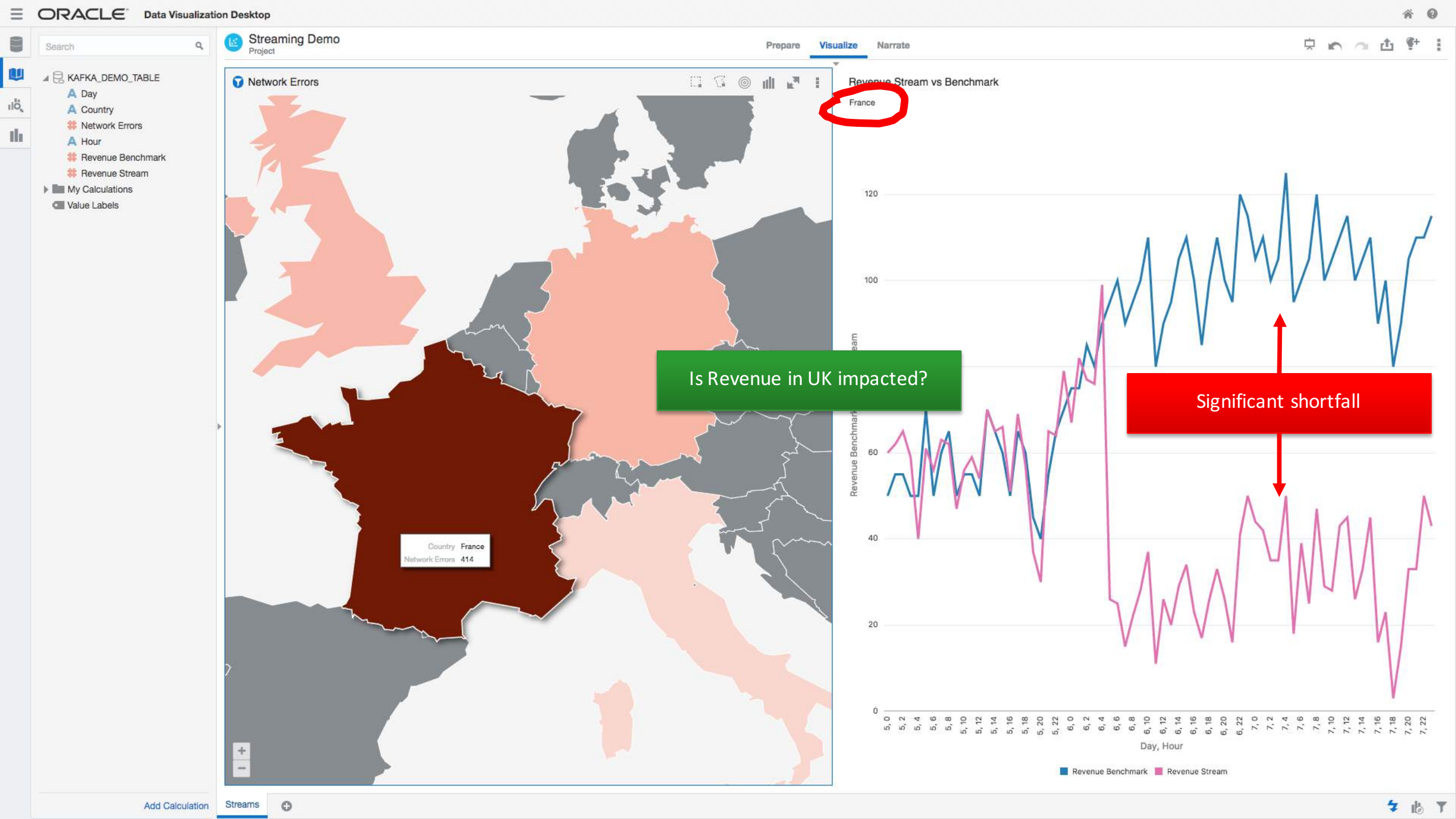
Prepare | Visualize | Narrate

Network Errors

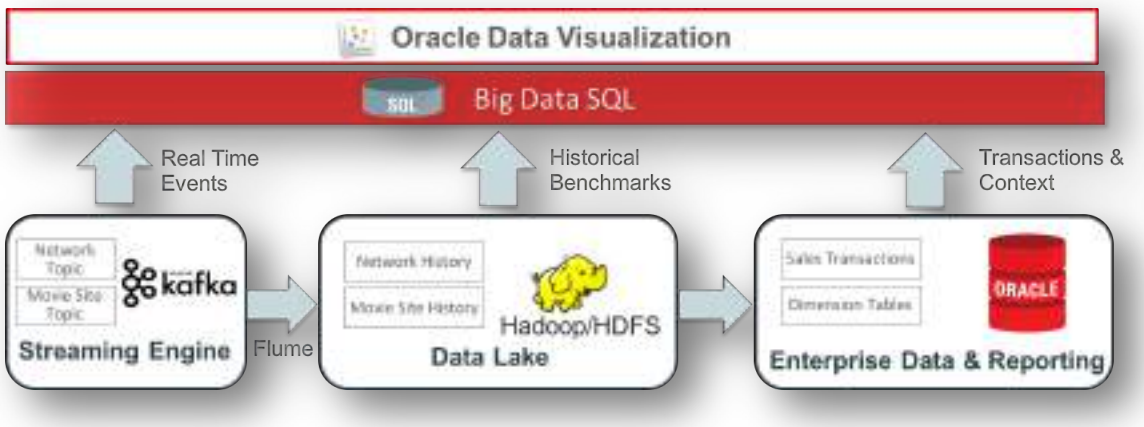
Revenue Stream vs Benchmark



Streams


ORACLE Data Visualization Desktop
Search
KAFKA_DEMO_TABLE
Day
Country
Network Errors
Hour
Revenue Benchmark
Revenue Stream
My Calculations
Value Labels
Add Calculation
Streaming Demo
Project
Prepare
Visualize
Narrate
Network Errors
Country France
Network Errors 414
Is Revenue in UK impacted?
Revenue Stream vs Benchmark
France
Revenue Benchmark
120
100
60
40
20
0
Significant shortfall
Day, Hour
Revenue Benchmark
Revenue Stream
Streams

Insights Achieved with Simplicity

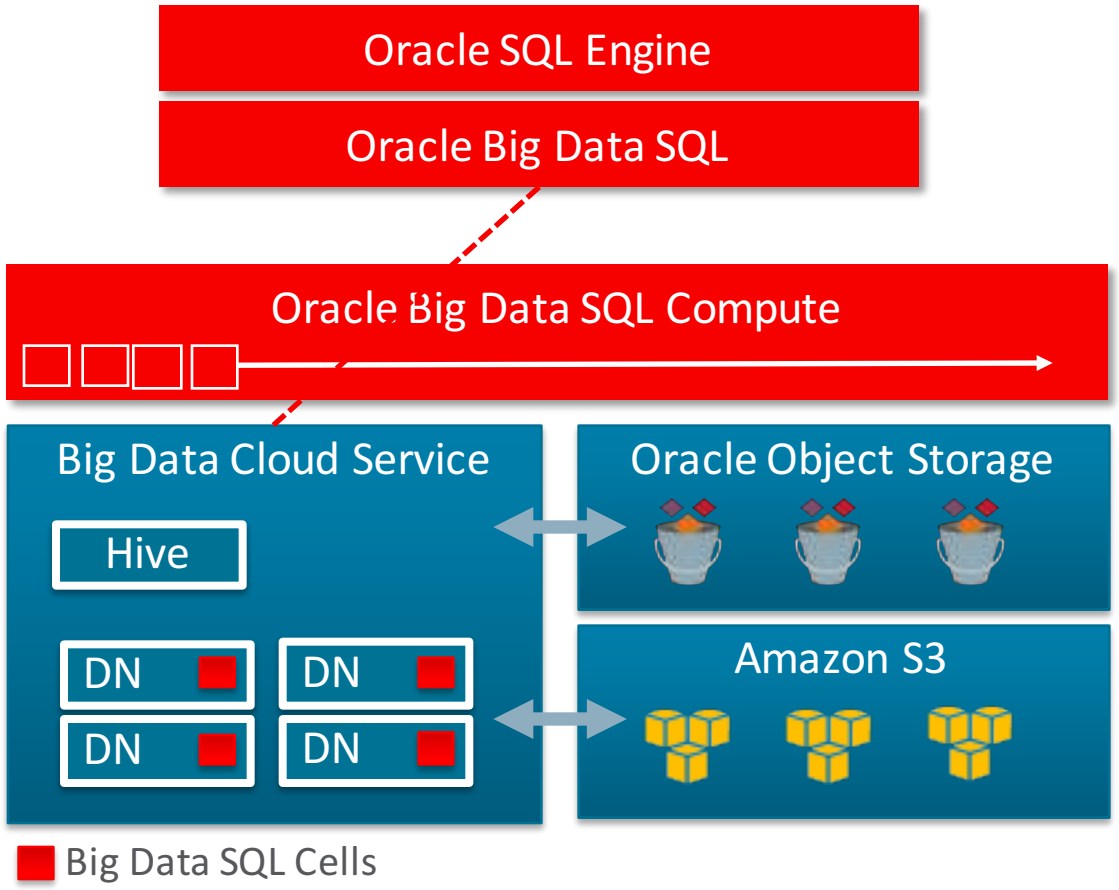


- Easily blend real time streams with history, benchmarks and context
  - Are we running at peak performance?
  - What is the opportunity cost of our current network latency?
- Any application realizes benefit
  - Use Oracle SQL and APIs over all data
- Ensure data is secure
  - Leverage Oracle advanced security

Copyright © 2017, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

Efficiently Query Data Lake

- Data local processing for best performance
- Query Object Store directly without need for Hadoop Cluster
- Elastically scale based on workload
- Optimize ground to cloud



Copyright © 2017, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

Optimized Drivers for Common File Types

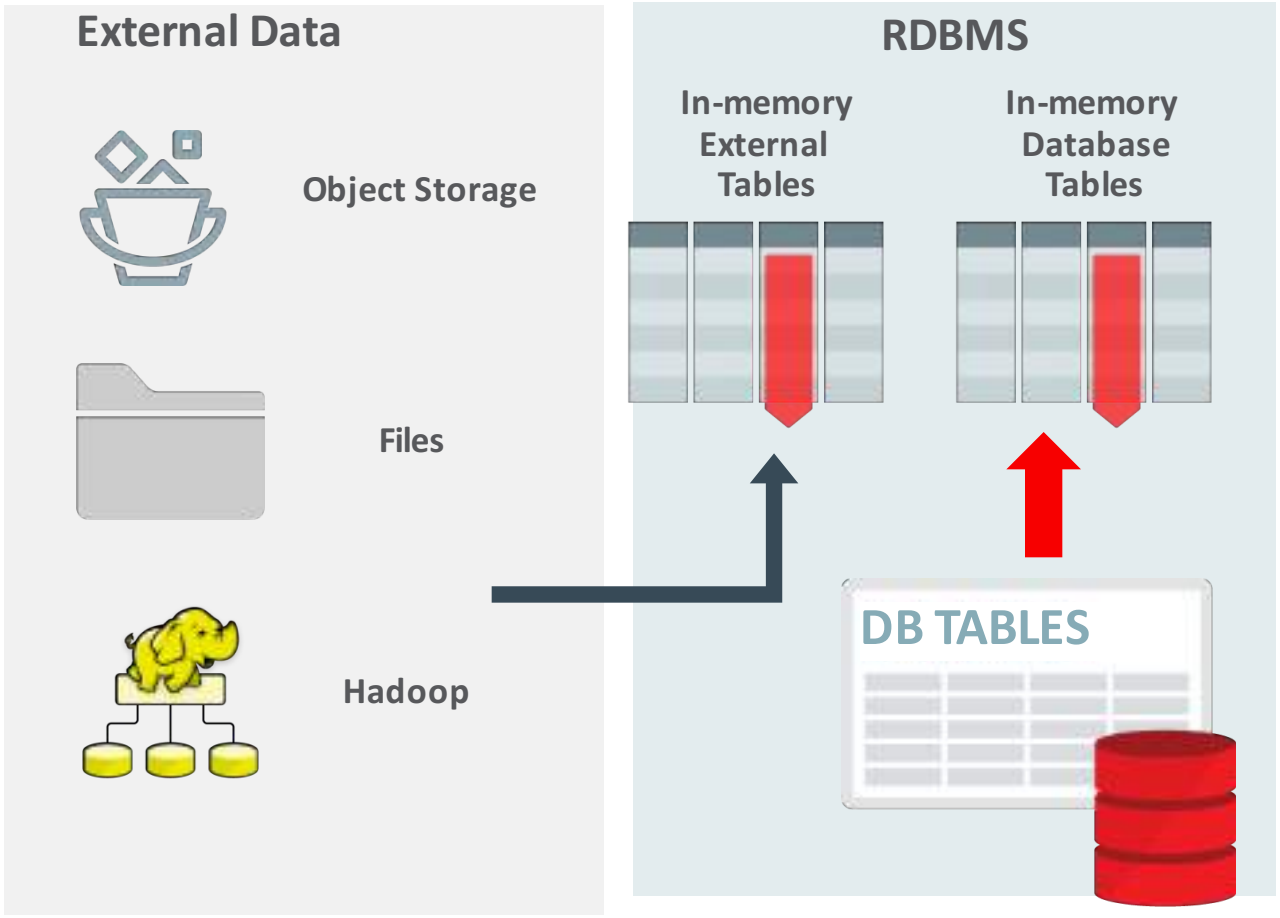
- Continue to push intelligent processing to cluster
  - Use open source when Oracle optimizations unavailable
- Examples:
  - Oracle Parquet Driver
  - C-based filter and parse of CLOBs (e.g. JSON documents)

Oracle Parquet Driver
Up to **8x** faster than
Hive Default Parquet Driver

Copyright © 2017, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

In-Memory and Columnar Processing

Fast Analytics on Big Data Sources



- In-Memory For External Tables builds in-memory column cache of data outside the DB for ultra-fast analytics on external data
- All In-Memory Optimizations apply
  - Vector processing, JSON expressions extend transparently to external data
  - Enhanced distributed processing across Hadoop cluster
- Up to 100X faster

Copyright © 2017, Oracle and/or its affiliates. All rights reserved. |