

ORACLE®

# CON-5465



October 1–5, 2017  
SAN FRANCISCO, CA

## Filling your Data Lake with potable data using Oracle Data Integration

Mike Matthews  
Senior Director, Product Management

Jayant Mahto  
Senior Product Manager

October 2<sup>nd</sup> 2017

**ORACLE**

# 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.

# Oracle Cloud Platform



Develop & Deploy



Integrate & Extend



Publish & Engage



Analyze & Predict



Secure & Manage

Innovate with a  
**Comprehensive, Open,  
Integrated and Hybrid**  
Cloud Platform  
that is  
**Highly Scalable, Secure  
and Globally Available**

# Oracle Cloud Platform

Comprehensive

Open





Integrated

Hybrid

Oracle  
Public Cloud



Oracle  
Data  
Center

-  Data Management
-  Application Development
-  Enterprise Integration
-  Data Integration

-  Analytics and Big Data
-  Content & Experience
-  Identity & Security
-  Systems Management

Oracle Cloud  
at Customer



Your  
Data  
Center

**Built on High Performant Oracle Cloud Infrastructure**

# Oracle Cloud Platform Momentum

**14,000+**

Oracle  
Cloud Platform  
Customers



**3,000+**

Apps in the  
Oracle Cloud  
Marketplace



**\$1.4 Billion**

FY17 Oracle Cloud  
Platform  
Revenue  
(60% YoY Growth)



**10 PaaS**

Categories where  
Oracle is a **Leader**  
According to  
Industry  
Analysts



# Oracle Cloud Platform for Integration

## Application and Data Integration

Complete

Simplified

Open





A large black pipe extends from the left side of the frame, pouring a thick, red, muddy stream into a body of water. The water is a vibrant blue, and the surrounding area is a lush green field. The scene is set against a clear blue sky. The text "Data Lake... or Data Swamp?" is overlaid in white on the image.

# Data Lake... or Data Swamp?



# Key Success Factors for your Data Lake



**Timely** access to data



**Flexibility** to extract and work the data as needed



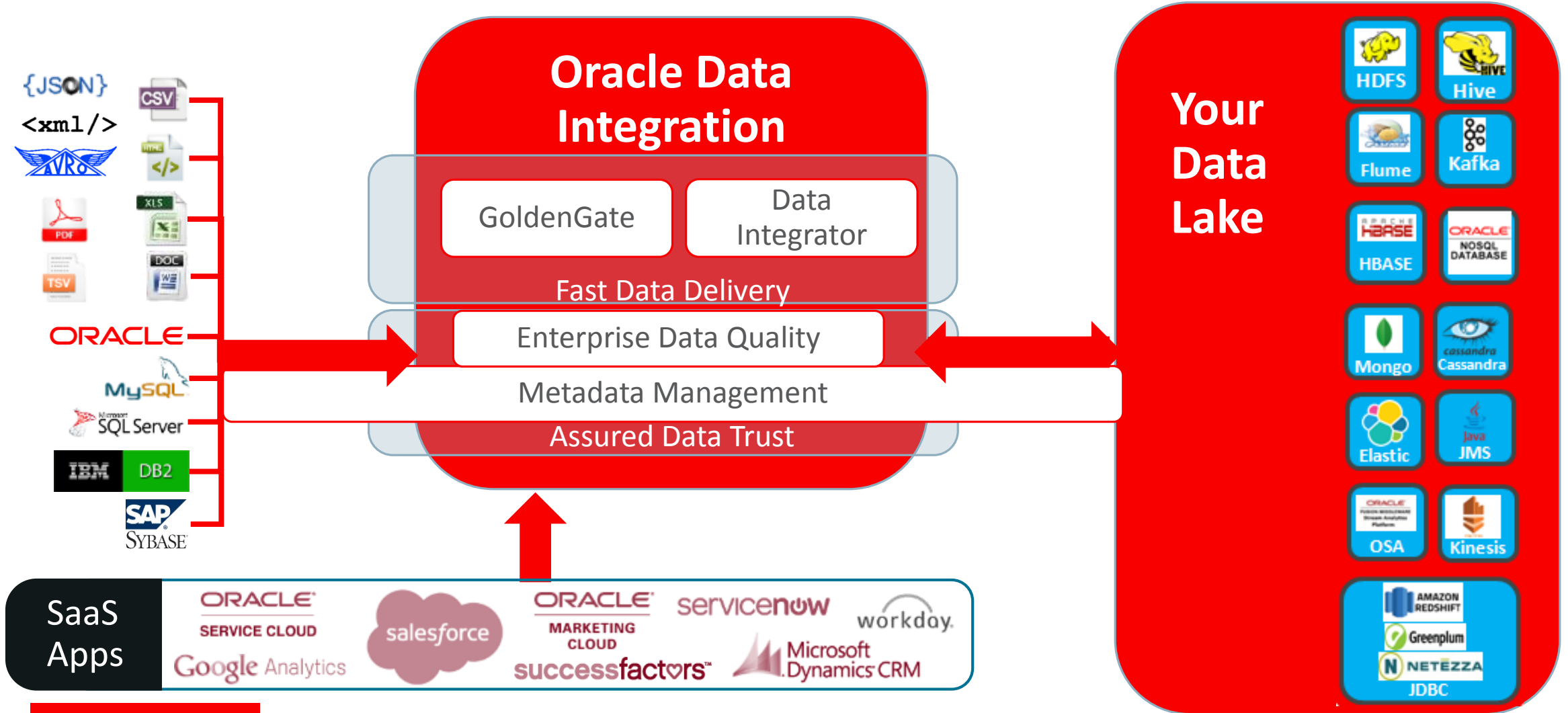
**Trust** in the quality of the data



Ability to **find** and **understand** the available data

Source: **Knowledgent** - <https://knowledgent.com/whitepaper/design-successful-data-lake/>

# Reference Architecture with Oracle Data Integration



# Key Success Factors for your Data Lake



**Timely** access to data



**Flexibility** to extract and work the data as needed



**Trust** in the quality of the data



Ability to **find** and **understand** the available data

Source: **Knowledgent** - <https://knowledgent.com/whitepaper/design-successful-data-lake/>

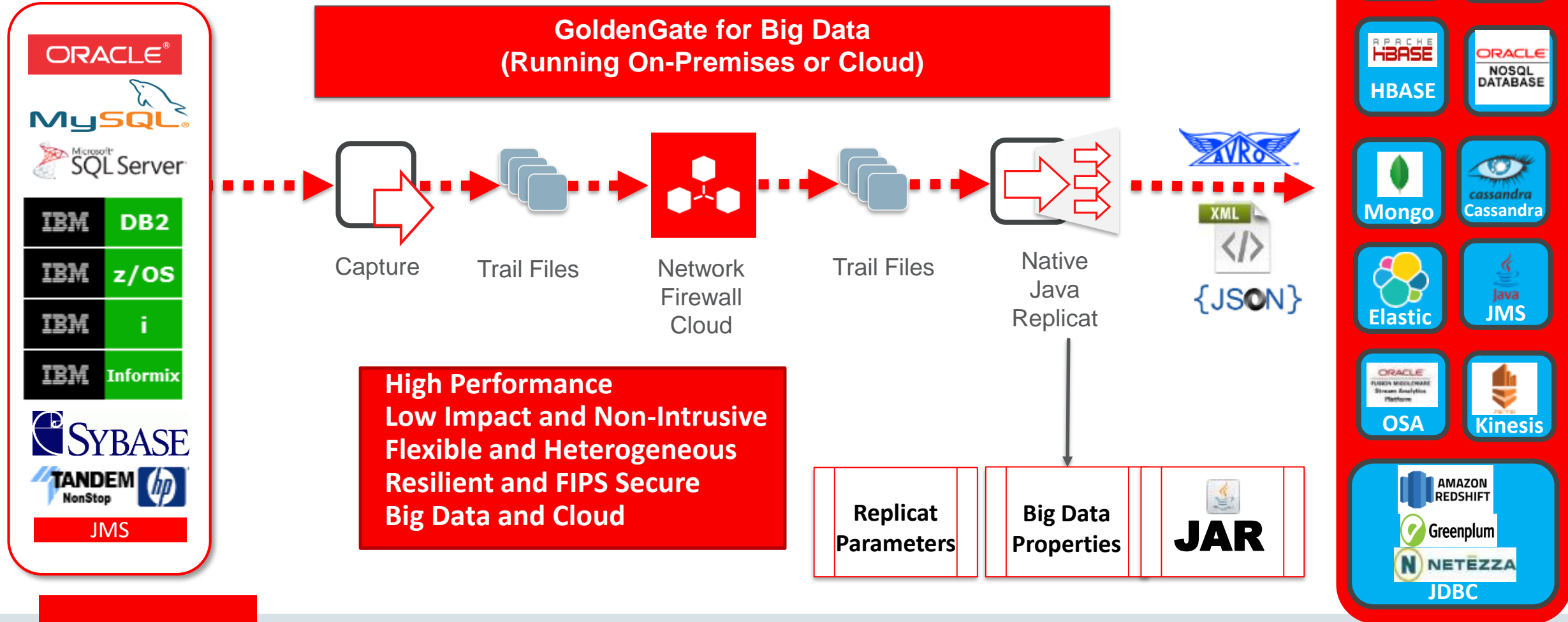
# Why GoldenGate?

- The Sushi Principle – ‘Data is best served raw’
- Some of the biggest data lakes use Oracle GoldenGate’s change data capture capability for real-time ingestion from source databases
- Traditional normalization, aggregation and schematization are skipped to simplify data flows and improve timeliness and performance



# Oracle GoldenGate for Big Data

## Modular & Pluggable Architecture



# Key Success Factors for your Data Lake



**Timely** access to data



**Flexibility** to extract and work the data as needed



**Trust** in the quality of the data



Ability to **find** and **understand** the available data

Source: **Knowledgent** - <https://knowledgent.com/whitepaper/design-successful-data-lake/>



# Continued Focus on Our Vision:

## Integrate Any Data Shape, Speed, Action, Volume & Location

Any Data Shape

Polyglot



Any Data Speed

Lambda



Any Data Action

Dataflow | Pipes



Any Data Volume

Open Source Platforms



Any Data Location

Cloud Infrastructure

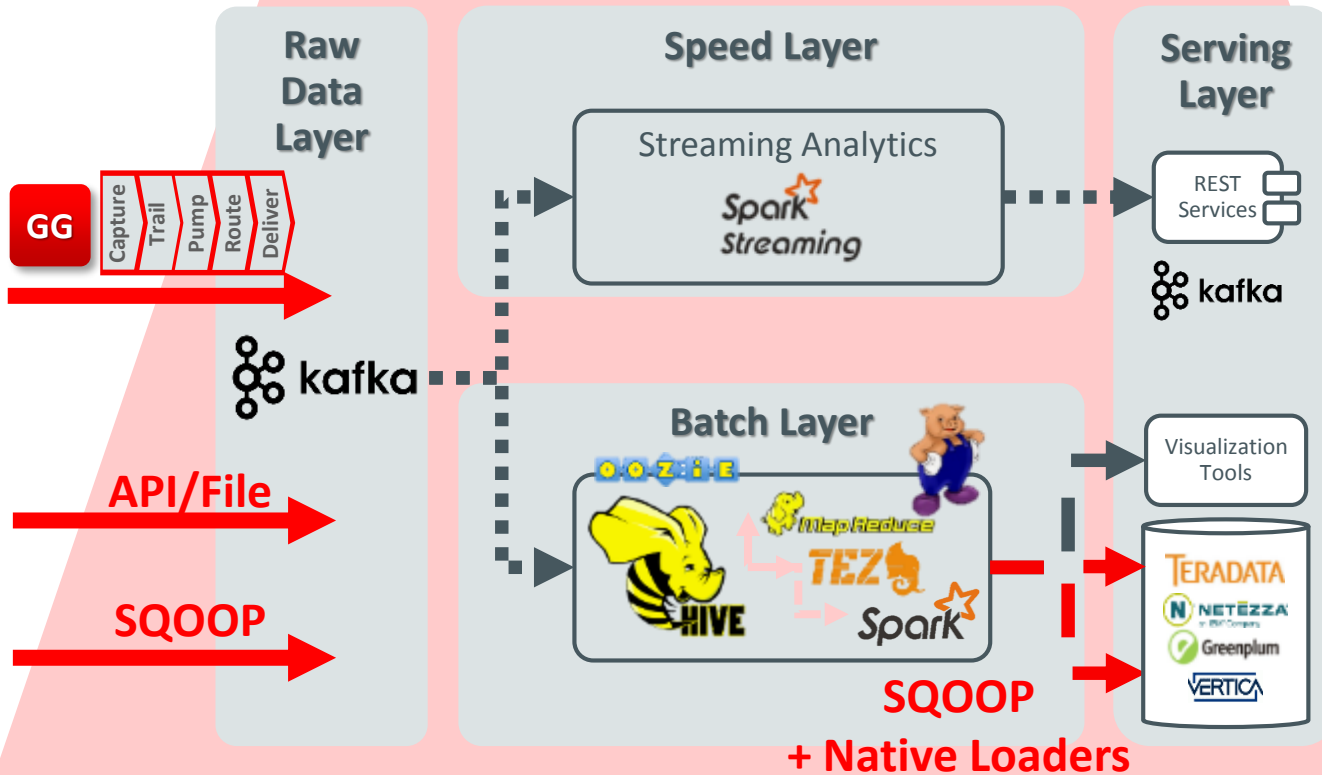


# Why Oracle Data Integrator?

- To provide true analytical flexibility and accuracy, some data re-shaping may be needed, especially as Data Lakes are increasingly working with **Master Data** as well as **Transactional Data**
- ODI's EL-T architecture can be very important when working with large volumes
- This may be done reading from a Data Lake and writing to a Data Warehouse
- ODI can also pushdown data transformations into the Data Lake

# Big Data Transformation with Data Integrator

## Oracle Data Integrator



## Data Integrator for Big Data

- ✓ Batch data ingestion with Sqoop, native loaders & Oozie
- ✓ Generate data transformations in Hive, Pig, Spark & Spark Streaming
- ✓ Extract data into external DBs, Files or Cloud

## Benefits

- ✓ **No ETL Engine** native E-LT execution, 1000s of references
- ✓ **Zero Footprint** does not require any Oracle install on cluster
- ✓ **Loosely Coupled** design time means you can reuse mapping logic in many big data languages

# Key Success Factors for your Data Lake



**Timely** access to data



**Flexibility** to extract and work the data as needed



**Trust** in the quality of the data



Ability to **find** and **understand** the available data

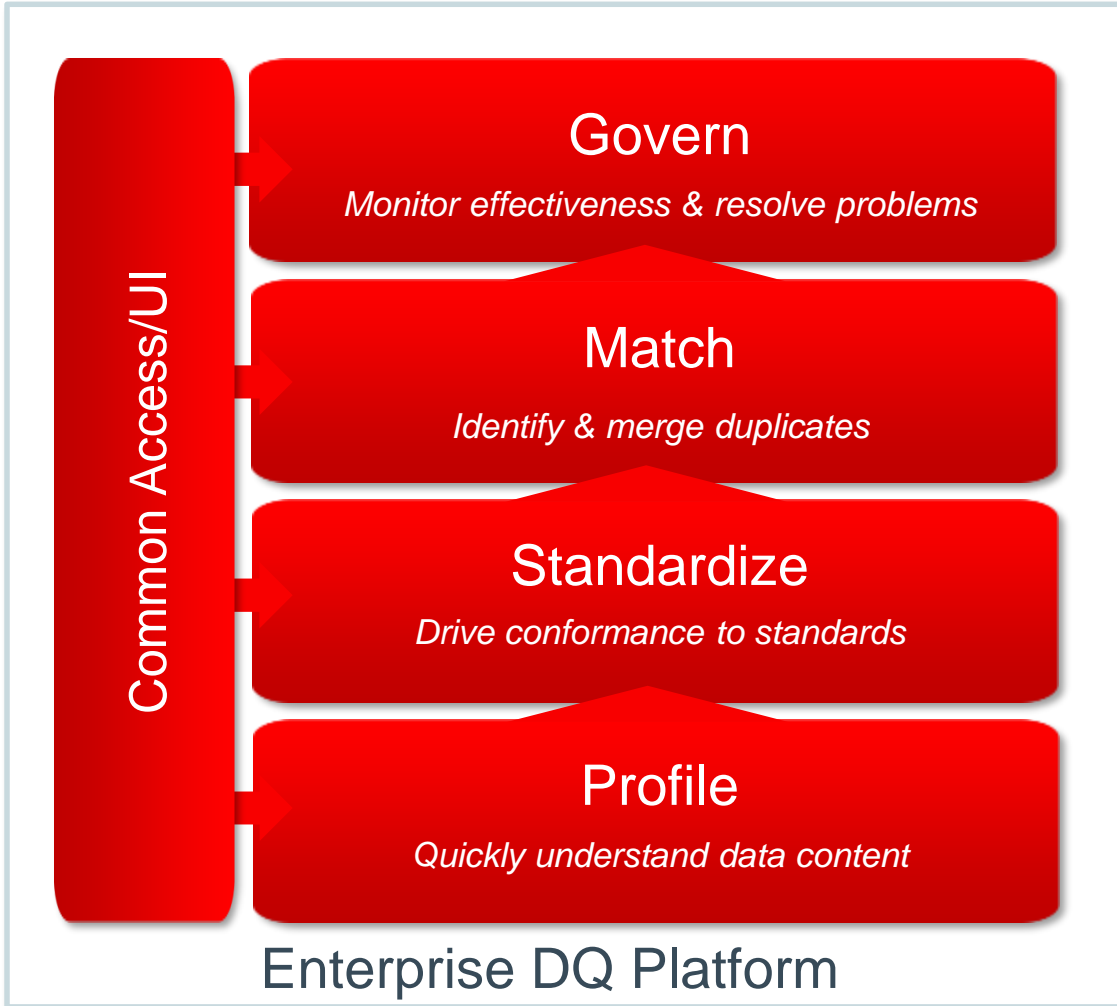
Source: **Knowledgent** - <https://knowledgent.com/whitepaper/design-successful-data-lake/>

# Some data can only be trusted if it is prepared

- Data Consumers need access to **Master Data** as well as **Transactional Data**
- Relating the two can be very powerful...
- ... but this is where raw data can be poisonous to strong business analytics
  - **Incomplete records**
  - **Hard-to-find Duplicates**
  - **Out-of-date information**
  - **Inconsistencies in data capture**



# Why Oracle Enterprise Data Quality?



Market-leading usability for all types of data



Unparalleled time-to-value



High performance engine



Out-of-the-box global knowledge-base



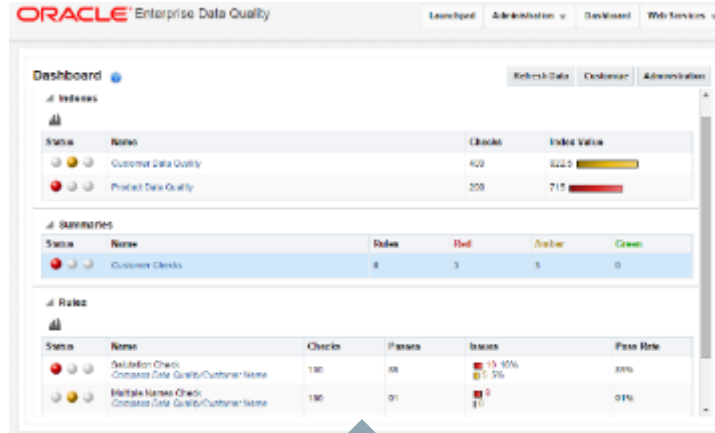
Foundation for governance program



# EDQ · Collaborative Data Quality Governance

## Data Analysts

- Immediate Data Insight
- Reusable DQ Services and Rules
- Transparent, self-documenting configuration

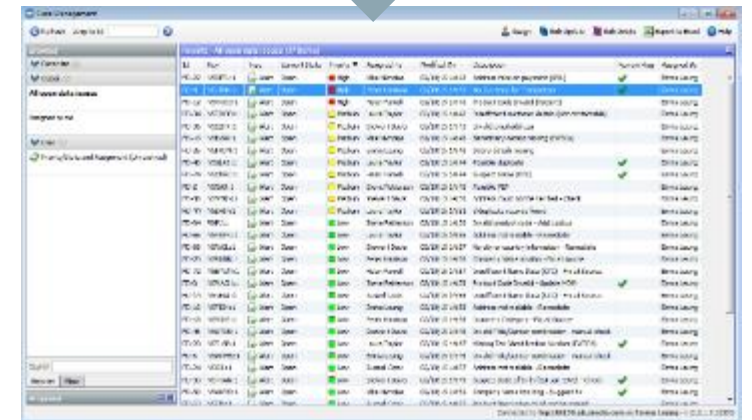
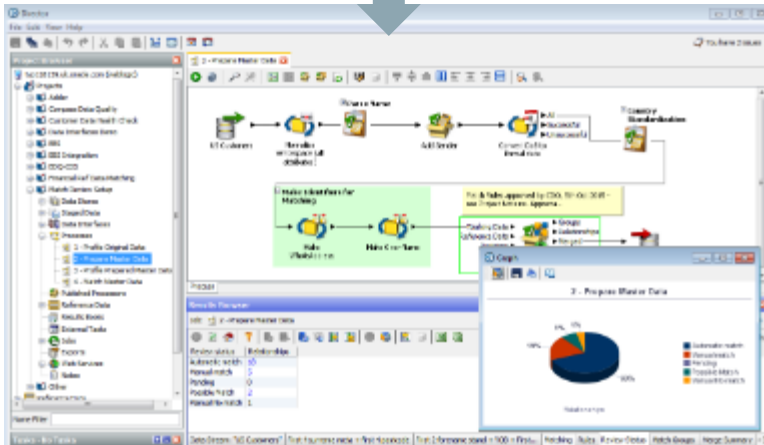


## Data Stewards

- Flexible Data Review and Remediation options in EDQ Case Management
- Integrated with DQ Rules
- Fully audited with comments, attachments, history, reports

## Data Stakeholders

- Zero Training EDQ Dashboard
- View by Data Asset, Data Domain, Rule
- Trend Analysis



# Key Success Factors for your Data Lake



**Timely** access to data



**Flexibility** to extract and work the data as needed



**Trust** in the quality of the data



Ability to **find** and **understand** the available data

Source: **Knowledgent** - <https://knowledgent.com/whitepaper/design-successful-data-lake/>

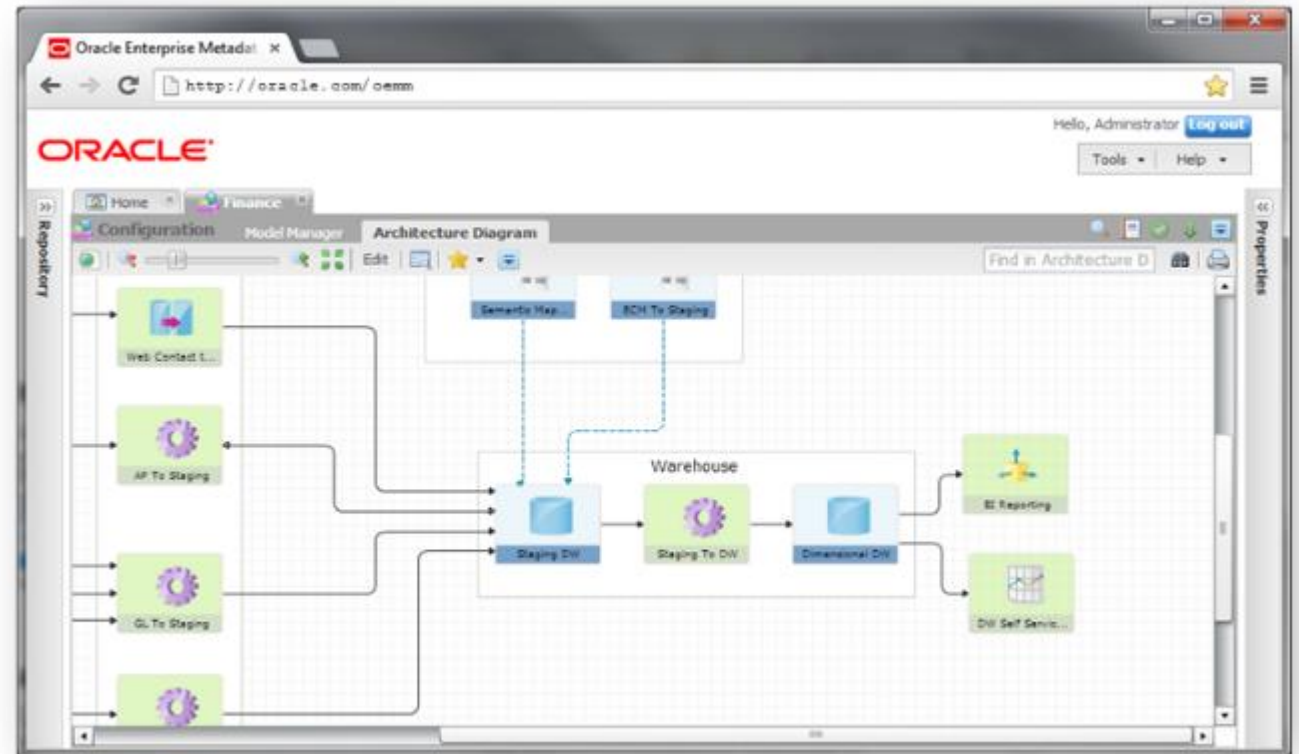
# Why Metadata Management for the Data Lake?

## Without Metadata Management

- x Silos of Data known only to their owners
- x No documentation
- x Duplicate effort and inefficient usage
- x No data usage analysis

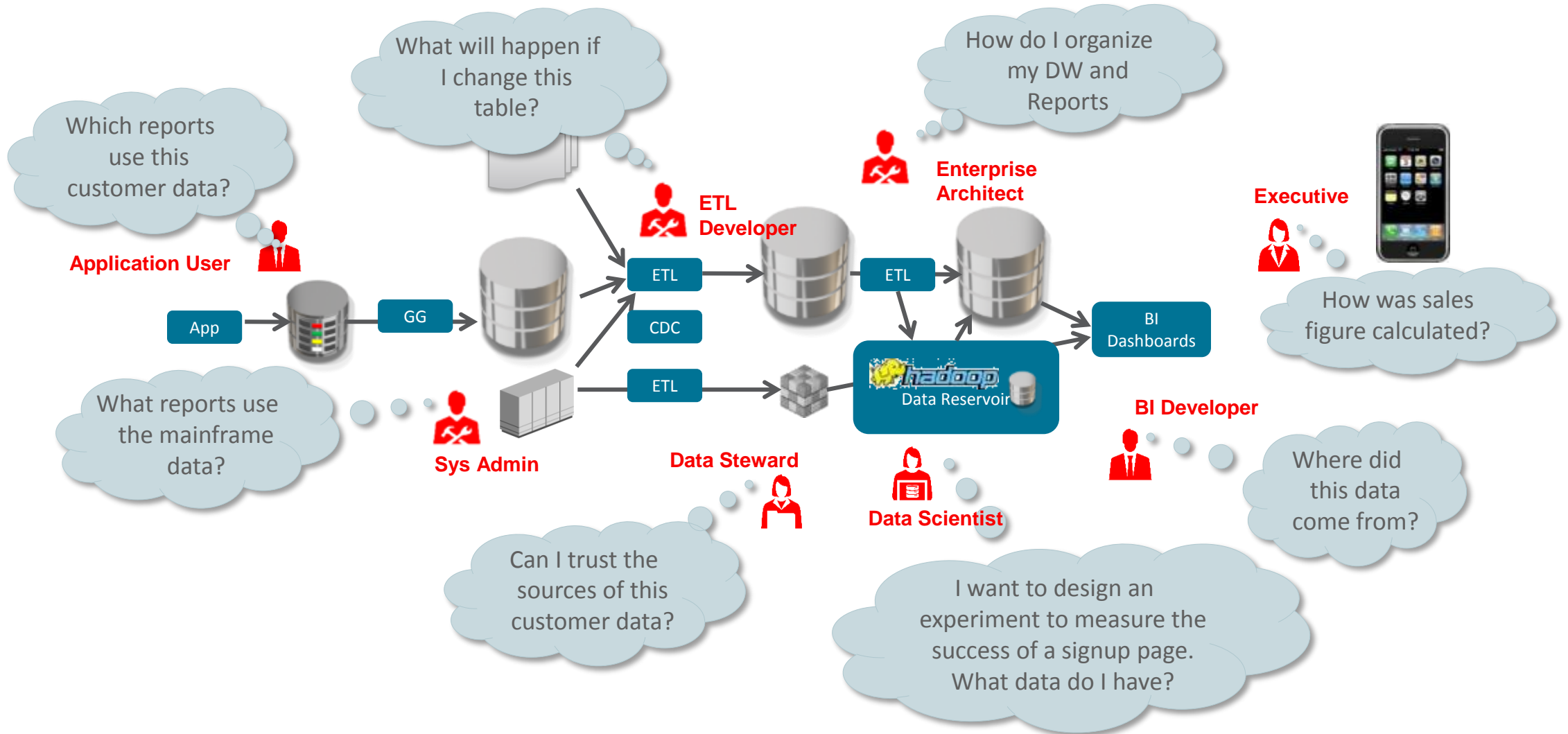
## With Metadata Management:

- ✓ Searchable
- ✓ Enriched with documentation
- ✓ Shared knowledge
- ✓ Lineage/impact analysis
- ✓ Semantic analysis



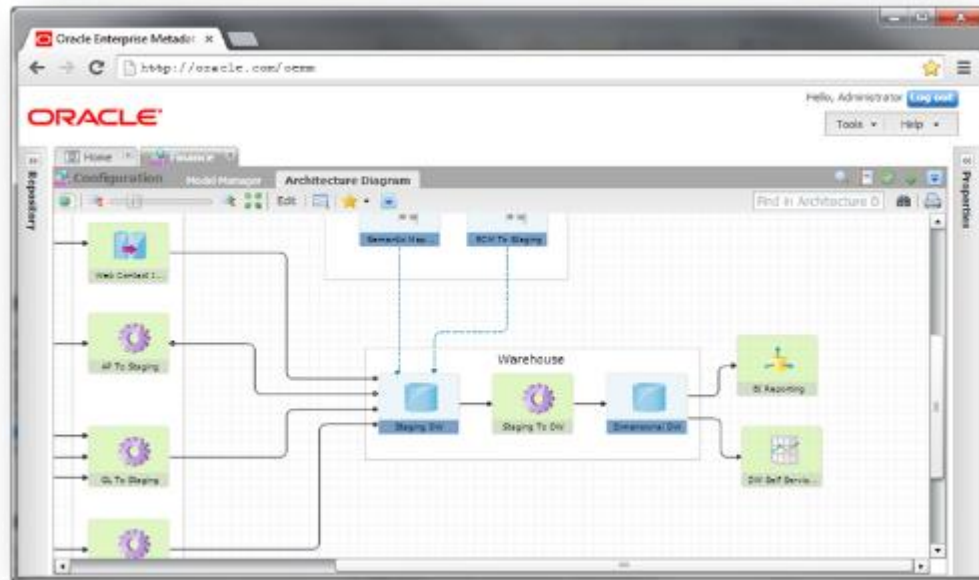
# Value of Enterprise Metadata Management

**Solves significant pain points for wide variety of business consumers and technical staff**



# Find and Understand your Data

- **Metadata Management** – horizontal and semantic data lineage for all data sources
- **Business Glossary** – simple tools to catalog, link and collaborate on business terms



- ✓ Business Data Catalog
- ✓ Report to Source Lineage
- ✓ Impact Analysis
- ✓ Audit, Versioning & Diff Reports
- ✓ Social/Collaboration Features
- ✓ Annotations and Tagging
- ✓ Comprehensive Harvesting
  - ✓ 3<sup>rd</sup> Party BI Metadata
  - ✓ 3<sup>rd</sup> Party ETL Metadata
  - ✓ 3<sup>rd</sup> Party DB Metadata
  - ✓ 3<sup>rd</sup> Party Modeling Tools
  - ✓ Big Data Metadata
  - ✓ Metadata Standards

# What does **Potable Data** mean?

- **Quickly and Easily Consumable** and **Trusted**
  - You can use GoldenGate to make data more quickly available, streamed into (and through) the Lake using CDC
  - You can use ODI to make the data easier to consume
  - Trust is not only about 'how good it is', but **knowing** how good it is (or not), and **where it came from**
  - You can use EDQ to add Data Quality dimensions to your data as it is streamed into the Lake...and the analytics tools you already use to tell you how good the data is
  - You can use OEMM to **understand** the data, and **where it comes from**



# Get a sneak peek at cutting-edge data integration designs and receive a free gift!

- Oracle is constantly developing new software and features that will make your work easier, and Oracle's User Experience team would love to get your feedback on new data integration designs.
- Feedback sessions will take place at a date and time of your own choice.
- You can take part via webconference, from the comfort and convenience of your own office.
- If you're interested, please fill out the 1-page form at <http://bit.ly/2vIH1Sg>  
uppercase I lowercase l
- To show our appreciation, we will post all participants their choice from a wide selection of thank-you gifts.

# Data Integration Programme – [FOCUS ON DOC LINK](#)

## Presentations on:

Oracle  
Data Integration  
Platform Cloud

Oracle  
Data Integrator

Oracle  
GoldenGate

Oracle  
Enterprise  
Data Quality

Oracle  
Enterprise  
Metadata  
Management

## Hands-on Labs:

Oracle GoldenGate  
Real-Time Data Replication  
in the Cloud  
HOL7715

Oracle  
Enterprise Data Quality  
HOL7653

ODI and OGG  
for Big Data  
HOL7708

Oracle  
Data Integration  
Platform Cloud  
HOL7673

## Demo Stations:

The EXchange  
Integration Area  
- Moscone West

The EXchange  
Data Management Area  
- Moscone West

The EXchange  
Analytics & Big Data Area  
- Moscone West

# Data Integration Programme – FOCUS ON DOC LINK

## Sunday, October 1

- Lift and Shift Workloads to Cloud with Oracle Data Integration Platform Cloud [SUN6653]
- Data Movement between On-Prem, Fusion ERP Cloud, Fusion HCM Cloud and Salesforce [SUN7286]
- Accelerate Migration to Cloud Infrastructure with Data Integration Platform [SUN6896]

## Monday, October 2

- Oracle Data Integration Platform Strategy and Roadmap [CON6646]
- **Filling Your Data Lake with Potable Data, Using Data Integration [CON5465]**
- GoldenGate : Deep Dive into Automating OGG using the new Microservices [CON6569]
- Oracle Data Integration Platform: Foundation for Cloud Integration [CON6650]
- Oracle Data Integration Platform Empowers Enterprise Grade Big Data Solutions [CON6893]
- Oracle Data Integration Platform Cloud Deep Dive [CON6651]
- Oracle GoldenGate Cloud Service: Real-Time Data Replication in the Cloud [HOL7715]

## Tuesday, October 3

- Oracle Data Integrator Product Update and Strategy [CON6654]
- Oracle Enterprise Data Quality: Product Overview and Roadmap [CON6656]
- Accelerate Cloud On-Boarding Using Oracle GoldenGate Cloud Service [CON6894]
- Oracle Enterprise Data Quality for All Types of Data [HOL7653]
- Oracle Data Integration Platform: a Cornerstone for Big Data [CON6655]
- GoldenGate: MAA and Best Practices for Oracle GoldenGate Microservices [CON6570]
- Oracle GoldenGate Product Update and Strategy [CON6897]

## Wednesday, October 4

- A Practical Path to Enterprise Data Governance with Oracle Enterprise Data Quality [CON6657]
- Oracle Data Integrator and Oracle GoldenGate for Big Data [HOL7708]
- Introduction to Oracle Data Integration Platform Cloud [HOL7673]
- An Enterprise Databus: GoldenGate in the Cloud Working with Kafka and Spark (CON6895)
- GoldenGate: Best Practices & Deep Dive on OGG 12.3 Microservices at Cloud [CON6568]
- Oracle GoldenGate for Big Data [CON6898]
- Oracle Data Integration Platform Cloud Service Governance Edition [CON6652]

# Connect with Oracle Integration



**Oracle Data Integration**



**@OracleDI**



**[Blogs.oracle.com/DataIntegration/](https://blogs.oracle.com/DataIntegration/)**



**Oracle Data Integration**



**Oracle FMW**



**@OracleIntegrate**



**[Blogs.oracle.com/Integration/](https://blogs.oracle.com/Integration/)**



**Oracle SOA**

# Stay Informed During and After OpenWorld



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



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



# Converged Infrastructure Forum

*Tuesday, Oct 3 from 6:30-9pm*

*SF MOMA*

*RSVP Required:*

<https://www.oracle.com/goto/Openworld/CIEventOct3>





ORACLE®

# Integrated Cloud

## Applications & Platform Services