

Oracle Cloud Infrastructure Data Catalog

**Data discovery and unified metadata
management in Oracle Cloud for the
modern data warehouse**





Making better use of data than ever before

As organizations embark on their cloud-centric analytics journeys with Oracle Cloud, they have a variety of data sources and services such as data lakes, data warehouses, analytics, and data science.

The Modern Data Warehouse Platform

Oracle Apps,
Fusion SaaS, NetSuite, EBS,
Peoplesoft, JDE, SAP,
Salesforce, Workday



Oracle Azure SQL,
CosmosDB...
AWS DynamoDB, RDS...
Google Bigtable, Firebase...



Oracle
Data Cloud



Logs Webclicks



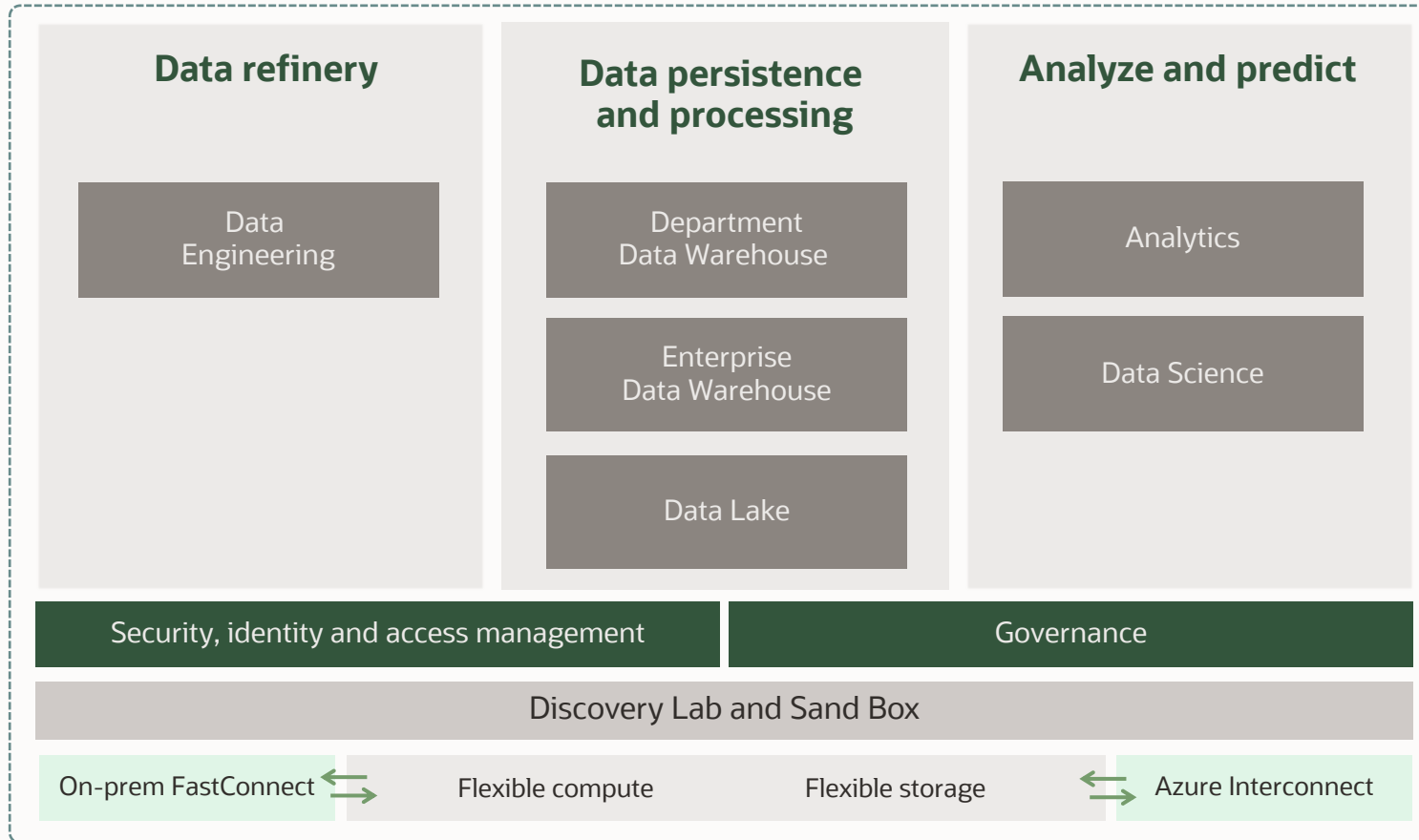
Events Streams



Media



Files Object Stores



Oracle Apps
Data Warehouse

Augmented
analytics

Dashboards
and reports

Machine Learning models

Data-driven
Apps

AI enabled
devices

Data
monetization

Discover

Ingest

Transform

Curate

Analyze, learn and predict

Measure and act



The challenges in gaining even more value from data

When data is spread across multiple sources in Oracle Cloud and on-premises, it becomes difficult for data producers and data consumers to understand what's available and derive value from that data.

Here's what makes it all so challenging:

Difficulty in finding the right data for analytics

- Lack of holistic view of data assets
- Reliance on tribal knowledge
- Business context of data is not easily available in consuming applications

No shared metadata for data lakes

- No data dictionary
- Manual schema definition
- Inability to share data models across applications

No support for data governance

- Unclear data ownership
- Lack of common business concepts
- Proliferation of sensitive data

The solution

A data catalog for discovery and unified metadata management in Oracle Cloud

Oracle Cloud Infrastructure (OCI) Data Catalog helps data professionals find data in Oracle Cloud and beyond by using a data asset inventory based on technical and business metadata with automated harvesting, a business glossary, and metadata curation. As a data catalog optimized for Oracle, it enables Oracle customers to gain more value from data and workloads in Oracle Cloud.



What can OCI Data Catalog do?



Metadata harvesting

- Searchable data asset inventory
- OCI Object Storage, Oracle Autonomous Database
- Oracle Database, MySQL, Hive, and Kafka on OCI and on-prem



Metadata curation

- Business glossaries with terms and categories
- Tags for annotations
- Link assets to business terms, tags



Search and browse

- Collaborative environment
- Search based on technical names, business terms, tags
- Browse based on data assets hierarchy



Optimized for Oracle Cloud

- Secure, scalable, serverless cloud-native
- REST APIs and SDKs in Java, Python, Ruby, and Go
- IAM-based policy management

What can you use OCI Data Catalog for?

OCI Data Catalog use cases

For data discovery

Quickly find data across OCI Object Storage, Oracle Autonomous Data Warehouse, and Oracle databases for analytics

For data governance

Manage a business glossary and associated technical metadata to help enable data governance

OCI Data Catalog

is a key component of your data management platform



How does OCI Data Catalog work?

Metadata – data about your data – is the key to OCI Data Catalog

OCI Data Catalog extracts, standardizes, and indexes technical metadata from connected data sources to create a trusted and searchable data asset inventory.

OCI Data Catalog also allows end users and subject matter experts to contribute their domain knowledge about data in the form of user annotations, tags, classifications, and business context. Data stewards can manage the organization's vocabulary in the form of a glossary and then establish links to technical metadata to provide a holistic view.

The metadata, combined with data management and search tools, is what helps data users find the data they need, discover information on available data, and gain information about the trustworthiness of data for different uses.

For an effective data catalog, you need to manage and store:

- **Technical metadata:** collected from enterprise systems
- **Business metadata:** contributed by users as annotation or business context
- **Operational metadata:** indicates data freshness and data usage, and connects everything together in a meaningful way

Oracle Cloud Infrastructure Data Catalog at a glance

Key features for self-service data discovery

OCI Data Catalog offers

Metadata harvesting from OCI sources

Sources supported on OCI

- Object Storage (CSV, ORC, Avro, Parquet, JSON, XLSX)
- Oracle Autonomous Transaction Processing and Oracle Autonomous Data Warehouse

Source supported on OCI and also on-premises

- Oracle Database
- MySQL
- Hive
- Kafka

Supported file types for Object Storage

- CSV, Excel
- ORC, Avro, Parquet
- JSON

Harvesting can be done on demand or on a set schedule

The screenshot displays the OCI Data Catalog interface. At the top, there is a 'Create Data Asset' button and a search bar. Below is a table of data assets:

Type	Name	Owner	Last Updated
<input type="checkbox"/>	aguij_oradb_asset	SystemAdmin	08-Sep-2019 20:04:46
<input type="checkbox"/>	DemoDB_dfml		
<input type="checkbox"/>	DSCI_WB_Hive		
<input type="checkbox"/>	Kafka_IOT_STR		
<input type="checkbox"/>	lastname_oradb_asset		
<input type="checkbox"/>	M		
<input type="checkbox"/>	o		

A modal window is open over the table, showing a three-step process: 'Select Connection', 'Select Entities', and 'Create Job'. The 'Select Entities' step is active, displaying 'Available Data Entities' (ALPHA, ANONYMOUS) and 'Selected Data Entities' (ALPHA, ORACLE SCHEMA).

Below the modal, a job completion summary is shown for 'Job : alpha_harvest - 10-Sep-2019 23:21:21'. The status is 'Succeeded'. The summary includes the following metrics:

Metric	Value
ENTITIES HARVESTED	32
ELAPSED SECS	1
FOLDERS HARVESTED	1



OCI Data Catalog offers

Built-in business glossary and metadata enrichment

- Manage a business glossary to help with building a vocabulary or business concepts taxonomy—the first step towards better data governance
- Define business terms and hierarchical categories
- Make use of free-form tags for user annotations
- Use assets links to business terms and tags to provide a holistic view of the data

The screenshot displays the OCI Data Catalog interface. On the left, a 'Hierarchy' tree shows the following structure:

- Finance Sample
 - CustomerCategory
 - Account Manager ID
 - Address
 - Address Line (highlighted)
 - City
 - Postal Code
 - State or Province

On the right, the 'Term: Address Line' details are shown, including 'Edit', 'Create Term', and 'Actions' buttons, and tabs for 'Summary', 'Linked Terms', and 'Linked Objects'. The 'Summary' tab is active, showing:

- Name: Address Line
- Owner:
- Status: ● New

In the foreground, a 'Link Term' dialog box is open, featuring a search bar and a table of terms:

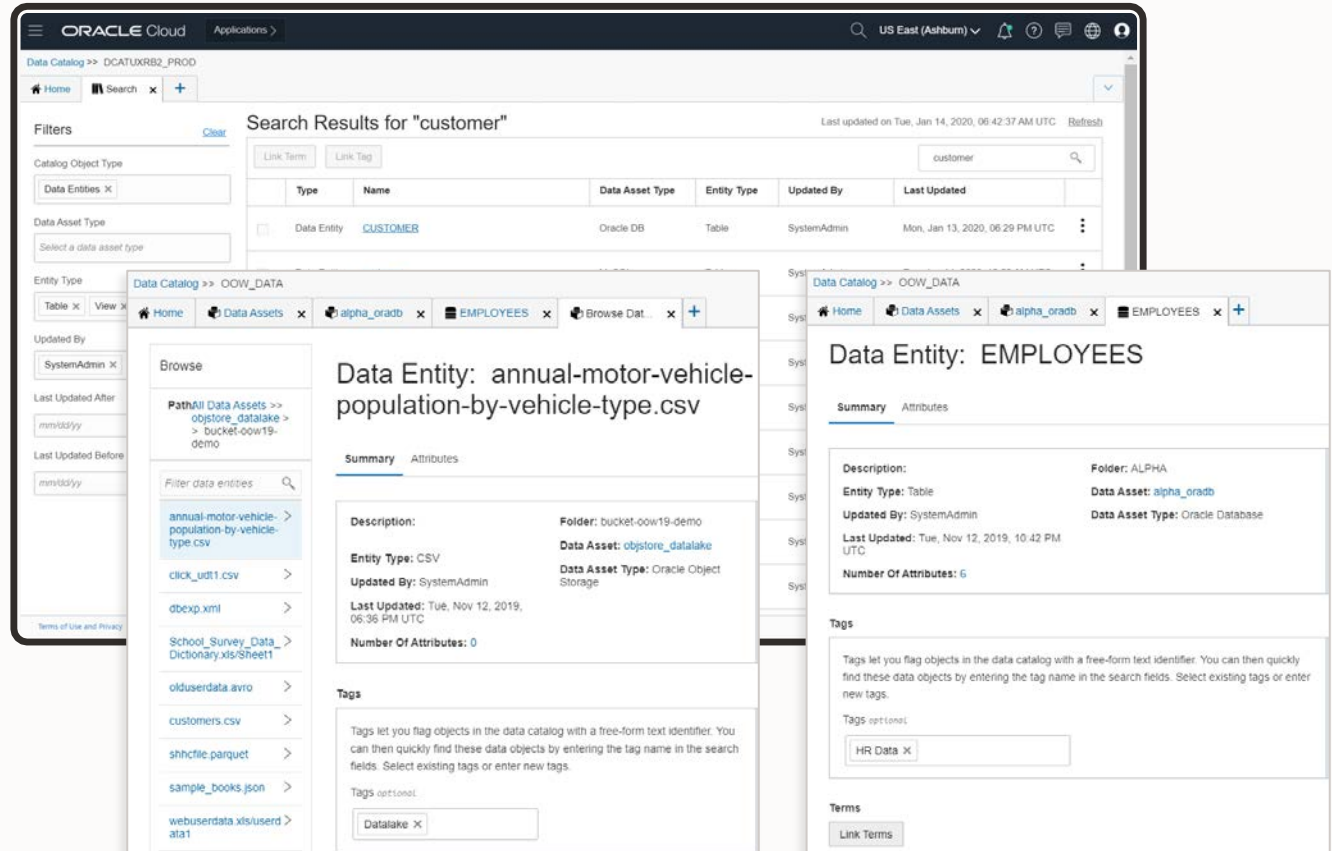
	Name	Glossary	Description
<input type="checkbox"/>	Account Manager ID	Finance Sample	Employee account..
<input type="checkbox"/>	Address Line	Finance Sample	Address Line typically contains street address. Ex..
<input type="checkbox"/>	City	Finance Sample	A city is a large human settlement.[4][5] Cities g..
<input type="checkbox"/>	Customer	Finance Sample	An ultimate customer may be a consumer as well, bu..
<input type="checkbox"/>	Date of Birth	Finance Sample	Date of Birth Needs proof..
<input type="checkbox"/>	Family Name	Finance Sample	A surname, family name, or last name is the portio..
<input type="checkbox"/>	Gender	Finance Sample	Customer Gender as identified by the customer..



OCI Data Catalog offers

Search and exploration

- Search data based on technical names, business terms, or tags
- View details of various objects
- Browse OCI Data Catalog based on data assets



OCI Data Catalog offers

Single collaborative environment for all users

- Homepage with helpful shortcuts and operational stats
- Quick Actions to manage data assets, glossaries, jobs, and schedules
- Popular tags and recently updated objects

The screenshot displays the Oracle Cloud Data Catalog interface. At the top, the Oracle Cloud logo and 'Applications' menu are visible. The breadcrumb path is 'Data Catalog >> OOW_DATA'. A search bar is present with the text 'Search within Data Catalog' and a 'Search' button. Below the search bar, it indicates 'Approximately 14.9K objects as of Wed, Jan 15, 2020, 03:58:31 PM UTC'. A navigation bar shows counts for various categories: Data Assets (11), Data Entities (2290), Attributes (12588), Glossaries (3), Terms (85), Jobs (27), and Schedules (1).

The main content area is divided into several sections:

- Learn:** Includes links for 'Data Catalog Overview', 'Product Tour', 'User Guide', and 'Rest API'.
- Data Assets by Type:** A pie chart showing the distribution of data assets by type. The legend indicates: Oracle DB (5) at 45.45%, MySQL (1) at 9.09%, Obj Store (2) at 9.09%, Hive (1) at 9.09%, Kafka (1) at 9.09%, and ATP (1) at 9.09%.
- Popular Catalog Tags:** A word cloud of tags including 'Clickstream', 'Sensitive', 'Sales', 'Customer', 'HR Data', 'Revenue', 'Datalake', 'Marketing', 'Order', 'Store', 'Library', 'cost', 'School', 'Product', and 'Official'.
- Quick Actions:** Includes 'Browse Data Assets' and 'Create Data Asset'.
- Recently Updated:** A table listing recently updated items.

Type	Name	Updated By	Last Updated
Glossary	Customer Order Sample	SystemAdmin	Mon, Jan 13, 2020, 05:32 AM UTC



OCI Data Catalog

Optimized for Oracle Cloud



Native OCI service

Secure, reliable, scalable serverless fully managed on Oracle Cloud



REST APIs and SDKs

Easily integrate OCI Data Catalog capabilities in other applications and services



Policy-based access

Manage access based on OCI IAM user groups



Get started today!

Sign up for a free trial and get started by visiting the Oracle Cloud Infrastructure Data Catalog webpage

oracle.com/big-data/data-catalog

ORACLE
Cloud Infrastructure

Copyright © 2020, Oracle and/or its affiliates. All rights reserved.
Oracle and Java are registered trademarks of Oracle and/or its affiliates.
Other names may be trademarks of their respective owners.

