Oracle Cloud Infrastructure
Data Catalog

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

Making better use of data than ever before
The Modern Data Warehouse Platform

Data refinery
- Data Engineering

Data persistence and processing
- Department Data Warehouse
- Enterprise Data Warehouse
- Data Lake

Analyze and predict
- Analytics
- Data Science

Security, identity and access management
Governance

Discovery Lab and Sand Box
- On-prem FastConnect
- Flexible compute
- Flexible storage
- Azure Interconnect

Discover Ingest Transform Curate Analyze, learn and predict Measure and act

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

Oracle Apps, Oracle Data Warehouse, Augmented analytics, Dashboards and reports, Machine Learning models, Data-driven Apps, AI enabled devices, Data monetization.
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
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
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
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