Data Catalog

A self-service data discovery and governance solution
Making better use of data than ever before

In today's world of big data, advanced analytics, and cloud, there are more opportunities than ever before—and data is powering them all. Data-driven organizations understand the benefits of maximizing insight from their data assets for data consumers and data producers.
But locating trusted data within an organization, gaining a unified view of data to understand whether it’s fit for use, and providing access to the right people all involve time-consuming and manual processes. In addition, the increasingly complex data landscape and government regulations require organizations to instill better data governance.

Here’s what makes it all so challenging:

Data estates and application landscapes have expanded from on premises to cloud to hybrid to multi-cloud.

Data and knowledge about data are trapped in unorganized silos which makes it hard to share information and assets.

Multiple users such as data analysts, data scientists, data stewards, data engineers, and app developers are all data hungry and need to collaborate more than ever before.

Complex government data regulations like GDPR and California Consumer Privacy Act mean organizations must have strict requirements on access, traceability, and transparency for private data.
To tackle the challenges of enterprise data management, you need to create a unified metadata management system that serves as a source of reference for information assets in the organization. This source of reference must be verified, complete, high quality, and secured. That’s where a data catalog comes in: a single collaborative solution for data professionals to collect, organize, find, access, enrich, and activate technical, business, and operational metadata to support self-service data discovery and governance for trusted data assets in Oracle Cloud and beyond.

Here’s what you deserve from a data catalog:

**Search and discovery**
Quickly find and explore data assets in the organization in a self-service manner and assess the suitability of data assets from both a technical and business perspective.

**Transparency and traceability**
Understand where the data came from and gain a comprehensive view of data activity to support governance and auditability.

**Automation and scale**
Use the power of AI/ML in the background to automate data catalog and data management tasks to improve productivity and handle the ever-increasing scale.
What is the Oracle Cloud Infrastructure Data Catalog vision?
Self-service data discovery and governance

Search and discovery
- Metadata harvesting from on-premises and cloud systems
- Semantic search, data profiling, lineage and impact analysis, and data relationships

Curation
- Enterprise business glossary, including approval workflows
- Tagging, user annotations, social collaboration, ratings, comments, associations, and links

Intelligence
- AI/ML based recommendations
- Auto-discovery, auto-classification, auto-association, auto-tagging, and data similarity

Enterprise-class
- Oracle Cloud Infrastructure-native, REST APIs, and SDKs
- Hive metastore for Oracle Cloud Infrastructure Object Store
- Value-added integration with other Oracle Cloud Infrastructure services in data management, analytics, and apps
Oracle Cloud Infrastructure Data Catalog extracts, standardizes, and indexes technical metadata from connected data sources to create a trusted and searchable data asset inventory.

Oracle Cloud Infrastructure 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 gauge the trustworthiness of data for different uses.
Oracle Cloud Infrastructure Data Catalog key features
Self-service data discovery and governance solution

**Search and explore**
Find information you need using an intuitive user interface

**Annotate with terms and tags**
Enrich harvested metadata by linking to business terms and freeform tags

**Manage business glossary**
Create business glossaries, categories, and terms

**Experience enterprise class**
Use REST APIs and SDKs to integrate Oracle Cloud Infrastructure Data Catalog with other applications

Copyright © 2020, Oracle and/or its affiliates. All rights reserved.
Oracle Cloud Infrastructure Data Catalog offers
Metadata harvesting from Oracle Cloud Infrastructure sources

Metadata harvesting from Oracle Cloud Infrastructure services:
• Oracle Cloud Infrastructure Object Storage
• Oracle Database
• Oracle Autonomous Transaction Processing
• Oracle Autonomous Data Warehouse
• Oracle MySQL Cloud Service
• Hive
• Kafka

Supported file types for Object Storage
• CSV, Excel
• ORC, Avro, Parquet
• JSON

On-demand or scheduled harvesting
Oracle Cloud Infrastructure 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
- Make use of freeform tags for user annotations
- Use asset links to business terms and tags to provide a holistic view of the data
Oracle Cloud Infrastructure Data Catalog offers
Search and exploration

• Search data based on technical names, business terms, or tags
• View details of various objects
• Browse Oracle Cloud Infrastructure Data Catalog based on data assets
Oracle Cloud Infrastructure Data Catalog offers
Single collaborative environment for all users

- Home page with helpful shortcuts and operational stats
- Search and browse
- Quick Actions to manage data assets, glossaries, jobs, and schedules
- Popular tags and recently updated objects
Oracle Cloud Infrastructure Data Catalog offers
An enterprise-class managed service

Native Oracle Cloud Infrastructure service
Fully managed service on Oracle Cloud Infrastructure

REST APIs and SDKs
Easily integrate Oracle Cloud Infrastructure Data Catalog capabilities in other applications

Policy-based access
Manage access based on Oracle Cloud Infrastructure IAM user groups
Get started today!

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

oracle.com/big-data/data-catalog