ORACLE®

BIG DATA PREPARATION CLOUD SERVICE

Prepare Your Data
Ingest, Transform, Enrich, and Publish in the Cloud

ORACLE®

What is It?

If you have complex business data sets, and you would like to turn it into meaningful information by taking advantage of automated data preparation and processing, then Oracle Big Data Preparation Cloud Service is the Oracle Cloud service for you.

Oracle Big Data Preparation Cloud Service gives you an easy-to-use way to work with your data. With its coordinated features, you can automate, streamline, and guide the error-prone process of data ingestion, preparation, repair, enrichment, and governance without costly manual intervention. This service is all about presenting an easy-to-use way to interact and work with your data.

How It Works

Oracle Big Data Preparation Cloud Service is available in Oracle Cloud as a massively scalable service built on Hadoop/Spark, and enhanced with Natural Language Processing and Reference Dataset Knowledge Service. It provides an intuitive and interactive user experience, guiding you with a machine learning-driven recommendation engine, so you can significantly reduce the amount of time needed to ingest and prepare new data sets for multiple downstream processes.

Your complex business operations are made easy, and any error-prone setups and configurations are resolved or avoided. Using Oracle Big Data Preparation Cloud Service, you can shape and improve your business data, and render your ecosystem simple, scalable, and automated.

Perception vs. Reality

When you work with your business data, the perception is that you spend most of your time on analytics. But in reality, you will devote much more time and effort on importing, profiling, cleansing, repairing, standardizing, and enriching your data. With Oracle Big Data Preparation Cloud Service, data preparation time and costs are reduced to a fraction of that of traditional methods.

ORACLE®

Recognize Yourself?

Can you answer "Yes" to any of these questions?

- Does your enterprise work with a mishmash of data sources?

- Is your content inconsistent, incomplete, and in a variety of formats, and is it taking you weeks to process your data and write custom scripts to clean up the mess?

- Do you want an efficient strategy to harvest and analyze data from social media and sales transactions?

- Do you have only a vague idea of the categories of information that your data might provide or its value?

What can Oracle Big Data Preparation Cloud Service do for you?

- Mitigates the challenges of data processing and preparation, lifting the burden of repairing, classifying, and publishing your data in the cloud.

- Provides you with a large set of data repair, transformation, and enrichment options that require zero coding or scripting.

- Enables you to see data transformations and the result of script automation in real time with a set of smart and interactive tools and features.
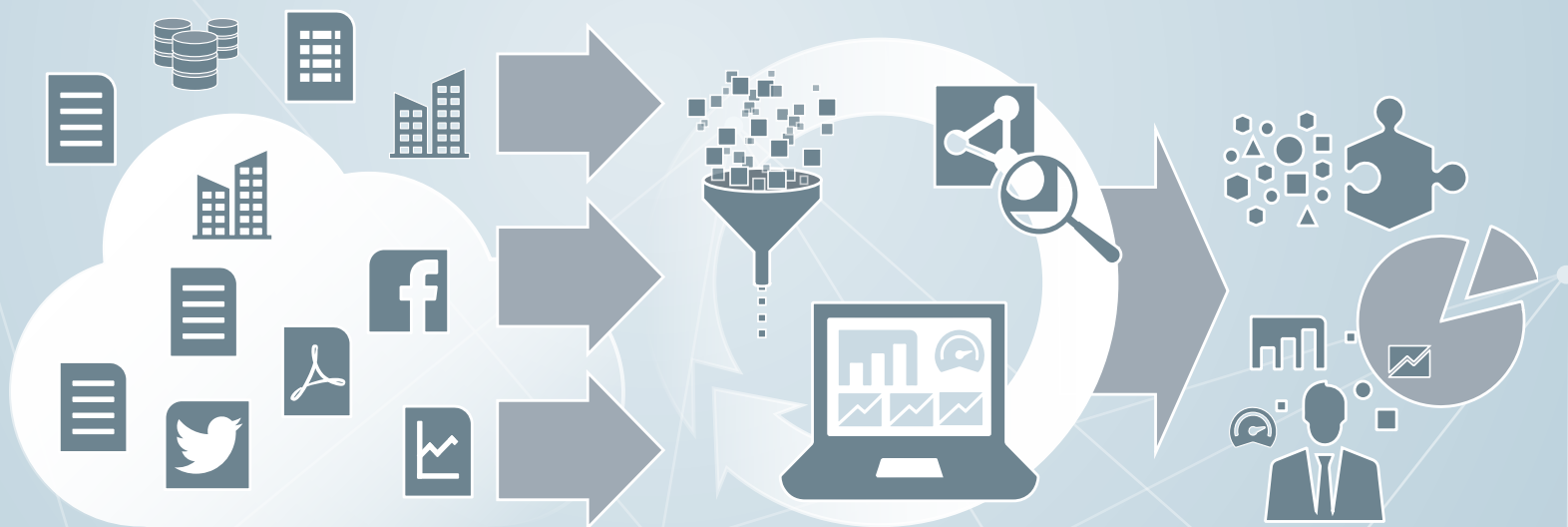
ORACLE®

What You Can Do

Move from gathering your unstructured, uncorrelated, and often unfit high volume data to clean and enhanced data sets. How? Use Oracle Big Data Preparation Cloud service to seamlessly complete the tasks of repairing and publishing your data toward downstream processes such as enterprise ETL, enterprise reporting, and data recovery.
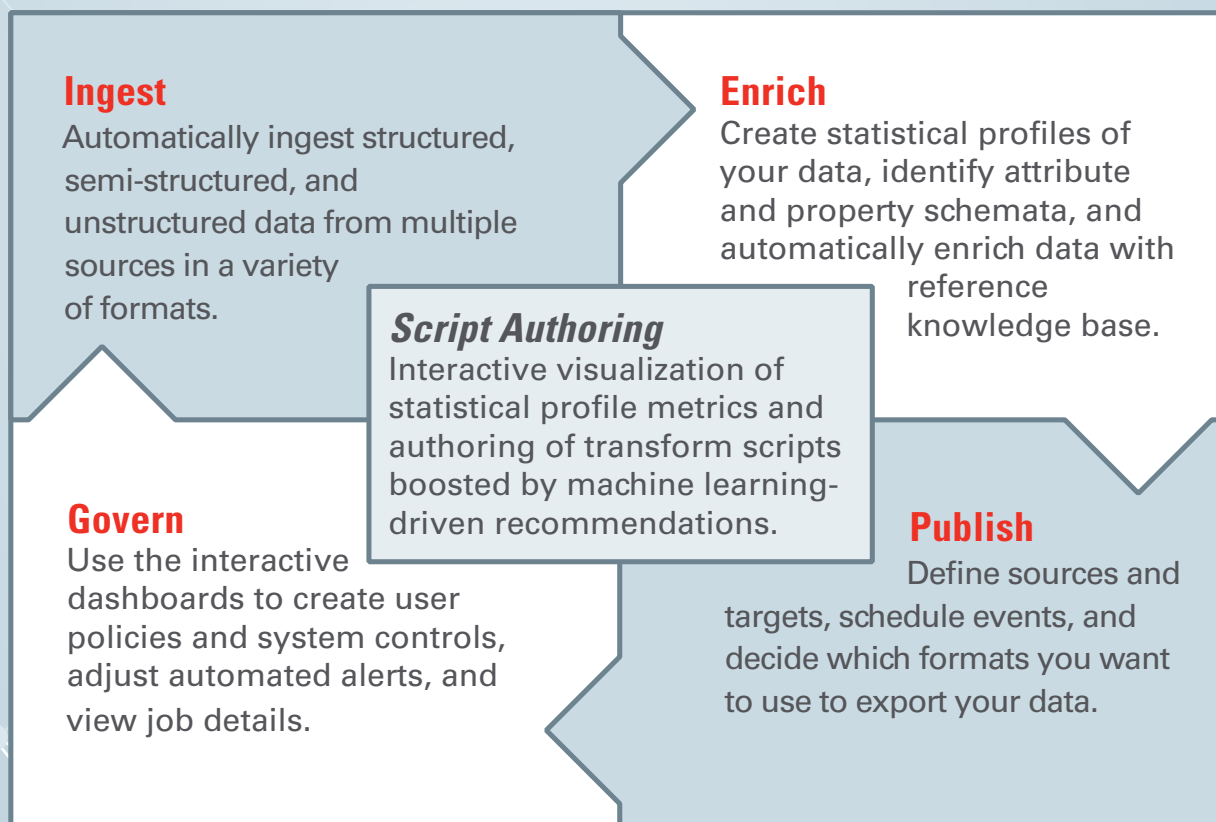
ORACLE®

Development Life Cycle

Using the complete set of capabilities of Oracle Big Data Preparation Cloud Service, you can make the most of the lifecycle features available to you. Clean, enrich, and export your data using the powerful features of interactive script authoring.

Ingest
Automatically ingest structured, semi-structured, and unstructured data from multiple sources in a variety of formats.

Enrich
Create statistical profiles of your data, identify attribute and property schemata, and automatically enrich data with reference knowledge base.

Script Authoring
Interactive visualization of statistical profile metrics and authoring of transform scripts boosted by machine learning-driven recommendations.

Govern
Use the interactive dashboards to create user policies and system controls, adjust automated alerts, and view job details.

Publish
Define sources and targets, schedule events, and decide which formats you want to use to export your data.

ORACLE®

Ingest

What are your data sources? Are they office documents, social media, or clickstream logs? If so, you need to ingest your data before you can effectively analyze and enrich it. To make sense of all the data you have, you must define a structure and correlate the disparate data sets. This important step involves both understanding and standardizing your data.

Here are just a few examples of what you can do to ingest and mend your data:

- **Statistical Profiling:** Create standard statistical analysis of numerical data, and frequency and term analysis of text data.

- **Process:** Handle multiple formats of data sources, whether their content is structured, semi-structured, or unstructured.

- **Cleanse:** Remove nonessential characters and standardize date formats.

- **Repair:** Find and fix inconsistencies.

- **Detect Schema:** Identify schema and metadata that is explicitly defined in headers, fields, or tags.

- **Identify Duplicates**: Find and flag duplicates in your data so you can reduce the size of your data pool.
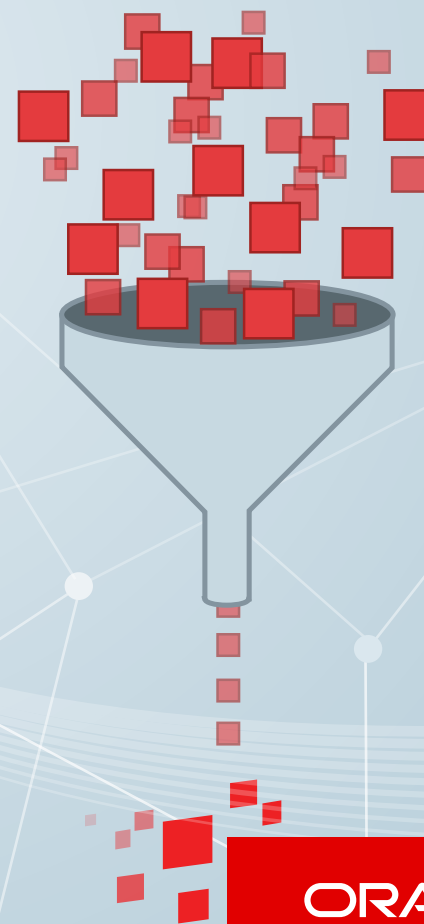
ORACLE®

Enrich

After you've cleansed your data, you can leverage any patterns and knowledge-based classifications to understand the domains found in your data sets.

Use the wide variety of known categories and vast array of reference data sets that Oracle Big Data Preparation Cloud Service has to analyze and recognize content without relying on any metadata.

After you've classified your data sets, accept the system recommendations to enrich your data sets with related entities from the reference knowledge service, and extract embedded entities found in your data. This semantically enriches and correlates your data.
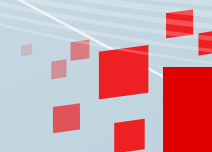
ORACLE®

Publish

You can publish your data using the powerful features of the Oracle Big Data Preparation Cloud Service Policies page. You can benefit from the service's knowledge graph API, visualize time lines for processed data sets, and drill down to individual job details, and publish interactively through a native scheduler or through the provided RESTful APIs.

You can create policies to run transforms on a schedule or an event such as monitoring a directory for new files to be processed. Oracle Big Data Preparation Cloud Service gives you the flexibility to run policies at regular intervals or at specific points in time.
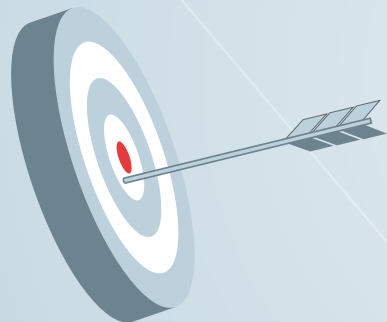
Sources/Targets
You can process data automatically from a number of sources in many different file formats, and you can manage your data sets using direct integration with Oracle Storage Cloud Service.

Execution
You can invoke service executions and customize them through the service's user interface, and you can schedule them according to your needs.

Formats
You can publish prepared data sets to a wide range of targets in a rich set of different formats. This feature allows you to make data available to multiple heterogeneous downstream BI, Analytics, and Enterprise ETL processes.

ORACLE®

Govern

As you ingest, enrich, and publish your data, Oracle Big Data Preparation Cloud Service provides a user interface-driven, intuitive Dashboard page to monitor all transform activity on your data sets.

The Dashboard page provides real-time metrics about the rows processed and automated alerts and errors. This information can be filtered by different time slices—from 24 hours up to 30 days. By using the Dashboard's pan and zoom features, you can see what's running at any time.

The Dashboard page also lets you view the types of files being ingested and the size of these files. You can take advantage of the graphical views of job listings and their details. Every aspect of data set transformation is exposed to you in Oracle Big Data Preparation Cloud Service.

ORACLE®

Automate

Reuse Transform Scripts

After you've created transform scripts via our recommendation-driven process, you can reuse those scripts to prepare new data sets in a fast, continuous process.
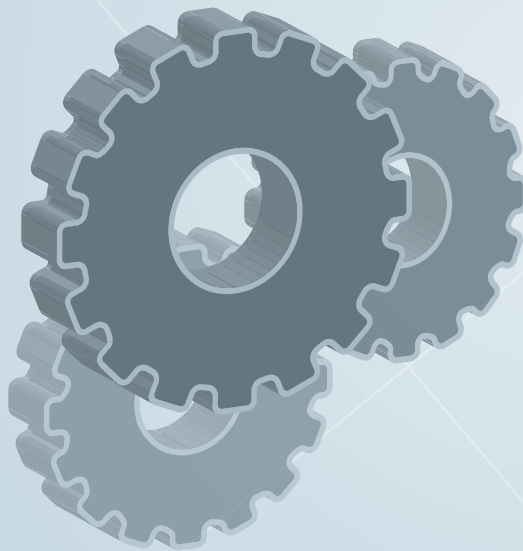
Automate the Process

Oracle Big Data Preparation Cloud Service gives you flexibility to automate the process. First, you can use the scheduler to set your transformations to run on a daily, weekly, or monthly basis against a pre-determined data source. Second, a rich set of RESTful APIs are available for you to automate the entire data preparation process, from file movement to preparation to publishing.

ORACLE®

Get Started

Subscribe

Set up an Oracle Public Cloud Services account for Oracle Big Data Preparation Cloud Service. You can register for a free 30-day trial or create a purchase order from the Oracle Cloud website (cloud.oracle.com) or the Oracle Store (shop.oracle.com) using your Oracle.com account.

If you need help creating a purchase order, an Oracle sales representative can do that for you.

For more information, see Trial and Paid Subscriptions for Oracle Cloud Services and visit the Oracle Cloud website at cloud.oracle.com.

Learn More

- Product Overview: cloud.oracle.com/bigdatapreparation
- User Documentation: Oracle Help Center

Attend Oracle Cloud Events

See events.oracle.com and blogs.oracle.com/cloud for information about Oracle Cloud events.

Join the Community

- Oracle Cloud Community: cloud.oracle.com
- Developers: cloud.oracle.com/developer

ORACLE®
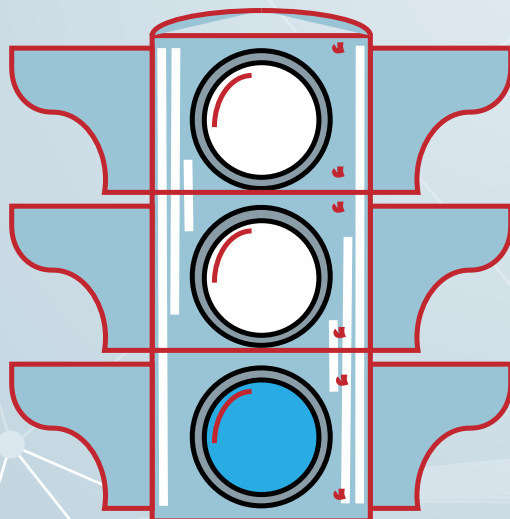
Safe Harbor

The preceding 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®

Follow us at:

Oracle Cloud on Facebook   Oracle Cloud on Twitter   Oracle Cloud on YouTube

Oracle Big Data Preparation Cloud Service Website