What is OCI DI

OCI DI stands for Oracle Cloud Infrastructure Data Integration. It is a fully managed, serverless extract, transform, load (ETL) service offered by Oracle as part of its cloud platform. OCI DI helps you design and automate data movement and transformation workflows across different data sources?on-premises or in the cloud.

Key Features:

- Visual Data Flows: Drag-and-drop interface to design ETL logic without writing code.
- Pipelines: Orchestration layer that schedules and sequences data flows, conditionals, and tasks.
- Built-in Connectors: Supports connections to Oracle databases, Object Storage, ADW, Big Data, MySQL, and 3rd party sources like Snowflake, Amazon S3, etc.
- Transformations: Rich set of transformations such as joins, filters, aggregations, lookups, etc.
- Parameterization: Use parameters for reusability and dynamic data flow logic.
- Monitoring & Logging: View run history, logs, and status of data integrations.

Use Cases:

- Migrate data from on-premises to Oracle Autonomous Data Warehouse (ADW)
- Data ingestion from multiple sources to a data lake
- Data preparation and transformation before feeding into analytics platforms
- Automating complex ETL workflows

Quick Example: Load CSV Data into Oracle ADW

Goal: Load employee data from a CSV file in OCI Object Storage into an Oracle Autonomous Data Warehouse (ADW), with basic transformation.

Steps:

1. Create a Data Asset for:
   - Source: Object Storage (CSV file)
   - Target: ADW

2. Create a Data Flow:
   - Drag the CSV file as the Source.
   - Add a Select or Filter transformation to remove null records or change column names.

- Add the ADW table as the Target.
- Map the columns from source to target.

3. Create a Pipeline (optional):
- Include the data flow task.
- Schedule it to run daily or on-demand.

4. Publish and Run.

Use Cases:

1. Data Migration - Move legacy data from on-premise Oracle DB to Oracle ADW using DI with minimal downtime.
2. Daily ETL for Reporting - Automate daily ingestion, transformation, and loading of sales or customer data into a reporting DB like ADW.
3. Data Lake Ingestion - Extract from various sources and load into OCI Data Lake.
4. Data Quality Pipelines - Clean and standardize data before analytics.
5. Event-Driven Processing - Trigger pipelines when new files are uploaded to Object Storage.

Example: Export from Oracle ADW to CSV in Object Storage

Use Case:

Export data from Oracle Autonomous Data Warehouse (ADW) to a CSV file in OCI Object Storage.

Steps in OCI Data Integration:

1. Create Data Assets:
- Source Data Asset: Oracle ADW (provide connection details)
- Target Data Asset: OCI Object Storage (bucket location for CSV)

2. Create a Data Flow:
- Source: Use a table or SQL query from ADW (e.g., SELECT * FROM EMPLOYEES)
- Transform (Optional): Add filters or rename columns
- Target: Object Storage, set format to CSV and file name (e.g., employees_export.csv)
- Map Columns: Ensure mapping between source and target

3. Create a Pipeline (Optional):

- Add Data Flow as a task
- Set up scheduling or conditional logic

4. Publish and Run:

- Run the pipeline or data flow to generate the CSV in Object Storage

Result:

CSV file (e.g., employees_export.csv) appears in the specified Object Storage bucket.