50

DBMS_CLOUD_PIPELINE

The `DBMS_CLOUD_PIPELINE` package allows you to create data pipelines for loading and exporting data in the cloud. This package supports continuous incremental data load of files in object store into the database. `DBMS_CLOUD_PIPELINE` also supports continuous incremental export of table data or query results from the database to object store based on a timestamp column.

Summary of DBMS_CLOUD_PIPELINE Subprograms

This table summarizes the subprograms included in the `DBMS_CLOUD_PIPELINE` package.

> **✎ Note:**
>
> You can use the DBMS_CLOUD_PIPELINE package only with Autonomous Database versions 19.22 or later.

Subprogram	Description
CREATE_PIPELINE Procedure	Creates a new data pipeline.
DROP_PIPELINE Procedure	Drops an existing data pipeline.
RESET_PIPELINE Procedure	Resets the tracking state of a data pipeline. Use reset pipeline to restart the pipeline from the initial state of data load or export. Optionally reset pipeline can purge data in the database or in object store, depending on the type of pipeline.
RUN_PIPELINE_ONCE Procedure	Performs an on-demand run of the pipeline in the current foreground session, instead of a scheduled job.
SET_ATTRIBUTE Procedure	Sets pipeline attributes. There are two overloaded procedures, one to set a single attribute and another to set multiple attributes using a JSON document of attribute name/value pairs
START_PIPELINE Procedure	Starts the data pipeline. When a pipeline is started, the pipeline operation will continuously run in a scheduled job according to the "interval" configured in pipeline attributes.
STOP_PIPELINE Procedure	Stops the data pipeline. When a pipeline is stopped, no future jobs are scheduled for the pipeline.

CREATE_PIPELINE Procedure

The procedure creates a new data pipeline.

Syntax

```
DBMS_CLOUD_PIPELINE.CREATE_PIPELINE(
        pipeline_name          IN   VARCHAR2,
        pipeline_type          IN   VARCHAR2,
        attributes             IN   CLOB         DEFAULT NULL,
```

```
        description          IN   VARCHAR2    DEFAULT NULL
);
```

Parameters

Parameter	Description
`pipeline_name`	Specifies a name for the pipeline. The pipeline name must follow the naming rules of Oracle SQL identifiers.
	This parameter is mandatory.
`pipeline_type`	Specifies the pipeline type.
	Valid values: `LOAD`, `EXPORT`
	This parameter is mandatory.
`attributes`	Pipeline attributes in JSON format.
	Default value: `NULL`
	See DBMS_CLOUD_PIPELINE Attributes for more information.
`description`	Description for the pipeline.
	Default value: `NULL`

DROP_PIPELINE Procedure

The procedure drops an existing data pipeline. If a pipeline has been started, then it must be stopped before it can be dropped.

Syntax

```
DBMS_CLOUD_PIPELINE.DROP_PIPELINE(
        pipeline_name          IN   VARCHAR2,
        force                  IN   BOOLEAN DEFAULT FALSE
 );
```

Parameters

Parameter	Description
`pipeline_name`	Specifies a pipeline name.
	This parameter is mandatory.
`force`	Forcibly drop a pipeline, even if it is in started state.
	Valid values: `TRUE`, `FALSE`
	Default value: `FALSE`

Usage Note

• In order to drop a pipeline that is in started state, set the `force` parameter to `TRUE`.

RESET_PIPELINE Procedure

Resets the tracking state of a data pipeline. Use reset pipeline to restart the pipeline from the initial state of data load or export. Optionally reset pipeline can purge data in the database or in

object store, depending on the type of pipeline. A data pipeline must be in stopped state to reset it.

Syntax

```
DBMS_CLOUD_PIPELINE.RESET_PIPELINE(
       pipeline_name        IN   VARCHAR2,
       purge_data           IN   BOOLEAN DEFAULT FALSE
 );
```

Parameters

Parameter	Description
pipeline_name	Specifies a name for the pipeline. This parameter is mandatory.
purge_data	Purge data applies for either a load pipeline or an export pipeline: • For a load pipeline, when TRUE, truncate the data in database table. • For an export pipeline, when TRUE, delete the files in the object store location. Valid values: TRUE, FALSE Default value: FALSE

Usage Notes

• A data pipeline must be in stopped state to reset it. See STOP_PIPELINE Procedure for more information.

• For a load pipeline, resetting the pipeline clears the record of the files being loaded by the pipeline. When you call either START_PIPELINE or RUN_PIPELINE_ONCE after resetting a load pipeline, the pipeline repeats the data load and includes all the files present in the object store location.

  When purge_data is set to TRUE, DBMS_CLOUD_PIPELINE.RESET_PIPELINE does the following:

  – Truncates the data in the pipeline's database table you specify with table_name attribute.

  – Drops the pipeline's status table, and the pipeline's bad file table and error table (if they exist).

• For an export pipeline, resetting the pipeline clears the last tracked data in the database table. When you call either START_PIPELINE or RUN_PIPELINE_ONCE after resetting an export pipeline, the pipeline repeats exporting data from the table or query.

  When purge_data set to TRUE, DBMS_CLOUD_PIPELINE.RESET_PIPELINE deletes existing files in the object store location specified with the location attribute.

RUN_PIPELINE_ONCE Procedure

This procedure performs an on-demand run of the pipeline in the current foreground session, instead of a running in a scheduled job. Use `DBMS_CLOUD_PIPELINE.RUN_PIPELINE_ONCE` to test a pipeline before you start the pipeline as a continuous job.

Syntax

```
DBMS_CLOUD_PIPELINE.RUN_PIPELINE_ONCE(
        pipeline_name  IN  VARCHAR2
);
```

Parameters

Parameter	Description
pipeline_name	Specifies a name for the pipeline to run.
	This parameter is mandatory.

Usage Notes

• After you perform a test run of a pipeline you can reset the pipeline state using `DBMS_CLOUD_PIPELINE.RESET_PIPELINE`. This allows you to reset the pipeline state before you start the pipeline in a scheduled job.

• If a pipeline is in the started state, then it cannot be run in the foreground session.

SET_ATTRIBUTE Procedure

This procedure sets pipeline attributes. There are two overloaded procedures, one to set a single attribute and another to set multiple attributes using a JSON document of attribute name/value pairs.

Syntax

```
PROCEDURE DBMS_CLOUD_PIPELINE.SET_ATTRIBUTE(
      pipeline_name        IN    VARCHAR2,
      attribute_name       IN    VARCHAR2,
      attribute_value      IN    CLOB
);

PROCEDURE DBMS_CLOUD_PIPELINE.SET_ATTRIBUTE(
      pipeline_name        IN    VARCHAR2,
      attributes           IN    CLOB
);
```

Parameters

Parameter	Description
pipeline_name	Specifies a name for the pipeline to set attributes.
	This parameter is mandatory.

Parameter	Description
attribute_name	Specifies the attribute name for the attribute to be set. See DBMS_CLOUD_PIPELINE Attributes for more information.
attribute_value	Specifies the value for the pipeline attribute to set. See DBMS_CLOUD_PIPELINE Attributes for more information.
attributes	Specifies a JSON document containing attribute names and values. See DBMS_CLOUD_PIPELINE Attributes for more information.

Usage Note

- When you use `DBMS_CLOUD_PIPELINE.SET_ATTRIBUTE` to set multiple attributes with the `attributes` parameter, all the existing attributes are deleted and overwritten with the specified attributes from the JSON document.

START_PIPELINE Procedure

The procedure starts the data pipeline. When a pipeline is started, the pipeline operation runs continuously in a scheduled job according to the `interval` configured with the pipeline attributes.

Syntax

```
DBMS_CLOUD_PIPELINE.START_PIPELINE(
      pipeline_name        IN   VARCHAR2,
      start_date           IN   TIMESTAMP WITH TIME ZONE DEFAULT NULL
);
```

Parameters

Parameter	Description
pipeline_name	Specifies a name for the pipeline. This parameter is mandatory.
start_date	Specifies the starting date for the pipeline job. Default value: NULL

Usage Notes

- By default, a pipeline job begins immediately, as soon as the pipeline is started. To start a pipeline job at a later time specify a valid date or timestamp using the `start_date` parameter.

- See DBMS_CLOUD_PIPELINE Attributes for information on the pipeline `interval` and other pipeline attributes.

STOP_PIPELINE Procedure

The procedure stops the data pipeline. When a pipeline is stopped, no future jobs are scheduled for the pipeline.

Syntax

```
DBMS_CLOUD_PIPELINE.STOP_PIPELINE(
     pipeline_name          IN   VARCHAR2,
     force                  IN   BOOLEAN  DEFAULTFALSE
);
```

Parameters

Parameter	Description
pipeline_name	Specifies a name for the pipeline. This parameter is mandatory.
force	If force parameter is passed as TRUE, then it will terminate any running jobs for the pipeline. Valid values: TRUE, FALSE Default value: FALSE

DBMS_CLOUD_PIPELINE Attributes

Attributes help to control and configure the behavior of a data pipeline.

Attributes

> **Note:**
>
> As indicated in the **Pipeline Type** column, depending on the pipeline type LOAD or EXPORT, a pipeline supports a different set of attributes.

Attribute Name	Description	Pipeline Type	Modifiable After Pipeline Starts
credential_name	The name of the credential to access the source Cloud Object Storage. Default value: NULL. If you do not supply a credential_name, credential_name is set to NULL.	LOAD, EXPORT	Yes

Attribute Name	Description	Pipeline Type	Modifiable After Pipeline Starts
field_list	Identifies the fields in the source files and their data types. This argument's syntax is the same as the field_list clause in regular Oracle external tables. Default value: NULL The default value specifies the fields and their data types are determined by the columns in the table specified in table_name attribute.	LOAD	Yes
format	The options describing the format for the type of pipeline. • For a load pipeline, see DBMS_CLOUD Package Format Options. • For an export pipeline, see DBMS_CLOUD Package Format Options for EXPORT_DATA Datapump format is not supported for an export pipeline. This attribute is mandatory for both LOAD and EXPORT pipelines.	LOAD, EXPORT	Yes
interval	The time interval in minutes between consecutive executions of scheduled pipeline job. Default value: 15 minutes	LOAD, EXPORT	Yes
key_column	A timestamp or date column in the specified table or query for exporting newer data continuously to object store. The last execution timestamp or date is tracked by Export pipeline and compared with the value in the key_column to identify new data for exporting to object store. Default value: NULL If key_column is not specified for an export pipeline, the entire contents of the table or query are uploaded to object store in each pipeline job execution.	EXPORT	No
location	Specifies a URI that points to an Object Storage location. The format of the URI depends on the Cloud Object Storage service. See DBMS_CLOUD URI Formats for more information. This attribute is mandatory for both LOAD and EXPORT pipelines.	LOAD, EXPORT	No

Attribute Name	Description	Pipeline Type	Modifiable After Pipeline Starts
query	Specifies a `SELECT` statement so that only the required data is exported. The query determines the contents of the files you export as text files (CSV, JSON, Parquet, or XML) or dump files. For example: `SELECT warehouse_id, quantity FROM inventories` Default value: `NULL` For an export pipeline, either `table_name` or `query` is mandatory.	`EXPORT`	No
table_name	Specifies the name of the target table for loading or exporting data. For a load pipeline `table_name` is mandatory. For an export pipeline, either `table_name` or `query` is mandatory.	`LOAD, EXPORT`	No
table_owner	The name of the schema where the target table resides for loading or exporting data. Default value: `NULL` With a `NULL` value the target table is in the same schema as the user running the procedure.	`LOAD, EXPORT`	No

DBMS_CLOUD_PIPELINE Views

The `DBMS_CLOUD_PIPELINE` package uses the following views.

- DBA_CLOUD_PIPELINES_View
- USER_CLOUD_PIPELINES_View
- DBA_CLOUD_PIPELINE_HISTORY View
- USER_CLOUD_PIPELINE_HISTORY View
- DBA_CLOUD_PIPELINE_ATTRIBUTES View
- USER_CLOUD_PIPELINE_ATTRIBUTES View