

DBMS_HADOOP

The `DBMS_HADOOP` package provides a PL/SQL procedure called `CREATE_EXTDDL_FOR_HIVE()`, that creates an Oracle external table for a given hive table.

Big Data SQL needs to be correctly set up for `DBMS_HADOOP` to work.

This chapter contains the following topics:

- [DBMS_HADOOP Overview](#)
- [DBMS_HADOOP Security Model](#)
- [Summary of DBMS_HADOOP Subprograms](#)

DBMS_HADOOP Overview

The `DBMS_HADOOP` package provides two procedures for creating an Oracle external table and for synchronizing the Oracle external table partitions.

These procedures are:

- `CREATE_EXTDDL_FOR_HIVE()` — creates an Oracle external table for a given hive table
- `SYNCHRONIZE_PARTITIONS_FOR_HIVE()` — helps to synchronize the Oracle external table partitions with those in the corresponding hive table

DBMS_HADOOP Security Model

Users must have `ALTER` privileges on the table for modifying external table parameters. In addition the `ALTER` privileges, users must have `READ` privileges for the directory object that contains the external data source and `WRITE` privileges for the directory objects containing bad, log, and discard files.

Similar privileges are applicable to the partitioned external tables.

Summary of DBMS_HADOOP Subprograms

`DBMS_HADOOP` includes the `CREATE_EXTDDL_FOR_HIVE` procedure and `SYNC_PARTITIONS_FOR_HIVE` procedure subprograms.

Table 95-1 DBMS_HADOOP Subprograms

| Subprogram | Description |
|---|---|
| CREATE_EXTDDL_FOR_HIVE Procedure | Given a hive table name, creates a text of DDL that can be executed to create an external table corresponding to the hive table |
| CREATE_HYBRID_PARTNED_TABLE Procedure | Merges a given partitioned Hive table with an Oracle partitioned table, and replaces it with the merged table. |

Table 95-1 (Cont.) DBMS_HADOOP Subprograms

| Subprogram | Description |
|---|---|
| SYNCHRONIZE_PARTITIONS_FOR_HIVE Procedure | Synchronizes the existing partitioning definition of a hive table in Oracle catalog |

CREATE_EXTDDL_FOR_HIVE Procedure

This procedure creates an Oracle external table for a given hive table.

Syntax

```
DBMS_HADOOP.CREATE_EXTDDL_FOR_HIVE (
  cluster_id      IN  VARCHAR2,
  db_name         IN  VARCHAR2 := NULL,
  hive_table_name IN  VARCHAR2,
  hive_partition  IN  BOOLEAN,
  table_name      IN  VARCHAR2 := NULL,
  perform_DDL     IN  BOOLEAN := FALSE,
  text_of_DDL     OUT CLOB);
```

Parameters

Table 95-2 CREATE_EXTDDL_FOR_HIVE Procedure Parameters

| Parameter | Description |
|-----------------|---|
| cluster_id | Hadoop cluster ID |
| hive_db_name | Database where the hive table is located |
| hive_table_name | Name of the hive table |
| hive_partition | If this argument is TRUE and if the original hive table is partitioned, the corresponding Oracle external table will also be partitioned using the same partition key(s). If the original hive table is not partitioned, hive_partition=TRUE has no effect. If hive_partition=FALSE, the corresponding Oracle external table will not be partitioned even if the original hive table is partitioned. |
| table_name | Name of the Oracle external table |
| perform_DDL | If this argument is TRUE, the external table will be automatically created. Otherwise, only the textual representation of the DDL statement will be generated and returned in text_of_DDL . |
| text_of_DDL | If the argument perform_DDL is FALSE, only the textual representation of the DDL statement will be generated and returned in text_of_DDL . |

CREATE_HYBRID_PARTNED_TABLE Procedure

This procedure merges the partitioned Hive table with an Oracle partitioned table, and replace it with the merged table.

Syntax

```
DBMS_HADOOP.CREATE_HYBRID_PARTNED_TABLE (
    cluster_id          IN      VARCHAR2,
    db_name             IN      VARCHAR2,
    hive_table_name     IN      VARCHAR2,
    table_name          IN      VARCHAR2,
    table_owner         IN      VARCHAR2,
    perform_ddl         IN      BOOLEAN := TRUE,
    text_of_ddl         OUT     CLOB);
```

Parameters

Table 95-3 CREATE_HYBRID_PARTNED_TABLE Procedure Parameters

| Parameter | Description |
|-----------------|---|
| cluster_id | Hadoop cluster ID |
| db_name | Database where the partitioned Hive table is located |
| hive_table_name | Name of the partitioned Hive table |
| table_name | Name of the partitioned Oracle table |
| table_owner | The owner of the partitioned Oracle table |
| perform_DDL | If the value of this parameter is <code>TRUE</code> , the external table is automatically created. If the value of this parameter is <code>FALSE</code> , only the textual representation of the DDL statement is generated and returned in <code>text_of_DDL</code> . |
| text_of_DDL | If the argument <code>perform_DDL</code> is <code>FALSE</code> , only the textual representation of the DDL statement is generated and returned in <code>text_of_DDL</code> . |

SYNCHRONIZE_PARTITIONS_FOR_HIVE Procedure

This procedure synchronizes the Oracle external table partitions with those in the corresponding hive table.

Syntax

```
DBMS_HADOOP.SYNCHRONIZE_PARTITIONS_FOR_HIVE (
    table_name      IN  VARCHAR2,
    table_owner     IN  VARCHAR2);
```

Parameters

Table 95-4 SYNCHRONIZE_PARTITIONS_FOR_HIVE Procedure Parameters

| Parameter | Description |
|-------------|-----------------------|
| table_name | Oracle external table |
| table_owner | Schema name |