

Databricks: Customer-Managed UDAF Installation

You must manually install AtScale UDAFs if you choose "Customer Managed" as the Custom Function Installation Mode during data warehouse configuration.

To facilitate the manual installation and upgrade of AtScale UDAF functions for a Databricks data warehouse, AtScale 2020.3.0 and up allows administrators to specify the Custom Function Installation Mode when configuring an AtScale data warehouse connection. The default mode is "None." Choosing "Customer Managed" gives the responsibility of installing and upgrading the AtScale UDAFs to the Databricks administrator.

The following instructions explain how a data warehouse administrator can install or upgrade AtScale UDAFs on a Databricks cluster.

Databricks: Before You Begin

1. Set the AtScale Data Warehouse Custom Function Installation Mode to "Customer Managed" after registering the HLL functions, when configuring the AtScale data warehouse connection. See [Adding Databricks Data Warehouses](#)
2. Go to the [Installer Packages](#) document and download the JAR file for Databricks. Then install it on your Databricks environment. For more information on this process, see [Libraries](#).
3. Go to the Databricks user interface, and get the name and location of the JAR file by selecting **Clusters > Click your cluster name > Libraries**.



Note: In the examples below, it is assumed that the location is: [dbfs:/FileStore/jars/honeybee-hive-4.0.68.jar](#)

Databricks: Register The HLL Functions

1. Create the UDAF schema where the HLL/Quantile functions will be stored. The UDAFSCHEMA is required to be the same as the one from the Add a Data Warehouse dialog in the Data Warehouse configuration. This step can be skipped if you are upgrading from an older version of Honeybee.

```
CREATE SCHEMA {UDAFSCHEMA}
```

2. Replace `{UDAFSCHEMA}` with the actual name of the schema from previous step and execute the following query:

```
USE {UDAFSCHEMA}
```

3. Execute the following queries to drop previous versions of HLL if any are present:

```

DROP FUNCTION IF EXISTS atscale_honeybee_version

DROP FUNCTION IF EXISTS hll_estimate

DROP FUNCTION IF EXISTS hll_aggregate

DROP FUNCTION IF EXISTS hll_aggregate_estimate

DROP FUNCTION IF EXISTS hll_aggregate_merge

```

4. Execute the following queries to create the HLL functions; make sure you use the correct location from your environment, as explained in step 3 in the [Databricks: Before You Begin](#) section above.

```

CREATE FUNCTION atscale_honeybee_version AS 'com.atscale.honeybee.HoneyBeeVersionUDF' USING JAR
'dbfs:/FileStore/jars/honeybee-hive-4.0.68.jar'

CREATE FUNCTION hll_estimate AS 'com.atscale.honeybee.hyperloglog.HyperLogLogEstimateUDA' USING JAR
'dbfs:/FileStore/jars/honeybee-hive-4.0.68.jar'

CREATE FUNCTION hll_aggregate_merge AS 'com.atscale.honeybee.hyperloglog.HyperLogLogAggregateMergeUDA'
USING JAR 'dbfs:/FileStore/jars/honeybee-hive-4.0.68.jar'

CREATE FUNCTION hll_aggregate_estimate AS
'com.atscale.honeybee.hyperloglog.HyperLogLogAggregateEstimateUDA' USING JAR 'dbfs:/FileStore/jars/honeybee-
hive-4.0.68.jar'

CREATE FUNCTION hll_aggregate AS 'com.atscale.honeybee.hyperloglog.HyperLogLogAggregateUDA' USING JAR
'dbfs:/FileStore/jars/honeybee-hive-4.0.68.jar'

```

5. Execute the following queries to drop previous versions of Quantile if any are present:

```

DROP FUNCTION IF EXISTS quantile_sketch

DROP FUNCTION IF EXISTS quantile_estimate

DROP FUNCTION IF EXISTS quantile_sketch_merge

DROP FUNCTION IF EXISTS quantilefromsketch

```

6. Create the Quantile functions; again, make sure you use the correct location from your environment, as explained in step 3 in the [Databricks: Before You Begin](#) section above.:

```

CREATE FUNCTION quantile_sketch AS 'com.atscale.honeybee.quantile.QuantileAggregateUDA' USING JAR
'dbfs:/FileStore/jars/honeybee-hive-4.0.68.jar'

CREATE FUNCTION quantile_estimate AS 'com.atscale.honeybee.quantile.QuantileEstimateUDA' USING JAR
'dbfs:/FileStore/jars/honeybee-hive-4.0.68.jar'

CREATE FUNCTION quantile_sketch_merge AS 'com.atscale.honeybee.quantile.QuantileMergeUDA' USING JAR
'dbfs:/FileStore/jars/honeybee-hive-4.0.68.jar'

CREATE FUNCTION quantilefromsketch AS 'com.atscale.honeybee.quantile.QuantileEstimateFromSketchUDF' USING
JAR 'dbfs:/FileStore/jars/honeybee-hive-4.0.68.jar'

```

7. Execute the following SQL query against your Databricks database. If the query returns HLL sketch, the HLL functions were successfully installed:

```
SELECT hll_estimate(1)
```

8. Set the AtScale Data Warehouse Custom Function Installation Mode to "Customer Managed" when configuring your Databricks data warehouse.
9. From the Databricks warehouse dialog, set the UDAF schema to the schema used in step 1.