Triggering Aggregate Rebuilds

You can trigger a rebuilding of aggregate instances whenever a specified file in the distributed file system of your data warehouse is updated. Aggregate rebuilds are triggered per cube.

About This Task

Many companies have ETL (extract, transform, load) jobs that they routinely run to cleanse raw data and load it into a data warehouse. As a result of these ETL processes, the jobs may write to a file or log to a distributed file system (in a path in a bucket, if your data warehouse is Google BigQuery or Amazon Redshift; in HDFS, if your data warehouse is Hadoop) to signal that the job is done.

In AtScale, you can configure a cube to watch a path in the distributed file system. When the system detects that the file is overwritten, it would start a batch rebuild of all of the cube's aggregate tables.

Procedure

1. Choose **Projects** from the Design Center. Find and click on the project name. Expand the project name under the Published section and click on the cube name.
2. Click **Build**. This is where you can edit schedules and set trigger files.
3. To set a trigger file, select **Triggers** from the **Build** drop-down menu. This action launches the Set Trigger File pop-up dialog.
4. In the **Paths** section of the **Set Trigger File** dialog, specify the data warehouse that the trigger file is in.
5. Then, specify the absolute path in the distributed file system as the directory where the trigger file is located. For example, this absolute path specifies the trigger file as etl.log:

   ```
   /user/atscale/some/path/etl.log
   ```

   - ▲ If your data warehouse is Google BigQuery, this absolute path starts at the root level of your Google Cloud bucket.
   - ▲ If your data warehouse is Amazon Redshift, this absolute path starts at the root level of your Amazon S3 bucket.
   - ▲ If your data warehouse is Hadoop, this absolute path starts at the root level of your cluster.

6. Click **Save**.

Results

AtScale will monitor the specified file and rebuild the cube's aggregates whenever there is a change to that file location.

What To Do Next

After an initial build or rebuild is triggered, you can go to the **Aggregates** page to view the results. Choose **Aggregates** from the main navigation, and select **Build History**. You can also open this page by clicking **View History** in the menu for a published cube.

This page shows whether builds are queued, running, or successful. It also shows whether builds failed.

If a build is running, you can cancel it, if you need to, by clicking **Cancel Build** in the entry for the build.

In an entry for a build, you can click the **View Instances** link to find out information about the instances that were or were not built.

Related concepts

How Aggregate Tables Are Populated With Data