Archive Cassandra Data

This functionality can be used to archive your Cassandra data, and archiving process use map-reduce jobs to achieve the task, therefore it supports for archiving large amount of data using cluster of hadoop nodes. You can run your archive process manually or you can schedule it using a cron expression. Follow the instructions below to use it.

- **Log in to BAM’s management console and select ”Home -> Configure -> Archive Data”**

- **Manually archive data between specified date range.**

![Cassandra Data Archive Configuration](image)

**Configuration parameters**

- **Stream name** - In BAM data model stream name map to a Cassandra Column Family. In here you need to provide the stream name, then data stored under that stream name can be archived.
- **Version** - Version of the stream. If you have different versions under same stream name (We don't encourage it), then this can be used to archive data specific to some version.
- **Username** - Cassandra username (same as BAM username)
- **Password** - Cassandra password (same as BAM password)
- **Date range**
  - **From** - Start date - ex: 25/01/2013 00:00:00 AM
  - **To** - End date - ex: 03/02/2013 00:00:00 AM
- **External Cassandra cluster**
  - **Connection URL** - connection details of Cassandra cluster. ex:- 10.100.60.150:9160,10.100.60.151:9160

**Schedule archive process**

![Cassandra Data Archive Configuration](image)

**Additional Configuration parameters**

- **No of days** - Keeps only last 'n' no of days data in the Column Family ex :- According to above
configuration it will only keep last 90 days of data. Data older than 90 days will archive.

- Cron expression - Archive process can be scheduled using the given cron expression. ex:- According to above configuration, archive job runs every day in the midnight.

**Archive Column Family Details**

Name of the archive column family will be original column family name + _arch

**Change schedule time later**

Once you submit the configuration to run the archive process, it will create a hive script and execute it. You can see the hive script in “Analytics -> List” (For manual process, it only executes the hive query and doesn’t save it).

Now you can change the schedule time using "schedule script" option.