Apache CarbonData 2.1.1 Release

Apache CarbonData community is pleased to announce the release of the Version 2.1.1 in The Apache Software Foundation (ASF).

CarbonData is a high-performance data solution that supports various data analytic scenarios, including BI analysis, ad-hoc SQL query, fast filter lookup on detail record, streaming analytics, and so on. CarbonData has been deployed in many enterprise production environments, in one of the largest scenarios, it supports queries on a single table with 3PB data (more than 5 trillion records) with response time less than 3 seconds!

We encourage you to use the release https://archive.apache.org/dist/carbondata/2.1.1/, and feedback through the CarbonData user mailing lists!

This release note provides information on the new features, improvements, and bug fixes of this release.

What’s New in CarbonData Version 2.1.1?

In CarbonData 2.1.1, 78 JIRA tickets related to improvements, and bugs have been resolved. Please find the summary of the important features that are developed with this release.

Support MERGE INTO SQL Syntax

CarbonData now supports MERGE INTO SQL syntax along with the API support. This will help the users to write CDC job and merge job using SQL also now.

Geospatial index algorithm improvement and UDFs enhancement

CarbonData now added new UDFs related to polygon for geospatial queries and previous quadtree forming algorithm as improved along with removed few configurations to improve usability.

Adding global sort support for SI segments data files merge operation.

CarbonData now uses global sort if the SI table is a global sort table for within a segment SI merge operation.

Size control of minor compaction

If some huge segment doesn’t want to participate in minor compaction, the user can configure a threshold to skip segments above threshold size for minor compaction.

Clean files become data trash manager

To avoid accidental data deletes, CarbonData has introduced a trash folder to keep accidentally deleted segments. so that users can review it and restore it back.

Note:

Fix error when loading string field with high cardinality (local dictionary fallback issue [CARBONDATA-4084])

If user is using local dictionary created from CarbonData 2.0.0 or 2.0.1 or 2.1.0, in case of local dictionary fallback scenario. while writing the local dictionary column, LV has become LLV. which was causing extra characters in the query output.
So, In your environment if this issue is observed, we suggest to drop those data / recreate a table and load the data again with carbonData 2.1.1 jars.

Please find the detailed JIRA list here.