Apache CarbonData 1.5.4 Release

Apache CarbonData community is pleased to announce the release of the Version 1.5.4 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/1.5.4/, 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 1.5.4?

CarbonData 1.5.4 intention was to move closer to unified analytics. We have added new binary datatype to store binary objects like images. We have also allowed users to change sort columns of an existing table for better flexibility as per user needs. we are now compacting the segments which are loaded using range sort.

In this version of CarbonData, around 13 JIRA tickets related to new features, improvements, and bugs have been resolved. Following are the summary.

CarbonData Core

Support Altering SORT_COLUMNS Property on the table

Previously, the user can configure the sort columns during table creation only, it restricts the user to load the data with same sort columns even though his query scenarios are changed.

From this version, we support altering the sort columns even after the table is created.

Support Configurable Page Size

This version allows the user to configure the page size, it gives the control of memory utilization during reading and loading data especially for complex, varchar, and binary datatypes.

Support Binary Data Type

It is useful to store big objects and binary objects like images.

Supported Compaction on Range Sorted Segments

The segments which are loaded with range sort scope will now be compacted using the range compaction.

Behaviour Change

None

Please find the detailed JIRA list: https://issues.apache.org/jira/secure/ReleaseNote.jspa?projectId=12320220&version=12345388

Sub-task

- [CARBONDATA-3362] - Document Update for Scenario

Bug

- [CARBONDATA-3331] - Database index size is more than overall index size in SHOW METADATA command
- [CARBONDATA-3334] - multiple segment files created for partition table for one segment
- [CARBONDATA-3344] - Fix Drop column not present in table
- [CARBONDATA-3353] - Fix MinMax Pruning for Measure column in case of Legacy store
- [CARBONDATA-3360] - NullPointerException in clean files operation
- [CARBONDATA-3369] - Fix issues during concurrent execution of Create table If not exists
- [CARBONDATA-3371] - Compaction show ArrayIndexOutOfBoundsException after sort_columns modification
- [CARBONDATA-3375] - GC Oversead limit exceeded error for huge data in Range Compaction
- [CARBONDATA-3377] - String Type Column with huge strings and null values fails Range Compaction
- [CARBONDATA-3386] - Concurrent Merge index and query is failing
**Improvement**

- [CARBONDATA-3001] - Propose configurable page size in MB (via carbon property)
- [CARBONDATA-3374] - Optimize documentation and fix some spell errors.