Apache CarbonData 2.3.0 Release

Apache CarbonData community is pleased to announce the release of the Version 2.3.0 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://dist.apache.org/repos/dist/release/carbondata/2.3.0/, 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 2.3.0?

In CarbonData 2.3.0, 48 JIRA tickets related to improvements, and bugs have been resolved. Please find the summary of the important features that are developed with this release.

Alter schema for complex columns

Support for alter operations on complex columns like map, struct and array has been added in this release

Support for Dynamic Partition Pruning for Spark-3.1 to enhance performance

Carbon now support DPP for spark 3.1, which can help to speed up partition query performance.

Introduce Streamer tool for Carbondata

Introduce streamer tool which can help ingest incrementally from various commonly used sources like kafka, DFS etc.

Support spatial index creation using DataFrame

Spatial Indexes can now be created using spark DataFrames