

Apache CarbonData 1.1.0 Release

Apache CarbonData community is pleased to announce the release of the Version 1.1.0 in The Apache Software Foundation (ASF). CarbonData is a new BigData native file format for faster interactive query using advanced columnar storage, index, compression, and encoding techniques to improve computing efficiency. In turn it will help to speedup queries an order of magnitude faster over PetaBytes of data.

We encourage everyone to download the release <https://archive.apache.org/dist/carbondata/1.1.0/>, and [feedback through the CarbonData user mailing lists](#)

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

What's New in Version 1.1.0?

In this version of CarbonData, following are the new features added for performance improvements, compatibility, and usability of CarbonData.

Introducing V3 Data Format

Benefits:

- Improves the query performance by ~20% to 50%.
- Improves the sequential IO by using larger size blocklets, this helps in reading larger data at once to memory.
- Introduced pages with size of 32000 each for every column inside the blocklets, and min/max is maintained for each page to improve the filter queries.
- Improved compression/decompression for row pages.

Alter Table Support

Benefits:

- Renaming of existing table.
- Adding a new column for existing table.
- Removing of new column for existing table.
- Upcasting of data type from INT to BIGINT or decimal precision from lower to higher.

Batch Sort Support for Data Loading

Benefits: Batch sort makes sort step as non blocking step, and capable of sorting whole batch in memory and converts to CarbonData file.

Improved Single Pass

Benefits: Improved Single Pass load by upgrading to latest Netty framework, and launched dictionary client for each loading thread.

Range Filter Support

Benefits: Range filters combines the between filters to one filter to improve the filter performance.

Improvements on Large Cluster

Benefits:

- No more parallel loading of dictionary metadata in executor. Now dictionary metadata is loaded only once after all tasks inside executor uses it.
- Minimized file operations to avoid multiple namenode calls during query

Please find the detailed JIRA list : <https://issues.apache.org/jira/secure/ReleaseNote.jspa?projectId=12320220&version=12338987>

Sub-task

- [\[CARBONDATA-661\]](#) - misc cleanup in carbon core
- [\[CARBONDATA-681\]](#) - CSVReader related code improvement
- [\[CARBONDATA-777\]](#) - Alter table support for spark 2.1
- [\[CARBONDATA-780\]](#) - Alter table support for compaction through sort step
- [\[CARBONDATA-805\]](#) - Fix groupid,package name,Class name issues
- [\[CARBONDATA-856\]](#) - [Documentation] Alter Table - TABLE RENAME
- [\[CARBONDATA-857\]](#) - [Documentation] Alter Table - ADD COLUMNS
- [\[CARBONDATA-894\]](#) - Add license header for vectorreader/AddColumnTestCases,ChangeDataTypeTestCases,DropColumnTestCases

Bug

- [\[CARBONDATA-325\]](#) - Create table with columns contains spaces in name.

- [CARBONDATA-400] - [Bad Records] Load data is fail and displaying the string value in beeline as exception
- [CARBONDATA-424] - Data Load will fail for badrecord and "bad_records_action" is fail
- [CARBONDATA-530] - Query with ordery by and limit is not optimized properly
- [CARBONDATA-642] - Delete Subquery is not working while creating and loading 2 tables
- [CARBONDATA-643] - When we are passing ALL_DICTIONARY_PATH' in load query ,it is throwing null pointer exception.
- [CARBONDATA-672] - Complex data type is not working while fetching it from Database
- [CARBONDATA-678] - Corr function is not working for double datatype.
- [CARBONDATA-680] - Add stats like rows processed in each step. And also fix unsafe sort enable issue.
- [CARBONDATA-682] - Fix license header for FloatDataTypeTestCase.scala and DateTypeTest.scala
- [CARBONDATA-684] - Improve test sufficiency and code coverage of carbondata-core module
- [CARBONDATA-685] - Able to create table with spaces using carbon source
- [CARBONDATA-688] - Abnormal behaviour of double datatype when used in DICTIONARY_INCLUDE and filtering null values
- [CARBONDATA-690] - Carbon data load fails with default option for USE_KETTLE(False)
- [CARBONDATA-691] - After Compaction records count are mismatched.
- [CARBONDATA-692] - Support scalar subquery in carbon
- [CARBONDATA-696] - NPE when select query run on measure having double data type without fraction.
- [CARBONDATA-697] - single_pass is not used while doing data load
- [CARBONDATA-700] - invalid example of no_inverted_index in carbondata ddl docs
- [CARBONDATA-702] - Created carbondata repository with adding format jar for facilitating compile
- [CARBONDATA-703] - Update build command after optimizing thrift compile issues
- [CARBONDATA-704] - data mismatch between hive and carbondata after loading for bigint values
- [CARBONDATA-705] - Make the partition distribution as configurable and keep spark distribution as default
- [CARBONDATA-706] - Multiple OR operators does not work properly in carbondata
- [CARBONDATA-707] - Less (<) than operator does not work properly in carbondata.
- [CARBONDATA-708] - Between operator does not work properly in carbondata.
- [CARBONDATA-709] - Incorrect documentation for bucketing in ddl section
- [CARBONDATA-711] - Inconsistent data load when single_pass='true'
- [CARBONDATA-712] - 'BAD_RECORDS_ACTION'='REDIRECT' is not working properly.
- [CARBONDATA-716] - Invalid hdfs lock path when load data if config viewfs
- [CARBONDATA-718] - All files have to contain Apache license header
- [CARBONDATA-731] - Enhance and correct quick start and installation guides
- [CARBONDATA-732] - User unable to execute the select/Load query using thrift server.
- [CARBONDATA-733] - Fixed testcase failure issue
- [CARBONDATA-734] - Can't create parquet/orc table with CarbonSession
- [CARBONDATA-735] - Dictionary Loading performance issue with multiple task in single node
- [CARBONDATA-736] - Dictionary Loading issue in Decoder
- [CARBONDATA-738] - Able to load dataframe with boolean type in a carbon table but with null values
- [CARBONDATA-739] - Avoid creating multiple instances of DirectDictionary in DictionaryBasedResultCollector
- [CARBONDATA-740] - Add logger for rows processed while closing in AbstractDataLoadProcessorStep
- [CARBONDATA-741] - Remove the unnecessary classes from carbondata
- [CARBONDATA-743] - Remove the abundant class CarbonFilters.scala
- [CARBONDATA-751] - Adding Header and making footer optional
- [CARBONDATA-752] - creating complex type gives exception
- [CARBONDATA-753] - Fix Date and Timestamp format issues
- [CARBONDATA-756] - RLE encoding isse
- [CARBONDATA-760] - Should to avoid ERROR log for successful select query
- [CARBONDATA-762] - modify all schemaName->databaseName, cubeName->tableName
- [CARBONDATA-766] - Size based blocklet for V3
- [CARBONDATA-770] - Filter Query not null data mismatch issue
- [CARBONDATA-771] - Dataloading fails in V3 format for TPC-DS data.
- [CARBONDATA-774] - Not like operator does not work properly in carbondata
- [CARBONDATA-783] - Loading data with Single Pass 'true' option is throwing an exception
- [CARBONDATA-786] - Data mismatch if the data data is loaded across blocklet groups
- [CARBONDATA-787] - Fixed Memory leak in Offheap Query + added statistics for V3
- [CARBONDATA-788] - Like operator is not working properly
- [CARBONDATA-791] - Exists queries of TPC-DS are failing in carbon
- [CARBONDATA-793] - Count with null values is giving wrong result.
- [CARBONDATA-794] - Numeric dimension column value should be validated for the bad record
- [CARBONDATA-795] - Table Rename command is changing the db name of provided table to current db
- [CARBONDATA-796] - Drop database command is deleting all the carbon files from database folder even if the user does not provide cascade
- [CARBONDATA-797] - Data loss for BigInt datatype if data contains long max and min values
- [CARBONDATA-798] - Update Bad Records folder name during table rename
- [CARBONDATA-800] - ArrayIndexOutOfBoundsException thrown when block size is specified as 2048 MB
- [CARBONDATA-801] - [Documentation] Examples format to be fixed
- [CARBONDATA-802] - Select query is throwing exception if new dictionary column is added without any default value
- [CARBONDATA-803] - Incorrect results returned by not equal to filter on dictionary column with numeric data type
- [CARBONDATA-804] - Update file structure info as per V3 format definition
- [CARBONDATA-809] - Union with alias is returning wrong result.
- [CARBONDATA-811] - Refactor dictionary based result collector class
- [CARBONDATA-814] - bad record log file writing is not correct
- [CARBONDATA-818] - The file_name stored in carbonindex is wrong
- [CARBONDATA-820] - Redundant BitSet created in data load
- [CARBONDATA-821] - Remove Kettle related code and flow from carbon.
- [CARBONDATA-827] - Query statistics log format is incorrect
- [CARBONDATA-828] - Fix length issue of model.dimensions in CarbonGlobalDictionaryGenerateRDD
- [CARBONDATA-829] - DICTIONARY_EXCLUDE is not working when using Spark Datasource DDL
- [CARBONDATA-830] - Incorrect schedule for NewCarbonDataLoadRDD
- [CARBONDATA-832] - Data loading is failing with duplicate header column in csv file
- [CARBONDATA-838] - Alter table add decimal column with default precision and scale is failing in parser.

- [CARBONDATA-839] - Table lock file is not getting deleted after table rename is successful
- [CARBONDATA-843] - null pointer exception is thrown when floor operation is done on decimal column
- [CARBONDATA-845] - Insert Select into same table is not working
- [CARBONDATA-847] - Select query not working properly after alter.
- [CARBONDATA-849] - if alter table ddl is executed on non existing table, then error message is wrong.
- [CARBONDATA-850] - Fix the comment definition issues of CarbonData thrift files
- [CARBONDATA-860] - Carbon with Spark2.1, select query with filter on dictionary column & order by dictionary/measure with limit is failing
- [CARBONDATA-862] - USE_KETTLE option described in [dml-operation-on-carbondata.md](#) document doesn't work
- [CARBONDATA-865] - Remove configurations for Kettle from master/docs/[installation-guide.md](#)
- [CARBONDATA-866] - remove kettle configuration from master/docs/[configuration-parameters.md](#)
- [CARBONDATA-868] - Select query on decimal datatype is not working fine after adding decimal column using alter
- [CARBONDATA-870] - Folders and files not getting cleaned up created locally during data load operation
- [CARBONDATA-871] - If locktype is not configured and store type is HDFS set HDFS lock as default
- [CARBONDATA-873] - Drop table command throwing table already exists exception
- [CARBONDATA-874] - select * from table order by limit query is failing
- [CARBONDATA-875] - create database ddl is creating the database folder with case sensitive name.
- [CARBONDATA-877] - String datatype is throwing an error when included in Dictionary_Exclude in a alter query
- [CARBONDATA-880] - when explain extended is done on a query then store path is getting printed to the user.
- [CARBONDATA-881] - Load status is successful even though system is fail to write status into tablestatus file
- [CARBONDATA-885] - Inconsistent usage of " " in queries in ddl operations on Carbondata
- [CARBONDATA-890] - For Spark 2.1 LRU cache size at driver is getting configured with the executor lru cache size.
- [CARBONDATA-891] - Fix compilation issue of LocalFileLockTest generate new folder "carbon.store"
- [CARBONDATA-892] - IndexOutOfBound exception while running query with 2nd level sub-query
- [CARBONDATA-893] - MR testcase hangs in Hadoop 2.7.2 version profile
- [CARBONDATA-897] - Redundant Fields Inside * **Global Dictionary Configurations** in [Configuration-parameters.md](#)
- [CARBONDATA-898] - When select query and alter table rename table is triggered concurrently, NullPointerException is getting thrown
- [CARBONDATA-900] - Is null query on a newly added measure column is not returning proper results
- [CARBONDATA-903] - data load is not failing even though bad records exists in the data in case of unsafe sort or batch sort
- [CARBONDATA-907] - The grammar for DELETE SEGMENT FOR DATE in website is not correct
- [CARBONDATA-909] - Single pass option in dataframe writer
- [CARBONDATA-911] - Exception raised while creating table using bucketing example in docs
- [CARBONDATA-915] - Call getAll dictionary from codegen of dictionary decoder to improve dictionary load performance
- [CARBONDATA-916] - Major compaction is failing
- [CARBONDATA-919] - result_size query stats is not giving proper row count if vector reader is enabled.
- [CARBONDATA-923] - InsertInto read from one row not working
- [CARBONDATA-925] - CarbonEnv is static & shared among all the Sessions. Cached relation in 1 session is not getting refreshed when another session is adding/dropping column
- [CARBONDATA-930] - Drop table named 'is' throwing exception
- [CARBONDATA-931] - Exception in BigDecimal unsafe store
- [CARBONDATA-932] - Variable length filter query is failing with empty data
- [CARBONDATA-934] - Cast Filter Expression Pushdown in Carbon
- [CARBONDATA-943] - Failing Mathematical functional in spark 2.1 is not displaying proper error message
- [CARBONDATA-949] - Compaction gives NullPointerException after alter table query
- [CARBONDATA-953] - Add validations to Unsafe dataload. And control the data added to threads
- [CARBONDATA-955] - CacheProvider test fails
- [CARBONDATA-957] - Table not found exception in rename table after lock acquire failure
- [CARBONDATA-958] - Schema modified time not updated in modifiedtime.mdt when dictionary column is updated to no-dictionary column due to high cardinality
- [CARBONDATA-960] - Unsafe merge sort is not working properly
- [CARBONDATA-963] - Fixed data mismatch issue and memory leak issue
- [CARBONDATA-964] - Add FAQ-How Carbon will behave when execute insert operation in abnormal scenarios?
- [CARBONDATA-965] - dataload fail message is not correct when there is no good data to load
- [CARBONDATA-967] - select * with order by and limit for join not working
- [CARBONDATA-968] - Alter temp store location and decimal data type incorrect result display correction
- [CARBONDATA-970] - invalid tasks are getting referred even after files are cleaned from memory
- [CARBONDATA-971] - Select query with where condition is failing
- [CARBONDATA-972] - Concurrent ADD COLUMN operation fails
- [CARBONDATA-974] - Index file loading performance issue in case of large cluster
- [CARBONDATA-975] - remove unreasonable code
- [CARBONDATA-976] - Wrong entry getting deleted from schemaEvolution during alter revert
- [CARBONDATA-978] - Range Filter Evaluation Bug
- [CARBONDATA-981] - Configuration can't find HIVE_CONNECTION_URL in yarn-client mode
- [CARBONDATA-984] - change word from Schema to Schema
- [CARBONDATA-985] - Remove unnecessary .show method call in test cases
- [CARBONDATA-986] - Add alter table example
- [CARBONDATA-987] - Can not delete lock file when drop table
- [CARBONDATA-990] - Installing and Configuring CarbonData instruction wrong
- [CARBONDATA-992] - Fix error log in Example module
- [CARBONDATA-1001] - Data type change should support int to long conversion
- [CARBONDATA-1004] - Broadcast join is not happening in spark 2.1
- [CARBONDATA-1005] - Data load does not load all rows when data size is multiples of page size
- [CARBONDATA-1006] - Range filter test case is failing in current master
- [CARBONDATA-1007] - Current unsafe sort does not keep pointers in memory
- [CARBONDATA-1009] - Select statement is not working with empty string in where clause.
- [CARBONDATA-1011] - select * doesn't work after adding column of date type
- [CARBONDATA-1013] - Unexpected characters displays in results while using join query.
- [CARBONDATA-1019] - Like Filter Pushdown
- [CARBONDATA-1022] - Get errors while do "select" query in Spark-shell
- [CARBONDATA-1023] - Able to do load from dataframe with byte data type in carbon table

- [\[CARBONDATA-1027\]](#) - insert into/data load failing for numeric dictionary included column having null value
- [\[CARBONDATA-1033\]](#) - using column with array<string> type bucket table is created but exception thrown when select performed
- [\[CARBONDATA-1037\]](#) - Select query is not returning any data when we query on New Table after Alter Table rename operation
- [\[CARBONDATA-1045\]](#) - Mismatch in message display with insert and load operation on failure due to bad records in update operation

Improvement

- [\[CARBONDATA-683\]](#) - Reduce test time
- [\[CARBONDATA-686\]](#) - Extend period coverage in NOTICE
- [\[CARBONDATA-687\]](#) - Updated Documentation for New Features in Release 1.0.0
- [\[CARBONDATA-694\]](#) - Optimize quick start document through adding hdfs as storepath
- [\[CARBONDATA-695\]](#) - Create CarbonDataFrameExample in example/spark2
- [\[CARBONDATA-701\]](#) - There is a memory leak issue in no kettle loading flow
- [\[CARBONDATA-714\]](#) - DOCUMENTATION - How to handle the bad records
- [\[CARBONDATA-715\]](#) - Optimize Single pass data load
- [\[CARBONDATA-726\]](#) - Update with V3 format for better IO and processing optimization.
- [\[CARBONDATA-730\]](#) - unsupported type: DecimalType
- [\[CARBONDATA-742\]](#) - Add batch sort to improve the loading performance
- [\[CARBONDATA-744\]](#) - The property "spark.carbon.custom.distribution" should be change to carbon.custom.block.distribution and should be part of CarbonProperties
- [\[CARBONDATA-746\]](#) - Support spark-sql CLI for spark2.1 carbon integration
- [\[CARBONDATA-747\]](#) - Add simple performance test for spark2.1 carbon integration
- [\[CARBONDATA-748\]](#) - "between and" filter query is very slow
- [\[CARBONDATA-758\]](#) - remove kettle related code in CarbonExample.scala
- [\[CARBONDATA-769\]](#) - Support Codegen in CarbonDictionaryDecoder
- [\[CARBONDATA-775\]](#) - Update Documentation for Supported Datatypes
- [\[CARBONDATA-781\]](#) - Some SegmentProperties objects occupy too much memory in driver
- [\[CARBONDATA-784\]](#) - Make configurable empty data to be treated as bad record or not And Expose BAD_RECORDS_ACTION default value to be configurable from out side.
- [\[CARBONDATA-790\]](#) - Added statistics for exclusive carbon read(I/O) and scan time
- [\[CARBONDATA-792\]](#) - Range Filter Optimization
- [\[CARBONDATA-799\]](#) - change word from currenr to current
- [\[CARBONDATA-812\]](#) - make vectorized reader as default reader
- [\[CARBONDATA-822\]](#) - Add unsafe sort for bucketing feature
- [\[CARBONDATA-823\]](#) - Refactory of data write step
- [\[CARBONDATA-846\]](#) - Add support to revert changes to alter table commands if there is a failure while executing the changes on hive.
- [\[CARBONDATA-863\]](#) - Support creation and deletion of dictionary files through RDD during alter add and drop
- [\[CARBONDATA-878\]](#) - Inconsistent stylin in [quick-start.md](#) file
- [\[CARBONDATA-884\]](#) - [Documentation] information on assembly jar to be provided in Quick Start
- [\[CARBONDATA-887\]](#) - lazy rdd iterator for InsertInto
- [\[CARBONDATA-914\]](#) - Clear BTree and Dictionary instances from LRU cache on table drop
- [\[CARBONDATA-926\]](#) - Set the max column from options in csv parser settings
- [\[CARBONDATA-927\]](#) - Show segment in data management doc
- [\[CARBONDATA-928\]](#) - Add link to configuration parameters in docs
- [\[CARBONDATA-944\]](#) - Fix wrong log info during drop table in spark-shell
- [\[CARBONDATA-966\]](#) - char and varchar should be supported in ALTER ADD COLUMNS
- [\[CARBONDATA-969\]](#) - Don't persist rdd because it is only use once
- [\[CARBONDATA-993\]](#) - Remove confused variable fs & Fixed spelling mistakes
- [\[CARBONDATA-1044\]](#) - Rename incubator-carbondata to carbondata in docs and readme

Task

- [\[CARBONDATA-659\]](#) - Should add WhitespaceAround and ParenPad to javastyle
- [\[CARBONDATA-676\]](#) - Correcting spelling mistakes and removed unnecessary methods
- [\[CARBONDATA-719\]](#) - Add a release guide in documentation
- [\[CARBONDATA-1020\]](#) - Remove incubator info as top level project