



Spark Release 2.4.7

Spark 2.4.7 is a maintenance release containing stability, correctness, and security fixes. This release is based on the branch-2.4 maintenance branch of Spark. We strongly recommend all 2.4 users to upgrade to this stable release.

Notable changes

[SPARK-28818] - FrequentItems applies an incorrect schema to the resulting dataframe when nulls are present

[SPARK-31511] - Make BytesToBytesMap iterator() thread-safe

[SPARK-31703] - Changes made by SPARK-26985 break reading parquet files correctly in BigEndian architectures (AIX + LinuxPPC64)

[SPARK-31854] - Different results of query execution with wholestage codegen on and off

[SPARK-31903] - toPandas with Arrow enabled doesn't show metrics in Query UI.

[SPARK-31923] - Event log cannot be generated when some internal accumulators use unexpected types

[SPARK-31935] - Hadoop file system config should be effective in data source options

[SPARK-31941] - Handling the exception in SparkUI for getSparkUser method

[SPARK-31967] - Loading jobs UI page takes 40 seconds

[SPARK-31968] - write.partitionBy() creates duplicate subdirectories when user provides duplicate columns

[SPARK-31980] - Spark sequence() fails if start and end of range are identical dates

[SPARK-31997] - Should drop test_udtf table when SingleSessionSuite completed

[SPARK-32000] - Fix the flaky testcase for partially launched task in barrier-mode.

[SPARK-32003] - Shuffle files for lost executor are not unregistered if fetch failure occurs after executor is lost

[SPARK-32024] - Disk usage tracker went negative in HistoryServerDiskManager

[SPARK-32028] - App id link in history summary page point to wrong application attempt

[SPARK-32034] - Port HIVE-14817: Shutdown the SessionManager timeoutChecker thread properly upon shutdown

[SPARK-32044] - [SS] 2.4 Kafka continuous processing print mislead initial offsets log

[SPARK-32098] - Use iloc for positional slicing instead of direct slicing in createDataFrame with Arrow

[SPARK-32115] - Incorrect results for SUBSTRING when overflow

[SPARK-32131] - Fix AnalysisException messages at UNION/INTERSECT/EXCEPT/MINUS operations

[SPARK-32167] - nullability of GetArrayStructFields is incorrect

[SPARK-32214] - The type conversion function generated in makeFromJava for "other" type uses a wrong variable.

[SPARK-32238] - Use Utils.getSimpleName to avoid hitting Malformed class name in ScalaUDF

[SPARK-32280] - AnalysisException thrown when query contains several JOINS

[SPARK-32300] - toPandas with no partitions should work

[SPARK-32344] - Unevaluable expr is set to FIRST/LAST ignoreNullsExpr in distinct aggregates

[SPARK-32364] - Use CaseInsensitiveMap for DataFrameReader/Writer options

[SPARK-32372] - "Resolved attribute(s) XXX missing" after dudup conflict references

[SPARK-32377] - CaseInsensitiveMap should be deterministic for addition

[SPARK-32609] - Incorrect exchange reuse with DataSourceV2

[SPARK-32672] - Data corruption in some cached compressed boolean columns

[SPARK-32693] - Compare two dataframes with same schema except nullable property

[SPARK-32771] - The example of expressions.Aggregator in Javadoc / Scaladoc is wrong

[SPARK-32810] - CSV/JSON data sources should avoid globbing paths when inferring schema

[SPARK-32812] - Run tests script for Python fails in certain environments

Dependency Changes

Known issues

You can consult JIRA for the detailed changes (<https://s.apache.org/v2.4.7-release-notes>).

We would like to acknowledge all community members for contributing patches to this release.

Spark News Archive (</news/>)

Latest News

Spark 3.5.5 released (</news/spark-3-5-5-released.html>) (Feb 27, 2025)

Spark 3.5.4 released (</news/spark-3-5-4-released.html>) (Dec 20, 2024)

Spark 3.4.4 released (</news/spark-3-4-4-released.html>) (Oct 27, 2024)

Preview release of Spark 4.0 (</news/spark-4.0.0-preview2.html>) (Sep 26, 2024)

Archive (</news/index.html>)



(<https://www.apache.org/events/current-event.html>)

DOWNLOAD SPARK ([/DOWNLOADS.HTML](/downloads.html))

Built-in Libraries:

SQL and DataFrames (</sql/>)

Spark Streaming (</streaming/>)

[MLlib \(machine learning\) \(/mllib/\)](#)

[GraphX \(graph\) \(/graphx/\)](#)

[Third-Party Projects \(/third-party-projects.html\)](#)

Apache Spark, Spark, Apache, the Apache feather logo, and the Apache Spark project logo are either registered trademarks or trademarks of The Apache Software Foundation in the United States and other countries. See guidance on use of Apache Spark trademarks ([/trademarks.html](#)). All other marks mentioned may be trademarks or registered trademarks of their respective owners. Copyright © 2018 The Apache Software Foundation, Licensed under the Apache License, Version 2.0 (<https://www.apache.org/licenses/>).