

Spark 1.1.1发布

Spark 1.1.1于美国时间的2014年11月26日正式发布。基于branch-1.1分支，主要修复了一些bug。推荐所有的1.1.0用户更新到这个稳定版本。本次更新共有55位开发者参与。

spark.shuffle.manager仍然使用Hash作为默认值，说明了SORT的Shuffle还不怎么成熟。等待1.2版本吧。

Fixes

Spark

1.1.1修复了

几个组件的bug。在下面

将会列出一些代表性的bug。同时，你也可以到[这里](#)查看所有修复的bug。

Spark Core

这是Spark core模块修复的重要bug。

Avoid many small spills in external data structures (SPARK-4480)

Memory leak in connection manager timeout thread (SPARK-4393)

Incorrect of channel read return value may lead to data truncation (SPARK-4107)

Stream corruption exceptions observed in sort-based shuffle (SPARK-3948)

Integer overflow in sort-based shuffle key comparison (SPARK-3032)

Lack of thread safety in Hadoop configuration usage in Spark (SPARK-2546)

SQL

这是SQL模块修复的重要bug。

Wrong Parquet filters are created for all inequality predicates with literals on the left hand side (SPARK-4468)

Support backticks in aliases (SPARK-3708 and SPARK-3834)

ColumnValue types do not match in Spark rows vs Hive rows (SPARK-3704)

PySpark

这是PySpark模块修复的重要bug。

Fix sortByKey on empty RDD (SPARK-4304)

Avoid using the same random seed for all partitions (SPARK-4148)

Avoid OOMs when take() is run on empty partitions (SPARK-3211)

MLlib

这是MLlib模块修复的重要bug。

KryoException caused by ALS.trainImplicit in PySpark (SPARK-3990)

Streaming

Block replication continuously fails if target is down (SPARK-3495)

Block replication may choose driver as target (SPARK-3496)

GraphX

Ensure VertexRDD.apply uses mergeFunc (SPARK-2062)

本博客文章除特别声明，全部都是原创！
原创文章版权归过往记忆大数据（[过往记忆](#)）所有，未经许可不得转载。
本文链接: [【】](#) ()