Learning_Pyspark_part2

August 30, 2022

```
[1]: #The sudo apt-get update command is used to download package information from
     ↪all configured sources
     !sudo apt update

     # innstall java
     !apt-get install openjdk-8-jdk-headless -qq > /dev/null

     # install spark
     !wget -q https://dlcdn.apache.org/spark/spark-3.3.0/spark-3.3.0-bin-hadoop3.tgz

     # unzip the spark file to the current folder
     !tar -xvf spark-3.3.0-bin-hadoop3.tgz

     # install findspark using pip
     !pip install -q findspark

     #For python users, also install pyspark using the following command.
     !pip install pyspark
```

```
Get:1 https://cloud.r-project.org/bin/linux/ubuntu bionic-cran40/ InRelease
[3,626 B]
Hit:2 http://archive.ubuntu.com/ubuntu bionic InRelease
Get:3 http://archive.ubuntu.com/ubuntu bionic-updates InRelease [88.7 kB]
Get:4 http://security.ubuntu.com/ubuntu bionic-security InRelease [88.7 kB]
Ign:5 https://developer.download.nvidia.com/compute/machine-
learning/repos/ubuntu1804/x86_64  InRelease
Get:6 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu1804/x86_64
InRelease [1,581 B]
Hit:7 https://developer.download.nvidia.com/compute/machine-
learning/repos/ubuntu1804/x86_64  Release
Get:8 http://ppa.launchpad.net/c2d4u.team/c2d4u4.0+/ubuntu bionic InRelease
[15.9 kB]
Get:9 http://archive.ubuntu.com/ubuntu bionic-backports InRelease [74.6 kB]
Get:10 https://cloud.r-project.org/bin/linux/ubuntu bionic-cran40/ Packages
[91.1 kB]
Hit:11 http://ppa.launchpad.net/cran/libgit2/ubuntu bionic InRelease
Get:12
```

```
https://developer.download.nvidia.com/compute/cuda/repos/ubuntu1804/x86_64
Packages [912 kB]
Hit:13 http://ppa.launchpad.net/deadsnakes/ppa/ubuntu bionic InRelease
Hit:15 http://ppa.launchpad.net/graphics-drivers/ppa/ubuntu bionic InRelease
Get:16 http://archive.ubuntu.com/ubuntu bionic-updates/universe amd64 Packages
[2,311 kB]
Get:17 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 Packages
[3,369 kB]
Get:18 http://ppa.launchpad.net/c2d4u.team/c2d4u4.0+/ubuntu bionic/main Sources
[2,094 kB]
Get:19 http://security.ubuntu.com/ubuntu bionic-security/universe amd64 Packages
[1,534 kB]
Get:20 http://security.ubuntu.com/ubuntu bionic-security/main amd64 Packages
[2,939 kB]
Get:21 http://ppa.launchpad.net/c2d4u.team/c2d4u4.0+/ubuntu bionic/main amd64
Packages [1,073 kB]
Fetched 14.6 MB in 3s (4,616 kB/s)
Reading package lists… Done
Building dependency tree
Reading state information… Done
42 packages can be upgraded. Run 'apt list --upgradable' to see them.
spark-3.3.0-bin-hadoop3/
spark-3.3.0-bin-hadoop3/NOTICE
spark-3.3.0-bin-hadoop3/kubernetes/
spark-3.3.0-bin-hadoop3/kubernetes/tests/
spark-3.3.0-bin-hadoop3/kubernetes/tests/python_executable_check.py
spark-3.3.0-bin-hadoop3/kubernetes/tests/autoscale.py
spark-3.3.0-bin-hadoop3/kubernetes/tests/worker_memory_check.py
spark-3.3.0-bin-hadoop3/kubernetes/tests/py_container_checks.py
spark-3.3.0-bin-hadoop3/kubernetes/tests/decommissioning.py
spark-3.3.0-bin-hadoop3/kubernetes/tests/pyfiles.py
spark-3.3.0-bin-hadoop3/kubernetes/tests/decommissioning_cleanup.py
spark-3.3.0-bin-hadoop3/kubernetes/dockerfiles/
spark-3.3.0-bin-hadoop3/kubernetes/dockerfiles/spark/
spark-3.3.0-bin-hadoop3/kubernetes/dockerfiles/spark/decom.sh
spark-3.3.0-bin-hadoop3/kubernetes/dockerfiles/spark/entrypoint.sh
spark-3.3.0-bin-hadoop3/kubernetes/dockerfiles/spark/bindings/
spark-3.3.0-bin-hadoop3/kubernetes/dockerfiles/spark/bindings/R/
spark-3.3.0-bin-hadoop3/kubernetes/dockerfiles/spark/bindings/R/Dockerfile
spark-3.3.0-bin-hadoop3/kubernetes/dockerfiles/spark/bindings/python/
spark-3.3.0-bin-hadoop3/kubernetes/dockerfiles/spark/bindings/python/Dockerfile
spark-3.3.0-bin-hadoop3/kubernetes/dockerfiles/spark/Dockerfile
spark-3.3.0-bin-hadoop3/kubernetes/dockerfiles/spark/Dockerfile.java17
spark-3.3.0-bin-hadoop3/jars/
spark-3.3.0-bin-hadoop3/jars/spark-network-shuffle_2.12-3.3.0.jar
spark-3.3.0-bin-hadoop3/jars/shapeless_2.12-2.3.7.jar
spark-3.3.0-bin-hadoop3/jars/arrow-format-7.0.0.jar
spark-3.3.0-bin-hadoop3/jars/metrics-jvm-4.2.7.jar
```

```
spark-3.3.0-bin-hadoop3/jars/spark-tags_2.12-3.3.0.jar
spark-3.3.0-bin-hadoop3/jars/okhttp-3.12.12.jar
spark-3.3.0-bin-hadoop3/jars/hive-shims-common-2.3.9.jar
spark-3.3.0-bin-hadoop3/jars/spire_2.12-0.17.0.jar
spark-3.3.0-bin-hadoop3/jars/logging-interceptor-3.12.12.jar
spark-3.3.0-bin-hadoop3/jars/hive-serde-2.3.9.jar
spark-3.3.0-bin-hadoop3/jars/objenesis-3.2.jar
spark-3.3.0-bin-hadoop3/jars/spark-streaming_2.12-3.3.0.jar
spark-3.3.0-bin-hadoop3/jars/avro-1.11.0.jar
spark-3.3.0-bin-hadoop3/jars/zookeeper-3.6.2.jar
spark-3.3.0-bin-hadoop3/jars/spark-kvstore_2.12-3.3.0.jar
spark-3.3.0-bin-hadoop3/jars/kryo-shaded-4.0.2.jar
spark-3.3.0-bin-hadoop3/jars/hive-jdbc-2.3.9.jar
spark-3.3.0-bin-hadoop3/jars/jackson-annotations-2.13.3.jar
spark-3.3.0-bin-hadoop3/jars/commons-logging-1.1.3.jar
spark-3.3.0-bin-hadoop3/jars/commons-compiler-3.0.16.jar
spark-3.3.0-bin-hadoop3/jars/spark-sql_2.12-3.3.0.jar
spark-3.3.0-bin-hadoop3/jars/jdo-api-3.0.1.jar
spark-3.3.0-bin-hadoop3/jars/pickle-1.2.jar
spark-3.3.0-bin-hadoop3/jars/JLargeArrays-1.5.jar
spark-3.3.0-bin-hadoop3/jars/kubernetes-model-apiextensions-5.12.2.jar
spark-3.3.0-bin-hadoop3/jars/parquet-common-1.12.2.jar
spark-3.3.0-bin-hadoop3/jars/httpcore-4.4.14.jar
spark-3.3.0-bin-hadoop3/jars/commons-codec-1.15.jar
spark-3.3.0-bin-hadoop3/jars/parquet-format-structures-1.12.2.jar
spark-3.3.0-bin-hadoop3/jars/commons-io-2.11.0.jar
spark-3.3.0-bin-hadoop3/jars/stax-api-1.0.1.jar
spark-3.3.0-bin-hadoop3/jars/hive-storage-api-2.7.2.jar
spark-3.3.0-bin-hadoop3/jars/kubernetes-model-certificates-5.12.2.jar
spark-3.3.0-bin-hadoop3/jars/zookeeper-jute-3.6.2.jar
spark-3.3.0-bin-hadoop3/jars/spark-yarn_2.12-3.3.0.jar
spark-3.3.0-bin-hadoop3/jars/commons-collections4-4.4.jar
spark-3.3.0-bin-hadoop3/jars/jline-2.14.6.jar
spark-3.3.0-bin-hadoop3/jars/dropwizard-metrics-hadoop-
metrics2-reporter-0.1.2.jar
spark-3.3.0-bin-hadoop3/jars/commons-pool-1.5.4.jar
spark-3.3.0-bin-hadoop3/jars/snakeyaml-1.30.jar
spark-3.3.0-bin-hadoop3/jars/xbean-asm9-shaded-4.20.jar
spark-3.3.0-bin-hadoop3/jars/lz4-java-1.8.0.jar
spark-3.3.0-bin-hadoop3/jars/cats-kernel_2.12-2.1.1.jar
spark-3.3.0-bin-hadoop3/jars/jta-1.1.jar
spark-3.3.0-bin-hadoop3/jars/istack-commons-runtime-3.0.8.jar
spark-3.3.0-bin-hadoop3/jars/jackson-module-scala_2.12-2.13.3.jar
spark-3.3.0-bin-hadoop3/jars/hive-common-2.3.9.jar
spark-3.3.0-bin-hadoop3/jars/shims-0.9.25.jar
spark-3.3.0-bin-hadoop3/jars/compress-lzf-1.1.jar
spark-3.3.0-bin-hadoop3/jars/commons-collections-3.2.2.jar
spark-3.3.0-bin-hadoop3/jars/audience-annotations-0.5.0.jar
```

```
spark-3.3.0-bin-hadoop3/jars/joda-time-2.10.13.jar
spark-3.3.0-bin-hadoop3/jars/jpam-1.1.jar
spark-3.3.0-bin-hadoop3/jars/jakarta.annotation-api-1.3.5.jar
spark-3.3.0-bin-hadoop3/jars/kubernetes-model-coordination-5.12.2.jar
spark-3.3.0-bin-hadoop3/jars/tink-1.6.1.jar
spark-3.3.0-bin-hadoop3/jars/metrics-json-4.2.7.jar
spark-3.3.0-bin-hadoop3/jars/spark-kubernetes_2.12-3.3.0.jar
spark-3.3.0-bin-hadoop3/jars/jersey-common-2.34.jar
spark-3.3.0-bin-hadoop3/jars/kubernetes-model-common-5.12.2.jar
spark-3.3.0-bin-hadoop3/jars/spark-graphx_2.12-3.3.0.jar
spark-3.3.0-bin-hadoop3/jars/parquet-hadoop-1.12.2.jar
spark-3.3.0-bin-hadoop3/jars/curator-recipes-2.13.0.jar
spark-3.3.0-bin-hadoop3/jars/zstd-jni-1.5.2-1.jar
spark-3.3.0-bin-hadoop3/jars/commons-math3-3.6.1.jar
spark-3.3.0-bin-hadoop3/jars/spark-tags_2.12-3.3.0-tests.jar
spark-3.3.0-bin-hadoop3/jars/orc-shims-1.7.4.jar
spark-3.3.0-bin-hadoop3/jars/netty-transport-4.1.74.Final.jar
spark-3.3.0-bin-hadoop3/jars/netty-tcnative-classes-2.0.48.Final.jar
spark-3.3.0-bin-hadoop3/jars/spark-launcher_2.12-3.3.0.jar
spark-3.3.0-bin-hadoop3/jars/antlr4-runtime-4.8.jar
spark-3.3.0-bin-hadoop3/jars/javax.jdo-3.2.0-m3.jar
spark-3.3.0-bin-hadoop3/jars/spark-hive-thriftserver_2.12-3.3.0.jar
spark-3.3.0-bin-hadoop3/jars/paranamer-2.8.jar
spark-3.3.0-bin-hadoop3/jars/spark-network-common_2.12-3.3.0.jar
spark-3.3.0-bin-hadoop3/jars/netty-all-4.1.74.Final.jar
spark-3.3.0-bin-hadoop3/jars/arrow-memory-core-7.0.0.jar
spark-3.3.0-bin-hadoop3/jars/orc-core-1.7.4.jar
spark-3.3.0-bin-hadoop3/jars/kubernetes-model-storageclass-5.12.2.jar
spark-3.3.0-bin-hadoop3/jars/flatbuffers-java-1.12.0.jar
spark-3.3.0-bin-hadoop3/jars/kubernetes-model-autoscaling-5.12.2.jar
spark-3.3.0-bin-hadoop3/jars/json4s-core_2.12-3.7.0-M11.jar
spark-3.3.0-bin-hadoop3/jars/stream-2.9.6.jar
spark-3.3.0-bin-hadoop3/jars/log4j-core-2.17.2.jar
spark-3.3.0-bin-hadoop3/jars/arpack-2.2.1.jar
spark-3.3.0-bin-hadoop3/jars/avro-ipc-1.11.0.jar
spark-3.3.0-bin-hadoop3/jars/hive-shims-2.3.9.jar
spark-3.3.0-bin-hadoop3/jars/datanucleus-api-jdo-4.2.4.jar
spark-3.3.0-bin-hadoop3/jars/gson-2.2.4.jar
spark-3.3.0-bin-hadoop3/jars/jersey-container-servlet-2.34.jar
spark-3.3.0-bin-hadoop3/jars/commons-compress-1.21.jar
spark-3.3.0-bin-hadoop3/jars/curator-client-2.13.0.jar
spark-3.3.0-bin-hadoop3/jars/opencsv-2.3.jar
spark-3.3.0-bin-hadoop3/jars/hive-exec-2.3.9-core.jar
spark-3.3.0-bin-hadoop3/jars/jsr305-3.0.0.jar
spark-3.3.0-bin-hadoop3/jars/netty-transport-native-epoll-4.1.74.Final-
linux-x86_64.jar
spark-3.3.0-bin-hadoop3/jars/derby-10.14.2.0.jar
spark-3.3.0-bin-hadoop3/jars/spark-catalyst_2.12-3.3.0.jar
```

```
spark-3.3.0-bin-hadoop3/jars/commons-dbcp-1.4.jar
spark-3.3.0-bin-hadoop3/jars/scala-reflect-2.12.15.jar
spark-3.3.0-bin-hadoop3/jars/okio-1.14.0.jar
spark-3.3.0-bin-hadoop3/jars/hk2-api-2.6.1.jar
spark-3.3.0-bin-hadoop3/jars/netty-transport-native-epoll-4.1.74.Final-linux-
aarch_64.jar
spark-3.3.0-bin-hadoop3/jars/hive-shims-0.23-2.3.9.jar
spark-3.3.0-bin-hadoop3/jars/spark-hive_2.12-3.3.0.jar
spark-3.3.0-bin-hadoop3/jars/jakarta.servlet-api-4.0.3.jar
spark-3.3.0-bin-hadoop3/jars/commons-lang3-3.12.0.jar
spark-3.3.0-bin-hadoop3/jars/automaton-1.11-8.jar
spark-3.3.0-bin-hadoop3/jars/commons-lang-2.6.jar
spark-3.3.0-bin-hadoop3/jars/kubernetes-model-core-5.12.2.jar
spark-3.3.0-bin-hadoop3/jars/ivy-2.5.0.jar
spark-3.3.0-bin-hadoop3/jars/hk2-utils-2.6.1.jar
spark-3.3.0-bin-hadoop3/jars/parquet-column-1.12.2.jar
spark-3.3.0-bin-hadoop3/jars/kubernetes-model-events-5.12.2.jar
spark-3.3.0-bin-hadoop3/jars/kubernetes-model-node-5.12.2.jar
spark-3.3.0-bin-hadoop3/jars/velocity-1.5.jar
spark-3.3.0-bin-hadoop3/jars/netty-buffer-4.1.74.Final.jar
spark-3.3.0-bin-hadoop3/jars/httpclient-4.5.13.jar
spark-3.3.0-bin-hadoop3/jars/JTransforms-3.1.jar
spark-3.3.0-bin-hadoop3/jars/netty-handler-4.1.74.Final.jar
spark-3.3.0-bin-hadoop3/jars/json4s-scalap_2.12-3.7.0-M11.jar
spark-3.3.0-bin-hadoop3/jars/slf4j-api-1.7.32.jar
spark-3.3.0-bin-hadoop3/jars/kubernetes-model-rbac-5.12.2.jar
spark-3.3.0-bin-hadoop3/jars/spark-repl_2.12-3.3.0.jar
spark-3.3.0-bin-hadoop3/jars/kubernetes-model-discovery-5.12.2.jar
spark-3.3.0-bin-hadoop3/jars/RoaringBitmap-0.9.25.jar
spark-3.3.0-bin-hadoop3/jars/jersey-server-2.34.jar
spark-3.3.0-bin-hadoop3/jars/netty-transport-classes-kqueue-4.1.74.Final.jar
spark-3.3.0-bin-hadoop3/jars/aircompressor-0.21.jar
spark-3.3.0-bin-hadoop3/jars/arrow-vector-7.0.0.jar
spark-3.3.0-bin-hadoop3/jars/jackson-core-2.13.3.jar
spark-3.3.0-bin-hadoop3/jars/jcl-over-slf4j-1.7.32.jar
spark-3.3.0-bin-hadoop3/jars/metrics-core-4.2.7.jar
spark-3.3.0-bin-hadoop3/jars/spark-core_2.12-3.3.0.jar
spark-3.3.0-bin-hadoop3/jars/jakarta.validation-api-2.0.2.jar
spark-3.3.0-bin-hadoop3/jars/jul-to-slf4j-1.7.32.jar
spark-3.3.0-bin-hadoop3/jars/netty-transport-native-kqueue-4.1.74.Final-
osx-x86_64.jar
spark-3.3.0-bin-hadoop3/jars/guava-14.0.1.jar
spark-3.3.0-bin-hadoop3/jars/parquet-jackson-1.12.2.jar
spark-3.3.0-bin-hadoop3/jars/zjsonpatch-0.3.0.jar
spark-3.3.0-bin-hadoop3/jars/hive-service-rpc-3.1.2.jar
spark-3.3.0-bin-hadoop3/jars/metrics-graphite-4.2.7.jar
spark-3.3.0-bin-hadoop3/jars/oro-2.0.8.jar
spark-3.3.0-bin-hadoop3/jars/commons-text-1.9.jar
```

```
spark-3.3.0-bin-hadoop3/jars/libfb303-0.9.3.jar
spark-3.3.0-bin-hadoop3/jars/core-1.1.2.jar
spark-3.3.0-bin-hadoop3/jars/avro-mapred-1.11.0.jar
spark-3.3.0-bin-hadoop3/jars/datanucleus-rdbms-4.1.19.jar
spark-3.3.0-bin-hadoop3/jars/super-csv-2.2.0.jar
spark-3.3.0-bin-hadoop3/jars/hadoop-yarn-server-web-proxy-3.3.2.jar
spark-3.3.0-bin-hadoop3/jars/scala-compiler-2.12.15.jar
spark-3.3.0-bin-hadoop3/jars/log4j-api-2.17.2.jar
spark-3.3.0-bin-hadoop3/jars/osgi-resource-locator-1.0.3.jar
spark-3.3.0-bin-hadoop3/jars/spark-sketch_2.12-3.3.0.jar
spark-3.3.0-bin-hadoop3/jars/hadoop-client-api-3.3.2.jar
spark-3.3.0-bin-hadoop3/jars/arrow-memory-netty-7.0.0.jar
spark-3.3.0-bin-hadoop3/jars/json-1.8.jar
spark-3.3.0-bin-hadoop3/jars/kubernetes-model-policy-5.12.2.jar
spark-3.3.0-bin-hadoop3/jars/antlr-runtime-3.5.2.jar
spark-3.3.0-bin-hadoop3/jars/jackson-datatype-jsr310-2.13.3.jar
spark-3.3.0-bin-hadoop3/jars/jackson-databind-2.13.3.jar
spark-3.3.0-bin-hadoop3/jars/spark-mllib_2.12-3.3.0.jar
spark-3.3.0-bin-hadoop3/jars/netty-transport-classes-epoll-4.1.74.Final.jar
spark-3.3.0-bin-hadoop3/jars/threeten-extra-1.5.0.jar
spark-3.3.0-bin-hadoop3/jars/kubernetes-client-5.12.2.jar
spark-3.3.0-bin-hadoop3/jars/scala-library-2.12.15.jar
spark-3.3.0-bin-hadoop3/jars/kubernetes-model-metrics-5.12.2.jar
spark-3.3.0-bin-hadoop3/jars/snappy-java-1.1.8.4.jar
spark-3.3.0-bin-hadoop3/jars/activation-1.1.1.jar
spark-3.3.0-bin-hadoop3/jars/annotations-17.0.0.jar
spark-3.3.0-bin-hadoop3/jars/commons-crypto-1.1.0.jar
spark-3.3.0-bin-hadoop3/jars/log4j-1.2-api-2.17.2.jar
spark-3.3.0-bin-hadoop3/jars/libthrift-0.12.0.jar
spark-3.3.0-bin-hadoop3/jars/HikariCP-2.5.1.jar
spark-3.3.0-bin-hadoop3/jars/generex-1.0.2.jar
spark-3.3.0-bin-hadoop3/jars/leveldbjni-all-1.8.jar
spark-3.3.0-bin-hadoop3/jars/jaxb-runtime-2.3.2.jar
spark-3.3.0-bin-hadoop3/jars/rocksdbjni-6.20.3.jar
spark-3.3.0-bin-hadoop3/jars/netty-transport-native-unix-common-4.1.74.Final.jar
spark-3.3.0-bin-hadoop3/jars/lapack-2.2.1.jar
spark-3.3.0-bin-hadoop3/jars/jackson-core-asl-1.9.13.jar
spark-3.3.0-bin-hadoop3/jars/spark-mllib-local_2.12-3.3.0.jar
spark-3.3.0-bin-hadoop3/jars/hive-cli-2.3.9.jar
spark-3.3.0-bin-hadoop3/jars/hive-beeline-2.3.9.jar
spark-3.3.0-bin-hadoop3/jars/scala-xml_2.12-1.2.0.jar
spark-3.3.0-bin-hadoop3/jars/hk2-locator-2.6.1.jar
spark-3.3.0-bin-hadoop3/jars/kubernetes-model-apps-5.12.2.jar
spark-3.3.0-bin-hadoop3/jars/metrics-jmx-4.2.7.jar
spark-3.3.0-bin-hadoop3/jars/jackson-mapper-asl-1.9.13.jar
spark-3.3.0-bin-hadoop3/jars/blas-2.2.1.jar
spark-3.3.0-bin-hadoop3/jars/ST4-4.0.4.jar
spark-3.3.0-bin-hadoop3/jars/jackson-dataformat-yaml-2.13.3.jar
```

```
spark-3.3.0-bin-hadoop3/jars/chill_2.12-0.10.0.jar
spark-3.3.0-bin-hadoop3/jars/json4s-jackson_2.12-3.7.0-M11.jar
spark-3.3.0-bin-hadoop3/jars/netty-resolver-4.1.74.Final.jar
spark-3.3.0-bin-hadoop3/jars/breeze-macros_2.12-1.2.jar
spark-3.3.0-bin-hadoop3/jars/xz-1.8.jar
spark-3.3.0-bin-hadoop3/jars/minlog-1.3.0.jar
spark-3.3.0-bin-hadoop3/jars/jersey-hk2-2.34.jar
spark-3.3.0-bin-hadoop3/jars/netty-common-4.1.74.Final.jar
spark-3.3.0-bin-hadoop3/jars/jakarta.ws.rs-api-2.1.6.jar
spark-3.3.0-bin-hadoop3/jars/spark-unsafe_2.12-3.3.0.jar
spark-3.3.0-bin-hadoop3/jars/jersey-container-servlet-core-2.34.jar
spark-3.3.0-bin-hadoop3/jars/aopalliance-repackaged-2.6.1.jar
spark-3.3.0-bin-hadoop3/jars/netty-transport-native-kqueue-4.1.74.Final-osx-
aarch_64.jar
spark-3.3.0-bin-hadoop3/jars/bonecp-0.8.0.RELEASE.jar
spark-3.3.0-bin-hadoop3/jars/jodd-core-3.5.2.jar
spark-3.3.0-bin-hadoop3/jars/orc-mapreduce-1.7.4.jar
spark-3.3.0-bin-hadoop3/jars/spire-util_2.12-0.17.0.jar
spark-3.3.0-bin-hadoop3/jars/protobuf-java-2.5.0.jar
spark-3.3.0-bin-hadoop3/jars/jersey-client-2.34.jar
spark-3.3.0-bin-hadoop3/jars/datanucleus-core-4.1.17.jar
spark-3.3.0-bin-hadoop3/jars/janino-3.0.16.jar
spark-3.3.0-bin-hadoop3/jars/spire-platform_2.12-0.17.0.jar
spark-3.3.0-bin-hadoop3/jars/javolution-5.5.1.jar
spark-3.3.0-bin-hadoop3/jars/transaction-api-1.1.jar
spark-3.3.0-bin-hadoop3/jars/mesos-1.4.3-shaded-protobuf.jar
spark-3.3.0-bin-hadoop3/jars/curator-framework-2.13.0.jar
spark-3.3.0-bin-hadoop3/jars/parquet-encoding-1.12.2.jar
spark-3.3.0-bin-hadoop3/jars/jakarta.inject-2.6.1.jar
spark-3.3.0-bin-hadoop3/jars/hive-llap-common-2.3.9.jar
spark-3.3.0-bin-hadoop3/jars/commons-cli-1.5.0.jar
spark-3.3.0-bin-hadoop3/jars/jakarta.xml.bind-api-2.3.2.jar
spark-3.3.0-bin-hadoop3/jars/kubernetes-model-flowcontrol-5.12.2.jar
spark-3.3.0-bin-hadoop3/jars/algebra_2.12-2.0.1.jar
spark-3.3.0-bin-hadoop3/jars/hadoop-shaded-guava-1.1.1.jar
spark-3.3.0-bin-hadoop3/jars/arpack_combined_all-0.1.jar
spark-3.3.0-bin-hadoop3/jars/hive-vector-code-gen-2.3.9.jar
spark-3.3.0-bin-hadoop3/jars/kubernetes-model-admissionregistration-5.12.2.jar
spark-3.3.0-bin-hadoop3/jars/spark-mesos_2.12-3.3.0.jar
spark-3.3.0-bin-hadoop3/jars/json4s-ast_2.12-3.7.0-M11.jar
spark-3.3.0-bin-hadoop3/jars/log4j-slf4j-impl-2.17.2.jar
spark-3.3.0-bin-hadoop3/jars/javassist-3.25.0-GA.jar
spark-3.3.0-bin-hadoop3/jars/scala-collection-compat_2.12-2.1.1.jar
spark-3.3.0-bin-hadoop3/jars/kubernetes-model-networking-5.12.2.jar
spark-3.3.0-bin-hadoop3/jars/py4j-0.10.9.5.jar
spark-3.3.0-bin-hadoop3/jars/netty-codec-4.1.74.Final.jar
spark-3.3.0-bin-hadoop3/jars/hadoop-client-runtime-3.3.2.jar
spark-3.3.0-bin-hadoop3/jars/univocity-parsers-2.9.1.jar
```

```
spark-3.3.0-bin-hadoop3/jars/hive-shims-scheduler-2.3.9.jar
spark-3.3.0-bin-hadoop3/jars/kubernetes-model-extensions-5.12.2.jar
spark-3.3.0-bin-hadoop3/jars/breeze_2.12-1.2.jar
spark-3.3.0-bin-hadoop3/jars/chill-java-0.10.0.jar
spark-3.3.0-bin-hadoop3/jars/spire-macros_2.12-0.17.0.jar
spark-3.3.0-bin-hadoop3/jars/scala-parser-combinators_2.12-1.1.2.jar
spark-3.3.0-bin-hadoop3/jars/hive-metastore-2.3.9.jar
spark-3.3.0-bin-hadoop3/jars/kubernetes-model-scheduling-5.12.2.jar
spark-3.3.0-bin-hadoop3/jars/kubernetes-model-batch-5.12.2.jar
spark-3.3.0-bin-hadoop3/data/
spark-3.3.0-bin-hadoop3/data/mllib/
spark-3.3.0-bin-hadoop3/data/mllib/sample_lda_data.txt
spark-3.3.0-bin-hadoop3/data/mllib/sample_libsvm_data.txt
spark-3.3.0-bin-hadoop3/data/mllib/sample_svm_data.txt
spark-3.3.0-bin-hadoop3/data/mllib/sample_multiclass_classification_data.txt
spark-3.3.0-bin-hadoop3/data/mllib/sample_linear_regression_data.txt
spark-3.3.0-bin-hadoop3/data/mllib/sample_lda_libsvm_data.txt
spark-3.3.0-bin-hadoop3/data/mllib/sample_fpgrowth.txt
spark-3.3.0-bin-hadoop3/data/mllib/sample_binary_classification_data.txt
spark-3.3.0-bin-hadoop3/data/mllib/sample_isotonic_regression_libsvm_data.txt
spark-3.3.0-bin-hadoop3/data/mllib/sample_movielens_data.txt
spark-3.3.0-bin-hadoop3/data/mllib/als/
spark-3.3.0-bin-hadoop3/data/mllib/als/test.data
spark-3.3.0-bin-hadoop3/data/mllib/als/sample_movielens_ratings.txt
spark-3.3.0-bin-hadoop3/data/mllib/sample_kmeans_data.txt
spark-3.3.0-bin-hadoop3/data/mllib/pic_data.txt
spark-3.3.0-bin-hadoop3/data/mllib/images/
spark-3.3.0-bin-hadoop3/data/mllib/images/origin/
spark-3.3.0-bin-hadoop3/data/mllib/images/origin/kittens/
spark-3.3.0-bin-hadoop3/data/mllib/images/origin/kittens/54893.jpg
spark-3.3.0-bin-hadoop3/data/mllib/images/origin/kittens/not-image.txt
spark-3.3.0-bin-hadoop3/data/mllib/images/origin/kittens/DP802813.jpg
spark-3.3.0-bin-hadoop3/data/mllib/images/origin/kittens/29.5.a_b_EGDP022204.jpg
spark-3.3.0-bin-hadoop3/data/mllib/images/origin/kittens/DP153539.jpg
spark-3.3.0-bin-hadoop3/data/mllib/images/origin/license.txt
spark-3.3.0-bin-hadoop3/data/mllib/images/origin/multi-channel/
spark-3.3.0-bin-hadoop3/data/mllib/images/origin/multi-channel/BGRA.png
spark-3.3.0-bin-hadoop3/data/mllib/images/origin/multi-channel/BGRA_alpha_60.png
spark-3.3.0-bin-hadoop3/data/mllib/images/origin/multi-channel/grayscale.jpg
spark-3.3.0-bin-hadoop3/data/mllib/images/origin/multi-channel/chr30.4.184.jpg
spark-3.3.0-bin-hadoop3/data/mllib/images/license.txt
spark-3.3.0-bin-hadoop3/data/mllib/ridge-data/
spark-3.3.0-bin-hadoop3/data/mllib/ridge-data/lpsa.data
spark-3.3.0-bin-hadoop3/data/mllib/kmeans_data.txt
spark-3.3.0-bin-hadoop3/data/mllib/streaming_kmeans_data_test.txt
spark-3.3.0-bin-hadoop3/data/mllib/pagerank_data.txt
spark-3.3.0-bin-hadoop3/data/mllib/gmm_data.txt
spark-3.3.0-bin-hadoop3/data/graphx/
```

```
spark-3.3.0-bin-hadoop3/data/graphx/users.txt
spark-3.3.0-bin-hadoop3/data/graphx/followers.txt
spark-3.3.0-bin-hadoop3/data/streaming/
spark-3.3.0-bin-hadoop3/data/streaming/AFINN-111.txt
spark-3.3.0-bin-hadoop3/R/
spark-3.3.0-bin-hadoop3/R/lib/
spark-3.3.0-bin-hadoop3/R/lib/sparkr.zip
spark-3.3.0-bin-hadoop3/R/lib/SparkR/
spark-3.3.0-bin-hadoop3/R/lib/SparkR/tests/
spark-3.3.0-bin-hadoop3/R/lib/SparkR/tests/testthat/
spark-3.3.0-bin-hadoop3/R/lib/SparkR/tests/testthat/test_basic.R
spark-3.3.0-bin-hadoop3/R/lib/SparkR/DESCRIPTION
spark-3.3.0-bin-hadoop3/R/lib/SparkR/profile/
spark-3.3.0-bin-hadoop3/R/lib/SparkR/profile/shell.R
spark-3.3.0-bin-hadoop3/R/lib/SparkR/profile/general.R
spark-3.3.0-bin-hadoop3/R/lib/SparkR/INDEX
spark-3.3.0-bin-hadoop3/R/lib/SparkR/Meta/
spark-3.3.0-bin-hadoop3/R/lib/SparkR/Meta/features.rds
spark-3.3.0-bin-hadoop3/R/lib/SparkR/Meta/links.rds
spark-3.3.0-bin-hadoop3/R/lib/SparkR/Meta/nsInfo.rds
spark-3.3.0-bin-hadoop3/R/lib/SparkR/Meta/package.rds
spark-3.3.0-bin-hadoop3/R/lib/SparkR/Meta/vignette.rds
spark-3.3.0-bin-hadoop3/R/lib/SparkR/Meta/Rd.rds
spark-3.3.0-bin-hadoop3/R/lib/SparkR/Meta/hsearch.rds
spark-3.3.0-bin-hadoop3/R/lib/SparkR/help/
spark-3.3.0-bin-hadoop3/R/lib/SparkR/help/SparkR.rdb
spark-3.3.0-bin-hadoop3/R/lib/SparkR/help/aliases.rds
spark-3.3.0-bin-hadoop3/R/lib/SparkR/help/SparkR.rdx
spark-3.3.0-bin-hadoop3/R/lib/SparkR/help/AnIndex
spark-3.3.0-bin-hadoop3/R/lib/SparkR/help/paths.rds
spark-3.3.0-bin-hadoop3/R/lib/SparkR/R/
spark-3.3.0-bin-hadoop3/R/lib/SparkR/R/SparkR.rdb
spark-3.3.0-bin-hadoop3/R/lib/SparkR/R/SparkR.rdx
spark-3.3.0-bin-hadoop3/R/lib/SparkR/R/SparkR
spark-3.3.0-bin-hadoop3/R/lib/SparkR/NAMESPACE
spark-3.3.0-bin-hadoop3/R/lib/SparkR/doc/
spark-3.3.0-bin-hadoop3/R/lib/SparkR/doc/sparkr-vignettes.Rmd
spark-3.3.0-bin-hadoop3/R/lib/SparkR/doc/index.html
spark-3.3.0-bin-hadoop3/R/lib/SparkR/doc/sparkr-vignettes.html
spark-3.3.0-bin-hadoop3/R/lib/SparkR/doc/sparkr-vignettes.R
spark-3.3.0-bin-hadoop3/R/lib/SparkR/html/
spark-3.3.0-bin-hadoop3/R/lib/SparkR/html/00Index.html
spark-3.3.0-bin-hadoop3/R/lib/SparkR/html/R.css
spark-3.3.0-bin-hadoop3/R/lib/SparkR/worker/
spark-3.3.0-bin-hadoop3/R/lib/SparkR/worker/worker.R
spark-3.3.0-bin-hadoop3/R/lib/SparkR/worker/daemon.R
spark-3.3.0-bin-hadoop3/README.md
spark-3.3.0-bin-hadoop3/RELEASE
```

```
spark-3.3.0-bin-hadoop3/yarn/
spark-3.3.0-bin-hadoop3/yarn/spark-3.3.0-yarn-shuffle.jar
spark-3.3.0-bin-hadoop3/LICENSE
spark-3.3.0-bin-hadoop3/sbin/
spark-3.3.0-bin-hadoop3/sbin/start-workers.sh
spark-3.3.0-bin-hadoop3/sbin/stop-mesos-shuffle-service.sh
spark-3.3.0-bin-hadoop3/sbin/start-master.sh
spark-3.3.0-bin-hadoop3/sbin/workers.sh
spark-3.3.0-bin-hadoop3/sbin/start-worker.sh
spark-3.3.0-bin-hadoop3/sbin/spark-config.sh
spark-3.3.0-bin-hadoop3/sbin/start-history-server.sh
spark-3.3.0-bin-hadoop3/sbin/start-slaves.sh
spark-3.3.0-bin-hadoop3/sbin/spark-daemon.sh
spark-3.3.0-bin-hadoop3/sbin/stop-worker.sh
spark-3.3.0-bin-hadoop3/sbin/stop-mesos-dispatcher.sh
spark-3.3.0-bin-hadoop3/sbin/decommission-worker.sh
spark-3.3.0-bin-hadoop3/sbin/start-mesos-shuffle-service.sh
spark-3.3.0-bin-hadoop3/sbin/decommission-slave.sh
spark-3.3.0-bin-hadoop3/sbin/slaves.sh
spark-3.3.0-bin-hadoop3/sbin/stop-history-server.sh
spark-3.3.0-bin-hadoop3/sbin/start-thriftserver.sh
spark-3.3.0-bin-hadoop3/sbin/stop-thriftserver.sh
spark-3.3.0-bin-hadoop3/sbin/start-slave.sh
spark-3.3.0-bin-hadoop3/sbin/start-all.sh
spark-3.3.0-bin-hadoop3/sbin/stop-slave.sh
spark-3.3.0-bin-hadoop3/sbin/spark-daemons.sh
spark-3.3.0-bin-hadoop3/sbin/stop-workers.sh
spark-3.3.0-bin-hadoop3/sbin/stop-slaves.sh
spark-3.3.0-bin-hadoop3/sbin/stop-all.sh
spark-3.3.0-bin-hadoop3/sbin/start-mesos-dispatcher.sh
spark-3.3.0-bin-hadoop3/sbin/stop-master.sh
spark-3.3.0-bin-hadoop3/examples/
spark-3.3.0-bin-hadoop3/examples/src/
spark-3.3.0-bin-hadoop3/examples/src/main/
spark-3.3.0-bin-hadoop3/examples/src/main/r/
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/survreg.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/glm.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/prefixSpan.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/lm_with_elastic_net.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/powerIterationClustering.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/lda.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/kstest.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/isoreg.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/ml.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/naiveBayes.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/fmRegressor.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/mlp.R
```

```
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/als.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/kmeans.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/svmLinear.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/fmClassifier.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/logit.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/randomForest.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/gbt.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/decisionTree.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/gaussianMixture.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/bisectingKmeans.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/ml/fpm.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/dataframe.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/RSparkSQLExample.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/data-manipulation.R
spark-3.3.0-bin-hadoop3/examples/src/main/r/streaming/
spark-3.3.0-bin-
hadoop3/examples/src/main/r/streaming/structured_network_wordcount.R
spark-3.3.0-bin-hadoop3/examples/src/main/scala/
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/SparkPageRank.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/Skewed
GroupByTest.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/Accumu
latorMetricsTest.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/HdfsTest.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Wor
d2VecExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Dev
eloperApiExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Min
HashLSHExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Ind
exToStringExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Bis
ectingKMeansExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Cor
relationExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Mul
ticlassLogisticRegressionWithElasticNetExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Vec
torSizeHintExample.scala
spark-3.3.0-bin-
```

hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/NGramExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Buc
ketedRandomProjectionLSHExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Bin
arizerExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Ran
domForestExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Sta
ndardScalerExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Log
isticRegressionExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Fea
tureHasherExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/One
VsRestExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Sto
pWordsRemoverExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Lin
earSVCExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Gau
ssianMixtureExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Max
AbsScalerExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Var
ianceThresholdSelectorExample.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/GBTExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Iso
tonicRegressionExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Mod
elSelectionViaCrossValidationExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Est
imatorTransformerParamExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Dec
isionTreeRegressionExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Ran
domForestRegressorExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Pre
fixSpanExample.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/KMeansExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Mod
elSelectionViaTrainValidationSplitExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Log
isticRegressionSummaryExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Mul
tilayerPerceptronClassifierExample.scala
spark-3.3.0-bin-

hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/PCAExample.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/DCTExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/FPG
rowthExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Ele
mentwiseProductExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Gen
eralizedLinearRegressionExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Uni
variateFeatureSelectorExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Cou
ntVectorizerExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Min
MaxScalerExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Chi
SquareTestExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Log
isticRegressionWithElasticNetExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/FMR
egressorExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Vec
torSlicerExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Lin
earRegressionExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Str
ingIndexerExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Gra
dientBoostedTreeRegressorExample.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/ALSExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Dec
isionTreeClassificationExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Pol
ynomialExpansionExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/FMC
lassifierExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Una
ryTransformerExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Imp
uterExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Tok
enizerExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Dec
isionTreeExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Gra
dientBoostedTreeClassifierExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Pow

erIterationClusteringExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/SQL
TransformerExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Sum
marizerExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Dat
aFrameExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Buc
ketizerExample.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/TfIdfExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/AFT
SurvivalRegressionExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Qua
ntileDiscretizerExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/One
HotEncoderExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Pip
elineExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Int
eractionExample.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/LDAExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/RFo
rmulaExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Vec
torIndexerExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Vec
torAssemblerExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Nor
malizerExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Chi
SqSelectorExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Nai
veBayesExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Ran
domForestClassifierExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Lin
earRegressionWithElasticNetExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/ml/Rob
ustScalerExample.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/LogQuery.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/SparkR
emoteFileTest.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/Except
ionHandlingTest.scala
spark-3.3.0-bin-

hadoop3/examples/src/main/scala/org/apache/spark/examples/pythonconverters/
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/python
converters/AvroConverters.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/BroadcastTest.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/SparkHdfsLR.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
Word2VecExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
GradientBoostedTreesRunner.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
RandomRDDGeneration.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
PMMLModelExportExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
BisectingKMeansExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
DecisionTreeRunner.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
MultiLabelMetricsExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
CorrelationsExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
LBFGSExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
TFIDFExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
StandardScalerExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
LogisticRegressionWithLBFGSExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
StreamingTestExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
BinaryClassification.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
DenseKMeans.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
GradientBoostingClassificationExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
RankingMetricsExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
GaussianMixtureExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
MultivariateSummarizer.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
IsotonicRegressionExample.scala

spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
PCAOnSourceVectorExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
Correlations.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
HypothesisTestingExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
DecisionTreeRegressionExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
AssociationRulesExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
StreamingLogisticRegression.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/SVDExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
PrefixSpanExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
KMeansExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
SparseNaiveBayes.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
AbstractParams.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
BinaryClassificationMetricsExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
SampledRDDs.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
FPGrowthExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
ElementwiseProductExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
RecommendationExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
TallSkinnySVD.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
HypothesisTestingKolmogorovSmirnovTestExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
GradientBoostingRegressionExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
StreamingKMeansExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
CosineSimilarity.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
RandomForestRegressionExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
StreamingLinearRegressionExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
StratifiedSamplingExample.scala

spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
DecisionTreeClassificationExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
TallSkinnyPCA.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
RandomForestClassificationExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
SimpleFPGrowth.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
LatentDirichletAllocationExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
KernelDensityEstimationExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
PowerIterationClusteringExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
SVMWithSGDExample.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/LDAExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
MovieLensALS.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
NormalizerExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
MulticlassMetricsExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
ChiSqSelectorExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
SummaryStatisticsExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
PCAOnRowMatrixExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/mllib/
NaiveBayesExample.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/SparkTC.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/Driver
SubmissionTest.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/MultiB
roadcastTest.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/SparkLR.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/SparkPi.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/graphx/
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/graphx
/ConnectedComponentsExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/graphx
/LiveJournalPageRank.scala

spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/graphx
/AggregateMessagesExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/graphx
/SSSPExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/graphx
/SynthBenchmark.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/graphx
/TriangleCountingExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/graphx
/PageRankExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/graphx
/ComprehensiveExample.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/graphx/Analytics.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/SparkKMeans.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/streaming/
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/streaming/clickstream/
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/stream
ing/clickstream/PageViewGenerator.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/stream
ing/clickstream/PageViewStream.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/stream
ing/DirectKafkaWordCount.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/stream
ing/HdfsWordCount.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/stream
ing/DirectKerberizedKafkaWordCount.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/stream
ing/RecoverableNetworkWordCount.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/stream
ing/NetworkWordCount.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/stream
ing/QueueStream.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/stream
ing/SqlNetworkWordCount.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/stream
ing/CustomReceiver.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/stream
ing/RawNetworkGrep.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/stream
ing/StatefulNetworkWordCount.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/stream
ing/StreamingExamples.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/LocalLR.scala

```
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/LocalKMeans.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/GroupByTest.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/LocalALS.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/sql/
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/sql/Si
mpleTypedAggregator.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/sql/Us
erDefinedUntypedAggregation.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/sql/hive/
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/sql/hi
ve/SparkHiveExample.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/sql/RDDRelation.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/sql/Sp
arkSQLExample.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/sql/SQ
LDataSourceExample.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/sql/streaming/
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/sql/st
reaming/StructuredKerberizedKafkaWordCount.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/sql/st
reaming/StructuredComplexSessionization.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/sql/st
reaming/StructuredNetworkWordCount.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/sql/st
reaming/StructuredNetworkWordCountWindowed.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/sql/st
reaming/StructuredKafkaWordCount.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/sql/st
reaming/StructuredSessionization.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/sql/Us
erDefinedTypedAggregation.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/sql/Us
erDefinedScalar.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/sql/jdbc/
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/sql/jd
bc/ExampleJdbcConnectionProvider.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/Simple
SkewedGroupByTest.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/extensions/
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/extens
```

```
ions/SparkSessionExtensionsTest.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/extens
ions/SessionExtensionsWithoutLoader.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/extens
ions/SessionExtensionsWithLoader.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/extens
ions/AgeExample.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/LocalFileLR.scala
spark-3.3.0-bin-hadoop3/examples/src/main/scala/org/apache/spark/examples/MiniRe
adWriteTest.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/LocalPi.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/DFSReadWriteTest.scala
spark-3.3.0-bin-
hadoop3/examples/src/main/scala/org/apache/spark/examples/SparkALS.scala
spark-3.3.0-bin-hadoop3/examples/src/main/resources/
spark-3.3.0-bin-hadoop3/examples/src/main/resources/people.json
spark-3.3.0-bin-hadoop3/examples/src/main/resources/users.avro
spark-3.3.0-bin-hadoop3/examples/src/main/resources/people.csv
spark-3.3.0-bin-hadoop3/examples/src/main/resources/users.parquet
spark-3.3.0-bin-hadoop3/examples/src/main/resources/META-INF/
spark-3.3.0-bin-hadoop3/examples/src/main/resources/META-INF/services/
spark-3.3.0-bin-hadoop3/examples/src/main/resources/META-
INF/services/org.apache.spark.sql.SparkSessionExtensionsProvider
spark-3.3.0-bin-hadoop3/examples/src/main/resources/META-
INF/services/org.apache.spark.sql.jdbc.JdbcConnectionProvider
spark-3.3.0-bin-hadoop3/examples/src/main/resources/users.orc
spark-3.3.0-bin-hadoop3/examples/src/main/resources/dir1/
spark-3.3.0-bin-hadoop3/examples/src/main/resources/dir1/file1.parquet
spark-3.3.0-bin-hadoop3/examples/src/main/resources/dir1/dir2/
spark-3.3.0-bin-hadoop3/examples/src/main/resources/dir1/dir2/file2.parquet
spark-3.3.0-bin-hadoop3/examples/src/main/resources/dir1/file3.json
spark-3.3.0-bin-hadoop3/examples/src/main/resources/user.avsc
spark-3.3.0-bin-hadoop3/examples/src/main/resources/full_user.avsc
spark-3.3.0-bin-hadoop3/examples/src/main/resources/kv1.txt
spark-3.3.0-bin-hadoop3/examples/src/main/resources/people.txt
spark-3.3.0-bin-hadoop3/examples/src/main/resources/employees.json
spark-3.3.0-bin-hadoop3/examples/src/main/java/
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
StringIndexerExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
```

ChiSqSelectorExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
LabeledDocument.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
LogisticRegressionWithElasticNetExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
PolynomialExpansionExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
TfIdfExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
BucketizerExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
LogisticRegressionSummaryExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
SummarizerExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
QuantileDiscretizerExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
BinarizerExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
OneHotEncoderExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
StandardScalerExample.java
spark-3.3.0-bin-
hadoop3/examples/src/main/java/org/apache/spark/examples/ml/JavaDCTExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
DecisionTreeRegressionExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
RandomForestClassifierExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
OneVsRestExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
GeneralizedLinearRegressionExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
GaussianMixtureExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
Word2VecExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
NormalizerExample.java
spark-3.3.0-bin-
hadoop3/examples/src/main/java/org/apache/spark/examples/ml/JavaLDAExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
CountVectorizerExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
MultilayerPerceptronClassifierExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
MinMaxScalerExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java

CorrelationExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
NGramExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
ImputerExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
MinHashLSHExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
VectorSizeHintExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
FMClassifierExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
FeatureHasherExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
ModelSelectionViaTrainValidationSplitExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
SQLTransformerExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
KMeansExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
VectorAssemblerExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
GradientBoostedTreeRegressorExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
PrefixSpanExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
FMRegressorExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
NaiveBayesExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
MulticlassLogisticRegressionWithElasticNetExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
ModelSelectionViaCrossValidationExample.java
spark-3.3.0-bin-
hadoop3/examples/src/main/java/org/apache/spark/examples/ml/JavaDocument.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
UnivariateFeatureSelectorExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
GradientBoostedTreeClassifierExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
RandomForestRegressorExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
AFTSurvivalRegressionExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
IsotonicRegressionExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
ElementwiseProductExample.java
spark-3.3.0-bin-

hadoop3/examples/src/main/java/org/apache/spark/examples/ml/JavaALSExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
BucketedRandomProjectionLSHExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
VarianceThresholdSelectorExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
VectorSlicerExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
TokenizerExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
RFormulaExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
EstimatorTransformerParamExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
PowerIterationClusteringExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
BisectingKMeansExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
MaxAbsScalerExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
DecisionTreeClassificationExample.java
spark-3.3.0-bin-
hadoop3/examples/src/main/java/org/apache/spark/examples/ml/JavaPCAExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
PipelineExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
StopWordsRemoverExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
IndexToStringExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
FPGrowthExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
VectorIndexerExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
InteractionExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
ChiSquareTestExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
RobustScalerExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
LinearRegressionWithElasticNetExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/ml/Java
LinearSVCExample.java
spark-3.3.0-bin-
hadoop3/examples/src/main/java/org/apache/spark/examples/JavaWordCount.java
spark-3.3.0-bin-
hadoop3/examples/src/main/java/org/apache/spark/examples/JavaHdfsLR.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/JavaSta

tusTrackerDemo.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/J
avaChiSqSelectorExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/J
avaHypothesisTestingKolmogorovSmirnovTestExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/J
avaRandomForestRegressionExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/J
avaSimpleFPGrowth.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/J
avaRandomForestClassificationExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/J
avaMulticlassClassificationMetricsExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/J
avaKernelDensityEstimationExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/J
avaSummaryStatisticsExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/J
avaDecisionTreeRegressionExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/J
avaGradientBoostingClassificationExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/J
avaGaussianMixtureExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/J
avaStratifiedSamplingExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/J
avaSVDExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/J
avaAssociationRulesExample.java
spark-3.3.0-bin-
hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/JavaALS.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/J
avaKMeansExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/J
avaSVMWithSGDExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/J
avaPrefixSpanExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/J
avaNaiveBayesExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/J
avaLogisticRegressionWithLBFGSExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/J
avaLBFGSExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/J
avaGradientBoostingRegressionExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/J
avaRankingMetricsExample.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/JavaIsotonicRegressionExample.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/JavaElementwiseProductExample.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/JavaCorrelationsExample.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/JavaPowerIterationClusteringExample.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/JavaBisectingKMeansExample.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/JavaDecisionTreeClassificationExample.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/JavaPCAExample.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/JavaRecommendationExample.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/JavaLatentDirichletAllocationExample.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/JavaHypothesisTestingExample.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/JavaMultiLabelClassificationMetricsExample.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/JavaBinaryClassificationMetricsExample.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/mllib/JavaStreamingTestExample.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/JavaPageRank.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/JavaLogQuery.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/streaming/

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/streaming/JavaStatefulNetworkWordCount.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/streaming/JavaCustomReceiver.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/streaming/JavaRecord.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/streaming/JavaNetworkWordCount.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/streaming/JavaDirectKerberizedKafkaWordCount.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/streaming/JavaDirectKafkaWordCount.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/streaming/JavaSqlNetworkWordCount.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/streaming/JavaRecoverableNetworkWordCount.java

spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/streaming/JavaQueueStream.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/JavaSparkPi.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/sql/
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/sql/JavaUserDefinedScalar.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/sql/hive/
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/sql/hive/JavaSparkHiveExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/sql/JavaUserDefinedTypedAggregation.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/sql/JavaSQLDataSourceExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/sql/JavaSparkSQLExample.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/sql/JavaUserDefinedUntypedAggregation.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/sql/streaming/
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/sql/streaming/JavaStructuredKafkaWordCount.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/sql/streaming/JavaStructuredComplexSessionization.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/sql/streaming/JavaStructuredNetworkWordCountWindowed.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/sql/streaming/JavaStructuredSessionization.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/sql/streaming/JavaStructuredNetworkWordCount.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/sql/streaming/JavaStructuredKerberizedKafkaWordCount.java
spark-3.3.0-bin-hadoop3/examples/src/main/java/org/apache/spark/examples/JavaTC.java
spark-3.3.0-bin-hadoop3/examples/src/main/scripts/
spark-3.3.0-bin-hadoop3/examples/src/main/scripts/getGpusResources.sh
spark-3.3.0-bin-hadoop3/examples/src/main/python/
spark-3.3.0-bin-hadoop3/examples/src/main/python/kmeans.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/dct_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/count_vectorizer_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/chisq_selector_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/gradient_boosted_tree_regressor_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/tf_idf_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/cross_validator.py
spark-3.3.0-bin-

hadoop3/examples/src/main/python/ml/linear_regression_with_elastic_net.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/normalizer_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/fm_regressor_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/ml/univariate_feature_selector_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/ml/polynomial_expansion_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/pipeline_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/ml/generalized_linear_regression_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/ml/multilayer_perceptron_classification.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/bisecting_k_means_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/chi_square_test_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/train_validation_split.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/stopwords_remover_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/linearsvc.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/lda_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/ml/random_forest_regressor_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/vector_assembler_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/word2vec_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/string_indexer_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/ml/decision_tree_classification_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/index_to_string_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/bucketizer_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/vector_size_hint_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/vector_indexer_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/tokenizer_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/robust_scaler_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/standard_scaler_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/kmeans_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/ml/power_iteration_clustering_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/naive_bayes_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/pca_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/aft_survival_regression.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/min_max_scaler_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/ml/elementwise_product_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/dataframe_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/n_gram_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/rformula_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/one_vs_rest_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/als_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/onehot_encoder_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/vector_slicer_example.py

```
spark-3.3.0-bin-
hadoop3/examples/src/main/python/ml/logistic_regression_with_elastic_net.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/sql_transformer.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/multiclass_logistic_regressi
on_with_elastic_net.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/summarizer_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/ml/quantile_discretizer_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/feature_hasher_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/imputer_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/ml/bucketed_random_projection_lsh_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/ml/estimator_transformer_param_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/ml/gradient_boosted_tree_classifier_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/ml/logistic_regression_summary_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/ml/decision_tree_regression_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/__init__,py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/fm_classifier_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/gaussian_mixture_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/fpgrowth_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/ml/variance_threshold_selector_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/ml/random_forest_classifier_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/correlation_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/min_hash_lsh_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/binarizer_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/prefixspan_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/max_abs_scaler_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/ml/isotonic_regression_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/ml/interaction_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/logistic_regression.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/__init__.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/als.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/mllib/
spark-3.3.0-bin-hadoop3/examples/src/main/python/mllib/word2vec.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/mllib/correlations_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/mllib/gradient_boosting_classif
ication_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/kernel_density_estimation_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/hypothesis_testing_example.py
```

```
spark-3.3.0-bin-hadoop3/examples/src/main/python/mllib/svd_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/mllib/kmeans.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/latent_dirichlet_allocation_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/mllib/tf_idf_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/mllib/correlations.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/random_forest_regression_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/mllib/normalizer_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/mllib/logistic_regression.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/multi_label_metrics_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/gradient_boosting_regression_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/random_forest_classification_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/mllib/__init__.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/streaming_linear_regression_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/mllib/sampled_rdds.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/bisecting_k_means_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/logistic_regression_with_lbfgs_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/mllib/k_means_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/streaming_k_means_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/mllib/word2vec_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/multi_class_metrics_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/decision_tree_classification_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/summary_statistics_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/binary_classification_metrics_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/standard_scaler_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/power_iteration_clustering_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/mllib/gaussian_mixture_model.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/regression_metrics_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/mllib/naive_bayes_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/elementwise_product_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/linear_regression_with_sgd_example.py
```

```
spark-3.3.0-bin-hadoop3/examples/src/main/python/mllib/pca_rowmatrix_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/mllib/hypothesis_testing_kolmog
orov_smirnov_test_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/mllib/random_rdd_generation.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/decision_tree_regression_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/gaussian_mixture_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/mllib/recommendation_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/mllib/fpgrowth_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/ranking_metrics_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/mllib/svm_with_sgd_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/isotonic_regression_example.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/mllib/stratified_sampling_example.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/wordcount.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/status_api_demo.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/pagerank.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/sort.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/transitive_closure.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/avro_inputformat.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/pi.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/streaming/
spark-3.3.0-bin-hadoop3/examples/src/main/python/streaming/queue_stream.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/streaming/__init__.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/streaming/stateful_network_wordcount.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/streaming/network_wordjoinsentiments.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/streaming/sql_network_wordcount.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/streaming/network_wordcount.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/streaming/hdfs_wordcount.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/streaming/recoverable_network_wordcount.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/sql/
spark-3.3.0-bin-hadoop3/examples/src/main/python/sql/datasource.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/sql/__init__.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/sql/hive.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/sql/arrow.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/sql/streaming/
spark-3.3.0-bin-hadoop3/examples/src/main/python/sql/streaming/structured_networ
k_wordcount_windowed.py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/sql/streaming/structured_network_wordcount.py
spark-3.3.0-bin-
```

```
hadoop3/examples/src/main/python/sql/streaming/structured_sessionization.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/sql/streaming/__init__,py
spark-3.3.0-bin-
hadoop3/examples/src/main/python/sql/streaming/structured_kafka_wordcount.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/sql/basic.py
spark-3.3.0-bin-hadoop3/examples/src/main/python/parquet_inputformat.py
spark-3.3.0-bin-hadoop3/examples/jars/
spark-3.3.0-bin-hadoop3/examples/jars/spark-examples_2.12-3.3.0.jar
spark-3.3.0-bin-hadoop3/examples/jars/scopt_2.12-3.7.1.jar
spark-3.3.0-bin-hadoop3/conf/
spark-3.3.0-bin-hadoop3/conf/metrics.properties.template
spark-3.3.0-bin-hadoop3/conf/workers.template
spark-3.3.0-bin-hadoop3/conf/fairscheduler.xml.template
spark-3.3.0-bin-hadoop3/conf/spark-defaults.conf.template
spark-3.3.0-bin-hadoop3/conf/spark-env.sh.template
spark-3.3.0-bin-hadoop3/conf/log4j2.properties.template
spark-3.3.0-bin-hadoop3/bin/
spark-3.3.0-bin-hadoop3/bin/sparkR.cmd
spark-3.3.0-bin-hadoop3/bin/sparkR
spark-3.3.0-bin-hadoop3/bin/spark-submit
spark-3.3.0-bin-hadoop3/bin/pyspark2.cmd
spark-3.3.0-bin-hadoop3/bin/spark-class
spark-3.3.0-bin-hadoop3/bin/pyspark.cmd
spark-3.3.0-bin-hadoop3/bin/spark-submit2.cmd
spark-3.3.0-bin-hadoop3/bin/load-spark-env.cmd
spark-3.3.0-bin-hadoop3/bin/spark-sql
spark-3.3.0-bin-hadoop3/bin/docker-image-tool.sh
spark-3.3.0-bin-hadoop3/bin/find-spark-home.cmd
spark-3.3.0-bin-hadoop3/bin/load-spark-env.sh
spark-3.3.0-bin-hadoop3/bin/pyspark
spark-3.3.0-bin-hadoop3/bin/spark-shell.cmd
spark-3.3.0-bin-hadoop3/bin/spark-shell2.cmd
spark-3.3.0-bin-hadoop3/bin/spark-submit.cmd
spark-3.3.0-bin-hadoop3/bin/beeline.cmd
spark-3.3.0-bin-hadoop3/bin/find-spark-home
spark-3.3.0-bin-hadoop3/bin/spark-class.cmd
spark-3.3.0-bin-hadoop3/bin/sparkR2.cmd
spark-3.3.0-bin-hadoop3/bin/beeline
spark-3.3.0-bin-hadoop3/bin/spark-class2.cmd
spark-3.3.0-bin-hadoop3/bin/spark-sql.cmd
spark-3.3.0-bin-hadoop3/bin/run-example
spark-3.3.0-bin-hadoop3/bin/spark-shell
spark-3.3.0-bin-hadoop3/bin/run-example.cmd
spark-3.3.0-bin-hadoop3/bin/spark-sql2.cmd
spark-3.3.0-bin-hadoop3/python/
spark-3.3.0-bin-hadoop3/python/.gitignore
spark-3.3.0-bin-hadoop3/python/run-tests-with-coverage
spark-3.3.0-bin-hadoop3/python/mypy.ini
```

```
spark-3.3.0-bin-hadoop3/python/MANIFEST.in
spark-3.3.0-bin-hadoop3/python/README.md
spark-3.3.0-bin-hadoop3/python/test_coverage/
spark-3.3.0-bin-hadoop3/python/test_coverage/coverage_daemon.py
spark-3.3.0-bin-hadoop3/python/test_coverage/conf/
spark-3.3.0-bin-hadoop3/python/test_coverage/conf/spark-defaults.conf
spark-3.3.0-bin-hadoop3/python/test_coverage/sitecustomize.py
spark-3.3.0-bin-hadoop3/python/run-tests.py
spark-3.3.0-bin-hadoop3/python/setup.py
spark-3.3.0-bin-hadoop3/python/test_support/
spark-3.3.0-bin-hadoop3/python/test_support/userlibrary.py
spark-3.3.0-bin-hadoop3/python/test_support/hello/
spark-3.3.0-bin-hadoop3/python/test_support/hello/sub_hello/
spark-3.3.0-bin-hadoop3/python/test_support/hello/sub_hello/sub_hello.txt
spark-3.3.0-bin-hadoop3/python/test_support/hello/hello.txt
spark-3.3.0-bin-hadoop3/python/test_support/userlib-0.1.zip
spark-3.3.0-bin-hadoop3/python/test_support/SimpleHTTPServer.py
spark-3.3.0-bin-hadoop3/python/test_support/sql/
spark-3.3.0-bin-hadoop3/python/test_support/sql/people.json
spark-3.3.0-bin-hadoop3/python/test_support/sql/people_array.json
spark-3.3.0-bin-hadoop3/python/test_support/sql/people_array_utf16le.json
spark-3.3.0-bin-hadoop3/python/test_support/sql/text-test.txt
spark-3.3.0-bin-hadoop3/python/test_support/sql/ages.csv
spark-3.3.0-bin-hadoop3/python/test_support/sql/parquet_partitioned/
spark-3.3.0-bin-hadoop3/python/test_support/sql/parquet_partitioned/year=2015/
spark-3.3.0-bin-
hadoop3/python/test_support/sql/parquet_partitioned/year=2015/month=10/
spark-3.3.0-bin-
hadoop3/python/test_support/sql/parquet_partitioned/year=2015/month=10/day=26/
spark-3.3.0-bin-hadoop3/python/test_support/sql/parquet_partitioned/year=2015/mo
nth=10/day=26/.part-r-00005.gz.parquet.crc
spark-3.3.0-bin-hadoop3/python/test_support/sql/parquet_partitioned/year=2015/mo
nth=10/day=26/part-r-00005.gz.parquet
spark-3.3.0-bin-
hadoop3/python/test_support/sql/parquet_partitioned/year=2015/month=10/day=25/
spark-3.3.0-bin-hadoop3/python/test_support/sql/parquet_partitioned/year=2015/mo
nth=10/day=25/.part-r-00002.gz.parquet.crc
spark-3.3.0-bin-hadoop3/python/test_support/sql/parquet_partitioned/year=2015/mo
nth=10/day=25/.part-r-00004.gz.parquet.crc
spark-3.3.0-bin-hadoop3/python/test_support/sql/parquet_partitioned/year=2015/mo
nth=10/day=25/part-r-00004.gz.parquet
spark-3.3.0-bin-hadoop3/python/test_support/sql/parquet_partitioned/year=2015/mo
nth=10/day=25/part-r-00002.gz.parquet
spark-3.3.0-bin-
hadoop3/python/test_support/sql/parquet_partitioned/year=2015/month=9/
spark-3.3.0-bin-
hadoop3/python/test_support/sql/parquet_partitioned/year=2015/month=9/day=1/
spark-3.3.0-bin-hadoop3/python/test_support/sql/parquet_partitioned/year=2015/mo
```

```
nth=9/day=1/.part-r-00007.gz.parquet.crc
spark-3.3.0-bin-hadoop3/python/test_support/sql/parquet_partitioned/year=2015/mo
nth=9/day=1/part-r-00007.gz.parquet
spark-3.3.0-bin-hadoop3/python/test_support/sql/parquet_partitioned/_metadata
spark-3.3.0-bin-hadoop3/python/test_support/sql/parquet_partitioned/year=2014/
spark-3.3.0-bin-
hadoop3/python/test_support/sql/parquet_partitioned/year=2014/month=9/
spark-3.3.0-bin-
hadoop3/python/test_support/sql/parquet_partitioned/year=2014/month=9/day=1/
spark-3.3.0-bin-hadoop3/python/test_support/sql/parquet_partitioned/year=2014/mo
nth=9/day=1/.part-r-00008.gz.parquet.crc
spark-3.3.0-bin-hadoop3/python/test_support/sql/parquet_partitioned/year=2014/mo
nth=9/day=1/part-r-00008.gz.parquet
spark-3.3.0-bin-
hadoop3/python/test_support/sql/parquet_partitioned/_common_metadata
spark-3.3.0-bin-hadoop3/python/test_support/sql/parquet_partitioned/_SUCCESS
spark-3.3.0-bin-hadoop3/python/test_support/sql/orc_partitioned/
spark-3.3.0-bin-hadoop3/python/test_support/sql/orc_partitioned/b=1/
spark-3.3.0-bin-hadoop3/python/test_support/sql/orc_partitioned/b=1/c=1/
spark-3.3.0-bin-hadoop3/python/test_support/sql/orc_partitioned/b=1/c=1/part-r-0
0000-829af031-b970-49d6-ad39-30460a0be2c8.orc
spark-3.3.0-bin-hadoop3/python/test_support/sql/orc_partitioned/b=1/c=1/.part-r-
00000-829af031-b970-49d6-ad39-30460a0be2c8.orc.crc
spark-3.3.0-bin-hadoop3/python/test_support/sql/orc_partitioned/b=0/
spark-3.3.0-bin-hadoop3/python/test_support/sql/orc_partitioned/b=0/c=0/
spark-3.3.0-bin-hadoop3/python/test_support/sql/orc_partitioned/b=0/c=0/part-r-0
0000-829af031-b970-49d6-ad39-30460a0be2c8.orc
spark-3.3.0-bin-hadoop3/python/test_support/sql/orc_partitioned/b=0/c=0/.part-r-
00000-829af031-b970-49d6-ad39-30460a0be2c8.orc.crc
spark-3.3.0-bin-hadoop3/python/test_support/sql/orc_partitioned/_SUCCESS
spark-3.3.0-bin-hadoop3/python/test_support/sql/ages_newlines.csv
spark-3.3.0-bin-hadoop3/python/test_support/sql/streaming/
spark-3.3.0-bin-hadoop3/python/test_support/sql/streaming/text-test.txt
spark-3.3.0-bin-hadoop3/python/test_support/sql/people1.json
spark-3.3.0-bin-hadoop3/python/pyspark/
spark-3.3.0-bin-hadoop3/python/pyspark/tests/
spark-3.3.0-bin-hadoop3/python/pyspark/tests/test_rddbarrier.py
spark-3.3.0-bin-hadoop3/python/pyspark/tests/test_worker.py
spark-3.3.0-bin-hadoop3/python/pyspark/tests/test_serializers.py
spark-3.3.0-bin-hadoop3/python/pyspark/tests/test_util.py
spark-3.3.0-bin-hadoop3/python/pyspark/tests/test_rdd.py
spark-3.3.0-bin-hadoop3/python/pyspark/tests/test_statcounter.py
spark-3.3.0-bin-hadoop3/python/pyspark/tests/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/tests/test_broadcast.py
spark-3.3.0-bin-hadoop3/python/pyspark/tests/test_appsubmit.py
spark-3.3.0-bin-hadoop3/python/pyspark/tests/test_profiler.py
spark-3.3.0-bin-hadoop3/python/pyspark/tests/test_pin_thread.py
spark-3.3.0-bin-hadoop3/python/pyspark/tests/test_install_spark.py
```

```
spark-3.3.0-bin-hadoop3/python/pyspark/tests/typing/
spark-3.3.0-bin-hadoop3/python/pyspark/tests/typing/test_resultiterable.yml
spark-3.3.0-bin-hadoop3/python/pyspark/tests/typing/test_context.yml
spark-3.3.0-bin-hadoop3/python/pyspark/tests/typing/test_rdd.yml
spark-3.3.0-bin-hadoop3/python/pyspark/tests/typing/test_core.yml
spark-3.3.0-bin-hadoop3/python/pyspark/tests/test_shuffle.py
spark-3.3.0-bin-hadoop3/python/pyspark/tests/test_join.py
spark-3.3.0-bin-hadoop3/python/pyspark/tests/test_taskcontext.py
spark-3.3.0-bin-hadoop3/python/pyspark/tests/test_context.py
spark-3.3.0-bin-hadoop3/python/pyspark/tests/test_readwrite.py
spark-3.3.0-bin-hadoop3/python/pyspark/tests/test_conf.py
spark-3.3.0-bin-hadoop3/python/pyspark/tests/test_daemon.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_series.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_utils.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_dataframe.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/indexes/
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/indexes/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/indexes/test_timedelta.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/indexes/test_category.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/indexes/test_datetime.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/indexes/test_base.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_window.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_indexops_spark.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_stats.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_extension.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_default_index.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_expanding.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_series_datetime.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_indexing.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_typedef.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_numpy_compat.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_dataframe_spark_io.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_repr.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_ops_on_diff_frames_groupby_expanding.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_namespace.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_ops_on_diff_frames_groupby.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_series_conversion.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_internal.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_series_string.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_groupby.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_csv.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_ops_on_diff_frames_groupby_rolling.py
```

```
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_dataframe_conversion.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/data_type_ops/
spark-3.3.0-bin-
hadoop3/python/pyspark/pandas/tests/data_type_ops/test_categorical_ops.py
spark-3.3.0-bin-
hadoop3/python/pyspark/pandas/tests/data_type_ops/test_complex_ops.py
spark-3.3.0-bin-
hadoop3/python/pyspark/pandas/tests/data_type_ops/test_boolean_ops.py
spark-3.3.0-bin-
hadoop3/python/pyspark/pandas/tests/data_type_ops/test_string_ops.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/data_type_ops/__init__.py
spark-3.3.0-bin-
hadoop3/python/pyspark/pandas/tests/data_type_ops/testing_utils.py
spark-3.3.0-bin-
hadoop3/python/pyspark/pandas/tests/data_type_ops/test_udt_ops.py
spark-3.3.0-bin-
hadoop3/python/pyspark/pandas/tests/data_type_ops/test_num_ops.py
spark-3.3.0-bin-
hadoop3/python/pyspark/pandas/tests/data_type_ops/test_date_ops.py
spark-3.3.0-bin-
hadoop3/python/pyspark/pandas/tests/data_type_ops/test_timedelta_ops.py
spark-3.3.0-bin-
hadoop3/python/pyspark/pandas/tests/data_type_ops/test_null_ops.py
spark-3.3.0-bin-
hadoop3/python/pyspark/pandas/tests/data_type_ops/test_datetime_ops.py
spark-3.3.0-bin-
hadoop3/python/pyspark/pandas/tests/data_type_ops/test_binary_ops.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/data_type_ops/test_base.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_frame_spark.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_config.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_categorical.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/plot/
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/plot/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/plot/test_frame_plot.py
spark-3.3.0-bin-
hadoop3/python/pyspark/pandas/tests/plot/test_frame_plot_matplotlib.py
spark-3.3.0-bin-
hadoop3/python/pyspark/pandas/tests/plot/test_frame_plot_plotly.py
spark-3.3.0-bin-
hadoop3/python/pyspark/pandas/tests/plot/test_series_plot_matplotlib.py
spark-3.3.0-bin-
hadoop3/python/pyspark/pandas/tests/plot/test_series_plot_plotly.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/plot/test_series_plot.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_ops_on_diff_frames.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_rolling.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_reshape.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_sql.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/tests/test_spark_functions.py
```

```
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/sql_processor.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/numpy_compat.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/usage_logging/
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/usage_logging/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/usage_logging/usage_logger.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/spark/
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/spark/functions.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/spark/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/spark/accessors.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/spark/utils.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/base.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/sql_formatter.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/frame.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/window.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/indexes/
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/indexes/base.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/indexes/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/indexes/timedelta.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/indexes/numeric.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/indexes/datetimes.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/indexes/category.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/indexes/multi.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/namespace.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/mlflow.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/indexing.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/strings.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/extensions.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/accessors.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/config.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/groupby.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/internal.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/missing/
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/missing/frame.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/missing/window.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/missing/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/missing/groupby.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/missing/series.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/missing/indexes.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/missing/common.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/ml.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/series.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/_typing.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/datetimes.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/utils.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/data_type_ops/
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/data_type_ops/complex_ops.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/data_type_ops/binary_ops.py
```

```
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/data_type_ops/base.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/data_type_ops/categorical_ops.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/data_type_ops/string_ops.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/data_type_ops/timedelta_ops.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/data_type_ops/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/data_type_ops/date_ops.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/data_type_ops/udt_ops.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/data_type_ops/null_ops.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/data_type_ops/datetime_ops.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/data_type_ops/num_ops.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/data_type_ops/boolean_ops.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/plot/
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/plot/core.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/plot/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/plot/plotly.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/plot/matplotlib.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/typedef/
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/typedef/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/typedef/typehints.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/categorical.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/generic.py
spark-3.3.0-bin-hadoop3/python/pyspark/pandas/exceptions.py
spark-3.3.0-bin-hadoop3/python/pyspark/_typing.pyi
spark-3.3.0-bin-hadoop3/python/pyspark/testing/
spark-3.3.0-bin-hadoop3/python/pyspark/testing/mlutils.py
spark-3.3.0-bin-hadoop3/python/pyspark/testing/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/testing/mllibutils.py
spark-3.3.0-bin-hadoop3/python/pyspark/testing/pandasutils.py
spark-3.3.0-bin-hadoop3/python/pyspark/testing/utils.py
spark-3.3.0-bin-hadoop3/python/pyspark/testing/sqlutils.py
spark-3.3.0-bin-hadoop3/python/pyspark/testing/streamingutils.py
spark-3.3.0-bin-hadoop3/python/pyspark/accumulators.py
spark-3.3.0-bin-hadoop3/python/pyspark/rddsampler.py
spark-3.3.0-bin-hadoop3/python/pyspark/install.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/test_algorithms.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/test_evaluation.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/test_util.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/test_feature.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/test_pipeline.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/test_wrapper.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/test_tuning.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/test_persistence.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/test_param.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/typing/
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/typing/test_clustering.yaml
```

```
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/typing/test_regression.yml
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/typing/test_param.yml
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/typing/test_evaluation.yml
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/typing/test_classification.yml
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/typing/test_readable.yml
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/typing/test_feature.yml
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/test_training_summary.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/test_linalg.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/test_image.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/test_stat.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tests/test_base.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/_typing.pyi
spark-3.3.0-bin-hadoop3/python/pyspark/ml/functions.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tuning.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/pipeline.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/base.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/feature.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/stat.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/image.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/classification.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/recommendation.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/regression.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/param/
spark-3.3.0-bin-hadoop3/python/pyspark/ml/param/_shared_params_code_gen.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/param/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/param/shared.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/tree.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/fpm.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/wrapper.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/clustering.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/common.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/linalg/
spark-3.3.0-bin-hadoop3/python/pyspark/ml/linalg/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/evaluation.py
spark-3.3.0-bin-hadoop3/python/pyspark/ml/util.py
spark-3.3.0-bin-hadoop3/python/pyspark/find_spark_home.py
spark-3.3.0-bin-hadoop3/python/pyspark/serializers.py
spark-3.3.0-bin-hadoop3/python/pyspark/java_gateway.py
spark-3.3.0-bin-hadoop3/python/pyspark/traceback_utils.py
spark-3.3.0-bin-hadoop3/python/pyspark/resource/
spark-3.3.0-bin-hadoop3/python/pyspark/resource/tests/
spark-3.3.0-bin-hadoop3/python/pyspark/resource/tests/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/resource/tests/test_resources.py
spark-3.3.0-bin-hadoop3/python/pyspark/resource/profile.py
spark-3.3.0-bin-hadoop3/python/pyspark/resource/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/resource/information.py
spark-3.3.0-bin-hadoop3/python/pyspark/resource/requests.py
```

```
spark-3.3.0-bin-hadoop3/python/pyspark/conf.py
spark-3.3.0-bin-hadoop3/python/pyspark/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/__pycache__/
spark-3.3.0-bin-hadoop3/python/pyspark/__pycache__/install.cpython-38.pyc
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/tests/
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/tests/test_algorithms.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/tests/test_streaming_algorithms.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/tests/test_util.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/tests/test_feature.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/tests/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/tests/test_linalg.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/tests/test_stat.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/_typing.pyi
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/recommendation.pyi
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/feature.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/classification.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/random.pyi
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/recommendation.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/regression.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/tree.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/fpm.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/random.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/stat/
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/stat/distribution.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/stat/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/stat/KernelDensity.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/stat/test.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/stat/_statistics.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/clustering.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/common.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/linalg/
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/linalg/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/linalg/distributed.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/evaluation.py
spark-3.3.0-bin-hadoop3/python/pyspark/mllib/util.py
spark-3.3.0-bin-hadoop3/python/pyspark/resultiterable.py
spark-3.3.0-bin-hadoop3/python/pyspark/profiler.py
spark-3.3.0-bin-hadoop3/python/pyspark/statcounter.py
spark-3.3.0-bin-hadoop3/python/pyspark/join.py
spark-3.3.0-bin-hadoop3/python/pyspark/daemon.py
spark-3.3.0-bin-hadoop3/python/pyspark/rdd.py
spark-3.3.0-bin-hadoop3/python/pyspark/context.py
spark-3.3.0-bin-hadoop3/python/pyspark/instrumentation_utils.py
spark-3.3.0-bin-hadoop3/python/pyspark/cloudpickle/
spark-3.3.0-bin-hadoop3/python/pyspark/cloudpickle/cloudpickle_fast.py
spark-3.3.0-bin-hadoop3/python/pyspark/cloudpickle/__init__.py
```

```
spark-3.3.0-bin-hadoop3/python/pyspark/cloudpickle/compat.py
spark-3.3.0-bin-hadoop3/python/pyspark/cloudpickle/cloudpickle.py
spark-3.3.0-bin-hadoop3/python/pyspark/version.py
spark-3.3.0-bin-hadoop3/python/pyspark/py.typed
spark-3.3.0-bin-hadoop3/python/pyspark/files.py
spark-3.3.0-bin-hadoop3/python/pyspark/worker.py
spark-3.3.0-bin-hadoop3/python/pyspark/shell.py
spark-3.3.0-bin-hadoop3/python/pyspark/streaming/
spark-3.3.0-bin-hadoop3/python/pyspark/streaming/tests/
spark-3.3.0-bin-hadoop3/python/pyspark/streaming/tests/test_listener.py
spark-3.3.0-bin-hadoop3/python/pyspark/streaming/tests/test_kinesis.py
spark-3.3.0-bin-hadoop3/python/pyspark/streaming/tests/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/streaming/tests/test_dstream.py
spark-3.3.0-bin-hadoop3/python/pyspark/streaming/tests/test_context.py
spark-3.3.0-bin-hadoop3/python/pyspark/streaming/dstream.py
spark-3.3.0-bin-hadoop3/python/pyspark/streaming/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/streaming/kinesis.py
spark-3.3.0-bin-hadoop3/python/pyspark/streaming/listener.py
spark-3.3.0-bin-hadoop3/python/pyspark/streaming/context.py
spark-3.3.0-bin-hadoop3/python/pyspark/streaming/util.py
spark-3.3.0-bin-hadoop3/python/pyspark/status.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_functions.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_readwriter.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_utils.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_pandas_grouped_map.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_dataframe.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_pandas_map.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_udf.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_streaming.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_serde.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_pandas_udf_window.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_group.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_pandas_udf.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_pandas_cogrouped_map.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_pandas_udf_grouped_agg.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_pandas_udf_scalar.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/typing/
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/typing/test_column.yml
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/typing/test_session.yml
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/typing/test_dataframe.yml
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/typing/test_functions.yml
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/typing/test_readwriter.yml
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/typing/test_udf.yml
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_udf_profiler.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_pandas_udf_typehints_with_
```

```
future_annotations.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_catalog.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_datasources.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_pandas_udf_typehints.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_arrow_map.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_types.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_column.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_context.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_conf.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_arrow.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/tests/test_session.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/pandas/
spark-3.3.0-bin-hadoop3/python/pyspark/sql/pandas/functions.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/pandas/serializers.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/pandas/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/pandas/typehints.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/pandas/map_ops.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/pandas/types.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/pandas/group_ops.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/pandas/functions.pyi
spark-3.3.0-bin-hadoop3/python/pyspark/sql/pandas/utils.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/pandas/conversion.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/pandas/_typing/
spark-3.3.0-bin-hadoop3/python/pyspark/sql/pandas/_typing/__init__.pyi
spark-3.3.0-bin-hadoop3/python/pyspark/sql/pandas/_typing/protocols/
spark-3.3.0-bin-hadoop3/python/pyspark/sql/pandas/_typing/protocols/__init__.pyi
spark-3.3.0-bin-hadoop3/python/pyspark/sql/pandas/_typing/protocols/series.pyi
spark-3.3.0-bin-hadoop3/python/pyspark/sql/pandas/_typing/protocols/frame.pyi
spark-3.3.0-bin-hadoop3/python/pyspark/sql/_typing.pyi
spark-3.3.0-bin-hadoop3/python/pyspark/sql/functions.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/readwriter.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/catalog.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/sql_formatter.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/window.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/udf.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/conf.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/session.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/column.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/group.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/context.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/types.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/observation.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/dataframe.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/avro/
spark-3.3.0-bin-hadoop3/python/pyspark/sql/avro/functions.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/avro/__init__.py
spark-3.3.0-bin-hadoop3/python/pyspark/sql/utils.py
```

```
spark-3.3.0-bin-hadoop3/python/pyspark/sql/streaming.py
spark-3.3.0-bin-hadoop3/python/pyspark/python/
spark-3.3.0-bin-hadoop3/python/pyspark/python/pyspark/
spark-3.3.0-bin-hadoop3/python/pyspark/python/pyspark/shell.py
spark-3.3.0-bin-hadoop3/python/pyspark/shuffle.py
spark-3.3.0-bin-hadoop3/python/pyspark/taskcontext.py
spark-3.3.0-bin-hadoop3/python/pyspark/_globals.py
spark-3.3.0-bin-hadoop3/python/pyspark/broadcast.py
spark-3.3.0-bin-hadoop3/python/pyspark/util.py
spark-3.3.0-bin-hadoop3/python/pyspark/storagelevel.py
spark-3.3.0-bin-hadoop3/python/.coveragerc
spark-3.3.0-bin-hadoop3/python/docs/
spark-3.3.0-bin-hadoop3/python/docs/make2.bat
spark-3.3.0-bin-hadoop3/python/docs/source/
spark-3.3.0-bin-hadoop3/python/docs/source/reference/
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.sql/
spark-3.3.0-bin-
hadoop3/python/docs/source/reference/pyspark.sql/configuration.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.sql/window.rst
spark-3.3.0-bin-
hadoop3/python/docs/source/reference/pyspark.sql/spark_session.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.sql/index.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.sql/grouping.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.sql/catalog.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.sql/column.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.sql/io.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.sql/observation.rst
spark-3.3.0-bin-
hadoop3/python/docs/source/reference/pyspark.sql/core_classes.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.sql/row.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.sql/avro.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.sql/functions.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.sql/dataframe.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.sql/data_types.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/index.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.ml.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.ss/
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.ss/index.rst
spark-3.3.0-bin-
hadoop3/python/docs/source/reference/pyspark.ss/query_management.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.ss/io.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.ss/core_classes.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.pandas/
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.pandas/window.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.pandas/index.rst
spark-3.3.0-bin-
hadoop3/python/docs/source/reference/pyspark.pandas/extensions.rst
```

```
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.pandas/ml.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.pandas/frame.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.pandas/io.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.pandas/series.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.pandas/indexing.rst
spark-3.3.0-bin-
hadoop3/python/docs/source/reference/pyspark.pandas/general_functions.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.pandas/groupby.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.mllib.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.streaming.rst
spark-3.3.0-bin-hadoop3/python/docs/source/reference/pyspark.resource.rst
spark-3.3.0-bin-hadoop3/python/docs/source/index.rst
spark-3.3.0-bin-hadoop3/python/docs/source/getting_started/
spark-3.3.0-bin-hadoop3/python/docs/source/getting_started/index.rst
spark-3.3.0-bin-hadoop3/python/docs/source/getting_started/quickstart_ps.ipynb
spark-3.3.0-bin-hadoop3/python/docs/source/getting_started/quickstart_df.ipynb
spark-3.3.0-bin-hadoop3/python/docs/source/getting_started/install.rst
spark-3.3.0-bin-hadoop3/python/docs/source/conf.py
spark-3.3.0-bin-hadoop3/python/docs/source/_templates/
spark-3.3.0-bin-hadoop3/python/docs/source/_templates/autosummary/
spark-3.3.0-bin-hadoop3/python/docs/source/_templates/autosummary/class.rst
spark-3.3.0-bin-
hadoop3/python/docs/source/_templates/autosummary/class_with_docs.rst
spark-3.3.0-bin-hadoop3/python/docs/source/_static/
spark-3.3.0-bin-hadoop3/python/docs/source/_static/copybutton.js
spark-3.3.0-bin-hadoop3/python/docs/source/_static/css/
spark-3.3.0-bin-hadoop3/python/docs/source/_static/css/pyspark.css
spark-3.3.0-bin-hadoop3/python/docs/source/development/
spark-3.3.0-bin-hadoop3/python/docs/source/development/setting_ide.rst
spark-3.3.0-bin-hadoop3/python/docs/source/development/index.rst
spark-3.3.0-bin-hadoop3/python/docs/source/development/debugging.rst
spark-3.3.0-bin-hadoop3/python/docs/source/development/testing.rst
spark-3.3.0-bin-hadoop3/python/docs/source/development/contributing.rst
spark-3.3.0-bin-hadoop3/python/docs/source/user_guide/
spark-3.3.0-bin-hadoop3/python/docs/source/user_guide/arrow_pandas.rst
spark-3.3.0-bin-hadoop3/python/docs/source/user_guide/index.rst
spark-3.3.0-bin-hadoop3/python/docs/source/user_guide/pandas_on_spark/
spark-3.3.0-bin-hadoop3/python/docs/source/user_guide/pandas_on_spark/types.rst
spark-3.3.0-bin-
hadoop3/python/docs/source/user_guide/pandas_on_spark/best_practices.rst
spark-3.3.0-bin-
hadoop3/python/docs/source/user_guide/pandas_on_spark/options.rst
spark-3.3.0-bin-
hadoop3/python/docs/source/user_guide/pandas_on_spark/pandas_pyspark.rst
spark-3.3.0-bin-hadoop3/python/docs/source/user_guide/pandas_on_spark/index.rst
spark-3.3.0-bin-
hadoop3/python/docs/source/user_guide/pandas_on_spark/typehints.rst
spark-3.3.0-bin-
```

```
hadoop3/python/docs/source/user_guide/pandas_on_spark/transform_apply.rst
spark-3.3.0-bin-
hadoop3/python/docs/source/user_guide/pandas_on_spark/supported_pandas_api.rst
spark-3.3.0-bin-
hadoop3/python/docs/source/user_guide/pandas_on_spark/from_to_dbms.rst
spark-3.3.0-bin-hadoop3/python/docs/source/user_guide/pandas_on_spark/faq.rst
spark-3.3.0-bin-hadoop3/python/docs/source/user_guide/python_packaging.rst
spark-3.3.0-bin-hadoop3/python/docs/source/user_guide/sql/
spark-3.3.0-bin-hadoop3/python/docs/source/user_guide/sql/arrow_pandas.rst
spark-3.3.0-bin-hadoop3/python/docs/source/user_guide/sql/index.rst
spark-3.3.0-bin-hadoop3/python/docs/source/migration_guide/
spark-3.3.0-bin-
hadoop3/python/docs/source/migration_guide/pyspark_2.3.0_to_2.3.1_above.rst
spark-3.3.0-bin-hadoop3/python/docs/source/migration_guide/index.rst
spark-3.3.0-bin-
hadoop3/python/docs/source/migration_guide/pyspark_2.4_to_3.0.rst
spark-3.3.0-bin-
hadoop3/python/docs/source/migration_guide/pyspark_1.0_1.2_to_1.3.rst
spark-3.3.0-bin-
hadoop3/python/docs/source/migration_guide/pyspark_3.2_to_3.3.rst
spark-3.3.0-bin-
hadoop3/python/docs/source/migration_guide/pyspark_2.3_to_2.4.rst
spark-3.3.0-bin-hadoop3/python/docs/source/migration_guide/koalas_to_pyspark.rst
spark-3.3.0-bin-
hadoop3/python/docs/source/migration_guide/pyspark_1.4_to_1.5.rst
spark-3.3.0-bin-
hadoop3/python/docs/source/migration_guide/pyspark_2.2_to_2.3.rst
spark-3.3.0-bin-
hadoop3/python/docs/source/migration_guide/pyspark_3.1_to_3.2.rst
spark-3.3.0-bin-hadoop3/python/docs/make.bat
spark-3.3.0-bin-hadoop3/python/docs/Makefile
spark-3.3.0-bin-hadoop3/python/lib/
spark-3.3.0-bin-hadoop3/python/lib/PY4J_LICENSE.txt
spark-3.3.0-bin-hadoop3/python/lib/py4j-0.10.9.5-src.zip
spark-3.3.0-bin-hadoop3/python/lib/pyspark.zip
spark-3.3.0-bin-hadoop3/python/run-tests
spark-3.3.0-bin-hadoop3/python/setup.cfg
spark-3.3.0-bin-hadoop3/python/dist/
spark-3.3.0-bin-hadoop3/python/pyspark.egg-info/
spark-3.3.0-bin-hadoop3/python/pyspark.egg-info/dependency_links.txt
spark-3.3.0-bin-hadoop3/python/pyspark.egg-info/PKG-INFO
spark-3.3.0-bin-hadoop3/python/pyspark.egg-info/top_level.txt
spark-3.3.0-bin-hadoop3/python/pyspark.egg-info/requires.txt
spark-3.3.0-bin-hadoop3/python/pyspark.egg-info/SOURCES.txt
spark-3.3.0-bin-hadoop3/licenses/
spark-3.3.0-bin-hadoop3/licenses/LICENSE-respond.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-sbt-launch-lib.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-antlr.txt
```

```
spark-3.3.0-bin-hadoop3/licenses/LICENSE-dagre-d3.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-pyrolite.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-sorttable.js.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-janino.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-protobuf.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-jquery.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-scopt.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-netlib.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-d3.min.js.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-graphlib-dot.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-AnchorJS.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-datatables.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-pmml-model.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-paranamer.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-jakarta-ws-rs-api
spark-3.3.0-bin-hadoop3/licenses/LICENSE-dnsjava.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-jakarta.xml.bind-api.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-jakarta-annotation-api
spark-3.3.0-bin-hadoop3/licenses/LICENSE-CC0.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-jodd.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-f2j.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-machinist.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-javolution.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-modernizr.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-spire.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-leveldbjni.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-join.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-zstd-jni.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-slf4j.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-arpack.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-jsp-api.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-JTransforms.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-JLargeArrays.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-bootstrap.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-reflectasm.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-javassist.html
spark-3.3.0-bin-hadoop3/licenses/LICENSE-zstd.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-json-formatter.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-matchMedia-polyfill.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-scala.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-jakarta.activation-api.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-automaton.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-javax-transaction-transaction-api.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-jaxb-runtime.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-minlog.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-mustache.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-xmlenc.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-jline.txt
```

```
spark-3.3.0-bin-hadoop3/licenses/LICENSE-istack-commons-runtime.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-py4j.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-vis-timeline.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-blas.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-re2j.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-kryo.txt
spark-3.3.0-bin-hadoop3/licenses/LICENSE-cloudpickle.txt
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-
wheels/public/simple/
Collecting pyspark
  Downloading pyspark-3.3.0.tar.gz (281.3 MB)
      |                        | 281.3 MB 49 kB/s
Collecting py4j==0.10.9.5
  Downloading py4j-0.10.9.5-py2.py3-none-any.whl (199 kB)
      |                        | 199 kB 68.0 MB/s
Building wheels for collected packages: pyspark
  Building wheel for pyspark (setup.py) … done
  Created wheel for pyspark: filename=pyspark-3.3.0-py2.py3-none-any.whl
size=281764026
sha256=444144b0402497edbd03feee8f92ebe45875fbaac1f30f92d69c3b212275c3e0
  Stored in directory: /root/.cache/pip/wheels/7a/8e/1b/f73a52650d2e5f337708d9f6
a1750d451a7349a867f928b885
Successfully built pyspark
Installing collected packages: py4j, pyspark
Successfully installed py4j-0.10.9.5 pyspark-3.3.0
```

```python
[2]: # set your spark folder to your system path environment.
     import os
     os.environ["JAVA_HOME"] = "/usr/lib/jvm/java-8-openjdk-amd64"
     os.environ["SPARK_HOME"] = "/content/spark-3.3.0-bin-hadoop3"
```

```python
[3]: #Run a local spark session to test your installation:

     import findspark
     findspark.init()

     #to know the location where Spark is installed
     findspark.find()

     from pyspark.sql import SparkSession
     spark = SparkSession.builder.master("local[*]").getOrCreate()
```

```python
[ ]: from google.colab import files
     files.upload()
```

```python
[5]: !unzip '/content/Data.zip'
```

```
Archive:  /content/Data.zip
```

46

```
  creating: Data/
  inflating: Data/location_temp.csv
  inflating: Data/server_name.csv
  inflating: Data/utilization.csv
  inflating: Data/utlization.json
```

[6]:
```python
data_path = '/content/Data/'
```

[7]:
```python
from pyspark.sql import SparkSession
```

[8]:
```python
spark = SparkSession \
    .builder \
    .appName("Spark SQL Query Dataframes") \
    .getOrCreate()
```

[31]:
```python
file_path = data_path + "/utilization.csv"
df1 = spark.read.format("csv").option("header", "false").load(file_path)
```

[32]:
```python
df1.head()
```

[32]:
```
Row(_c0='03/05/2019 08:06:14', _c1='100', _c2='0.57', _c3='0.51', _c4='47')
```

[33]:
```python
df1.columns
```

[33]:
```
['_c0', '_c1', '_c2', '_c3', '_c4']
```

[35]:
```python
df2 = df1.withColumnRenamed("_c0", "event_datetime") \
        .withColumnRenamed ("_c1", "server_id")          \
        .withColumnRenamed("_c2", "cpu_utilization")  \
        .withColumnRenamed("_c3", "free_memory")        \
        .withColumnRenamed("_c4", "session_count")
```

[40]:
```python
json_df2_path = data_path + "/utilization.json"
df2.write.mode('overwrite').json(json_df2_path)
```

[41]:
```python
json_df2_path = data_path + "/utilization.json"
df = spark.read.format("json").load(json_df2_path)
```

[42]:
```python
df.show(5)
```

```
+---------------+-------------------+-----------+---------+-------------+
|cpu_utilization|     event_datetime|free_memory|server_id|session_count|
+---------------+-------------------+-----------+---------+-------------+
|           0.57|03/05/2019 08:06:14|       0.51|      100|           47|
|           0.47|03/05/2019 08:11:14|       0.62|      100|           43|
|           0.56|03/05/2019 08:16:14|       0.57|      100|           62|
|           0.57|03/05/2019 08:21:14|       0.56|      100|           50|
|           0.35|03/05/2019 08:26:14|       0.46|      100|           43|
```

```
+--------------+------------------+----------+---------+-------------+
```
only showing top 5 rows

[43]: `df.printSchema()`

```
root
 |-- cpu_utilization: string (nullable = true)
 |-- event_datetime: string (nullable = true)
 |-- free_memory: string (nullable = true)
 |-- server_id: string (nullable = true)
 |-- session_count: string (nullable = true)
```

[44]: `df.count()`

[44]: 500000

[45]: `df.createOrReplaceTempView("utilization")`

[46]: `df_sql = spark.sql("SELECT * FROM utilization LIMIT 10")`
`df_sql.show()`

```
+--------------+------------------+----------+---------+-------------+
|cpu_utilization|    event_datetime|free_memory|server_id|session_count|
+--------------+------------------+----------+---------+-------------+
|          0.57|03/05/2019 08:06:14|      0.51|      100|           47|
|          0.47|03/05/2019 08:11:14|      0.62|      100|           43|
|          0.56|03/05/2019 08:16:14|      0.57|      100|           62|
|          0.57|03/05/2019 08:21:14|      0.56|      100|           50|
|          0.35|03/05/2019 08:26:14|      0.46|      100|           43|
|          0.41|03/05/2019 08:31:14|      0.58|      100|           48|
|          0.57|03/05/2019 08:36:14|      0.35|      100|           58|
|          0.41|03/05/2019 08:41:14|      0.40|      100|           58|
|          0.53|03/05/2019 08:46:14|      0.35|      100|           62|
|          0.51|03/05/2019 08:51:14|      0.60|      100|           45|
+--------------+------------------+----------+---------+-------------+
```

[47]: `df_sql.count()`

[47]: 10

[48]: `df_sql = spark.sql("SELECT server_id, session_count FROM utilization LIMIT 10")`
`df_sql.show()`

```
+---------+-------------+
|server_id|session_count|
```

```
+---------+------------+
|      100|          47|
|      100|          43|
|      100|          62|
|      100|          50|
|      100|          43|
|      100|          48|
|      100|          58|
|      100|          58|
|      100|          62|
|      100|          45|
+---------+------------+
```

[50]:
```
df_sql = spark.sql("SELECT server_id as sid, session_count as sc FROM␣
    ↪utilization")
df_sql.show(5)
```

```
+---+---+
|sid| sc|
+---+---+
|100| 47|
|100| 43|
|100| 62|
|100| 50|
|100| 43|
+---+---+
only showing top 5 rows
```

[51]:
```
df_sql = spark.sql("SELECT * FROM utilization WHERE server_id = 120")
df_sql.show(10)
```

```
+---------------+-------------------+-----------+---------+-------------+
|cpu_utilization|     event_datetime|free_memory|server_id|session_count|
+---------------+-------------------+-----------+---------+-------------+
|           0.66|03/05/2019 08:06:48|       0.31|      120|           54|
|           0.58|03/05/2019 08:11:48|       0.38|      120|           64|
|           0.55|03/05/2019 08:16:48|       0.61|      120|           54|
|           0.70|03/05/2019 08:21:48|       0.35|      120|           80|
|           0.60|03/05/2019 08:26:48|       0.39|      120|           71|
|           0.53|03/05/2019 08:31:48|       0.35|      120|           49|
|           0.73|03/05/2019 08:36:48|       0.42|      120|           73|
|           0.41|03/05/2019 08:41:48|       0.60|      120|           72|
|           0.62|03/05/2019 08:46:48|       0.57|      120|           57|
|           0.67|03/05/2019 08:51:48|       0.44|      120|           78|
+---------------+-------------------+-----------+---------+-------------+
only showing top 10 rows
```

```
[52]: df_sql.count()
```

```
[52]: 10000
```

```
[53]: df_sql = spark.sql("SELECT server_id, session_count FROM utilization WHERE␣
      ↪session_count > 70")
      df_sql.show(10)
```

```
+---------+-------------+
|server_id|session_count|
+---------+-------------+
|      100|           71|
|      100|           71|
|      100|           71|
|      100|           71|
|      100|           72|
|      100|           72|
|      100|           71|
|      100|           71|
|      100|           71|
|      100|           72|
+---------+-------------+
only showing top 10 rows
```

```
[54]: df_sql.count()
```

```
[54]: 239659
```

```
[55]: df_sql = spark.sql("SELECT server_id, session_count FROM utilization WHERE␣
      ↪session_count > 70 AND server_id = 120")
      df_sql.show(10)
```

```
+---------+-------------+
|server_id|session_count|
+---------+-------------+
|      120|           80|
|      120|           71|
|      120|           73|
|      120|           72|
|      120|           78|
|      120|           73|
|      120|           78|
|      120|           73|
|      120|           74|
|      120|           78|
```

```
+---------+-------------+
```
only showing top 10 rows

[56]: df_sql.count()

[56]: 2733

[57]:
```
df_sql = spark.sql("SELECT server_id, session_count \
                    FROM utilization \
                    WHERE session_count > 70 AND server_id = 120 \
                    ORDER BY session_count DESC")
df_sql.show(10)
```

```
+---------+-------------+
|server_id|session_count|
+---------+-------------+
|      120|           80|
|      120|           80|
|      120|           80|
|      120|           80|
|      120|           80|
|      120|           80|
|      120|           80|
|      120|           80|
|      120|           80|
|      120|           80|
+---------+-------------+
```
only showing top 10 rows

[59]:
```
df_sql = spark.sql("SELECT distinct server_id, session_count \
                    FROM utilization \
                    WHERE session_count > 70 \
                    ORDER BY session_count DESC")
df_sql.show(10)
```

```
+---------+-------------+
|server_id|session_count|
+---------+-------------+
|      103|           99|
|      102|           99|
|      101|           99|
|      108|           99|
|      149|           99|
|      145|           99|
|      148|           99|
|      137|           99|
```

```
|      133|          99|
|      113|          99|
+---------+-------------+
only showing top 10 rows
```

[60]:
```
df_sql = spark.sql("SELECT distinct server_id, session_count \
                    FROM utilization \
                    WHERE session_count > 70 \
                    ORDER BY server_id DESC")
df_sql.show(10)
```

```
+---------+-------------+
|server_id|session_count|
+---------+-------------+
|      149|          95|
|      149|          97|
|      149|          81|
|      149|          86|
|      149|          77|
|      149|          82|
|      149|          85|
|      149|          88|
|      149|          87|
|      149|          90|
+---------+-------------+
only showing top 10 rows
```

[61]:
```
df_sql = spark.sql("SELECT count(*) \
                    FROM utilization \
                    WHERE session_count > 70")
df_sql.show()
```

```
+--------+
|count(1)|
+--------+
|  239659|
+--------+
```

[63]:
```
df_sql = spark.sql("SELECT server_id, count(*) \
                    FROM utilization \
                    WHERE session_count > 70 \
                    GROUP BY server_id")
df_sql.show(10)
```

```
+---------+--------+
```

```
|server_id|count(1)|
+---------+--------+
|      125|    2843|
|      101|    9808|
|      112|    7425|
|      113|    9418|
|      107|    5646|
|      110|    2826|
|      100|     391|
|      126|    6365|
|      120|    2733|
|      118|    7913|
+---------+--------+
only showing top 10 rows
```

```
[66]: df_sql = spark.sql("SELECT server_id, count(*) \
                          FROM utilization \
                          WHERE session_count > 70 \
                          GROUP BY server_id \
                          ORDER BY count(*) DESC")
      df_sql.show(10)
```

```
+---------+--------+
|server_id|count(1)|
+---------+--------+
|      101|    9808|
|      113|    9418|
|      145|    9304|
|      103|    8744|
|      102|    8586|
|      133|    8583|
|      108|    8375|
|      149|    8288|
|      137|    8248|
|      148|    8027|
+---------+--------+
only showing top 10 rows
```

```
[67]: df_sql = spark.sql("SELECT server_id, min(session_count), avg(session_count),␣
      ↪max(session_count) \
                          FROM utilization \
                          WHERE session_count > 70 \
                          GROUP BY server_id \
                          ORDER BY count(*) DESC")
      df_sql.show(10)
```

```
+---------+-----------------+-----------------+-----------------+
|server_id|min(session_count)|avg(session_count)|max(session_count)|
+---------+-----------------+-----------------+-----------------+
|      101|              100| 87.66557911908646|               99|
|      113|              100| 86.96262476109577|               99|
|      145|              100| 86.97732158211522|               99|
|      103|              100| 85.76372369624886|               99|
|      102|              100| 85.71150710458886|               99|
|      133|              100| 85.46720260981009|               99|
|      108|              100|  85.1219104477612|               99|
|      149|               71|  84.9612693050193|               99|
|      137|               71|  85.0061833171678|               99|
|      148|               71| 84.70350068518749|               99|
+---------+-----------------+-----------------+-----------------+
only showing top 10 rows
```

[68]:
```python
df_sql = spark.sql("SELECT server_id, min(session_count),␣
↪round(avg(session_count),2), max(session_count) \
              FROM utilization \
              WHERE session_count > 70 \
              GROUP BY server_id \
              ORDER BY count(*) DESC")
df_sql.show(10)
```

```
+---------+-----------------+-------------------------+-----------------+
|server_id|min(session_count)|round(avg(session_count), 2)|max(session_count)|
+---------+-----------------+-------------------------+-----------------+
|      101|              100|                    87.67|               99|
|      113|              100|                    86.96|               99|
|      145|              100|                    86.98|               99|
|      103|              100|                    85.76|               99|
|      102|              100|                    85.71|               99|
|      133|              100|                    85.47|               99|
|      108|              100|                    85.12|               99|
|      149|               71|                    84.96|               99|
|      137|               71|                    85.01|               99|
|      148|               71|                     84.7|               99|
+---------+-----------------+-------------------------+-----------------+
only showing top 10 rows
```

[69]:
```python
csv_df_path = data_path + "/server_name.csv"
df_server = spark.read.csv(csv_df_path,header=True)
```

[70]:
```python
df_server.show(10)
```

```
+---------+-----------+
```

```
|server_id|server_name|
+---------+-----------+
|      100| 100 Server|
|      101| 101 Server|
|      102| 102 Server|
|      103| 103 Server|
|      104| 104 Server|
|      105| 105 Server|
|      106| 106 Server|
|      107| 107 Server|
|      108| 108 Server|
|      109| 109 Server|
+---------+-----------+
only showing top 10 rows
```

[71]:
```
df_server.createOrReplaceTempView("server_name")
```

[72]:
```
df_count = spark.sql("SELECT DISTINCT server_id FROM utilization ORDER BY␣
 ↪server_id")
df_count.show(10)
```

```
+---------+
|server_id|
+---------+
|      100|
|      101|
|      102|
|      103|
|      104|
|      105|
|      106|
|      107|
|      108|
|      109|
+---------+
only showing top 10 rows
```

[73]:
```
spark.sql("SELECT min(server_id), max(server_id) FROM utilization").show()
```

```
+--------------+--------------+
|min(server_id)|max(server_id)|
+--------------+--------------+
|           100|           149|
+--------------+--------------+
```

```
[76]: spark.sql("SELECT * FROM server_name").show(5)

      +---------+-----------+
      |server_id|server_name|
      +---------+-----------+
      |      100| 100 Server|
      |      101| 101 Server|
      |      102| 102 Server|
      |      103| 103 Server|
      |      104| 104 Server|
      +---------+-----------+
      only showing top 5 rows
```

```
[77]: df_join = spark.sql("SELECT u.server_id, sn.server_name, u.session_count \
                          FROM utilization u \
                          INNER JOIN server_name sn \
                          ON sn.server_id = u.server_id")
      df_join.show(5)
```

```
      +---------+-----------+-------------+
      |server_id|server_name|session_count|
      +---------+-----------+-------------+
      |      100| 100 Server|           47|
      |      100| 100 Server|           43|
      |      100| 100 Server|           62|
      |      100| 100 Server|           50|
      |      100| 100 Server|           43|
      +---------+-----------+-------------+
      only showing top 5 rows
```

```
[78]: from pyspark.sql import Row
```

```
[83]: df_dup = spark.sparkContext.parallelize([Row(server_name='101 Server',␣
      ↪cpu_utilization=85, session_count=80), \
                             Row(server_name='101 Server', cpu_utilization=80,␣
      ↪session_count=90),
                             Row(server_name='102 Server', cpu_utilization=85,␣
      ↪session_count=80),
                             Row(server_name='102 Server', cpu_utilization=85,␣
      ↪session_count=80),
                             Row(server_name='103 Server', cpu_utilization=87,␣
      ↪session_count=82),
                             Row(server_name='103 Server', cpu_utilization=87,␣
      ↪session_count=82),
```

```
                          Row(server_name='104 Server', cpu_utilization=88,
    ↪session_count=81),
                          Row(server_name='104 Server', cpu_utilization=88,
    ↪session_count=81)]).toDF()
```

[84]: `df_dup.show()`

```
+----------+---------------+-------------+
|server_name|cpu_utilization|session_count|
+----------+---------------+-------------+
| 101 Server|             85|           80|
| 101 Server|             80|           90|
| 102 Server|             85|           80|
| 102 Server|             85|           80|
| 103 Server|             87|           82|
| 103 Server|             87|           82|
| 104 Server|             88|           81|
| 104 Server|             88|           81|
+----------+---------------+-------------+
```

[85]: `df_dup.drop_duplicates().show()`

```
+----------+---------------+-------------+
|server_name|cpu_utilization|session_count|
+----------+---------------+-------------+
| 101 Server|             85|           80|
| 101 Server|             80|           90|
| 102 Server|             85|           80|
| 103 Server|             87|           82|
| 104 Server|             88|           81|
+----------+---------------+-------------+
```

[86]: `df_dup.drop_duplicates(['server_name']).show()`

```
+----------+---------------+-------------+
|server_name|cpu_utilization|session_count|
+----------+---------------+-------------+
| 102 Server|             85|           80|
| 101 Server|             85|           80|
| 104 Server|             88|           81|
| 103 Server|             87|           82|
+----------+---------------+-------------+
```

```
[87]: df = spark.sparkContext.parallelize([Row(server_name='101 Server',␣
      ↪cpu_utilization=85, session_count=80), \
                              Row(server_name='101 Server', cpu_utilization=80,␣
      ↪session_count=90),
                              Row(server_name='102 Server', cpu_utilization=85,␣
      ↪session_count=40),
                              Row(server_name='103 Server', cpu_utilization=70,␣
      ↪session_count=80),
                              Row(server_name='104 Server', cpu_utilization=60,␣
      ↪session_count=80)]).toDF()
```

```
[88]: df.show()
```

```
+-----------+---------------+-------------+
|server_name|cpu_utilization|session_count|
+-----------+---------------+-------------+
| 101 Server|             85|           80|
| 101 Server|             80|           90|
| 102 Server|             85|           40|
| 103 Server|             70|           80|
| 104 Server|             60|           80|
+-----------+---------------+-------------+
```

```
[90]: from pyspark.sql.functions import lit
      from pyspark.sql.types import StringType
```

```
[91]: df_na = df.withColumn('na_col', lit(None).cast(StringType()))
```

```
[92]: df.show()
```

```
+-----------+---------------+-------------+
|server_name|cpu_utilization|session_count|
+-----------+---------------+-------------+
| 101 Server|             85|           80|
| 101 Server|             80|           90|
| 102 Server|             85|           40|
| 103 Server|             70|           80|
| 104 Server|             60|           80|
+-----------+---------------+-------------+
```

```
[93]: df_na = df.withColumn('na_col', lit(None).cast(StringType()))
```

```
[94]: df_na.show()
```

```
+-----------+---------------+-------------+------+
```

```
|server_name|cpu_utilization|session_count|na_col|
+----------+---------------+-------------+------+
| 101 Server|             85|           80|  null|
| 101 Server|             80|           90|  null|
| 102 Server|             85|           40|  null|
| 103 Server|             70|           80|  null|
| 104 Server|             60|           80|  null|
+----------+---------------+-------------+------+
```

[95]: `df_na.fillna('A').show()`

```
+----------+---------------+-------------+------+
|server_name|cpu_utilization|session_count|na_col|
+----------+---------------+-------------+------+
| 101 Server|             85|           80|     A|
| 101 Server|             80|           90|     A|
| 102 Server|             85|           40|     A|
| 103 Server|             70|           80|     A|
| 104 Server|             60|           80|     A|
+----------+---------------+-------------+------+
```

[96]: `df2 = df_na.fillna('A').union(df_na)`

[97]: `df2.show()`

```
+----------+---------------+-------------+------+
|server_name|cpu_utilization|session_count|na_col|
+----------+---------------+-------------+------+
| 101 Server|             85|           80|     A|
| 101 Server|             80|           90|     A|
| 102 Server|             85|           40|     A|
| 103 Server|             70|           80|     A|
| 104 Server|             60|           80|     A|
| 101 Server|             85|           80|  null|
| 101 Server|             80|           90|  null|
| 102 Server|             85|           40|  null|
| 103 Server|             70|           80|  null|
| 104 Server|             60|           80|  null|
+----------+---------------+-------------+------+
```

[98]: `df2.na.drop().show()`

```
+----------+---------------+-------------+------+
|server_name|cpu_utilization|session_count|na_col|
+----------+---------------+-------------+------+
```

```
| 101 Server|              85|           80|    A|
| 101 Server|              80|           90|    A|
| 102 Server|              85|           40|    A|
| 103 Server|              70|           80|    A|
| 104 Server|              60|           80|    A|
+----------+--------------+------------+------+
```

[99]: `df2.createOrReplaceTempView("na_table")`

[100]: `spark.sql("SELECT * FROM na_table").show()`

```
+----------+--------------+------------+------+
|server_name|cpu_utilization|session_count|na_col|
+----------+--------------+------------+------+
| 101 Server|              85|           80|    A|
| 101 Server|              80|           90|    A|
| 102 Server|              85|           40|    A|
| 103 Server|              70|           80|    A|
| 104 Server|              60|           80|    A|
| 101 Server|              85|           80| null|
| 101 Server|              80|           90| null|
| 102 Server|              85|           40| null|
| 103 Server|              70|           80| null|
| 104 Server|              60|           80| null|
+----------+--------------+------------+------+
```

[ ]: