"C:\Program Files\Java\jdk1.8.0\_66\bin\java" "-javaagent:F:\JetBrains\IntelliJ IDEA Community Edition 2017.1.5\lib\idea\_rt.jar=54644:F:\JetBrains\IntelliJ IDEA Community Edition 2017.1.5\bin" -Dfile.encoding=UTF-8 -classpath "C:\Program Files\Java\jdk1.8.0\_66\jre\lib\charsets.jar;C:\Program Files\Java\jdk1.8.0\_66\jre\lib\deploy.jar;C:\Program Files\Java\jdk1.8.0\_66\jre\lib\ext\access-bridge-64.jar;C:\Program Files\Java\jdk1.8.0\_66\jre\lib\ext\cldrdata.jar;C:\Program Files\Java\jdk1.8.0\_66\jre\lib\ext\dnsns.jar;C:\Program Files\Java\jdk1.8.0\_66\jre\lib\ext\jaccess.jar;C:\Program Files\Java\jdk1.8.0\_66\jre\lib\ext\jfxrt.jar;C:\Program Files\Java\jdk1.8.0\_66\jre\lib\ext\localedata.jar;C:\Program Files\Java\jdk1.8.0\_66\jre\lib\ext\nashorn.jar;C:\Program Files\Java\jdk1.8.0\_66\jre\lib\ext\sunec.jar;C:\Program Files\Java\jdk1.8.0\_66\jre\lib\ext\sunjce\_provider.jar;C:\Program Files\Java\jdk1.8.0\_66\jre\lib\ext\sunmscapi.jar;C:\Program Files\Java\jdk1.8.0\_66\jre\lib\ext\sunpkcs11.jar;C:\Program Files\Java\jdk1.8.0\_66\jre\lib\ext\zipfs.jar;C:\Program Files\Java\jdk1.8.0\_66\jre\lib\javaws.jar;C:\Program Files\Java\jdk1.8.0\_66\jre\lib\jce.jar;C:\Program Files\Java\jdk1.8.0\_66\jre\lib\jfr.jar;C:\Program Files\Java\jdk1.8.0\_66\jre\lib\jfxswt.jar;C:\Program Files\Java\jdk1.8.0\_66\jre\lib\jsse.jar;C:\Program Files\Java\jdk1.8.0\_66\jre\lib\management-agent.jar;C:\Program Files\Java\jdk1.8.0\_66\jre\lib\plugin.jar;C:\Program Files\Java\jdk1.8.0\_66\jre\lib\resources.jar;C:\Program Files\Java\jdk1.8.0\_66\jre\lib\rt.jar;G:\IdeaProjects\SparkPractice1\target\scala-2.11\classes;C:\Users\Inspiron\.ivy2\cache\antlr\antlr\jars\antlr-2.7.7.jar;C:\Users\Inspiron\.ivy2\cache\xmlenc\xmlenc\jars\xmlenc-0.52.jar;C:\Users\Inspiron\.ivy2\cache\stax\stax-api\jars\stax-api-1.0.1.jar;C:\Users\Inspiron\.ivy2\cache\oro\oro\jars\oro-2.0.8.jar;C:\Users\Inspiron\.ivy2\cache\org.xerial.snappy\snappy-java\bundles\snappy-java-1.1.2.4.jar;C:\Users\Inspiron\.ivy2\cache\org.tukaani\xz\jars\xz-1.0.jar;C:\Users\Inspiron\.ivy2\cache\org.spire-math\spire\_2.11\jars\spire\_2.11-0.7.4.jar;C:\Users\Inspiron\.ivy2\cache\org.spire-math\spire-macros\_2.11\jars\spire-macros\_2.11-0.7.4.jar;C:\Users\Inspiron\.ivy2\cache\org.spark-project.spark\unused\jars\unused-1.0.0.jar;C:\Users\Inspiron\.ivy2\cache\org.spark-project.hive\hive-metastore\jars\hive-metastore-1.2.1.spark2.jar;C:\Users\Inspiron\.ivy2\cache\org.spark-project.hive\hive-exec\jars\hive-exec-1.2.1.spark2.jar;C:\Users\Inspiron\.ivy2\cache\org.sonatype.sisu.inject\cglib\jars\cglib-2.2.1-v20090111.jar;C:\Users\Inspiron\.ivy2\cache\org.slf4j\slf4j-log4j12\jars\slf4j-log4j12-1.7.16.jar;C:\Users\Inspiron\.ivy2\cache\org.slf4j\slf4j-api\jars\slf4j-api-1.7.16.jar;C:\Users\Inspiron\.ivy2\cache\org.slf4j\jul-to-slf4j\jars\jul-to-slf4j-1.7.16.jar;C:\Users\Inspiron\.ivy2\cache\org.slf4j\jcl-over-slf4j\jars\jcl-over-slf4j-1.7.16.jar;C:\Users\Inspiron\.ivy2\cache\org.scalatest\scalatest\_2.11\bundles\scalatest\_2.11-2.2.6.jar;C:\Users\Inspiron\.ivy2\cache\org.scalanlp\breeze\_2.11\jars\breeze\_2.11-0.11.2.jar;C:\Users\Inspiron\.ivy2\cache\org.scalanlp\breeze-macros\_2.11\jars\breeze-macros\_2.11-0.11.2.jar;C:\Users\Inspiron\.ivy2\cache\org.scala-lang.modules\scala-xml\_2.11\bundles\scala-xml\_2.11-1.0.4.jar;C:\Users\Inspiron\.ivy2\cache\org.scala-lang.modules\scala-parser-combinators\_2.11\bundles\scala-parser-combinators\_2.11-1.0.4.jar;C:\Users\Inspiron\.ivy2\cache\org.scala-lang\scalap\jars\scalap-2.11.8.jar;C:\Users\Inspiron\.ivy2\cache\org.scala-lang\scala-reflect\jars\scala-reflect-2.11.8.jar;C:\Users\Inspiron\.ivy2\cache\org.scala-lang\scala-library\jars\scala-library-2.11.8.jar;C:\Users\Inspiron\.ivy2\cache\org.scala-lang\scala-compiler\jars\scala-compiler-2.11.8.jar;C:\Users\Inspiron\.ivy2\cache\org.roaringbitmap\RoaringBitmap\bundles\RoaringBitmap-0.5.11.jar;C:\Users\Inspiron\.ivy2\cache\org.objenesis\objenesis\jars\objenesis-2.1.jar;C:\Users\Inspiron\.ivy2\cache\org.mortbay.jetty\jetty-util\jars\jetty-util-6.1.26.jar;C:\Users\Inspiron\.ivy2\cache\org.json4s\json4s-jackson\_2.11\jars\json4s-jackson\_2.11-3.2.11.jar;C:\Users\Inspiron\.ivy2\cache\org.json4s\json4s-core\_2.11\jars\json4s-core\_2.11-3.2.11.jar;C:\Users\Inspiron\.ivy2\cache\org.json4s\json4s-ast\_2.11\jars\json4s-ast\_2.11-3.2.11.jar;C:\Users\Inspiron\.ivy2\cache\org.json\json\jars\json-20090211.jar;C:\Users\Inspiron\.ivy2\cache\org.jpmml\pmml-schema\jars\pmml-schema-1.2.15.jar;C:\Users\Inspiron\.ivy2\cache\org.jpmml\pmml-model\jars\pmml-model-1.2.15.jar;C:\Users\Inspiron\.ivy2\cache\org.jodd\jodd-core\jars\jodd-core-3.5.2.jar;C:\Users\Inspiron\.ivy2\cache\org.javassist\javassist\bundles\javassist-3.18.1-GA.jar;C:\Users\Inspiron\.ivy2\cache\org.iq80.snappy\snappy\jars\snappy-0.2.jar;C:\Users\Inspiron\.ivy2\cache\org.glassfish.jersey.media\jersey-media-jaxb\jars\jersey-media-jaxb-2.22.2.jar;C:\Users\Inspiron\.ivy2\cache\org.glassfish.jersey.core\jersey-server\jars\jersey-server-2.22.2.jar;C:\Users\Inspiron\.ivy2\cache\org.glassfish.jersey.core\jersey-common\jars\jersey-common-2.22.2.jar;C:\Users\Inspiron\.ivy2\cache\org.glassfish.jersey.core\jersey-client\jars\jersey-client-2.22.2.jar;C:\Users\Inspiron\.ivy2\cache\org.glassfish.jersey.containers\jersey-container-servlet-core\jars\jersey-container-servlet-core-2.22.2.jar;C:\Users\Inspiron\.ivy2\cache\org.glassfish.jersey.containers\jersey-container-servlet\jars\jersey-container-servlet-2.22.2.jar;C:\Users\Inspiron\.ivy2\cache\org.glassfish.jersey.bundles.repackaged\jersey-guava\bundles\jersey-guava-2.22.2.jar;C:\Users\Inspiron\.ivy2\cache\org.glassfish.hk2.external\javax.inject\jars\javax.inject-2.4.0-b34.jar;C:\Users\Inspiron\.ivy2\cache\org.glassfish.hk2.external\aopalliance-repackaged\jars\aopalliance-repackaged-2.4.0-b34.jar;C:\Users\Inspiron\.ivy2\cache\org.glassfish.hk2\osgi-resource-locator\jars\osgi-resource-locator-1.0.1.jar;C:\Users\Inspiron\.ivy2\cache\org.glassfish.hk2\hk2-utils\jars\hk2-utils-2.4.0-b34.jar;C:\Users\Inspiron\.ivy2\cache\org.glassfish.hk2\hk2-locator\jars\hk2-locator-2.4.0-b34.jar;C:\Users\Inspiron\.ivy2\cache\org.glassfish.hk2\hk2-api\jars\hk2-api-2.4.0-b34.jar;C:\Users\Inspiron\.ivy2\cache\org.fusesource.leveldbjni\leveldbjni-all\bundles\leveldbjni-all-1.8.jar;C:\Users\Inspiron\.ivy2\cache\org.datanucleus\datanucleus-rdbms\jars\datanucleus-rdbms-3.2.9.jar;C:\Users\Inspiron\.ivy2\cache\org.datanucleus\datanucleus-core\jars\datanucleus-core-3.2.10.jar;C:\Users\Inspiron\.ivy2\cache\org.datanucleus\datanucleus-api-jdo\jars\datanucleus-api-jdo-3.2.6.jar;C:\Users\Inspiron\.ivy2\cache\org.codehaus.janino\janino\jars\janino-2.7.8.jar;C:\Users\Inspiron\.ivy2\cache\org.codehaus.janino\commons-compiler\jars\commons-compiler-2.7.8.jar;C:\Users\Inspiron\.ivy2\cache\org.codehaus.jackson\jackson-mapper-asl\jars\jackson-mapper-asl-1.9.13.jar;C:\Users\Inspiron\.ivy2\cache\org.codehaus.jackson\jackson-core-asl\jars\jackson-core-asl-1.9.13.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.zookeeper\zookeeper\jars\zookeeper-3.4.5.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.xmlbeans\xmlbeans\jars\xmlbeans-2.6.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.xbean\xbean-asm5-shaded\bundles\xbean-asm5-shaded-4.4.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.thrift\libthrift\jars\libthrift-0.9.2.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.thrift\libfb303\jars\libfb303-0.9.2.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.spark\spark-unsafe\_2.11\jars\spark-unsafe\_2.11-2.0.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.spark\spark-tags\_2.11\jars\spark-tags\_2.11-2.0.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.spark\spark-streaming\_2.11\jars\spark-streaming\_2.11-2.0.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.spark\spark-sql\_2.11\jars\spark-sql\_2.11-2.0.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.spark\spark-sketch\_2.11\jars\spark-sketch\_2.11-2.0.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.spark\spark-network-shuffle\_2.11\jars\spark-network-shuffle\_2.11-2.0.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.spark\spark-network-common\_2.11\jars\spark-network-common\_2.11-2.0.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.spark\spark-mllib\_2.11\jars\spark-mllib\_2.11-2.0.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.spark\spark-mllib-local\_2.11\jars\spark-mllib-local\_2.11-2.0.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.spark\spark-launcher\_2.11\jars\spark-launcher\_2.11-2.0.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.spark\spark-hive\_2.11\jars\spark-hive\_2.11-2.0.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.spark\spark-graphx\_2.11\jars\spark-graphx\_2.11-2.0.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.spark\spark-core\_2.11\jars\spark-core\_2.11-2.0.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.spark\spark-catalyst\_2.11\jars\spark-catalyst\_2.11-2.0.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.poi\poi-ooxml-schemas\jars\poi-ooxml-schemas-3.14.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.poi\poi-ooxml\jars\poi-ooxml-3.14.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.poi\poi\jars\poi-3.14.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.parquet\parquet-jackson\jars\parquet-jackson-1.7.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.parquet\parquet-hadoop\jars\parquet-hadoop-1.7.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.parquet\parquet-generator\jars\parquet-generator-1.7.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.parquet\parquet-format\jars\parquet-format-2.3.0-incubating.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.parquet\parquet-encoding\jars\parquet-encoding-1.7.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.parquet\parquet-common\jars\parquet-common-1.7.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.parquet\parquet-column\jars\parquet-column-1.7.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.mesos\mesos\jars\mesos-0.21.1-shaded-protobuf.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.ivy\ivy\jars\ivy-2.4.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.httpcomponents\httpcore\jars\httpcore-4.4.4.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.httpcomponents\httpclient\jars\httpclient-4.5.2.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.hadoop\hadoop-yarn-server-common\jars\hadoop-yarn-server-common-2.2.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.hadoop\hadoop-yarn-common\jars\hadoop-yarn-common-2.2.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.hadoop\hadoop-yarn-client\jars\hadoop-yarn-client-2.2.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.hadoop\hadoop-yarn-api\jars\hadoop-yarn-api-2.2.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.hadoop\hadoop-mapreduce-client-shuffle\jars\hadoop-mapreduce-client-shuffle-2.2.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.hadoop\hadoop-mapreduce-client-jobclient\jars\hadoop-mapreduce-client-jobclient-2.2.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.hadoop\hadoop-mapreduce-client-core\jars\hadoop-mapreduce-client-core-2.2.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.hadoop\hadoop-mapreduce-client-common\jars\hadoop-mapreduce-client-common-2.2.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.hadoop\hadoop-mapreduce-client-app\jars\hadoop-mapreduce-client-app-2.2.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.hadoop\hadoop-hdfs\jars\hadoop-hdfs-2.2.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.hadoop\hadoop-common\jars\hadoop-common-2.2.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.hadoop\hadoop-client\jars\hadoop-client-2.2.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.hadoop\hadoop-auth\jars\hadoop-auth-2.2.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.hadoop\hadoop-annotations\jars\hadoop-annotations-2.2.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.derby\derby\jars\derby-10.10.2.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.curator\curator-recipes\bundles\curator-recipes-2.4.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.curator\curator-framework\bundles\curator-framework-2.4.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.curator\curator-client\bundles\curator-client-2.4.0.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.commons\commons-math3\jars\commons-math3-3.4.1.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.commons\commons-math\jars\commons-math-2.1.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.commons\commons-lang3\jars\commons-lang3-3.3.2.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.commons\commons-compress\jars\commons-compress-1.4.1.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.calcite\calcite-linq4j\jars\calcite-linq4j-1.2.0-incubating.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.calcite\calcite-core\jars\calcite-core-1.2.0-incubating.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.calcite\calcite-avatica\jars\calcite-avatica-1.2.0-incubating.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.avro\avro-mapred\jars\avro-mapred-1.7.7-hadoop2.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.avro\avro-ipc\jars\avro-ipc-1.7.7-tests.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.avro\avro-ipc\jars\avro-ipc-1.7.7.jar;C:\Users\Inspiron\.ivy2\cache\org.apache.avro\avro\bundles\avro-1.7.7.jar;C:\Users\Inspiron\.ivy2\cache\org.antlr\stringtemplate\jars\stringtemplate-3.2.1.jar;C:\Users\Inspiron\.ivy2\cache\org.antlr\antlr4-runtime\jars\antlr4-runtime-4.5.3.jar;C:\Users\Inspiron\.ivy2\cache\org.antlr\antlr-runtime\jars\antlr-runtime-3.4.jar;C:\Users\Inspiron\.ivy2\cache\org.antlr\ST4\jars\ST4-4.0.4.jar;C:\Users\Inspiron\.ivy2\cache\net.sourceforge.f2j\arpack\_combined\_all\jars\arpack\_combined\_all-0.1.jar;C:\Users\Inspiron\.ivy2\cache\net.sf.py4j\py4j\jars\py4j-0.10.1.jar;C:\Users\Inspiron\.ivy2\cache\net.sf.opencsv\opencsv\jars\opencsv-2.3.jar;C:\Users\Inspiron\.ivy2\cache\net.razorvine\pyrolite\jars\pyrolite-4.9.jar;C:\Users\Inspiron\.ivy2\cache\net.jpountz.lz4\lz4\jars\lz4-1.3.0.jar;C:\Users\Inspiron\.ivy2\cache\net.java.dev.jets3t\jets3t\jars\jets3t-0.7.1.jar;C:\Users\Inspiron\.ivy2\cache\net.hydromatic\eigenbase-properties\bundles\eigenbase-properties-1.1.5.jar;C:\Users\Inspiron\.ivy2\cache\log4j\log4j\bundles\log4j-1.2.17.jar;C:\Users\Inspiron\.ivy2\cache\log4j\apache-log4j-extras\bundles\apache-log4j-extras-1.2.17.jar;C:\Users\Inspiron\.ivy2\cache\joda-time\joda-time\jars\joda-time-2.9.3.jar;C:\Users\Inspiron\.ivy2\cache\jline\jline\jars\jline-2.12.jar;C:\Users\Inspiron\.ivy2\cache\javolution\javolution\bundles\javolution-5.5.1.jar;C:\Users\Inspiron\.ivy2\cache\javax.ws.rs\javax.ws.rs-api\jars\javax.ws.rs-api-2.0.1.jar;C:\Users\Inspiron\.ivy2\cache\javax.validation\validation-api\jars\validation-api-1.1.0.Final.jar;C:\Users\Inspiron\.ivy2\cache\javax.transaction\jta\jars\jta-1.1.jar;C:\Users\Inspiron\.ivy2\cache\javax.servlet\javax.servlet-api\jars\javax.servlet-api-3.1.0.jar;C:\Users\Inspiron\.ivy2\cache\javax.jdo\jdo-api\jars\jdo-api-3.0.1.jar;C:\Users\Inspiron\.ivy2\cache\javax.inject\javax.inject\jars\javax.inject-1.jar;C:\Users\Inspiron\.ivy2\cache\javax.annotation\javax.annotation-api\jars\javax.annotation-api-1.2.jar;C:\Users\Inspiron\.ivy2\cache\io.netty\netty-all\jars\netty-all-4.0.29.Final.jar;C:\Users\Inspiron\.ivy2\cache\io.netty\netty\bundles\netty-3.8.0.Final.jar;C:\Users\Inspiron\.ivy2\cache\io.dropwizard.metrics\metrics-jvm\bundles\metrics-jvm-3.1.2.jar;C:\Users\Inspiron\.ivy2\cache\io.dropwizard.metrics\metrics-json\bundles\metrics-json-3.1.2.jar;C:\Users\Inspiron\.ivy2\cache\io.dropwizard.metrics\metrics-graphite\bundles\metrics-graphite-3.1.2.jar;C:\Users\Inspiron\.ivy2\cache\io.dropwizard.metrics\metrics-core\bundles\metrics-core-3.1.2.jar;C:\Users\Inspiron\.ivy2\cache\commons-pool\commons-pool\jars\commons-pool-1.5.4.jar;C:\Users\Inspiron\.ivy2\cache\commons-net\commons-net\jars\commons-net-2.2.jar;C:\Users\Inspiron\.ivy2\cache\commons-logging\commons-logging\jars\commons-logging-1.2.jar;C:\Users\Inspiron\.ivy2\cache\commons-lang\commons-lang\jars\commons-lang-2.6.jar;C:\Users\Inspiron\.ivy2\cache\commons-io\commons-io\jars\commons-io-2.4.jar;C:\Users\Inspiron\.ivy2\cache\commons-httpclient\commons-httpclient\jars\commons-httpclient-3.1.jar;C:\Users\Inspiron\.ivy2\cache\commons-digester\commons-digester\jars\commons-digester-1.8.jar;C:\Users\Inspiron\.ivy2\cache\commons-dbcp\commons-dbcp\jars\commons-dbcp-1.4.jar;C:\Users\Inspiron\.ivy2\cache\commons-configuration\commons-configuration\jars\commons-configuration-1.6.jar;C:\Users\Inspiron\.ivy2\cache\commons-collections\commons-collections\jars\commons-collections-3.2.1.jar;C:\Users\Inspiron\.ivy2\cache\commons-codec\commons-codec\jars\commons-codec-1.10.jar;C:\Users\Inspiron\.ivy2\cache\commons-cli\commons-cli\jars\commons-cli-1.2.jar;C:\Users\Inspiron\.ivy2\cache\commons-beanutils\commons-beanutils-core\jars\commons-beanutils-core-1.8.0.jar;C:\Users\Inspiron\.ivy2\cache\commons-beanutils\commons-beanutils\jars\commons-beanutils-1.7.0.jar;C:\Users\Inspiron\.ivy2\cache\com.univocity\univocity-parsers\jars\univocity-parsers-2.1.1.jar;C:\Users\Inspiron\.ivy2\cache\com.twitter\parquet-hadoop-bundle\jars\parquet-hadoop-bundle-1.6.0.jar;C:\Users\Inspiron\.ivy2\cache\com.twitter\chill\_2.11\jars\chill\_2.11-0.8.0.jar;C:\Users\Inspiron\.ivy2\cache\com.twitter\chill-java\jars\chill-java-0.8.0.jar;C:\Users\Inspiron\.ivy2\cache\com.thoughtworks.paranamer\paranamer\jars\paranamer-2.6.jar;C:\Users\Inspiron\.ivy2\cache\com.ning\compress-lzf\bundles\compress-lzf-1.0.3.jar;C:\Users\Inspiron\.ivy2\cache\com.jolbox\bonecp\bundles\bonecp-0.8.0.RELEASE.jar;C:\Users\Inspiron\.ivy2\cache\com.googlecode.javaewah\JavaEWAH\jars\JavaEWAH-0.3.2.jar;C:\Users\Inspiron\.ivy2\cache\com.google.protobuf\protobuf-java\bundles\protobuf-java-2.5.0.jar;C:\Users\Inspiron\.ivy2\cache\com.google.inject\guice\jars\guice-3.0.jar;C:\Users\Inspiron\.ivy2\cache\com.google.guava\guava\bundles\guava-14.0.1.jar;C:\Users\Inspiron\.ivy2\cache\com.google.code.findbugs\jsr305\jars\jsr305-1.3.9.jar;C:\Users\Inspiron\.ivy2\cache\com.github.virtuald\curvesapi\jars\curvesapi-1.03.jar;C:\Users\Inspiron\.ivy2\cache\com.github.rwl\jtransforms\jars\jtransforms-2.4.0.jar;C:\Users\Inspiron\.ivy2\cache\com.github.fommil.netlib\core\jars\core-1.1.2.jar;C:\Users\Inspiron\.ivy2\cache\com.fasterxml.jackson.module\jackson-module-scala\_2.11\bundles\jackson-module-scala\_2.11-2.6.5.jar;C:\Users\Inspiron\.ivy2\cache\com.fasterxml.jackson.module\jackson-module-paranamer\bundles\jackson-module-paranamer-2.6.5.jar;C:\Users\Inspiron\.ivy2\cache\com.fasterxml.jackson.core\jackson-databind\bundles\jackson-databind-2.6.5.jar;C:\Users\Inspiron\.ivy2\cache\com.fasterxml.jackson.core\jackson-core\bundles\jackson-core-2.6.5.jar;C:\Users\Inspiron\.ivy2\cache\com.fasterxml.jackson.core\jackson-annotations\bundles\jackson-annotations-2.6.5.jar;C:\Users\Inspiron\.ivy2\cache\com.esotericsoftware\minlog\bundles\minlog-1.3.0.jar;C:\Users\Inspiron\.ivy2\cache\com.esotericsoftware\kryo-shaded\bundles\kryo-shaded-3.0.3.jar;C:\Users\Inspiron\.ivy2\cache\com.crealytics\spark-excel\_2.10\jars\spark-excel\_2.10-0.8.3.jar;C:\Users\Inspiron\.ivy2\cache\com.clearspring.analytics\stream\jars\stream-2.7.0.jar;C:\Users\Inspiron\.ivy2\cache\aopalliance\aopalliance\jars\aopalliance-1.0.jar" Session17Assignment2New

Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties

17/08/03 16:57:34 INFO SparkContext: Running Spark version 2.0.0

17/08/03 16:57:35 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

17/08/03 16:57:35 INFO SecurityManager: Changing view acls to: Inspiron

17/08/03 16:57:35 INFO SecurityManager: Changing modify acls to: Inspiron

17/08/03 16:57:35 INFO SecurityManager: Changing view acls groups to:

17/08/03 16:57:35 INFO SecurityManager: Changing modify acls groups to:

17/08/03 16:57:35 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(Inspiron); groups with view permissions: Set(); users with modify permissions: Set(Inspiron); groups with modify permissions: Set()

17/08/03 16:57:36 INFO Utils: Successfully started service 'sparkDriver' on port 54665.

17/08/03 16:57:37 INFO SparkEnv: Registering MapOutputTracker

17/08/03 16:57:37 INFO SparkEnv: Registering BlockManagerMaster

17/08/03 16:57:37 INFO DiskBlockManager: Created local directory at C:\Users\Inspiron\AppData\Local\Temp\blockmgr-09f6e5a8-6987-4fa8-8deb-45f1eb59e671

17/08/03 16:57:37 INFO MemoryStore: MemoryStore started with capacity 349.2 MB

17/08/03 16:57:38 INFO SparkEnv: Registering OutputCommitCoordinator

17/08/03 16:57:38 INFO Utils: Successfully started service 'SparkUI' on port 4040.

17/08/03 16:57:38 INFO SparkUI: Bound SparkUI to 0.0.0.0, and started at http://192.168.56.1:4040

17/08/03 16:57:39 INFO Executor: Starting executor ID driver on host localhost

17/08/03 16:57:39 INFO Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 54675.

17/08/03 16:57:39 INFO NettyBlockTransferService: Server created on 192.168.56.1:54675

17/08/03 16:57:39 INFO BlockManagerMaster: Registering BlockManager BlockManagerId(driver, 192.168.56.1, 54675)

17/08/03 16:57:39 INFO BlockManagerMasterEndpoint: Registering block manager 192.168.56.1:54675 with 349.2 MB RAM, BlockManagerId(driver, 192.168.56.1, 54675)

17/08/03 16:57:39 INFO BlockManagerMaster: Registered BlockManager BlockManagerId(driver, 192.168.56.1, 54675)

17/08/03 16:57:40 INFO MemoryStore: Block broadcast\_0 stored as values in memory (estimated size 107.7 KB, free 349.1 MB)

17/08/03 16:57:41 INFO MemoryStore: Block broadcast\_0\_piece0 stored as bytes in memory (estimated size 10.2 KB, free 349.1 MB)

17/08/03 16:57:41 INFO BlockManagerInfo: Added broadcast\_0\_piece0 in memory on 192.168.56.1:54675 (size: 10.2 KB, free: 349.2 MB)

17/08/03 16:57:41 INFO SparkContext: Created broadcast 0 from textFile at Session17Assignment2New.scala:62

17/08/03 16:57:41 ERROR Shell: Failed to locate the winutils binary in the hadoop binary path

java.io.IOException: Could not locate executable null\bin\winutils.exe in the Hadoop binaries.

at org.apache.hadoop.util.Shell.getQualifiedBinPath(Shell.java:278)

at org.apache.hadoop.util.Shell.getWinUtilsPath(Shell.java:300)

at org.apache.hadoop.util.Shell.<clinit>(Shell.java:293)

at org.apache.hadoop.util.StringUtils.<clinit>(StringUtils.java:76)

at org.apache.hadoop.mapred.FileInputFormat.setInputPaths(FileInputFormat.java:362)

at org.apache.spark.SparkContext$$anonfun$hadoopFile$1$$anonfun$29.apply(SparkContext.scala:985)

at org.apache.spark.SparkContext$$anonfun$hadoopFile$1$$anonfun$29.apply(SparkContext.scala:985)

at org.apache.spark.rdd.HadoopRDD$$anonfun$getJobConf$6.apply(HadoopRDD.scala:177)

at org.apache.spark.rdd.HadoopRDD$$anonfun$getJobConf$6.apply(HadoopRDD.scala:177)

at scala.Option.map(Option.scala:146)

at org.apache.spark.rdd.HadoopRDD.getJobConf(HadoopRDD.scala:177)

at org.apache.spark.rdd.HadoopRDD.getPartitions(HadoopRDD.scala:196)

at org.apache.spark.rdd.RDD$$anonfun$partitions$2.apply(RDD.scala:248)

at org.apache.spark.rdd.RDD$$anonfun$partitions$2.apply(RDD.scala:246)

at scala.Option.getOrElse(Option.scala:121)

at org.apache.spark.rdd.RDD.partitions(RDD.scala:246)

at org.apache.spark.rdd.MapPartitionsRDD.getPartitions(MapPartitionsRDD.scala:35)

at org.apache.spark.rdd.RDD$$anonfun$partitions$2.apply(RDD.scala:248)

at org.apache.spark.rdd.RDD$$anonfun$partitions$2.apply(RDD.scala:246)

at scala.Option.getOrElse(Option.scala:121)

at org.apache.spark.rdd.RDD.partitions(RDD.scala:246)

at org.apache.spark.SparkContext.runJob(SparkContext.scala:1911)

at org.apache.spark.rdd.RDD$$anonfun$foreach$1.apply(RDD.scala:875)

at org.apache.spark.rdd.RDD$$anonfun$foreach$1.apply(RDD.scala:873)

at org.apache.spark.rdd.RDDOperationScope$.withScope(RDDOperationScope.scala:151)

at org.apache.spark.rdd.RDDOperationScope$.withScope(RDDOperationScope.scala:112)

at org.apache.spark.rdd.RDD.withScope(RDD.scala:358)

at org.apache.spark.rdd.RDD.foreach(RDD.scala:873)

at Session17Assignment2New$.delayedEndpoint$Session17Assignment2New$1(Session17Assignment2New.scala:64)

at Session17Assignment2New$delayedInit$body.apply(Session17Assignment2New.scala:38)

at scala.Function0$class.apply$mcV$sp(Function0.scala:34)

at scala.runtime.AbstractFunction0.apply$mcV$sp(AbstractFunction0.scala:12)

at scala.App$$anonfun$main$1.apply(App.scala:76)

at scala.App$$anonfun$main$1.apply(App.scala:76)

at scala.collection.immutable.List.foreach(List.scala:381)

at scala.collection.generic.TraversableForwarder$class.foreach(TraversableForwarder.scala:35)

at scala.App$class.main(App.scala:76)

at Session17Assignment2New$.main(Session17Assignment2New.scala:38)

at Session17Assignment2New.main(Session17Assignment2New.scala)

17/08/03 16:57:41 INFO FileInputFormat: Total input paths to process : 1

17/08/03 16:57:41 INFO SparkContext: Starting job: foreach at Session17Assignment2New.scala:64

17/08/03 16:57:42 INFO DAGScheduler: Got job 0 (foreach at Session17Assignment2New.scala:64) with 1 output partitions

17/08/03 16:57:42 INFO DAGScheduler: Final stage: ResultStage 0 (foreach at Session17Assignment2New.scala:64)

17/08/03 16:57:42 INFO DAGScheduler: Parents of final stage: List()

17/08/03 16:57:42 INFO DAGScheduler: Missing parents: List()

17/08/03 16:57:42 INFO DAGScheduler: Submitting ResultStage 0 (file:///G:/ACADGILD/course material/Hadoop/Sessions/Session 17/Assignments/Assignment2/17.2\_Dataset.txt MapPartitionsRDD[1] at textFile at Session17Assignment2New.scala:62), which has no missing parents

17/08/03 16:57:42 INFO MemoryStore: Block broadcast\_1 stored as values in memory (estimated size 3.0 KB, free 349.1 MB)

17/08/03 16:57:42 INFO MemoryStore: Block broadcast\_1\_piece0 stored as bytes in memory (estimated size 1905.0 B, free 349.1 MB)

17/08/03 16:57:42 INFO BlockManagerInfo: Added broadcast\_1\_piece0 in memory on 192.168.56.1:54675 (size: 1905.0 B, free: 349.2 MB)

17/08/03 16:57:42 INFO SparkContext: Created broadcast 1 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:42 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 0 (file:///G:/ACADGILD/course material/Hadoop/Sessions/Session 17/Assignments/Assignment2/17.2\_Dataset.txt MapPartitionsRDD[1] at textFile at Session17Assignment2New.scala:62)

17/08/03 16:57:42 INFO TaskSchedulerImpl: Adding task set 0.0 with 1 tasks

17/08/03 16:57:42 INFO TaskSetManager: Starting task 0.0 in stage 0.0 (TID 0, localhost, partition 0, PROCESS\_LOCAL, 5439 bytes)

17/08/03 16:57:42 INFO Executor: Running task 0.0 in stage 0.0 (TID 0)

17/08/03 16:57:42 INFO HadoopRDD: Input split: file:/G:/ACADGILD/course material/Hadoop/Sessions/Session 17/Assignments/Assignment2/17.2\_Dataset.txt:0+622

17/08/03 16:57:42 INFO deprecation: mapred.tip.id is deprecated. Instead, use mapreduce.task.id

17/08/03 16:57:42 INFO deprecation: mapred.task.id is deprecated. Instead, use mapreduce.task.attempt.id

17/08/03 16:57:42 INFO deprecation: mapred.task.is.map is deprecated. Instead, use mapreduce.task.ismap

17/08/03 16:57:42 INFO deprecation: mapred.task.partition is deprecated. Instead, use mapreduce.task.partition

17/08/03 16:57:42 INFO deprecation: mapred.job.id is deprecated. Instead, use mapreduce.job.id

Mathew,science,grade-3,45,12

Mathew,history,grade-2,55,13

Mark,maths,grade-2,23,13

Mark,science,grade-1,76,13

John,history,grade-1,14,12

John,maths,grade-2,74,13

Lisa,science,grade-1,24,12

Lisa,history,grade-3,86,13

Andrew,maths,grade-1,34,13

Andrew,science,grade-3,26,14

Andrew,history,grade-1,74,12

Mathew,science,grade-2,55,12

Mathew,history,grade-2,87,12

Mark,maths,grade-1,92,13

Mark,science,grade-2,12,12

John,history,grade-1,67,13

John,maths,grade-1,35,11

Lisa,science,grade-2,24,13

Lisa,history,grade-2,98,15

Andrew,maths,grade-1,23,16

Andrew,science,grade-3,44,14

Andrew,history,grade-2,77,11

17/08/03 16:57:42 INFO Executor: Finished task 0.0 in stage 0.0 (TID 0). 916 bytes result sent to driver

17/08/03 16:57:42 INFO TaskSetManager: Finished task 0.0 in stage 0.0 (TID 0) in 266 ms on localhost (1/1)

17/08/03 16:57:42 INFO TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool

17/08/03 16:57:42 INFO DAGScheduler: ResultStage 0 (foreach at Session17Assignment2New.scala:64) finished in 0.328 s

17/08/03 16:57:42 INFO DAGScheduler: Job 0 finished: foreach at Session17Assignment2New.scala:64, took 0.711841 s

<<<<<<<<<<<<---------- TUPLED RDD ------------>>>>>>>>>>>>

17/08/03 16:57:42 INFO SparkContext: Starting job: foreach at Session17Assignment2New.scala:78

17/08/03 16:57:42 INFO DAGScheduler: Got job 1 (foreach at Session17Assignment2New.scala:78) with 1 output partitions

17/08/03 16:57:42 INFO DAGScheduler: Final stage: ResultStage 1 (foreach at Session17Assignment2New.scala:78)

17/08/03 16:57:42 INFO DAGScheduler: Parents of final stage: List()

17/08/03 16:57:42 INFO DAGScheduler: Missing parents: List()

17/08/03 16:57:42 INFO DAGScheduler: Submitting ResultStage 1 (MapPartitionsRDD[2] at map at Session17Assignment2New.scala:68), which has no missing parents

17/08/03 16:57:42 INFO MemoryStore: Block broadcast\_2 stored as values in memory (estimated size 3.2 KB, free 349.1 MB)

17/08/03 16:57:42 INFO MemoryStore: Block broadcast\_2\_piece0 stored as bytes in memory (estimated size 1999.0 B, free 349.1 MB)

17/08/03 16:57:42 INFO BlockManagerInfo: Added broadcast\_2\_piece0 in memory on 192.168.56.1:54675 (size: 1999.0 B, free: 349.2 MB)

17/08/03 16:57:42 INFO SparkContext: Created broadcast 2 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:42 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 1 (MapPartitionsRDD[2] at map at Session17Assignment2New.scala:68)

17/08/03 16:57:42 INFO TaskSchedulerImpl: Adding task set 1.0 with 1 tasks

17/08/03 16:57:42 INFO TaskSetManager: Starting task 0.0 in stage 1.0 (TID 1, localhost, partition 0, PROCESS\_LOCAL, 5439 bytes)

17/08/03 16:57:42 INFO Executor: Running task 0.0 in stage 1.0 (TID 1)

(Mathew,science,grade-3,45,12)

(Mathew,history,grade-2,55,13)

(Mark,maths,grade-2,23,13)

(Mark,science,grade-1,76,13)

(John,history,grade-1,14,12)

17/08/03 16:57:42 INFO HadoopRDD: Input split: file:/G:/ACADGILD/course material/Hadoop/Sessions/Session 17/Assignments/Assignment2/17.2\_Dataset.txt:0+622

(John,maths,grade-2,74,13)

(Lisa,science,grade-1,24,12)

(Lisa,history,grade-3,86,13)

(Andrew,maths,grade-1,34,13)

(Andrew,science,grade-3,26,14)

(Andrew,history,grade-1,74,12)

(Mathew,science,grade-2,55,12)

(Mathew,history,grade-2,87,12)

(Mark,maths,grade-1,92,13)

(Mark,science,grade-2,12,12)

(John,history,grade-1,67,13)

(John,maths,grade-1,35,11)

(Lisa,science,grade-2,24,13)

(Lisa,history,grade-2,98,15)

(Andrew,maths,grade-1,23,16)

(Andrew,science,grade-3,44,14)

(Andrew,history,grade-2,77,11)

17/08/03 16:57:42 INFO Executor: Finished task 0.0 in stage 1.0 (TID 1). 916 bytes result sent to driver

17/08/03 16:57:42 INFO TaskSetManager: Finished task 0.0 in stage 1.0 (TID 1) in 31 ms on localhost (1/1)

17/08/03 16:57:42 INFO TaskSchedulerImpl: Removed TaskSet 1.0, whose tasks have all completed, from pool

17/08/03 16:57:42 INFO DAGScheduler: ResultStage 1 (foreach at Session17Assignment2New.scala:78) finished in 0.031 s

17/08/03 16:57:42 INFO DAGScheduler: Job 1 finished: foreach at Session17Assignment2New.scala:78, took 0.062198 s

17/08/03 16:57:42 INFO SparkContext: Starting job: count at Session17Assignment2New.scala:86

17/08/03 16:57:42 INFO DAGScheduler: Got job 2 (count at Session17Assignment2New.scala:86) with 1 output partitions

17/08/03 16:57:42 INFO DAGScheduler: Final stage: ResultStage 2 (count at Session17Assignment2New.scala:86)

17/08/03 16:57:42 INFO DAGScheduler: Parents of final stage: List()

17/08/03 16:57:42 INFO DAGScheduler: Missing parents: List()

17/08/03 16:57:42 INFO DAGScheduler: Submitting ResultStage 2 (MapPartitionsRDD[2] at map at Session17Assignment2New.scala:68), which has no missing parents

17/08/03 16:57:42 INFO MemoryStore: Block broadcast\_3 stored as values in memory (estimated size 3.1 KB, free 349.1 MB)

17/08/03 16:57:42 INFO MemoryStore: Block broadcast\_3\_piece0 stored as bytes in memory (estimated size 1968.0 B, free 349.1 MB)

17/08/03 16:57:42 INFO BlockManagerInfo: Added broadcast\_3\_piece0 in memory on 192.168.56.1:54675 (size: 1968.0 B, free: 349.2 MB)

17/08/03 16:57:42 INFO SparkContext: Created broadcast 3 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:42 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 2 (MapPartitionsRDD[2] at map at Session17Assignment2New.scala:68)

17/08/03 16:57:42 INFO TaskSchedulerImpl: Adding task set 2.0 with 1 tasks

17/08/03 16:57:42 INFO TaskSetManager: Starting task 0.0 in stage 2.0 (TID 2, localhost, partition 0, PROCESS\_LOCAL, 5355 bytes)

17/08/03 16:57:42 INFO Executor: Running task 0.0 in stage 2.0 (TID 2)

17/08/03 16:57:42 INFO HadoopRDD: Input split: file:/G:/ACADGILD/course material/Hadoop/Sessions/Session 17/Assignments/Assignment2/17.2\_Dataset.txt:0+622

17/08/03 16:57:42 INFO Executor: Finished task 0.0 in stage 2.0 (TID 2). 962 bytes result sent to driver

17/08/03 16:57:42 INFO TaskSetManager: Finished task 0.0 in stage 2.0 (TID 2) in 31 ms on localhost (1/1)

17/08/03 16:57:42 INFO TaskSchedulerImpl: Removed TaskSet 2.0, whose tasks have all completed, from pool

17/08/03 16:57:42 INFO DAGScheduler: ResultStage 2 (count at Session17Assignment2New.scala:86) finished in 0.031 s

17/08/03 16:57:42 INFO DAGScheduler: Job 2 finished: count at Session17Assignment2New.scala:86, took 0.060353 s

Total number of rows present : 22

17/08/03 16:57:42 INFO SparkContext: Starting job: toStream at Session17Assignment2New.scala:96

17/08/03 16:57:42 INFO DAGScheduler: Got job 3 (toStream at Session17Assignment2New.scala:96) with 1 output partitions

17/08/03 16:57:42 INFO DAGScheduler: Final stage: ResultStage 3 (toStream at Session17Assignment2New.scala:96)

17/08/03 16:57:42 INFO DAGScheduler: Parents of final stage: List()

17/08/03 16:57:42 INFO DAGScheduler: Missing parents: List()

17/08/03 16:57:42 INFO DAGScheduler: Submitting ResultStage 3 (MapPartitionsRDD[2] at map at Session17Assignment2New.scala:68), which has no missing parents

17/08/03 16:57:42 INFO MemoryStore: Block broadcast\_4 stored as values in memory (estimated size 3.4 KB, free 349.1 MB)

17/08/03 16:57:42 INFO MemoryStore: Block broadcast\_4\_piece0 stored as bytes in memory (estimated size 2035.0 B, free 349.1 MB)

17/08/03 16:57:42 INFO BlockManagerInfo: Added broadcast\_4\_piece0 in memory on 192.168.56.1:54675 (size: 2035.0 B, free: 349.2 MB)

17/08/03 16:57:42 INFO SparkContext: Created broadcast 4 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:42 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 3 (MapPartitionsRDD[2] at map at Session17Assignment2New.scala:68)

17/08/03 16:57:42 INFO TaskSchedulerImpl: Adding task set 3.0 with 1 tasks

17/08/03 16:57:42 INFO TaskSetManager: Starting task 0.0 in stage 3.0 (TID 3, localhost, partition 0, PROCESS\_LOCAL, 5355 bytes)

17/08/03 16:57:42 INFO Executor: Running task 0.0 in stage 3.0 (TID 3)

17/08/03 16:57:42 INFO HadoopRDD: Input split: file:/G:/ACADGILD/course material/Hadoop/Sessions/Session 17/Assignments/Assignment2/17.2\_Dataset.txt:0+622

<<<<<<<<--------- Distinct Subjects in entire School are as follows --------->>>>>>>>>>

science

history

maths

Distinct Number of Subjects present in entire school are : 3

17/08/03 16:57:42 INFO Executor: Finished task 0.0 in stage 3.0 (TID 3). 1873 bytes result sent to driver

17/08/03 16:57:42 INFO TaskSetManager: Finished task 0.0 in stage 3.0 (TID 3) in 31 ms on localhost (1/1)

17/08/03 16:57:42 INFO TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool

17/08/03 16:57:42 INFO DAGScheduler: ResultStage 3 (toStream at Session17Assignment2New.scala:96) finished in 0.031 s

17/08/03 16:57:42 INFO DAGScheduler: Job 3 finished: toStream at Session17Assignment2New.scala:96, took 0.044425 s

17/08/03 16:57:42 INFO SparkContext: Starting job: toStream at Session17Assignment2New.scala:115

17/08/03 16:57:42 INFO DAGScheduler: Got job 4 (toStream at Session17Assignment2New.scala:115) with 1 output partitions

17/08/03 16:57:42 INFO DAGScheduler: Final stage: ResultStage 4 (toStream at Session17Assignment2New.scala:115)

17/08/03 16:57:42 INFO DAGScheduler: Parents of final stage: List()

17/08/03 16:57:42 INFO DAGScheduler: Missing parents: List()

17/08/03 16:57:42 INFO DAGScheduler: Submitting ResultStage 4 (MapPartitionsRDD[2] at map at Session17Assignment2New.scala:68), which has no missing parents

17/08/03 16:57:42 INFO MemoryStore: Block broadcast\_5 stored as values in memory (estimated size 3.4 KB, free 349.1 MB)

17/08/03 16:57:42 INFO MemoryStore: Block broadcast\_5\_piece0 stored as bytes in memory (estimated size 2035.0 B, free 349.1 MB)

17/08/03 16:57:42 INFO BlockManagerInfo: Added broadcast\_5\_piece0 in memory on 192.168.56.1:54675 (size: 2035.0 B, free: 349.2 MB)

17/08/03 16:57:42 INFO SparkContext: Created broadcast 5 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:42 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 4 (MapPartitionsRDD[2] at map at Session17Assignment2New.scala:68)

17/08/03 16:57:42 INFO TaskSchedulerImpl: Adding task set 4.0 with 1 tasks

17/08/03 16:57:42 INFO TaskSetManager: Starting task 0.0 in stage 4.0 (TID 4, localhost, partition 0, PROCESS\_LOCAL, 5355 bytes)

17/08/03 16:57:42 INFO Executor: Running task 0.0 in stage 4.0 (TID 4)

17/08/03 16:57:42 INFO HadoopRDD: Input split: file:/G:/ACADGILD/course material/Hadoop/Sessions/Session 17/Assignments/Assignment2/17.2\_Dataset.txt:0+622

<<<<<----- Students in the school, whose name is Mathew and marks is 55 ----->>>>>

(Mathew,history,grade-2,55,13)

(Mathew,science,grade-2,55,12)

Count of the number of students in the school, whose name is Mathew and marks is 55 : 2

17/08/03 16:57:42 INFO Executor: Finished task 0.0 in stage 4.0 (TID 4). 1873 bytes result sent to driver

17/08/03 16:57:42 INFO TaskSetManager: Finished task 0.0 in stage 4.0 (TID 4) in 16 ms on localhost (1/1)

17/08/03 16:57:42 INFO TaskSchedulerImpl: Removed TaskSet 4.0, whose tasks have all completed, from pool

17/08/03 16:57:42 INFO DAGScheduler: ResultStage 4 (toStream at Session17Assignment2New.scala:115) finished in 0.016 s

17/08/03 16:57:42 INFO DAGScheduler: Job 4 finished: toStream at Session17Assignment2New.scala:115, took 0.036869 s

<<<<---- Count of students per grade in the school are as follows ---->>>>

17/08/03 16:57:43 INFO SparkContext: Starting job: foreach at Session17Assignment2New.scala:153

17/08/03 16:57:43 INFO DAGScheduler: Registering RDD 3 (groupBy at Session17Assignment2New.scala:140)

17/08/03 16:57:43 INFO DAGScheduler: Registering RDD 5 (sortBy at Session17Assignment2New.scala:144)

17/08/03 16:57:43 INFO DAGScheduler: Got job 5 (foreach at Session17Assignment2New.scala:153) with 1 output partitions

17/08/03 16:57:43 INFO DAGScheduler: Final stage: ResultStage 7 (foreach at Session17Assignment2New.scala:153)

17/08/03 16:57:43 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 6)

17/08/03 16:57:43 INFO DAGScheduler: Missing parents: List(ShuffleMapStage 6)

17/08/03 16:57:43 INFO DAGScheduler: Submitting ShuffleMapStage 5 (MapPartitionsRDD[3] at groupBy at Session17Assignment2New.scala:140), which has no missing parents

17/08/03 16:57:43 INFO BlockManagerInfo: Removed broadcast\_1\_piece0 on 192.168.56.1:54675 in memory (size: 1905.0 B, free: 349.2 MB)

17/08/03 16:57:43 INFO BlockManagerInfo: Removed broadcast\_2\_piece0 on 192.168.56.1:54675 in memory (size: 1999.0 B, free: 349.2 MB)

17/08/03 16:57:43 INFO BlockManagerInfo: Removed broadcast\_3\_piece0 on 192.168.56.1:54675 in memory (size: 1968.0 B, free: 349.2 MB)

17/08/03 16:57:43 INFO MemoryStore: Block broadcast\_6 stored as values in memory (estimated size 5.0 KB, free 349.1 MB)

17/08/03 16:57:43 INFO BlockManagerInfo: Removed broadcast\_4\_piece0 on 192.168.56.1:54675 in memory (size: 2035.0 B, free: 349.2 MB)

17/08/03 16:57:43 INFO MemoryStore: Block broadcast\_6\_piece0 stored as bytes in memory (estimated size 2.7 KB, free 349.1 MB)

17/08/03 16:57:43 INFO BlockManagerInfo: Added broadcast\_6\_piece0 in memory on 192.168.56.1:54675 (size: 2.7 KB, free: 349.2 MB)

17/08/03 16:57:43 INFO SparkContext: Created broadcast 6 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:43 INFO DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 5 (MapPartitionsRDD[3] at groupBy at Session17Assignment2New.scala:140)

17/08/03 16:57:43 INFO TaskSchedulerImpl: Adding task set 5.0 with 1 tasks

17/08/03 16:57:43 INFO BlockManagerInfo: Removed broadcast\_5\_piece0 on 192.168.56.1:54675 in memory (size: 2035.0 B, free: 349.2 MB)

17/08/03 16:57:43 INFO TaskSetManager: Starting task 0.0 in stage 5.0 (TID 5, localhost, partition 0, PROCESS\_LOCAL, 5429 bytes)

17/08/03 16:57:43 INFO Executor: Running task 0.0 in stage 5.0 (TID 5)

17/08/03 16:57:43 INFO HadoopRDD: Input split: file:/G:/ACADGILD/course material/Hadoop/Sessions/Session 17/Assignments/Assignment2/17.2\_Dataset.txt:0+622

17/08/03 16:57:43 INFO Executor: Finished task 0.0 in stage 5.0 (TID 5). 1155 bytes result sent to driver

17/08/03 16:57:43 INFO TaskSetManager: Finished task 0.0 in stage 5.0 (TID 5) in 78 ms on localhost (1/1)

17/08/03 16:57:43 INFO TaskSchedulerImpl: Removed TaskSet 5.0, whose tasks have all completed, from pool

17/08/03 16:57:43 INFO DAGScheduler: ShuffleMapStage 5 (groupBy at Session17Assignment2New.scala:140) finished in 0.078 s

17/08/03 16:57:43 INFO DAGScheduler: looking for newly runnable stages

17/08/03 16:57:43 INFO DAGScheduler: running: Set()

17/08/03 16:57:43 INFO DAGScheduler: waiting: Set(ShuffleMapStage 6, ResultStage 7)

17/08/03 16:57:43 INFO DAGScheduler: failed: Set()

17/08/03 16:57:43 INFO DAGScheduler: Submitting ShuffleMapStage 6 (MapPartitionsRDD[5] at sortBy at Session17Assignment2New.scala:144), which has no missing parents

17/08/03 16:57:43 INFO MemoryStore: Block broadcast\_7 stored as values in memory (estimated size 6.5 KB, free 349.1 MB)

17/08/03 16:57:43 INFO MemoryStore: Block broadcast\_7\_piece0 stored as bytes in memory (estimated size 3.3 KB, free 349.1 MB)

17/08/03 16:57:43 INFO BlockManagerInfo: Added broadcast\_7\_piece0 in memory on 192.168.56.1:54675 (size: 3.3 KB, free: 349.2 MB)

17/08/03 16:57:43 INFO SparkContext: Created broadcast 7 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:43 INFO DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 6 (MapPartitionsRDD[5] at sortBy at Session17Assignment2New.scala:144)

17/08/03 16:57:43 INFO TaskSchedulerImpl: Adding task set 6.0 with 1 tasks

17/08/03 16:57:43 INFO TaskSetManager: Starting task 0.0 in stage 6.0 (TID 6, localhost, partition 0, ANY, 5130 bytes)

17/08/03 16:57:43 INFO Executor: Running task 0.0 in stage 6.0 (TID 6)

17/08/03 16:57:43 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:43 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 16 ms

17/08/03 16:57:43 INFO Executor: Finished task 0.0 in stage 6.0 (TID 6). 1879 bytes result sent to driver

17/08/03 16:57:43 INFO TaskSetManager: Finished task 0.0 in stage 6.0 (TID 6) in 125 ms on localhost (1/1)

17/08/03 16:57:43 INFO TaskSchedulerImpl: Removed TaskSet 6.0, whose tasks have all completed, from pool

17/08/03 16:57:43 INFO DAGScheduler: ShuffleMapStage 6 (sortBy at Session17Assignment2New.scala:144) finished in 0.125 s

17/08/03 16:57:43 INFO DAGScheduler: looking for newly runnable stages

17/08/03 16:57:43 INFO DAGScheduler: running: Set()

17/08/03 16:57:43 INFO DAGScheduler: waiting: Set(ResultStage 7)

17/08/03 16:57:43 INFO DAGScheduler: failed: Set()

17/08/03 16:57:43 INFO DAGScheduler: Submitting ResultStage 7 (MapPartitionsRDD[8] at map at Session17Assignment2New.scala:144), which has no missing parents

17/08/03 16:57:43 INFO MemoryStore: Block broadcast\_8 stored as values in memory (estimated size 3.5 KB, free 349.1 MB)

17/08/03 16:57:43 INFO MemoryStore: Block broadcast\_8\_piece0 stored as bytes in memory (estimated size 2024.0 B, free 349.1 MB)

17/08/03 16:57:43 INFO BlockManagerInfo: Added broadcast\_8\_piece0 in memory on 192.168.56.1:54675 (size: 2024.0 B, free: 349.2 MB)

17/08/03 16:57:43 INFO SparkContext: Created broadcast 8 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:43 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 7 (MapPartitionsRDD[8] at map at Session17Assignment2New.scala:144)

17/08/03 16:57:43 INFO TaskSchedulerImpl: Adding task set 7.0 with 1 tasks

17/08/03 16:57:43 INFO TaskSetManager: Starting task 0.0 in stage 7.0 (TID 7, localhost, partition 0, ANY, 5141 bytes)

17/08/03 16:57:43 INFO Executor: Running task 0.0 in stage 7.0 (TID 7)

17/08/03 16:57:43 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:43 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

(grade-1,9)

(grade-2,9)

(grade-3,4)

17/08/03 16:57:43 INFO Executor: Finished task 0.0 in stage 7.0 (TID 7). 1553 bytes result sent to driver

17/08/03 16:57:43 INFO TaskSetManager: Finished task 0.0 in stage 7.0 (TID 7) in 63 ms on localhost (1/1)

17/08/03 16:57:43 INFO TaskSchedulerImpl: Removed TaskSet 7.0, whose tasks have all completed, from pool

17/08/03 16:57:43 INFO DAGScheduler: ResultStage 7 (foreach at Session17Assignment2New.scala:153) finished in 0.063 s

17/08/03 16:57:43 INFO DAGScheduler: Job 5 finished: foreach at Session17Assignment2New.scala:153, took 0.382031 s

17/08/03 16:57:43 INFO SparkContext: Starting job: foreach at Session17Assignment2New.scala:180

17/08/03 16:57:43 INFO DAGScheduler: Registering RDD 9 (groupBy at Session17Assignment2New.scala:159)

17/08/03 16:57:43 INFO DAGScheduler: Registering RDD 12 (sortBy at Session17Assignment2New.scala:163)

17/08/03 16:57:43 INFO DAGScheduler: Got job 6 (foreach at Session17Assignment2New.scala:180) with 1 output partitions

17/08/03 16:57:43 INFO DAGScheduler: Final stage: ResultStage 10 (foreach at Session17Assignment2New.scala:180)

17/08/03 16:57:43 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 9)

17/08/03 16:57:43 INFO DAGScheduler: Missing parents: List(ShuffleMapStage 9)

17/08/03 16:57:43 INFO DAGScheduler: Submitting ShuffleMapStage 8 (MapPartitionsRDD[9] at groupBy at Session17Assignment2New.scala:159), which has no missing parents

17/08/03 16:57:43 INFO MemoryStore: Block broadcast\_9 stored as values in memory (estimated size 5.0 KB, free 349.1 MB)

17/08/03 16:57:43 INFO MemoryStore: Block broadcast\_9\_piece0 stored as bytes in memory (estimated size 2.7 KB, free 349.1 MB)

17/08/03 16:57:43 INFO BlockManagerInfo: Added broadcast\_9\_piece0 in memory on 192.168.56.1:54675 (size: 2.7 KB, free: 349.2 MB)

17/08/03 16:57:43 INFO SparkContext: Created broadcast 9 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:43 INFO DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 8 (MapPartitionsRDD[9] at groupBy at Session17Assignment2New.scala:159)

17/08/03 16:57:43 INFO TaskSchedulerImpl: Adding task set 8.0 with 1 tasks

17/08/03 16:57:43 INFO TaskSetManager: Starting task 0.0 in stage 8.0 (TID 8, localhost, partition 0, PROCESS\_LOCAL, 5429 bytes)

17/08/03 16:57:43 INFO Executor: Running task 0.0 in stage 8.0 (TID 8)

17/08/03 16:57:43 INFO HadoopRDD: Input split: file:/G:/ACADGILD/course material/Hadoop/Sessions/Session 17/Assignments/Assignment2/17.2\_Dataset.txt:0+622

17/08/03 16:57:43 INFO Executor: Finished task 0.0 in stage 8.0 (TID 8). 1242 bytes result sent to driver

17/08/03 16:57:43 INFO TaskSetManager: Finished task 0.0 in stage 8.0 (TID 8) in 46 ms on localhost (1/1)

17/08/03 16:57:43 INFO TaskSchedulerImpl: Removed TaskSet 8.0, whose tasks have all completed, from pool

17/08/03 16:57:43 INFO DAGScheduler: ShuffleMapStage 8 (groupBy at Session17Assignment2New.scala:159) finished in 0.046 s

17/08/03 16:57:43 INFO DAGScheduler: looking for newly runnable stages

17/08/03 16:57:43 INFO DAGScheduler: running: Set()

17/08/03 16:57:43 INFO DAGScheduler: waiting: Set(ShuffleMapStage 9, ResultStage 10)

17/08/03 16:57:43 INFO DAGScheduler: failed: Set()

17/08/03 16:57:43 INFO DAGScheduler: Submitting ShuffleMapStage 9 (MapPartitionsRDD[12] at sortBy at Session17Assignment2New.scala:163), which has no missing parents

17/08/03 16:57:43 INFO MemoryStore: Block broadcast\_10 stored as values in memory (estimated size 6.7 KB, free 349.0 MB)

17/08/03 16:57:43 INFO MemoryStore: Block broadcast\_10\_piece0 stored as bytes in memory (estimated size 3.4 KB, free 349.0 MB)

17/08/03 16:57:43 INFO BlockManagerInfo: Added broadcast\_10\_piece0 in memory on 192.168.56.1:54675 (size: 3.4 KB, free: 349.2 MB)

17/08/03 16:57:43 INFO SparkContext: Created broadcast 10 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:43 INFO DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 9 (MapPartitionsRDD[12] at sortBy at Session17Assignment2New.scala:163)

17/08/03 16:57:43 INFO TaskSchedulerImpl: Adding task set 9.0 with 1 tasks

17/08/03 16:57:43 INFO TaskSetManager: Starting task 0.0 in stage 9.0 (TID 9, localhost, partition 0, ANY, 5130 bytes)

17/08/03 16:57:43 INFO Executor: Running task 0.0 in stage 9.0 (TID 9)

17/08/03 16:57:43 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:43 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

17/08/03 16:57:43 INFO Executor: Finished task 0.0 in stage 9.0 (TID 9). 1879 bytes result sent to driver

17/08/03 16:57:43 INFO TaskSetManager: Finished task 0.0 in stage 9.0 (TID 9) in 47 ms on localhost (1/1)

17/08/03 16:57:43 INFO TaskSchedulerImpl: Removed TaskSet 9.0, whose tasks have all completed, from pool

17/08/03 16:57:43 INFO DAGScheduler: ShuffleMapStage 9 (sortBy at Session17Assignment2New.scala:163) finished in 0.047 s

17/08/03 16:57:43 INFO DAGScheduler: looking for newly runnable stages

17/08/03 16:57:43 INFO DAGScheduler: running: Set()

17/08/03 16:57:43 INFO DAGScheduler: waiting: Set(ResultStage 10)

17/08/03 16:57:43 INFO DAGScheduler: failed: Set()

17/08/03 16:57:43 INFO DAGScheduler: Submitting ResultStage 10 (MapPartitionsRDD[15] at map at Session17Assignment2New.scala:177), which has no missing parents

17/08/03 16:57:43 INFO MemoryStore: Block broadcast\_11 stored as values in memory (estimated size 3.6 KB, free 349.0 MB)

17/08/03 16:57:43 INFO MemoryStore: Block broadcast\_11\_piece0 stored as bytes in memory (estimated size 2.0 KB, free 349.0 MB)

17/08/03 16:57:43 INFO BlockManagerInfo: Added broadcast\_11\_piece0 in memory on 192.168.56.1:54675 (size: 2.0 KB, free: 349.2 MB)

17/08/03 16:57:43 INFO SparkContext: Created broadcast 11 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:43 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 10 (MapPartitionsRDD[15] at map at Session17Assignment2New.scala:177)

17/08/03 16:57:43 INFO TaskSchedulerImpl: Adding task set 10.0 with 1 tasks

17/08/03 16:57:43 INFO TaskSetManager: Starting task 0.0 in stage 10.0 (TID 10, localhost, partition 0, ANY, 5141 bytes)

17/08/03 16:57:43 INFO Executor: Running task 0.0 in stage 10.0 (TID 10)

(Andrew,grade-2,77,1)

(Andrew,grade-1,131,3)

(Andrew,grade-3,70,2)

(John,grade-1,116,3)

(John,grade-2,74,1)

(Lisa,grade-1,24,1)

(Lisa,grade-2,122,2)

(Lisa,grade-3,86,1)

(Mark,grade-2,35,2)

(Mark,grade-1,168,2)

(Mathew,grade-3,45,1)

(Mathew,grade-2,197,3)

17/08/03 16:57:43 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:43 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

17/08/03 16:57:43 INFO Executor: Finished task 0.0 in stage 10.0 (TID 10). 1466 bytes result sent to driver

17/08/03 16:57:43 INFO TaskSetManager: Finished task 0.0 in stage 10.0 (TID 10) in 15 ms on localhost (1/1)

17/08/03 16:57:43 INFO TaskSchedulerImpl: Removed TaskSet 10.0, whose tasks have all completed, from pool

17/08/03 16:57:43 INFO DAGScheduler: ResultStage 10 (foreach at Session17Assignment2New.scala:180) finished in 0.015 s

<<<<---- Average of each student per grade (Note - Mathew is grade-1, is different from Mathew in some other grade!) ---->>>>

17/08/03 16:57:43 INFO DAGScheduler: Job 6 finished: foreach at Session17Assignment2New.scala:180, took 0.172351 s

17/08/03 16:57:43 INFO SparkContext: Starting job: foreach at Session17Assignment2New.scala:189

17/08/03 16:57:43 INFO MapOutputTrackerMaster: Size of output statuses for shuffle 3 is 147 bytes

17/08/03 16:57:43 INFO MapOutputTrackerMaster: Size of output statuses for shuffle 2 is 147 bytes

17/08/03 16:57:43 INFO DAGScheduler: Got job 7 (foreach at Session17Assignment2New.scala:189) with 1 output partitions

17/08/03 16:57:43 INFO DAGScheduler: Final stage: ResultStage 13 (foreach at Session17Assignment2New.scala:189)

17/08/03 16:57:43 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 12)

17/08/03 16:57:43 INFO DAGScheduler: Missing parents: List()

17/08/03 16:57:43 INFO DAGScheduler: Submitting ResultStage 13 (MapPartitionsRDD[16] at map at Session17Assignment2New.scala:183), which has no missing parents

17/08/03 16:57:43 INFO MemoryStore: Block broadcast\_12 stored as values in memory (estimated size 3.7 KB, free 349.0 MB)

17/08/03 16:57:43 INFO MemoryStore: Block broadcast\_12\_piece0 stored as bytes in memory (estimated size 2.1 KB, free 349.0 MB)

17/08/03 16:57:43 INFO BlockManagerInfo: Added broadcast\_12\_piece0 in memory on 192.168.56.1:54675 (size: 2.1 KB, free: 349.2 MB)

17/08/03 16:57:43 INFO SparkContext: Created broadcast 12 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:43 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 13 (MapPartitionsRDD[16] at map at Session17Assignment2New.scala:183)

17/08/03 16:57:43 INFO TaskSchedulerImpl: Adding task set 13.0 with 1 tasks

(Andrew,grade-2,77.0)

(Andrew,grade-1,43.666668)

(Andrew,grade-3,35.0)

(John,grade-1,38.666668)

(John,grade-2,74.0)

(Lisa,grade-1,24.0)

(Lisa,grade-2,61.0)

(Lisa,grade-3,86.0)

(Mark,grade-2,17.5)

(Mark,grade-1,84.0)

(Mathew,grade-3,45.0)

(Mathew,grade-2,65.666664)

17/08/03 16:57:43 INFO TaskSetManager: Starting task 0.0 in stage 13.0 (TID 11, localhost, partition 0, ANY, 5141 bytes)

17/08/03 16:57:43 INFO Executor: Running task 0.0 in stage 13.0 (TID 11)

17/08/03 16:57:43 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:43 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

17/08/03 16:57:43 INFO Executor: Finished task 0.0 in stage 13.0 (TID 11). 1466 bytes result sent to driver

17/08/03 16:57:43 INFO TaskSetManager: Finished task 0.0 in stage 13.0 (TID 11) in 15 ms on localhost (1/1)

17/08/03 16:57:43 INFO TaskSchedulerImpl: Removed TaskSet 13.0, whose tasks have all completed, from pool

17/08/03 16:57:43 INFO DAGScheduler: ResultStage 13 (foreach at Session17Assignment2New.scala:189) finished in 0.031 s

17/08/03 16:57:43 INFO DAGScheduler: Job 7 finished: foreach at Session17Assignment2New.scala:189, took 0.049531 s

17/08/03 16:57:43 INFO SparkContext: Starting job: foreach at Session17Assignment2New.scala:199

17/08/03 16:57:44 INFO DAGScheduler: Registering RDD 17 (groupBy at Session17Assignment2New.scala:193)

17/08/03 16:57:44 INFO DAGScheduler: Registering RDD 20 (sortBy at Session17Assignment2New.scala:195)

17/08/03 16:57:44 INFO DAGScheduler: Got job 8 (foreach at Session17Assignment2New.scala:199) with 1 output partitions

17/08/03 16:57:44 INFO DAGScheduler: Final stage: ResultStage 16 (foreach at Session17Assignment2New.scala:199)

17/08/03 16:57:44 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 15)

17/08/03 16:57:44 INFO DAGScheduler: Missing parents: List(ShuffleMapStage 15)

17/08/03 16:57:44 INFO DAGScheduler: Submitting ShuffleMapStage 14 (MapPartitionsRDD[17] at groupBy at Session17Assignment2New.scala:193), which has no missing parents

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_13 stored as values in memory (estimated size 5.0 KB, free 349.0 MB)

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_13\_piece0 stored as bytes in memory (estimated size 2.7 KB, free 349.0 MB)

17/08/03 16:57:44 INFO BlockManagerInfo: Added broadcast\_13\_piece0 in memory on 192.168.56.1:54675 (size: 2.7 KB, free: 349.2 MB)

17/08/03 16:57:44 INFO SparkContext: Created broadcast 13 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:44 INFO DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 14 (MapPartitionsRDD[17] at groupBy at Session17Assignment2New.scala:193)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Adding task set 14.0 with 1 tasks

17/08/03 16:57:44 INFO TaskSetManager: Starting task 0.0 in stage 14.0 (TID 12, localhost, partition 0, PROCESS\_LOCAL, 5429 bytes)

17/08/03 16:57:44 INFO Executor: Running task 0.0 in stage 14.0 (TID 12)

17/08/03 16:57:44 INFO HadoopRDD: Input split: file:/G:/ACADGILD/course material/Hadoop/Sessions/Session 17/Assignments/Assignment2/17.2\_Dataset.txt:0+622

17/08/03 16:57:44 INFO Executor: Finished task 0.0 in stage 14.0 (TID 12). 1155 bytes result sent to driver

17/08/03 16:57:44 INFO TaskSetManager: Finished task 0.0 in stage 14.0 (TID 12) in 31 ms on localhost (1/1)

17/08/03 16:57:44 INFO DAGScheduler: ShuffleMapStage 14 (groupBy at Session17Assignment2New.scala:193) finished in 0.031 s

17/08/03 16:57:44 INFO DAGScheduler: looking for newly runnable stages

17/08/03 16:57:44 INFO DAGScheduler: running: Set()

17/08/03 16:57:44 INFO DAGScheduler: waiting: Set(ShuffleMapStage 15, ResultStage 16)

17/08/03 16:57:44 INFO DAGScheduler: failed: Set()

17/08/03 16:57:44 INFO DAGScheduler: Submitting ShuffleMapStage 15 (MapPartitionsRDD[20] at sortBy at Session17Assignment2New.scala:195), which has no missing parents

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_14 stored as values in memory (estimated size 6.7 KB, free 349.0 MB)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Removed TaskSet 14.0, whose tasks have all completed, from pool

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_14\_piece0 stored as bytes in memory (estimated size 3.4 KB, free 349.0 MB)

17/08/03 16:57:44 INFO BlockManagerInfo: Added broadcast\_14\_piece0 in memory on 192.168.56.1:54675 (size: 3.4 KB, free: 349.2 MB)

17/08/03 16:57:44 INFO SparkContext: Created broadcast 14 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:44 INFO DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 15 (MapPartitionsRDD[20] at sortBy at Session17Assignment2New.scala:195)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Adding task set 15.0 with 1 tasks

17/08/03 16:57:44 INFO TaskSetManager: Starting task 0.0 in stage 15.0 (TID 13, localhost, partition 0, ANY, 5130 bytes)

17/08/03 16:57:44 INFO Executor: Running task 0.0 in stage 15.0 (TID 13)

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

17/08/03 16:57:44 INFO Executor: Finished task 0.0 in stage 15.0 (TID 13). 1879 bytes result sent to driver

17/08/03 16:57:44 INFO TaskSetManager: Finished task 0.0 in stage 15.0 (TID 13) in 32 ms on localhost (1/1)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Removed TaskSet 15.0, whose tasks have all completed, from pool

17/08/03 16:57:44 INFO DAGScheduler: ShuffleMapStage 15 (sortBy at Session17Assignment2New.scala:195) finished in 0.032 s

17/08/03 16:57:44 INFO DAGScheduler: looking for newly runnable stages

17/08/03 16:57:44 INFO DAGScheduler: running: Set()

17/08/03 16:57:44 INFO DAGScheduler: waiting: Set(ResultStage 16)

17/08/03 16:57:44 INFO DAGScheduler: failed: Set()

17/08/03 16:57:44 INFO DAGScheduler: Submitting ResultStage 16 (MapPartitionsRDD[23] at map at Session17Assignment2New.scala:198), which has no missing parents

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_15 stored as values in memory (estimated size 3.6 KB, free 349.0 MB)

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_15\_piece0 stored as bytes in memory (estimated size 2.1 KB, free 349.0 MB)

17/08/03 16:57:44 INFO BlockManagerInfo: Added broadcast\_15\_piece0 in memory on 192.168.56.1:54675 (size: 2.1 KB, free: 349.2 MB)

17/08/03 16:57:44 INFO SparkContext: Created broadcast 15 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:44 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 16 (MapPartitionsRDD[23] at map at Session17Assignment2New.scala:198)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Adding task set 16.0 with 1 tasks

17/08/03 16:57:44 INFO TaskSetManager: Starting task 0.0 in stage 16.0 (TID 14, localhost, partition 0, ANY, 5141 bytes)

17/08/03 16:57:44 INFO Executor: Running task 0.0 in stage 16.0 (TID 14)

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

(Andrew,science,70,2)

(Andrew,maths,57,2)

(Andrew,history,151,2)

(John,maths,109,2)

(John,history,81,2)

(Lisa,history,184,2)

(Lisa,science,48,2)

(Mark,maths,115,2)

(Mark,science,88,2)

(Mathew,science,100,2)

(Mathew,history,142,2)

17/08/03 16:57:44 INFO Executor: Finished task 0.0 in stage 16.0 (TID 14). 1466 bytes result sent to driver

17/08/03 16:57:44 INFO TaskSetManager: Finished task 0.0 in stage 16.0 (TID 14) in 16 ms on localhost (1/1)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Removed TaskSet 16.0, whose tasks have all completed, from pool

17/08/03 16:57:44 INFO DAGScheduler: ResultStage 16 (foreach at Session17Assignment2New.scala:199) finished in 0.016 s

17/08/03 16:57:44 INFO DAGScheduler: Job 8 finished: foreach at Session17Assignment2New.scala:199, took 0.121773 s

<<<<---- Average score of students in each subject across all grades ---->>>>

17/08/03 16:57:44 INFO SparkContext: Starting job: foreach at Session17Assignment2New.scala:204

17/08/03 16:57:44 INFO MapOutputTrackerMaster: Size of output statuses for shuffle 5 is 147 bytes

17/08/03 16:57:44 INFO MapOutputTrackerMaster: Size of output statuses for shuffle 4 is 147 bytes

17/08/03 16:57:44 INFO DAGScheduler: Got job 9 (foreach at Session17Assignment2New.scala:204) with 1 output partitions

17/08/03 16:57:44 INFO DAGScheduler: Final stage: ResultStage 19 (foreach at Session17Assignment2New.scala:204)

17/08/03 16:57:44 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 18)

17/08/03 16:57:44 INFO DAGScheduler: Missing parents: List()

17/08/03 16:57:44 INFO DAGScheduler: Submitting ResultStage 19 (MapPartitionsRDD[24] at map at Session17Assignment2New.scala:202), which has no missing parents

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_16 stored as values in memory (estimated size 3.7 KB, free 349.0 MB)

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_16\_piece0 stored as bytes in memory (estimated size 2.1 KB, free 349.0 MB)

17/08/03 16:57:44 INFO BlockManagerInfo: Added broadcast\_16\_piece0 in memory on 192.168.56.1:54675 (size: 2.1 KB, free: 349.2 MB)

17/08/03 16:57:44 INFO SparkContext: Created broadcast 16 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:44 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 19 (MapPartitionsRDD[24] at map at Session17Assignment2New.scala:202)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Adding task set 19.0 with 1 tasks

17/08/03 16:57:44 INFO TaskSetManager: Starting task 0.0 in stage 19.0 (TID 15, localhost, partition 0, ANY, 5141 bytes)

17/08/03 16:57:44 INFO Executor: Running task 0.0 in stage 19.0 (TID 15)

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

(Andrew,science,35.0)

(Andrew,maths,28.5)

(Andrew,history,75.5)

(John,maths,54.5)

(John,history,40.5)

(Lisa,history,92.0)

(Lisa,science,24.0)

(Mark,maths,57.5)

(Mark,science,44.0)

(Mathew,science,50.0)

(Mathew,history,71.0)

17/08/03 16:57:44 INFO Executor: Finished task 0.0 in stage 19.0 (TID 15). 1474 bytes result sent to driver

17/08/03 16:57:44 INFO TaskSetManager: Finished task 0.0 in stage 19.0 (TID 15) in 15 ms on localhost (1/1)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Removed TaskSet 19.0, whose tasks have all completed, from pool

17/08/03 16:57:44 INFO DAGScheduler: ResultStage 19 (foreach at Session17Assignment2New.scala:204) finished in 0.015 s

17/08/03 16:57:44 INFO DAGScheduler: Job 9 finished: foreach at Session17Assignment2New.scala:204, took 0.046730 s

17/08/03 16:57:44 INFO SparkContext: Starting job: foreach at Session17Assignment2New.scala:217

17/08/03 16:57:44 INFO DAGScheduler: Registering RDD 25 (groupBy at Session17Assignment2New.scala:209)

17/08/03 16:57:44 INFO DAGScheduler: Registering RDD 28 (sortBy at Session17Assignment2New.scala:212)

17/08/03 16:57:44 INFO DAGScheduler: Got job 10 (foreach at Session17Assignment2New.scala:217) with 1 output partitions

17/08/03 16:57:44 INFO DAGScheduler: Final stage: ResultStage 22 (foreach at Session17Assignment2New.scala:217)

17/08/03 16:57:44 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 21)

17/08/03 16:57:44 INFO DAGScheduler: Missing parents: List(ShuffleMapStage 21)

17/08/03 16:57:44 INFO DAGScheduler: Submitting ShuffleMapStage 20 (MapPartitionsRDD[25] at groupBy at Session17Assignment2New.scala:209), which has no missing parents

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_17 stored as values in memory (estimated size 5.1 KB, free 349.0 MB)

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_17\_piece0 stored as bytes in memory (estimated size 2.7 KB, free 349.0 MB)

17/08/03 16:57:44 INFO BlockManagerInfo: Added broadcast\_17\_piece0 in memory on 192.168.56.1:54675 (size: 2.7 KB, free: 349.2 MB)

17/08/03 16:57:44 INFO SparkContext: Created broadcast 17 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:44 INFO DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 20 (MapPartitionsRDD[25] at groupBy at Session17Assignment2New.scala:209)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Adding task set 20.0 with 1 tasks

17/08/03 16:57:44 INFO TaskSetManager: Starting task 0.0 in stage 20.0 (TID 16, localhost, partition 0, PROCESS\_LOCAL, 5429 bytes)

17/08/03 16:57:44 INFO Executor: Running task 0.0 in stage 20.0 (TID 16)

17/08/03 16:57:44 INFO HadoopRDD: Input split: file:/G:/ACADGILD/course material/Hadoop/Sessions/Session 17/Assignments/Assignment2/17.2\_Dataset.txt:0+622

17/08/03 16:57:44 INFO Executor: Finished task 0.0 in stage 20.0 (TID 16). 1155 bytes result sent to driver

17/08/03 16:57:44 INFO TaskSetManager: Finished task 0.0 in stage 20.0 (TID 16) in 31 ms on localhost (1/1)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Removed TaskSet 20.0, whose tasks have all completed, from pool

17/08/03 16:57:44 INFO DAGScheduler: ShuffleMapStage 20 (groupBy at Session17Assignment2New.scala:209) finished in 0.031 s

17/08/03 16:57:44 INFO DAGScheduler: looking for newly runnable stages

17/08/03 16:57:44 INFO DAGScheduler: running: Set()

17/08/03 16:57:44 INFO DAGScheduler: waiting: Set(ShuffleMapStage 21, ResultStage 22)

17/08/03 16:57:44 INFO DAGScheduler: failed: Set()

17/08/03 16:57:44 INFO DAGScheduler: Submitting ShuffleMapStage 21 (MapPartitionsRDD[28] at sortBy at Session17Assignment2New.scala:212), which has no missing parents

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_18 stored as values in memory (estimated size 6.8 KB, free 349.0 MB)

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_18\_piece0 stored as bytes in memory (estimated size 3.5 KB, free 349.0 MB)

17/08/03 16:57:44 INFO BlockManagerInfo: Added broadcast\_18\_piece0 in memory on 192.168.56.1:54675 (size: 3.5 KB, free: 349.2 MB)

17/08/03 16:57:44 INFO SparkContext: Created broadcast 18 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:44 INFO DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 21 (MapPartitionsRDD[28] at sortBy at Session17Assignment2New.scala:212)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Adding task set 21.0 with 1 tasks

17/08/03 16:57:44 INFO TaskSetManager: Starting task 0.0 in stage 21.0 (TID 17, localhost, partition 0, ANY, 5130 bytes)

17/08/03 16:57:44 INFO Executor: Running task 0.0 in stage 21.0 (TID 17)

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

17/08/03 16:57:44 INFO Executor: Finished task 0.0 in stage 21.0 (TID 17). 1792 bytes result sent to driver

17/08/03 16:57:44 INFO TaskSetManager: Finished task 0.0 in stage 21.0 (TID 17) in 31 ms on localhost (1/1)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Removed TaskSet 21.0, whose tasks have all completed, from pool

17/08/03 16:57:44 INFO DAGScheduler: ShuffleMapStage 21 (sortBy at Session17Assignment2New.scala:212) finished in 0.031 s

17/08/03 16:57:44 INFO DAGScheduler: looking for newly runnable stages

17/08/03 16:57:44 INFO DAGScheduler: running: Set()

17/08/03 16:57:44 INFO DAGScheduler: waiting: Set(ResultStage 22)

17/08/03 16:57:44 INFO DAGScheduler: failed: Set()

17/08/03 16:57:44 INFO DAGScheduler: Submitting ResultStage 22 (MapPartitionsRDD[31] at map at Session17Assignment2New.scala:216), which has no missing parents

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_19 stored as values in memory (estimated size 3.7 KB, free 349.0 MB)

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_19\_piece0 stored as bytes in memory (estimated size 2.1 KB, free 349.0 MB)

17/08/03 16:57:44 INFO BlockManagerInfo: Added broadcast\_19\_piece0 in memory on 192.168.56.1:54675 (size: 2.1 KB, free: 349.2 MB)

(Andrew,history,grade-2,77,1)

(Andrew,history,grade-1,74,1)

(Andrew,maths,grade-1,57,2)

(Andrew,science,grade-3,70,2)

(John,history,grade-1,81,2)

(John,maths,grade-2,74,1)

(John,maths,grade-1,35,1)

(Lisa,history,grade-3,86,1)

(Lisa,history,grade-2,98,1)

(Lisa,science,grade-2,24,1)

(Lisa,science,grade-1,24,1)

(Mark,maths,grade-2,23,1)

(Mark,maths,grade-1,92,1)

(Mark,science,grade-2,12,1)

(Mark,science,grade-1,76,1)

(Mathew,history,grade-2,142,2)

(Mathew,science,grade-3,45,1)

(Mathew,science,grade-2,55,1)

17/08/03 16:57:44 INFO SparkContext: Created broadcast 19 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:44 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 22 (MapPartitionsRDD[31] at map at Session17Assignment2New.scala:216)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Adding task set 22.0 with 1 tasks

17/08/03 16:57:44 INFO TaskSetManager: Starting task 0.0 in stage 22.0 (TID 18, localhost, partition 0, ANY, 5141 bytes)

17/08/03 16:57:44 INFO Executor: Running task 0.0 in stage 22.0 (TID 18)

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

17/08/03 16:57:44 INFO Executor: Finished task 0.0 in stage 22.0 (TID 18). 1466 bytes result sent to driver

17/08/03 16:57:44 INFO TaskSetManager: Finished task 0.0 in stage 22.0 (TID 18) in 16 ms on localhost (1/1)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Removed TaskSet 22.0, whose tasks have all completed, from pool

17/08/03 16:57:44 INFO DAGScheduler: ResultStage 22 (foreach at Session17Assignment2New.scala:217) finished in 0.016 s

17/08/03 16:57:44 INFO DAGScheduler: Job 10 finished: foreach at Session17Assignment2New.scala:217, took 0.126341 s

<<<<---- Average score of students in each subject per grade ---->>>>

17/08/03 16:57:44 INFO SparkContext: Starting job: foreach at Session17Assignment2New.scala:222

17/08/03 16:57:44 INFO MapOutputTrackerMaster: Size of output statuses for shuffle 7 is 147 bytes

17/08/03 16:57:44 INFO MapOutputTrackerMaster: Size of output statuses for shuffle 6 is 147 bytes

17/08/03 16:57:44 INFO DAGScheduler: Got job 11 (foreach at Session17Assignment2New.scala:222) with 1 output partitions

17/08/03 16:57:44 INFO DAGScheduler: Final stage: ResultStage 25 (foreach at Session17Assignment2New.scala:222)

17/08/03 16:57:44 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 24)

17/08/03 16:57:44 INFO DAGScheduler: Missing parents: List()

17/08/03 16:57:44 INFO DAGScheduler: Submitting ResultStage 25 (MapPartitionsRDD[32] at map at Session17Assignment2New.scala:220), which has no missing parents

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_20 stored as values in memory (estimated size 3.8 KB, free 349.0 MB)

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_20\_piece0 stored as bytes in memory (estimated size 2.1 KB, free 349.0 MB)

17/08/03 16:57:44 INFO BlockManagerInfo: Added broadcast\_20\_piece0 in memory on 192.168.56.1:54675 (size: 2.1 KB, free: 349.2 MB)

17/08/03 16:57:44 INFO SparkContext: Created broadcast 20 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:44 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 25 (MapPartitionsRDD[32] at map at Session17Assignment2New.scala:220)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Adding task set 25.0 with 1 tasks

17/08/03 16:57:44 INFO TaskSetManager: Starting task 0.0 in stage 25.0 (TID 19, localhost, partition 0, ANY, 5141 bytes)

17/08/03 16:57:44 INFO Executor: Running task 0.0 in stage 25.0 (TID 19)

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

17/08/03 16:57:44 INFO Executor: Finished task 0.0 in stage 25.0 (TID 19). 1466 bytes result sent to driver

(Andrew,history,grade-2,77.0)

(Andrew,history,grade-1,74.0)

(Andrew,maths,grade-1,28.5)

(Andrew,science,grade-3,35.0)

(John,history,grade-1,40.5)

(John,maths,grade-2,74.0)

(John,maths,grade-1,35.0)

(Lisa,history,grade-3,86.0)

(Lisa,history,grade-2,98.0)

(Lisa,science,grade-2,24.0)

(Lisa,science,grade-1,24.0)

(Mark,maths,grade-2,23.0)

(Mark,maths,grade-1,92.0)

(Mark,science,grade-2,12.0)

(Mark,science,grade-1,76.0)

(Mathew,history,grade-2,71.0)

(Mathew,science,grade-3,45.0)

(Mathew,science,grade-2,55.0)

17/08/03 16:57:44 INFO TaskSetManager: Finished task 0.0 in stage 25.0 (TID 19) in 31 ms on localhost (1/1)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Removed TaskSet 25.0, whose tasks have all completed, from pool

17/08/03 16:57:44 INFO DAGScheduler: ResultStage 25 (foreach at Session17Assignment2New.scala:222) finished in 0.031 s

17/08/03 16:57:44 INFO DAGScheduler: Job 11 finished: foreach at Session17Assignment2New.scala:222, took 0.046343 s

17/08/03 16:57:44 INFO SparkContext: Starting job: foreach at Session17Assignment2New.scala:236

17/08/03 16:57:44 INFO DAGScheduler: Registering RDD 33 (groupBy at Session17Assignment2New.scala:227)

17/08/03 16:57:44 INFO DAGScheduler: Registering RDD 37 (sortBy at Session17Assignment2New.scala:232)

17/08/03 16:57:44 INFO DAGScheduler: Got job 12 (foreach at Session17Assignment2New.scala:236) with 1 output partitions

17/08/03 16:57:44 INFO DAGScheduler: Final stage: ResultStage 28 (foreach at Session17Assignment2New.scala:236)

17/08/03 16:57:44 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 27)

17/08/03 16:57:44 INFO DAGScheduler: Missing parents: List(ShuffleMapStage 27)

17/08/03 16:57:44 INFO DAGScheduler: Submitting ShuffleMapStage 26 (MapPartitionsRDD[33] at groupBy at Session17Assignment2New.scala:227), which has no missing parents

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_21 stored as values in memory (estimated size 5.0 KB, free 349.0 MB)

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_21\_piece0 stored as bytes in memory (estimated size 2.7 KB, free 349.0 MB)

17/08/03 16:57:44 INFO BlockManagerInfo: Added broadcast\_21\_piece0 in memory on 192.168.56.1:54675 (size: 2.7 KB, free: 349.1 MB)

17/08/03 16:57:44 INFO SparkContext: Created broadcast 21 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:44 INFO DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 26 (MapPartitionsRDD[33] at groupBy at Session17Assignment2New.scala:227)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Adding task set 26.0 with 1 tasks

17/08/03 16:57:44 INFO TaskSetManager: Starting task 0.0 in stage 26.0 (TID 20, localhost, partition 0, PROCESS\_LOCAL, 5429 bytes)

17/08/03 16:57:44 INFO Executor: Running task 0.0 in stage 26.0 (TID 20)

17/08/03 16:57:44 INFO HadoopRDD: Input split: file:/G:/ACADGILD/course material/Hadoop/Sessions/Session 17/Assignments/Assignment2/17.2\_Dataset.txt:0+622

17/08/03 16:57:44 INFO Executor: Finished task 0.0 in stage 26.0 (TID 20). 1242 bytes result sent to driver

17/08/03 16:57:44 INFO TaskSetManager: Finished task 0.0 in stage 26.0 (TID 20) in 47 ms on localhost (1/1)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Removed TaskSet 26.0, whose tasks have all completed, from pool

17/08/03 16:57:44 INFO DAGScheduler: ShuffleMapStage 26 (groupBy at Session17Assignment2New.scala:227) finished in 0.047 s

17/08/03 16:57:44 INFO DAGScheduler: looking for newly runnable stages

17/08/03 16:57:44 INFO DAGScheduler: running: Set()

17/08/03 16:57:44 INFO DAGScheduler: waiting: Set(ShuffleMapStage 27, ResultStage 28)

17/08/03 16:57:44 INFO DAGScheduler: failed: Set()

17/08/03 16:57:44 INFO DAGScheduler: Submitting ShuffleMapStage 27 (MapPartitionsRDD[37] at sortBy at Session17Assignment2New.scala:232), which has no missing parents

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_22 stored as values in memory (estimated size 6.9 KB, free 349.0 MB)

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_22\_piece0 stored as bytes in memory (estimated size 3.5 KB, free 349.0 MB)

17/08/03 16:57:44 INFO BlockManagerInfo: Added broadcast\_22\_piece0 in memory on 192.168.56.1:54675 (size: 3.5 KB, free: 349.1 MB)

17/08/03 16:57:44 INFO SparkContext: Created broadcast 22 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:44 INFO DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 27 (MapPartitionsRDD[37] at sortBy at Session17Assignment2New.scala:232)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Adding task set 27.0 with 1 tasks

17/08/03 16:57:44 INFO TaskSetManager: Starting task 0.0 in stage 27.0 (TID 21, localhost, partition 0, ANY, 5130 bytes)

17/08/03 16:57:44 INFO Executor: Running task 0.0 in stage 27.0 (TID 21)

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

17/08/03 16:57:44 INFO Executor: Finished task 0.0 in stage 27.0 (TID 21). 1792 bytes result sent to driver

17/08/03 16:57:44 INFO TaskSetManager: Finished task 0.0 in stage 27.0 (TID 21) in 31 ms on localhost (1/1)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Removed TaskSet 27.0, whose tasks have all completed, from pool

17/08/03 16:57:44 INFO DAGScheduler: ShuffleMapStage 27 (sortBy at Session17Assignment2New.scala:232) finished in 0.047 s

17/08/03 16:57:44 INFO DAGScheduler: looking for newly runnable stages

17/08/03 16:57:44 INFO DAGScheduler: running: Set()

17/08/03 16:57:44 INFO DAGScheduler: waiting: Set(ResultStage 28)

17/08/03 16:57:44 INFO DAGScheduler: failed: Set()

17/08/03 16:57:44 INFO DAGScheduler: Submitting ResultStage 28 (MapPartitionsRDD[40] at map at Session17Assignment2New.scala:235), which has no missing parents

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_23 stored as values in memory (estimated size 3.6 KB, free 349.0 MB)

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_23\_piece0 stored as bytes in memory (estimated size 2.1 KB, free 349.0 MB)

17/08/03 16:57:44 INFO BlockManagerInfo: Added broadcast\_23\_piece0 in memory on 192.168.56.1:54675 (size: 2.1 KB, free: 349.1 MB)

17/08/03 16:57:44 INFO SparkContext: Created broadcast 23 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:44 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 28 (MapPartitionsRDD[40] at map at Session17Assignment2New.scala:235)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Adding task set 28.0 with 1 tasks

17/08/03 16:57:44 INFO TaskSetManager: Starting task 0.0 in stage 28.0 (TID 22, localhost, partition 0, ANY, 5141 bytes)

17/08/03 16:57:44 INFO Executor: Running task 0.0 in stage 28.0 (TID 22)

(Andrew,grade-2,77,1)

(John,grade-2,74,1)

(Lisa,grade-2,122,2)

(Mark,grade-2,35,2)

(Mathew,grade-2,197,3)

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

17/08/03 16:57:44 INFO Executor: Finished task 0.0 in stage 28.0 (TID 22). 1474 bytes result sent to driver

17/08/03 16:57:44 INFO TaskSetManager: Finished task 0.0 in stage 28.0 (TID 22) in 15 ms on localhost (1/1)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Removed TaskSet 28.0, whose tasks have all completed, from pool

17/08/03 16:57:44 INFO DAGScheduler: ResultStage 28 (foreach at Session17Assignment2New.scala:236) finished in 0.015 s

17/08/03 16:57:44 INFO DAGScheduler: Job 12 finished: foreach at Session17Assignment2New.scala:236, took 0.170642 s

<<<<---- Students having grade-2 and average score greater than 50 ---->>>>

17/08/03 16:57:44 INFO SparkContext: Starting job: foreach at Session17Assignment2New.scala:242

17/08/03 16:57:44 INFO MapOutputTrackerMaster: Size of output statuses for shuffle 9 is 147 bytes

17/08/03 16:57:44 INFO MapOutputTrackerMaster: Size of output statuses for shuffle 8 is 147 bytes

17/08/03 16:57:44 INFO DAGScheduler: Got job 13 (foreach at Session17Assignment2New.scala:242) with 1 output partitions

17/08/03 16:57:44 INFO DAGScheduler: Final stage: ResultStage 31 (foreach at Session17Assignment2New.scala:242)

17/08/03 16:57:44 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 30)

17/08/03 16:57:44 INFO DAGScheduler: Missing parents: List()

17/08/03 16:57:44 INFO DAGScheduler: Submitting ResultStage 31 (MapPartitionsRDD[42] at filter at Session17Assignment2New.scala:239), which has no missing parents

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_24 stored as values in memory (estimated size 4.0 KB, free 348.9 MB)

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_24\_piece0 stored as bytes in memory (estimated size 2.2 KB, free 348.9 MB)

17/08/03 16:57:44 INFO BlockManagerInfo: Added broadcast\_24\_piece0 in memory on 192.168.56.1:54675 (size: 2.2 KB, free: 349.1 MB)

17/08/03 16:57:44 INFO SparkContext: Created broadcast 24 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:44 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 31 (MapPartitionsRDD[42] at filter at Session17Assignment2New.scala:239)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Adding task set 31.0 with 1 tasks

17/08/03 16:57:44 INFO TaskSetManager: Starting task 0.0 in stage 31.0 (TID 23, localhost, partition 0, ANY, 5141 bytes)

17/08/03 16:57:44 INFO Executor: Running task 0.0 in stage 31.0 (TID 23)

(Andrew,grade-2,77.0)

(John,grade-2,74.0)

(Lisa,grade-2,61.0)

(Mathew,grade-2,65.666664)

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

17/08/03 16:57:44 INFO Executor: Finished task 0.0 in stage 31.0 (TID 23). 1474 bytes result sent to driver

17/08/03 16:57:44 INFO TaskSetManager: Finished task 0.0 in stage 31.0 (TID 23) in 15 ms on localhost (1/1)

17/08/03 16:57:44 INFO DAGScheduler: ResultStage 31 (foreach at Session17Assignment2New.scala:242) finished in 0.015 s

17/08/03 16:57:44 INFO TaskSchedulerImpl: Removed TaskSet 31.0, whose tasks have all completed, from pool

17/08/03 16:57:44 INFO DAGScheduler: Job 13 finished: foreach at Session17Assignment2New.scala:242, took 0.027774 s

<<<<<<------ Average Score per student\_name across all grades ------>>>>>>

17/08/03 16:57:44 INFO SparkContext: Starting job: foreach at Session17Assignment2New.scala:273

17/08/03 16:57:44 INFO DAGScheduler: Registering RDD 43 (groupBy at Session17Assignment2New.scala:262)

17/08/03 16:57:44 INFO DAGScheduler: Registering RDD 46 (sortBy at Session17Assignment2New.scala:265)

17/08/03 16:57:44 INFO DAGScheduler: Got job 14 (foreach at Session17Assignment2New.scala:273) with 1 output partitions

17/08/03 16:57:44 INFO DAGScheduler: Final stage: ResultStage 34 (foreach at Session17Assignment2New.scala:273)

17/08/03 16:57:44 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 33)

17/08/03 16:57:44 INFO DAGScheduler: Missing parents: List(ShuffleMapStage 33)

17/08/03 16:57:44 INFO DAGScheduler: Submitting ShuffleMapStage 32 (MapPartitionsRDD[43] at groupBy at Session17Assignment2New.scala:262), which has no missing parents

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_25 stored as values in memory (estimated size 5.0 KB, free 348.9 MB)

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_25\_piece0 stored as bytes in memory (estimated size 2.7 KB, free 348.9 MB)

17/08/03 16:57:44 INFO BlockManagerInfo: Added broadcast\_25\_piece0 in memory on 192.168.56.1:54675 (size: 2.7 KB, free: 349.1 MB)

17/08/03 16:57:44 INFO SparkContext: Created broadcast 25 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:44 INFO DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 32 (MapPartitionsRDD[43] at groupBy at Session17Assignment2New.scala:262)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Adding task set 32.0 with 1 tasks

17/08/03 16:57:44 INFO TaskSetManager: Starting task 0.0 in stage 32.0 (TID 24, localhost, partition 0, PROCESS\_LOCAL, 5430 bytes)

17/08/03 16:57:44 INFO Executor: Running task 0.0 in stage 32.0 (TID 24)

17/08/03 16:57:44 INFO HadoopRDD: Input split: file:/G:/ACADGILD/course material/Hadoop/Sessions/Session 17/Assignments/Assignment2/17.2\_Dataset.txt:0+622

17/08/03 16:57:44 INFO Executor: Finished task 0.0 in stage 32.0 (TID 24). 1155 bytes result sent to driver

17/08/03 16:57:44 INFO TaskSetManager: Finished task 0.0 in stage 32.0 (TID 24) in 16 ms on localhost (1/1)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Removed TaskSet 32.0, whose tasks have all completed, from pool

17/08/03 16:57:44 INFO DAGScheduler: ShuffleMapStage 32 (groupBy at Session17Assignment2New.scala:262) finished in 0.032 s

17/08/03 16:57:44 INFO DAGScheduler: looking for newly runnable stages

17/08/03 16:57:44 INFO DAGScheduler: running: Set()

17/08/03 16:57:44 INFO DAGScheduler: waiting: Set(ShuffleMapStage 33, ResultStage 34)

17/08/03 16:57:44 INFO DAGScheduler: failed: Set()

17/08/03 16:57:44 INFO DAGScheduler: Submitting ShuffleMapStage 33 (MapPartitionsRDD[46] at sortBy at Session17Assignment2New.scala:265), which has no missing parents

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_26 stored as values in memory (estimated size 6.6 KB, free 348.9 MB)

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_26\_piece0 stored as bytes in memory (estimated size 3.4 KB, free 348.9 MB)

17/08/03 16:57:44 INFO BlockManagerInfo: Added broadcast\_26\_piece0 in memory on 192.168.56.1:54675 (size: 3.4 KB, free: 349.1 MB)

17/08/03 16:57:44 INFO SparkContext: Created broadcast 26 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:44 INFO DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 33 (MapPartitionsRDD[46] at sortBy at Session17Assignment2New.scala:265)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Adding task set 33.0 with 1 tasks

17/08/03 16:57:44 INFO TaskSetManager: Starting task 0.0 in stage 33.0 (TID 25, localhost, partition 0, ANY, 5131 bytes)

17/08/03 16:57:44 INFO Executor: Running task 0.0 in stage 33.0 (TID 25)

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

17/08/03 16:57:44 INFO Executor: Finished task 0.0 in stage 33.0 (TID 25). 1792 bytes result sent to driver

17/08/03 16:57:44 INFO TaskSetManager: Finished task 0.0 in stage 33.0 (TID 25) in 47 ms on localhost (1/1)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Removed TaskSet 33.0, whose tasks have all completed, from pool

17/08/03 16:57:44 INFO DAGScheduler: ShuffleMapStage 33 (sortBy at Session17Assignment2New.scala:265) finished in 0.047 s

17/08/03 16:57:44 INFO DAGScheduler: looking for newly runnable stages

17/08/03 16:57:44 INFO DAGScheduler: running: Set()

17/08/03 16:57:44 INFO DAGScheduler: waiting: Set(ResultStage 34)

17/08/03 16:57:44 INFO DAGScheduler: failed: Set()

17/08/03 16:57:44 INFO DAGScheduler: Submitting ResultStage 34 (MapPartitionsRDD[50] at map at Session17Assignment2New.scala:271), which has no missing parents

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_27 stored as values in memory (estimated size 3.7 KB, free 348.9 MB)

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_27\_piece0 stored as bytes in memory (estimated size 2.1 KB, free 348.9 MB)

17/08/03 16:57:44 INFO BlockManagerInfo: Added broadcast\_27\_piece0 in memory on 192.168.56.1:54675 (size: 2.1 KB, free: 349.1 MB)

17/08/03 16:57:44 INFO SparkContext: Created broadcast 27 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:44 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 34 (MapPartitionsRDD[50] at map at Session17Assignment2New.scala:271)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Adding task set 34.0 with 1 tasks

17/08/03 16:57:44 INFO TaskSetManager: Starting task 0.0 in stage 34.0 (TID 26, localhost, partition 0, ANY, 5142 bytes)

17/08/03 16:57:44 INFO Executor: Running task 0.0 in stage 34.0 (TID 26)

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

17/08/03 16:57:44 INFO Executor: Finished task 0.0 in stage 34.0 (TID 26). 1466 bytes result sent to driver

(Andrew,46.333332)

(John,47.5)

(Lisa,58.0)

(Mark,50.75)

(Mathew,60.5)

17/08/03 16:57:44 INFO TaskSetManager: Finished task 0.0 in stage 34.0 (TID 26) in 16 ms on localhost (1/1)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Removed TaskSet 34.0, whose tasks have all completed, from pool

17/08/03 16:57:44 INFO DAGScheduler: ResultStage 34 (foreach at Session17Assignment2New.scala:273) finished in 0.016 s

17/08/03 16:57:44 INFO DAGScheduler: Job 14 finished: foreach at Session17Assignment2New.scala:273, took 0.127713 s

<<<<<<------ Average score per student\_name per grade ------>>>>>>

17/08/03 16:57:44 INFO SparkContext: Starting job: foreach at Session17Assignment2New.scala:284

17/08/03 16:57:44 INFO MapOutputTrackerMaster: Size of output statuses for shuffle 3 is 147 bytes

17/08/03 16:57:44 INFO MapOutputTrackerMaster: Size of output statuses for shuffle 2 is 147 bytes

17/08/03 16:57:44 INFO DAGScheduler: Got job 15 (foreach at Session17Assignment2New.scala:284) with 1 output partitions

17/08/03 16:57:44 INFO DAGScheduler: Final stage: ResultStage 37 (foreach at Session17Assignment2New.scala:284)

17/08/03 16:57:44 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 36)

17/08/03 16:57:44 INFO DAGScheduler: Missing parents: List()

17/08/03 16:57:44 INFO DAGScheduler: Submitting ResultStage 37 (MapPartitionsRDD[51] at map at Session17Assignment2New.scala:282), which has no missing parents

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_28 stored as values in memory (estimated size 3.9 KB, free 348.9 MB)

(Andrew,77.0)

(Andrew,43.666668)

(Andrew,35.0)

(John,38.666668)

(John,74.0)

(Lisa,24.0)

(Lisa,61.0)

(Lisa,86.0)

(Mark,17.5)

(Mark,84.0)

(Mathew,45.0)

(Mathew,65.666664)

17/08/03 16:57:44 INFO MemoryStore: Block broadcast\_28\_piece0 stored as bytes in memory (estimated size 2.1 KB, free 348.9 MB)

17/08/03 16:57:44 INFO BlockManagerInfo: Added broadcast\_28\_piece0 in memory on 192.168.56.1:54675 (size: 2.1 KB, free: 349.1 MB)

17/08/03 16:57:44 INFO SparkContext: Created broadcast 28 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:44 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 37 (MapPartitionsRDD[51] at map at Session17Assignment2New.scala:282)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Adding task set 37.0 with 1 tasks

17/08/03 16:57:44 INFO TaskSetManager: Starting task 0.0 in stage 37.0 (TID 27, localhost, partition 0, ANY, 5142 bytes)

17/08/03 16:57:44 INFO Executor: Running task 0.0 in stage 37.0 (TID 27)

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:44 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

17/08/03 16:57:44 INFO Executor: Finished task 0.0 in stage 37.0 (TID 27). 1466 bytes result sent to driver

17/08/03 16:57:44 INFO TaskSetManager: Finished task 0.0 in stage 37.0 (TID 27) in 15 ms on localhost (1/1)

17/08/03 16:57:44 INFO TaskSchedulerImpl: Removed TaskSet 37.0, whose tasks have all completed, from pool

17/08/03 16:57:44 INFO DAGScheduler: ResultStage 37 (foreach at Session17Assignment2New.scala:284) finished in 0.015 s

17/08/03 16:57:44 INFO DAGScheduler: Job 15 finished: foreach at Session17Assignment2New.scala:284, took 0.029649 s

17/08/03 16:57:45 INFO BlockManagerInfo: Removed broadcast\_18\_piece0 on 192.168.56.1:54675 in memory (size: 3.5 KB, free: 349.1 MB)

17/08/03 16:57:45 INFO SparkContext: Starting job: isEmpty at Session17Assignment2New.scala:291

17/08/03 16:57:45 INFO MapOutputTrackerMaster: Size of output statuses for shuffle 11 is 147 bytes

17/08/03 16:57:45 INFO MapOutputTrackerMaster: Size of output statuses for shuffle 10 is 147 bytes

17/08/03 16:57:45 INFO DAGScheduler: Registering RDD 52 (intersection at Session17Assignment2New.scala:288)

17/08/03 16:57:45 INFO DAGScheduler: Registering RDD 53 (intersection at Session17Assignment2New.scala:288)

17/08/03 16:57:45 INFO DAGScheduler: Got job 16 (isEmpty at Session17Assignment2New.scala:291) with 1 output partitions

17/08/03 16:57:45 INFO DAGScheduler: Final stage: ResultStage 44 (isEmpty at Session17Assignment2New.scala:291)

17/08/03 16:57:45 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 43, ShuffleMapStage 40)

17/08/03 16:57:45 INFO DAGScheduler: Missing parents: List(ShuffleMapStage 43, ShuffleMapStage 40)

17/08/03 16:57:45 INFO DAGScheduler: Submitting ShuffleMapStage 40 (MapPartitionsRDD[52] at intersection at Session17Assignment2New.scala:288), which has no missing parents

17/08/03 16:57:45 INFO MemoryStore: Block broadcast\_29 stored as values in memory (estimated size 3.8 KB, free 348.9 MB)

17/08/03 16:57:45 INFO MemoryStore: Block broadcast\_29\_piece0 stored as bytes in memory (estimated size 2.1 KB, free 348.9 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Added broadcast\_29\_piece0 in memory on 192.168.56.1:54675 (size: 2.1 KB, free: 349.1 MB)

17/08/03 16:57:45 INFO SparkContext: Created broadcast 29 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:45 INFO DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 40 (MapPartitionsRDD[52] at intersection at Session17Assignment2New.scala:288)

17/08/03 16:57:45 INFO TaskSchedulerImpl: Adding task set 40.0 with 1 tasks

17/08/03 16:57:45 INFO DAGScheduler: Submitting ShuffleMapStage 43 (MapPartitionsRDD[53] at intersection at Session17Assignment2New.scala:288), which has no missing parents

17/08/03 16:57:45 INFO MemoryStore: Block broadcast\_30 stored as values in memory (estimated size 4.0 KB, free 348.9 MB)

17/08/03 16:57:45 INFO MemoryStore: Block broadcast\_30\_piece0 stored as bytes in memory (estimated size 2.2 KB, free 348.9 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Removed broadcast\_19\_piece0 on 192.168.56.1:54675 in memory (size: 2.1 KB, free: 349.1 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Added broadcast\_30\_piece0 in memory on 192.168.56.1:54675 (size: 2.2 KB, free: 349.1 MB)

17/08/03 16:57:45 INFO SparkContext: Created broadcast 30 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:45 INFO DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 43 (MapPartitionsRDD[53] at intersection at Session17Assignment2New.scala:288)

17/08/03 16:57:45 INFO TaskSchedulerImpl: Adding task set 43.0 with 1 tasks

17/08/03 16:57:45 INFO TaskSetManager: Starting task 0.0 in stage 40.0 (TID 28, localhost, partition 0, ANY, 5131 bytes)

17/08/03 16:57:45 INFO Executor: Running task 0.0 in stage 40.0 (TID 28)

17/08/03 16:57:45 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:45 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

17/08/03 16:57:45 INFO BlockManagerInfo: Removed broadcast\_20\_piece0 on 192.168.56.1:54675 in memory (size: 2.1 KB, free: 349.1 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Removed broadcast\_21\_piece0 on 192.168.56.1:54675 in memory (size: 2.7 KB, free: 349.1 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Removed broadcast\_22\_piece0 on 192.168.56.1:54675 in memory (size: 3.5 KB, free: 349.1 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Removed broadcast\_23\_piece0 on 192.168.56.1:54675 in memory (size: 2.1 KB, free: 349.1 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Removed broadcast\_24\_piece0 on 192.168.56.1:54675 in memory (size: 2.2 KB, free: 349.1 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Removed broadcast\_25\_piece0 on 192.168.56.1:54675 in memory (size: 2.7 KB, free: 349.1 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Removed broadcast\_26\_piece0 on 192.168.56.1:54675 in memory (size: 3.4 KB, free: 349.2 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Removed broadcast\_27\_piece0 on 192.168.56.1:54675 in memory (size: 2.1 KB, free: 349.2 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Removed broadcast\_28\_piece0 on 192.168.56.1:54675 in memory (size: 2.1 KB, free: 349.2 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Removed broadcast\_6\_piece0 on 192.168.56.1:54675 in memory (size: 2.7 KB, free: 349.2 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Removed broadcast\_7\_piece0 on 192.168.56.1:54675 in memory (size: 3.3 KB, free: 349.2 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Removed broadcast\_8\_piece0 on 192.168.56.1:54675 in memory (size: 2024.0 B, free: 349.2 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Removed broadcast\_9\_piece0 on 192.168.56.1:54675 in memory (size: 2.7 KB, free: 349.2 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Removed broadcast\_10\_piece0 on 192.168.56.1:54675 in memory (size: 3.4 KB, free: 349.2 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Removed broadcast\_11\_piece0 on 192.168.56.1:54675 in memory (size: 2.0 KB, free: 349.2 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Removed broadcast\_12\_piece0 on 192.168.56.1:54675 in memory (size: 2.1 KB, free: 349.2 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Removed broadcast\_13\_piece0 on 192.168.56.1:54675 in memory (size: 2.7 KB, free: 349.2 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Removed broadcast\_14\_piece0 on 192.168.56.1:54675 in memory (size: 3.4 KB, free: 349.2 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Removed broadcast\_15\_piece0 on 192.168.56.1:54675 in memory (size: 2.1 KB, free: 349.2 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Removed broadcast\_16\_piece0 on 192.168.56.1:54675 in memory (size: 2.1 KB, free: 349.2 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Removed broadcast\_17\_piece0 on 192.168.56.1:54675 in memory (size: 2.7 KB, free: 349.2 MB)

17/08/03 16:57:45 INFO Executor: Finished task 0.0 in stage 40.0 (TID 28). 1792 bytes result sent to driver

17/08/03 16:57:45 INFO TaskSetManager: Starting task 0.0 in stage 43.0 (TID 29, localhost, partition 0, ANY, 5131 bytes)

17/08/03 16:57:45 INFO TaskSetManager: Finished task 0.0 in stage 40.0 (TID 28) in 578 ms on localhost (1/1)

17/08/03 16:57:45 INFO Executor: Running task 0.0 in stage 43.0 (TID 29)

17/08/03 16:57:45 INFO TaskSchedulerImpl: Removed TaskSet 40.0, whose tasks have all completed, from pool

17/08/03 16:57:45 INFO DAGScheduler: ShuffleMapStage 40 (intersection at Session17Assignment2New.scala:288) finished in 0.578 s

17/08/03 16:57:45 INFO DAGScheduler: looking for newly runnable stages

17/08/03 16:57:45 INFO DAGScheduler: running: Set(ShuffleMapStage 43)

17/08/03 16:57:45 INFO DAGScheduler: waiting: Set(ResultStage 44)

17/08/03 16:57:45 INFO DAGScheduler: failed: Set()

17/08/03 16:57:45 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:45 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

17/08/03 16:57:45 INFO Executor: Finished task 0.0 in stage 43.0 (TID 29). 1800 bytes result sent to driver

17/08/03 16:57:45 INFO TaskSetManager: Finished task 0.0 in stage 43.0 (TID 29) in 16 ms on localhost (1/1)

17/08/03 16:57:45 INFO TaskSchedulerImpl: Removed TaskSet 43.0, whose tasks have all completed, from pool

17/08/03 16:57:45 INFO DAGScheduler: ShuffleMapStage 43 (intersection at Session17Assignment2New.scala:288) finished in 0.578 s

17/08/03 16:57:45 INFO DAGScheduler: looking for newly runnable stages

17/08/03 16:57:45 INFO DAGScheduler: running: Set()

17/08/03 16:57:45 INFO DAGScheduler: waiting: Set(ResultStage 44)

17/08/03 16:57:45 INFO DAGScheduler: failed: Set()

17/08/03 16:57:45 INFO DAGScheduler: Submitting ResultStage 44 (MapPartitionsRDD[57] at intersection at Session17Assignment2New.scala:288), which has no missing parents

17/08/03 16:57:45 INFO MemoryStore: Block broadcast\_31 stored as values in memory (estimated size 3.4 KB, free 349.1 MB)

17/08/03 16:57:45 INFO MemoryStore: Block broadcast\_31\_piece0 stored as bytes in memory (estimated size 1970.0 B, free 349.1 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Added broadcast\_31\_piece0 in memory on 192.168.56.1:54675 (size: 1970.0 B, free: 349.2 MB)

17/08/03 16:57:45 INFO SparkContext: Created broadcast 31 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:45 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 44 (MapPartitionsRDD[57] at intersection at Session17Assignment2New.scala:288)

17/08/03 16:57:45 INFO TaskSchedulerImpl: Adding task set 44.0 with 1 tasks

17/08/03 16:57:45 INFO TaskSetManager: Starting task 0.0 in stage 44.0 (TID 30, localhost, partition 0, PROCESS\_LOCAL, 5205 bytes)

17/08/03 16:57:45 INFO Executor: Running task 0.0 in stage 44.0 (TID 30)

17/08/03 16:57:45 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:45 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

17/08/03 16:57:45 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:45 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

<<<<<<------ No such record found -------->>>>>>

17/08/03 16:57:45 INFO Executor: Finished task 0.0 in stage 44.0 (TID 30). 1462 bytes result sent to driver

17/08/03 16:57:45 INFO TaskSetManager: Finished task 0.0 in stage 44.0 (TID 30) in 16 ms on localhost (1/1)

17/08/03 16:57:45 INFO TaskSchedulerImpl: Removed TaskSet 44.0, whose tasks have all completed, from pool

17/08/03 16:57:45 INFO DAGScheduler: ResultStage 44 (isEmpty at Session17Assignment2New.scala:291) finished in 0.031 s

17/08/03 16:57:45 INFO DAGScheduler: Job 16 finished: isEmpty at Session17Assignment2New.scala:291, took 0.641620 s

17/08/03 16:57:45 INFO SparkContext: Starting job: foreach at Session17Assignment2New.scala:301

17/08/03 16:57:45 INFO MapOutputTrackerMaster: Size of output statuses for shuffle 5 is 147 bytes

17/08/03 16:57:45 INFO MapOutputTrackerMaster: Size of output statuses for shuffle 4 is 147 bytes

17/08/03 16:57:45 INFO DAGScheduler: Got job 17 (foreach at Session17Assignment2New.scala:301) with 1 output partitions

17/08/03 16:57:45 INFO DAGScheduler: Final stage: ResultStage 47 (foreach at Session17Assignment2New.scala:301)

17/08/03 16:57:45 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 46)

17/08/03 16:57:45 INFO DAGScheduler: Missing parents: List()

17/08/03 16:57:45 INFO DAGScheduler: Submitting ResultStage 47 (MapPartitionsRDD[24] at map at Session17Assignment2New.scala:202), which has no missing parents

17/08/03 16:57:45 INFO MemoryStore: Block broadcast\_32 stored as values in memory (estimated size 3.7 KB, free 349.1 MB)

17/08/03 16:57:45 INFO MemoryStore: Block broadcast\_32\_piece0 stored as bytes in memory (estimated size 2.1 KB, free 349.1 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Added broadcast\_32\_piece0 in memory on 192.168.56.1:54675 (size: 2.1 KB, free: 349.2 MB)

17/08/03 16:57:45 INFO SparkContext: Created broadcast 32 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:45 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 47 (MapPartitionsRDD[24] at map at Session17Assignment2New.scala:202)

17/08/03 16:57:45 INFO TaskSchedulerImpl: Adding task set 47.0 with 1 tasks

(Andrew,science,35.0)

(Andrew,maths,28.5)

(Andrew,history,75.5)

(John,maths,54.5)

(John,history,40.5)

(Lisa,history,92.0)

(Lisa,science,24.0)

(Mark,maths,57.5)

(Mark,science,44.0)

(Mathew,science,50.0)

(Mathew,history,71.0)

17/08/03 16:57:45 INFO TaskSetManager: Starting task 0.0 in stage 47.0 (TID 31, localhost, partition 0, ANY, 5142 bytes)

17/08/03 16:57:45 INFO Executor: Running task 0.0 in stage 47.0 (TID 31)

17/08/03 16:57:45 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:45 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

17/08/03 16:57:45 INFO Executor: Finished task 0.0 in stage 47.0 (TID 31). 1387 bytes result sent to driver

17/08/03 16:57:45 INFO TaskSetManager: Finished task 0.0 in stage 47.0 (TID 31) in 16 ms on localhost (1/1)

17/08/03 16:57:45 INFO TaskSchedulerImpl: Removed TaskSet 47.0, whose tasks have all completed, from pool

17/08/03 16:57:45 INFO DAGScheduler: ResultStage 47 (foreach at Session17Assignment2New.scala:301) finished in 0.016 s

17/08/03 16:57:45 INFO DAGScheduler: Job 17 finished: foreach at Session17Assignment2New.scala:301, took 0.017330 s

(Andrew,grade-2,77.0)

(Andrew,grade-1,43.666668)

(Andrew,grade-3,35.0)

(John,grade-1,38.666668)

(John,grade-2,74.0)

(Lisa,grade-1,24.0)

(Lisa,grade-2,61.0)

(Lisa,grade-3,86.0)

(Mark,grade-2,17.5)

(Mark,grade-1,84.0)

(Mathew,grade-3,45.0)

(Mathew,grade-2,65.666664)

17/08/03 16:57:45 INFO SparkContext: Starting job: foreach at Session17Assignment2New.scala:305

17/08/03 16:57:45 INFO MapOutputTrackerMaster: Size of output statuses for shuffle 3 is 147 bytes

17/08/03 16:57:45 INFO MapOutputTrackerMaster: Size of output statuses for shuffle 2 is 147 bytes

17/08/03 16:57:45 INFO DAGScheduler: Got job 18 (foreach at Session17Assignment2New.scala:305) with 1 output partitions

17/08/03 16:57:45 INFO DAGScheduler: Final stage: ResultStage 50 (foreach at Session17Assignment2New.scala:305)

17/08/03 16:57:45 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 49)

17/08/03 16:57:45 INFO DAGScheduler: Missing parents: List()

17/08/03 16:57:45 INFO DAGScheduler: Submitting ResultStage 50 (MapPartitionsRDD[16] at map at Session17Assignment2New.scala:183), which has no missing parents

17/08/03 16:57:45 INFO MemoryStore: Block broadcast\_33 stored as values in memory (estimated size 3.7 KB, free 349.1 MB)

17/08/03 16:57:45 INFO MemoryStore: Block broadcast\_33\_piece0 stored as bytes in memory (estimated size 2.1 KB, free 349.1 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Added broadcast\_33\_piece0 in memory on 192.168.56.1:54675 (size: 2.1 KB, free: 349.2 MB)

17/08/03 16:57:45 INFO SparkContext: Created broadcast 33 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:45 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 50 (MapPartitionsRDD[16] at map at Session17Assignment2New.scala:183)

17/08/03 16:57:45 INFO TaskSchedulerImpl: Adding task set 50.0 with 1 tasks

17/08/03 16:57:45 INFO TaskSetManager: Starting task 0.0 in stage 50.0 (TID 32, localhost, partition 0, ANY, 5142 bytes)

17/08/03 16:57:45 INFO Executor: Running task 0.0 in stage 50.0 (TID 32)

17/08/03 16:57:45 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:45 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

17/08/03 16:57:45 INFO Executor: Finished task 0.0 in stage 50.0 (TID 32). 1387 bytes result sent to driver

17/08/03 16:57:45 INFO TaskSetManager: Finished task 0.0 in stage 50.0 (TID 32) in 16 ms on localhost (1/1)

17/08/03 16:57:45 INFO TaskSchedulerImpl: Removed TaskSet 50.0, whose tasks have all completed, from pool

17/08/03 16:57:45 INFO DAGScheduler: ResultStage 50 (foreach at Session17Assignment2New.scala:305) finished in 0.016 s

17/08/03 16:57:45 INFO DAGScheduler: Job 18 finished: foreach at Session17Assignment2New.scala:305, took 0.018033 s

17/08/03 16:57:45 INFO SparkContext: Starting job: isEmpty at Session17Assignment2New.scala:314

17/08/03 16:57:45 INFO DAGScheduler: Registering RDD 60 (intersection at Session17Assignment2New.scala:312)

17/08/03 16:57:45 INFO DAGScheduler: Registering RDD 61 (intersection at Session17Assignment2New.scala:312)

17/08/03 16:57:45 INFO DAGScheduler: Got job 19 (isEmpty at Session17Assignment2New.scala:314) with 1 output partitions

17/08/03 16:57:45 INFO DAGScheduler: Final stage: ResultStage 57 (isEmpty at Session17Assignment2New.scala:314)

17/08/03 16:57:45 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 56, ShuffleMapStage 53)

17/08/03 16:57:45 INFO DAGScheduler: Missing parents: List(ShuffleMapStage 56, ShuffleMapStage 53)

17/08/03 16:57:45 INFO DAGScheduler: Submitting ShuffleMapStage 53 (MapPartitionsRDD[60] at intersection at Session17Assignment2New.scala:312), which has no missing parents

17/08/03 16:57:45 INFO MemoryStore: Block broadcast\_34 stored as values in memory (estimated size 4.0 KB, free 349.1 MB)

17/08/03 16:57:45 INFO MemoryStore: Block broadcast\_34\_piece0 stored as bytes in memory (estimated size 2.2 KB, free 349.1 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Added broadcast\_34\_piece0 in memory on 192.168.56.1:54675 (size: 2.2 KB, free: 349.2 MB)

17/08/03 16:57:45 INFO SparkContext: Created broadcast 34 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:45 INFO DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 53 (MapPartitionsRDD[60] at intersection at Session17Assignment2New.scala:312)

17/08/03 16:57:45 INFO TaskSchedulerImpl: Adding task set 53.0 with 1 tasks

17/08/03 16:57:45 INFO DAGScheduler: Submitting ShuffleMapStage 56 (MapPartitionsRDD[61] at intersection at Session17Assignment2New.scala:312), which has no missing parents

17/08/03 16:57:45 INFO TaskSetManager: Starting task 0.0 in stage 53.0 (TID 33, localhost, partition 0, ANY, 5131 bytes)

17/08/03 16:57:45 INFO Executor: Running task 0.0 in stage 53.0 (TID 33)

17/08/03 16:57:45 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:45 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

17/08/03 16:57:45 INFO MemoryStore: Block broadcast\_35 stored as values in memory (estimated size 4.0 KB, free 349.0 MB)

17/08/03 16:57:45 INFO MemoryStore: Block broadcast\_35\_piece0 stored as bytes in memory (estimated size 2.2 KB, free 349.0 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Added broadcast\_35\_piece0 in memory on 192.168.56.1:54675 (size: 2.2 KB, free: 349.2 MB)

17/08/03 16:57:45 INFO SparkContext: Created broadcast 35 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:45 INFO DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 56 (MapPartitionsRDD[61] at intersection at Session17Assignment2New.scala:312)

17/08/03 16:57:45 INFO TaskSchedulerImpl: Adding task set 56.0 with 1 tasks

17/08/03 16:57:45 INFO Executor: Finished task 0.0 in stage 53.0 (TID 33). 1792 bytes result sent to driver

17/08/03 16:57:45 INFO TaskSetManager: Starting task 0.0 in stage 56.0 (TID 34, localhost, partition 0, ANY, 5131 bytes)

17/08/03 16:57:45 INFO Executor: Running task 0.0 in stage 56.0 (TID 34)

17/08/03 16:57:45 INFO TaskSetManager: Finished task 0.0 in stage 53.0 (TID 33) in 172 ms on localhost (1/1)

17/08/03 16:57:45 INFO TaskSchedulerImpl: Removed TaskSet 53.0, whose tasks have all completed, from pool

17/08/03 16:57:45 INFO DAGScheduler: ShuffleMapStage 53 (intersection at Session17Assignment2New.scala:312) finished in 0.172 s

17/08/03 16:57:45 INFO DAGScheduler: looking for newly runnable stages

17/08/03 16:57:45 INFO DAGScheduler: running: Set(ShuffleMapStage 56)

17/08/03 16:57:45 INFO DAGScheduler: waiting: Set(ResultStage 57)

17/08/03 16:57:45 INFO DAGScheduler: failed: Set()

17/08/03 16:57:45 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:45 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

17/08/03 16:57:45 INFO Executor: Finished task 0.0 in stage 56.0 (TID 34). 1792 bytes result sent to driver

17/08/03 16:57:45 INFO TaskSetManager: Finished task 0.0 in stage 56.0 (TID 34) in 16 ms on localhost (1/1)

17/08/03 16:57:45 INFO TaskSchedulerImpl: Removed TaskSet 56.0, whose tasks have all completed, from pool

17/08/03 16:57:45 INFO DAGScheduler: ShuffleMapStage 56 (intersection at Session17Assignment2New.scala:312) finished in 0.172 s

17/08/03 16:57:45 INFO DAGScheduler: looking for newly runnable stages

17/08/03 16:57:45 INFO DAGScheduler: running: Set()

17/08/03 16:57:45 INFO DAGScheduler: waiting: Set(ResultStage 57)

17/08/03 16:57:45 INFO DAGScheduler: failed: Set()

17/08/03 16:57:45 INFO DAGScheduler: Submitting ResultStage 57 (MapPartitionsRDD[65] at intersection at Session17Assignment2New.scala:312), which has no missing parents

17/08/03 16:57:45 INFO MemoryStore: Block broadcast\_36 stored as values in memory (estimated size 3.4 KB, free 349.0 MB)

17/08/03 16:57:45 INFO MemoryStore: Block broadcast\_36\_piece0 stored as bytes in memory (estimated size 1969.0 B, free 349.0 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Added broadcast\_36\_piece0 in memory on 192.168.56.1:54675 (size: 1969.0 B, free: 349.2 MB)

17/08/03 16:57:45 INFO SparkContext: Created broadcast 36 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:45 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 57 (MapPartitionsRDD[65] at intersection at Session17Assignment2New.scala:312)

17/08/03 16:57:45 INFO TaskSchedulerImpl: Adding task set 57.0 with 1 tasks

17/08/03 16:57:45 INFO TaskSetManager: Starting task 0.0 in stage 57.0 (TID 35, localhost, partition 0, PROCESS\_LOCAL, 5205 bytes)

17/08/03 16:57:45 INFO Executor: Running task 0.0 in stage 57.0 (TID 35)

17/08/03 16:57:45 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:45 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

17/08/03 16:57:45 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:45 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

17/08/03 16:57:45 INFO Executor: Finished task 0.0 in stage 57.0 (TID 35). 1611 bytes result sent to driver

17/08/03 16:57:45 INFO TaskSetManager: Finished task 0.0 in stage 57.0 (TID 35) in 15 ms on localhost (1/1)

17/08/03 16:57:45 INFO TaskSchedulerImpl: Removed TaskSet 57.0, whose tasks have all completed, from pool

17/08/03 16:57:45 INFO DAGScheduler: ResultStage 57 (isEmpty at Session17Assignment2New.scala:314) finished in 0.015 s

17/08/03 16:57:45 INFO DAGScheduler: Job 19 finished: isEmpty at Session17Assignment2New.scala:314, took 0.207115 s

17/08/03 16:57:45 INFO SparkContext: Starting job: foreach at Session17Assignment2New.scala:316

17/08/03 16:57:45 INFO MapOutputTrackerMaster: Size of output statuses for shuffle 5 is 147 bytes

17/08/03 16:57:45 INFO MapOutputTrackerMaster: Size of output statuses for shuffle 4 is 147 bytes

17/08/03 16:57:45 INFO MapOutputTrackerMaster: Size of output statuses for shuffle 14 is 147 bytes

17/08/03 16:57:45 INFO MapOutputTrackerMaster: Size of output statuses for shuffle 3 is 147 bytes

17/08/03 16:57:45 INFO MapOutputTrackerMaster: Size of output statuses for shuffle 2 is 147 bytes

17/08/03 16:57:45 INFO MapOutputTrackerMaster: Size of output statuses for shuffle 15 is 147 bytes

17/08/03 16:57:45 INFO DAGScheduler: Got job 20 (foreach at Session17Assignment2New.scala:316) with 1 output partitions

17/08/03 16:57:45 INFO DAGScheduler: Final stage: ResultStage 64 (foreach at Session17Assignment2New.scala:316)

17/08/03 16:57:45 INFO DAGScheduler: Parents of final stage: List(ShuffleMapStage 63, ShuffleMapStage 60)

17/08/03 16:57:45 INFO DAGScheduler: Missing parents: List()

17/08/03 16:57:45 INFO DAGScheduler: Submitting ResultStage 64 (MapPartitionsRDD[65] at intersection at Session17Assignment2New.scala:312), which has no missing parents

17/08/03 16:57:45 INFO MemoryStore: Block broadcast\_37 stored as values in memory (estimated size 3.3 KB, free 349.0 MB)

17/08/03 16:57:45 INFO MemoryStore: Block broadcast\_37\_piece0 stored as bytes in memory (estimated size 1935.0 B, free 349.0 MB)

17/08/03 16:57:45 INFO BlockManagerInfo: Added broadcast\_37\_piece0 in memory on 192.168.56.1:54675 (size: 1935.0 B, free: 349.2 MB)

17/08/03 16:57:45 INFO SparkContext: Created broadcast 37 from broadcast at DAGScheduler.scala:1012

17/08/03 16:57:45 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 64 (MapPartitionsRDD[65] at intersection at Session17Assignment2New.scala:312)

17/08/03 16:57:45 INFO TaskSchedulerImpl: Adding task set 64.0 with 1 tasks

17/08/03 16:57:45 INFO TaskSetManager: Starting task 0.0 in stage 64.0 (TID 36, localhost, partition 0, PROCESS\_LOCAL, 5205 bytes)

17/08/03 16:57:45 INFO Executor: Running task 0.0 in stage 64.0 (TID 36)

17/08/03 16:57:46 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:46 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

(Andrew,35.0)

(Lisa,24.0)

17/08/03 16:57:46 INFO ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks

17/08/03 16:57:46 INFO ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms

17/08/03 16:57:46 INFO Executor: Finished task 0.0 in stage 64.0 (TID 36). 1474 bytes result sent to driver

17/08/03 16:57:46 INFO TaskSetManager: Finished task 0.0 in stage 64.0 (TID 36) in 16 ms on localhost (1/1)

17/08/03 16:57:46 INFO TaskSchedulerImpl: Removed TaskSet 64.0, whose tasks have all completed, from pool

17/08/03 16:57:46 INFO DAGScheduler: ResultStage 64 (foreach at Session17Assignment2New.scala:316) finished in 0.016 s

17/08/03 16:57:46 INFO DAGScheduler: Job 20 finished: foreach at Session17Assignment2New.scala:316, took 0.034823 s

17/08/03 16:57:46 INFO SparkContext: Invoking stop() from shutdown hook

17/08/03 16:57:46 INFO SparkUI: Stopped Spark web UI at http://192.168.56.1:4040

17/08/03 16:57:46 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!

17/08/03 16:57:48 INFO MemoryStore: MemoryStore cleared

17/08/03 16:57:48 INFO BlockManager: BlockManager stopped

17/08/03 16:57:48 INFO BlockManagerMaster: BlockManagerMaster stopped

17/08/03 16:57:48 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!

17/08/03 16:57:48 INFO SparkContext: Successfully stopped SparkContext

17/08/03 16:57:48 INFO ShutdownHookManager: Shutdown hook called

17/08/03 16:57:48 INFO ShutdownHookManager: Deleting directory C:\Users\Inspiron\AppData\Local\Temp\spark-c1cb9f8b-39de-45dc-afb5-024ec53f7f84

Process finished with exit code 0