20/02/29 00:56:12 WARN util.Utils: Your hostname, Snehs-MacBook-Pro.local resolves to a loopback address: 127.0.0.1; using 10.110.20.230 instead (on interface en0)

20/02/29 00:56:12 WARN util.Utils: Set SPARK\_LOCAL\_IP if you need to bind to another address

20/02/29 00:56:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

20/02/29 00:56:13 INFO spark.SparkContext: Running Spark version 2.4.5

20/02/29 00:56:13 INFO spark.SparkContext: Submitted application: RepjoinDS

20/02/29 00:56:13 INFO spark.SecurityManager: Changing view acls to: snehgurdasani

20/02/29 00:56:13 INFO spark.SecurityManager: Changing modify acls to: snehgurdasani

20/02/29 00:56:13 INFO spark.SecurityManager: Changing view acls groups to:

20/02/29 00:56:13 INFO spark.SecurityManager: Changing modify acls groups to:

20/02/29 00:56:13 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(snehgurdasani); groups with view permissions: Set(); users with modify permissions: Set(snehgurdasani); groups with modify permissions: Set()

20/02/29 00:56:13 INFO util.Utils: Successfully started service 'sparkDriver' on port 57894.

20/02/29 00:56:13 INFO spark.SparkEnv: Registering MapOutputTracker

20/02/29 00:56:13 INFO spark.SparkEnv: Registering BlockManagerMaster

20/02/29 00:56:14 INFO storage.BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information

20/02/29 00:56:14 INFO storage.BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up

20/02/29 00:56:14 INFO storage.DiskBlockManager: Created local directory at /private/var/folders/nj/f9l6vhs50h975cdq0gphfn500000gn/T/blockmgr-c1136449-b9b2-4425-a19f-290e54dcd70e

20/02/29 00:56:14 INFO memory.MemoryStore: MemoryStore started with capacity 366.3 MB

20/02/29 00:56:14 INFO spark.SparkEnv: Registering OutputCommitCoordinator

20/02/29 00:56:14 INFO util.log: Logging initialized @3376ms

20/02/29 00:56:14 INFO server.Server: jetty-9.3.z-SNAPSHOT, build timestamp: unknown, git hash: unknown

20/02/29 00:56:14 INFO server.Server: Started @3473ms

20/02/29 00:56:14 INFO server.AbstractConnector: Started ServerConnector@7dac3fd8{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}

20/02/29 00:56:14 INFO util.Utils: Successfully started service 'SparkUI' on port 4040.

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@52c8295b{/jobs,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@581d969c{/jobs/json,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@22db8f4{/jobs/job,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@1d572e62{/jobs/job/json,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@29caf222{/stages,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@46cf05f7{/stages/json,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@5851bd4f{/stages/stage,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@3caa4757{/stages/stage/json,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@69c43e48{/stages/pool,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@1804f60d{/stages/pool/json,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@3a80515c{/storage,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@547e29a4{/storage/json,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@1c807b1d{/storage/rdd,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@238b521e{/storage/rdd/json,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@1b39fd82{/environment,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@3e2fc448{/environment/json,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@21680803{/executors,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@588ab592{/executors/json,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@c8b96ec{/executors/threadDump,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@4cc61eb1{/executors/threadDump/json,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@2d8f2f3a{/static,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@6ebd78d1{/,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@436390f4{/api,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@3228d990{/jobs/job/kill,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@54e7391d{/stages/stage/kill,null,AVAILABLE,@Spark}

20/02/29 00:56:14 INFO ui.SparkUI: Bound SparkUI to 0.0.0.0, and started at http://10.110.20.230:4040

20/02/29 00:56:14 INFO spark.SparkContext: Added JAR file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/spark-demo.jar at spark://10.110.20.230:57894/jars/spark-demo.jar with timestamp 1582955774434

20/02/29 00:56:14 INFO executor.Executor: Starting executor ID driver on host localhost

20/02/29 00:56:14 INFO util.Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 57896.

20/02/29 00:56:14 INFO netty.NettyBlockTransferService: Server created on 10.110.20.230:57896

20/02/29 00:56:14 INFO storage.BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy

20/02/29 00:56:14 INFO storage.BlockManagerMaster: Registering BlockManager BlockManagerId(driver, 10.110.20.230, 57896, None)

20/02/29 00:56:14 INFO storage.BlockManagerMasterEndpoint: Registering block manager 10.110.20.230:57896 with 366.3 MB RAM, BlockManagerId(driver, 10.110.20.230, 57896, None)

20/02/29 00:56:14 INFO storage.BlockManagerMaster: Registered BlockManager BlockManagerId(driver, 10.110.20.230, 57896, None)

20/02/29 00:56:14 INFO storage.BlockManager: Initialized BlockManager: BlockManagerId(driver, 10.110.20.230, 57896, None)

20/02/29 00:56:14 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@6d2d99fc{/metrics/json,null,AVAILABLE,@Spark}

20/02/29 00:56:15 INFO memory.MemoryStore: Block broadcast\_0 stored as values in memory (estimated size 331.2 KB, free 366.0 MB)

20/02/29 00:56:15 INFO memory.MemoryStore: Block broadcast\_0\_piece0 stored as bytes in memory (estimated size 27.9 KB, free 365.9 MB)

20/02/29 00:56:15 INFO storage.BlockManagerInfo: Added broadcast\_0\_piece0 in memory on 10.110.20.230:57896 (size: 27.9 KB, free: 366.3 MB)

20/02/29 00:56:15 INFO spark.SparkContext: Created broadcast 0 from textFile at RepjoinDS.scala:27

20/02/29 00:56:15 WARN spark.SparkContext: Using an existing SparkContext; some configuration may not take effect.

20/02/29 00:56:17 INFO internal.SharedState: Setting hive.metastore.warehouse.dir ('null') to the value of spark.sql.warehouse.dir ('file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/spark-warehouse').

20/02/29 00:56:17 INFO internal.SharedState: Warehouse path is 'file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/spark-warehouse'.

20/02/29 00:56:17 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@4b9c411{/SQL,null,AVAILABLE,@Spark}

20/02/29 00:56:17 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@41e9f86{/SQL/json,null,AVAILABLE,@Spark}

20/02/29 00:56:17 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@19fec3d6{/SQL/execution,null,AVAILABLE,@Spark}

20/02/29 00:56:17 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@5287ba5f{/SQL/execution/json,null,AVAILABLE,@Spark}

20/02/29 00:56:17 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@39266403{/static/sql,null,AVAILABLE,@Spark}

20/02/29 00:56:18 INFO state.StateStoreCoordinatorRef: Registered StateStoreCoordinator endpoint

20/02/29 00:56:19 INFO codegen.CodeGenerator: Code generated in 182.982858 ms

20/02/29 00:56:19 INFO codegen.CodeGenerator: Code generated in 182.981396 ms

20/02/29 00:56:19 INFO codegen.CodeGenerator: Code generated in 183.102124 ms

20/02/29 00:56:19 INFO spark.ContextCleaner: Cleaned accumulator 0

20/02/29 00:56:19 INFO mapred.FileInputFormat: Total input files to process : 1

20/02/29 00:56:19 INFO spark.SparkContext: Starting job: run at ThreadPoolExecutor.java:1149

20/02/29 00:56:19 INFO spark.SparkContext: Starting job: run at ThreadPoolExecutor.java:1149

20/02/29 00:56:19 INFO scheduler.DAGScheduler: Got job 0 (run at ThreadPoolExecutor.java:1149) with 40 output partitions

20/02/29 00:56:19 INFO scheduler.DAGScheduler: Final stage: ResultStage 0 (run at ThreadPoolExecutor.java:1149)

20/02/29 00:56:19 INFO scheduler.DAGScheduler: Parents of final stage: List()

20/02/29 00:56:19 INFO scheduler.DAGScheduler: Missing parents: List()

20/02/29 00:56:19 INFO scheduler.DAGScheduler: Submitting ResultStage 0 (MapPartitionsRDD[9] at run at ThreadPoolExecutor.java:1149), which has no missing parents

20/02/29 00:56:19 INFO memory.MemoryStore: Block broadcast\_1 stored as values in memory (estimated size 10.6 KB, free 365.9 MB)

20/02/29 00:56:19 INFO memory.MemoryStore: Block broadcast\_1\_piece0 stored as bytes in memory (estimated size 5.1 KB, free 365.9 MB)

20/02/29 00:56:19 INFO storage.BlockManagerInfo: Added broadcast\_1\_piece0 in memory on 10.110.20.230:57896 (size: 5.1 KB, free: 366.3 MB)

20/02/29 00:56:19 INFO spark.SparkContext: Created broadcast 1 from broadcast at DAGScheduler.scala:1163

20/02/29 00:56:19 INFO scheduler.DAGScheduler: Submitting 40 missing tasks from ResultStage 0 (MapPartitionsRDD[9] at run at ThreadPoolExecutor.java:1149) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))

20/02/29 00:56:19 INFO scheduler.TaskSchedulerImpl: Adding task set 0.0 with 40 tasks

20/02/29 00:56:19 INFO scheduler.DAGScheduler: Got job 1 (run at ThreadPoolExecutor.java:1149) with 40 output partitions

20/02/29 00:56:19 INFO scheduler.DAGScheduler: Final stage: ResultStage 1 (run at ThreadPoolExecutor.java:1149)

20/02/29 00:56:19 INFO scheduler.DAGScheduler: Parents of final stage: List()

20/02/29 00:56:19 INFO scheduler.DAGScheduler: Missing parents: List()

20/02/29 00:56:19 INFO scheduler.DAGScheduler: Submitting ResultStage 1 (MapPartitionsRDD[8] at run at ThreadPoolExecutor.java:1149), which has no missing parents

20/02/29 00:56:19 INFO memory.MemoryStore: Block broadcast\_2 stored as values in memory (estimated size 10.6 KB, free 365.9 MB)

20/02/29 00:56:19 INFO memory.MemoryStore: Block broadcast\_2\_piece0 stored as bytes in memory (estimated size 5.0 KB, free 365.9 MB)

20/02/29 00:56:19 INFO storage.BlockManagerInfo: Added broadcast\_2\_piece0 in memory on 10.110.20.230:57896 (size: 5.0 KB, free: 366.3 MB)

20/02/29 00:56:19 INFO spark.SparkContext: Created broadcast 2 from broadcast at DAGScheduler.scala:1163

20/02/29 00:56:19 INFO scheduler.DAGScheduler: Submitting 40 missing tasks from ResultStage 1 (MapPartitionsRDD[8] at run at ThreadPoolExecutor.java:1149) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))

20/02/29 00:56:19 INFO scheduler.TaskSchedulerImpl: Adding task set 1.0 with 40 tasks

20/02/29 00:56:19 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 0.0 (TID 0, localhost, executor driver, partition 0, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:19 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 0.0 (TID 1, localhost, executor driver, partition 1, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:19 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 0.0 (TID 2, localhost, executor driver, partition 2, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:19 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 0.0 (TID 3, localhost, executor driver, partition 3, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:19 INFO executor.Executor: Running task 2.0 in stage 0.0 (TID 2)

20/02/29 00:56:19 INFO executor.Executor: Running task 0.0 in stage 0.0 (TID 0)

20/02/29 00:56:19 INFO executor.Executor: Running task 1.0 in stage 0.0 (TID 1)

20/02/29 00:56:19 INFO executor.Executor: Running task 3.0 in stage 0.0 (TID 3)

20/02/29 00:56:19 INFO executor.Executor: Fetching spark://10.110.20.230:57894/jars/spark-demo.jar with timestamp 1582955774434

20/02/29 00:56:19 INFO client.TransportClientFactory: Successfully created connection to /10.110.20.230:57894 after 47 ms (0 ms spent in bootstraps)

20/02/29 00:56:19 INFO util.Utils: Fetching spark://10.110.20.230:57894/jars/spark-demo.jar to /private/var/folders/nj/f9l6vhs50h975cdq0gphfn500000gn/T/spark-6dd8c03c-e456-4363-8bd5-13fe598d3ed8/userFiles-d65e49af-04f6-48c4-87b6-0adb8f143417/fetchFileTemp1435461397018149715.tmp

20/02/29 00:56:19 INFO executor.Executor: Adding file:/private/var/folders/nj/f9l6vhs50h975cdq0gphfn500000gn/T/spark-6dd8c03c-e456-4363-8bd5-13fe598d3ed8/userFiles-d65e49af-04f6-48c4-87b6-0adb8f143417/spark-demo.jar to class loader

20/02/29 00:56:20 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:67108864+33554432

20/02/29 00:56:20 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:0+33554432

20/02/29 00:56:20 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:33554432+33554432

20/02/29 00:56:20 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:100663296+33554432

20/02/29 00:56:21 INFO executor.Executor: Finished task 3.0 in stage 0.0 (TID 3). 9716 bytes result sent to driver

20/02/29 00:56:21 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 0.0 (TID 4, localhost, executor driver, partition 4, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:21 INFO executor.Executor: Running task 4.0 in stage 0.0 (TID 4)

20/02/29 00:56:21 INFO executor.Executor: Finished task 2.0 in stage 0.0 (TID 2). 12478 bytes result sent to driver

20/02/29 00:56:21 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 0.0 (TID 3) in 2164 ms on localhost (executor driver) (1/40)

20/02/29 00:56:21 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:134217728+33554432

20/02/29 00:56:21 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 0.0 (TID 5, localhost, executor driver, partition 5, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:21 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 0.0 (TID 2) in 2174 ms on localhost (executor driver) (2/40)

20/02/29 00:56:21 INFO executor.Executor: Running task 5.0 in stage 0.0 (TID 5)

20/02/29 00:56:21 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:167772160+33554432

20/02/29 00:56:21 INFO executor.Executor: Finished task 1.0 in stage 0.0 (TID 1). 26040 bytes result sent to driver

20/02/29 00:56:21 INFO executor.Executor: Finished task 0.0 in stage 0.0 (TID 0). 36512 bytes result sent to driver

20/02/29 00:56:21 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 0.0 (TID 6, localhost, executor driver, partition 6, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:21 INFO executor.Executor: Running task 6.0 in stage 0.0 (TID 6)

20/02/29 00:56:21 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 0.0 (TID 7, localhost, executor driver, partition 7, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:21 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:201326592+33554432

20/02/29 00:56:21 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 0.0 (TID 0) in 2232 ms on localhost (executor driver) (3/40)

20/02/29 00:56:21 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 0.0 (TID 1) in 2214 ms on localhost (executor driver) (4/40)

20/02/29 00:56:21 INFO executor.Executor: Running task 7.0 in stage 0.0 (TID 7)

20/02/29 00:56:21 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:234881024+33554432

20/02/29 00:56:23 INFO executor.Executor: Finished task 6.0 in stage 0.0 (TID 6). 9781 bytes result sent to driver

20/02/29 00:56:23 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 0.0 (TID 8, localhost, executor driver, partition 8, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:23 INFO executor.Executor: Running task 8.0 in stage 0.0 (TID 8)

20/02/29 00:56:23 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 0.0 (TID 6) in 1259 ms on localhost (executor driver) (5/40)

20/02/29 00:56:23 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:268435456+33554432

20/02/29 00:56:23 INFO executor.Executor: Finished task 4.0 in stage 0.0 (TID 4). 8888 bytes result sent to driver

20/02/29 00:56:23 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 0.0 (TID 9, localhost, executor driver, partition 9, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:23 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 0.0 (TID 4) in 1312 ms on localhost (executor driver) (6/40)

20/02/29 00:56:23 INFO executor.Executor: Running task 9.0 in stage 0.0 (TID 9)

20/02/29 00:56:23 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:301989888+33554432

20/02/29 00:56:23 INFO executor.Executor: Finished task 7.0 in stage 0.0 (TID 7). 3123 bytes result sent to driver

20/02/29 00:56:23 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 0.0 (TID 10, localhost, executor driver, partition 10, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:23 INFO executor.Executor: Running task 10.0 in stage 0.0 (TID 10)

20/02/29 00:56:23 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 0.0 (TID 7) in 1270 ms on localhost (executor driver) (7/40)

20/02/29 00:56:23 INFO executor.Executor: Finished task 5.0 in stage 0.0 (TID 5). 13768 bytes result sent to driver

20/02/29 00:56:23 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:335544320+33554432

20/02/29 00:56:23 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 0.0 (TID 11, localhost, executor driver, partition 11, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:23 INFO executor.Executor: Running task 11.0 in stage 0.0 (TID 11)

20/02/29 00:56:23 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 0.0 (TID 5) in 1312 ms on localhost (executor driver) (8/40)

20/02/29 00:56:23 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:369098752+33554432

20/02/29 00:56:24 INFO executor.Executor: Finished task 8.0 in stage 0.0 (TID 8). 3329 bytes result sent to driver

20/02/29 00:56:24 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 0.0 (TID 12, localhost, executor driver, partition 12, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:24 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 0.0 (TID 8) in 1236 ms on localhost (executor driver) (9/40)

20/02/29 00:56:24 INFO executor.Executor: Running task 12.0 in stage 0.0 (TID 12)

20/02/29 00:56:24 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:402653184+33554432

20/02/29 00:56:24 INFO executor.Executor: Finished task 9.0 in stage 0.0 (TID 9). 7112 bytes result sent to driver

20/02/29 00:56:24 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 0.0 (TID 13, localhost, executor driver, partition 13, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:24 INFO executor.Executor: Running task 13.0 in stage 0.0 (TID 13)

20/02/29 00:56:24 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 0.0 (TID 9) in 1255 ms on localhost (executor driver) (10/40)

20/02/29 00:56:24 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:436207616+33554432

20/02/29 00:56:24 INFO executor.Executor: Finished task 10.0 in stage 0.0 (TID 10). 9348 bytes result sent to driver

20/02/29 00:56:24 INFO scheduler.TaskSetManager: Starting task 14.0 in stage 0.0 (TID 14, localhost, executor driver, partition 14, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:24 INFO executor.Executor: Running task 14.0 in stage 0.0 (TID 14)

20/02/29 00:56:24 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 0.0 (TID 10) in 1254 ms on localhost (executor driver) (11/40)

20/02/29 00:56:24 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:469762048+33554432

20/02/29 00:56:24 INFO executor.Executor: Finished task 11.0 in stage 0.0 (TID 11). 13422 bytes result sent to driver

20/02/29 00:56:24 INFO scheduler.TaskSetManager: Starting task 15.0 in stage 0.0 (TID 15, localhost, executor driver, partition 15, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:24 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 0.0 (TID 11) in 1283 ms on localhost (executor driver) (12/40)

20/02/29 00:56:24 INFO executor.Executor: Running task 15.0 in stage 0.0 (TID 15)

20/02/29 00:56:24 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:503316480+33554432

20/02/29 00:56:25 INFO executor.Executor: Finished task 12.0 in stage 0.0 (TID 12). 30612 bytes result sent to driver

20/02/29 00:56:25 INFO scheduler.TaskSetManager: Starting task 16.0 in stage 0.0 (TID 16, localhost, executor driver, partition 16, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:25 INFO executor.Executor: Running task 16.0 in stage 0.0 (TID 16)

20/02/29 00:56:25 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 0.0 (TID 12) in 1295 ms on localhost (executor driver) (13/40)

20/02/29 00:56:25 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:536870912+33554432

20/02/29 00:56:25 INFO executor.Executor: Finished task 15.0 in stage 0.0 (TID 15). 7102 bytes result sent to driver

20/02/29 00:56:25 INFO scheduler.TaskSetManager: Starting task 17.0 in stage 0.0 (TID 17, localhost, executor driver, partition 17, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:25 INFO executor.Executor: Running task 17.0 in stage 0.0 (TID 17)

20/02/29 00:56:25 INFO scheduler.TaskSetManager: Finished task 15.0 in stage 0.0 (TID 15) in 1255 ms on localhost (executor driver) (14/40)

20/02/29 00:56:25 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:570425344+33554432

20/02/29 00:56:25 INFO executor.Executor: Finished task 14.0 in stage 0.0 (TID 14). 55717 bytes result sent to driver

20/02/29 00:56:25 INFO scheduler.TaskSetManager: Starting task 18.0 in stage 0.0 (TID 18, localhost, executor driver, partition 18, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:25 INFO executor.Executor: Running task 18.0 in stage 0.0 (TID 18)

20/02/29 00:56:25 INFO scheduler.TaskSetManager: Finished task 14.0 in stage 0.0 (TID 14) in 1297 ms on localhost (executor driver) (15/40)

20/02/29 00:56:25 INFO executor.Executor: Finished task 13.0 in stage 0.0 (TID 13). 48569 bytes result sent to driver

20/02/29 00:56:25 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:603979776+33554432

20/02/29 00:56:25 INFO scheduler.TaskSetManager: Starting task 19.0 in stage 0.0 (TID 19, localhost, executor driver, partition 19, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:25 INFO scheduler.TaskSetManager: Finished task 13.0 in stage 0.0 (TID 13) in 1311 ms on localhost (executor driver) (16/40)

20/02/29 00:56:25 INFO executor.Executor: Running task 19.0 in stage 0.0 (TID 19)

20/02/29 00:56:25 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:637534208+33554432

20/02/29 00:56:26 INFO executor.Executor: Finished task 16.0 in stage 0.0 (TID 16). 18822 bytes result sent to driver

20/02/29 00:56:26 INFO scheduler.TaskSetManager: Starting task 20.0 in stage 0.0 (TID 20, localhost, executor driver, partition 20, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:26 INFO executor.Executor: Running task 20.0 in stage 0.0 (TID 20)

20/02/29 00:56:26 INFO scheduler.TaskSetManager: Finished task 16.0 in stage 0.0 (TID 16) in 1231 ms on localhost (executor driver) (17/40)

20/02/29 00:56:26 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:671088640+33554432

20/02/29 00:56:27 INFO executor.Executor: Finished task 17.0 in stage 0.0 (TID 17). 6803 bytes result sent to driver

20/02/29 00:56:27 INFO scheduler.TaskSetManager: Starting task 21.0 in stage 0.0 (TID 21, localhost, executor driver, partition 21, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:27 INFO executor.Executor: Running task 21.0 in stage 0.0 (TID 21)

20/02/29 00:56:27 INFO scheduler.TaskSetManager: Finished task 17.0 in stage 0.0 (TID 17) in 1283 ms on localhost (executor driver) (18/40)

20/02/29 00:56:27 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:704643072+33554432

20/02/29 00:56:27 INFO executor.Executor: Finished task 19.0 in stage 0.0 (TID 19). 32883 bytes result sent to driver

20/02/29 00:56:27 INFO scheduler.TaskSetManager: Starting task 22.0 in stage 0.0 (TID 22, localhost, executor driver, partition 22, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:27 INFO scheduler.TaskSetManager: Finished task 19.0 in stage 0.0 (TID 19) in 1289 ms on localhost (executor driver) (19/40)

20/02/29 00:56:27 INFO executor.Executor: Running task 22.0 in stage 0.0 (TID 22)

20/02/29 00:56:27 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:738197504+33554432

20/02/29 00:56:27 INFO executor.Executor: Finished task 18.0 in stage 0.0 (TID 18). 13318 bytes result sent to driver

20/02/29 00:56:27 INFO scheduler.TaskSetManager: Starting task 23.0 in stage 0.0 (TID 23, localhost, executor driver, partition 23, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:27 INFO executor.Executor: Running task 23.0 in stage 0.0 (TID 23)

20/02/29 00:56:27 INFO scheduler.TaskSetManager: Finished task 18.0 in stage 0.0 (TID 18) in 1336 ms on localhost (executor driver) (20/40)

20/02/29 00:56:27 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:771751936+33554432

20/02/29 00:56:28 INFO executor.Executor: Finished task 20.0 in stage 0.0 (TID 20). 17019 bytes result sent to driver

20/02/29 00:56:28 INFO scheduler.TaskSetManager: Starting task 24.0 in stage 0.0 (TID 24, localhost, executor driver, partition 24, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:28 INFO executor.Executor: Running task 24.0 in stage 0.0 (TID 24)

20/02/29 00:56:28 INFO scheduler.TaskSetManager: Finished task 20.0 in stage 0.0 (TID 20) in 1262 ms on localhost (executor driver) (21/40)

20/02/29 00:56:28 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:805306368+33554432

20/02/29 00:56:28 INFO executor.Executor: Finished task 21.0 in stage 0.0 (TID 21). 11100 bytes result sent to driver

20/02/29 00:56:28 INFO scheduler.TaskSetManager: Starting task 25.0 in stage 0.0 (TID 25, localhost, executor driver, partition 25, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:28 INFO executor.Executor: Running task 25.0 in stage 0.0 (TID 25)

20/02/29 00:56:28 INFO scheduler.TaskSetManager: Finished task 21.0 in stage 0.0 (TID 21) in 1280 ms on localhost (executor driver) (22/40)

20/02/29 00:56:28 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:838860800+33554432

20/02/29 00:56:28 INFO executor.Executor: Finished task 22.0 in stage 0.0 (TID 22). 454156 bytes result sent to driver

20/02/29 00:56:28 INFO scheduler.TaskSetManager: Starting task 26.0 in stage 0.0 (TID 26, localhost, executor driver, partition 26, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:28 INFO executor.Executor: Running task 26.0 in stage 0.0 (TID 26)

20/02/29 00:56:28 INFO scheduler.TaskSetManager: Finished task 22.0 in stage 0.0 (TID 22) in 1493 ms on localhost (executor driver) (23/40)

20/02/29 00:56:28 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:872415232+33554432

20/02/29 00:56:28 INFO executor.Executor: Finished task 23.0 in stage 0.0 (TID 23). 1371 bytes result sent to driver

20/02/29 00:56:28 INFO scheduler.TaskSetManager: Starting task 27.0 in stage 0.0 (TID 27, localhost, executor driver, partition 27, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:28 INFO scheduler.TaskSetManager: Finished task 23.0 in stage 0.0 (TID 23) in 1476 ms on localhost (executor driver) (24/40)

20/02/29 00:56:28 INFO executor.Executor: Running task 27.0 in stage 0.0 (TID 27)

20/02/29 00:56:28 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:905969664+33554432

20/02/29 00:56:29 INFO executor.Executor: Finished task 24.0 in stage 0.0 (TID 24). 1371 bytes result sent to driver

20/02/29 00:56:29 INFO scheduler.TaskSetManager: Starting task 28.0 in stage 0.0 (TID 28, localhost, executor driver, partition 28, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:29 INFO executor.Executor: Running task 28.0 in stage 0.0 (TID 28)

20/02/29 00:56:29 INFO scheduler.TaskSetManager: Finished task 24.0 in stage 0.0 (TID 24) in 1410 ms on localhost (executor driver) (25/40)

20/02/29 00:56:29 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:939524096+33554432

20/02/29 00:56:29 INFO executor.Executor: Finished task 25.0 in stage 0.0 (TID 25). 1371 bytes result sent to driver

20/02/29 00:56:29 INFO scheduler.TaskSetManager: Starting task 29.0 in stage 0.0 (TID 29, localhost, executor driver, partition 29, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:29 INFO executor.Executor: Running task 29.0 in stage 0.0 (TID 29)

20/02/29 00:56:29 INFO scheduler.TaskSetManager: Finished task 25.0 in stage 0.0 (TID 25) in 1439 ms on localhost (executor driver) (26/40)

20/02/29 00:56:29 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:973078528+33554432

20/02/29 00:56:29 INFO executor.Executor: Finished task 26.0 in stage 0.0 (TID 26). 1371 bytes result sent to driver

20/02/29 00:56:29 INFO scheduler.TaskSetManager: Starting task 30.0 in stage 0.0 (TID 30, localhost, executor driver, partition 30, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:29 INFO executor.Executor: Running task 30.0 in stage 0.0 (TID 30)

20/02/29 00:56:29 INFO scheduler.TaskSetManager: Finished task 26.0 in stage 0.0 (TID 26) in 1316 ms on localhost (executor driver) (27/40)

20/02/29 00:56:29 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1006632960+33554432

20/02/29 00:56:29 INFO executor.Executor: Finished task 27.0 in stage 0.0 (TID 27). 1371 bytes result sent to driver

20/02/29 00:56:29 INFO scheduler.TaskSetManager: Starting task 31.0 in stage 0.0 (TID 31, localhost, executor driver, partition 31, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:29 INFO executor.Executor: Running task 31.0 in stage 0.0 (TID 31)

20/02/29 00:56:29 INFO scheduler.TaskSetManager: Finished task 27.0 in stage 0.0 (TID 27) in 1314 ms on localhost (executor driver) (28/40)

20/02/29 00:56:29 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1040187392+33554432

20/02/29 00:56:30 INFO executor.Executor: Finished task 28.0 in stage 0.0 (TID 28). 1371 bytes result sent to driver

20/02/29 00:56:30 INFO scheduler.TaskSetManager: Starting task 32.0 in stage 0.0 (TID 32, localhost, executor driver, partition 32, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:30 INFO executor.Executor: Running task 32.0 in stage 0.0 (TID 32)

20/02/29 00:56:30 INFO scheduler.TaskSetManager: Finished task 28.0 in stage 0.0 (TID 28) in 1268 ms on localhost (executor driver) (29/40)

20/02/29 00:56:30 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1073741824+33554432

20/02/29 00:56:31 INFO executor.Executor: Finished task 29.0 in stage 0.0 (TID 29). 1371 bytes result sent to driver

20/02/29 00:56:31 INFO scheduler.TaskSetManager: Starting task 33.0 in stage 0.0 (TID 33, localhost, executor driver, partition 33, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:31 INFO executor.Executor: Running task 33.0 in stage 0.0 (TID 33)

20/02/29 00:56:31 INFO scheduler.TaskSetManager: Finished task 29.0 in stage 0.0 (TID 29) in 1280 ms on localhost (executor driver) (30/40)

20/02/29 00:56:31 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1107296256+33554432

20/02/29 00:56:31 INFO executor.Executor: Finished task 30.0 in stage 0.0 (TID 30). 1371 bytes result sent to driver

20/02/29 00:56:31 INFO scheduler.TaskSetManager: Starting task 34.0 in stage 0.0 (TID 34, localhost, executor driver, partition 34, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:31 INFO executor.Executor: Running task 34.0 in stage 0.0 (TID 34)

20/02/29 00:56:31 INFO scheduler.TaskSetManager: Finished task 30.0 in stage 0.0 (TID 30) in 1265 ms on localhost (executor driver) (31/40)

20/02/29 00:56:31 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1140850688+33554432

20/02/29 00:56:31 INFO executor.Executor: Finished task 31.0 in stage 0.0 (TID 31). 1371 bytes result sent to driver

20/02/29 00:56:31 INFO scheduler.TaskSetManager: Starting task 35.0 in stage 0.0 (TID 35, localhost, executor driver, partition 35, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:31 INFO executor.Executor: Running task 35.0 in stage 0.0 (TID 35)

20/02/29 00:56:31 INFO scheduler.TaskSetManager: Finished task 31.0 in stage 0.0 (TID 31) in 1270 ms on localhost (executor driver) (32/40)

20/02/29 00:56:31 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1174405120+33554432

20/02/29 00:56:32 INFO executor.Executor: Finished task 32.0 in stage 0.0 (TID 32). 1371 bytes result sent to driver

20/02/29 00:56:32 INFO scheduler.TaskSetManager: Starting task 36.0 in stage 0.0 (TID 36, localhost, executor driver, partition 36, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:32 INFO scheduler.TaskSetManager: Finished task 32.0 in stage 0.0 (TID 32) in 1272 ms on localhost (executor driver) (33/40)

20/02/29 00:56:32 INFO executor.Executor: Running task 36.0 in stage 0.0 (TID 36)

20/02/29 00:56:32 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1207959552+33554432

20/02/29 00:56:32 INFO executor.Executor: Finished task 33.0 in stage 0.0 (TID 33). 1371 bytes result sent to driver

20/02/29 00:56:32 INFO scheduler.TaskSetManager: Starting task 37.0 in stage 0.0 (TID 37, localhost, executor driver, partition 37, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:32 INFO executor.Executor: Running task 37.0 in stage 0.0 (TID 37)

20/02/29 00:56:32 INFO scheduler.TaskSetManager: Finished task 33.0 in stage 0.0 (TID 33) in 1287 ms on localhost (executor driver) (34/40)

20/02/29 00:56:32 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1241513984+33554432

20/02/29 00:56:32 INFO executor.Executor: Finished task 34.0 in stage 0.0 (TID 34). 1371 bytes result sent to driver

20/02/29 00:56:32 INFO scheduler.TaskSetManager: Starting task 38.0 in stage 0.0 (TID 38, localhost, executor driver, partition 38, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:32 INFO executor.Executor: Running task 38.0 in stage 0.0 (TID 38)

20/02/29 00:56:32 INFO scheduler.TaskSetManager: Finished task 34.0 in stage 0.0 (TID 34) in 1280 ms on localhost (executor driver) (35/40)

20/02/29 00:56:32 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1275068416+33554432

20/02/29 00:56:32 INFO executor.Executor: Finished task 35.0 in stage 0.0 (TID 35). 1371 bytes result sent to driver

20/02/29 00:56:32 INFO scheduler.TaskSetManager: Starting task 39.0 in stage 0.0 (TID 39, localhost, executor driver, partition 39, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:32 INFO executor.Executor: Running task 39.0 in stage 0.0 (TID 39)

20/02/29 00:56:32 INFO scheduler.TaskSetManager: Finished task 35.0 in stage 0.0 (TID 35) in 1288 ms on localhost (executor driver) (36/40)

20/02/29 00:56:32 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1308622848+10658977

20/02/29 00:56:32 INFO executor.Executor: Finished task 39.0 in stage 0.0 (TID 39). 1371 bytes result sent to driver

20/02/29 00:56:32 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 1.0 (TID 40, localhost, executor driver, partition 0, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:32 INFO executor.Executor: Running task 0.0 in stage 1.0 (TID 40)

20/02/29 00:56:32 INFO scheduler.TaskSetManager: Finished task 39.0 in stage 0.0 (TID 39) in 404 ms on localhost (executor driver) (37/40)

20/02/29 00:56:32 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:0+33554432

20/02/29 00:56:33 INFO executor.Executor: Finished task 36.0 in stage 0.0 (TID 36). 1371 bytes result sent to driver

20/02/29 00:56:33 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 1.0 (TID 41, localhost, executor driver, partition 1, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:33 INFO scheduler.TaskSetManager: Finished task 36.0 in stage 0.0 (TID 36) in 1266 ms on localhost (executor driver) (38/40)

20/02/29 00:56:33 INFO executor.Executor: Running task 1.0 in stage 1.0 (TID 41)

20/02/29 00:56:33 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:33554432+33554432

20/02/29 00:56:33 INFO executor.Executor: Finished task 37.0 in stage 0.0 (TID 37). 1371 bytes result sent to driver

20/02/29 00:56:33 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 1.0 (TID 42, localhost, executor driver, partition 2, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:33 INFO executor.Executor: Running task 2.0 in stage 1.0 (TID 42)

20/02/29 00:56:33 INFO scheduler.TaskSetManager: Finished task 37.0 in stage 0.0 (TID 37) in 1273 ms on localhost (executor driver) (39/40)

20/02/29 00:56:33 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:67108864+33554432

20/02/29 00:56:33 INFO executor.Executor: Finished task 38.0 in stage 0.0 (TID 38). 1371 bytes result sent to driver

20/02/29 00:56:33 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 1.0 (TID 43, localhost, executor driver, partition 3, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:33 INFO executor.Executor: Running task 3.0 in stage 1.0 (TID 43)

20/02/29 00:56:33 INFO scheduler.TaskSetManager: Finished task 38.0 in stage 0.0 (TID 38) in 1249 ms on localhost (executor driver) (40/40)

20/02/29 00:56:33 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool

20/02/29 00:56:33 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:100663296+33554432

20/02/29 00:56:33 INFO scheduler.DAGScheduler: ResultStage 0 (run at ThreadPoolExecutor.java:1149) finished in 14.032 s

20/02/29 00:56:33 INFO scheduler.DAGScheduler: Job 0 finished: run at ThreadPoolExecutor.java:1149, took 14.087008 s

20/02/29 00:56:33 INFO codegen.CodeGenerator: Code generated in 14.098594 ms

20/02/29 00:56:33 INFO memory.MemoryStore: Block broadcast\_3 stored as values in memory (estimated size 4.1 MB, free 361.8 MB)

20/02/29 00:56:33 INFO memory.MemoryStore: Block broadcast\_3\_piece0 stored as bytes in memory (estimated size 1163.1 KB, free 360.7 MB)

20/02/29 00:56:33 INFO storage.BlockManagerInfo: Added broadcast\_3\_piece0 in memory on 10.110.20.230:57896 (size: 1163.1 KB, free: 365.1 MB)

20/02/29 00:56:33 INFO spark.SparkContext: Created broadcast 3 from run at ThreadPoolExecutor.java:1149

20/02/29 00:56:34 INFO executor.Executor: Finished task 0.0 in stage 1.0 (TID 40). 36512 bytes result sent to driver

20/02/29 00:56:34 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 1.0 (TID 44, localhost, executor driver, partition 4, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:34 INFO executor.Executor: Running task 4.0 in stage 1.0 (TID 44)

20/02/29 00:56:34 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 1.0 (TID 40) in 1398 ms on localhost (executor driver) (1/40)

20/02/29 00:56:34 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:134217728+33554432

20/02/29 00:56:34 INFO executor.Executor: Finished task 1.0 in stage 1.0 (TID 41). 26040 bytes result sent to driver

20/02/29 00:56:34 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 1.0 (TID 45, localhost, executor driver, partition 5, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:34 INFO executor.Executor: Running task 5.0 in stage 1.0 (TID 45)

20/02/29 00:56:34 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 1.0 (TID 41) in 1370 ms on localhost (executor driver) (2/40)

20/02/29 00:56:34 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:167772160+33554432

20/02/29 00:56:34 INFO executor.Executor: Finished task 2.0 in stage 1.0 (TID 42). 12478 bytes result sent to driver

20/02/29 00:56:34 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 1.0 (TID 46, localhost, executor driver, partition 6, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:34 INFO executor.Executor: Running task 6.0 in stage 1.0 (TID 46)

20/02/29 00:56:34 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 1.0 (TID 42) in 1366 ms on localhost (executor driver) (3/40)

20/02/29 00:56:34 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:201326592+33554432

20/02/29 00:56:34 INFO executor.Executor: Finished task 3.0 in stage 1.0 (TID 43). 9673 bytes result sent to driver

20/02/29 00:56:34 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 1.0 (TID 47, localhost, executor driver, partition 7, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:34 INFO executor.Executor: Running task 7.0 in stage 1.0 (TID 47)

20/02/29 00:56:34 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 1.0 (TID 43) in 1346 ms on localhost (executor driver) (4/40)

20/02/29 00:56:35 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:234881024+33554432

20/02/29 00:56:35 INFO executor.Executor: Finished task 4.0 in stage 1.0 (TID 44). 8888 bytes result sent to driver

20/02/29 00:56:35 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 1.0 (TID 48, localhost, executor driver, partition 8, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:35 INFO executor.Executor: Running task 8.0 in stage 1.0 (TID 48)

20/02/29 00:56:35 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 1.0 (TID 44) in 1259 ms on localhost (executor driver) (5/40)

20/02/29 00:56:35 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:268435456+33554432

20/02/29 00:56:36 INFO executor.Executor: Finished task 5.0 in stage 1.0 (TID 45). 13725 bytes result sent to driver

20/02/29 00:56:36 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 1.0 (TID 49, localhost, executor driver, partition 9, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:36 INFO executor.Executor: Running task 9.0 in stage 1.0 (TID 49)

20/02/29 00:56:36 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 1.0 (TID 45) in 1281 ms on localhost (executor driver) (6/40)

20/02/29 00:56:36 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:301989888+33554432

20/02/29 00:56:36 INFO executor.Executor: Finished task 6.0 in stage 1.0 (TID 46). 9738 bytes result sent to driver

20/02/29 00:56:36 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 1.0 (TID 50, localhost, executor driver, partition 10, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:36 INFO executor.Executor: Running task 10.0 in stage 1.0 (TID 50)

20/02/29 00:56:36 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 1.0 (TID 46) in 1275 ms on localhost (executor driver) (7/40)

20/02/29 00:56:36 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:335544320+33554432

20/02/29 00:56:36 INFO executor.Executor: Finished task 7.0 in stage 1.0 (TID 47). 3166 bytes result sent to driver

20/02/29 00:56:36 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 1.0 (TID 51, localhost, executor driver, partition 11, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:36 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 1.0 (TID 47) in 1251 ms on localhost (executor driver) (8/40)

20/02/29 00:56:36 INFO executor.Executor: Running task 11.0 in stage 1.0 (TID 51)

20/02/29 00:56:36 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:369098752+33554432

20/02/29 00:56:36 INFO executor.Executor: Finished task 8.0 in stage 1.0 (TID 48). 3329 bytes result sent to driver

20/02/29 00:56:36 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 1.0 (TID 52, localhost, executor driver, partition 12, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:36 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 1.0 (TID 48) in 1257 ms on localhost (executor driver) (9/40)

20/02/29 00:56:36 INFO executor.Executor: Running task 12.0 in stage 1.0 (TID 52)

20/02/29 00:56:36 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:402653184+33554432

20/02/29 00:56:37 INFO executor.Executor: Finished task 9.0 in stage 1.0 (TID 49). 7112 bytes result sent to driver

20/02/29 00:56:37 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 1.0 (TID 53, localhost, executor driver, partition 13, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:37 INFO executor.Executor: Running task 13.0 in stage 1.0 (TID 53)

20/02/29 00:56:37 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 1.0 (TID 49) in 1230 ms on localhost (executor driver) (10/40)

20/02/29 00:56:37 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:436207616+33554432

20/02/29 00:56:37 INFO executor.Executor: Finished task 10.0 in stage 1.0 (TID 50). 9262 bytes result sent to driver

20/02/29 00:56:37 INFO scheduler.TaskSetManager: Starting task 14.0 in stage 1.0 (TID 54, localhost, executor driver, partition 14, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:37 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 1.0 (TID 50) in 1229 ms on localhost (executor driver) (11/40)

20/02/29 00:56:37 INFO executor.Executor: Running task 14.0 in stage 1.0 (TID 54)

20/02/29 00:56:37 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:469762048+33554432

20/02/29 00:56:37 INFO executor.Executor: Finished task 11.0 in stage 1.0 (TID 51). 13422 bytes result sent to driver

20/02/29 00:56:37 INFO scheduler.TaskSetManager: Starting task 15.0 in stage 1.0 (TID 55, localhost, executor driver, partition 15, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:37 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 1.0 (TID 51) in 1248 ms on localhost (executor driver) (12/40)

20/02/29 00:56:37 INFO executor.Executor: Running task 15.0 in stage 1.0 (TID 55)

20/02/29 00:56:37 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:503316480+33554432

20/02/29 00:56:38 INFO executor.Executor: Finished task 12.0 in stage 1.0 (TID 52). 30612 bytes result sent to driver

20/02/29 00:56:38 INFO scheduler.TaskSetManager: Starting task 16.0 in stage 1.0 (TID 56, localhost, executor driver, partition 16, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:38 INFO executor.Executor: Running task 16.0 in stage 1.0 (TID 56)

20/02/29 00:56:38 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 1.0 (TID 52) in 1289 ms on localhost (executor driver) (13/40)

20/02/29 00:56:38 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:536870912+33554432

20/02/29 00:56:38 INFO executor.Executor: Finished task 13.0 in stage 1.0 (TID 53). 48569 bytes result sent to driver

20/02/29 00:56:38 INFO scheduler.TaskSetManager: Starting task 17.0 in stage 1.0 (TID 57, localhost, executor driver, partition 17, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:38 INFO executor.Executor: Running task 17.0 in stage 1.0 (TID 57)

20/02/29 00:56:38 INFO scheduler.TaskSetManager: Finished task 13.0 in stage 1.0 (TID 53) in 1283 ms on localhost (executor driver) (14/40)

20/02/29 00:56:38 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:570425344+33554432

20/02/29 00:56:38 INFO executor.Executor: Finished task 15.0 in stage 1.0 (TID 55). 7102 bytes result sent to driver

20/02/29 00:56:38 INFO scheduler.TaskSetManager: Starting task 18.0 in stage 1.0 (TID 58, localhost, executor driver, partition 18, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:38 INFO scheduler.TaskSetManager: Finished task 15.0 in stage 1.0 (TID 55) in 1251 ms on localhost (executor driver) (15/40)

20/02/29 00:56:38 INFO executor.Executor: Running task 18.0 in stage 1.0 (TID 58)

20/02/29 00:56:38 INFO executor.Executor: Finished task 14.0 in stage 1.0 (TID 54). 55674 bytes result sent to driver

20/02/29 00:56:38 INFO scheduler.TaskSetManager: Starting task 19.0 in stage 1.0 (TID 59, localhost, executor driver, partition 19, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:38 INFO scheduler.TaskSetManager: Finished task 14.0 in stage 1.0 (TID 54) in 1298 ms on localhost (executor driver) (16/40)

20/02/29 00:56:38 INFO executor.Executor: Running task 19.0 in stage 1.0 (TID 59)

20/02/29 00:56:38 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:603979776+33554432

20/02/29 00:56:38 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:637534208+33554432

20/02/29 00:56:39 INFO executor.Executor: Finished task 16.0 in stage 1.0 (TID 56). 18822 bytes result sent to driver

20/02/29 00:56:39 INFO scheduler.TaskSetManager: Starting task 20.0 in stage 1.0 (TID 60, localhost, executor driver, partition 20, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:39 INFO scheduler.TaskSetManager: Finished task 16.0 in stage 1.0 (TID 56) in 1239 ms on localhost (executor driver) (17/40)

20/02/29 00:56:39 INFO executor.Executor: Running task 20.0 in stage 1.0 (TID 60)

20/02/29 00:56:39 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:671088640+33554432

20/02/29 00:56:39 INFO executor.Executor: Finished task 17.0 in stage 1.0 (TID 57). 6803 bytes result sent to driver

20/02/29 00:56:39 INFO scheduler.TaskSetManager: Starting task 21.0 in stage 1.0 (TID 61, localhost, executor driver, partition 21, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:39 INFO scheduler.TaskSetManager: Finished task 17.0 in stage 1.0 (TID 57) in 1219 ms on localhost (executor driver) (18/40)

20/02/29 00:56:39 INFO executor.Executor: Running task 21.0 in stage 1.0 (TID 61)

20/02/29 00:56:39 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:704643072+33554432

20/02/29 00:56:40 INFO executor.Executor: Finished task 18.0 in stage 1.0 (TID 58). 13361 bytes result sent to driver

20/02/29 00:56:40 INFO scheduler.TaskSetManager: Starting task 22.0 in stage 1.0 (TID 62, localhost, executor driver, partition 22, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:40 INFO executor.Executor: Running task 22.0 in stage 1.0 (TID 62)

20/02/29 00:56:40 INFO scheduler.TaskSetManager: Finished task 18.0 in stage 1.0 (TID 58) in 1256 ms on localhost (executor driver) (19/40)

20/02/29 00:56:40 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:738197504+33554432

20/02/29 00:56:40 INFO executor.Executor: Finished task 19.0 in stage 1.0 (TID 59). 32883 bytes result sent to driver

20/02/29 00:56:40 INFO scheduler.TaskSetManager: Starting task 23.0 in stage 1.0 (TID 63, localhost, executor driver, partition 23, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:40 INFO executor.Executor: Running task 23.0 in stage 1.0 (TID 63)

20/02/29 00:56:40 INFO scheduler.TaskSetManager: Finished task 19.0 in stage 1.0 (TID 59) in 1291 ms on localhost (executor driver) (20/40)

20/02/29 00:56:40 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:771751936+33554432

20/02/29 00:56:40 INFO executor.Executor: Finished task 20.0 in stage 1.0 (TID 60). 17062 bytes result sent to driver

20/02/29 00:56:40 INFO scheduler.TaskSetManager: Starting task 24.0 in stage 1.0 (TID 64, localhost, executor driver, partition 24, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:40 INFO executor.Executor: Running task 24.0 in stage 1.0 (TID 64)

20/02/29 00:56:40 INFO scheduler.TaskSetManager: Finished task 20.0 in stage 1.0 (TID 60) in 1257 ms on localhost (executor driver) (21/40)

20/02/29 00:56:40 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:805306368+33554432

20/02/29 00:56:41 INFO executor.Executor: Finished task 21.0 in stage 1.0 (TID 61). 11100 bytes result sent to driver

20/02/29 00:56:41 INFO scheduler.TaskSetManager: Starting task 25.0 in stage 1.0 (TID 65, localhost, executor driver, partition 25, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:41 INFO scheduler.TaskSetManager: Finished task 21.0 in stage 1.0 (TID 61) in 1279 ms on localhost (executor driver) (22/40)

20/02/29 00:56:41 INFO executor.Executor: Running task 25.0 in stage 1.0 (TID 65)

20/02/29 00:56:41 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:838860800+33554432

20/02/29 00:56:41 INFO executor.Executor: Finished task 22.0 in stage 1.0 (TID 62). 454113 bytes result sent to driver

20/02/29 00:56:41 INFO scheduler.TaskSetManager: Starting task 26.0 in stage 1.0 (TID 66, localhost, executor driver, partition 26, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:41 INFO executor.Executor: Running task 26.0 in stage 1.0 (TID 66)

20/02/29 00:56:41 INFO scheduler.TaskSetManager: Finished task 22.0 in stage 1.0 (TID 62) in 1437 ms on localhost (executor driver) (23/40)

20/02/29 00:56:41 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:872415232+33554432

20/02/29 00:56:41 INFO executor.Executor: Finished task 23.0 in stage 1.0 (TID 63). 1371 bytes result sent to driver

20/02/29 00:56:41 INFO scheduler.TaskSetManager: Starting task 27.0 in stage 1.0 (TID 67, localhost, executor driver, partition 27, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:41 INFO scheduler.TaskSetManager: Finished task 23.0 in stage 1.0 (TID 63) in 1437 ms on localhost (executor driver) (24/40)

20/02/29 00:56:41 INFO executor.Executor: Running task 27.0 in stage 1.0 (TID 67)

20/02/29 00:56:41 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:905969664+33554432

20/02/29 00:56:41 INFO executor.Executor: Finished task 24.0 in stage 1.0 (TID 64). 1371 bytes result sent to driver

20/02/29 00:56:41 INFO scheduler.TaskSetManager: Starting task 28.0 in stage 1.0 (TID 68, localhost, executor driver, partition 28, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:41 INFO scheduler.TaskSetManager: Finished task 24.0 in stage 1.0 (TID 64) in 1457 ms on localhost (executor driver) (25/40)

20/02/29 00:56:41 INFO executor.Executor: Running task 28.0 in stage 1.0 (TID 68)

20/02/29 00:56:41 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:939524096+33554432

20/02/29 00:56:42 INFO executor.Executor: Finished task 25.0 in stage 1.0 (TID 65). 1414 bytes result sent to driver

20/02/29 00:56:42 INFO scheduler.TaskSetManager: Starting task 29.0 in stage 1.0 (TID 69, localhost, executor driver, partition 29, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:42 INFO executor.Executor: Running task 29.0 in stage 1.0 (TID 69)

20/02/29 00:56:42 INFO scheduler.TaskSetManager: Finished task 25.0 in stage 1.0 (TID 65) in 1454 ms on localhost (executor driver) (26/40)

20/02/29 00:56:42 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:973078528+33554432

20/02/29 00:56:42 INFO executor.Executor: Finished task 26.0 in stage 1.0 (TID 66). 1371 bytes result sent to driver

20/02/29 00:56:42 INFO scheduler.TaskSetManager: Starting task 30.0 in stage 1.0 (TID 70, localhost, executor driver, partition 30, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:42 INFO scheduler.TaskSetManager: Finished task 26.0 in stage 1.0 (TID 66) in 1426 ms on localhost (executor driver) (27/40)

20/02/29 00:56:42 INFO executor.Executor: Running task 30.0 in stage 1.0 (TID 70)

20/02/29 00:56:42 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1006632960+33554432

20/02/29 00:56:42 INFO executor.Executor: Finished task 27.0 in stage 1.0 (TID 67). 1371 bytes result sent to driver

20/02/29 00:56:42 INFO scheduler.TaskSetManager: Starting task 31.0 in stage 1.0 (TID 71, localhost, executor driver, partition 31, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:42 INFO scheduler.TaskSetManager: Finished task 27.0 in stage 1.0 (TID 67) in 1433 ms on localhost (executor driver) (28/40)

20/02/29 00:56:42 INFO executor.Executor: Running task 31.0 in stage 1.0 (TID 71)

20/02/29 00:56:42 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1040187392+33554432

20/02/29 00:56:43 INFO executor.Executor: Finished task 28.0 in stage 1.0 (TID 68). 1371 bytes result sent to driver

20/02/29 00:56:43 INFO scheduler.TaskSetManager: Starting task 32.0 in stage 1.0 (TID 72, localhost, executor driver, partition 32, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:43 INFO executor.Executor: Running task 32.0 in stage 1.0 (TID 72)

20/02/29 00:56:43 INFO scheduler.TaskSetManager: Finished task 28.0 in stage 1.0 (TID 68) in 1296 ms on localhost (executor driver) (29/40)

20/02/29 00:56:43 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1073741824+33554432

20/02/29 00:56:43 INFO executor.Executor: Finished task 29.0 in stage 1.0 (TID 69). 1371 bytes result sent to driver

20/02/29 00:56:43 INFO scheduler.TaskSetManager: Starting task 33.0 in stage 1.0 (TID 73, localhost, executor driver, partition 33, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:43 INFO executor.Executor: Running task 33.0 in stage 1.0 (TID 73)

20/02/29 00:56:43 INFO scheduler.TaskSetManager: Finished task 29.0 in stage 1.0 (TID 69) in 1291 ms on localhost (executor driver) (30/40)

20/02/29 00:56:43 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1107296256+33554432

20/02/29 00:56:44 INFO executor.Executor: Finished task 30.0 in stage 1.0 (TID 70). 1371 bytes result sent to driver

20/02/29 00:56:44 INFO scheduler.TaskSetManager: Starting task 34.0 in stage 1.0 (TID 74, localhost, executor driver, partition 34, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:44 INFO scheduler.TaskSetManager: Finished task 30.0 in stage 1.0 (TID 70) in 1285 ms on localhost (executor driver) (31/40)

20/02/29 00:56:44 INFO executor.Executor: Running task 34.0 in stage 1.0 (TID 74)

20/02/29 00:56:44 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1140850688+33554432

20/02/29 00:56:44 INFO executor.Executor: Finished task 31.0 in stage 1.0 (TID 71). 1371 bytes result sent to driver

20/02/29 00:56:44 INFO scheduler.TaskSetManager: Starting task 35.0 in stage 1.0 (TID 75, localhost, executor driver, partition 35, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:44 INFO executor.Executor: Running task 35.0 in stage 1.0 (TID 75)

20/02/29 00:56:44 INFO scheduler.TaskSetManager: Finished task 31.0 in stage 1.0 (TID 71) in 1287 ms on localhost (executor driver) (32/40)

20/02/29 00:56:44 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1174405120+33554432

20/02/29 00:56:44 INFO executor.Executor: Finished task 32.0 in stage 1.0 (TID 72). 1328 bytes result sent to driver

20/02/29 00:56:44 INFO scheduler.TaskSetManager: Starting task 36.0 in stage 1.0 (TID 76, localhost, executor driver, partition 36, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:44 INFO executor.Executor: Running task 36.0 in stage 1.0 (TID 76)

20/02/29 00:56:44 INFO scheduler.TaskSetManager: Finished task 32.0 in stage 1.0 (TID 72) in 1291 ms on localhost (executor driver) (33/40)

20/02/29 00:56:44 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1207959552+33554432

20/02/29 00:56:45 INFO executor.Executor: Finished task 33.0 in stage 1.0 (TID 73). 1371 bytes result sent to driver

20/02/29 00:56:45 INFO scheduler.TaskSetManager: Starting task 37.0 in stage 1.0 (TID 77, localhost, executor driver, partition 37, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:45 INFO executor.Executor: Running task 37.0 in stage 1.0 (TID 77)

20/02/29 00:56:45 INFO scheduler.TaskSetManager: Finished task 33.0 in stage 1.0 (TID 73) in 1283 ms on localhost (executor driver) (34/40)

20/02/29 00:56:45 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1241513984+33554432

20/02/29 00:56:45 INFO executor.Executor: Finished task 34.0 in stage 1.0 (TID 74). 1371 bytes result sent to driver

20/02/29 00:56:45 INFO scheduler.TaskSetManager: Starting task 38.0 in stage 1.0 (TID 78, localhost, executor driver, partition 38, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:45 INFO scheduler.TaskSetManager: Finished task 34.0 in stage 1.0 (TID 74) in 1271 ms on localhost (executor driver) (35/40)

20/02/29 00:56:45 INFO executor.Executor: Running task 38.0 in stage 1.0 (TID 78)

20/02/29 00:56:45 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1275068416+33554432

20/02/29 00:56:45 INFO executor.Executor: Finished task 35.0 in stage 1.0 (TID 75). 1371 bytes result sent to driver

20/02/29 00:56:45 INFO scheduler.TaskSetManager: Starting task 39.0 in stage 1.0 (TID 79, localhost, executor driver, partition 39, PROCESS\_LOCAL, 7463 bytes)

20/02/29 00:56:45 INFO executor.Executor: Running task 39.0 in stage 1.0 (TID 79)

20/02/29 00:56:45 INFO scheduler.TaskSetManager: Finished task 35.0 in stage 1.0 (TID 75) in 1268 ms on localhost (executor driver) (36/40)

20/02/29 00:56:45 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1308622848+10658977

20/02/29 00:56:45 INFO executor.Executor: Finished task 36.0 in stage 1.0 (TID 76). 1371 bytes result sent to driver

20/02/29 00:56:45 INFO scheduler.TaskSetManager: Finished task 36.0 in stage 1.0 (TID 76) in 1244 ms on localhost (executor driver) (37/40)

20/02/29 00:56:45 INFO executor.Executor: Finished task 39.0 in stage 1.0 (TID 79). 1371 bytes result sent to driver

20/02/29 00:56:45 INFO scheduler.TaskSetManager: Finished task 39.0 in stage 1.0 (TID 79) in 408 ms on localhost (executor driver) (38/40)

20/02/29 00:56:46 INFO executor.Executor: Finished task 37.0 in stage 1.0 (TID 77). 1328 bytes result sent to driver

20/02/29 00:56:46 INFO scheduler.TaskSetManager: Finished task 37.0 in stage 1.0 (TID 77) in 1203 ms on localhost (executor driver) (39/40)

20/02/29 00:56:46 INFO executor.Executor: Finished task 38.0 in stage 1.0 (TID 78). 1371 bytes result sent to driver

20/02/29 00:56:46 INFO scheduler.TaskSetManager: Finished task 38.0 in stage 1.0 (TID 78) in 1099 ms on localhost (executor driver) (40/40)

20/02/29 00:56:46 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 1.0, whose tasks have all completed, from pool

20/02/29 00:56:46 INFO scheduler.DAGScheduler: ResultStage 1 (run at ThreadPoolExecutor.java:1149) finished in 26.815 s

20/02/29 00:56:46 INFO scheduler.DAGScheduler: Job 1 finished: run at ThreadPoolExecutor.java:1149, took 26.947159 s

20/02/29 00:56:46 INFO codegen.CodeGenerator: Code generated in 10.265132 ms

20/02/29 00:56:46 INFO memory.MemoryStore: Block broadcast\_4 stored as values in memory (estimated size 8.0 MB, free 352.7 MB)

20/02/29 00:56:46 INFO memory.MemoryStore: Block broadcast\_4\_piece0 stored as bytes in memory (estimated size 2.2 MB, free 350.5 MB)

20/02/29 00:56:46 INFO storage.BlockManagerInfo: Added broadcast\_4\_piece0 in memory on 10.110.20.230:57896 (size: 2.2 MB, free: 362.9 MB)

20/02/29 00:56:46 INFO spark.SparkContext: Created broadcast 4 from run at ThreadPoolExecutor.java:1149

20/02/29 00:56:46 INFO codegen.CodeGenerator: Code generated in 22.586175 ms

20/02/29 00:56:46 INFO spark.SparkContext: Starting job: count at RepjoinDS.scala:48

20/02/29 00:56:46 INFO scheduler.DAGScheduler: Registering RDD 12 (count at RepjoinDS.scala:48) as input to shuffle 0

20/02/29 00:56:46 INFO scheduler.DAGScheduler: Got job 2 (count at RepjoinDS.scala:48) with 1 output partitions

20/02/29 00:56:46 INFO scheduler.DAGScheduler: Final stage: ResultStage 3 (count at RepjoinDS.scala:48)

20/02/29 00:56:46 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleMapStage 2)

20/02/29 00:56:46 INFO scheduler.DAGScheduler: Missing parents: List(ShuffleMapStage 2)

20/02/29 00:56:46 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 2 (MapPartitionsRDD[12] at count at RepjoinDS.scala:48), which has no missing parents

20/02/29 00:56:46 INFO memory.MemoryStore: Block broadcast\_5 stored as values in memory (estimated size 16.8 KB, free 350.4 MB)

20/02/29 00:56:46 INFO memory.MemoryStore: Block broadcast\_5\_piece0 stored as bytes in memory (estimated size 7.5 KB, free 350.4 MB)

20/02/29 00:56:46 INFO storage.BlockManagerInfo: Added broadcast\_5\_piece0 in memory on 10.110.20.230:57896 (size: 7.5 KB, free: 362.9 MB)

20/02/29 00:56:46 INFO spark.SparkContext: Created broadcast 5 from broadcast at DAGScheduler.scala:1163

20/02/29 00:56:46 INFO scheduler.DAGScheduler: Submitting 40 missing tasks from ShuffleMapStage 2 (MapPartitionsRDD[12] at count at RepjoinDS.scala:48) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))

20/02/29 00:56:46 INFO scheduler.TaskSchedulerImpl: Adding task set 2.0 with 40 tasks

20/02/29 00:56:46 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 2.0 (TID 80, localhost, executor driver, partition 0, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:46 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 2.0 (TID 81, localhost, executor driver, partition 1, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:46 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 2.0 (TID 82, localhost, executor driver, partition 2, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:46 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 2.0 (TID 83, localhost, executor driver, partition 3, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:46 INFO executor.Executor: Running task 0.0 in stage 2.0 (TID 80)

20/02/29 00:56:46 INFO executor.Executor: Running task 1.0 in stage 2.0 (TID 81)

20/02/29 00:56:46 INFO executor.Executor: Running task 2.0 in stage 2.0 (TID 82)

20/02/29 00:56:46 INFO executor.Executor: Running task 3.0 in stage 2.0 (TID 83)

20/02/29 00:56:46 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:33554432+33554432

20/02/29 00:56:46 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:100663296+33554432

20/02/29 00:56:46 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:0+33554432

20/02/29 00:56:46 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:67108864+33554432

20/02/29 00:56:48 INFO executor.Executor: Finished task 3.0 in stage 2.0 (TID 83). 1766 bytes result sent to driver

20/02/29 00:56:48 INFO executor.Executor: Finished task 2.0 in stage 2.0 (TID 82). 1766 bytes result sent to driver

20/02/29 00:56:48 INFO scheduler.TaskSetManager: Starting task 4.0 in stage 2.0 (TID 84, localhost, executor driver, partition 4, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:48 INFO executor.Executor: Running task 4.0 in stage 2.0 (TID 84)

20/02/29 00:56:48 INFO scheduler.TaskSetManager: Starting task 5.0 in stage 2.0 (TID 85, localhost, executor driver, partition 5, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:48 INFO executor.Executor: Running task 5.0 in stage 2.0 (TID 85)

20/02/29 00:56:48 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 2.0 (TID 83) in 1314 ms on localhost (executor driver) (1/40)

20/02/29 00:56:48 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 2.0 (TID 82) in 1314 ms on localhost (executor driver) (2/40)

20/02/29 00:56:48 INFO executor.Executor: Finished task 1.0 in stage 2.0 (TID 81). 1766 bytes result sent to driver

20/02/29 00:56:48 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 2.0 (TID 86, localhost, executor driver, partition 6, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:48 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:167772160+33554432

20/02/29 00:56:48 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 2.0 (TID 81) in 1317 ms on localhost (executor driver) (3/40)

20/02/29 00:56:48 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:134217728+33554432

20/02/29 00:56:48 INFO executor.Executor: Running task 6.0 in stage 2.0 (TID 86)

20/02/29 00:56:48 INFO executor.Executor: Finished task 0.0 in stage 2.0 (TID 80). 1766 bytes result sent to driver

20/02/29 00:56:48 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 2.0 (TID 87, localhost, executor driver, partition 7, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:48 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 2.0 (TID 80) in 1321 ms on localhost (executor driver) (4/40)

20/02/29 00:56:48 INFO executor.Executor: Running task 7.0 in stage 2.0 (TID 87)

20/02/29 00:56:48 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:201326592+33554432

20/02/29 00:56:48 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:234881024+33554432

20/02/29 00:56:49 INFO executor.Executor: Finished task 7.0 in stage 2.0 (TID 87). 1766 bytes result sent to driver

20/02/29 00:56:49 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 2.0 (TID 88, localhost, executor driver, partition 8, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:49 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 2.0 (TID 87) in 1284 ms on localhost (executor driver) (5/40)

20/02/29 00:56:49 INFO executor.Executor: Running task 8.0 in stage 2.0 (TID 88)

20/02/29 00:56:49 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:268435456+33554432

20/02/29 00:56:49 INFO executor.Executor: Finished task 6.0 in stage 2.0 (TID 86). 1766 bytes result sent to driver

20/02/29 00:56:49 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 2.0 (TID 89, localhost, executor driver, partition 9, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:49 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 2.0 (TID 86) in 1313 ms on localhost (executor driver) (6/40)

20/02/29 00:56:49 INFO executor.Executor: Running task 9.0 in stage 2.0 (TID 89)

20/02/29 00:56:49 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:301989888+33554432

20/02/29 00:56:49 INFO executor.Executor: Finished task 4.0 in stage 2.0 (TID 84). 1766 bytes result sent to driver

20/02/29 00:56:49 INFO scheduler.TaskSetManager: Starting task 10.0 in stage 2.0 (TID 90, localhost, executor driver, partition 10, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:49 INFO scheduler.TaskSetManager: Finished task 4.0 in stage 2.0 (TID 84) in 1336 ms on localhost (executor driver) (7/40)

20/02/29 00:56:49 INFO executor.Executor: Running task 10.0 in stage 2.0 (TID 90)

20/02/29 00:56:49 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:335544320+33554432

20/02/29 00:56:49 INFO executor.Executor: Finished task 5.0 in stage 2.0 (TID 85). 1766 bytes result sent to driver

20/02/29 00:56:49 INFO scheduler.TaskSetManager: Starting task 11.0 in stage 2.0 (TID 91, localhost, executor driver, partition 11, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:49 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 2.0 (TID 85) in 1338 ms on localhost (executor driver) (8/40)

20/02/29 00:56:49 INFO executor.Executor: Running task 11.0 in stage 2.0 (TID 91)

20/02/29 00:56:49 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:369098752+33554432

20/02/29 00:56:50 INFO executor.Executor: Finished task 8.0 in stage 2.0 (TID 88). 1809 bytes result sent to driver

20/02/29 00:56:50 INFO scheduler.TaskSetManager: Starting task 12.0 in stage 2.0 (TID 92, localhost, executor driver, partition 12, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:50 INFO scheduler.TaskSetManager: Finished task 8.0 in stage 2.0 (TID 88) in 1334 ms on localhost (executor driver) (9/40)

20/02/29 00:56:50 INFO executor.Executor: Running task 12.0 in stage 2.0 (TID 92)

20/02/29 00:56:50 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:402653184+33554432

20/02/29 00:56:50 INFO executor.Executor: Finished task 9.0 in stage 2.0 (TID 89). 1766 bytes result sent to driver

20/02/29 00:56:50 INFO scheduler.TaskSetManager: Starting task 13.0 in stage 2.0 (TID 93, localhost, executor driver, partition 13, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:50 INFO executor.Executor: Running task 13.0 in stage 2.0 (TID 93)

20/02/29 00:56:50 INFO scheduler.TaskSetManager: Finished task 9.0 in stage 2.0 (TID 89) in 1335 ms on localhost (executor driver) (10/40)

20/02/29 00:56:50 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:436207616+33554432

20/02/29 00:56:50 INFO executor.Executor: Finished task 10.0 in stage 2.0 (TID 90). 1766 bytes result sent to driver

20/02/29 00:56:50 INFO scheduler.TaskSetManager: Starting task 14.0 in stage 2.0 (TID 94, localhost, executor driver, partition 14, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:50 INFO scheduler.TaskSetManager: Finished task 10.0 in stage 2.0 (TID 90) in 1363 ms on localhost (executor driver) (11/40)

20/02/29 00:56:50 INFO executor.Executor: Running task 14.0 in stage 2.0 (TID 94)

20/02/29 00:56:50 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:469762048+33554432

20/02/29 00:56:50 INFO executor.Executor: Finished task 11.0 in stage 2.0 (TID 91). 1766 bytes result sent to driver

20/02/29 00:56:50 INFO scheduler.TaskSetManager: Starting task 15.0 in stage 2.0 (TID 95, localhost, executor driver, partition 15, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:50 INFO executor.Executor: Running task 15.0 in stage 2.0 (TID 95)

20/02/29 00:56:50 INFO scheduler.TaskSetManager: Finished task 11.0 in stage 2.0 (TID 91) in 1380 ms on localhost (executor driver) (12/40)

20/02/29 00:56:50 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:503316480+33554432

20/02/29 00:56:51 INFO executor.Executor: Finished task 12.0 in stage 2.0 (TID 92). 1766 bytes result sent to driver

20/02/29 00:56:51 INFO scheduler.TaskSetManager: Starting task 16.0 in stage 2.0 (TID 96, localhost, executor driver, partition 16, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:51 INFO executor.Executor: Running task 16.0 in stage 2.0 (TID 96)

20/02/29 00:56:51 INFO scheduler.TaskSetManager: Finished task 12.0 in stage 2.0 (TID 92) in 1365 ms on localhost (executor driver) (13/40)

20/02/29 00:56:51 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:536870912+33554432

20/02/29 00:56:52 INFO executor.Executor: Finished task 15.0 in stage 2.0 (TID 95). 1766 bytes result sent to driver

20/02/29 00:56:52 INFO scheduler.TaskSetManager: Starting task 17.0 in stage 2.0 (TID 97, localhost, executor driver, partition 17, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:52 INFO executor.Executor: Running task 17.0 in stage 2.0 (TID 97)

20/02/29 00:56:52 INFO scheduler.TaskSetManager: Finished task 15.0 in stage 2.0 (TID 95) in 1316 ms on localhost (executor driver) (14/40)

20/02/29 00:56:52 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:570425344+33554432

20/02/29 00:56:52 INFO executor.Executor: Finished task 13.0 in stage 2.0 (TID 93). 1766 bytes result sent to driver

20/02/29 00:56:52 INFO scheduler.TaskSetManager: Starting task 18.0 in stage 2.0 (TID 98, localhost, executor driver, partition 18, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:52 INFO scheduler.TaskSetManager: Finished task 13.0 in stage 2.0 (TID 93) in 1400 ms on localhost (executor driver) (15/40)

20/02/29 00:56:52 INFO executor.Executor: Running task 18.0 in stage 2.0 (TID 98)

20/02/29 00:56:52 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:603979776+33554432

20/02/29 00:56:52 INFO executor.Executor: Finished task 14.0 in stage 2.0 (TID 94). 1809 bytes result sent to driver

20/02/29 00:56:52 INFO scheduler.TaskSetManager: Starting task 19.0 in stage 2.0 (TID 99, localhost, executor driver, partition 19, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:52 INFO executor.Executor: Running task 19.0 in stage 2.0 (TID 99)

20/02/29 00:56:52 INFO scheduler.TaskSetManager: Finished task 14.0 in stage 2.0 (TID 94) in 1412 ms on localhost (executor driver) (16/40)

20/02/29 00:56:52 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:637534208+33554432

20/02/29 00:56:53 INFO executor.Executor: Finished task 16.0 in stage 2.0 (TID 96). 1766 bytes result sent to driver

20/02/29 00:56:53 INFO scheduler.TaskSetManager: Starting task 20.0 in stage 2.0 (TID 100, localhost, executor driver, partition 20, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:53 INFO executor.Executor: Running task 20.0 in stage 2.0 (TID 100)

20/02/29 00:56:53 INFO scheduler.TaskSetManager: Finished task 16.0 in stage 2.0 (TID 96) in 1294 ms on localhost (executor driver) (17/40)

20/02/29 00:56:53 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:671088640+33554432

20/02/29 00:56:53 INFO executor.Executor: Finished task 17.0 in stage 2.0 (TID 97). 1766 bytes result sent to driver

20/02/29 00:56:53 INFO scheduler.TaskSetManager: Starting task 21.0 in stage 2.0 (TID 101, localhost, executor driver, partition 21, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:53 INFO executor.Executor: Running task 21.0 in stage 2.0 (TID 101)

20/02/29 00:56:53 INFO scheduler.TaskSetManager: Finished task 17.0 in stage 2.0 (TID 97) in 1284 ms on localhost (executor driver) (18/40)

20/02/29 00:56:53 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:704643072+33554432

20/02/29 00:56:53 INFO executor.Executor: Finished task 18.0 in stage 2.0 (TID 98). 1766 bytes result sent to driver

20/02/29 00:56:53 INFO scheduler.TaskSetManager: Starting task 22.0 in stage 2.0 (TID 102, localhost, executor driver, partition 22, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:53 INFO executor.Executor: Running task 22.0 in stage 2.0 (TID 102)

20/02/29 00:56:53 INFO scheduler.TaskSetManager: Finished task 18.0 in stage 2.0 (TID 98) in 1314 ms on localhost (executor driver) (19/40)

20/02/29 00:56:53 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:738197504+33554432

20/02/29 00:56:53 INFO executor.Executor: Finished task 19.0 in stage 2.0 (TID 99). 1766 bytes result sent to driver

20/02/29 00:56:53 INFO scheduler.TaskSetManager: Starting task 23.0 in stage 2.0 (TID 103, localhost, executor driver, partition 23, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:53 INFO scheduler.TaskSetManager: Finished task 19.0 in stage 2.0 (TID 99) in 1385 ms on localhost (executor driver) (20/40)

20/02/29 00:56:53 INFO executor.Executor: Running task 23.0 in stage 2.0 (TID 103)

20/02/29 00:56:53 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:771751936+33554432

20/02/29 00:56:54 INFO executor.Executor: Finished task 20.0 in stage 2.0 (TID 100). 1766 bytes result sent to driver

20/02/29 00:56:54 INFO scheduler.TaskSetManager: Starting task 24.0 in stage 2.0 (TID 104, localhost, executor driver, partition 24, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:54 INFO executor.Executor: Running task 24.0 in stage 2.0 (TID 104)

20/02/29 00:56:54 INFO scheduler.TaskSetManager: Finished task 20.0 in stage 2.0 (TID 100) in 1308 ms on localhost (executor driver) (21/40)

20/02/29 00:56:54 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:805306368+33554432

20/02/29 00:56:54 INFO executor.Executor: Finished task 21.0 in stage 2.0 (TID 101). 1766 bytes result sent to driver

20/02/29 00:56:54 INFO scheduler.TaskSetManager: Starting task 25.0 in stage 2.0 (TID 105, localhost, executor driver, partition 25, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:54 INFO executor.Executor: Running task 25.0 in stage 2.0 (TID 105)

20/02/29 00:56:54 INFO scheduler.TaskSetManager: Finished task 21.0 in stage 2.0 (TID 101) in 1332 ms on localhost (executor driver) (22/40)

20/02/29 00:56:54 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:838860800+33554432

20/02/29 00:56:54 INFO executor.Executor: Finished task 23.0 in stage 2.0 (TID 103). 1766 bytes result sent to driver

20/02/29 00:56:54 INFO scheduler.TaskSetManager: Starting task 26.0 in stage 2.0 (TID 106, localhost, executor driver, partition 26, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:54 INFO executor.Executor: Running task 26.0 in stage 2.0 (TID 106)

20/02/29 00:56:54 INFO scheduler.TaskSetManager: Finished task 23.0 in stage 2.0 (TID 103) in 1362 ms on localhost (executor driver) (23/40)

20/02/29 00:56:54 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:872415232+33554432

20/02/29 00:56:55 INFO executor.Executor: Finished task 24.0 in stage 2.0 (TID 104). 1766 bytes result sent to driver

20/02/29 00:56:55 INFO scheduler.TaskSetManager: Starting task 27.0 in stage 2.0 (TID 107, localhost, executor driver, partition 27, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:55 INFO executor.Executor: Running task 27.0 in stage 2.0 (TID 107)

20/02/29 00:56:55 INFO scheduler.TaskSetManager: Finished task 24.0 in stage 2.0 (TID 104) in 1291 ms on localhost (executor driver) (24/40)

20/02/29 00:56:55 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:905969664+33554432

20/02/29 00:56:55 INFO executor.Executor: Finished task 25.0 in stage 2.0 (TID 105). 1766 bytes result sent to driver

20/02/29 00:56:55 INFO scheduler.TaskSetManager: Starting task 28.0 in stage 2.0 (TID 108, localhost, executor driver, partition 28, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:55 INFO executor.Executor: Running task 28.0 in stage 2.0 (TID 108)

20/02/29 00:56:55 INFO scheduler.TaskSetManager: Finished task 25.0 in stage 2.0 (TID 105) in 1315 ms on localhost (executor driver) (25/40)

20/02/29 00:56:55 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:939524096+33554432

20/02/29 00:56:56 INFO executor.Executor: Finished task 26.0 in stage 2.0 (TID 106). 1766 bytes result sent to driver

20/02/29 00:56:56 INFO scheduler.TaskSetManager: Starting task 29.0 in stage 2.0 (TID 109, localhost, executor driver, partition 29, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:56 INFO scheduler.TaskSetManager: Finished task 26.0 in stage 2.0 (TID 106) in 1321 ms on localhost (executor driver) (26/40)

20/02/29 00:56:56 INFO executor.Executor: Running task 29.0 in stage 2.0 (TID 109)

20/02/29 00:56:56 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:973078528+33554432

20/02/29 00:56:57 INFO executor.Executor: Finished task 27.0 in stage 2.0 (TID 107). 1766 bytes result sent to driver

20/02/29 00:56:57 INFO scheduler.TaskSetManager: Starting task 30.0 in stage 2.0 (TID 110, localhost, executor driver, partition 30, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:57 INFO scheduler.TaskSetManager: Finished task 27.0 in stage 2.0 (TID 107) in 1310 ms on localhost (executor driver) (27/40)

20/02/29 00:56:57 INFO executor.Executor: Running task 30.0 in stage 2.0 (TID 110)

20/02/29 00:56:57 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1006632960+33554432

20/02/29 00:56:57 INFO executor.Executor: Finished task 28.0 in stage 2.0 (TID 108). 1766 bytes result sent to driver

20/02/29 00:56:57 INFO scheduler.TaskSetManager: Starting task 31.0 in stage 2.0 (TID 111, localhost, executor driver, partition 31, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:57 INFO executor.Executor: Running task 31.0 in stage 2.0 (TID 111)

20/02/29 00:56:57 INFO scheduler.TaskSetManager: Finished task 28.0 in stage 2.0 (TID 108) in 1275 ms on localhost (executor driver) (28/40)

20/02/29 00:56:57 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1040187392+33554432

20/02/29 00:56:57 INFO executor.Executor: Finished task 29.0 in stage 2.0 (TID 109). 1723 bytes result sent to driver

20/02/29 00:56:57 INFO scheduler.TaskSetManager: Starting task 32.0 in stage 2.0 (TID 112, localhost, executor driver, partition 32, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:57 INFO executor.Executor: Running task 32.0 in stage 2.0 (TID 112)

20/02/29 00:56:57 INFO scheduler.TaskSetManager: Finished task 29.0 in stage 2.0 (TID 109) in 1295 ms on localhost (executor driver) (29/40)

20/02/29 00:56:57 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1073741824+33554432

20/02/29 00:56:57 INFO executor.Executor: Finished task 22.0 in stage 2.0 (TID 102). 1766 bytes result sent to driver

20/02/29 00:56:57 INFO scheduler.TaskSetManager: Starting task 33.0 in stage 2.0 (TID 113, localhost, executor driver, partition 33, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:57 INFO scheduler.TaskSetManager: Finished task 22.0 in stage 2.0 (TID 102) in 4366 ms on localhost (executor driver) (30/40)

20/02/29 00:56:57 INFO executor.Executor: Running task 33.0 in stage 2.0 (TID 113)

20/02/29 00:56:57 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1107296256+33554432

20/02/29 00:56:58 INFO executor.Executor: Finished task 30.0 in stage 2.0 (TID 110). 1766 bytes result sent to driver

20/02/29 00:56:58 INFO scheduler.TaskSetManager: Starting task 34.0 in stage 2.0 (TID 114, localhost, executor driver, partition 34, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:58 INFO executor.Executor: Running task 34.0 in stage 2.0 (TID 114)

20/02/29 00:56:58 INFO scheduler.TaskSetManager: Finished task 30.0 in stage 2.0 (TID 110) in 1413 ms on localhost (executor driver) (31/40)

20/02/29 00:56:58 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1140850688+33554432

20/02/29 00:56:58 INFO executor.Executor: Finished task 31.0 in stage 2.0 (TID 111). 1766 bytes result sent to driver

20/02/29 00:56:58 INFO scheduler.TaskSetManager: Starting task 35.0 in stage 2.0 (TID 115, localhost, executor driver, partition 35, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:58 INFO scheduler.TaskSetManager: Finished task 31.0 in stage 2.0 (TID 111) in 1463 ms on localhost (executor driver) (32/40)

20/02/29 00:56:58 INFO executor.Executor: Running task 35.0 in stage 2.0 (TID 115)

20/02/29 00:56:58 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1174405120+33554432

20/02/29 00:56:58 INFO executor.Executor: Finished task 32.0 in stage 2.0 (TID 112). 1766 bytes result sent to driver

20/02/29 00:56:58 INFO scheduler.TaskSetManager: Starting task 36.0 in stage 2.0 (TID 116, localhost, executor driver, partition 36, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:58 INFO executor.Executor: Running task 36.0 in stage 2.0 (TID 116)

20/02/29 00:56:58 INFO scheduler.TaskSetManager: Finished task 32.0 in stage 2.0 (TID 112) in 1453 ms on localhost (executor driver) (33/40)

20/02/29 00:56:58 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1207959552+33554432

20/02/29 00:56:59 INFO executor.Executor: Finished task 33.0 in stage 2.0 (TID 113). 1809 bytes result sent to driver

20/02/29 00:56:59 INFO scheduler.TaskSetManager: Starting task 37.0 in stage 2.0 (TID 117, localhost, executor driver, partition 37, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:59 INFO scheduler.TaskSetManager: Finished task 33.0 in stage 2.0 (TID 113) in 1417 ms on localhost (executor driver) (34/40)

20/02/29 00:56:59 INFO executor.Executor: Running task 37.0 in stage 2.0 (TID 117)

20/02/29 00:56:59 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1241513984+33554432

20/02/29 00:56:59 INFO executor.Executor: Finished task 34.0 in stage 2.0 (TID 114). 1766 bytes result sent to driver

20/02/29 00:56:59 INFO scheduler.TaskSetManager: Starting task 38.0 in stage 2.0 (TID 118, localhost, executor driver, partition 38, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:59 INFO scheduler.TaskSetManager: Finished task 34.0 in stage 2.0 (TID 114) in 1277 ms on localhost (executor driver) (35/40)

20/02/29 00:56:59 INFO executor.Executor: Running task 38.0 in stage 2.0 (TID 118)

20/02/29 00:56:59 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1275068416+33554432

20/02/29 00:56:59 INFO executor.Executor: Finished task 35.0 in stage 2.0 (TID 115). 1766 bytes result sent to driver

20/02/29 00:56:59 INFO scheduler.TaskSetManager: Starting task 39.0 in stage 2.0 (TID 119, localhost, executor driver, partition 39, PROCESS\_LOCAL, 7452 bytes)

20/02/29 00:56:59 INFO scheduler.TaskSetManager: Finished task 35.0 in stage 2.0 (TID 115) in 1272 ms on localhost (executor driver) (36/40)

20/02/29 00:56:59 INFO executor.Executor: Running task 39.0 in stage 2.0 (TID 119)

20/02/29 00:56:59 INFO rdd.HadoopRDD: Input split: file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/input/edges.csv:1308622848+10658977

20/02/29 00:57:00 INFO executor.Executor: Finished task 36.0 in stage 2.0 (TID 116). 1766 bytes result sent to driver

20/02/29 00:57:00 INFO scheduler.TaskSetManager: Finished task 36.0 in stage 2.0 (TID 116) in 1260 ms on localhost (executor driver) (37/40)

20/02/29 00:57:00 INFO executor.Executor: Finished task 39.0 in stage 2.0 (TID 119). 1723 bytes result sent to driver

20/02/29 00:57:00 INFO scheduler.TaskSetManager: Finished task 39.0 in stage 2.0 (TID 119) in 390 ms on localhost (executor driver) (38/40)

20/02/29 00:57:00 INFO executor.Executor: Finished task 37.0 in stage 2.0 (TID 117). 1766 bytes result sent to driver

20/02/29 00:57:00 INFO scheduler.TaskSetManager: Finished task 37.0 in stage 2.0 (TID 117) in 1239 ms on localhost (executor driver) (39/40)

20/02/29 00:57:00 INFO executor.Executor: Finished task 38.0 in stage 2.0 (TID 118). 1766 bytes result sent to driver

20/02/29 00:57:00 INFO scheduler.TaskSetManager: Finished task 38.0 in stage 2.0 (TID 118) in 1066 ms on localhost (executor driver) (40/40)

20/02/29 00:57:00 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 2.0, whose tasks have all completed, from pool

20/02/29 00:57:00 INFO scheduler.DAGScheduler: ShuffleMapStage 2 (count at RepjoinDS.scala:48) finished in 14.279 s

20/02/29 00:57:00 INFO scheduler.DAGScheduler: looking for newly runnable stages

20/02/29 00:57:00 INFO scheduler.DAGScheduler: running: Set()

20/02/29 00:57:00 INFO scheduler.DAGScheduler: waiting: Set(ResultStage 3)

20/02/29 00:57:00 INFO scheduler.DAGScheduler: failed: Set()

20/02/29 00:57:00 INFO scheduler.DAGScheduler: Submitting ResultStage 3 (MapPartitionsRDD[15] at count at RepjoinDS.scala:48), which has no missing parents

20/02/29 00:57:00 INFO memory.MemoryStore: Block broadcast\_6 stored as values in memory (estimated size 8.6 KB, free 350.4 MB)

20/02/29 00:57:00 INFO memory.MemoryStore: Block broadcast\_6\_piece0 stored as bytes in memory (estimated size 4.4 KB, free 350.4 MB)

20/02/29 00:57:00 INFO storage.BlockManagerInfo: Added broadcast\_6\_piece0 in memory on 10.110.20.230:57896 (size: 4.4 KB, free: 362.9 MB)

20/02/29 00:57:00 INFO spark.SparkContext: Created broadcast 6 from broadcast at DAGScheduler.scala:1163

20/02/29 00:57:00 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 3 (MapPartitionsRDD[15] at count at RepjoinDS.scala:48) (first 15 tasks are for partitions Vector(0))

20/02/29 00:57:00 INFO scheduler.TaskSchedulerImpl: Adding task set 3.0 with 1 tasks

20/02/29 00:57:00 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 3.0 (TID 120, localhost, executor driver, partition 0, ANY, 7246 bytes)

20/02/29 00:57:00 INFO executor.Executor: Running task 0.0 in stage 3.0 (TID 120)

20/02/29 00:57:00 INFO storage.ShuffleBlockFetcherIterator: Getting 40 non-empty blocks including 40 local blocks and 0 remote blocks

20/02/29 00:57:00 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 6 ms

20/02/29 00:57:01 INFO executor.Executor: Finished task 0.0 in stage 3.0 (TID 120). 1749 bytes result sent to driver

20/02/29 00:57:01 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 3.0 (TID 120) in 65 ms on localhost (executor driver) (1/1)

20/02/29 00:57:01 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool

20/02/29 00:57:01 INFO scheduler.DAGScheduler: ResultStage 3 (count at RepjoinDS.scala:48) finished in 0.072 s

20/02/29 00:57:01 INFO scheduler.DAGScheduler: Job 2 finished: count at RepjoinDS.scala:48, took 14.374921 s

20/02/29 00:57:01 INFO Configuration.deprecation: mapred.output.dir is deprecated. Instead, use mapreduce.output.fileoutputformat.outputdir

20/02/29 00:57:01 INFO io.HadoopMapRedCommitProtocol: Using output committer class org.apache.hadoop.mapred.FileOutputCommitter

20/02/29 00:57:01 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

20/02/29 00:57:01 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup \_temporary folders under output directory:false, ignore cleanup failures: false

20/02/29 00:57:01 INFO spark.SparkContext: Starting job: runJob at SparkHadoopWriter.scala:78

20/02/29 00:57:01 INFO scheduler.DAGScheduler: Got job 3 (runJob at SparkHadoopWriter.scala:78) with 4 output partitions

20/02/29 00:57:01 INFO scheduler.DAGScheduler: Final stage: ResultStage 4 (runJob at SparkHadoopWriter.scala:78)

20/02/29 00:57:01 INFO scheduler.DAGScheduler: Parents of final stage: List()

20/02/29 00:57:01 INFO scheduler.DAGScheduler: Missing parents: List()

20/02/29 00:57:01 INFO scheduler.DAGScheduler: Submitting ResultStage 4 (MapPartitionsRDD[17] at saveAsTextFile at RepjoinDS.scala:52), which has no missing parents

20/02/29 00:57:01 INFO memory.MemoryStore: Block broadcast\_7 stored as values in memory (estimated size 85.2 KB, free 350.3 MB)

20/02/29 00:57:01 INFO memory.MemoryStore: Block broadcast\_7\_piece0 stored as bytes in memory (estimated size 30.8 KB, free 350.3 MB)

20/02/29 00:57:01 INFO storage.BlockManagerInfo: Added broadcast\_7\_piece0 in memory on 10.110.20.230:57896 (size: 30.8 KB, free: 362.8 MB)

20/02/29 00:57:01 INFO spark.SparkContext: Created broadcast 7 from broadcast at DAGScheduler.scala:1163

20/02/29 00:57:01 INFO scheduler.DAGScheduler: Submitting 4 missing tasks from ResultStage 4 (MapPartitionsRDD[17] at saveAsTextFile at RepjoinDS.scala:52) (first 15 tasks are for partitions Vector(0, 1, 2, 3))

20/02/29 00:57:01 INFO scheduler.TaskSchedulerImpl: Adding task set 4.0 with 4 tasks

20/02/29 00:57:01 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 4.0 (TID 121, localhost, executor driver, partition 0, PROCESS\_LOCAL, 7347 bytes)

20/02/29 00:57:01 INFO scheduler.TaskSetManager: Starting task 1.0 in stage 4.0 (TID 122, localhost, executor driver, partition 1, PROCESS\_LOCAL, 7365 bytes)

20/02/29 00:57:01 INFO scheduler.TaskSetManager: Starting task 2.0 in stage 4.0 (TID 123, localhost, executor driver, partition 2, PROCESS\_LOCAL, 7347 bytes)

20/02/29 00:57:01 INFO scheduler.TaskSetManager: Starting task 3.0 in stage 4.0 (TID 124, localhost, executor driver, partition 3, PROCESS\_LOCAL, 7398 bytes)

20/02/29 00:57:01 INFO executor.Executor: Running task 2.0 in stage 4.0 (TID 123)

20/02/29 00:57:01 INFO executor.Executor: Running task 0.0 in stage 4.0 (TID 121)

20/02/29 00:57:01 INFO executor.Executor: Running task 3.0 in stage 4.0 (TID 124)

20/02/29 00:57:01 INFO executor.Executor: Running task 1.0 in stage 4.0 (TID 122)

20/02/29 00:57:01 INFO io.HadoopMapRedCommitProtocol: Using output committer class org.apache.hadoop.mapred.FileOutputCommitter

20/02/29 00:57:01 INFO io.HadoopMapRedCommitProtocol: Using output committer class org.apache.hadoop.mapred.FileOutputCommitter

20/02/29 00:57:01 INFO io.HadoopMapRedCommitProtocol: Using output committer class org.apache.hadoop.mapred.FileOutputCommitter

20/02/29 00:57:01 INFO io.HadoopMapRedCommitProtocol: Using output committer class org.apache.hadoop.mapred.FileOutputCommitter

20/02/29 00:57:01 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

20/02/29 00:57:01 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

20/02/29 00:57:01 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup \_temporary folders under output directory:false, ignore cleanup failures: false

20/02/29 00:57:01 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

20/02/29 00:57:01 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1

20/02/29 00:57:01 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup \_temporary folders under output directory:false, ignore cleanup failures: false

20/02/29 00:57:01 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup \_temporary folders under output directory:false, ignore cleanup failures: false

20/02/29 00:57:01 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup \_temporary folders under output directory:false, ignore cleanup failures: false

20/02/29 00:57:01 INFO output.FileOutputCommitter: Saved output of task 'attempt\_20200229005701\_0017\_m\_000000\_0' to file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/output/\_temporary/0/task\_20200229005701\_0017\_m\_000000

20/02/29 00:57:01 INFO output.FileOutputCommitter: Saved output of task 'attempt\_20200229005701\_0017\_m\_000001\_0' to file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/output/\_temporary/0/task\_20200229005701\_0017\_m\_000001

20/02/29 00:57:01 INFO output.FileOutputCommitter: Saved output of task 'attempt\_20200229005701\_0017\_m\_000002\_0' to file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/output/\_temporary/0/task\_20200229005701\_0017\_m\_000002

20/02/29 00:57:01 INFO output.FileOutputCommitter: Saved output of task 'attempt\_20200229005701\_0017\_m\_000003\_0' to file:/Users/snehgurdasani/Documents/CS4240\_Large\_Scale\_Data\_Processing/Homeworks/homework3/Spark-Demo-Rep-DS/output/\_temporary/0/task\_20200229005701\_0017\_m\_000003

20/02/29 00:57:01 INFO mapred.SparkHadoopMapRedUtil: attempt\_20200229005701\_0017\_m\_000000\_0: Committed

20/02/29 00:57:01 INFO mapred.SparkHadoopMapRedUtil: attempt\_20200229005701\_0017\_m\_000002\_0: Committed

20/02/29 00:57:01 INFO mapred.SparkHadoopMapRedUtil: attempt\_20200229005701\_0017\_m\_000001\_0: Committed

20/02/29 00:57:01 INFO mapred.SparkHadoopMapRedUtil: attempt\_20200229005701\_0017\_m\_000003\_0: Committed

20/02/29 00:57:01 INFO executor.Executor: Finished task 3.0 in stage 4.0 (TID 124). 1035 bytes result sent to driver

20/02/29 00:57:01 INFO executor.Executor: Finished task 1.0 in stage 4.0 (TID 122). 1035 bytes result sent to driver

20/02/29 00:57:01 INFO executor.Executor: Finished task 2.0 in stage 4.0 (TID 123). 949 bytes result sent to driver

20/02/29 00:57:01 INFO executor.Executor: Finished task 0.0 in stage 4.0 (TID 121). 949 bytes result sent to driver

20/02/29 00:57:01 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 4.0 (TID 124) in 69 ms on localhost (executor driver) (1/4)

20/02/29 00:57:01 INFO scheduler.TaskSetManager: Finished task 1.0 in stage 4.0 (TID 122) in 69 ms on localhost (executor driver) (2/4)

20/02/29 00:57:01 INFO scheduler.TaskSetManager: Finished task 2.0 in stage 4.0 (TID 123) in 69 ms on localhost (executor driver) (3/4)

20/02/29 00:57:01 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 4.0 (TID 121) in 84 ms on localhost (executor driver) (4/4)

20/02/29 00:57:01 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 4.0, whose tasks have all completed, from pool

20/02/29 00:57:01 INFO scheduler.DAGScheduler: ResultStage 4 (runJob at SparkHadoopWriter.scala:78) finished in 0.102 s

20/02/29 00:57:01 INFO scheduler.DAGScheduler: Job 3 finished: runJob at SparkHadoopWriter.scala:78, took 0.104679 s

20/02/29 00:57:01 INFO io.SparkHadoopWriter: Job job\_20200229005701\_0017 committed.

20/02/29 00:57:01 INFO spark.SparkContext: Invoking stop() from shutdown hook

20/02/29 00:57:01 INFO server.AbstractConnector: Stopped Spark@7dac3fd8{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}

20/02/29 00:57:01 INFO ui.SparkUI: Stopped Spark web UI at http://10.110.20.230:4040

20/02/29 00:57:01 INFO spark.MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!

20/02/29 00:57:01 INFO memory.MemoryStore: MemoryStore cleared

20/02/29 00:57:01 INFO storage.BlockManager: BlockManager stopped

20/02/29 00:57:01 INFO storage.BlockManagerMaster: BlockManagerMaster stopped

20/02/29 00:57:01 INFO scheduler.OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!

20/02/29 00:57:01 INFO spark.SparkContext: Successfully stopped SparkContext

20/02/29 00:57:01 INFO util.ShutdownHookManager: Shutdown hook called

20/02/29 00:57:01 INFO util.ShutdownHookManager: Deleting directory /private/var/folders/nj/f9l6vhs50h975cdq0gphfn500000gn/T/spark-6dd8c03c-e456-4363-8bd5-13fe598d3ed8

20/02/29 00:57:01 INFO util.ShutdownHookManager: Deleting directory /private/var/folders/nj/f9l6vhs50h975cdq0gphfn500000gn/T/spark-4de0b827-6d66-43e9-accb-ebeee4728bb1

(base) Snehs-MacBook-Pro:Spark-Demo-Rep-DS snehgurdasani$