

● ● ● CP caching-and-persistence main

Current File

⏮ ⚙ ⋮

👤 🔍 ⚙

air-quality-analysis [~/Documents/scala-project-2/scenario-1/air-qua]

Day17Task – WriteToKeyspaces.scala [Day17Task]

table-normalization – build.sbt [table-normalization]

caching-and-persistence – Main.scala [caching-and-persistence]

Project

⌵

caching-and-persistence ~/Docur

> .bsp

Main.scala

build.sbt

24

println("Without Caching:")

25

Main

Run

Main

Main

🔄

■

📷

📄

💬

⋮

↑

↓

⌵

⌴

📄

🗑

>

🔍 Elapsed

× ↶ Cc W .*

1/2

↑ ↓ 🔍 ⋮

×

| user_1000|2024-06-15T02:09:37Z|

|user_10000|2024-06-17T22:15:26Z|

| user_1001|2024-06-16T20:42:17Z|

| user_1002|2024-06-13T08:27:54Z|

| user_1004|2024-06-17T01:59:28Z|

| user_1005|2024-06-16T10:36:46Z|

| user_1008|2024-06-15T18:47:52Z|

| user_1009|2024-06-17T19:39:36Z|

| user_1010|2024-06-16T23:28:15Z|

| user_1013|2024-06-17T18:34:23Z|

| user_1016|2024-06-17T03:04:25Z|

| user_1018|2024-06-17T00:59:23Z|

| user_102|2024-06-17T21:07:38Z|

| user_1020|2024-06-17T06:47:02Z|

| user_1021|2024-06-16T21:11:32Z|

| user_1022|2024-06-17T08:42:54Z|

| user_1023|2024-06-12T19:17:55Z|

| user_1028|2024-06-17T15:48:43Z|

+-----+

only showing top 20 rows

Elapsed time: 8.12471475 seconds

24/06/20 11:51:06 INFO MapPartitionsRDD: Removing RDD 29 from persistence list

24/06/20 11:51:06 INFO BlockManager: Removing RDD 29

24/06/20 11:51:06 INFO SparkContext: Invoking stop() from shutdown hook

24/06/20 11:51:06 INFO SparkContext: SparkContext is stopping with exitCode 0.

24/06/20 11:51:06 INFO SparkUI: Stopped Spark web UI at http://17.76.52.140:4040

24/06/20 11:51:06 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!

24/06/20 11:51:06 INFO MemoryStore: MemoryStore cleared

24/06/20 11:51:06 INFO BlockManager: BlockManager stopped

24/06/20 11:51:06 INFO BlockManagerMaster: BlockManagerMaster stopped

24/06/20 11:51:06 INFO OutputCommitCoordinator\$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!

caching-and-persistence > src > main > scala > Main.scala

🔍 [T] ⌚ 27:61 LF UTF-8 2 spaces

● ● ● CP caching-and-persistence main

Current File

▶ ⚙️ ⋮

👤 🔍 ⚙️

air-quality-analysis [~/Documents/scala-project-2/scenario-1/air-qua]

Day17Task – WriteToKeyspaces.scala [Day17Task]

table-normalization – build.sbt [table-normalization]

caching-and-persistence – Main.scala [caching-and-persistence]

Project

▼ caching-and-persistence ~/Documents/scala-project-2/scenario-1/air-qua

> .bsp

> .idea

> project [caching-and-persistence]

▼ src

▼ main

▼ scala

⌚ Main

≡ user_logs.csv

> test

> target

🚫 .gitignore

📄 build.sbt

> External Libraries

> Scratches and Consoles

⌚ Main.scala × 📄 build.sbt

3 object Main extends App {

22 .csv(logFile)

23

24 println("Without Caching:")

25

26 time {

27 val actionsPerUserDF = logsDF.groupBy("user_id").count()

28 actionsPerUserDF.show()

29

30 val uniqueItemsPerUserDF = logsDF.groupBy("user_id").agg(F.countDistinct("item_id").alias("unique_items"))

31 uniqueItemsPerUserDF.show()

32

33 val mostRecentActionPerUserDF = logsDF.groupBy("user_id").agg(F.max("timestamp").alias("most_recent_action"))

34 mostRecentActionPerUserDF.show()

35 }

36

37 logsDF.cache()

38

39 time {

40 val actionsPerUserDFCache = logsDF.groupBy("user_id").count()

41 actionsPerUserDFCache.show()

42

43 val uniqueItemsPerUserDFCache = logsDF.groupBy("user_id").agg(F.countDistinct("item_id").alias("unique_items"))

44 uniqueItemsPerUserDFCache.show()

45

46 val mostRecentActionPerUserDFCache = logsDF.groupBy("user_id").agg(F.max("timestamp").alias("most_recent_action"))

47 mostRecentActionPerUserDFCache.show()

48 }

49

50 logsDF.unpersist()

51

Main

▶ Run Main × Main ×

🔄 📷 📄 🗨️ ⋮

↑ > 🔍 Elapsed

↓ only showing top 20 rows

>

📄 caching-and-persistence > src > main > scala > ⌚ Main.scala

🔗 [T] ⌚ 27:61 LF UTF-8 2 spaces 📄

● ● ● CP caching-and-persistence main

Current File

⏮ ⚙ ⋮

👤 🔍 ⚙

air-quality-analysis [~/Documents/scala-project-2/scenario-1/air-qua]

Day17Task – WriteToKeyspaces.scala [Day17Task]

table-normalization – build.sbt [table-normalization]

caching-and-persistence – Main.scala [caching-and-persistence]

Project

▼ caching-and-persistence ~/Docur

> .bsp

Main.scala × build.sbt

24 println("Without Caching:")

Main

Run Main Main

🔄 📷 ↗ 🗨 ⋮

↑ > 🔍 Elapsed

24/06/20 11:50:58 INFO DAGScheduler: Job 6 is finished. Cancelling potential speculative or zombie tasks for this job

24/06/20 11:50:58 INFO TaskSchedulerImpl: Killing all running tasks in stage 11: Stage finished

24/06/20 11:50:58 INFO DAGScheduler: Job 6 finished: show at Main.scala:34, took 0.251577 s

+-----+-----+

| user_id| most_recent_action|

+-----+-----+

| user_1|2024-06-14T06:25:50Z|

| user_10|2024-06-17T15:52:50Z|

| user_1000|2024-06-15T02:09:37Z|

|user_10000|2024-06-17T22:15:26Z|

| user_1001|2024-06-16T20:42:17Z|

| user_1002|2024-06-13T08:27:54Z|

| user_1004|2024-06-17T01:59:28Z|

| user_1005|2024-06-16T10:36:46Z|

| user_1008|2024-06-15T18:47:52Z|

| user_1009|2024-06-17T19:39:36Z|

| user_1010|2024-06-16T23:28:15Z|

| user_1013|2024-06-17T18:34:23Z|

| user_1016|2024-06-17T03:04:25Z|

| user_1018|2024-06-17T00:59:23Z|

| user_102|2024-06-17T21:07:38Z|

| user_1020|2024-06-17T06:47:02Z|

| user_1021|2024-06-16T21:11:32Z|

| user_1022|2024-06-17T08:42:54Z|

| user_1023|2024-06-12T19:17:55Z|

| user_1028|2024-06-17T15:48:43Z|

+-----+-----+

only showing top 20 rows

Elapsed time: 19.269400375 seconds

24/06/20 11:50:58 INFO FileSourceStrategy: Pushed Filters:

24/06/20 11:50:58 INFO FileSourceStrategy: Post-Scan Filters:

📄 🗑

🔧

📁

▶

⏮

⚙

🔍

👤

caching-and-persistence > src > main > scala > Main.scala

🔍 [T] ⌚ 27:61 LF UTF-8 2 spaces