Billing:

Billing

Job:

T=1:


```
23/03/21 01:51:21 INFO org.apache.spark.SparkEnv: Registering BlockManagerMaster
23/03/21 01:51:22 INFO org.apache.spark.SparkEnv: Registering OutputCommitCoordinator
23/03/21 01:51:22 INFO org.spark_project.jetty.util.log: Logging initialized @5664ms to org.spark_project.jetty.util.log.Slf4jLog
23/03/21 01:51:22 INFO org.spark_project.jetty.server.Server: jetty-9.4.z-SNAPSHOT; built: unknown; git: unknown; jvm 1.8.0_362-b09
23/03/21 01:51:22 INFO org.spark_project.jetty.server.Server: Started @5926ms
23/03/21 01:51:22 INFO org.spark_project.jetty.server.AbstractConnector: Started ServerConnector@78990400{HTTP/1.1, (http/1.1)}{0.0.0.0:44871}
23/03/21 01:51:26 INFO org.apache.hadoop.yarn.client.RMProxy: Connecting to ResourceManager at data301-2023-amh284-lab3-cluster-m/10.152.0.4:8032
23/03/21 01:51:26 INFO org.apache.hadoop.yarn.client.AHSProxy: Connecting to Application History server at data301-2023-amh284-lab3-cluster-m/10.152.0.4:10200
23/03/21 01:51:26 INFO org.apache.hadoop.conf.Configuration: resource-types.xml not found
23/03/21 01:51:26 INFO org.apache.hadoop.yarn.util.resource.ResourceUtils: Unable to find 'resource-types.xml'.
23/03/21 01:51:26 INFO org.apache.hadoop.yarn.util.resource.ResourceUtils: Adding resource type - name = memory-mb, units = Mi, type = COUNTABLE
23/03/21 01:51:26 INFO org.apache.hadoop.yarn.util.resource.ResourceUtils: Adding resource type - name = vcores, units = , type = COUNTABLE
23/03/21 01:51:31 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl: Submitted application application_1679363362615_0001
23/03/21 01:51:56 INFO org.apache.hadoop.mapred.FileInputFormat: Total input files to process : 1
[(('DAI62779', 'ELE17451'), 1592), (('FRO40251', 'SNA80324'), 1412), (('DAI75645', 'FRO40251'), 1254), (('FRO40251', 'GRO85051'), 1213), (('DAI62779', 'GRO73461'), 1139)]
elapsed time is 19.5792396068573
23/03/21 01:52:13 INFO org.spark_project.jetty.server.AbstractConnector: Stopped Spark@78990400{HTTP/1.1, (http/1.1)}{0.0.0.0:0}
```

Screenshots

T=4

```
Job [46727185d8584515a5a28a92e12ef321] submitted.
Waiting for job output...
23/03/21 23:54:51 INFO org.apache.spark.SparkEnv: Registering MapOutputTracker
23/03/21 23:54:51 INFO org.apache.spark.SparkEnv: Registering BlockManagerMaster
23/03/21 23:54:51 INFO org.apache.spark.SparkEnv: Registering OutputCommitCoordinator
23/03/21 23:54:51 INFO org.spark_project.jetty.util.log: Logging initialized @5237ms to org.spark_project.jetty.util.log.S]
23/03/21 23:54:52 INFO org.spark_project.jetty.server.Server: jetty-9.4.z-SNAPSHOT; built: unknown; git: unknown; jvm 1.8.0
23/03/21 23:54:52 INFO org.spark_project.jetty.server.Server: Started @5430ms
23/03/21 23:54:52 INFO org.spark_project.jetty.server.AbstractConnector: Started ServerConnector@5ee6aca5{HTTP/1.1, (http:/
23/03/21 23:54:54 INFO org.apache.hadoop.yarn.client.RMProxy: Connecting to ResourceManager at data301-2023-amh284-lab3-clu
23/03/21 23:54:54 INFO org.apache.hadoop.yarn.client.AHSProxy: Connecting to Application History server at data301-2023-amh
23/03/21 23:54:55 INFO org.apache.hadoop.conf.Configuration: resource-types.xml not found
23/03/21 23:54:55 INFO org.apache.hadoop.yarn.util.resource.ResourceUtils: Unable to find 'resource-types.xml'.
23/03/21 23:54:55 INFO org.apache.hadoop.yarn.util.resource.ResourceUtils: Adding resource type - name = memory-mb, units =
23/03/21 23:54:55 INFO org.apache.hadoop.yarn.util.resource.ResourceUtils: Adding resource type - name = vcores, units = ,
23/03/21 23:54:58 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl: Submitted application application_167944278921
23/03/21 23:55:12 INFO org.apache.hadoop.mapred.FileInputFormat: Total input files to process : 1
[(('DAI62779', 'ELE17451'), 1592), (('FRO40251', 'SNA80324'), 1412), (('DAI75645', 'FRO40251'), 1254), (('FRO40251', 'GRO85
elapsed time is 12.796269655227661
```
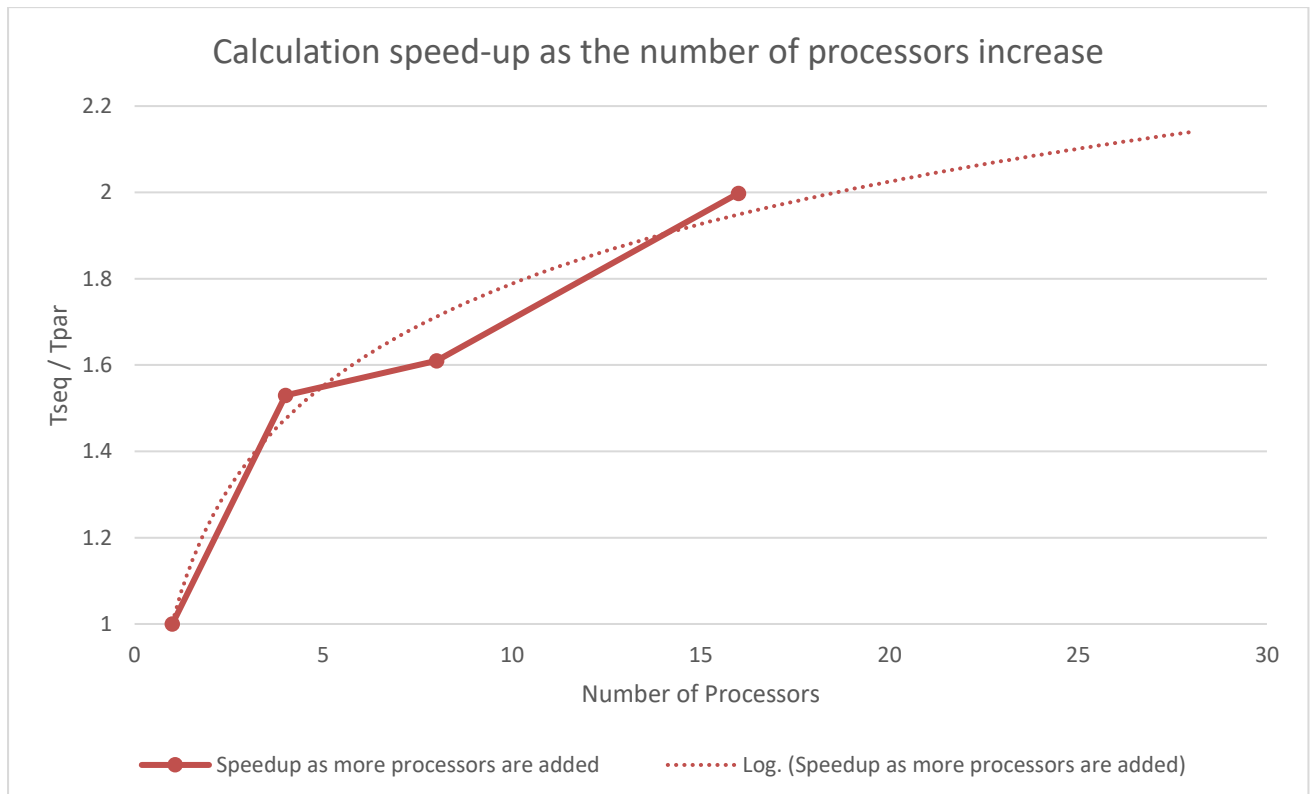
T=8

```
Job [290bc5cebeba4b0bb71f3f69c6ada3aa] submitted.
Waiting for job output...
23/03/21 23:44:48 INFO org.apache.spark.SparkEnv: Registering MapOutputTracker
23/03/21 23:44:48 INFO org.apache.spark.SparkEnv: Registering BlockManagerMaster
23/03/21 23:44:48 INFO org.apache.spark.SparkEnv: Registering OutputCommitCoordinator
23/03/21 23:44:48 INFO org.spark_project.jetty.util.log: Logging initialized @5093ms to org.spark_project.jetty.util.log.S]
23/03/21 23:44:48 INFO org.spark_project.jetty.server.Server: jetty-9.4.z-SNAPSHOT; built: unknown; git: unknown; jvm 1.8.0
23/03/21 23:44:48 INFO org.spark_project.jetty.server.Server: Started @5311ms
23/03/21 23:44:49 INFO org.spark_project.jetty.server.AbstractConnector: Started ServerConnector@2db400c{HTTP/1.1, (http/1
23/03/21 23:44:51 INFO org.apache.hadoop.yarn.client.RMProxy: Connecting to ResourceManager at data301-2023-amh284-lab3-cl
23/03/21 23:44:51 INFO org.apache.hadoop.yarn.client.AHSProxy: Connecting to Application History server at data301-2023-am
23/03/21 23:44:51 INFO org.apache.hadoop.conf.Configuration: resource-types.xml not found
23/03/21 23:44:51 INFO org.apache.hadoop.yarn.util.resource.ResourceUtils: Unable to find 'resource-types.xml'.
23/03/21 23:44:51 INFO org.apache.hadoop.yarn.util.resource.ResourceUtils: Adding resource type - name = memory-mb, units =
23/03/21 23:44:51 INFO org.apache.hadoop.yarn.util.resource.ResourceUtils: Adding resource type - name = vcores, units = ,
23/03/21 23:44:55 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl: Submitted application application_1679442192
23/03/21 23:45:07 INFO org.apache.hadoop.mapred.FileInputFormat: Total input files to process : 1
[(('DAI62779', 'ELE17451'), 1592), (('FRO40251', 'SNA80324'), 1412), (('DAI75645', 'FRO40251'), 1254), (('FRO40251', 'GRO8
elapsed time is 12.160754203796387
```

T=16

```
Job [e55a7c58ad3c47a0a844e514ffd59805] submitted.
Waiting for job output...
23/03/21 23:49:51 INFO org.apache.spark.SparkEnv: Registering MapOutputTracker
23/03/21 23:49:51 INFO org.apache.spark.SparkEnv: Registering BlockManagerMaster
23/03/21 23:49:51 INFO org.apache.spark.SparkEnv: Registering OutputCommitCoordinator
23/03/21 23:49:52 INFO org.spark_project.jetty.util.log: Logging initialized @4709ms to org.spark_project.jetty.util.log.S]
23/03/21 23:49:52 INFO org.spark_project.jetty.server.Server: jetty-9.4.z-SNAPSHOT; built: unknown; git: unknown; jvm 1.8.0
23/03/21 23:49:52 INFO org.spark_project.jetty.server.Server: Started @4879ms
23/03/21 23:49:52 INFO org.spark_project.jetty.server.AbstractConnector: Started ServerConnector@2d37148{HTTP/1.1, (http/1.
23/03/21 23:49:54 INFO org.apache.hadoop.yarn.client.RMProxy: Connecting to ResourceManager at data301-2023-amh284-lab3-clu
23/03/21 23:49:54 INFO org.apache.hadoop.yarn.client.AHSProxy: Connecting to Application History server at data301-2023-amh
23/03/21 23:49:54 INFO org.apache.hadoop.conf.Configuration: resource-types.xml not found
23/03/21 23:49:54 INFO org.apache.hadoop.yarn.util.resource.ResourceUtils: Unable to find 'resource-types.xml'.
23/03/21 23:49:54 INFO org.apache.hadoop.yarn.util.resource.ResourceUtils: Adding resource type - name = memory-mb, units =
23/03/21 23:49:54 INFO org.apache.hadoop.yarn.util.resource.ResourceUtils: Adding resource type - name = vcores, units = ,
23/03/21 23:49:58 INFO org.apache.hadoop.yarn.client.api.impl.YarnClientImpl: Submitted application application_167944250451
23/03/21 23:50:09 INFO org.apache.hadoop.mapred.FileInputFormat: Total input files to process : 1
[(('DAI62779', 'ELE17451'), 1592), (('FRO40251', 'SNA80324'), 1412), (('DAI75645', 'FRO40251'), 1254), (('FRO40251', 'GRO85
elapsed time is 9.799102067947388
23/03/21 23:50:18 INFO org.spark_project.jetty.server.AbstractConnector: Stopped Spark@2d37148{HTTP/1.1, (http/1.1)}{0.0.0.
```

Timing:

:)	Number of processors (T)			
	1	4	8	16
Time (sec)	19.579	12.796	12.161	9.799
Tseq ÷ Tpar	1	1.530088	1.609983	1.998061

It is difficult to know the trend line, as we have so few data points, however I believe the line will follow a logarithmic curve (doubling the cores doesn't halve the time), as the A-Priori confidence/interest algorithms require multiple expensive joins.