

## Billing:

### Billing

MY BILLING ACCOUNTS MY PROJECTS

**Filter** Enter property name or value ? Download

Name	ID	Billing account ↑	Billing account ID	Actions ?
data301-2023-amh284	data301-2023-amh284	Billing Account for Education	018CB7-3CFB2C-876BF6	⋮

**Billing**

- Billing account Billing Account for Education
- Overview**
- Cost management
- Reports
- Cost table
- Cost breakdown
- Budgets & alerts
- Billing export
- Cost optimisation
- Committed-use discounts (CUD)
- CUD analysis
- Pricing

**Overview**

**BILLING ACCOUNT OVERVIEW**

**Current month**  
1–21 March 2023

Month-to-date total cost <span style="color: #0072bc;">?</span> US\$0.00	End-of-month total cost (forecasted) <span style="color: #0072bc;">?</span> Not enough historical data to project cost
---	---

[View report](#)

**Cost trend**  
1 March 2022 – 31 March 2023

Average monthly total cost US\$0.00
--

**Billing health checks** ^  
Take a look at your account health results to avoid common billing-related issues and adopt our best practice recommendations.

[Learn more](#)

! 0 ! 1 ! 1

[View all health checks](#)

## Job:

### T=1:

[Job details](#) CLONE DELETE STOP REFRESH

Job ID	a2402545becc47829e3facf1d34274c1
Job UUID	72f66841-2765-39f7-8d51-3fe3d1ff16ea
Type	Dataproc job
Status	<span style="color: green;">✓</span> Succeeded

MONITORING
CONFIGURATION

i The charts below represent the metrics from the cluster that this job ran on, scoped to the time that this job was running. It is possible for more than one job to run on a cluster at a time, so for a job may lag behind the job run by several minutes.

SAVE AS DASHBOARD
RESET ZOOM
1 hour
6 hours
12 hours
1 day

YARN memory

5GiB

YARN pending memory

4GiB

Output
LINE WRAP: OFF

i Spark jobs take ~60 seconds to initialise resources.

```
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.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.Slf4jLog
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_292-b10
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://0.0.0.0:4040)}
23/03/21 23:54:54 INFO org.apache.hadoop.yarn.client.RMProxy: Connecting to ResourceManager at data301-2023-amh284-lab3-client/192.168.1.10:8032
23/03/21 23:54:54 INFO org.apache.hadoop.yarn.client.AHSProxy: Connecting to Application History server at data301-2023-amh284-lab3-client/192.168.1.10:1020
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.client.api.impl.YarnClientImpl: Submitted application application_16794427892
23/03/21 23:55:12 INFO org.apache.hadoop.mapred.FileInputFormat: Total input files to process : 1
[('DAI62779', 'ELE17451'), 1592], ('FR040251', 'SNA80324'), 1412], ('DAI75645', 'FR040251'), 1254], ('FR040251', 'GR085
elapsed time is 12.796269655227661
```

T=8

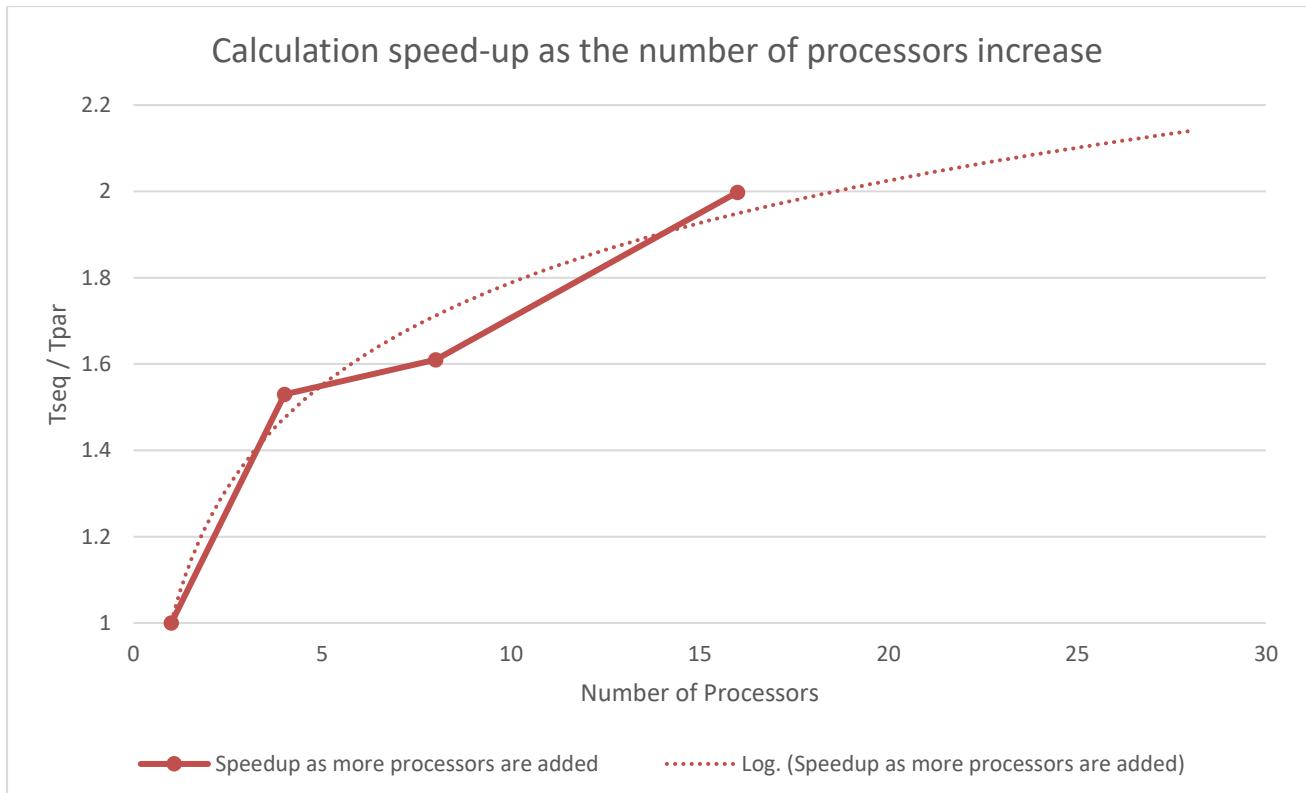
```
Job [290bc5cebeba4b0bb71f3f69c6ada3aa] submitted.
Waiting for job output...
23/03/21 23:44:48 INFO org.apache.spark.SparkEnv: Registering MapOutputTracker
23/03/21 23:44:48 INFO org.apache.spark.SparkEnv: Registering BlockManagerMaster
23/03/21 23:44:48 INFO org.apache.spark.SparkEnv: Registering OutputCommitCoordinator
23/03/21 23:44:48 INFO org.spark_project.jetty.util.log: Logging initialized @5093ms to org.spark_project.jetty.util.log.Slf4jLog
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_292-b10
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://0.0.0.0:4040)}
23/03/21 23:44:51 INFO org.apache.hadoop.yarn.client.RMProxy: Connecting to ResourceManager at data301-2023-amh284-lab3-client/192.168.1.10:8032
23/03/21 23:44:51 INFO org.apache.hadoop.yarn.client.AHSProxy: Connecting to Application History server at data301-2023-amh284-lab3-client/192.168.1.10:1020
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.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], ('FR040251', 'SNA80324'), 1412], ('DAI75645', 'FR040251'), 1254], ('FR040251', 'GR085
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.Slf4jLog
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_292-b10
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://0.0.0.0:4040)}
23/03/21 23:49:54 INFO org.apache.hadoop.yarn.client.RMProxy: Connecting to ResourceManager at data301-2023-amh284-lab3-client/192.168.1.10:8032
23/03/21 23:49:54 INFO org.apache.hadoop.yarn.client.AHSProxy: Connecting to Application History server at data301-2023-amh284-lab3-client/192.168.1.10:1020
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.client.api.impl.YarnClientImpl: Submitted application application_1679442504
23/03/21 23:50:09 INFO org.apache.hadoop.mapred.FileInputFormat: Total input files to process : 1
[('DAI62779', 'ELE17451'), 1592], ('FR040251', 'SNA80324'), 1412], ('DAI75645', 'FR040251'), 1254], ('FR040251', 'GR085
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://0.0.0.0:4040)}
```

**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.