CERN Log Analysis and Optimization Suggestion

Intel SSG/STO/PAC

Cluster Hardware Platform

- 3 types of nodes
- Node type 1: SandyBridge 8C16T
  - physical thread num : memory = 1 : 4GB
- Node type 2: BroadWell 12C24T
  - physical thread num : memory = 1 : 21GB
- Node type 3: BroadWell 10C20T
  - physical thread num : memory = 1 : 6.4GB

Node Type 1:
Node Count: 21
CPU: Intel(R) Xeon(R) CPU E5-2650 0 @ 2.00GHz
RAM: 64 GB
Network: 10Gbps x 1
Storage: 66.79 TB
Node Type 2:
Node Count: 12
CPU: Intel(R) Xeon(R) CPU E5-2650 v4 @ 2.20GHz
RAM: 503 GB
Network: 10 Gbps X 2
Storage: 260.95 TB
Node Type 3:
Node Count: 5
CPU: Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz
RAM: 128 GB
Network: 10 Gbps X 2
Storage: 260.95 TB

Log Summary

- Single application: 16 jobs, 32 stages
- Application not finished: stop at stage 32
- Stage 0 ~ 31 looks like CPU intensive

Stages for All Jobs

Active Stages: 1
Completed Stages: 32

Active Stages (1)

Stage Id ▼	Description		Submitted	Duration	Tasks: Succeeded/Total	Input	Output	Shuffle Read	Shuffle Write
32	save at DimuonReductionAODMultiDataset.scala:369	+details	Unknown	160.5 h	16825/25397 (36 failed)				

Completed Stages (32)

Stage Id ▼	Description		Submitted	Duration	Tasks: Succeeded/Total	Input	Output	Shuffle Read	Shuffle Write
31	count at DimuonReductionAODMultiDataset.scala:331	+details	2018/05/30 21:38:42	8 s	1/1			355.5 KB	
30	count at DimuonReductionAODMultiDataset.scala:331	+details	2018/05/30 21:27:30	11 min	6171/6171				355.5 KB
29	count at DimuonReductionAODMultiDataset.scala:331	+details	2018/05/30 21:27:09	10 s	1/1			245.0 KB	
28	count at DimuonReductionAODMultiDataset.scala:331	+details	2018/05/30 21:24:41	2.5 min	4252/4252				245.0 KB
27	count at DimuonReductionAODMultiDataset.scala:331	+details	2018/05/30 21:24:30	2 s	1/1			98.8 KB	
26	count at DimuonReductionAODMultiDataset.scala:331	+details	2018/05/30 21:22:44	1.8 min	1715/1715				98.8 KB

Stage 32 Issues

- Many tasks hang up
  - root cause still not very clear

- Concurrent multiple HDFS file writing
  - Trigger some errors

- data skew happens on output data size among tasks

Potential Issues of Other Stages

- Huge task execution time diversity
  - Data skew? (need input size to judge)
  - Slow machine?
  - Executor memory not enough? (4 tasks share 2g memory/executor)
- Tasks failed in stage
  - Executor lost due to heartbeat time out
- Max parallel executor number is 64
  - Cluster resource not fully used

Summary Metrics for 6171 Completed Tasks

Metric	Min	25th percentile	Median	75th percentile	Max
Duration	2 s	3 s	4 s	5 s	9.9 min
GC Time	34 ms	0.2 s	0.3 s	0.4 s	1 s
Shuffle Write Size / Records	58.0 B / 1	59.0 B / 1	59.0 B / 1	59.0 B / 1	59.0 B / 1

Index	ID	Attempt	Status ▲	Locality Level	Executor ID	Host	Launch Time	Duration	GC Time	Write Time	Shuffle Write Size / Records	Errors
189	2609	0	FAILED	PROCESS_LOCAL	96	ithdp1112.cern.ch stdout stderr	2018/05/30 21:01:07	2.4 min			0.0 B / 0	ExecutorLostFailure (executor 96 exited caused by one of the running tasks) Reason: Executor heartbeat timed out after 145734 ms
432	2852	0	FAILED	PROCESS_LOCAL	96	ithdp1112.cern.ch stdout stderr	2018/05/30 21:01:16	2.3 min			0.0 B / 0	ExecutorLostFailure (executor 96 exited caused by one of the running tasks) Reason: Executor heartbeat timed out after 145734 ms
433	2853	0	FAILED	PROCESS_LOCAL	96	ithdp1112.cern.ch stdout stderr	2018/05/30 21:01:16	2.3 min			0.0 B / 0	ExecutorLostFailure (executor 96 exited caused by one of the running tasks) Reason: Executor heartbeat timed out after 145734 ms
434	2854	0	FAILED	PROCESS_LOCAL	96	ithdp1112.cern.ch stdout stderr	2018/05/30 21:01:16	2.3 min			0.0 B / 0	ExecutorLostFailure (executor 96 exited caused by one of the running tasks) Reason: Executor heartbeat timed out after 145734 ms
2452	4872	0	FAILED	PROCESS_LOCAL	117	p05151113452631.cern.ch	2018/05/30 21:02:03	6.5 min			0.0 B / 0	ExecutorLostFailure (executor 117 exited caused by one of the running tasks) Reason: Executor heartbeat timed out after 153911 ms
2453	4873	0	FAILED	PROCESS_LOCAL	117	p05151113452631.cern.ch	2018/05/30 21:02:03	6.5 min			0.0 B / 0	ExecutorLostFailure (executor 117 exited caused by one of the running tasks) Reason: Executor heartbeat timed out after 153911 ms
2454	4874	0	FAILED	PROCESS_LOCAL	117	p05151113452631.cern.ch	2018/05/30 21:02:03	6.5 min			0.0 B / 0	ExecutorLostFailure (executor 117 exited caused by one of the running tasks) Reason: Executor heartbeat timed out after 153911 ms
2455	4875	0	FAILED	PROCESS_LOCAL	117	p05151113452631.cern.ch	2018/05/30 21:02:03	6.5 min			0.0 B / 0	ExecutorLostFailure (executor 117 exited caused by one of the running tasks) Reason: Executor heartbeat timed out after 153911 ms

Optimization Suggestion

- Node of type 1: disable HT
- Node of type 2: enable HT
- Node of type 3: enable HT
- Spark parameters setting
    - Spark.executor.memory   16g
    - Spark.executor.cores   4
    - Spark.executor.memory.overhead 4g
- Application:
    - If data skew exists, considering increase partition number or repartition
    - Stage 32 stores output in local file system or Hbase?

Node Type 1:

yarn.nodemanager.resource.memory-mb:
51200

yarn.nodemanager.resource.cpu-vcores:
8

Node Type 2:

yarn.nodemanager.resource.memory-mb:
131072

yarn.nodemanager.resource.cpu-vcores:
26

Node Type 3:

yarn.nodemanager.resource.memory-mb:
102400

yarn.nodemanager.resource.cpu-vcores:
20

Next Step Work

- Get spark history log and detail log of the same job
  - Identify the root cause of application failure
  - Provide parameters setting optimization package
- Cluster usage scenarios: single user with single application or multiple users with multiple applications
- Scale up/out analysis after successful execution

PAC Competitive Simulation

SSG/STO/PAC