Report

I declare that the assignment submitted on Elearning system is original except for source material explicitly acknowledged, and that the same or related material has not been previously submitted for another course. I also acknowledge that I am aware of University policy and regulations on honesty in academic work, and of the disciplinary guidelines and procedures applicable to breaches of such policy and regulations, as contained in the website http://www.cuhk.edu.hk/policy/academichonesty/.

Signed (Student Cao Yuhang ) Date: 2/11/2018
Name: Cao Yuhang SID: 1155092180

c. Test on 1 master, 4 slaves cluster, each vm has 2 cores 8 GB memory and 40 GB volume.

1) 8 mappers, 1 reducer

| Maximum mapper time | Minimum mapper time | Average mapper time | Maximum reducer time | Minimum reducer time | Average reducer time | Total job |
|---|---|---|---|---|---|---|
| 10mins, 55sec | 22sec | 3mins, 40sec | 3mins, 34sec | 3mins, 34mec | 3mins, 34sec | 15mins, 5sec |

2) 16 mappers, 2 reducers

| Maximum mapper time | Minimum mapper time | Average mapper time | Maximum reducer time | Minimum reducer time | Average reducer time | Total job |
|---|---|---|---|---|---|---|
| 8mins, 15sec | 7sec | 1mins, 52sec | 1mins, 50sec | 1mins, 46mec | 1mins, 48sec | 11mins, 11sec |

3) 32 mappers, 4 reducer

| Maximum mapper time | Minimum mapper time | Average mapper time | Maximum reducer time | Minimum reducer time | Average reducer time | Total job |
|---|---|---|---|---|---|---|
| 5mins, 6sec | 5sec | 1mins, 1sec | 55sec | 55sec | 55sec | 8mins, 16sec |

4) 64 mappers, 8 reducers

| Maximum mapper time | Minimum mapper time | Average mapper time | Maximum reducer time | Minimum reducer time | Average reducer time | Total job |
|---|---|---|---|---|---|---|
| 3mins, 55sec | 3sec | 31sec | 30sec | 28sec | 29sec | 7mins, 27sec |

5) 128 mappers, 16 reducers

| Maximum mapper time | Minimum mapper time | Average mapper time | Maximum reducer time | Minimum reducer time | Average reducer time | Total job |
|---|---|---|---|---|---|---|
| 3mins, 31sec | 3sec | 17sec | 18sec | 13sec | 15sec | 9mins, 6sec |

Explanation:

Just as Hadoop Wiki said: “Increasing the number of tasks increases the framework overhead, but increases load balancing and lowers the cost of failures.”

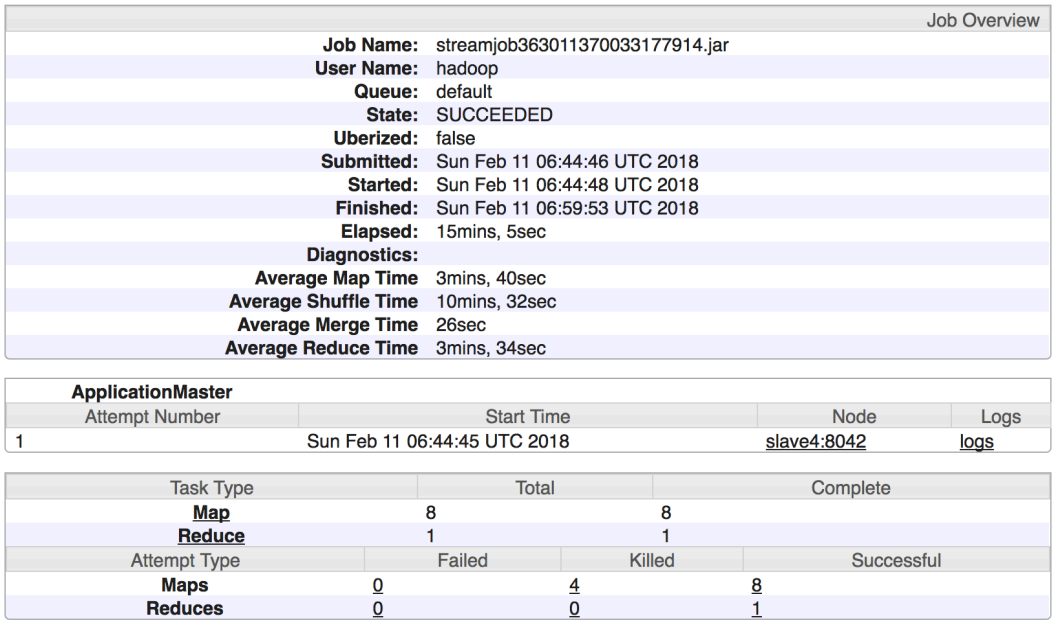
- In 1), 2), 3), 4) experiments, when I increase the number of mappers and reducers, the *total job time* is decreasing, which proves increasing the number of tasks if beneficial for parallelization and increasing performance.
- If we observe the *maximum mapper time* and *average mapper time*, *maximum reducer time*, *average reducer time*, we can find all metrics are decreasing, which prove increasing number of tasks is beneficial for load balancing and lowers the cost of failures.

But if we observe experiment 5), we can find even though almost all metrics are decreasing, but the *total job time* increases, which proves too much tasks will increase the framework overhead, I guess these may be some possible factors:

- Too many mappers and reducers will increase the start up time.
- Too many mappers will lead too many input splits, it may need more random seeks overhead for random seek is very slow.
- Too many mappers will violate data locality, the data need to transfer across different mappers, which increase the network traffic overhead.
- Too many reducers will increase data traffic in shuffle and sort state, because in shuffle and sort state the output of mappers need to be partitioned to different reducers.
- Too many reducers will increase the I/O operations since you need to create more files as each reducer create its own file.

Screen shot record:

1) 8 mappers, 1 reducer



Job Overview

| | |
|---|---|
| **Job Name:** | streamjob363011370033177914.jar |
| **User Name:** | hadoop |
| **Queue:** | default |
| **State:** | SUCCEEDED |
| **Uberized:** | false |
| **Submitted:** | Sun Feb 11 06:44:46 UTC 2018 |
| **Started:** | Sun Feb 11 06:44:48 UTC 2018 |
| **Finished:** | Sun Feb 11 06:59:53 UTC 2018 |
| **Elapsed:** | 15mins, 5sec |
| **Diagnostics:** | |
| **Average Map Time** | 3mins, 40sec |
| **Average Shuffle Time** | 10mins, 32sec |
| **Average Merge Time** | 26sec |
| **Average Reduce Time** | 3mins, 34sec |

**ApplicationMaster**

| Attempt Number | Start Time | Node | Logs |
|---|---|---|---|
| 1 | Sun Feb 11 06:44:45 UTC 2018 | slave4:8042 | logs |

| Task Type | Total | Complete |
|---|---|---|
| **Map** | 8 | 8 |
| **Reduce** | 1 | 1 |

| Attempt Type | Failed | Killed | Successful |
|---|---|---|---|
| **Maps** | 0 | 4 | 8 |
| **Reduces** | 0 | 0 | 1 |

2) 16 mappers, 2 reducers

| Job Overview | |
|---|---|
| **Job Name:** | streamjob1969245684752408941.jar |
| **User Name:** | hadoop |
| **Queue:** | default |
| **State:** | SUCCEEDED |
| **Uberized:** | false |
| **Submitted:** | Sun Feb 11 07:01:22 UTC 2018 |
| **Started:** | Sun Feb 11 07:01:22 UTC 2018 |
| **Finished:** | Sun Feb 11 07:12:34 UTC 2018 |
| **Elapsed:** | 11mins, 11sec |
| **Diagnostics:** | |
| **Average Map Time** | 1mins, 52sec |
| **Average Shuffle Time** | 6mins, 28sec |
| **Average Merge Time** | 59sec |
| **Average Reduce Time** | 1mins, 48sec |

| **ApplicationMaster** | | | |
|---|---|---|---|
| Attempt Number | Start Time | Node | Logs |
| 1 | Sun Feb 11 07:01:19 UTC 2018 | slave4:8042 | logs |

| Task Type | Total | Complete |
|---|---|---|
| **Map** | 16 | 16 |
| **Reduce** | 2 | 2 |

| Attempt Type | Failed | Killed | Successful |
|---|---|---|---|
| **Maps** | 0 | 3 | 16 |
| **Reduces** | 0 | 0 | 2 |

3) 32 mappers, 4 reducers

| Job Overview | |
|---|---|
| **Job Name:** | streamjob6211025324602395882.jar |
| **User Name:** | hadoop |
| **Queue:** | default |
| **State:** | SUCCEEDED |
| **Uberized:** | false |
| **Submitted:** | Sun Feb 11 07:13:47 UTC 2018 |
| **Started:** | Sun Feb 11 07:13:46 UTC 2018 |
| **Finished:** | Sun Feb 11 07:22:03 UTC 2018 |
| **Elapsed:** | 8mins, 16sec |
| **Diagnostics:** | |
| **Average Map Time** | 1mins, 1sec |
| **Average Shuffle Time** | 1mins, 33sec |
| **Average Merge Time** | 34sec |
| **Average Reduce Time** | 55sec |

| **ApplicationMaster** | | | |
|---|---|---|---|
| Attempt Number | Start Time | Node | Logs |
| 1 | Sun Feb 11 07:13:43 UTC 2018 | slave2:8042 | logs |

| Task Type | Total | Complete |
|---|---|---|
| **Map** | 32 | 32 |
| **Reduce** | 4 | 4 |

| Attempt Type | Failed | Killed | Successful |
|---|---|---|---|
| **Maps** | 0 | 0 | 32 |
| **Reduces** | 0 | 1 | 4 |

4) 64 mappers, 8 reducers

| Job Overview | |
|---|---|
| **Job Name:** | streamjob7887003209791181674.jar |
| **User Name:** | hadoop |
| **Queue:** | default |
| **State:** | SUCCEEDED |
| **Uberized:** | false |
| **Submitted:** | Sun Feb 11 07:22:36 UTC 2018 |
| **Started:** | Sun Feb 11 07:22:36 UTC 2018 |
| **Finished:** | Sun Feb 11 07:30:04 UTC 2018 |
| **Elapsed:** | 7mins, 27sec |
| **Diagnostics:** | |
| **Average Map Time** | 31sec |
| **Average Shuffle Time** | 32sec |
| **Average Merge Time** | 19sec |
| **Average Reduce Time** | 29sec |

| **ApplicationMaster** | | | |
|---|---|---|---|
| Attempt Number | Start Time | Node | Logs |
| 1 | Sun Feb 11 07:22:33 UTC 2018 | slave1:8042 | logs |

| Task Type | Total | Complete |
|---|---|---|
| **Map** | 64 | 64 |
| **Reduce** | 8 | 8 |

| Attempt Type | Failed | Killed | Successful |
|---|---|---|---|
| **Maps** | 0 | 0 | 64 |
| **Reduces** | 0 | 1 | 8 |

5) 128 mappers, 16 reducers

| Job Overview | |
|---|---|
| **Job Name:** | streamjob4837117934239272653.jar |
| **User Name:** | hadoop |
| **Queue:** | default |
| **State:** | SUCCEEDED |
| **Uberized:** | false |
| **Submitted:** | Tue Feb 13 01:53:37 UTC 2018 |
| **Started:** | Tue Feb 13 01:53:38 UTC 2018 |
| **Finished:** | Tue Feb 13 02:02:45 UTC 2018 |
| **Elapsed:** | 9mins, 6sec |
| **Diagnostics:** | |
| **Average Map Time** | 17sec |
| **Average Shuffle Time** | 25sec |
| **Average Merge Time** | 11sec |
| **Average Reduce Time** | 15sec |

| **ApplicationMaster** | | | |
|---|---|---|---|
| Attempt Number | Start Time | Node | Logs |
| 1 | Tue Feb 13 01:53:35 UTC 2018 | slave1:8042 | logs |

| Task Type | Total | Complete |
|---|---|---|
| **Map** | 128 | 128 |
| **Reduce** | 16 | 16 |

| Attempt Type | Failed | Killed | Successful |
|---|---|---|---|
| **Maps** | 0 | 0 | 128 |
| **Reduces** | 0 | 1 | 16 |