A Guide to The IE DIC Cluster

Cluster overview:

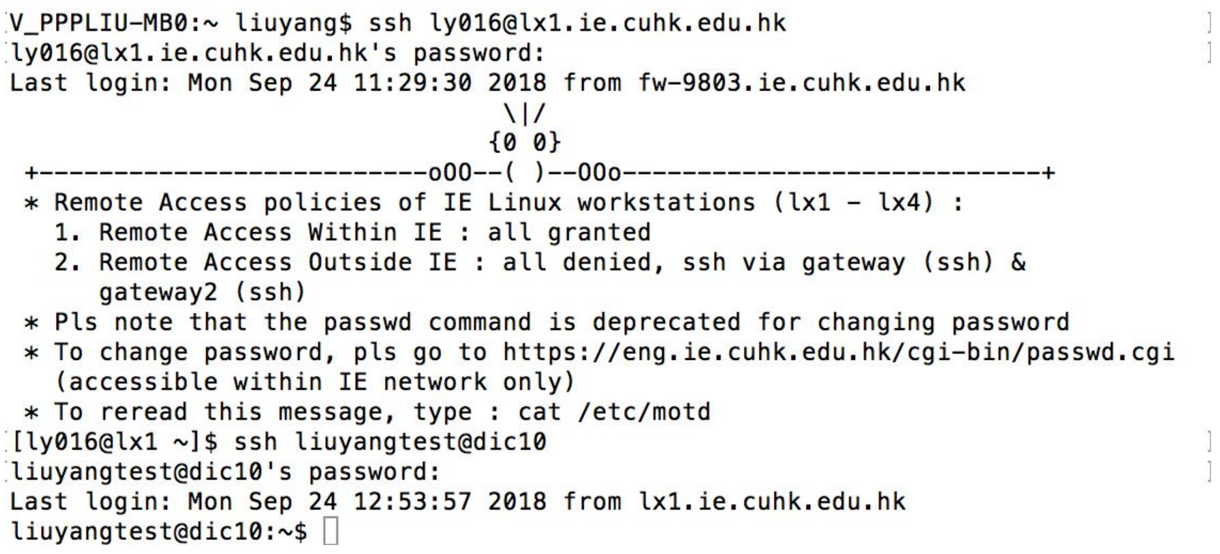
- 28 nodes
- Memory: 47 GB *2 + 30 GB*4 + 63 GB*3 + 16 GB*19 = 707 GB
- Virtual CPU Cores: 24*2 + 16*7 + 8*19 = 312 cores
- Disk: 44.2TB
- Resource management platform: YARN
- Installed applications: MapReduce, Hive, Pig, Storm,Spark

Account application:

- **(Old Version)** Send an email to yangliu476730@yahoo.com. If you are an IE student, please attach your IE account and Student ID. If you are a CS student, please attach your Student ID and I will apply a temporary IE account for you.
- **(New Version)** You can directly get your IE DIC account in the spreadsheet: https://docs.google.com/spreadsheets/d/15pAlrtEt4BvxwgGHoaYA8uO_NVyxrLDEJ4UiqMOBhVM/edit?usp=sharing

Cluster login:

- Login the cluster via: ssh user_id@dic10.ie.cuhk.edu.hk
- The IE DIC Cluster can only be accessed within the IE network, you can follow the link below to set up an IE VPN. http://mobitec.ie.cuhk.edu.hk/engg4030Fall2016/homework/vpn_setup.html



```
V_PPPLIU-MB0:~ liuyang$ ssh ly016@lx1.ie.cuhk.edu.hk
ly016@lx1.ie.cuhk.edu.hk's password:
Last login: Mon Sep 24 11:29:30 2018 from fw-9803.ie.cuhk.edu.hk
                                 \|/
                                {0 0}
 +--------------------------oOO--( )--OOo----------------------------+
 * Remote Access policies of IE Linux workstations (lx1 - lx4) :
   1. Remote Access Within IE : all granted
   2. Remote Access Outside IE : all denied, ssh via gateway (ssh) &
      gateway2 (ssh)
 * Pls note that the passwd command is deprecated for changing password
 * To change password, pls go to https://eng.ie.cuhk.edu.hk/cgi-bin/passwd.cgi
   (accessible within IE network only)
 * To reread this message, type : cat /etc/motd
[ly016@lx1 ~]$ ssh liuyangtest@dic10
liuyangtest@dic10's password:
Last login: Mon Sep 24 12:53:57 2018 from lx1.ie.cuhk.edu.hk
liuyangtest@dic10:~$ 
```

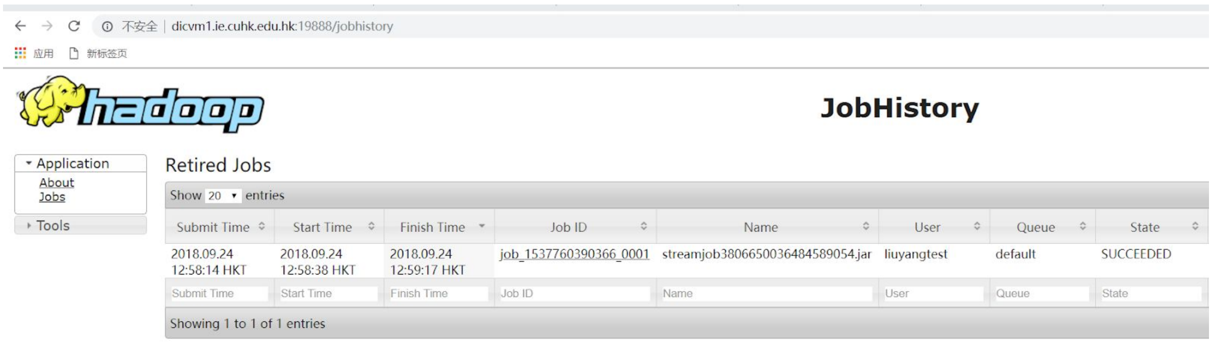
Submit a Python MapReduce job:

- Below is an example to login IE DIC Cluster and submit a Python Mapreduce job

```
[ly016@lx1 ~]$ ssh liuyangtest@dic10.ie.cuhk.edu.hk
liuyangtest@dic10:~$  hdfs dfs -mkdir input
liuyangtest@dic10:~$  hdfs dfs -put file input
liuyangtest@dic10:~$  vim mapper.py
liuyangtest@dic10:~$  vim reducer.py
liuyangtest@dic10:~$  chmod +x mapper.py reducer.py
liuyangtest@dic10:~$  hadoop jar
/usr/hdp/current/hadoop-mapreduce-client/hadoop-streaming.jar -file
mapper.py -mapper mapper.py -file reducer.py -reducer reducer.py -input
input/* -output output
```

Find the logs of MapReduce application:

- Users can find the logs of all applications in the cluster via the Web UI: *http://dicvm1.ie.cuhk.edu.hk:19888/*
- Users can find the details of a particular application via the Web UI: http://dicvm1.ie.cuhk.edu.hk:19888/jobhistory/job/job_1473396442288_0004 where *job_ 1473396442288_0004* is the ID of the job you created.
- The log information of an application includes: 1) how many containers are allocated; 2) the scheduling time and the completion time of each container; 3) the stderr file which can help you to find bugs of your codes.



JobHistory

Retired Jobs

Show 20 entries

Submit Time	Start Time	Finish Time	Job ID	Name	User	Queue	State
2018.09.24 12:58:14 HKT	2018.09.24 12:58:38 HKT	2018.09.24 12:59:17 HKT	job_1537760390366_0001	streamjob3806650036484589054.jar	liuyangtest	default	SUCCEEDED

Showing 1 to 1 of 1 entries

Submit a JAVA Storm job:

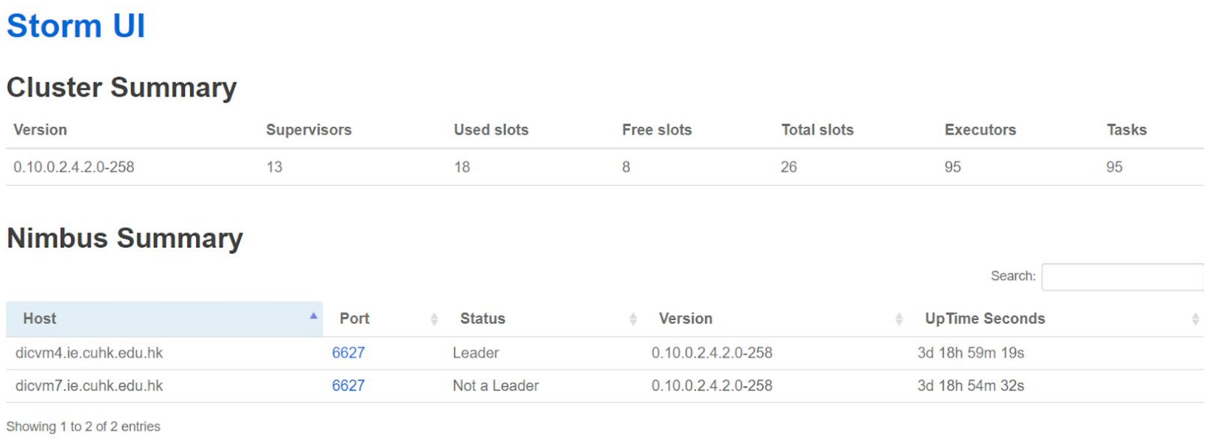
- Below is an example to login IE DIC Cluster and submit a JAVA Storm job

```
[ly016@lx1 ~]$ ssh liuyangtest@dic10.ie.cuhk.edu.hk
liuyangtest@dic10:~$  cd /usr/hdp/current/storm-client/contrib/storm-starter
liuyangtest@dic10:~$  storm jar storm-starter-topologies-0.10.0.2.4.2.0-258.jar
storm.starter.WordCountTopology
liuyangtest@dic10:~$  mvn -v
```

Find the logs of Storm application:

- Users can find the logs of all applications in the cluster via the Web UI:
  http://dicvm4.ie.cuhk.edu.hk:8744/index.html

- The log information of an application includes: how many workers, executors
- **Pay attension to the version of Storm. The version 0.9.x and version 1.x.x are different.**



Storm UI

Cluster Summary

Version	Supervisors	Used slots	Free slots	Total slots	Executors	Tasks
0.10.0.2.4.2.0-258	13	18	8	26	95	95

Nimbus Summary

Search:

Host	Port	Status	Version	UpTime Seconds
dicvm4.ie.cuhk.edu.hk	6627	Leader	0.10.0.2.4.2.0-258	3d 18h 59m 19s
dicvm7.ie.cuhk.edu.hk	6627	Not a Leader	0.10.0.2.4.2.0-258	3d 18h 54m 32s

Showing 1 to 2 of 2 entries

Submit a Hive and Pig job:

- The location mode is enough for our homework.
- Below is an example to login IE DIC Cluster and submit Hive and Pig jobs:

```
[ly016@lx1 ~]$ ssh liuyangtest@dic10.ie.cuhk.edu.hk
liuyangtest@dic10:~$  hive
liuyangtest@dic10:~$  pig
```