1. MongoDB 安裝

(1) 使用 mongoDB apt 的公開金鑰

```
sudo apt-key adv --keyserver hkp://keyserver.ubuntu.com:80 --recv 2930ADAE8CAF5059EE73BB4B58712A2291FA4AD5
```

```
bao@master:~$ sudo apt-key adv --keyserver hkp://keyserver.ubuntu.com:80 --recv
2930ADAE8CAF5059EE73BB4B58712A2291FA4AD5
Executing: /tmp/tmp.zhDpuFGTTg/gpg.1.sh --keyserver
hkp://keyserver.ubuntu.com:80
--recv
2930ADAE8CAF5059EE73BB4B58712A2291FA4AD5
gpg: requesting key 91FA4AD5 from hkp server keyserver.ubuntu.com
gpg: key 91FA4AD5: public key "MongoDB 3.6 Release Signing Key <packaging@mongod
b.com>" imported
gpg: Total number processed: 1
gpg:               imported: 1  (RSA: 1)
```

(2) 將金鑰加入

```
echo "deb [ arch=amd64,arm64 ] https://repo.mongodb.org/apt/ubuntu xenial/mongodb-org/3.6 multiverse" | sudo tee /etc/apt/sources.list.d/mongodb-org-3.6.list
```

(3) 更新

```
sudo apt-get update
```

(4) 安裝 mongodb

```
sudo apt-get install -y mongodb-org
```

(5) 啟動

```
sudo service mongod start
```

(6) 執行 mongodb

mongo

```
bao@master:~$ mongo
MongoDB shell version v3.6.17
connecting to: mongodb://127.0.0.1:27017/?gssapiServiceName=mongodb
Implicit session: session { "id" : UUID("324119e0-5429-46ed-aad7-417bca5d9c06")
}
MongoDB server version: 3.6.17
Server has startup warnings:
2020-02-27T15:50:53.888+0800 I STORAGE  [initandlisten]
2020-02-27T15:50:53.888+0800 I STORAGE  [initandlisten] ** WARNING: Using the XF
S filesystem is strongly recommended with the WiredTiger storage engine
2020-02-27T15:50:53.888+0800 I STORAGE  [initandlisten] **          See http://d
ochub.mongodb.org/core/prodnotes-filesystem
2020-02-27T15:50:54.355+0800 I CONTROL  [initandlisten]
2020-02-27T15:50:54.355+0800 I CONTROL  [initandlisten] ** WARNING: Access contr
ol is not enabled for the database.
2020-02-27T15:50:54.355+0800 I CONTROL  [initandlisten] **          Read and wri
te access to data and configuration is unrestricted.
2020-02-27T15:50:54.355+0800 I CONTROL  [initandlisten]
>
```

(7) Mongodb 操作

show dbs　查看有哪些資料庫

use your_number　使用資料庫(若無資料庫會新增一個)

```
> use 123456789
switched to db 123456789
```

輸入下列程式碼插入資料

```
db.restaurants.insert(
   {
      "address" : {
         "street" : "2 Avenue",
         "zipcode" : "10075",
         "building" : "1480",
         "coord" : [ -73.9557413, 40.7720266 ],
      },
      "borough" : "Manhattan",
      "cuisine" : "Italian",
      "grades" : [
```

```
        {
            "date" : ISODate("2014-10-01T00:00:00Z"),
            "grade" : "A",
            "score" : 11
        },
        {
            "date" : ISODate("2014-01-16T00:00:00Z"),
            "grade" : "B",
            "score" : 17
        }
    ],
    "name" : "Vella",
    "restaurant_id" : "41704620"
  }
)
```

```
...         "zipcode" : "10075",
...         "building" : "1480",
...         "coord" : [ -73.9557413, 40.7720266 ],
...       },
...       "borough" : "Manhattan",
...       "cuisine" : "Italian",
...       "grades" : [
...         {
...           "date" : ISODate("2014-10-01T00:00:00Z"),
...           "grade" : "A",
...           "score" : 11
...         },
...         {
...           "date" : ISODate("2014-01-16T00:00:00Z"),
...           "grade" : "B",
...           "score" : 17
...         }
...       ],
...       "name" : "Vella",
...       "restaurant_id" : "41704620"
...     }
... )
WriteResult({ "nInserted" : 1 })
```

show collections　　查看有哪些資料集合

```
> show collections
restaurants
```

db.restaurants.find() 顯示資料

```
> db.restaurants.find()
{ "_id" : ObjectId("5e5786a72d7c7c7616b87b3e"), "address" : { "street" : "2 Aven
ue", "zipcode" : "10075", "building" : "1480", "coord" : [ -73.9557413, 40.77202
66 ] }, "borough" : "Manhattan", "cuisine" : "Italian", "grades" : [ { "date" :
ISODate("2014-10-01T00:00:00Z"), "grade" : "A", "score" : 11 }, { "date" : ISODa
te("2014-01-16T00:00:00Z"), "grade" : "B", "score" : 17 } ], "name" : "Vella", "
restaurant_id" : "41704620" }
```

(補充)

條件查詢

db.restaurants.find({ "borough": "Manhattan" })

db.restaurants.find({ "address.zipcode": "10075" })

多條件查詢

db.restaurants.find({ "cuisine": "Italian", "address.zipcode": "10075" })

db.restaurants.find(

{ $or: [{ "cuisine": "Italian" }, { "address.zipcode": "10075" }] }

)

db.collection.drop() 刪除資料集合(collection 要填入集合的名字)

```
> db.getCollectionNames()
[ "restaurants" ]
> db.restaurants.drop()
true
```

use db_name 使用資料庫

db.dropDatabase() 刪除資料庫

exit 離開 mongo

```
> use test
switched to db test
> db.dropDatabase()
{ "ok" : 1 }
> exit
bye
```

2. Scala 安裝

(1) cd ~/Downloads

(2) wget https://www.scala-lang.org/files/archive/scala-2.12.1.tgz

(3) sudo tar xvf scala-2.12.1.tgz

(4) sudo mv scala-2.12.1 /opt/scala

(5) sudo gedit /etc/profile

#新增以下一行

export SCALA_HOME="/opt/scala"

#修改 export PATH，在最後面加上:$SCALA_HOME/bin

export PATH

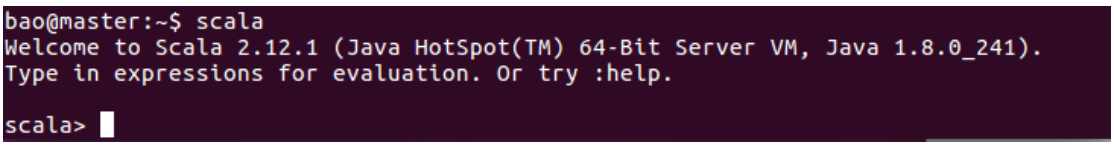
=$PATH:$JAVA_HOME/bin:$HADOOP_HOME/bin:$HADOOP_HOME/sbin:$

MAVEN_HOME/bin:$MAHOUT_HOME/bin:$SCALA_HOME/bin

(6) source /etc/profile

(7) 執行 scala

scala



(8)離開

:q

3. Spark 安裝

(1) cd ~/Downloads

(2) wget http://ftp.tc.edu.tw/pub/Apache/spark/spark-3.0.0-preview2/spark-3.0.0-preview2-bin-hadoop2.7.tgz

(3) sudo tar xvf spark-3.0.0-preview2-bin-hadoop2.7.tgz

(4) sudo mv spark-3.0.0-preview2-bin-hadoop2.7 /opt/spark

(5) sudo gedit /etc/profile

#新增以下一行

export SPARK_HOME="/opt/spark"

#修改 export PATH，在最後面加上:$SPARK_HOME /bin

export PATH

=$PATH:$JAVA_HOME/bin:$HADOOP_HOME/bin:$HADOOP_HOME/sbin:$

MAVEN_HOME/bin:$MAHOUT_HOME/bin:$SCALA_HOME/bin:$SPARK_H

OME/bin

(6) source /etc/profile

(7) 執行 spark

spark-shell

```
bao@master:~$ spark-shell
20/02/27 17:47:04 WARN NativeCodeLoader: Unable to load native-hadoop library fo
r your platform... using builtin-java classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLeve
l(newLevel).
Spark context Web UI available at http://master:4040
Spark context available as 'sc' (master = local[*], app id = local-1582796835189
).
Spark session available as 'spark'.
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /___/ .__/\_,_/_/ /_/\_\   version 3.0.0-preview2
      /_/

Using Scala version 2.12.10 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_241)
Type in expressions to have them evaluated.
Type :help for more information.

scala>
```

(8) 離開

:q

4. Mongo + Spark

(1) 下載 package 並配置(在終端機輸入)

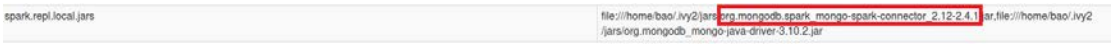
123456789 輸入自己的學號(db 名稱)

numbers 輸入集合的名稱

```
spark-shell --conf
"spark.mongodb.input.uri=mongodb://127.0.0.1/123456789.numbers?read
Preference=primaryPreferred" --conf
"spark.mongodb.output.uri=mongodb://127.0.0.1/123456789.numbers" --
packages org.mongodb.spark:mongo-spark-connector_2.12:2.4.1
```

可以透過 http://master:4040中的 Environment 去查看

org.mongodb.spark:mongo-spark-connector 版本



(2) 引入 mongoDB 和 Spark 連結檔案

```
import com.mongodb.spark._
```

利用迴圈將數字 1~10 寫入 mongoDB

```
import org.bson.Document
val documents = sc.parallelize((1 to 10).map(i => Document.parse(s"{test:
$i}")))
```

MongoSpark.save(documents)

```
scala> import com.mongodb.spark._
import com.mongodb.spark._

scala> import org.bson.Document
import org.bson.Document

scala> val documents = sc.parallelize((1 to 10).map(i => Document.parse(s"{test: $i}")
))
documents: org.apache.spark.rdd.RDD[org.bson.Document] = ParallelCollectionRDD[0] at p
arallelize at <console>:28

scala> MongoSpark.save(documents)
```

(3) Spark 讀取 mongoDB 並顯示內容

val rdd = MongoSpark.load(sc)

println(rdd.count)

println(rdd.first.toJson)

```
scala> val rdd = MongoSpark.load(sc)
20/02/27 19:44:38 WARN SparkSession$Builder: Using an existing SparkSession; some conf
iguration may not take effect.
rdd: com.mongodb.spark.rdd.MongoRDD[org.bson.Document] = MongoRDD[1] at RDD at MongoRD
D.scala:51

scala> println(rdd.count)
10

scala> println(rdd.first.toJson)
{"_id": {"$oid": "5e57aba07925f00502fad81e"}, "test": 1}
```

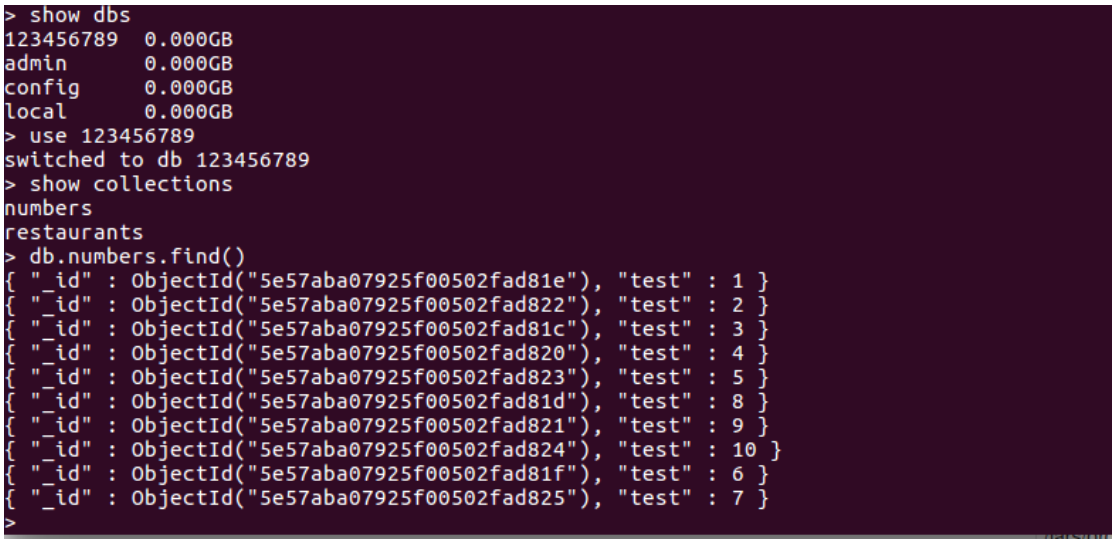
(4) rdd.collect() ：將 rdd 轉換成 Array

```
scala> rdd.collect()
res3: Array[org.bson.Document] = Array(Document{{_id=5e57aba07925f00502fad81e, test=1}
}, Document{{_id=5e57aba07925f00502fad822, test=2}}, Document{{_id=5e57aba07925f00502f
ad81c, test=3}}, Document{{_id=5e57aba07925f00502fad820, test=4}}, Document{{_id=5e57a
ba07925f00502fad823, test=5}}, Document{{_id=5e57aba07925f00502fad81d, test=8}}, Docum
ent{{_id=5e57aba07925f00502fad821, test=9}}, Document{{_id=5e57aba07925f00502fad824, t
est=10}}, Document{{_id=5e57aba07925f00502fad81f, test=6}}, Document{{_id=5e57aba07925
f00502fad825, test=7}})
```

(5) 離開 spark

:q

開啟 mongo 並展示如下圖



疑難排除

移除 mongodb 套件：

```
sudo apt-get purge mongodb-org*
```

移除資料目錄：

```
sudo rm -r /var/log/mongodb
sudo rm -r /var/lib/mongodb
```