HADOOP & NATURAL LANGUAGE PROCESSING (DATA SCIENCE FLAVOR) ALL IN ONE COURSE

GITHUB · https://github.com/nitinkaushik01/Hadoop-Natural-Language-Processing

CREATE DATABASE:

CREATE DATABASE MYDATABASE;

DROP DATABASE

DROP DATABASE MYDATABASE;

SHOW DATABASES

SHOW DATABASES;

SHOW TABLES

SHOW TABLES;

CREATE EXTERNAL TABLE

CREATE EXTERNAL TABLE CLUB_MMBRSHIP_EXTERNAL (
CLUB_ID INT,
NAME STRING,
MEMBER_FEE DECIMAL)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY '|'
LINES TERMINATED BY '\N'
STORED AS TEXTFILE
LOCATION '/USER/MYDATA';

CREATE INTERNAL TABLE

CREATE TABLE CLUB_MMBRSHIP_INTERNAL (
CLUB_ID INT,
NAME STRING,
MEMBERSHIP_FEE DECIMAL)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY '|'
LINES TERMINATED BY '\N'
STORED AS TEXTFILE
LOCATION '/USER/MYDATA';

DROP TABLE

```
DROP TABLE CLUB_MMBRSHIP_INTERNAL;
```

SET PROPERTIES FOR DYNAMIC PARTITIONED TABLE

```
SET HIVE.EXEC.DYNAMIC.PARTITION=TRUE
SET HIVE.EXEC.DYNAMIC.PARTITION.MODE=NONSTRICT
```

CREATE PARTITIONED TABLE

```
CREATE TABLE EMPLOYEE (
ID_NUM INT,
NAME STRING,
ROLE STRING
)
PARTITIONED BY (YOP INT)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY '|'
LINES TERMINATED BY '\N'
STORED AS TEXTFILE;
```

DROP A PARTITION

```
ALTER TABLE EMPLOYEE DROP
PARTITION (YEAROFPROMOTION='2014');
```

SHOW PARTITION(S)

```
SHOW PARTITIONS EMPLOYEE;
```

LOAD DATA FROM FLAT FILE (RESIDING AT HADOOP LOCAL DIRECTORY PATH) TO A TABLE

```
LOAD DATA LOCAL INPATH '/ROOT/EMPLOYEE.TXT' INTO TABLE EMPLOYEE_STAGING;
```

LOAD DATA FROM FLAT FILE (RESIDING ON HDFS DIRECTORY PATH) TO A TABLE

```
LOAD DATA INPATH '/USER/EMPLOYEE.TXT' INTO TABLE EMPLOYEE_STAGING;
```

INSERT OVERWRITE AN EXISTING TABLE TAKING DATA FROM ANOTHER TABLE

```
INSERT OVERWRITE TABLE EMPLOYEE PARTITION (YOP) SELECT ID_NUM, NAME, ROLE, YOP FROM
EMPLOYEE_STAGING
```

CREATE A VIEW

```
CREATE VIEW PROD_PRICE100 AS
```

```
SELECT * FROM PRODUCT
WHERE PRODUCT_PRICE > 9;
```

DROP A VIEW

```
DROP VIEW PROD_PRICE9;
```

CREATE AN INDEX

```
CREATE INDEX INEDX_ITEMID ON TABLE ORDER_ITEM(ITEM_ID)
AS 'ORG.APACHE.HADOOP.HIVE.QL.INDEX.COMPACT.COMPACTINDEXHANDLER' WITH DEFERRED
REBUILD;
```

DROP AN INDEX

```
DROP INDEX INDEX_ITEMID ON ORDER_ITEM;
```

CREATE A .CSV OR .TXT FILE (AT HADOOP LOCAL) TAKING DATA FROM SPECIFIC TABLE

```
INSERT OVERWRITE LOCAL DIRECTORY '/ROOT/MYDIRECTORY/POSITIVE.TXT'
SELECT MSG FROM ANALYZE_TWEETS WHERE SENTIMENT = 2;
```

CREATE A .CSV OR .TXT FILE (AT HDFS DIRECTORY PATH) TAKING DATA FROM SPECIFIC TABLE

```
INSERT OVERWRITE DIRECTORY '/ROOT/MYDIRECTORY/POSITIVE.TXT'
SELECT MSG FROM ANALYZE_TWEETS WHERE SENTIMENT = 2;
```