## Hive.sql

> CREATE DATABASE enhanceit

> LOCATION '/enhance\_dbs';

> USE enhanceit;

> CREATE EXTERNAL TABLE engineers(id INT, first\_name STRING, last\_name STRING, age INT, city STRING, country STRING, salary FLOAT, department INT)

ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'

STORED AS TEXTFILE

LOCATION '/enhance\_dbs/engineers';

> CREATE EXTERNAL TABLE managers

LIKE engineers

LOCATION '/enhance\_dbs/managers';

> CREATE EXTERNAL TABLE departments(

id INT,

name STRING)

LOCATION '/enhance\_dbs/departments';

> ALTER TABLE engineers

ADD COLUMNS (hire\_date DATE, renewal\_date DATE);

> LOAD DATA LOCAL INPATH '/home/maria\_dev/data\_Local/spark.SQL-Hive/engineers.csv' OVERWRITE INTO TABLE engineers;

> LOAD DATA LOCAL INPATH '/home/maria\_dev/data\_Local/spark.SQL-Hive/departments.csv' OVERWRITE INTO TABLE departments;

> LOAD DATA LOCAL INPATH '/home/maria\_dev/data\_Local/spark.SQL-Hive/managers.csv' OVERWRITE INTO TABLE managers;

**// Highest Paid Manager by Department**

Select max(m.salary) as salary, d.name

from department d, managers m

where d.id = m.id\_manager

group by d.name;

/**/ Top 3 Highest Paid Manager**

SELECT \* FROM

(

SELECT \*FROM managers

ORDER BY salary desc

)

WHERE rownum <= 3

ORDER BY salary ;

**// Top 2 Highest Paid Engineers**

SELECT \* FROM

(

SELECT \*FROM engineers

ORDER BY salary desc

)

WHERE rownum <= 3

ORDER BY salary ;

**//Highest Paid Emp by Manager**

Select max(e.salary) as salary, m.first\_name

from engineers e, managers m

where e.id = m.id

group by m.first\_name;