**1-0. Windows防火牆設定**

開始 > 控制台 > Windows防火牆 > 進階設定 > 輸入規則 > 新增規則 > 連接阜(O) > 特定連接阜1433 > 允許連線(A)> 下一步 > … > 完成



* 1. **Download PieTTY**

### 主機名稱或IP位置 : 10.120.28.xxx

### 可從windows 連進 linux的 terminal

<http://ntu.csie.org/~piaip/pietty/>



### 2-0. Installing the Microsoft SQL Server JDBC Driver

### Download the MSSQL Server JDBC driver [here](http://www.microsoft.com/en-us/download/details.aspx?displaylang=en&id=11774) and copy it to /var/lib/sqoop/ directory.

### http://www.microsoft.com/zh-tw/download/details.aspx?id=11774

$ tar -xf sqljdbc\_4.0.2206.100\_cht.tar.gz

$ sudo cp sqljdbc4.jar /var/lib/sqoop

**3-1. 在SQL Server內建立一個View**

### 包含1808部電影的前三個演員和第一週票房和上映年

create view information

AS

SELECT DISTINCT w.pkno, ROUND(LOG (w.weekendgross), 5) AS gross, ca.priority4c AS sort, ca.cast, w.year

FROM dbo.convHistRateByYear AS b

INNER JOIN

(SELECT pkno

FROM dbo.boxoffice

GROUP BY pkno

HAVING (SUM(weekendgross) >= 300000)) AS s

ON b.pkno = s.pkno

INNER JOIN

(SELECT pkno, title, weekendgross, CONVERT(char(4), date) AS year

FROM dbo.boxoffice

WHERE (weekend = 1)) AS w

ON s.pkno = w.pkno

INNER JOIN dbo.cast\_pri AS ca

ON b.pkno = ca.pkno

WHERE (0 < ca.priority4c)

AND (4 > ca.priority4c)

**3-2. Import data from local to HDFS**

### 將local的表格放入hdfs

$ hadoop fs -ls ./

$ hadoop fs -put ./data/USArrests.txt ./data/USArrests.txt

**3-3. Import data from SQL Server to HDFS**

### 將sqlserver表格放入hdfs

$ sqoop import --connect "jdbc:sqlserver://10.120.28.27:1433;username=sa;password=passw0rd;database=IMDB" \

--table information -m 1

$ hadoop fs -cat ./information/part-m-00000

**3-4. Import data from SQL Server to Hive**

### 將sqlserver表格放入hive

$ sqoop import --connect "jdbc:sqlserver://10.120.28.27:1433;username=sa;password=passw0rd;database=IMDB" \

--table information -m 1 --hive-import

### 注意hdfs內不能有同名稱的表格, 會衝突



### Query Editor (Hive)

> SELECT \* FROM information

**3-5. Import data from local to Hive**

### 將local表格放入hive

CREATE TABLE twice (id STRING, frequency INT)

ROW FORMAT DELIMITED

FIELDS TERMINATED by '\t'

LOAD DATA LOCAL INPATH 'home/cloudera/Desktop/hive/data/twice.txt' OVERWRITE INTO TABLE twice;

> SELECT \* FROM twice

> SELECT twice.frequency, COUNT(twice.id)

FROM twice

GROUP by twice.frequency

ORDER by twice.frequency ASC

**3-6. Import data from HDFS to HDFS**

### 將hdfs表格放入hive

> DROP TABLE IF EXISTS USArrests;

CREATE EXTERNAL TABLE USArrests(

Location string,

Murder float,

Assault float,

UrbanPop float,

Rape float

)

ROW FORMAT DELIMITED

FIELDS TERMINATED by ','

### 建立表格和上傳資料要分開執行

> LOAD DATA INPATH '/user/shiuanh/data/USArrests.csv' OVERWRITE INTO TABLE USArrests;

> SELECT \* FROM USArrests

**4-1. Export data from Hive to SQL Server**

### 先在sqlserver建立好information\_hive表格

****

$ sqoop export --connect "jdbc:sqlserver://10.120.28.27:1433;username=sa;password=passw0rd;database=IMDBComplete" -m -1 --table information\_hive --export-dir /user/hive/warehouse/information --input-fields-terminated-by '\0001'

