Retail Project Sales transtion

Data Sample (In xml format)

<tr><dt>01/23/2005</dt><amt>350000</amt><country>India</country><prodcut>bear</product></tr>

<tr><dt>01/27/2005</dt><amt>380000</amt><country>India</country><prodcut>visky</product></tr>

<tr><dt>02/12/2005</dt><amt>450000</amt><country>India</country><prodcut>Rum</product></tr>

<tr><dt>01/23/2006</dt><amt>500000</amt><country>USA</country><prodcut>bear</product></tr>

<tr><dt>01/27/2006</dt><amt>550000</amt><country>USA</country><prodcut>rum</product></tr>

<tr><dt>02/12/2006</dt><amt>650000</amt><country>USA</country><prodcut>Visky</product></tr>

<tr><dt>01/23/2006</dt><amt>500000</amt><country>China</country><prodcut>Beer</product></tr>

<tr><dt>01/27/2006</dt><amt>550000</amt><country>China</country><prodcut>Visky</product></tr>

<tr><dt>02/12/2006</dt><amt>650000</amt><country>China</country><prodcut>Rum</product></tr>

Insert the structured data into final data table.(we used here dynamic partition table)year--→ month-→date

```
sethive.exec.dynamic.partition=true
;
sethive.exec.dynamic.partition.mode
=nonstrict;
```

sample
dynamic partition

create table dypart(country string,product string, amtint, year int, month int, day int) partiotioned by(y int,mint, d int);

in back end y,m and d become the three directory.

when describe the dypart table we will see three extra column in the table actually these are partition column.

When we will generate report that time we can use the column in where clause. That retrieves the data fast.

Basically dynamic partition increase performance data retrieval.

Configuring the number of dynamic partition

sethive.exec.max.dynamic.partitions=100000

1)yearly sales report

2)Yearly sum for all country

3)Select yr, SUM(amt) from finaldata

group by yr;

4)Yearly sum for specified country

5)yearly report dumping in the yr_sales_rep

Meanwhile we can apply all aggregation function here.

6) monthly sales report of a perticular year

Dumping the data into mn_sales_rep

7)quarterly sales report of a particular year

8)For each quartetr report for each year

Write udf function for that.

And for all year

9) half yearly sales rep of " "

Report is dumped into all_hy_sales_rep

10) foreach year, monthly sales rep

11) foreach year, quarterly sales rep

Report is dumped into all_qrt_sales_rep

12) foreACH year, half yearl sales rep.

Report is dumped into fore_yr_sales_rep

13) multiple branches worldwide the fiscal year is diff from country to country

14) generate first 7 report according to country fiscal years;

According for month for fiscal year like india for jan to july

For indiaapril to dec.

15.) compare the quaatly sales of each country of the particaular year.

Done for all year----all_q1, all_q2

16.) in a specific quart which product made more bussiness

Compare all quarter according to product

17.) suggest the compnay (saler) for a specific product in which area concentrate more


AWS big data ecosystem
Collect
Process & Analyze
Visualize
Apps
App Server
Web Server
Transactional Data
Devices
iOS Devices
Android
Logging Frameworks
LOG4J
File Data
fluentd
Kinesis Producer
Stream Data
DynamoDB
S3
Kinesis
Data Pipeline
informatica
ATTUNITY
EMR
Redshift
COGNOS
SAP
Business Objects
tableau
sas
birst
pentaho
TIBCO
Jaspersoft
web services