|  |
| --- |
|  |
|  | Store raw data into hdfs location |
|  | Create a internal hive table "sales\_order\_csv" which will store csv data sales\_order\_csv .. make sure to skip header row while creating table |
|  |  |
|  | Load data from hdfs path into "sales\_order\_csv" |
|  |  |
|  | Create an internal hive table which will store data in ORC format "sales\_order\_orc" |
|  |  |
|  | Load data from "sales\_order\_csv" into "sales\_order\_orc" |
|  |  |
|  |  |
|  |  |
|  | Perform below menioned queries on "sales\_order\_orc" table : |
|  |  |
|  | 1. Calculatye total sales per year |
|  | 1. Find a product for which maximum orders were placed   select productline, sum(quantityordered) as total\_orders from sales\_order\_data\_orc group by productline limit 1; |
|  | 1. Calculate the total sales for each quarter   select qtr\_id, sum(sales) as total\_sales from sales\_order\_data\_orc group by qtr\_id; |
|  |  |
|  | 1. In which quarter sales was minimum   select qtr\_id, sum(sales) as total\_sales from sales\_order\_data\_orc group by qtr\_id limit 1; |
|  | 1. In which country sales was maximum and in which country sales was minimum   select country, sum(sales) as total\_sales from sales\_order\_data\_orc group by country limit1; |
|  | select country, sum(sales) as total\_sales from sales\_order\_data\_orc group by country order by total\_sales limit 1; |
|  | 1. Calculate quartelry sales for each city   select qtr\_id,city, sum(sales) as quarterly\_sales from sales\_order\_data\_orc group by qtr\_id,city; |
|  |  |
|  | 1. Find a month for each year in which maximum number of quantities were sold   select month\_id, year\_id, sum(sales) as total\_sales from sales\_order\_data\_orc group by month\_id,year\_id order by total\_sales desc; |