

Conducting a hands-on exercise using Hue, Hive, and Impala involves a series of steps from accessing data to performing analysis and visualization. Let's break down this process into detailed steps with examples.

Hands-On Exercise: Using Hue, Hive, and Impala

1. Access Data with Hue

- **Log into Hue:** Open your web browser and navigate to the Hue interface. Log in with your credentials.
- **Explore HDFS Data:**
 - Use the File Browser in Hue to navigate through the HDFS directories.
 - View or download files, upload new data, or create new directories.

2. Running Queries in Hue

Using Hive through Hue

- **Open Hive Editor:** In Hue, go to the Hive editor.
- **Create Hive Table:** If you don't already have a table, create one.

```
CREATE TABLE if not exists sales_data (  
    sale_id INT,  
    product_id INT,  
    quantity_sold INT,  
    sale_date DATE  
)  
ROW FORMAT DELIMITED  
FIELDS TERMINATED BY ','  
STORED AS TEXTFILE;
```

- **Load Data into Hive Table:** Load data from an HDFS file into the sales_data table.

```
LOAD DATA INPATH '/user/hive/sales_data.csv' INTO TABLE sales_data;
```

- **Run Hive Query:** Perform a query on the sales_data table.

```
SELECT product_id, SUM(quantity_sold) as total_sold  
FROM sales_data  
GROUP BY product_id;
```

Using Impala through Hue

- **Open Impala Editor:** Navigate to the Impala editor within Hue.
- **Run Impala Query:** Execute a query using Impala. For instance, querying the same sales_data table:

```
SELECT product_id, AVG(quantity_sold) as average_sold
FROM sales_data
GROUP BY product_id;
```

- **Refresh Impala Metadata:** If new data is loaded into Hive, use the `INVALIDATE METADATA` command in Impala to refresh its metadata.

3. Data Analysis and Visualization

- **Basic Analysis:** Use the query results in Hue for basic analysis. Hue displays query results that can be sorted and filtered.
- **Visualization:**
 - Hue offers basic visualization tools. For example, you can visualize the results of your queries in graphical formats like bar charts or pie charts.
 - To create a visualization, select the columns you want to visualize in the query results, then choose the type of chart or graph.
- **Advanced Analysis:**
 - For more advanced analysis, consider exporting the query results to a CSV file and using external tools like Python with Pandas or Jupyter notebooks.

Example Scenario: Sales Data Analysis

1. **Explore Sales Data:** Navigate to the sales data file in HDFS using the File Browser.
2. **Create Hive Table and Load Data:** As per the steps above.
3. **Perform Queries:**
 - Use Hive to calculate total sales per product.
 - Use Impala to calculate the average quantity sold per product.
4. **Visualize Results:** Create a bar chart in Hue to visualize the total and average sales per product.