1. **When was the last time you tried/learned something new? What did you do and how?**

One month back I started learning Tableau. I installed Tableau desktop version and made few

Graphs with the sample data and also learnt about storage techniques, filters, parameters in Tableau.

In the current project in the organization, I am learning and working on HDFS and HiveQL.

1. **What is Lambda architecture? Which problems does it solve?**

**Answer:** Lambda architecture loads the data into conventional ETL framework and real time processing framework in parallel. It skips data quality checks while loading data through real time processing framework. ETL framework manages the dataset and compute pre-aggregated views .While sending the query results back, data from final tables of ETL framework and real time processing framework are merged and sent to the user.

It solves the problem of unavailability or delay of real time data when real time availability is critical to business but quality of the data is not a major concern.

**Advantages -:**

1. Low latency
2. High throughput
3. Fault tolerant

**Disadvantage -:**

1. Complex to maintain two code for same logic

**Question 3) Please implement the SQL transformation code from session\_events to screen\_flows. How would you test it?**

I’ll set up the data as per the scenarios explained in the problem and check if the result set complies with the business rule each time. Please find the data and result set attached with the mail. **This query was tested in Hive QueryEditor**

**Question 4) What do you consider a modern data warehouse architecture?**

1. Modern data warehouse must have automated decision making on certain pre-defined business events.
2. Modern data warehouse must be able to load process all sort of data which includes text, images and must provide the analysis within justifiable latency.
3. Modern data warehouse must embrace data lakes in which data can be stored forever and at very low cost to build predictive models and data mining.
4. Modern data warehouse should have MPP databases with In-memory to efficiently process structured data and to perform iterative analysis.
5. Modern data warehouse should have a tool to easily query relational and non-relational data through a single query service.
6. Modern data warehouse must be available to lowest level of stakeholder in the business.

**How would you implement monitoring?**

1. Choose scheduling software which provides centralized view of all the jobs in data warehouse.
2. Identify the jobs and try to name the jobs according to its significance and frequency.
3. Find all the source application and owners of the jobs and type of inputs from all the sources.
4. Create and Classify users in database and give the user with appropriate priority and resources.
5. Classifying all the jobs according to business criticality and give more priority and resources to business critical users and jobs
6. Choose database which provides centralized information system about data warehouse health.
7. All the running jobs should be monitored on few parameters like 1) Memory 2) CPU 3) IO
8. To achieve SLA, we can use the back end databases of scheduling software to providing real time progress reports of the batch.
9. All the code change should be thoroughly reviewed and implemented.
10. Scheduled purging of data, backing up and capacity planning well in advance.
11. Proper documentation of all the changes after development and next versions.