1. List the execution time of the weblog aggregation query for Hive, SparkSQL, and SparkSQL on Parquet.

Hive: 121.904 seconds
SparkSQL: 28.566 seconds
Parquet: 7.831 seconds

2. How many jobs does Hive launch? Does SparkSQL launch jobs?

Hive launched 2 jobs.

SparkSQL does not appear to launch any jobs.

3. Write a query which joins weblogs_parquet to user_info and counts the top 5 locations. List the locations.

```
SELECT COUNT(user_id), location
FROM user_info
GROUP BY location
ORDER BY COUNT(user_id) DESC;
```

Top 5 locations:

```
19  Hamilton
18  Axis
17  La Fayette
17  Headland
17  Hazel Green
```

Ask Amit about joins