

W205 Section 3 Lab 3
Sanjay Dorairaj

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

Hive	Time taken: 57.946 seconds, Fetched: 50 row(s)
SparkSQL	Time taken: 11.715 seconds, Fetched 50 row(s)
SparkSQL on Parquet	Time taken: 3.673 seconds, Fetched 50 row(s)

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

- Hive launches 2 jobs.
- SparkSQL does not launch any jobs

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

```
select location,count(location) as locationcount from user_info inner
join weblogs_parquet on weblogs_parquet.user_id=user_info.userid
group by location order by locationcount desc limit 5;
```

```
La Fayette      49
Leeds 47
Blountsville   46
Hayden         45
Hamilton       45
```

```
Time taken: 2.673 seconds, Fetched 5 row(s)
```