

**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_info.location) as location,  
user_info.location FROM user_info  
INNER JOIN weblogs_parquet  
ON (user_info.user_id = weblogs_parquet.user_id)  
GROUP BY location  
ORDER BY location DESC limit 5;
```

Top 5 locations are bolded below:

```
49 La Fayette  
47 Leeds  
46 Blountsville  
45 Hayden  
45 Hamilton  
44 Angoon  
44 Chapman  
42 Hazel Green  
42 Greensboro  
41 Whittier  
41 Adger  
41 Albertville
```

...