1. **Pig vs MapReduce?**

* Pig Scans whole dataset, so not suited if we are looking for small portion of data.
* Pig is highly supportive
* Pig is bit slower than Java written MR jobs.
* Pig runs series of MR programs under its hood.

1. **What are different execution environments for Pig?**

**Local mode:** Runs on single JVM

**Map Reduce mode:** Runs on Hadoop cluster

1. How to start Pig?
2. What are bags and tuples in Pig Latin?
3. What are disadvantages of Pig?
4. What are disadvantages of Hive?
5. Hive vs Pig vs MR?
6. What do you understand by Schema on Read in case of Hive?
7. How to start Hive?
8. What is Metastore? Which databases are used as metastore?
9. Managed table vs External Table?
10. Hive Warehouse
11. What is partition in Hive?
12. What is Bucketing in Hive? When to use it?
13. Storage Format in Hive? Row Format and File Format?
14. SerDe Row Format.