Quick Ref: Tez

|  |  |  |
| --- | --- | --- |
| **S.No** | **Topic** | **Desc** |
|  | **WebRef** |  |
|  | Tez Configuration | <https://www.youtube.com/watch?v=ffvlaRvXpjs&index=54&list=PLf0swTFhTI8rG6Q4dr_FruqDdzu9M7AQ2> |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  | **General Info** |  |
| 1 | About Tez | 1. Tez is an application framework configured on top of YARN to allow complex directed acyclic graph (DAG) to process data. 2. Tez can also be used to run Apache Hive queries and Apache Pig Latin scripts. Hive queries or Pig Latin scripts will be run as a Tez Job 3. Tez will use only Yarn not MapReduce to process data. It will skip MapReduce and run its own way 4. Tez tries to fine tune Hive or Pig queries which spawns into multiple jobs by eliminating HDFS between dependent map reduce jobs 5. Tez Comes out of with Hortonworks box (ambari), but not with other vendors like cloudera. 6. Performance is better when we run complex hive jobs |
| 2 | Daemon process | Client Only (No daemon process) |
| 3 | Parameter files | Tez-site.xml  Tez-env.sh |
| 4 | Log files | Not applicable. Typically jobs will not be run on Tez directly instead they will be run either by using Pig or Hive. Hence look at the Hive or Pig log files for any issues |
| 5 | MapReduce vs Tez   1. Blue box is Map 2. Green box is Reduce 3. Tez eliminates Map in the middle layer |  |
| 6 |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  | **General Commands** |  |
| 1 | Configuration path | # After Tez configuration using Ambari (Hortonworks)  >cd /etc/tez/conf  # you can see tez-site.xml and tez-env.sh |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |