What is Hive? ⍰ Hive is a data warehousing infrastructure based on Apache Hadoop which provides SQL like language for querying and analyzing Big Data. ⍰ Hive provides a mechanism to project structure onto the data and query the data using SQL-like language called HiveQL ⍰ Hive uses MapReduce and HDFS for processing and storage/retrieval of data. ⍰ Hive is used for analyzing structured and semi-structured data. ⍰ SQL commands in Hive are called as HiveQL. ⍰ HiveQL gets converted to map reduce jobs by the Hive compiler. ⍰ Apache Hive supports Data Definition Language (DDL), Data Manipulation Language (DML) and User Defined Functions (UDF). ⍰ Hive is not designed for online transaction processing. It is best used for traditional data warehousing tasks.