In Previous Years Questions

Prerequisites

1. Java: A compatible version of Java (e.g., Java 8 or later) must be installed and configured.
2. Hadoop: Apache Hadoop must be installed and configured. Hive relies on Hadoop for distributed file storage and processing.
3. ZooKeeper: ZooKeeper is required for coordinating distributed operations within Hive.

Installation Steps

1. Download: Download the latest version of Apache Hive from the official website (https://hive.apache.org/).
2. Extract: Unzip the downloaded archive to a desired location.
3. Configuration:
    - Edit the hive-site.xml file located in the conf directory.
    - Configure the necessary properties, such as the Hadoop installation directory and ZooKeeper connection details.
    - Optionally, configure additional settings like Hive metastore database connection.
4. Environment Variables
    - Set the HIVE_HOME environment variable to point to the main Hive directory where the extracted files reside.
    - Add the $HIVE_HOME/bin directory to the PATH environment variable.
5. Initialization
    - Run the schematool -initSchema -dbType derby command to initialize the Hive metastore database.

Running Hive

1. Open a terminal window.
2. Navigate to the Hive installation directory using the cd command.
3. Start the Hive shell by running the hive command.

References:

- https://hive.apache.org/
- https://cwiki.apache.org/confluence/display/Hive/HiveQL