Cloud Computing And Big Data Beginners

This lesson covers the following topics:

- Spark
- Hadoop
- Kubernetes I'd like to talk about cloud computing, it seems we can't do without many tools: Hadoop, Hive, Hbase, ZooKeeper, Docker, Kubernetes, Spark, Kafka, MongoDB, Flink, Druid, Presto, Kylin, Elastic Search. Have you heard of them? I found these from job descriptions for a "Big Data Engineer"and "Distributed Backend Engineer". These are high-paying positions. Let's try installing them all and giving them a spin.

Getting Started with Spark

Spark is an open-source, distributed computing system used for big data processing and machine learning. It provides an interface for programming entire clusters with implicit data parallelism and fault tolerance.

To get started with Spark, you need to install it on your machine or cluster. Here's a step-by-step guide for installing Spark on a local machine using Homebrew on macOS:

1. Install Java Development Kit (JDK):

   ```
   brew install adoptopenjdk
   ```

2. Add JDK to environment variables:

   ```
   export JAVA_HOME=$(/usr/local/opt/openjdk@11/libexec)
   export PATH=$JAVA_HOME/bin:$PATH
   ```

3. Install Spark:

   ```
   brew install apache-spark
   ```

4. Add Spark to environment variables:

   ```
   export SPARK_HOME=$(brew --prefix apache-spark)
   export PATH=$SPARK_HOME/bin:$PATH
   ```

5. Verify the installation:

   ```
   spark-shell --version
   ```

   You should see the Spark version number displayed.

Now that Spark is installed, you can start using it for data processing and machine learning tasks. You can write Spark applications in Scala, Java, or Python. To learn more about Spark, check out the official documentation: https://spark.apache.org/docs/latest/index.html. The website states that `Spark` is an engine for analyzing large-scale data. `Spark` is a suite of libraries. It seems different from `Redis` which is divided into server and client. `Spark` is only used on the client side. Downloaded the latest version from the website, `spark-3.1.1-bin-hadoop3.2.tar`.