

Hướng dẫn cài đặt Apache Spark trên CentOS

Hướng dẫn cài đặt Apache Spark trên CentOS



1. Tải xuống Apache Spark từ trang web chính thức của Spark:

<https://spark.apache.org/downloads.html>

2. Giải nén tệp tar bằng lệnh sau:

```
tar -xvf spark-3.1.2-bin-hadoop3.2.tgz
```

Trong đó, `spark-3.1.2-bin-hadoop3.2.tgz` là tên tệp bạn đã tải xuống.

3. Di chuyển thư mục Spark vừa được giải nén bằng lệnh sau:

```
cd spark-3.1.2-bin-hadoop3.2
```

4. Tạo một biến môi trường cho Spark bằng cách thêm dòng sau vào tệp `~/.bashrc`:

```
export SPARK_HOME=/path/to/spark-3.1.2-bin-hadoop3.2
export PATH=$PATH:$SPARK_HOME/bin
```

Trong đó, `path/to` là đường dẫn đến thư mục chứa tệp Spark.

5. Tải xuống và cài đặt Java Development Kit (JDK) trên CentOS bằng lệnh sau:

```
sudo yum install java-1.8.0-openjdk-devel
```

6. Kiểm tra xem JDK đã được cài đặt chưa bằng lệnh sau:

```
java -version
```

Nếu JDK đã được cài đặt, bạn sẽ thấy thông tin về phiên bản JDK trên CentOS của bạn.

7. Kiểm tra xem Spark đã được cài đặt chưa bằng lệnh sau:

```
spark-shell
```

Nếu Spark đã được cài đặt thành công, bạn sẽ thấy giao diện dòng lệnh của Spark.

Đó là quá trình cài đặt Apache Spark trên CentOS. Bây giờ bạn đã sẵn sàng để bắt đầu sử dụng Spark để xử lý dữ liệu của mình.

Revision #2

Created 21 June 2023 06:39:50 by Laptrinh.vn

Updated 21 June 2023 06:41:52 by Laptrinh.vn