

아파치 스파크

스파크 데이터 프레임 파티션

스파크 데이터 프레임 파티션

spark.read.csv()

스파크

드라이버

세션

인-메모리

HDFS

스파크 데이터 프레임 파티션

스파크

드라이버

세션

M M M M M

사용자 정의 함수

HDFS

왜 스파크를 사용해야 하는가?

1. 추상화
2. 통합 플랫폼
3. 사용 편의성