

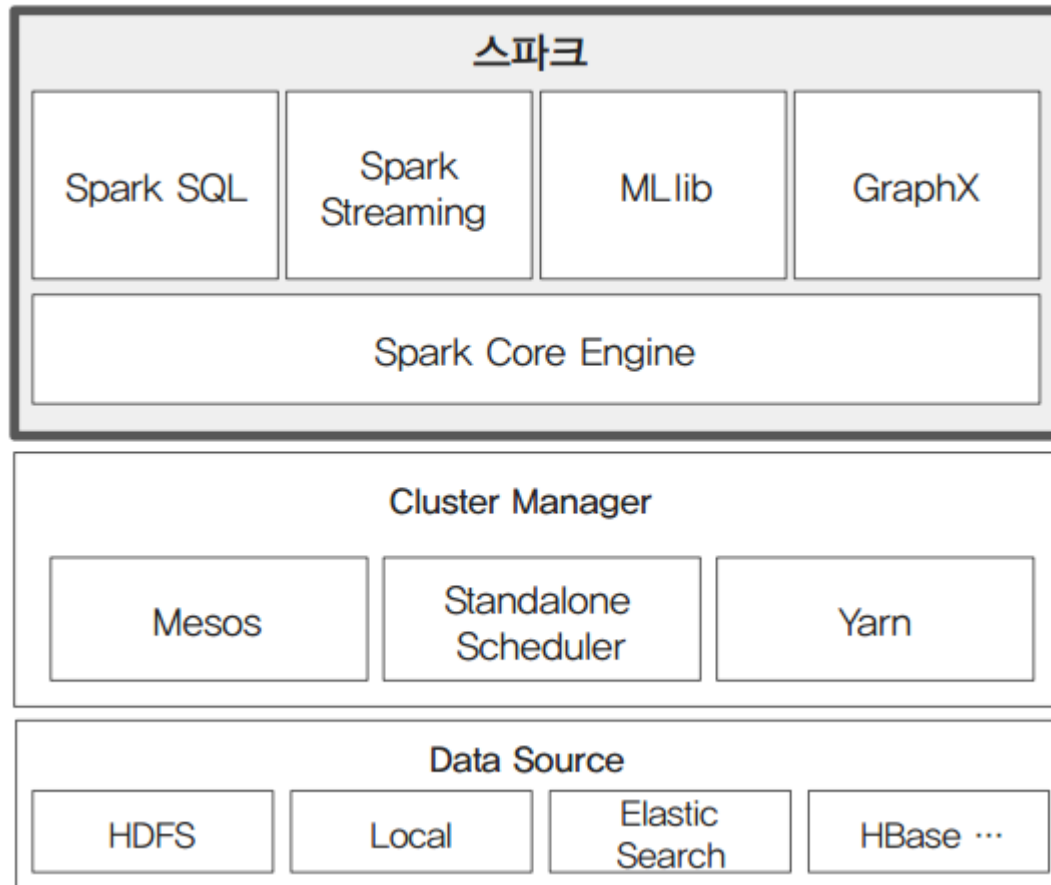
# 6.2 빅데이터 탐색에 활용되는 기술

## 스파크

공식 홈페이지	 <a href="http://spark.apache.org">http://spark.apache.org</a>	
주요 구성 요소	Spark RDD	스파크 프로그래밍의 기초 데이터셋 모델
	Spark Driver / Executors	Driver는 RDD 프로그램을 분산 노드에서 실행하기 위한 Task의 구성, 할당, 계획 등을 수립하고, Executor는 Task를 실행 관리하며, 분산 노드의 스토리지 및 메모리를 참조
	Spark Cluster Manager	스파크 실행 환경을 구성하는 클러스터 관리자로 Mesos, YARN, Spark Standalone이 있음
	Spark SQL	SQL 방식으로 스파크 RDD 프로그래밍을 지원
	Spark Streaming	스트리밍 데이터를 마이크로타임의 배치로 나누어 실시간 처리
	Spark MLib	스파크에서 머신러닝 프로그래밍(군집, 분류, 추천 등)을 지원
	Spark GraphX	다양한 유형의 네트워크(SNS, 하이퍼링크 등) 구조 분석을 지원
라이선스	Apache	
유사 프로젝트	Impala, Tajo, Tez	

## 6.2 빅데이터 탐색에 활용되는 기술

### 스파크 아키텍처



## 6.2 빅데이터 탐색에 활용되는 기술

### 스파크 활용 방안

