# Spark

عادل قاضی خانی
هیئت علمی گروه کامپیوتر دانشگاه امام رضا (ع)

# تعریف

- بستری برای برای پردازش توزیع شده روی **خوشه ای** [1] از کامپیوترها که سریع و همه منظوره است

1-cluster

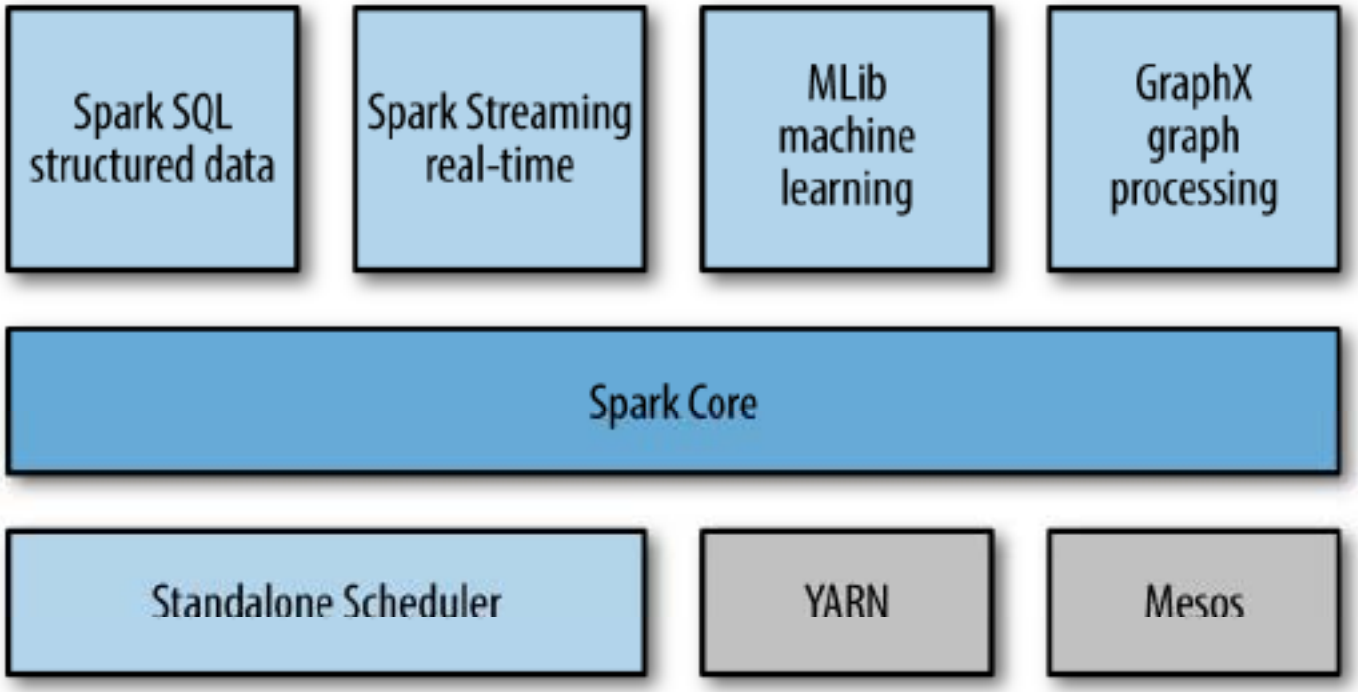
# تفاوتهای اصلی اسپارک و هدوپ

- **هدوپ** بستر کاملتری است، **اسپارک** HDFS ندارد
- **اسپارک** با **MapReduce** قابل مقایسه است
- در **اسپارک** بیشتر عملیات در **RAM** انجام می شود بر خلاف **MapReduce** که از **HDFS** استفاده می کند
- **معمولاً اسپارک در کنار هدوپ** نصب می شود

# تفاوتهای اصلی اسپارک و هدوپ

- **اسپارک** بدلیل استفاده بیشتر از **RAM** **هزینه بیشتری** دارد
- برای **مقابله با خطا**، هدوپ از Jobtracker و Tasktracker استفاده می کند، اسپارک از RDD
- هر دو **مقیاس پذیری** خوبی دارند. هدوپ در Yahoo ، ۴۲۰۰۰ گره، اسپارک ۸۰۰۰ گره

# ساختار اسپارک



# ساختار اسپارک

- **هسته اسپارک[1]**
- قابلیتهای پایه ای اسپارک از قبیل مدیریت عملیات، مدیریت حافظه، بازیابی پس از خطا، تعامل با سیستمهای ذخیره سازی، و ...

- محل دسترسی به API (*resilient distributed datasets*)RDD

1-Spark core

# ساختار اسپارک

- **زبان پرس و جوی ساخت‌یافته اسپارک[1]**
- برای کار با داده های ساخت‌یافته
- امکان استفاده از SQL
- امکان ترکیب SQL با تحلیل داده ها در یک برنامه

1-Spark SQL

# ساختار اسپارک

- **پردازش داده های جریانی اسپارک[1]**
- پردازش داده های جریانی

1-Spark Streaming

# ساختار اسپارک

- **کتابخانه یادگیری ماشین اسپارک (Spark MLlib)**
- کتابخانه ای برای یادگیری ماشین

- **پردازش گراف اسپارک (Spark GraphX)**
- کتابخانه ای برای پردازش گراف

# ساختار اسپارک

- **مدیریت خوشه**

- خود اسپارک
- HadoopYARN
- Mesos

# تولید نرم افزار در اسپارک

- در بستر **اسپارک** می توان **یک نرم افزار کاربردی برای پردازش و تحلیل داده های حجیم** تولید کرد.

- برای تولید نرم افزار از زبانهای ***Python,Scala,Java,SQL*** می توان استفاده کرد