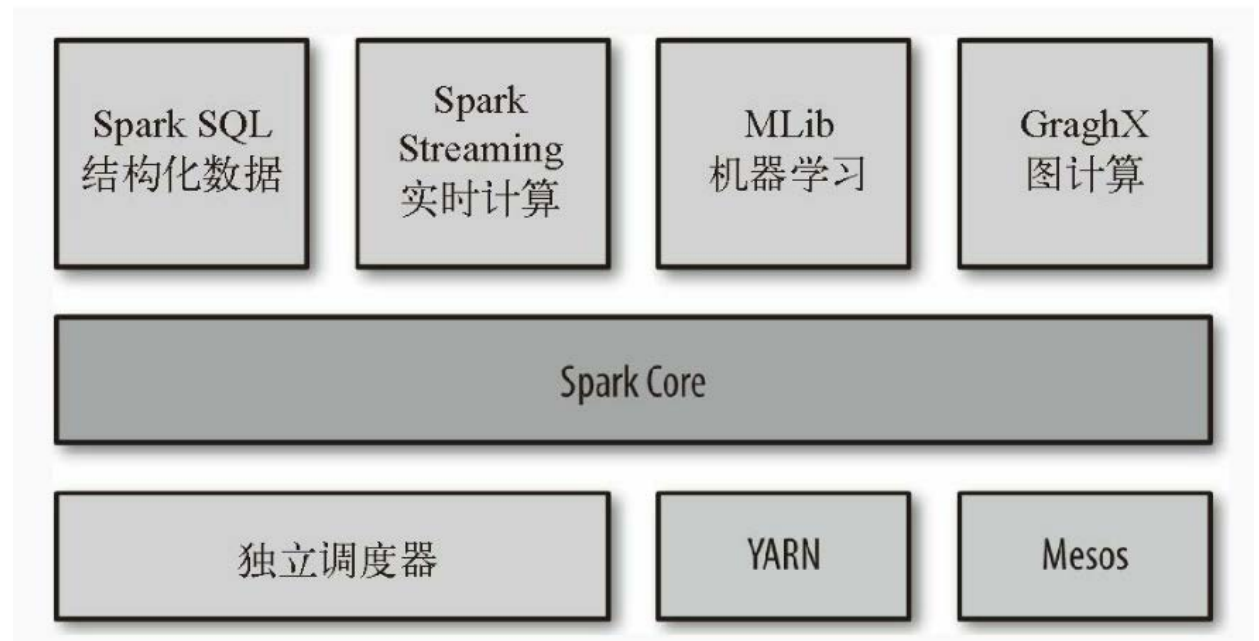


一本星火单词，助我搞定英语。
一本 Spark，助我搞定计算机！

Scala 也是一种 program language
<https://www.scala-lang.org/>



Spark Core

Spark Core 实现了 Spark 的基本功能，包括任务调度，内存管理，错误恢复，与存储系统交互等模块。

RDD: Resilient Distributed Dataset: 高适应性分布式数据集。

RDD 表示分布在多个计算机节点上可以并行操作的元素集合，是 Spark 主要的编程抽象。

Spark Core 提供了创建和操作这个集合的多个 API。

Spark SQL

Spark SQL 是 Spark 操作结构化数据的程序包。

Spark Streaming

Spark Streaming 是 Spark 提供的对实时数据进行流式运算的组件。

MLlib 机器学习组件

GraphX 操作图的程序库，可以进行并行的图计算。社交网络中的朋友关系图。

Spark 支持各种集群管理器。Cluster Manager。

包括 Hadoop YARN，Apache Mesos，以及 Spark 自带的一个简易调度器。

Spark 的用户和用途：

数据科学家与工程师。

数据的科学应用和数据的处理应用。

Hadoop 并非是 Spark 的必要条件，Spark 支持任何实现了 Hadoop 接口的存储系统。