Hadoop官网：https://hadoop.apache.org/

hadoop分为海量数据存储 HDFS 和海量数据分析 mapreduce

狭义的Hadoop：

Hadoop公用程序：支持其他Hadoop模块的公用程序。

Hadoop分布式文件系统（HDFS™): 提供对应用程序数据的高吞吐量访问的分布式文件系统。

Hadoop YARN：作业调度和集群资源管理的框架。

Hadoop MapReduce：一个基于YARN的大型数据集并行处理系统

广义的Hadoop：

**[Ambari™](https://ambari.apache.org/)**: 一个基于web的工具，用于配置、管理和监控Apache Hadoop集群，其中包括对Hadoop HDFS、Hadoop映射减少、蜂巢、H目录、HBase、动物管理员、Oozie、猪和Sqoop的支持。Ambari还提供了一个查看集群运行状况的仪表板，如热图，以及可视化查看MapReduce、猪和蜂巢应用程序的能力，以及以用户友好的方式诊断其性能特征的特性。

**[Avro™](https://avro.apache.org/)**: A data serialization system.

**[Cassandra™](https://cassandra.apache.org/)**: A scalable multi-master database with no single points of failure.

**[Chukwa™](https://chukwa.apache.org/)**: A data collection system for managing large distributed systems.

**[HBase™](https://hbase.apache.org/)**: A scalable, distributed database that supports structured data storage for large tables.

**[Hive™](https://hive.apache.org/)**: A data warehouse infrastructure that provides data summarization and ad hoc querying.

**[Mahout™](https://mahout.apache.org/)**: A Scalable machine learning and data mining library.

**[Ozone™](https://ozone.apache.org/)**: A scalable, redundant, and distributed object store for Hadoop.

**[Pig™](https://pig.apache.org/)**: A high-level data-flow language and execution framework for parallel computation.

**[Spark™](https://spark.apache.org/)**: A fast and general compute engine for Hadoop data. Spark provides a simple and expressive programming model that supports a wide range of applications, including ETL, machine learning, stream processing, and graph computation.

**[Submarine](https://submarine.apache.org/)**: A unified AI platform which allows engineers and data scientists to run Machine Learning and Deep Learning workload in distributed cluster.

**[Tez™](https://tez.apache.org/)**: A generalized data-flow programming framework, built on Hadoop YARN, which provides a powerful and flexible engine to execute an arbitrary DAG of tasks to process data for both batch and interactive use-cases. Tez is being adopted by Hive™, Pig™ and other frameworks in the Hadoop ecosystem, and also by other commercial software (e.g. ETL tools), to replace Hadoop™ MapReduce as the underlying execution engine.

**[ZooKeeper™](https://zookeeper.apache.org/)**: A high-performance coordination service for distributed applications.

Hadoop 2.7安装流程

配置IP地址修改主机名和映射文件

配置jdk环境变量