

## 大数据技术-第五章：YARN资源管理器 YARN资源管理器概述



# CONTENTS

---

**01.** Hadoop1.0存在的问题   **02** YARN的概述

**03** YARN在Hadoop生态系统中的位置



# 01

## Hadoop1.0存在的问题



## ➤ Hadoop 1.0 存在的问题

HDFS 1.0 存在的问题：

- **Namenode 单点故障：**集群的文件都是以“块(block)”的形式存储，并且为了容错，每个block有多个副本。namenode需要记录整个集群所有block及其副本的元数据信息（fsimage：文件目录结构，block和文件的映射关系等）和操作日志（edits），因此，在hadoop1.0框架中，namenode设计为单个节点，通常部署在一台性能强劲的计算机上，客户端上传的所有文件，都要先与namenode进行交互，由namenode管理。这样，namenode的可用性决定整个集群的存储系统的可用性。
- **Namenode 压力过大，内存受限：**在集群启动时，namenode需要将集群所有block的元数据信息加载到内存，这样，当文件越来越多时，namenode的内存压力也会遇到瓶颈。

## » Hadoop 1.0 存在的问题

MapReduce 1.0 存在的问题：

- JobTracker 单点故障：MapReduce 作为 hadoop 的计算框架，其作业的调度和资源分配是通过 JobTracker 完成。但是，因为资源管理是全局性的，因此，JobTracker 仍然是单个节点，通常部署在一台性能强劲的计算机上。一旦该节点失败，整个集群的作业就全部失败。
- 不支持 MapReduce 之外的其它计算框架：hadoop 1.0 仅能支持 MapReduce 计算框架，随着人们对实时性的要求，spark、storm 等计算框架也应运而生，但是 hadoop 1.0 并不能良好地支持。
- JobTracker 访问压力大，影响系统扩展性：在 hadoop 1.0 中，JobTracker 负责所有作业的调度、各个作业的生命周期、各个作业中所有 task 的跟踪和失败重启等。因此，当集群中作业很多时，所有作业的 map task 和 reduce task 都要与 JobTracker 进行交互。



## » Hadoop 1.0 存在的问题

为解决hadoop1.0中JobTracker单点故障问题，hadoop2.0开始将Hadoop 1.0中的JobTracker的集群资源分配和作业调度分为两个守护进程来控制，分别是Global Resource Manager(以下简称RM)和每个Application的ApplicationMaster(以下简称AM)，这也是YARN最重要的两个组件。

# 02

## YARN的概述



Apache Hadoop YARN (Yet Another Resource Negotiator, 另一种资源协调者) 是一种新的Hadoop资源管理器, 它是一个通用的资源管理系统, 可为上层应用提供统一的资源管理和调度服务, 它的引入为集群在资源利用率、资源的统一管理调度和数据共享等方面带来了巨大好处。

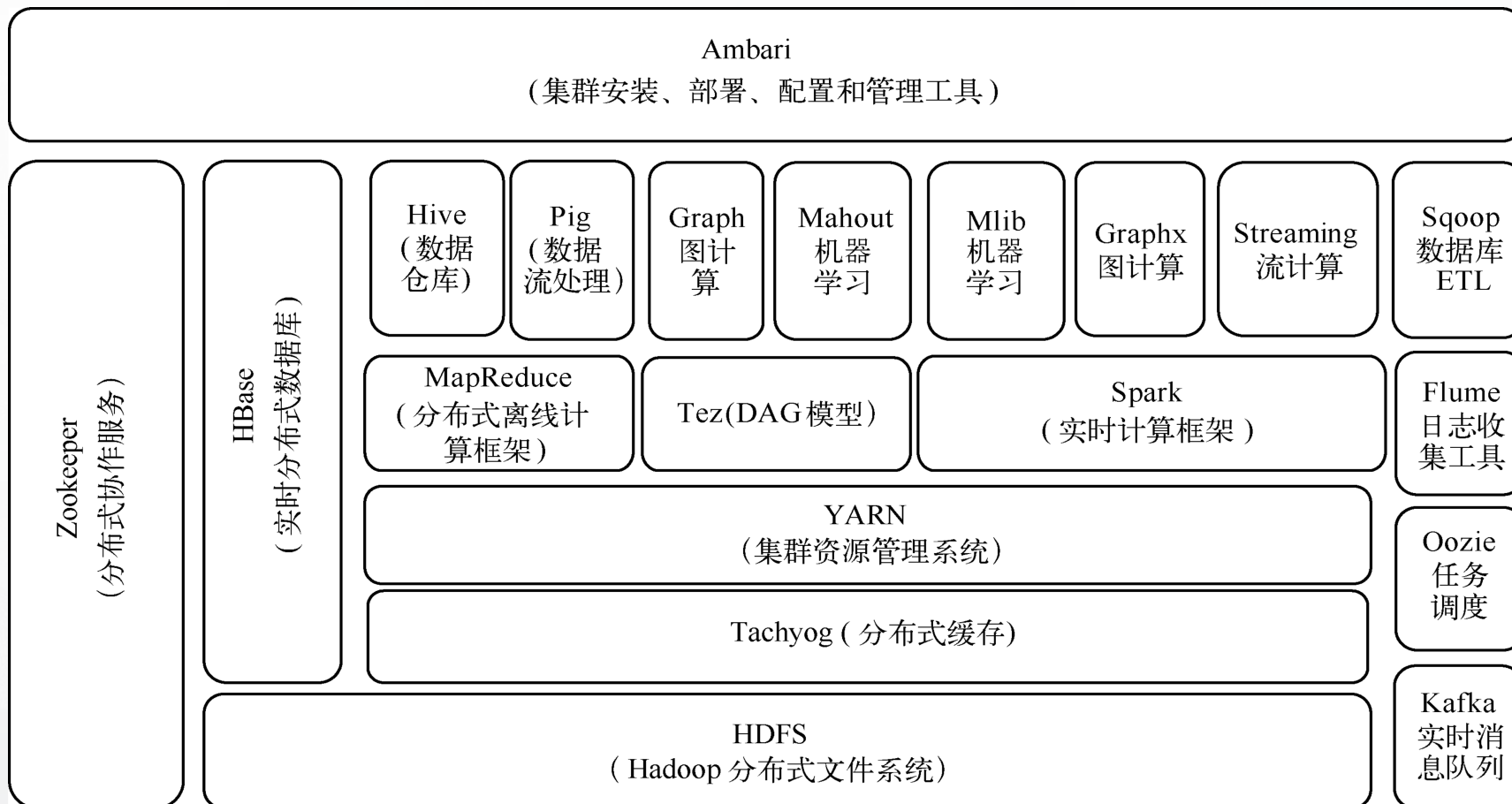


# 03

## YARN在Hadoop生态系统中的位置



# >> YARN在Hadoop生态系统中的位置



Turing AI 万维图灵 | 大数据系列课程

大数据

BIG  
DATA

智 / 能 / 科 / 技      放 / 眼 / 未 / 来

