

大数据技术架构主要包含大数据获取技术、分布式数据处理技术和大数据管理技术，以及大数据应用和服务技术。

(1) 大数据获取技术

目前，大数据获取的研究主要集中在数据采集、整合和清洗三个方面。数据采集技术实现数据源的获取，然后通过整合和清理技术保证数据质量。

(2) 分布式数据处理技术

分布式计算是随着分布式系统的发展而兴起的，其核心是将任务分解成许多小的部分，分配给多台计算机进行处理，通过并行工作的机制，达到节约整体计算时间，提高计算效率的目的。目前，主流的分布式计算系统有 Hadoop、Spark 和 Storm。Hadoop 常用于离线的复杂的大数据处理，Spark 常用于离线的快速的大数据处理，而 Storm 常用于在线的实时的大数据处理。

(3) 大数据管理技术

大数据管理技术主要集中在大数据存储、大数据协同和安全隐私等方面。

(4) 大数据应用和服务技术

大数据应用和服务技术主要包含分析应用技术和可视化技术。

4、区块链（掌握）

区块链技术具有多中心化存储、隐私保护、防篡改等特点，提供了开放、分散和容错的事务机制，成为新一代匿名在线支付、汇款和数字资产交易的核心，被广泛应用于各大交易平台，为金融、监管机构、科技创新、农业以及政治等领域带来了深刻的变革。

1. 技术基础

区块链概念可以理解为以非对称加密算法为基础，以改进的默克尔树（Merkle Tree）为数据结构，使用共识机制、点对点网络、智能合约等技术结合而成的一种分布式存储数据库技术。区块链分为公有链（Public Blockchain）、联盟链（Consortium Blockchain）、私有链（Private Blockchain）和混合链（Hybrid Blockchain）四大类。

一般来说，区块链的典型特征包括：**【口诀：多多识（时）字（智）不开心（信）】**

- **多中心化**：链上数据的验证、核算、存储、维护和传输等过程均依赖分布式系统结构，运用纯数学方法代替中心化组织机构在多个分布式节点之间构建信任关系，从而建立可信的分布式系统。

- **多方维护**：激励机制可确保分布式系统中的所有节点均可参与数据区块的验证过程，并通过共识机制选择特定节点将新产生的区块加入到区块链中。

- **时序数据**：区块链运用带有时间戳信息的链式结构来存储数据信息，为数据信息添加时间维度的属性，从而可实现数据信息的可追溯性。

- **智能合约**：区块链技术能够为用户提供灵活可变的脚本代码，以支持其创建新型的智能合约。

- **不可篡改**：在区块链系统中，因为相邻区块间后序区块可对前序区块进行验证，若篡改某一区块的数据信息，则需递归修改该区块及其所有后序区块的数据信息，然而每一次哈希的重新计算代价是巨大的，且须在有限时间内完成，因此可保障链上数据的不可篡改性。

- **开放共识**：在区块链网络中，每台物理设备均可作为该网络中的一个节点，任意节点可自由加入且拥有一份完整的数据库拷贝。

- **安全可信**：数据安全可通过基于非对称加密技术对链上数据进行加密来实现，分布式系统中各节点通过区块链共识算法所形成的算力来抵御外部攻击、保证链上数据不被篡改和伪造，从而具有较高的保密性、可信性和安全性。

2. 关键技术

(1) 分布式账本