

知 H3C大数据平台架构怎么理解?

厉梦如 2019-05-19 发表

问题描述

Q: H3C大数据平台架构怎么理解?

解决方法

A: H3C大数据平台包含4个部分:

第一部分是运维管理, 包括: 安装部署、配置管理、主机管理、用户管理、服务管理、监控告警和安全管理等。

第二部分是数据ETL, 即获取、转换、加载, 包括: 关系数据库连接Sqoop、日志采集Flume、ETL工具 Kettle。

第三部分是数据计算。MPP采用分析型分布式数据库, 存储高价值密度的结构化数据; Hadoop存储非结构化/半结构化数据和低价值密度结构化数据。计算结果都存到数据仓库, 数据仓库中的数据可直接用于分析和展示。数据仓库是面向主题的、集成的、稳定的且随时间不断变化的数据集合, 用以支持经营管理中的决策制定过程。

第四部分数据服务, 包括: 机器学习、数据挖掘、数据检索、数据可视化、即席分析、SQL和API, 为应用层提供服务和中间件调用