

1、 演讲主题；

大数据治理与成本管控实践

2、 演讲主题简介；

1).概要介绍：

随着智能化时代的到来，随着人工智能、云计算等技术的推动，业务复杂度持续增加，全球数据量正在无限制地扩展，而传统数仓方式面临着单机存储瓶颈与计算分析执行时间指数级增加等问题难以承载海量数据计算，同时数据变的越来越重要，对企业而言数据已经成为战略性资源，因此大数据架构使用 HDFS 等分布式架构作为存储引擎解决海量数据存储，使用 Hive/Spark 等分布式计算引擎解决数据计算问题，用大规模机器组成集群提供高效稳定的数据存储与超大并行计算能力，但面临纷杂数据的无限制生产，如何进行数据规范制作以及进行有效的数据治理与管控成为很多企业必须要面对的挑战。

本次分享将从数据的生产、计算、消费等方面介绍数据仓库与任务调度管理建设规范，通过建设元数据管理系统打通数据生成与消费的全链路流程，以及通过建设数据与任务的健康度评估数据资产。同时也介绍通过建设指标管理与数据质量等体系完善的数据评估模型进行存储与计算成本的管控与提升数据管理能力。

2).内容大纲：

1.大数据架构

2. 数据仓库系统建设

1) 数据仓库元数据管理

2) 数据仓库规范建设

3) 数据集市

3. 数据规范建设

3.1 OneData 系统设计

3.1.1 规范需求

3.1.2 全链路平台管理

3.2 指标系统设计

3.2.1 指标管理设计

3.2.2 维度管理

3.2.3 修饰词管理

3.3 数据地图简介

4. 数据管控

4.1 数据治理体系

4.1.1 元数据系统建设

4.1.2 基础数据管理系统

4.1.3 任务画像系统建设

4.2 数据质量保障体系

4.3 数据资产评估