

数据中台（银行金融）

① 统一-工具体系

统一数据的存储标准
日志存储标准
实时数据处理标准

“收集 - 治理 - 整合 - 消费”全数据链路

⑤ 全域数据体系

比如 统一-指标、用户标签等
OneID
oneData
Onservice

全域的数据体系

打通公司所有业务数据，形成统一-数据资产

建设数据中台典型支撑技术

① 开发平台

数据仓库开发与管理平台

大数据计算引擎
数据集成 / 同步 / 交换引擎

自助分析平台

发展趋势

增强分析工具

敏捷BI系统
利用BI工具中的机器学习和AI来自自动化数据准备工作以及帮助用户发现和分享见解

NLP搜索

NLP语音搜索
指计算机程序可理解非正式人类语言的能力
帮助用户通过文本查询高效获取数据见解

图形分析

指一组分析技术，它允许探索和映射数据集（例如企业、人员和交易等）之间的关系

② 服务平台

数据服务平台

③ 运维平台

指标管理平台

元数据管理平台

数据质量管理平台

数据安全管理平台

数据资产管理平台

② 基于工具体系构建数据仓库

“原始层 - 明细层 - 汇总层 - 应用层”

经典数仓架构

④ 服务业务

场景驱动数据产品与服务

风控应用

数据报表

实时指标等

数据中台 vs 数据湖

① 数据湖

汇集企业的各个数据源，以数据聚集、加工为目的的数据资源池
解决聚集问题，不可控需求导致数据加工异常繁重
数据的繁杂和混乱引入数据治理让数据加工更加举步维艰

② 数据中台

更多实现数据的管理，强调利用数据的能力，强调数据开发和高效的使用
数据资产管理可以对数据湖中的数据按照数据域方式进行管理并结合业务的逻辑实现整个数据模型的加工和开发
数据中台按分析的原子指标和派生指标方式做计算并存储在Maxcompute平台上

如有及时查询要求会同步分析结果数据给MPP或其他DB

③ 相关术语

	数据湖	数据中台
数据结构	结构化、半结构化、非结构化	结构化、非结构化
数据处理	只读为主	读写均有
架构设计	分布式大数据系统。以Hadoop存储和计算为主	数据加工处理基于Maxcompute或Hadoop Hive和HDFS为主，业务数据查询基于分布式MPP和KV数据库
数据管理和分析	无	有
数据资产管理	无	有
成熟的产品化开发管理工具	无	有
数据加工处理	以SQL、UDF为主，根据需求而定做数据加工，没有固定逻辑	阿里云数据中台方法论贯穿数据加工过程
使用者	数据专业开发人员	业务负责人，一线业务人员，一线技术人员或技术负责人

数据中台与数据湖

	应用场景	计算存储	产品功能	数据组织	价值体现
DW1.0	报表	数据库选型：Oracle、Teradata、Greenplum	PLSQL、厂商自研主要涵盖数据交换和数据开发	面向业务主题指标体系	面向业务人员的辅助决策
DW2.0	报表	Hadoop生态选型：开源Hadoop、CDH	开源：kettle、Sqoop、Azkaban、Hue、主要涵盖数据交换、数据开发、任务调度	面向业务主题指标体系（准实时、预测类指标）	面向业务人员的辅助决策
数据湖	无	Hadoop生态选型：开源Hadoop、CDH	开源：kettle、Sqoop	原始	无
数据中台（建设工具）	多元化场景：报表、商品推荐、精准广告展示、个性化信息推送、销量预测、客满评价等	多元化选型：Hadoop、Spark、Flink、Kylm、Greenplum	开源：kettle、Sqoop、Azkaban和自研，涵盖数据交换、数据开发（调度）、数据治理、数据服务（引擎*）等。	面向场景驱动的数据资产体系	面向业务系统的优化升级

数据平台迭代历史