

Architecture of Enterprise Applications 20

Data Lake

Haopeng Chen

***RE**liable, **IN**telligent and **SC**alable Systems Group (**REINS**)*

Shanghai Jiao Tong University

Shanghai, China

<http://reins.se.sjtu.edu.cn/~chenhp>

e-mail: chen-hp@sjtu.edu.cn

- Contents
 - PolarDB
 - MaxCompute
 - Data Lake
- Objectives
 - 能够了解主流云原生数据库的基本概念和特点，理解数据湖的特征的技术难点

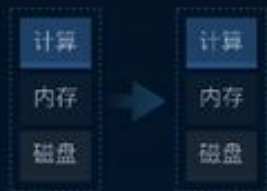
- 阿里巴巴自研的新一代云原生关系型数据库，存储与计算分离
- 为用户提供具备极致弹性、高性能、海量存储、安全可靠的数据服务
- 提供100T级别数据库容量，6倍于MySQL性能，重新定义云数据库标准

为了无法计算的价值 | 阿里云

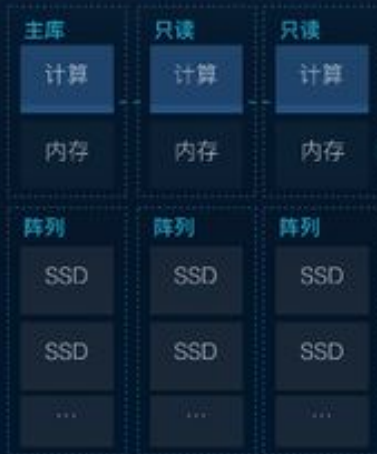
POLARDB

基于可靠且速度惊人的智慧存储系统

第一代 本地盘



第二代 集中式存储



第三代 分布式共享存储



POLARDB 性能冲击世界顶峰

为了无法计算的价值 | 阿里云

性能6倍于MySQL —— 100TB存储容量 —— 100%向下兼容MySQL 5.6

100万QPS



读性能

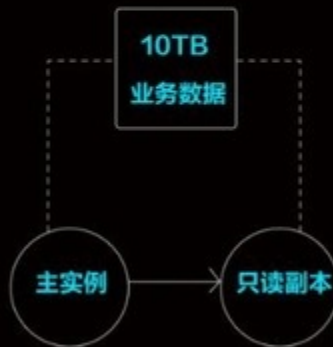
13万TPS



写性能

2分钟

过去需要70小时



创建只读副本

3分钟

全球范围



创建容灾实例

知乎 @裕裕同学

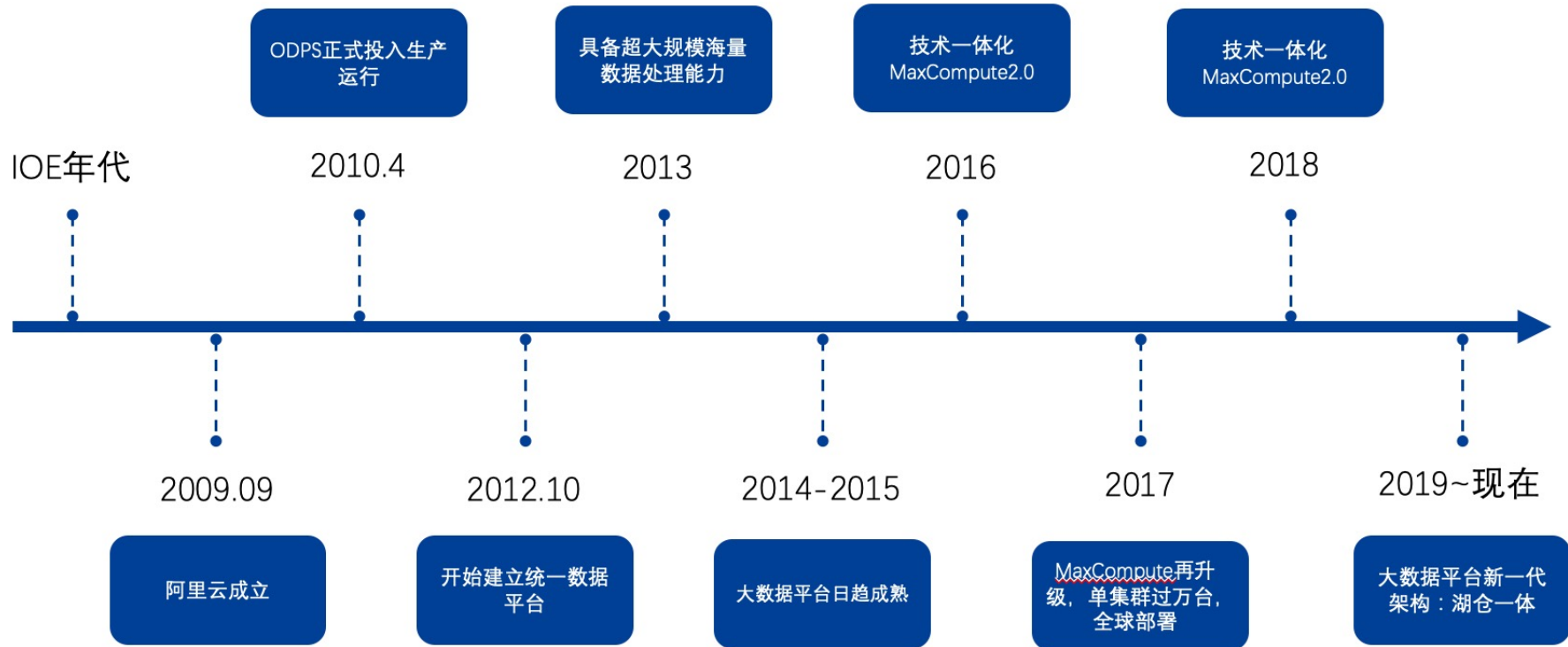
- 基于CPU和内存分离的分布式共享内存池，实现CPU和内存解耦
 - 多个CPU共享同一份内存，降低内存资源开销
- 基于分布式共享内存池的多点可写技术
 - 多个CPU可以共享同一个内存池来交换页面信息和事务信息，同时结合高性能RDMA网络和NVM
- 行列混存HTAP架构
 - 列存引擎使用Delta-Main的存储架构，直接在内存的Delta Store中支持DML的增删改操作
- One size fits all
 - 一个数据库即可满足现时多类数据库混合使用

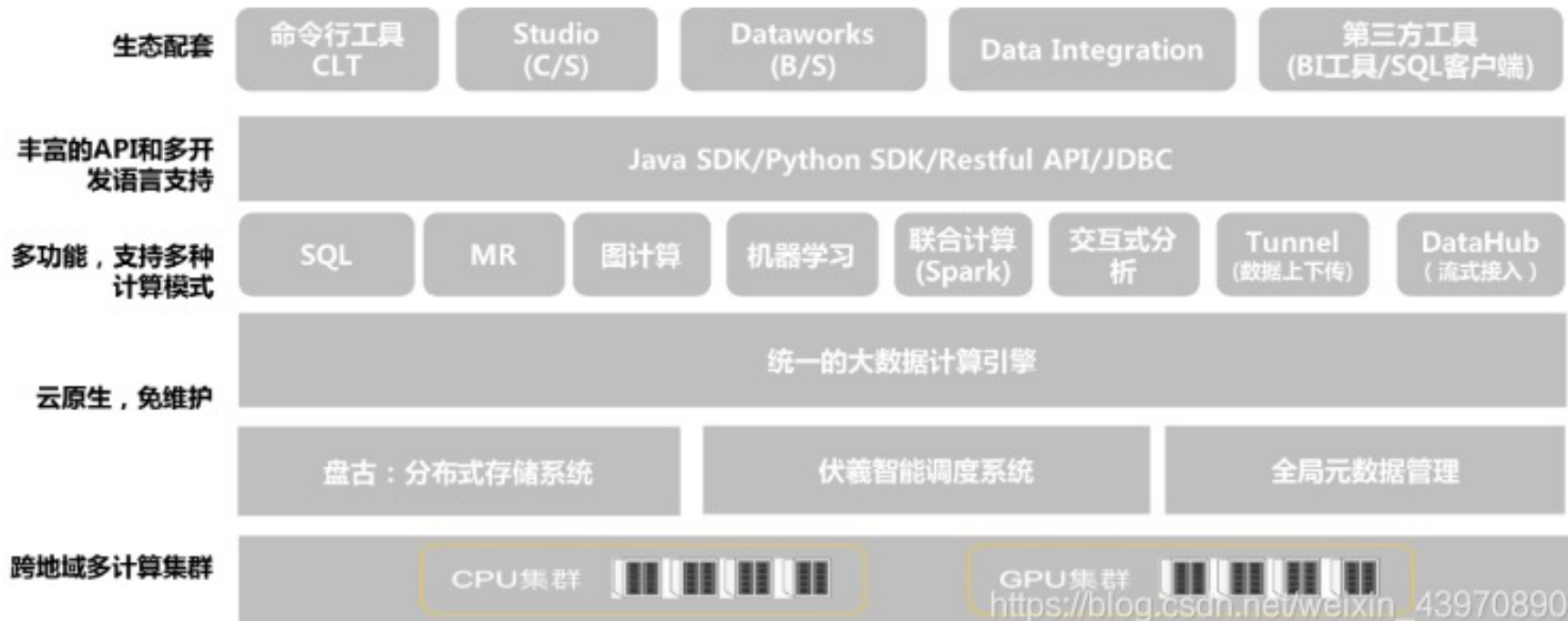
- MaxCompute 内部产品名为ODPS,
 - 是一个**高效能、低成本、完全托管**的**EB级**大数据计算服务
- 内部服务
 - 对内承担阿里最大的存储和计算量，是阿里数据的汇合点
- 对外服务
 - 在公有云和专有云上对外部客户提供服务
- 成熟的商用平台
 - 经过双十一的考验，具有完善的开发、部署、运维、基线保障等能力

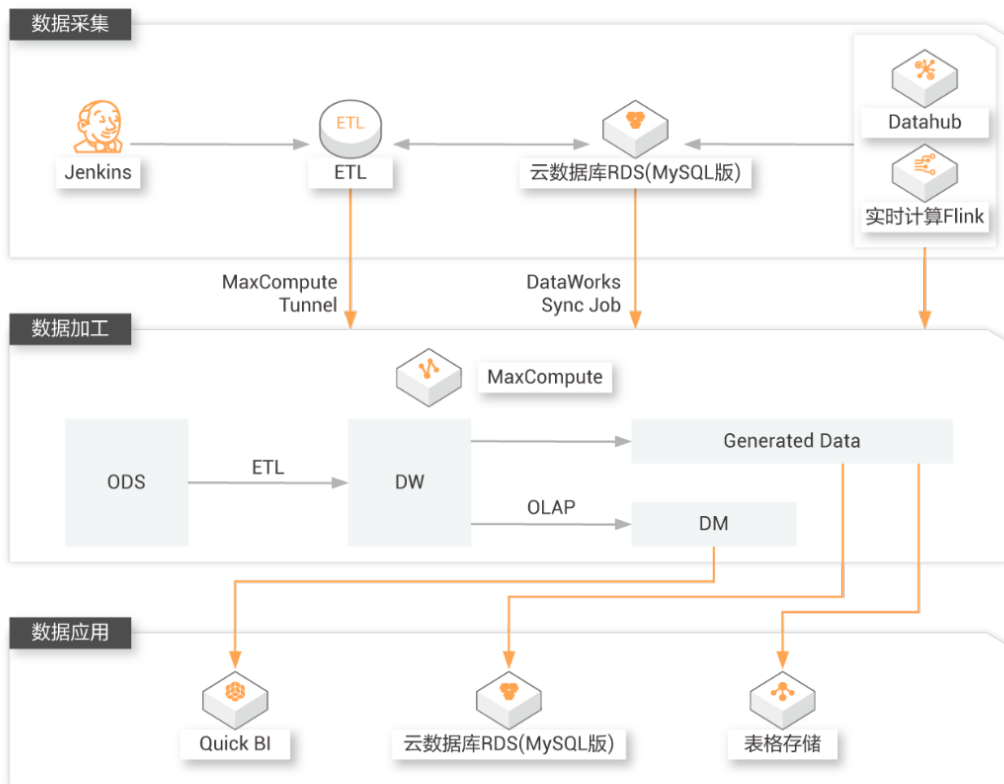
MaxCompute – 从ODPS到MaxCompute 2.0



REliable, INtelligent & Scalable Systems



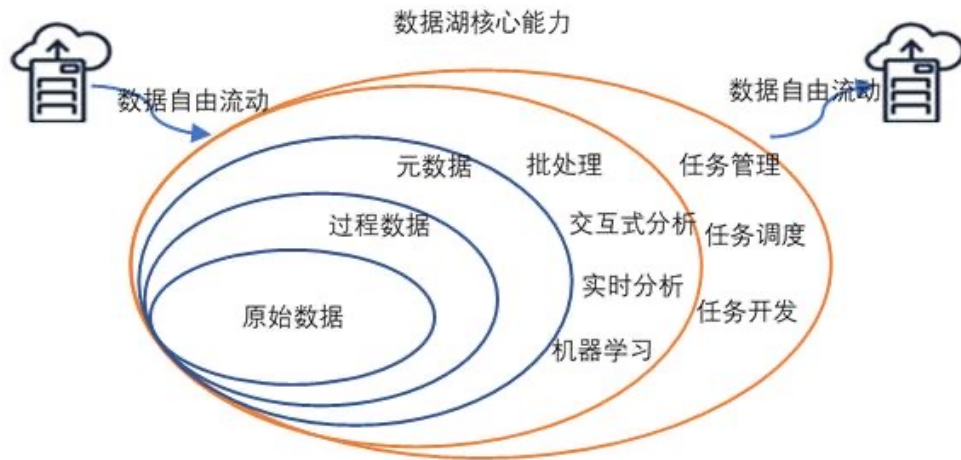




- 支持大规模计算存储
- 数据分布式存储，多副本冗余，数据存储对外仅开放表的操作接口
- 自研数据存储结构，表数据列式存储，默认高度压缩
- 支持外表，将存储在OSS对象存储、OTS表格存储的数据映射为二维表
- 支持Partition、Bucket的分区
- 底层不是HDFS，是阿里自研的盘古文件系统
- 存储与计算解耦，不需要仅仅为了存储扩大不必要的计算资源
- ...

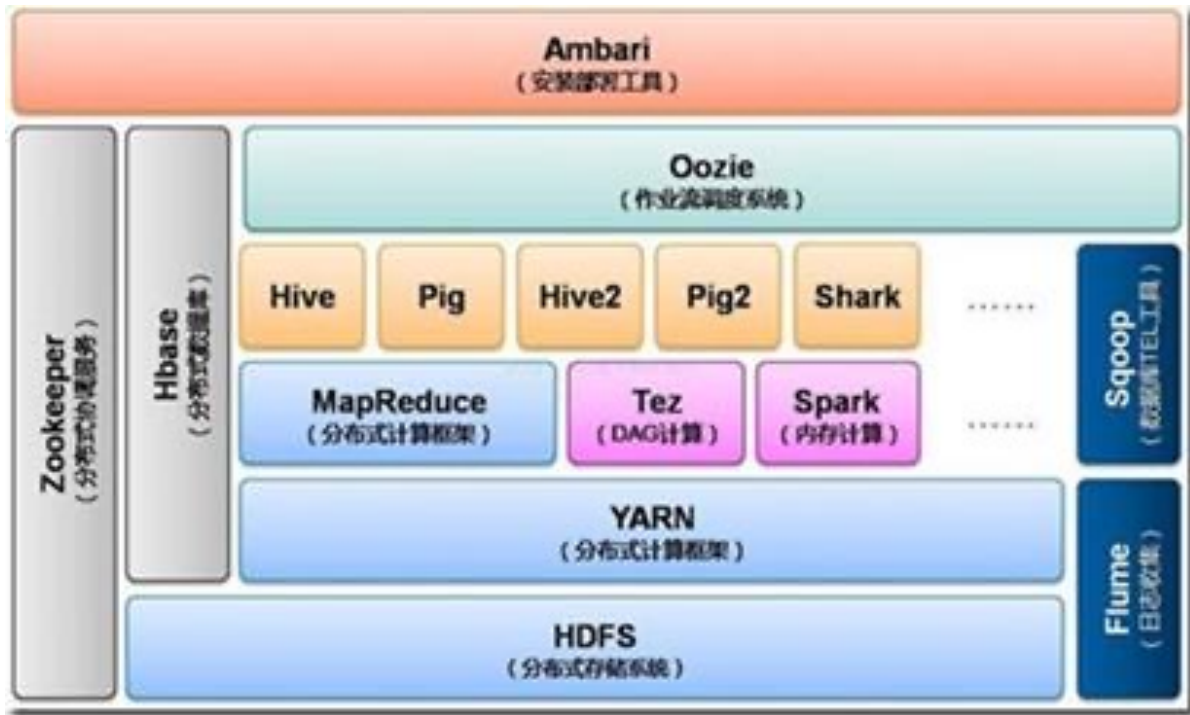
- MaxCompute SQL
 - TPC-DS 100% 支持, 同时语法高度兼容Hive
- MapReduce
 - 支持MapReduce编程接口
- MaxCompute Graph
 - 面向迭代的图计算处理框架, 用于PageRank、单源最短距离
- PyODPS
 - Python SDK
- Spark
 - Spark on MaxCompute, 使MaxCompute提供兼容开源的Spark计算服务

- Data lake是指使用大型二进制对象或文件这样的自然格式储存数据的系统，通常把所有的企业数据统一存储，既包括源系统中的原始副本，也包括转换后的数据
—— Wikipedia
- 是一种不断演进中、可扩展的大数据存储、处理、分析的基础设施
- 以数据为导向，实现任意来源、速度、规模、类型数据的全生命周期管理

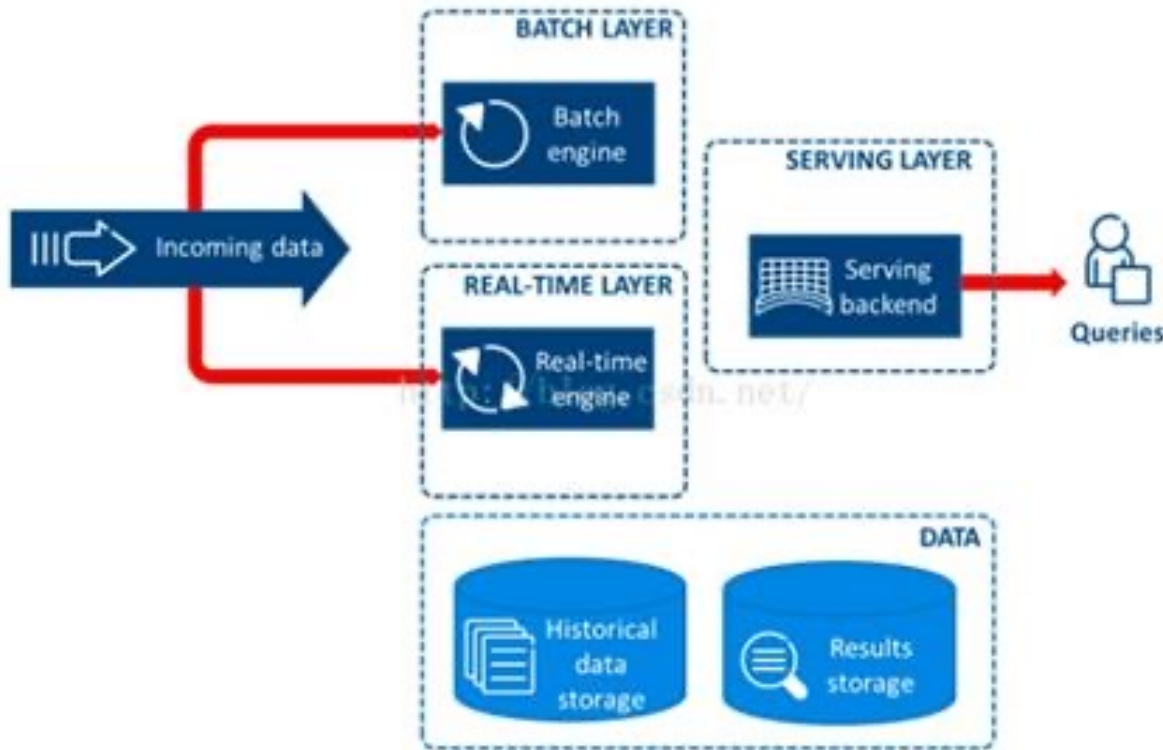


Data warehouse	Data lake
数据体系严格，提前建模	数据体系松散，事后建模
灵活性较低	灵活性较高
数据治理容易	数据治理困难
数据种类单一（结构化、半结构化）	数据种类丰富（结构化、半结构化、非结构化）
面向成熟数据的企业级分析与处理	面向异构数据的科学探查与价值挖掘
向特定引擎开放，易获得高度优化	向所有引擎开放，各引擎有限优化

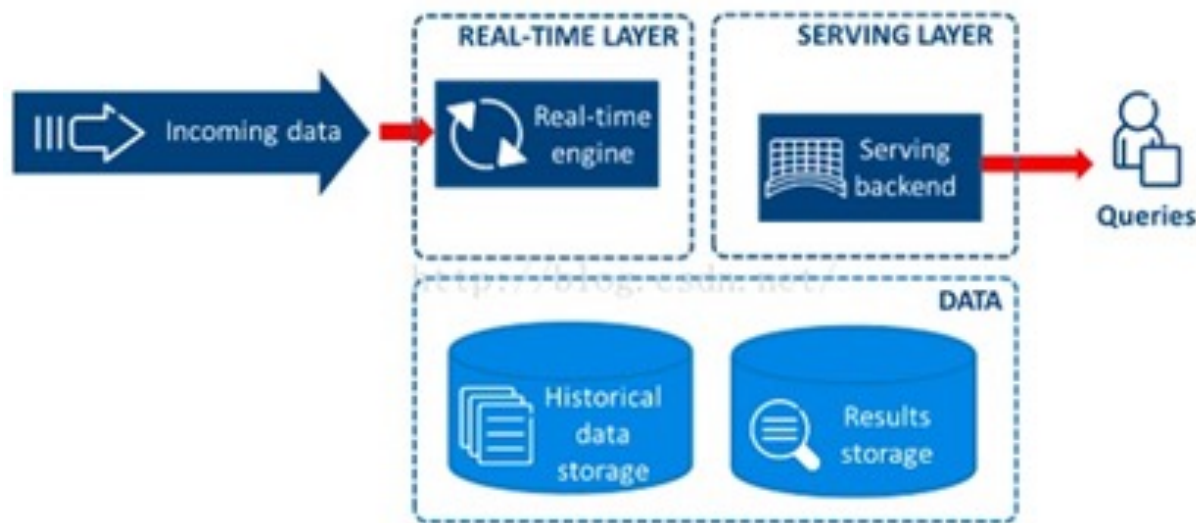
- 以HDFS为核心存储，以MapReduce为基本计算模型

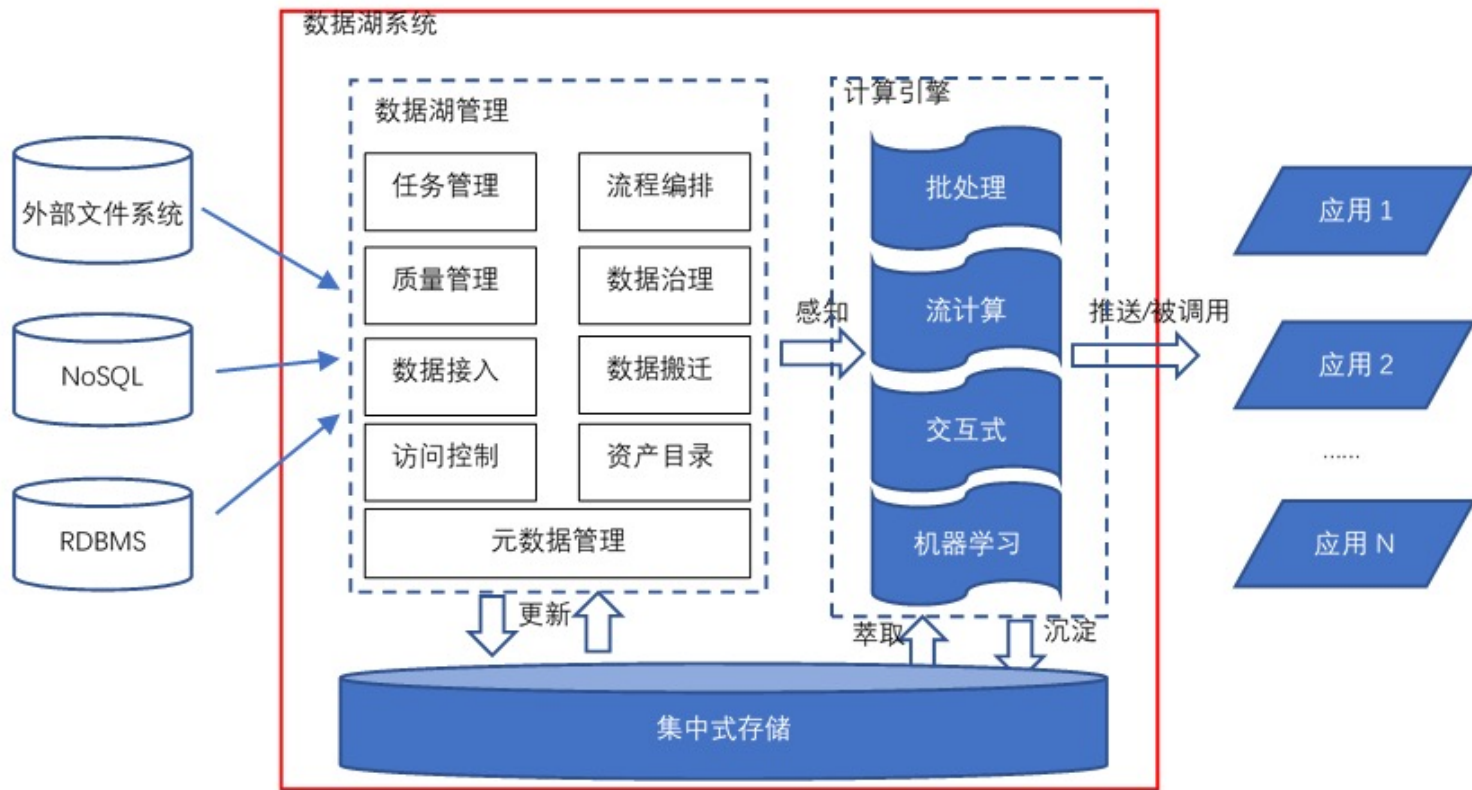


- HDFS批处理无法满足实时性要求高的场景，因此产生了lambda架构“流批一体”

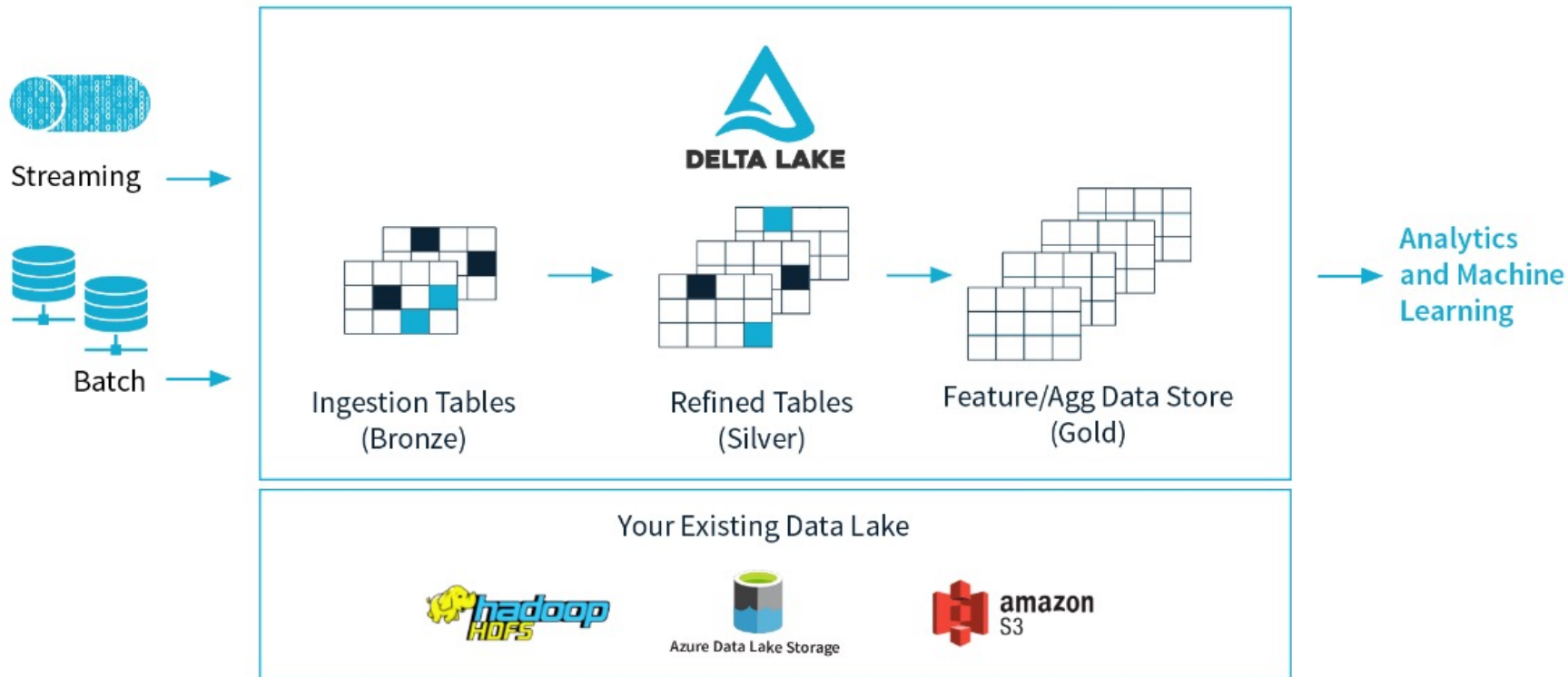


- 加大流计算的并发性与“时间窗口”，统一流/批处理



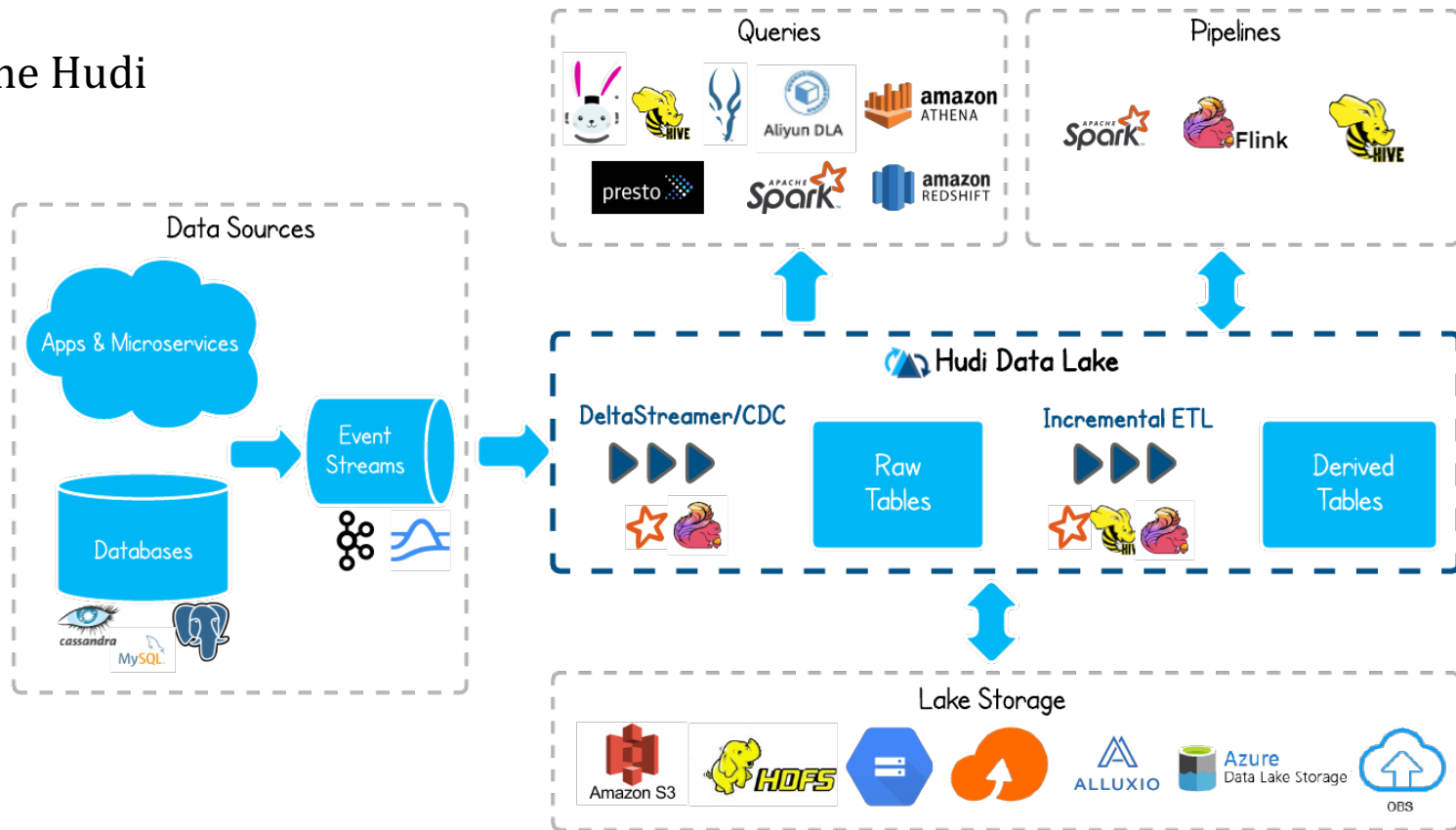


- Delta Lake



数据湖 – 湖仓一体(Lakehouse)

- Apache Hudi



- Apache Iceberg
- **Apache Iceberg is an open table format for huge analytic datasets.**
 - Iceberg adds tables to compute engines including Spark, Trino, PrestoDB, Flink and Hive using a high-performance table format that works just like a SQL table.



- 什么是PolarDB-阿里云帮助中心
 - https://help.aliyun.com/document_detail/58764.html
- 最了不起的国产数据库—PolarDB-知乎文章
 - <https://zhuanlan.zhihu.com/p/351877136>
- MaxCompute 2.0—从ODPS到MaxCompute
 - <https://developer.aliyun.com/article/64259>
- 什么是MaxCompute-阿里云帮助中心
 - https://help.aliyun.com/document_detail/27800.html
- 阿里云-湖仓一体
 - <https://www.aliyun.com/activity/bigdata/maxcompute-lakehouse?accounttraceid=32894aac682d4c6f86f56155bba99b22tdgk>

- 一文读懂Data Lake
 - <https://blog.csdn.net/u011598442/article/details/106610486/>
- 维基百科-data lake
 - https://en.wikipedia.org/wiki/Data_lake
- Delta lake
 - <https://delta.io/>
- Apache hudi
 - <https://hudi.apache.org/>
- Apache Iceberg
 - <https://iceberg.apache.org/>
- Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics
 - http://cidrdb.org/cidr2021/papers/cidr2021_paper17.pdf



Thank You!