

利用开源技术支撑 大规模数据存储与处理需求

周彦伟@极数云舟

内容简介：基于MySQL开源数据库实现数字化转型数据处理新引擎

- 数字化转型新需求
- MySQL的特点与优势
- 结合开源技术实现内核级升级
- 顺应需求实现边缘计算与中心分布式数据库

个人简介：周彦伟

- 极数云舟创始人、CEO
- 云舟数据库ArkDB总架构师
- 浙江大学硕士研究生校外导师
- 中国计算机行业协会数据库专委会会长
- Oracle ACE Director, All China MySQL User Group主席
- 出版：《MySQL运维内参》《Learning HBase（中文版）》《MySQL 8 Cookbook（中文版）》

大规模数据的需求

- **2021年**，《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》中“数字化”出现25次。
- **2020年**，中共中央国务院发布《关于构建更加完善的要素市场化配置体制机制的意见》，明确数据为生产要素、指出加快培育数据要素市场。
- **2020年**，国资委发布《关于加快推进国有企业数字化转型工作的通知》，面向国有企业数字化转型提出指导意见及具体步骤。
- **2020年5月13日下午**，国家发展改革委官网发布“数字化转型伙伴行动”倡议。

数字化转型要建设什么？

- 业务方向
 - 业务架构
 - 业务拓展
 - 业务创新
 - 业务融合
- 技术方向
 - 技术架构升级-从耦合到解耦
 - 技术体系升级-从集中式到分布式
 - 技术服务升级-从人工管理到自动运维
 - 数据地位升级-从附属支撑到核心竞争力

数字化转型中的数据建设

- 数据来源
 - 信息化数据
 - 物联网数据
 - 工业实时数据
 - 环境检测数据
- 数据技术
 - 数据接入
 - 数据存储
 - 数据处理
 - 数据管理

痛点

- 数据开发不规范
- 传统信息化孤岛林立，数据提供成本高，应用效率低
- 企业海量的高价值数据分散在各个业务系统中，前期由于没有数据标准导致数据接口复杂各异，难统一集中与整合
- 信息系统架构复杂，数据处理软件林立，技术壁垒制约发展
- 信息化数据、物联网数据、工业生产数据种类繁多，格式差异大
- 数据源及访问模式繁多(JDBC/ODBC/API/直连/文本文件)
- 数据库类型繁多(Oracle/MS/ArkDB/MySQL/DM/神通/金仓/南大通用/大数据)
- 数据库与数据系统缺乏统一管理和统一调度

关于MySQL

技术特点与优势

- 天生分层架构：MySQL Server+InnoDB Engine
- 插件式引擎思想：InnoDB、MyISAM、ArkDB
- 成熟的日志机制：REDO Log、Binlog
- 统一文件存储接口
- 轻量级线程处理模型

应用特点与优势

- 接入简单，编程灵活
- SQL简单，方便快捷
- 普及度高，装机量大
- 社区活跃，开源力量强大

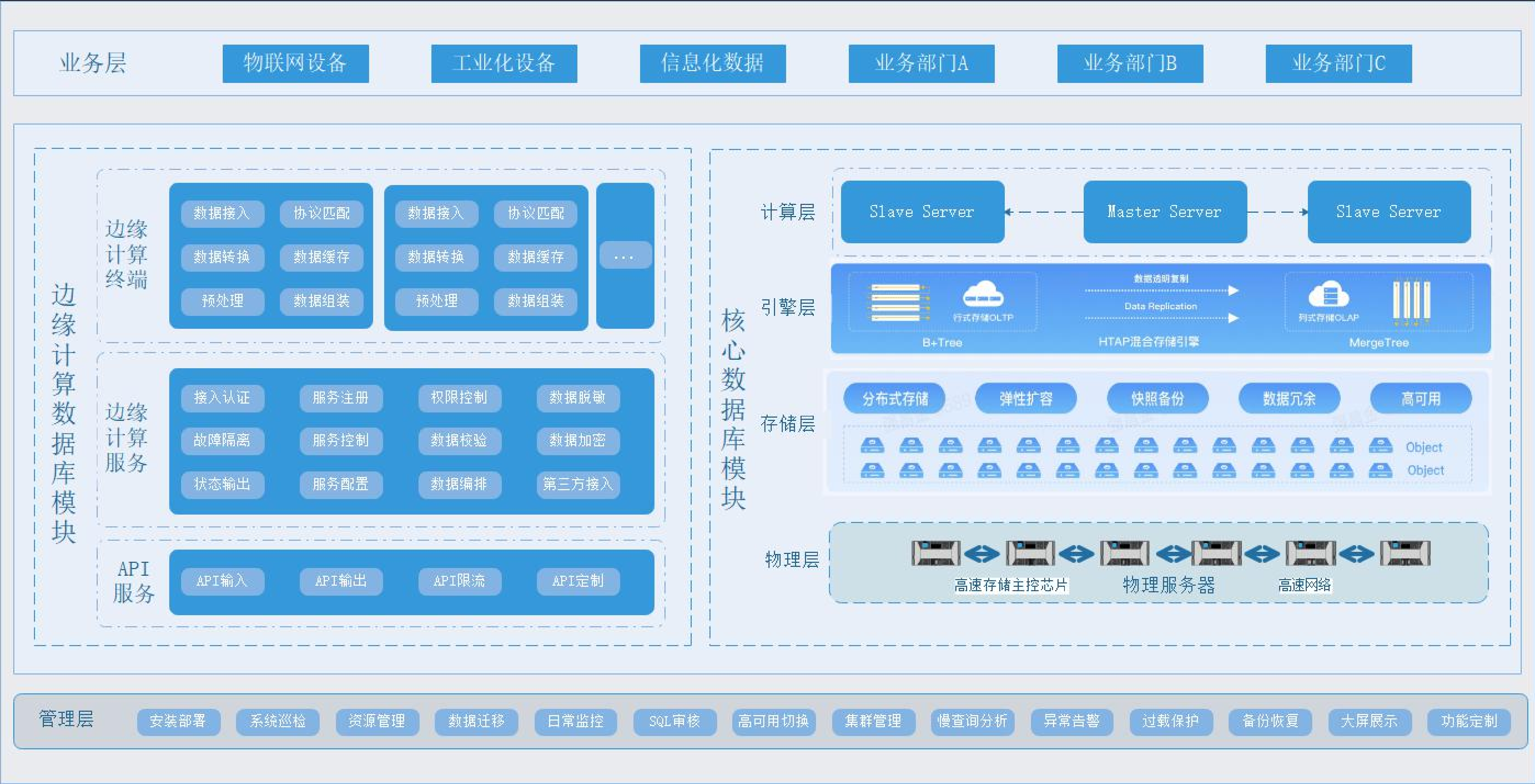
基于MySQL的技术架构拓展

- 云原生数据库思想指导
 - 高可用、高性能
 - 物理复制
 - 计算与存储分离
 - 弹性扩容
 - 并行计算
 - 分布式存储
- MySQL体系支撑
 - 数据接入与迁移
 - MySQL语法体系
 - 混合数据处理引擎
 - MySQL插件式引擎
 - 数据定制化写入
 - MySQL内核API
 - 数据管理
 - MySQL开源运维体系

边缘计算数据库探索

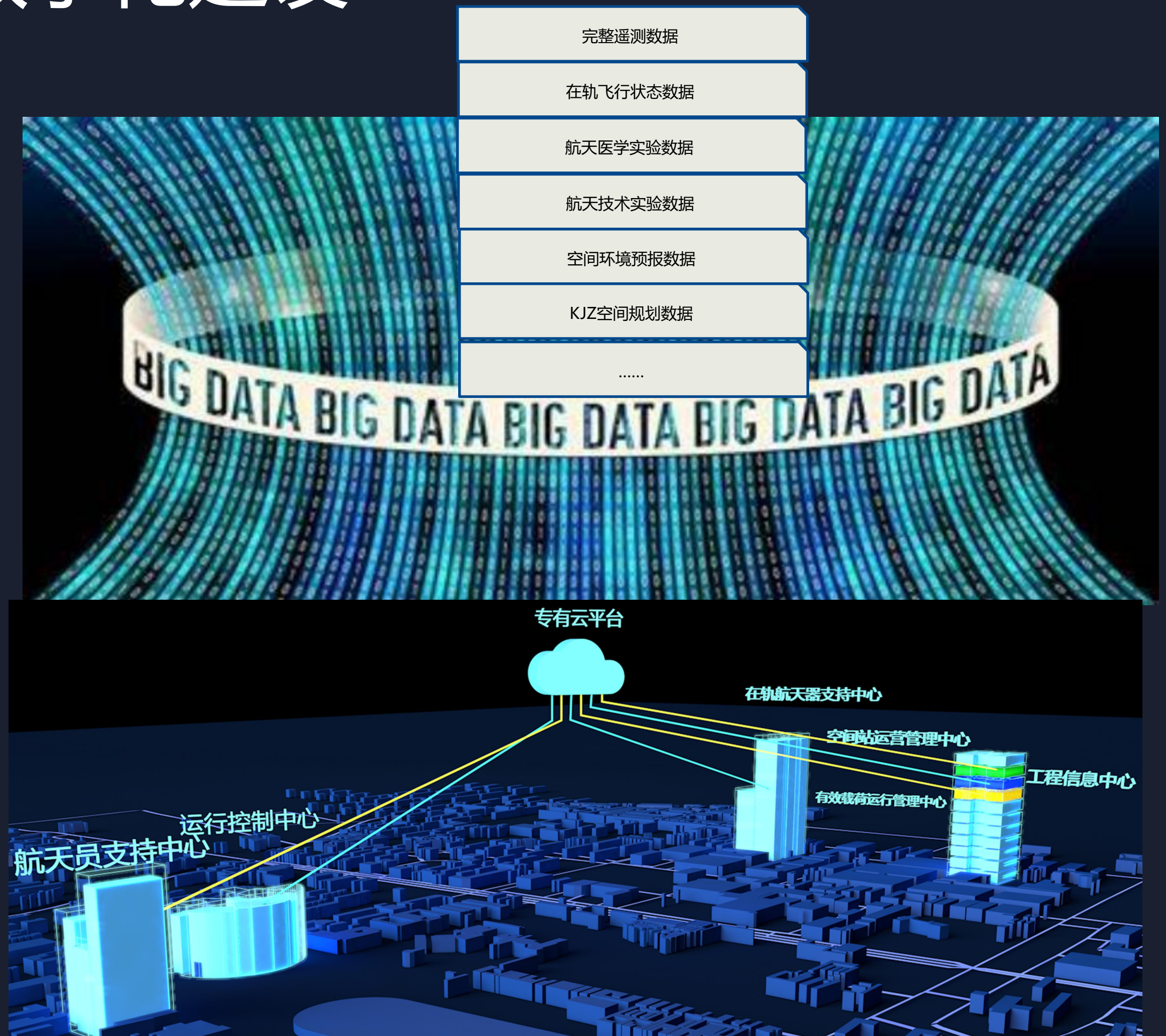
- 边缘计算终端
 - 协议匹配
 - 数据预处理
 - 数据缓存与续传
 - 数据加密与压缩
 - 数据API定制
- 边缘计算服务
 - 终端设备管理
 - 终端数据接入
 - 后端数据库对接
 - 数据服务化管理与支撑

技术架构详解



场景案例：载人航天工程空间站数字化建设

- 空间站运营期间，工程运营数据将包括：
 - 10多种任务类型
 - 5大支持中心在不同任务阶段产生的数百种数据
 - 异构数据源(五大中心使用的数据库类型繁多，包括Oracle、MS、MySQL、达梦数据库、神通数据库等)
- 在工程中心专有云平台，以ArkDB为核心，结合边缘计算与API接入，实现空间站多维数据融合与海量数据的存储与计算
 - 规划容量：300PB
 - 2021.4.29随着天和核心舱发射成功，系统已经正式上线运转
- 效果：实现了空间站数据化建设中，中心数据库从无到有的质变，目前随着核心舱等关键组件的发射成功，线上数据量已逐步积累，正在支撑载人航天事业的发展。





THANKS