

Oracle湖仓一体解决方案

公益讲座11: 00分准时开始, 请大家先浏览云技术微信公众号技术文章资料会在各群同步发布, 已入群客户请勿重复入群!



20-18

数据库和云讲座群



甲骨文云技术公众号

ORACLE

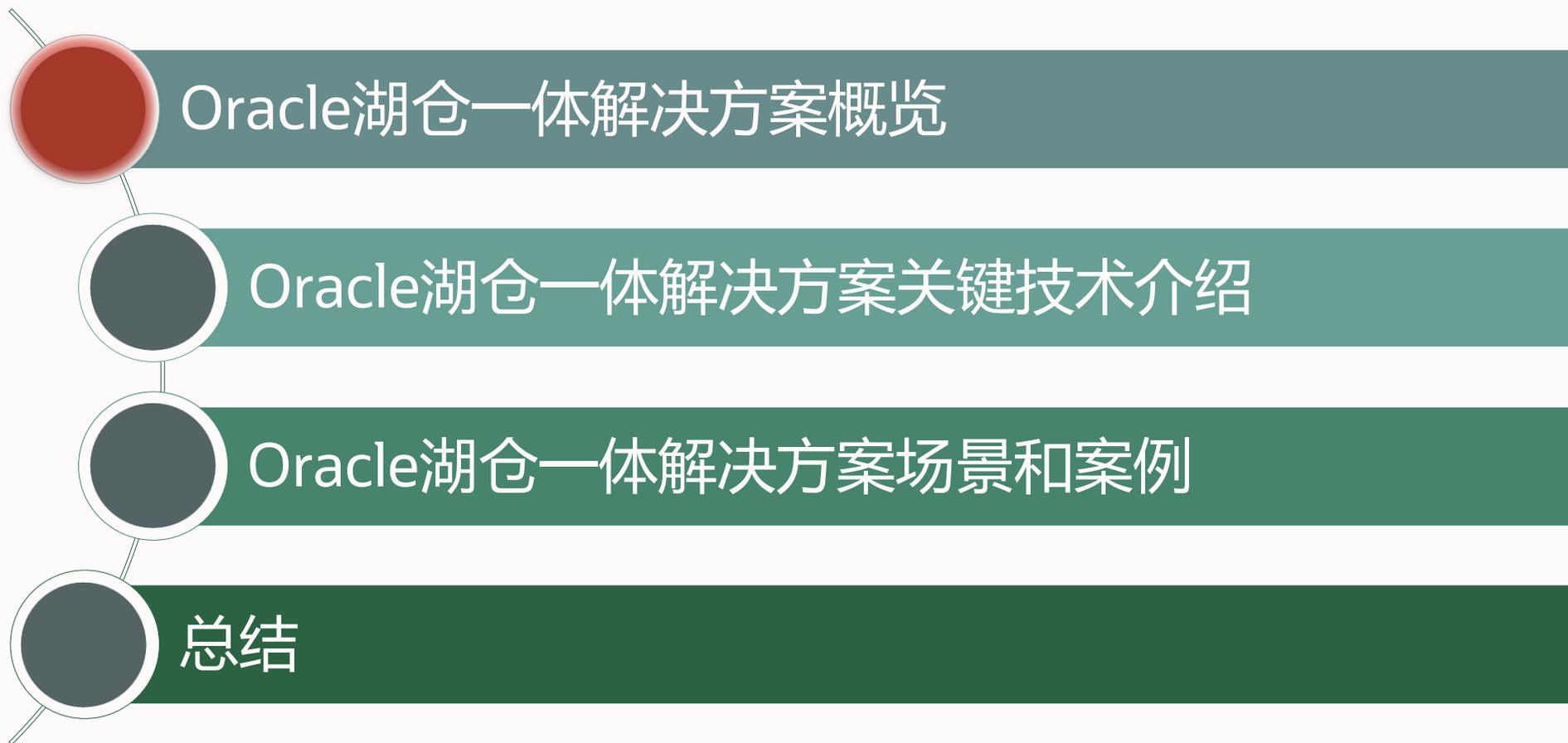
Oracle湖仓一体解决方案

刘群策

Oracle China



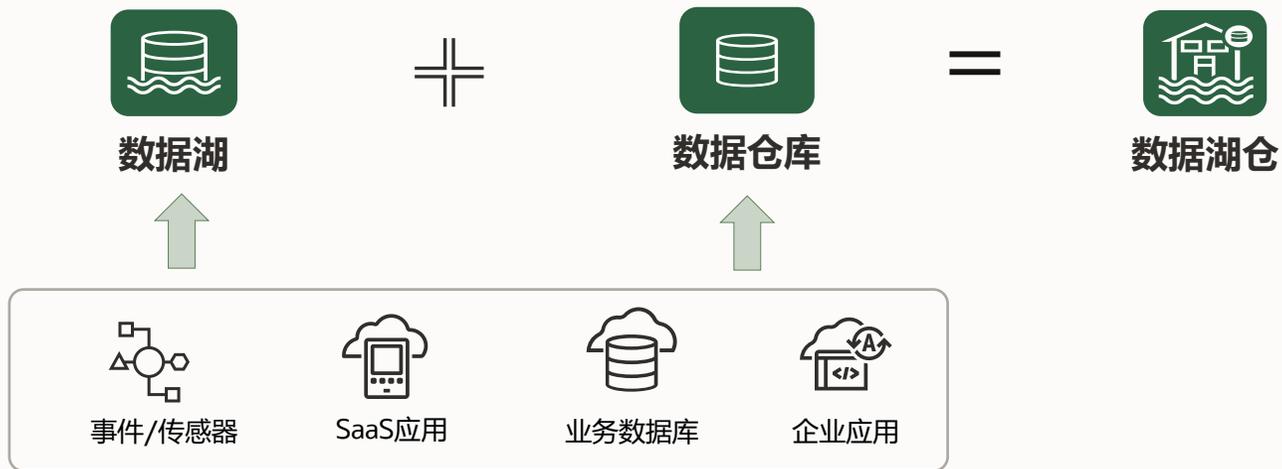
议程



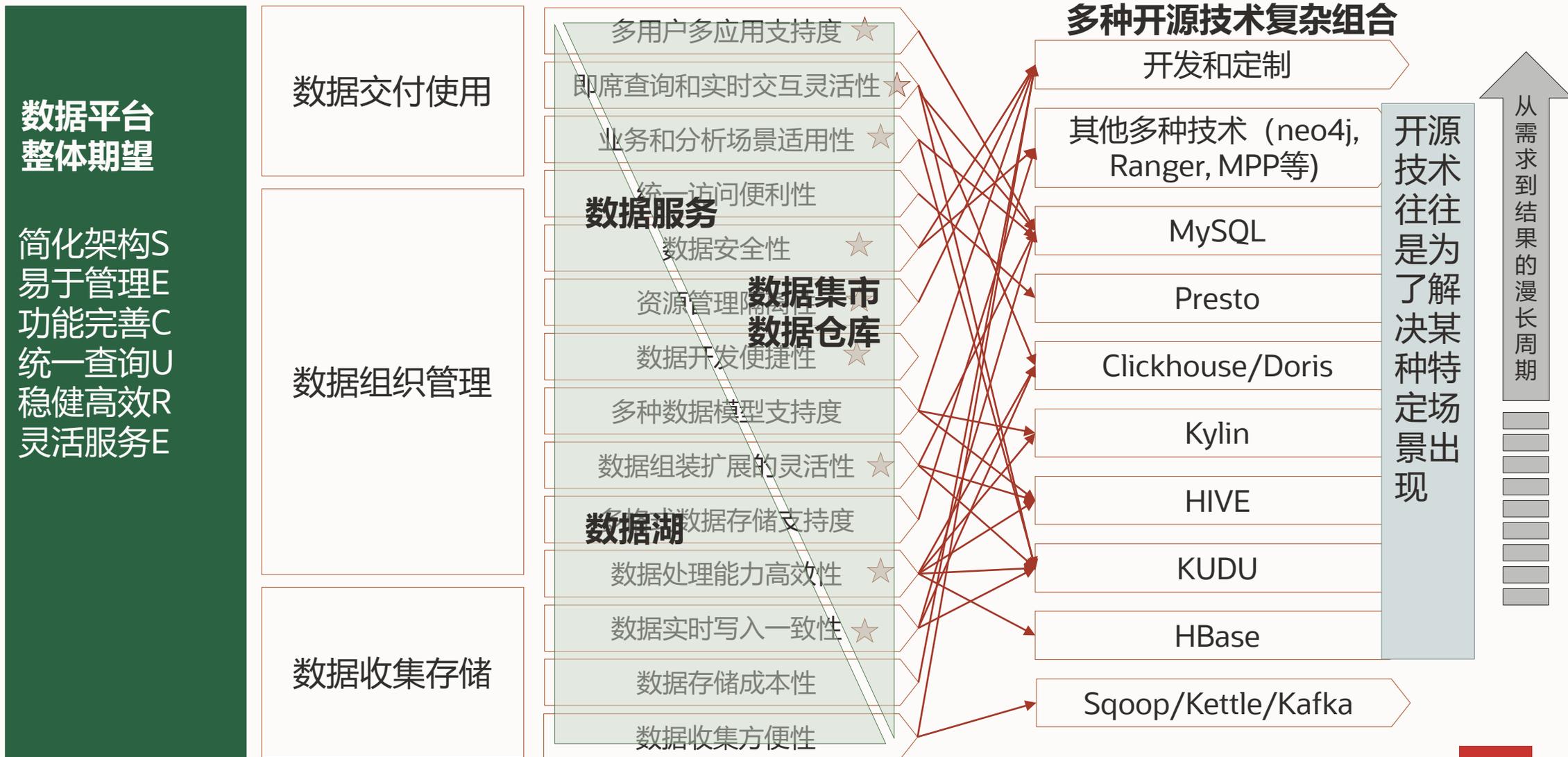
企业数据管理技术的发展

	数据湖	数据仓库
数据	各种格式	结构化, 业务处理
数据处理	一般不做处理, 用时再处理	高度处理后的精炼数据
任务	分享数据, 场景应用	为报告报表和查询数据优化
模式	数据使用时再按需配置和重配模型	按照事先定义的固定的配置或模型写入
用户	数据科学家, 分析人员	业务人员, 管理者

现有湖仓技术遇到的挑战
技术和业务的难于协同, 从业务需求转化为技术理解, 到结果交付和应用效果追踪困难
为适应各种数据服务要求, 需要多种技术结合, 或者数据多次复制, 架构复杂, 不一致
数据模型很难快速扩展响应业务频繁变化
从湖到仓要经过多步骤复杂ETL, 周期长

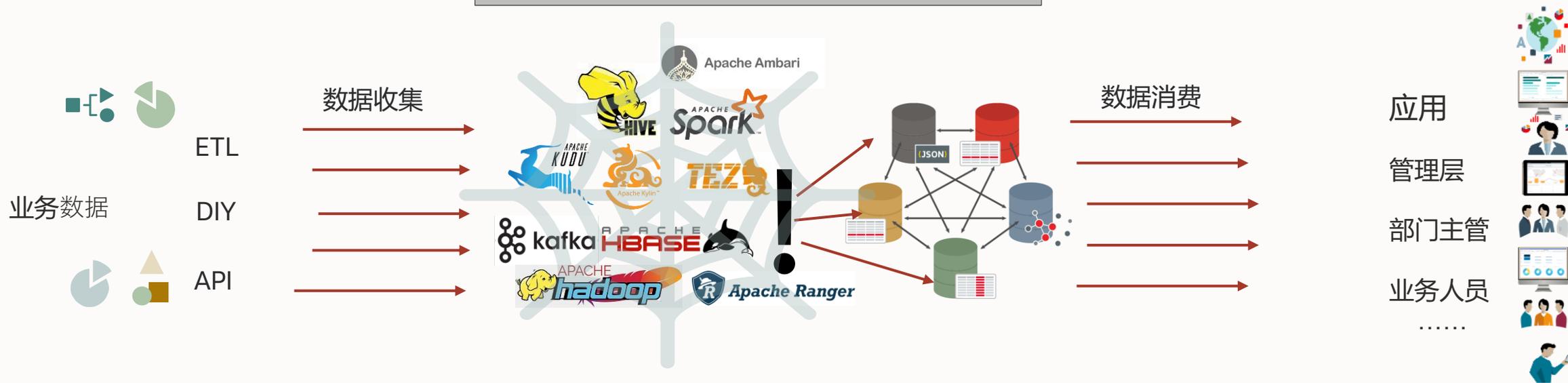


从业务对数据使用的期望来看数据平台能力的需求



企业级数据湖面临的挑战（基于开源和Hadoop）

以管理界面为主，技术架构纷繁复杂



多种硬件和组件复杂，技术挑战

数据多次处理和流转，时效性挑战

多种数据库应对使用场景，管控挑战

严重依赖第三方/人力集成和升级维护

软件众多复杂以及发展路线面临风险



Oracle解决方案提升企业级数据平台敏捷性

数据平台 整体期望

简化架构S
易于管理E
功能完善C
统一查询U
稳健高效R
灵活服务E

数据交付使用

多用户多应用支持度

即席查询和实时交互灵活性

业务和分析场景适用性

统一访问便利性

数据安全性

资源管理隔离性

数据开发便捷性

多种数据模型支持度

数据组装扩展的灵活性

多格式数据存储支持度

数据处理能力高效性

数据实时写入一致性

数据存储成本性(分层)

数据收集方便性

数据组织管理

数据收集存储

Oracle数据平台

敏捷可视化和智能分析

Oracle Big Data SQL

Oracle

Hadoop

多租户

多负载

多模型

多格式

APACHE
Spark

HIVE

APACHE
hadoop

kafka

ODI/OGG/OSA

Oracle Big data

Connectors

已有知识
经验重用

成熟的
企业级特性

可持续化
发展能力

自治
弹性



Oracle企业级融合数据平台（以SQL为主的湖仓一体）

兼顾快速启动和扩展，端到端提升数据应用敏捷性

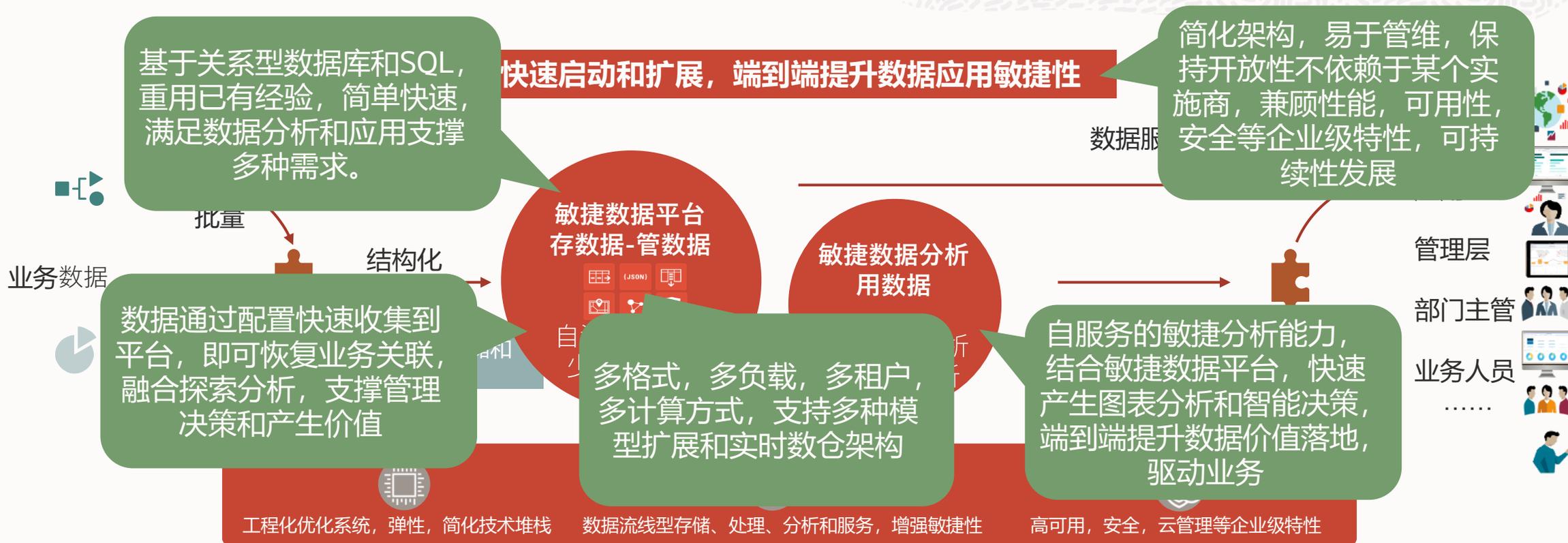


80%技术层面问题由平台智能解决

支撑业务人员探索数据的业务价值



Oracle企业级融合数据平台的价值 (以SQL为主的湖仓一体)

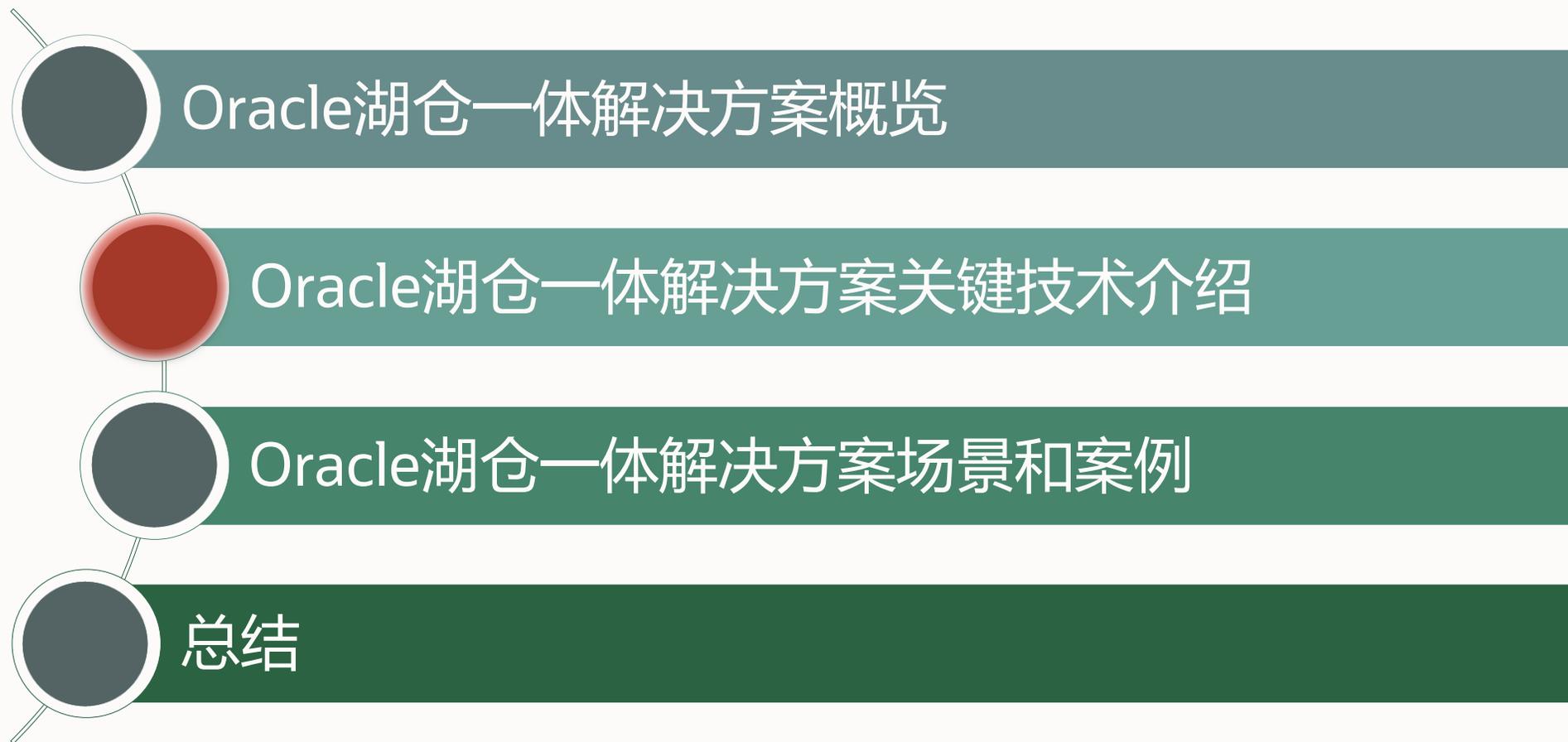


80%技术层面问题由平台智能解决

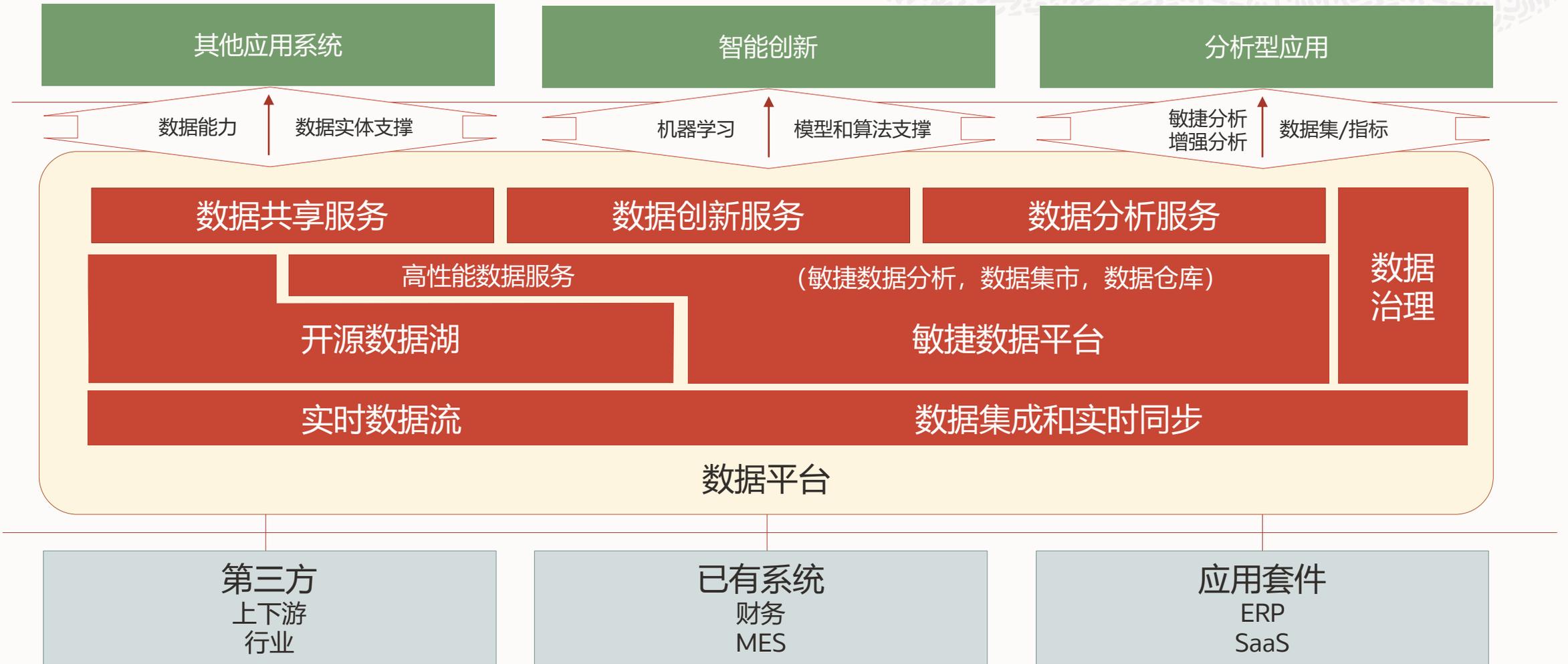
支撑业务人员探索数据的业务价值



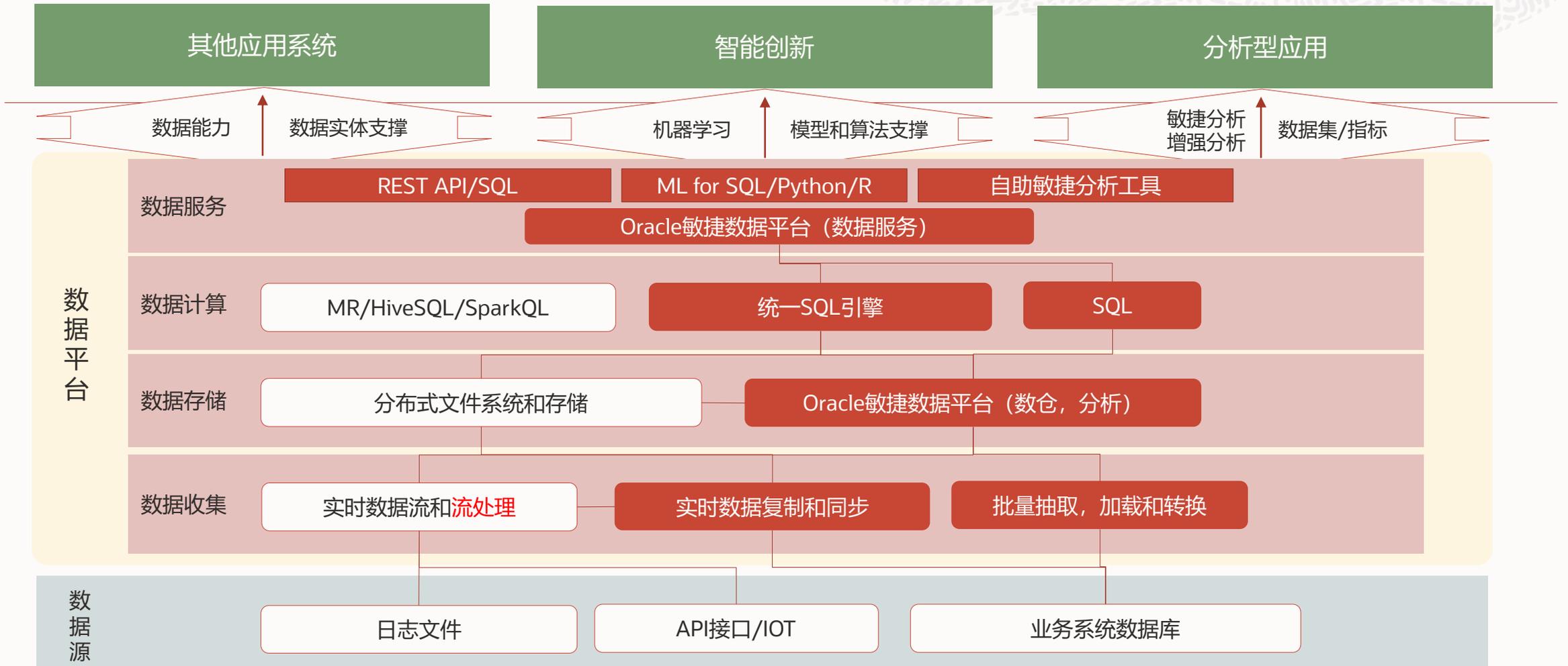
议程



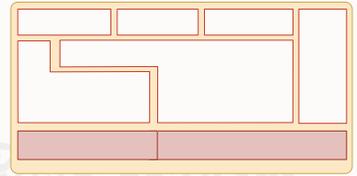
Oracle企业级数据平台整体架构 (SQL湖仓一体)



Oracle企业级数据平台整体架构—数据流转



Oracle数据平台：统一数据收集和同步



穿戴设备
APP
机器数据
行为数据



传感器
设备参数

已有
报告数据

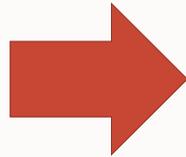


非结构化数
据

日志, JSON
生产记录



业务系统



直接从文件或
云存储加载

数据界面
填报补录

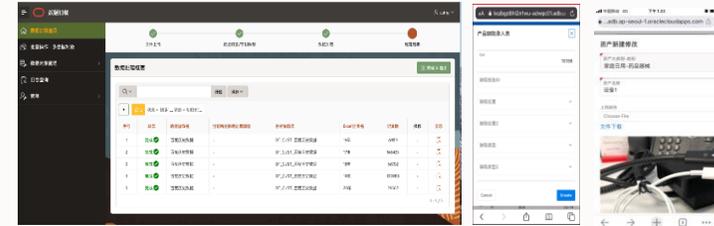
批量数据抽取,
加载和转换

实时数据复制
和同步

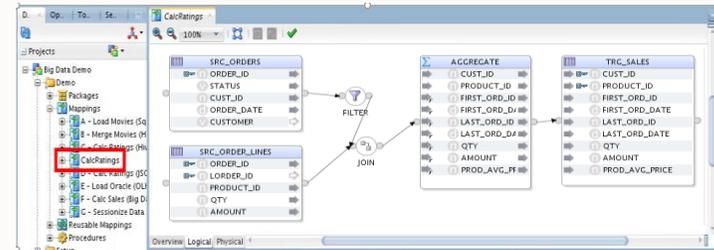
Web界面, 直接从本地或
云存储, 源数据库链接/
加载数据文件或表数据



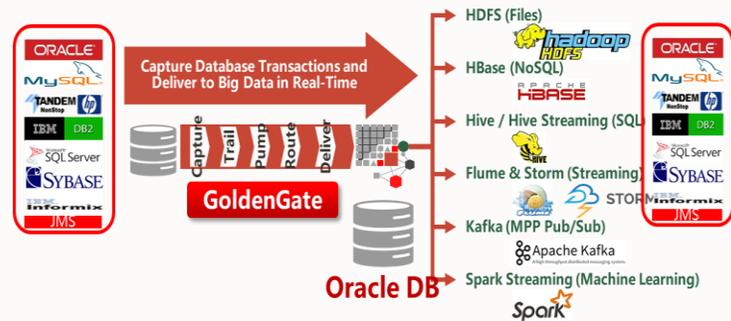
APEX: 低代码开发, 懂
DB+SQL即可快速构建
应用, 与Oracle DB一起
运行



ODI: ELT架构, 多数据源
和目标支持, 避免单点故
障, 一次映射多环境自动
脚本转换, 图形化

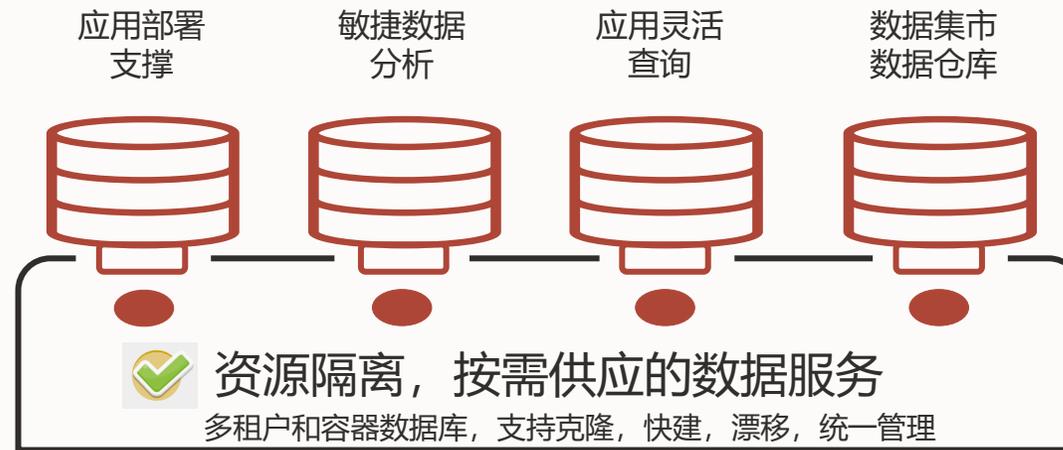


OGG: 无侵入实时数据复
制, 对源系统影响小, 支
持异构, 一对多, 多对一
等多种收集方式



Oracle数据平台提供数据源动力，满足企业多种数据需求

支撑企业级数据全生命周期管理和分析洞察的独特能力



混合负载支持

同时支持OLTP和OLAP, 优化内存计算和性能



内置ML算法

ML算法靠近数据, SQL加速数据相关性探查



快速数据加载

支持高性能数据实时写入, 快速规整



安全数据管控

灵活的细粒度安全控制, 自动加密, 动态脱敏



高效数据计算

支持多用户即席查询, 灵活条件交互, 快如闪电



多模数据支持

结构化, JSON, 地理信息统一管理, 简化开发



融合数据访问

跨Oracle, Hadoop和Kafka一个语句SQL查询



优化数据存储

混合列压缩, 热温冷数据自动分层存储, 降低成本

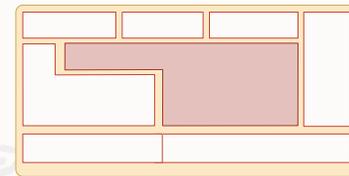


自治数据管理

内置AI运维能力, 简化工作量, 降低风险



Oracle数据平台高性能：混合负载支持，高速数据加载和灵活查询



现状和挑战：

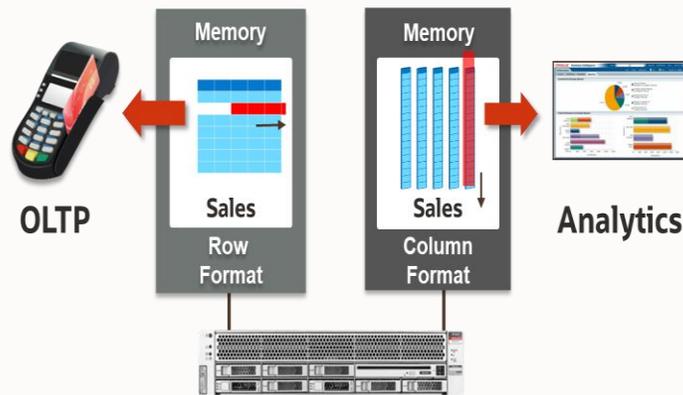
- 后端平台数据层层预处理，维度和过滤查询条件限制
- 分开多集群资源和安全隔离
- 通过开发、分流、数据冗余提升数据集时效，
- 数据冗余带来的数据核对工作量
- 模型扩展困难
- 业务使用依赖IT层层优化

Oracle数据平台内预置的性能优化能力：

- 透明内存计算提升OLTP/OLAP能力
- 支持高并发用户访问和查询
- 支持大量表联合即席查询和条件组合
- 支持自动性能优化，无需人工过多干预
- 支持资源弹性扩展，保证客户体验度
- 缩短从数据落地到供应的周期
- 直接参与OLTP生产，无需再进行数据推送同步等操作

Oracle卓越性能和弹性扩展能力

处理结果直接通过服务方式分享

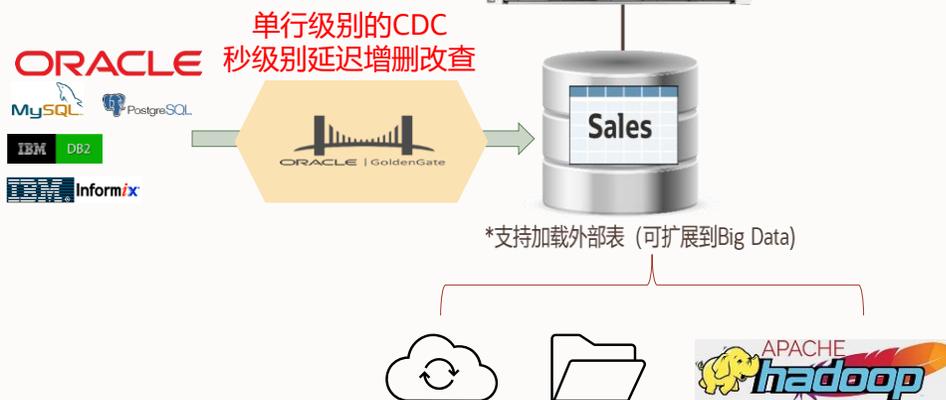


某高科技制造企业

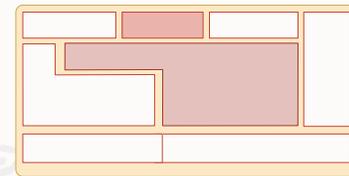
30X性能
>100GB
8~15 X 压缩

某零售企业

30X性能
400GB
~26X 压缩



Oracle数据平台敏捷应用：多模数据，敏捷数据分析和机器学习



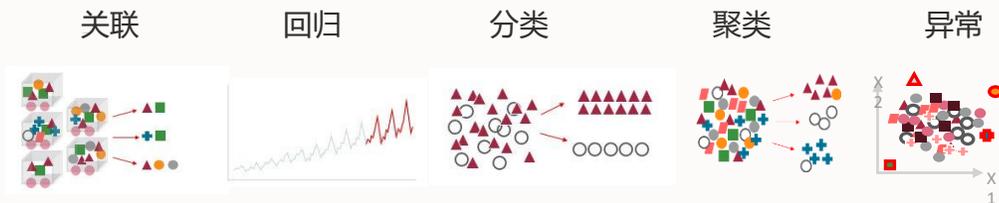
现状和挑战：

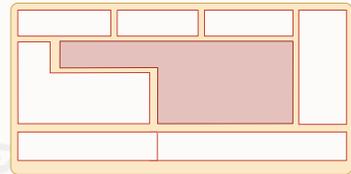
- 不同类型的数据采用不同技术，开发和管理复杂，层层处理，
- 专有工具和数据分开，频繁移动数据，周期长
- 往往单纯提供数据给业务，业务再使用Excel等工具自行处理，很难重用和分享

Oracle数据库提供的能力：

- 支持多种数据格式（结构化，JSON，地理数据，实体关系等）组织和分析，写入即可分析
JSON示例：`json_field: {"name": "小二", "age": 40, "gender": "男".....}`
`SELECT JSON_VALUE(json_field, '$.name') FROM json_tbl_customer ORDER BY JSON_VALUE(json_field, '$.age')`
- 提供靠近数据的机器学习算法，SQL/R/Python调用方式
- 支持IT与业务协同的敏捷数据分析，基于平台上分析，使用数据，提升平台整体价值

Oracle支持多种格式数据，内置机器学习和相关性洞察，支持IT与业务协同的敏捷数据分析





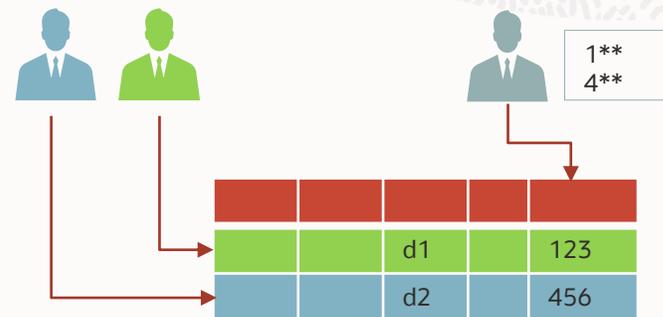
Oracle数据平台安全合规：数据层面的纵深数据防御

安全数据管控，融合数据访问

Oracle敏捷数据平台提供纵深防御和安全能力

现状和挑战：

- 需要在应用层使用大量代码开发实现细粒度的管控，工作量大，重用性差
- 有可能被恶意绕过，出现安全风险
- 需要多次复制数据集，预先配置，灵活性差



Oracle数据库安全能力：

- 数据加载阶段，自动加密保存，防止数据恶意泄露。
- 数据使用阶段，不同业务层级的角色看到不同数据
- 数据输出阶段，对某些字段脱敏，并且保持数据之间关系
- 数据处理阶段，使用不同数据技术能够实现统一的细粒度的安全管控

数据屏蔽



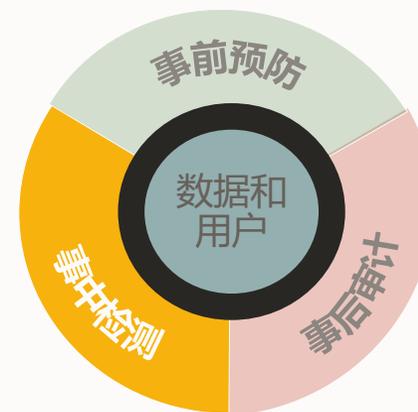
数据加密



访问审计



权限控制



Oracle数据平台数据服务：资源隔离和云化管理

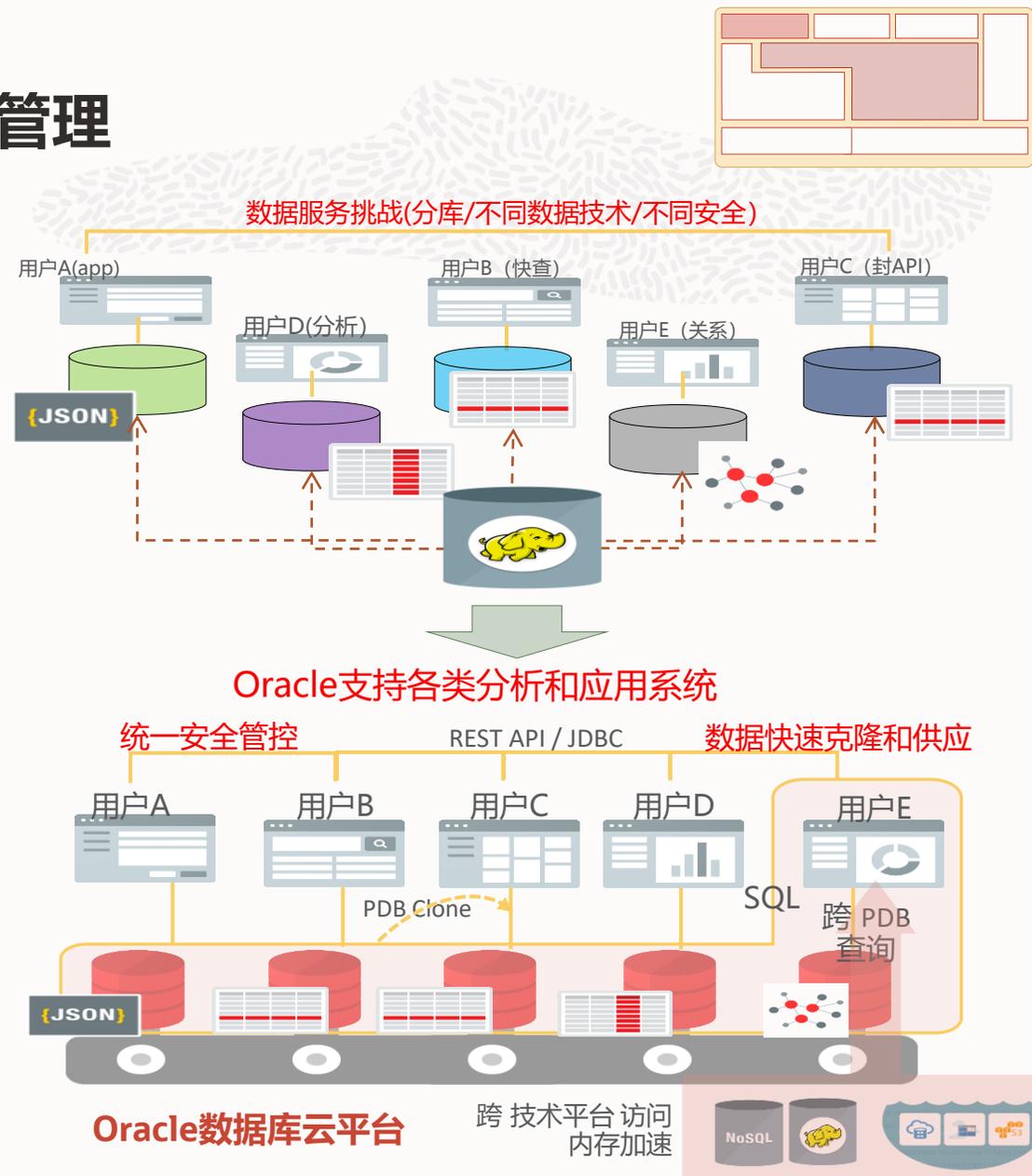
单一平台管理各种数据库实例，简化架构和管理

现状和挑战：

- 为了适应不同数据访问需求，使用各种技术或者专有数据库
- 数据多次复制，冗余，时间和存储成本
- 安全管控面临挑战

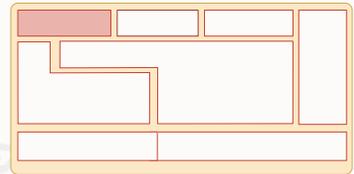
利用Oracle数据库多租户和云能力：

- 构建企业级数据库云平台，实现数据资源池，数据服务弹性供应，资源隔离
- 整个数据库统一管理，升级和补丁，统一备份
- 子数据库快速克隆创建，可独立控制权限和资源，独立恢复
- 基于BDS实现跨技术平台数据访问，可按需实现其他源数据映射
- 企业级特性：性能，安全，可用性，兼容性等



Oracle数据库低代码开发：支持丰富的数据应用

低代码开发，无需编程，懂SQL即可实现Web界面

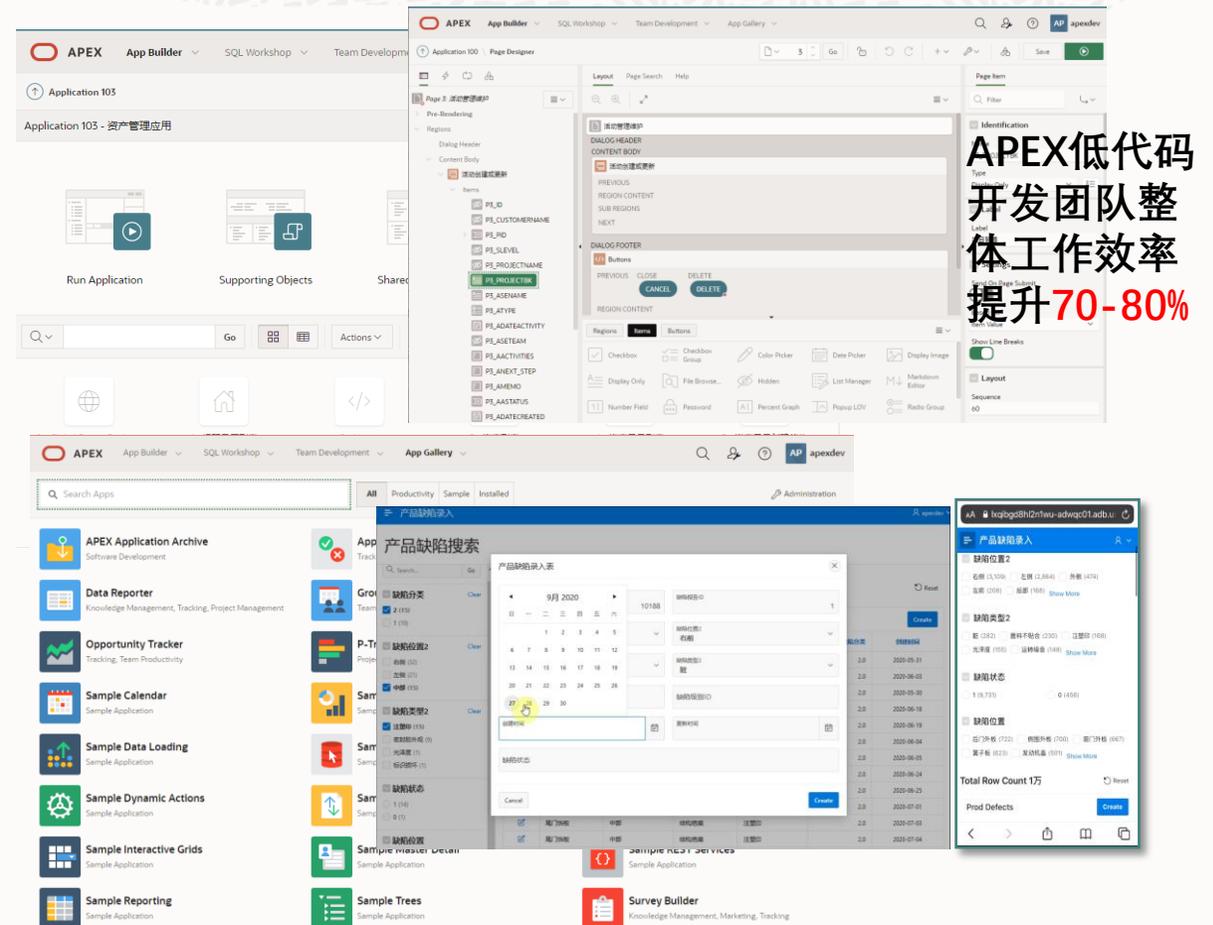


现状和挑战：

- 业务部门不断提出数据管理和应用需求，需要提供输入和展示界面响应
- 数据应用开发严重依赖高级技术人员或开发商，技术门槛高，周期长，成本高
- 应用和数据分离，多种工具和处理时间延后

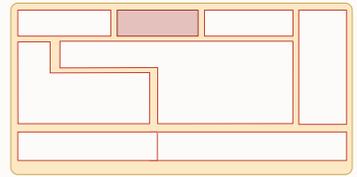
Oracle数据平台提供低代码应用开发框架和运行环境：

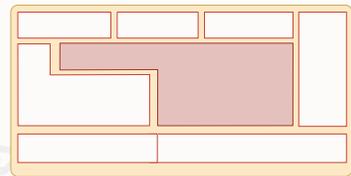
- 支持Web和移动开发
- 基于Web在线声明式开发，无须IDE
- 全栈技术支持(UI-数据访问-可视化-安全-多语言等)
- 与数据平台天然集成，管理和可视化数据
- 有效利用SQL技能和数据库功能，支持REST，易上手
- 数据录入/电子表格/数据共享多种应用场景
- 提供多种应用模板，一键创建即可使用



Oracle数据平台易用性：提供丰富的自服务管理开发工具

满足不同角色数据处理和管理需求





Oracle数据平台融合性：统一数据访问和安全管控

跨不同平台支持单一SQL查询，应用透明，丰富SQL支持

现状和挑战：

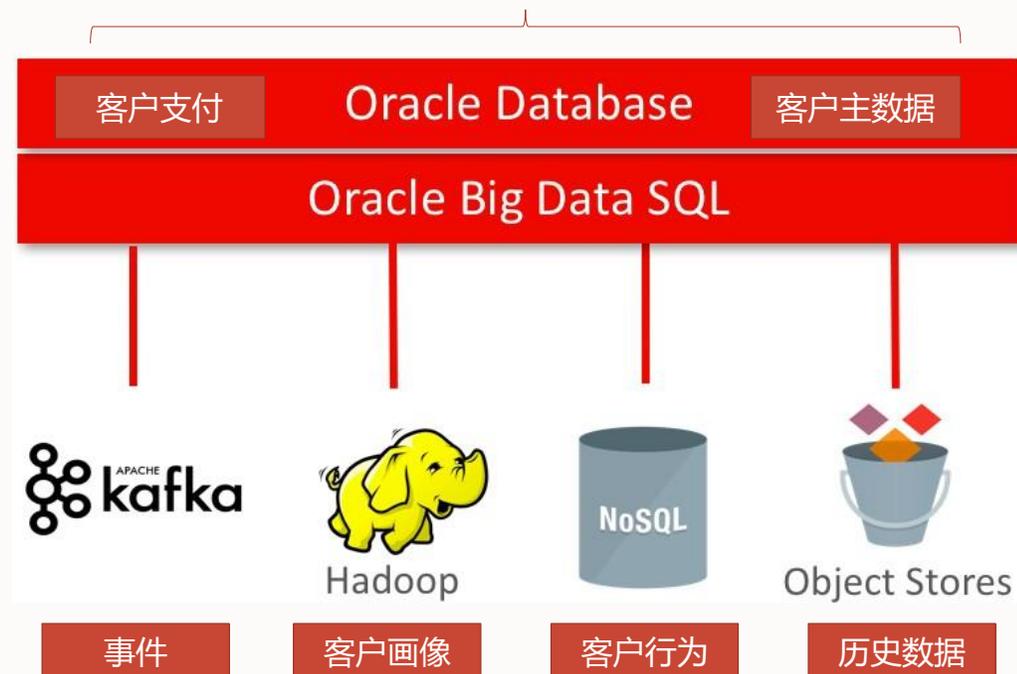
- 已有多个数据平台收集各种数据
- 为满足不同应用场景数据多次复制和加工处理
- 缺乏有效的安全管控和服务隔离
- 新技术增加了学习成本

Oracle数据平台支持使用Oracle SQL来统一访问数据，跨不同平台（Oracle DB/Hadoop/NoSQL/Kafka）：

- **保护现有投资**：不需要任何修改就可以直接运行Oracle SQL查询
- **卓越性能**：Hadoop和NoSQL上的智能扫描通过解析和智能地过滤各自节点数据大幅度提高性能
- **安全管控**：Oracle数据库安全特性扩展到新技术平台，控制对敏感数据的访问

通过Oracle数据库提供统一的数据访问服务

示例：客户行为实时查询和预测



Oracle敏捷数据分析：端到端加速数据价值实现

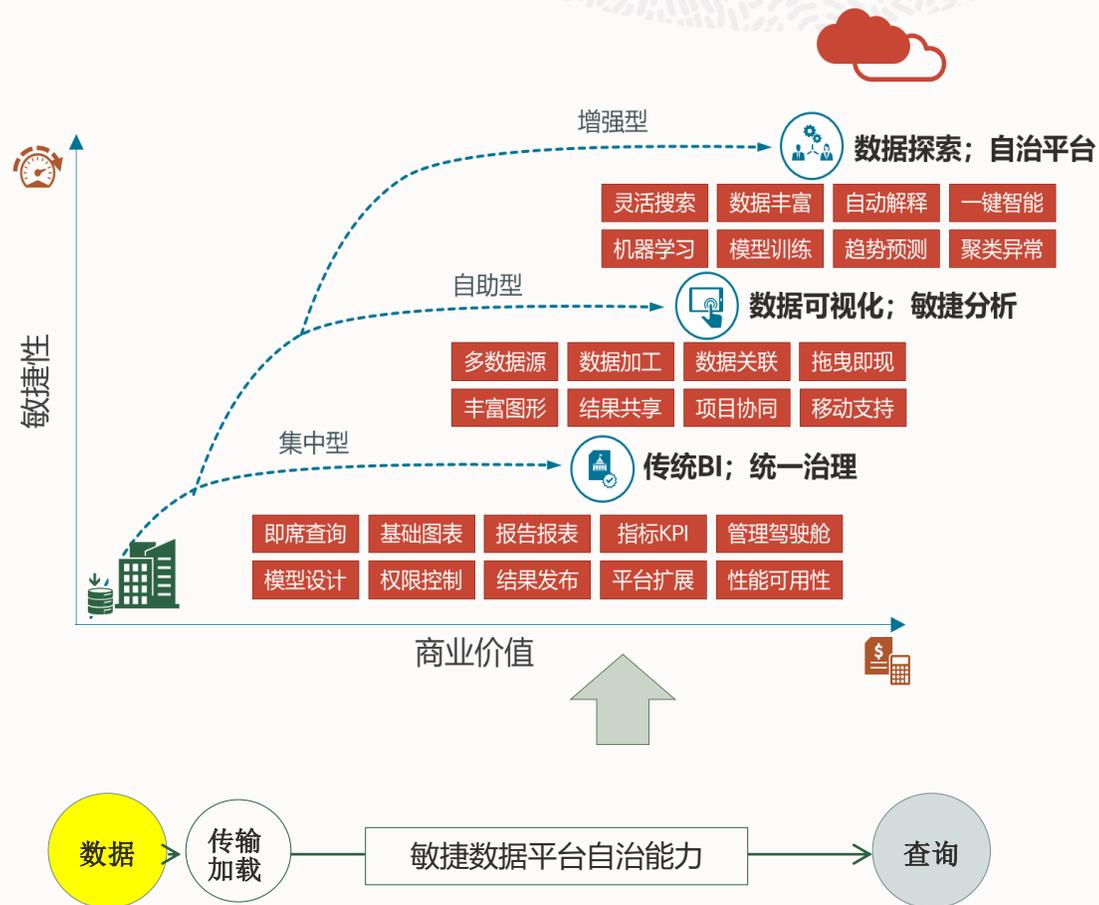
单一分析工具实现固定格式报表，敏捷可视化和机器学习，结合平台，加速数据探索和模型迭代

现状和挑战：

- 平台部门应对业务需求，按需提供数据，或开发界面，但对业务影响后期未知
- 不同需求使用不同工具，固定报表，自助可视化，机器学习，数据多份复制，
- 数据离开平台，很难统一安全管控
- 新的数据准备，模型扩展困难或周期长

利用Oracle敏捷数据分析能力：

- 简化IT交付数据工作，加速从数据落地，规整，建模，模型扩展或安全控制的过程
- 分析工具集成固定格式化报表，敏捷可视化和机器学习，无需改变使用习惯
- 支持从前期数据探索到模型固化的迭代过程，提升业务分析和数据团队的协同能力
- 从业务角度驱动数据分析价值快速变现



Oracle企业级融合数据平台（以SQL为主的湖仓一体）的落地实现

兼顾快速启动和扩展，端到端提升数据应用敏捷性



80%技术层面问题由平台智能解决

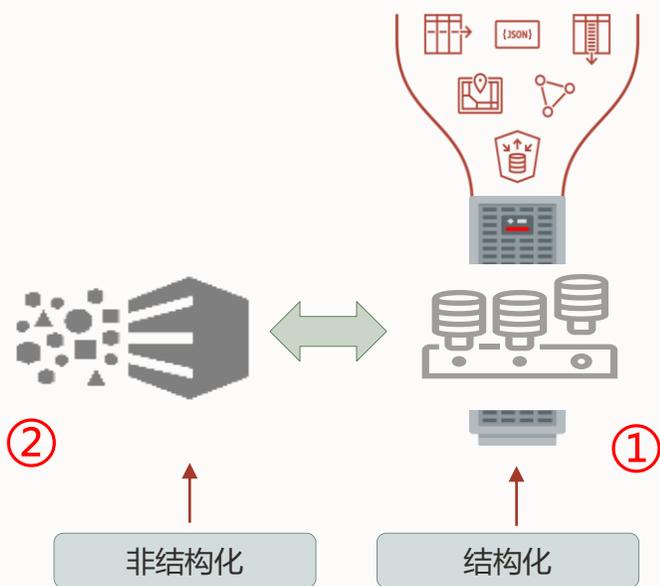
支撑业务人员探索数据的业务价值



Oracle数据平台满足企业多种架构快速启动需求

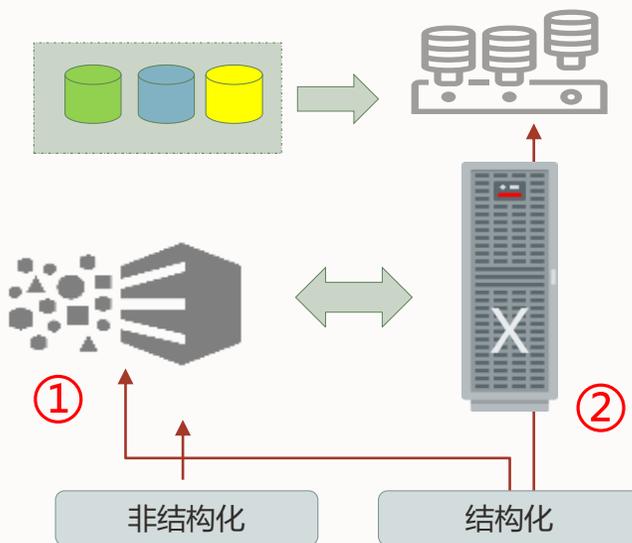
快速启动平台建设，整合和分析

- 结构化数据为主，支撑应用，备份，多模，多用途
- 未来大数据平台建设，集成和扩展



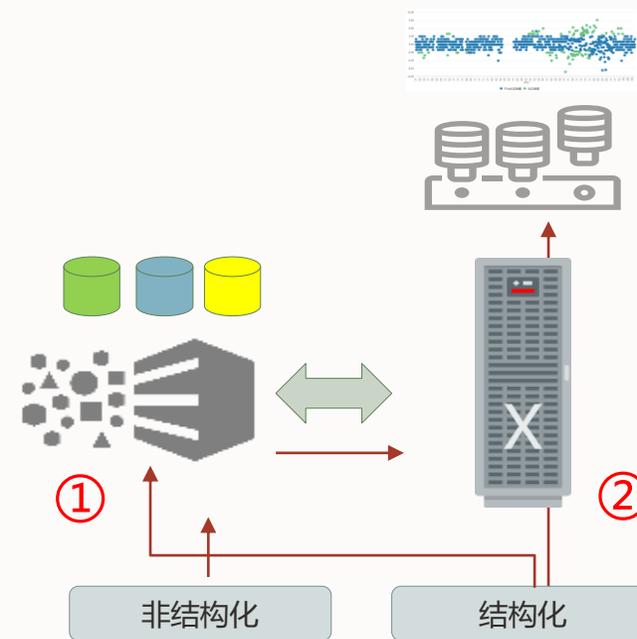
已有Hadoop，增强敏捷性

- 提升结构化数据服务和利用敏捷性
- 混合负载，灵活组合查询，数据智能，快速供应



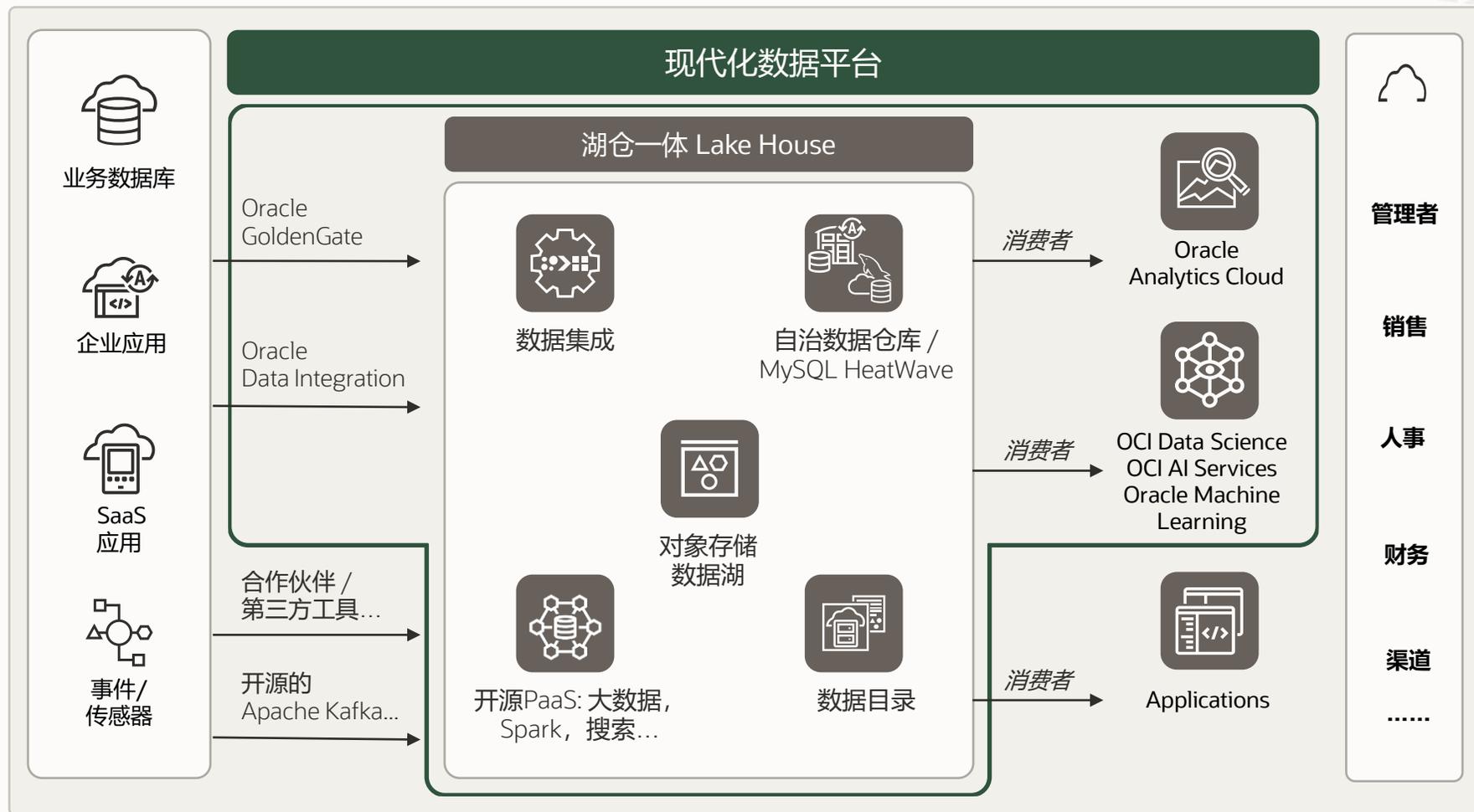
已有架构，敏捷业务分析

- 聚焦数据服务和敏捷智能分析
- 业务人员自服务使用，轻建模
- 技术快速供应，保证安全性和性能



Oracle云中新一代敏捷数据平台 --- 数据湖仓

整合数据仓库和数据湖，可做所有数据的综合分析，数据讲述业务的价值



将结构化数据与非结构化数据都集成在数据湖仓中统一管理

即可实现数仓高性能的数据存储和查询分析

又可以将数据湖中原始的低价值的快速存放在低价的对象存储中

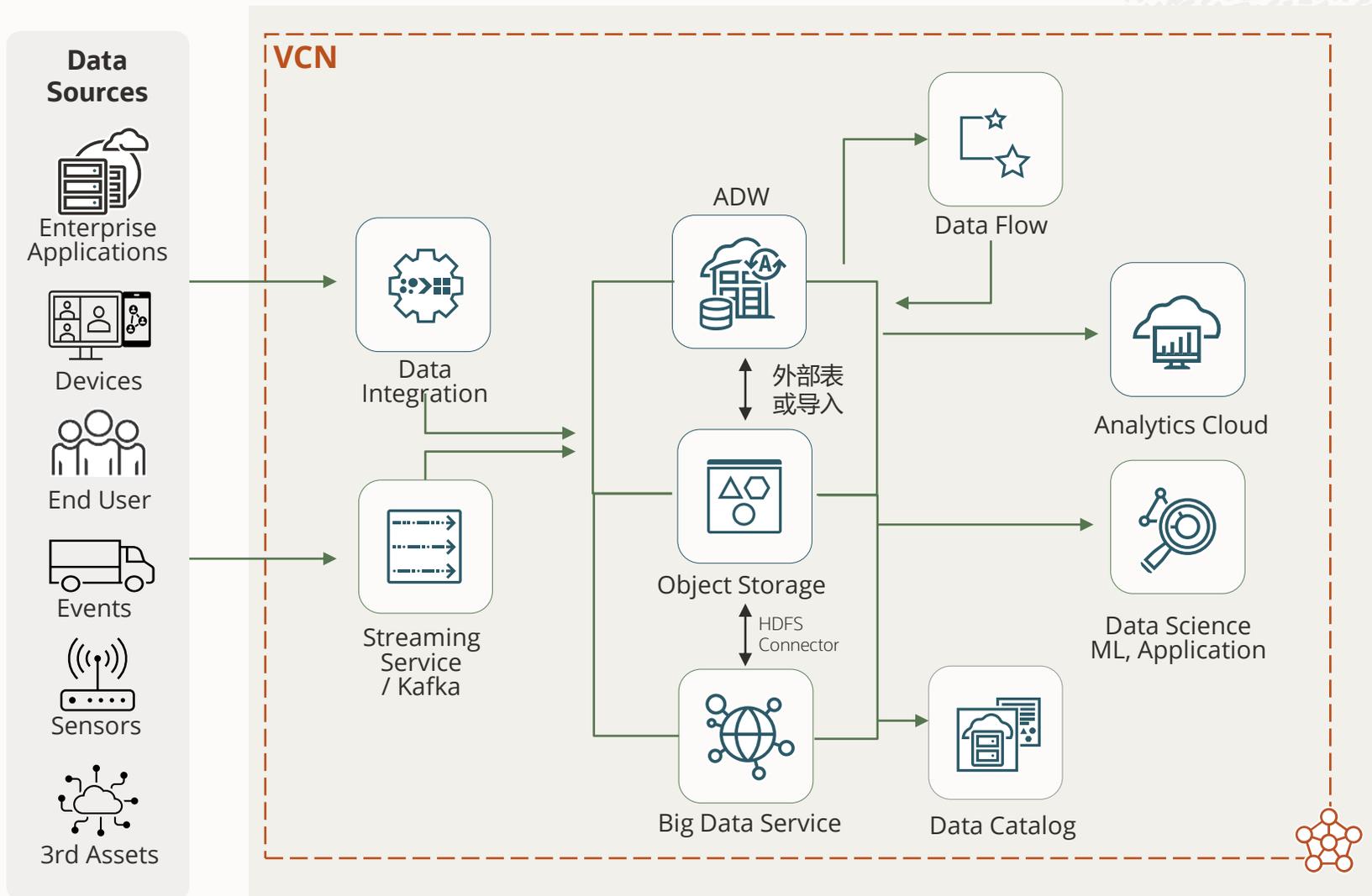
通过OCI数据集成去在数据湖和数仓之间移动整合数据

能实现全部数据的综合查询分析

- Oracle LakeHouse是新一代敏捷数据平台，不同格式数据不同方式收集，使用最佳方式组织，统一提供服务给各种角色，外部可通过SQL使用
- 如果只是简单的湖仓放在一起，原来数据之间的关联关系就没有了，全打乱了，再整理出头绪就会很难，这样成本很高，也会更复杂，不易展示出数据业务价值。



Oracle云中湖仓一体技术架构



Oracle ADW提供了简单，快速，弹性的企业级数据仓库/数据集市能力，自动优化数据管理和处理交付，易于管理和使用。可直接访问对象存储中各种类型数据查询或导入。

Data Flow提供了Spark离线计算服务，按资源使用计费，支持海量数据扫描计算和负责ETL，节省成本。

Data Integration支持按需集成多种数据源。**Streaming Service**支持实时数据流接收和订阅，与Kafka兼容，接收和发布实时数据流，例如事件，日志提取等，也可自建Kafka集群

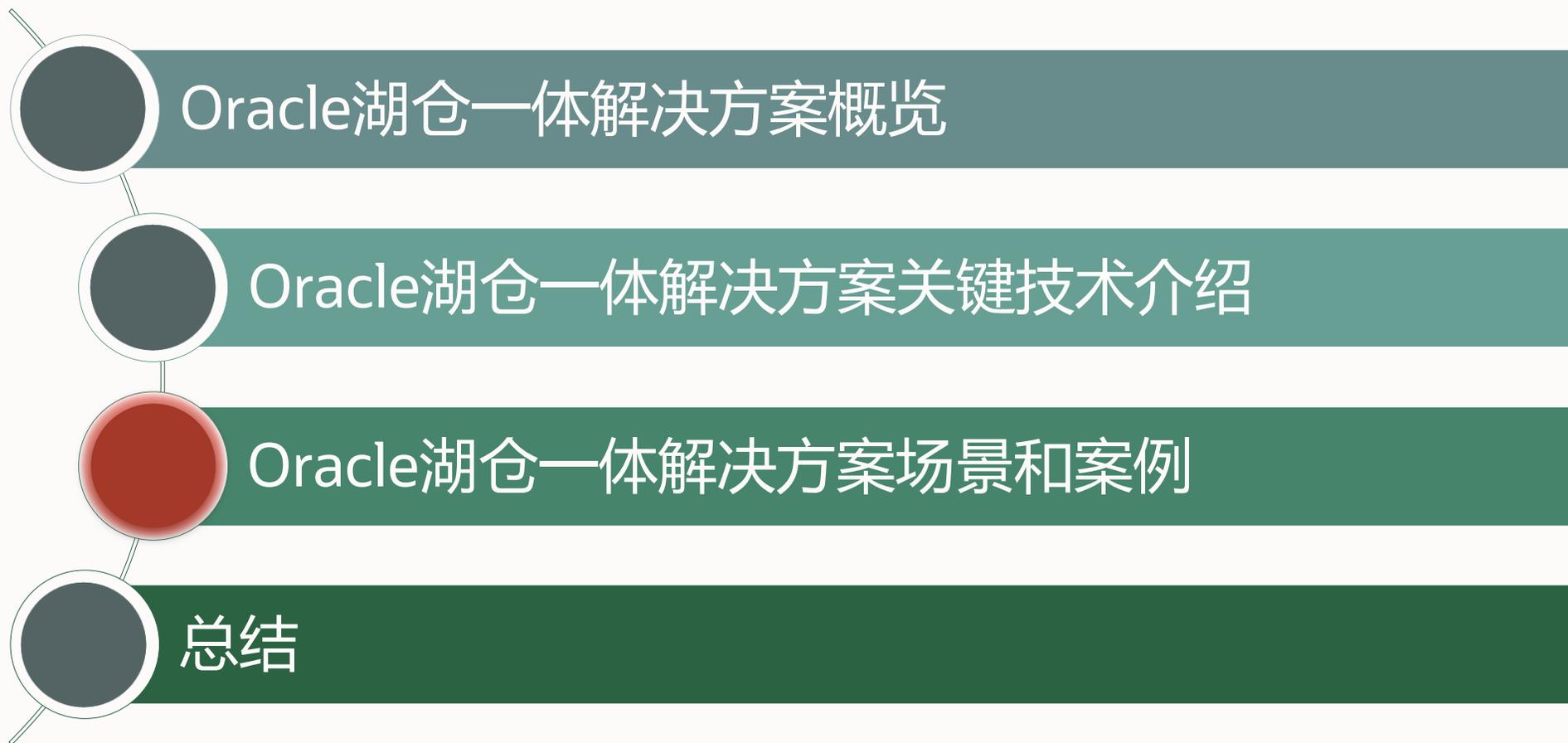
Big Data Service提供安全可靠，灵活扩展的Hadoop集群及相关组件，支持数据快速写入，简单查询，复杂计算。通过**HDFS Connector**，支持大数据与对象存储之间的集成。

Analytics Cloud提供经典商务智能，敏捷可视化和机器学习能力，便于快速探索和数据洞察，转化数据价值

Data Science提供了交互协同的机器学习环境，支持多种Python开发包和提供AutoML能力各组件、数据目录和数据湖之间的紧密集成，实现了与**对象存储的无缝查询**，OCI提供全面的**数据和隐私保护**，包括**透明数据加密**，**细粒度访问控制**和**敏感数据掩码或脱敏**。



议程



Oracle数据平台案例和价值

领先家电集团

数据服务和 分析平台

构建了产品研发、销售、零售、供应链、质量、财务售后等多方面分析，扩展到大数据协同查询分析，支持互联网+和智能制造。

- ETL处理提升>30倍
- 数据运算提升9-26倍
- 高性能和可用性
- 维护成本节省 80%
- 数据访问和查询时效性提升

高新制造企业

数据分析和 智能平台

基于销售历史数据和客服数据，构建机器学习模型，实现销售预测和模式识别，驱动交叉销售和向上销售，提供个性化服务：

- 转化率提升30%
- 满意度提升30%
- 调度成本降低40%
- 数据更新：实时

某高科技集团

数据服务和 查询平台

生产数据快速落地，按客户需求，从不同维度实现产品生产数据快速组合查询，及时响应客户质量追溯和管控需求

- 数据加载提升>15倍
- 数据查询提升5-30倍
- 数据压缩比8-15倍，节省成本
- 实施周期缩短 2/3
- 数据一致性保证溯源高效可靠

某省移动公司

数据共享 服务平台

构建信令大数据敏捷数据平台，与分布式实现数据融合，提升PB级数据服务的敏捷性，灵活条件组合查询，实时交互性和访问用户并行性。

- 数据压缩比>5倍
- 数据存储 > 3PB
- SQL扫描 > 100GB/s
- 收据抽取加速 5-10倍
- 支撑几十个服务商模型构建和应用部署

其他

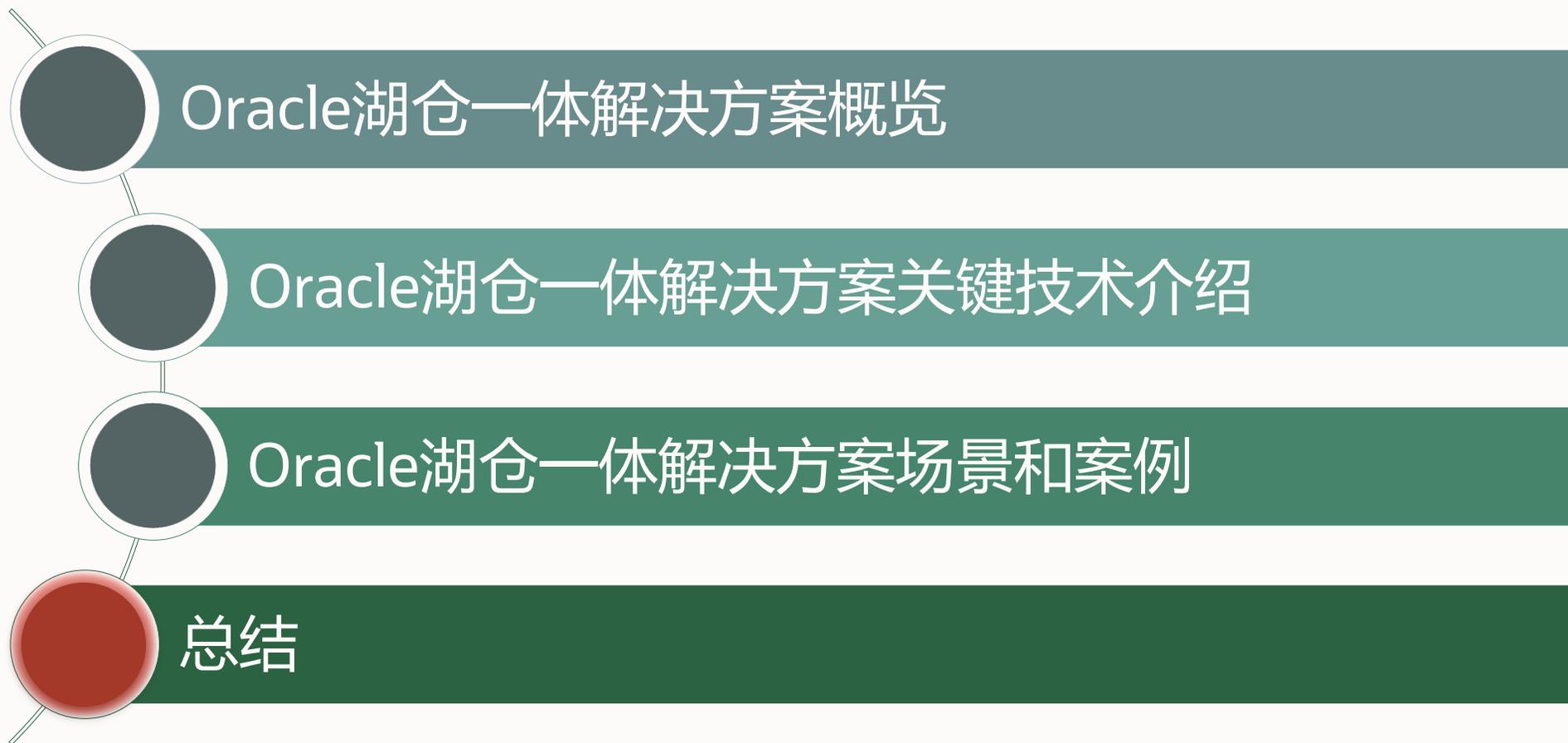
数据分析 和服务平台

应用于各行各业企业，支撑产品生命周期，客户生命周期管理和分析，提供数据分析和应用支撑，整体价值：

- 总运营成本减少47%
- IT基础管理效率提升 69%
- 非计划宕机时间减少 73%
- 上线时间缩短40%



议程



Oracle敏捷数据平台的独特优势

重用已有技术和经验

已有SQL处理和分析，模型，整合处理都可重用，融合数据访问能力无需太多学习新技术

可持续性发展

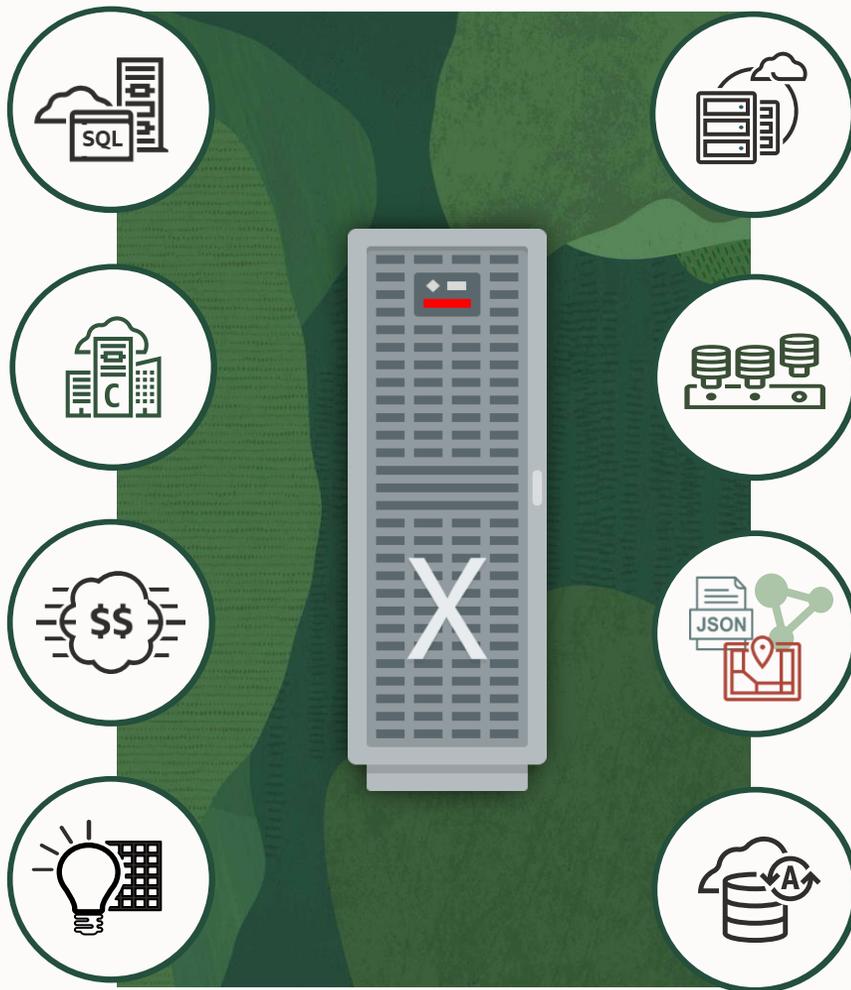
成熟的商业化软件，持续化创新发展，不依赖于开源社区和第三方，没有突然停滞风险和未知安全隐患

综合性价比高

不仅仅是软硬件成本，还需要考虑长期开发和维护支持相关的人力，时间，机会成本等

更好推动业务创新

从复杂底层技术解放人力，协同业务分析人员和科学家，专注于上层数据分析和创新



成熟的企业级特性

充分实践证明的企业级稳定性，高性能，可扩展性，混合负载能力，混合列压缩，开箱即用，预先配置和优化

云数据平台能力

数仓、集市、应用通过多租户可共享软硬件资源，节约资源、数据共享，同时资源隔离又能保证关键任务

多模，简化数据管理

支持空间信息、图数据、区块链关系数据的统一存储，无需关联多种数据库和数据复制，集成机器学习能力，简化管理和开发

统一安全管控

企业级统一数据安全管控能力，细粒度权限控制，责权分离，敏感数据动态加密，保证数据共享无泄露风险



ORACLE

来给数据库做个SPA Oracle SPA and RAT

数据库和云系列(八十)



王健

- 甲骨文资深解决方案工程师
- 专注数据库和数据领域。数据库从业多年，在运维、性能调优等方面有丰富经验

内容简介

- SPA 与 SPA 概述
- 客户体验及反馈
- 最佳实践简介



直播时间: 8月26日 11:00 - 12:00

扫描二维码注册并安装手机Zoom进入直播

Zoom ID: 976 6962 5763 密码: 98039717



数据库和云讲座群

20-18



甲骨文云技术公众号



技术专家1V1深入交流

