

Break New Ground

# 实时洞察，掌控先机

数据集成平台云激活实时洞察

李杰

Oracle数据集成解决方案专家

Integration Solution Specialist Team, APAC

2019年4月

*ORACLE  
CODE*

# Safe Harbor Statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, timing, and pricing of any features or functionality described for Oracle's products may change and remains at the sole discretion of Oracle Corporation.

# 数字化创新是企业发展的新动能

数字技术改变生活方式

消费者更加倾向通过数字渠道购买产品



大数据



数字化带来个性化的用户体验

人工智能

移动互联网

安全依然是企业关注的重点



物联网

云计算

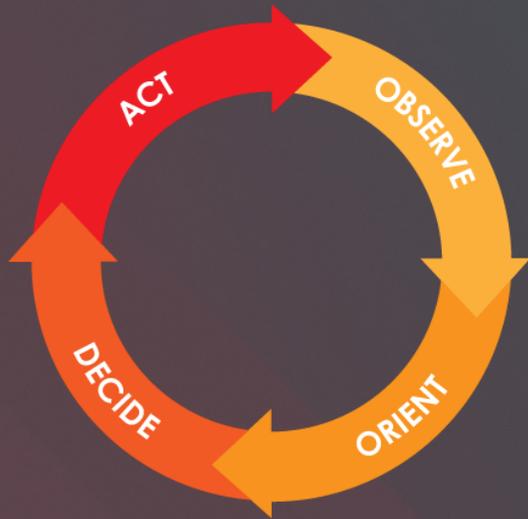


数字化带来“万物互联”

# 数据资产化对数据集成提出了更高的要求

## 更敏捷的信息闭环

更快、更精准基于数据驱动的决策



## 处理来自不同数据源的不同种类的数据

更快、更好地处理不同种类的数据



## 数据安全与合规

更安全、更合规的数据采集、处理和存储

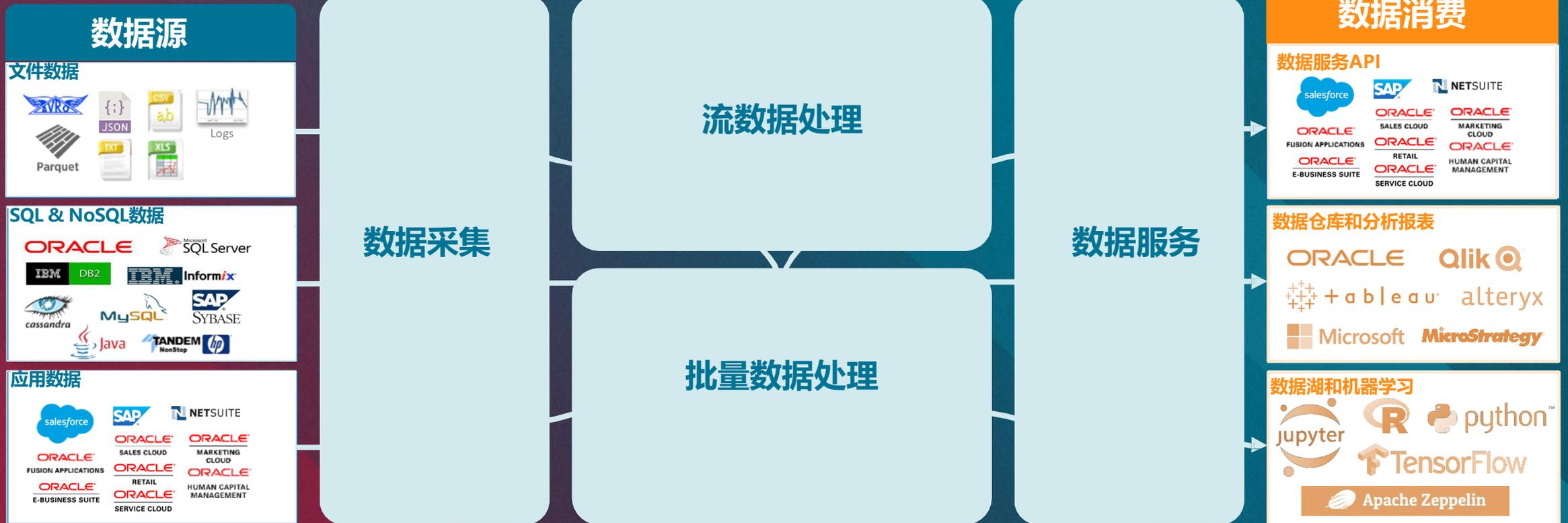


# 新一代数据平台

实时采集任何数据

流数据和批量数据处理融合

多种方式消费数据



# 实时洞察的组成部分

## 实时洞察

实时数据采集



低延迟、高吞吐数据采集

实时数据处理



流数据处理

实时数据查询



自治数据仓库

实时数据同步



实时数据同步



# 实时数据采集

## 实时采集任何数据

### 数据源

#### 文件数据



#### SQL & NoSQL数据



#### 应用数据



### 数据采集



### 数据抽取

## GoldenGate+Kafka的优势

实时

异构

易用

可扩展

安全可靠

# Oracle GoldenGate – 企业级实时数据采集解决方案

NON-RELATIONAL DATA

KERNEL INTEGRATIONS

REMOTE CAPTURE

SIMPLIFICATION

MICROSERVICES

CONTAINERS

MONITORING

STREAM ANALYTICS

CLOUD

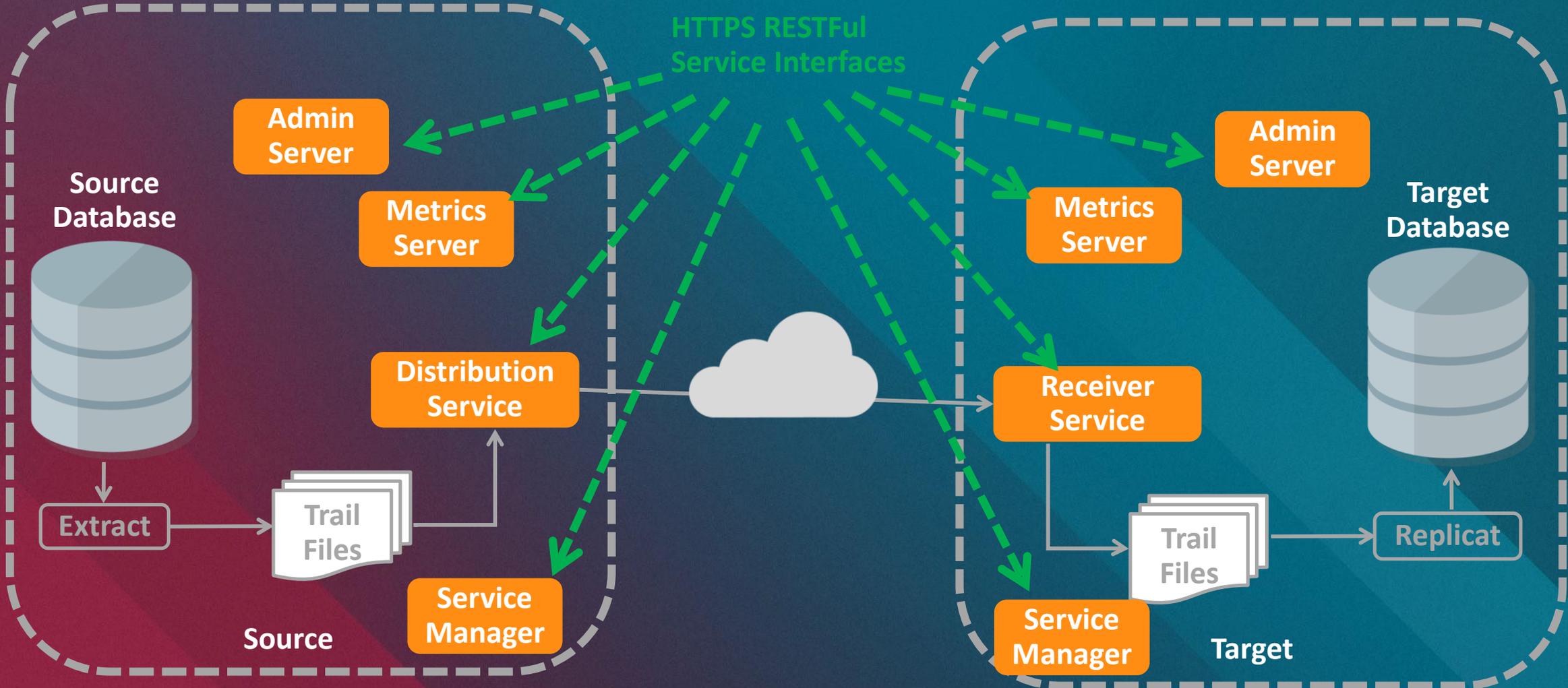
SUBSCRIPTIONS

Oracle  
GoldenGate:  
Innovations for  
Another 20  
Years

SUMMER, 2018

ORACLE

# GoldenGate微服务架构支持通过API管理、配置、监控作业





# 实时数据处理

## 流数据和批量数据处理融合

### 流数据处理

Stream Analytics



Spark Streaming

### 批量数据处理



对象存储  
HDFS

### 流数据处理场景

- 网站实时流量分析
- 实时营销
- 实时反欺诈

### 开源流计算引擎

- Flink
- Spark Streaming
- Storm
- Kafka Streams
- Samza

### Stream Analytics

- 基于开源的Kafka、Spark Streaming
- 端到端的企业级流数据处理解决方案
- 覆盖数据采集、消息发布订阅、流计算引擎、图形化开发工具、可视化能力

# 流数据处理应用场景



## 金融服务

- 实时反欺诈
- 实时风控
- 实时营销



## 交通运输

- 公共交通
- 集装箱追踪
- 车辆管理



## 运营商

- 视频分析
- 网络管理
- 安全操作



## 公共安全&医疗健康

- 平安城市
- 医疗设备监控



## 零售

- 个性化推荐
- 降价优化
- 动态定价
- 货架管理

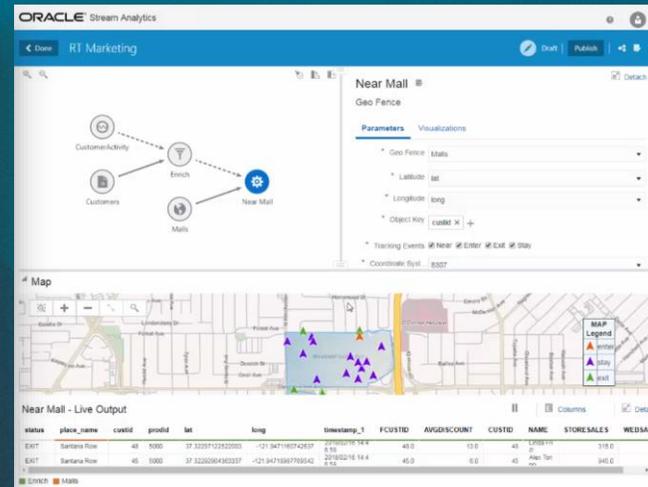
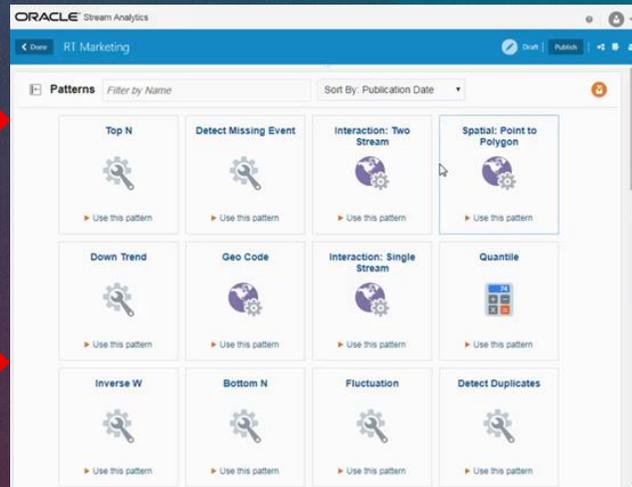


## 制造业

- 智能库存
- 品质管控
- 预测性维护



# Oracle Stream Analytics - 企业级流数据处理解决方案



Events / Messages

Dimensional

Micro-Batch



## 实时数据采集

实时采集数据到Stream Analytics  
直接从Kafka读取数据

## 选择处理模式

丰富的预置模式集可以显著提高  
开发人员的效率和价值实现时间

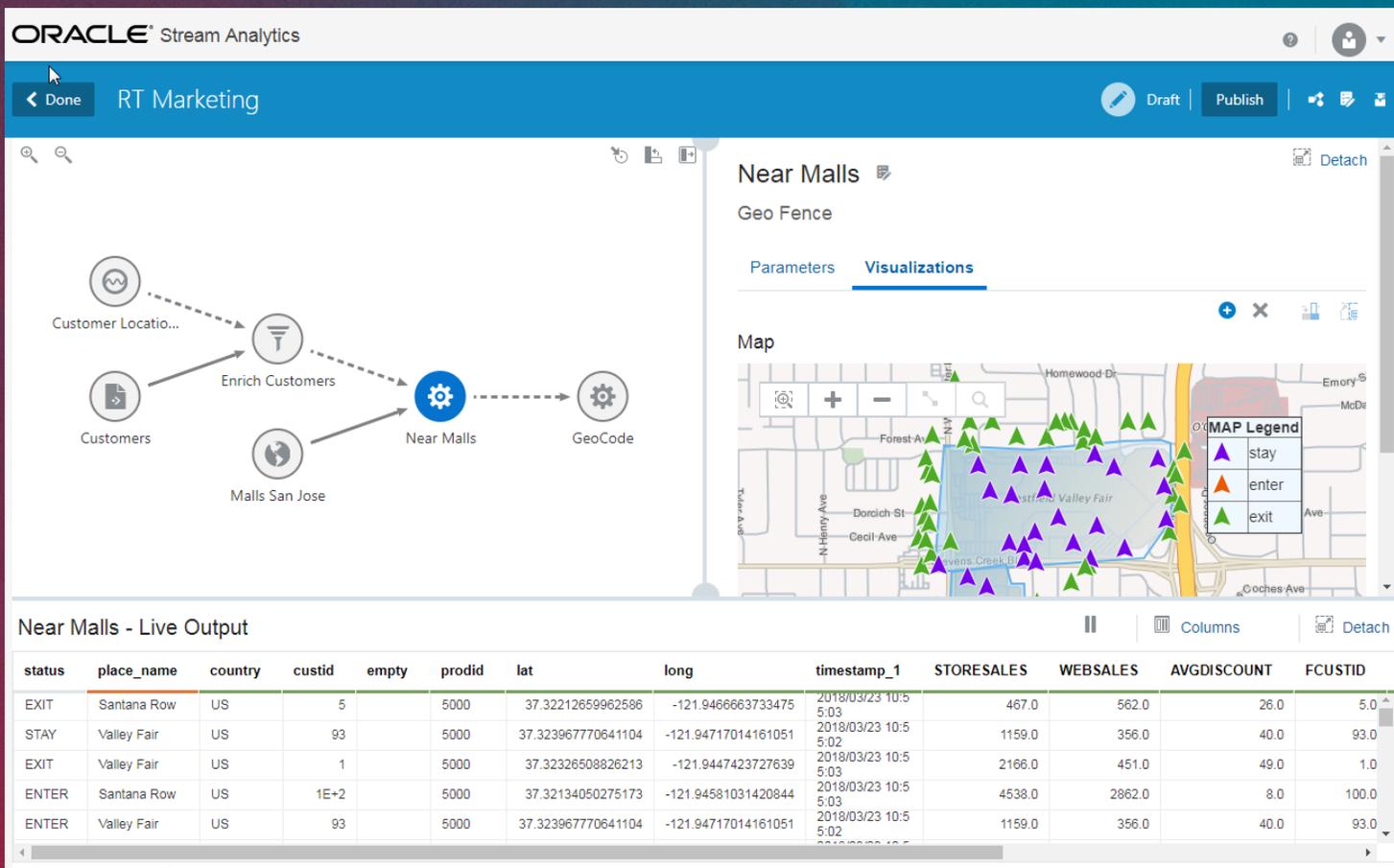
## 构建事件管道

工具可以轻松利用地理围栏、机  
器学习和数据流中的其他查找数  
据

## 提供数据服务

可以将数据发送到Kafka、数据库、  
或者到缓冲层作为后续ETL作业的  
数据源

# Oracle Stream Analytics实时营销场景



1

关联客户LBS信息和用户画像

2

设置商圈地理围栏和识别方式

3

实时在线预测

4

圈出消费者发送个性化推荐

# 批量数据处理

## 流数据和批量数据处理融合

### 流数据处理

Stream Analytics



### 批量数据处理



对象存储  
HDFS

### 批量数据处理场景

- 数据清洗转换
- 数据湖加载
- 分布式机器学习

### 批量数据处理

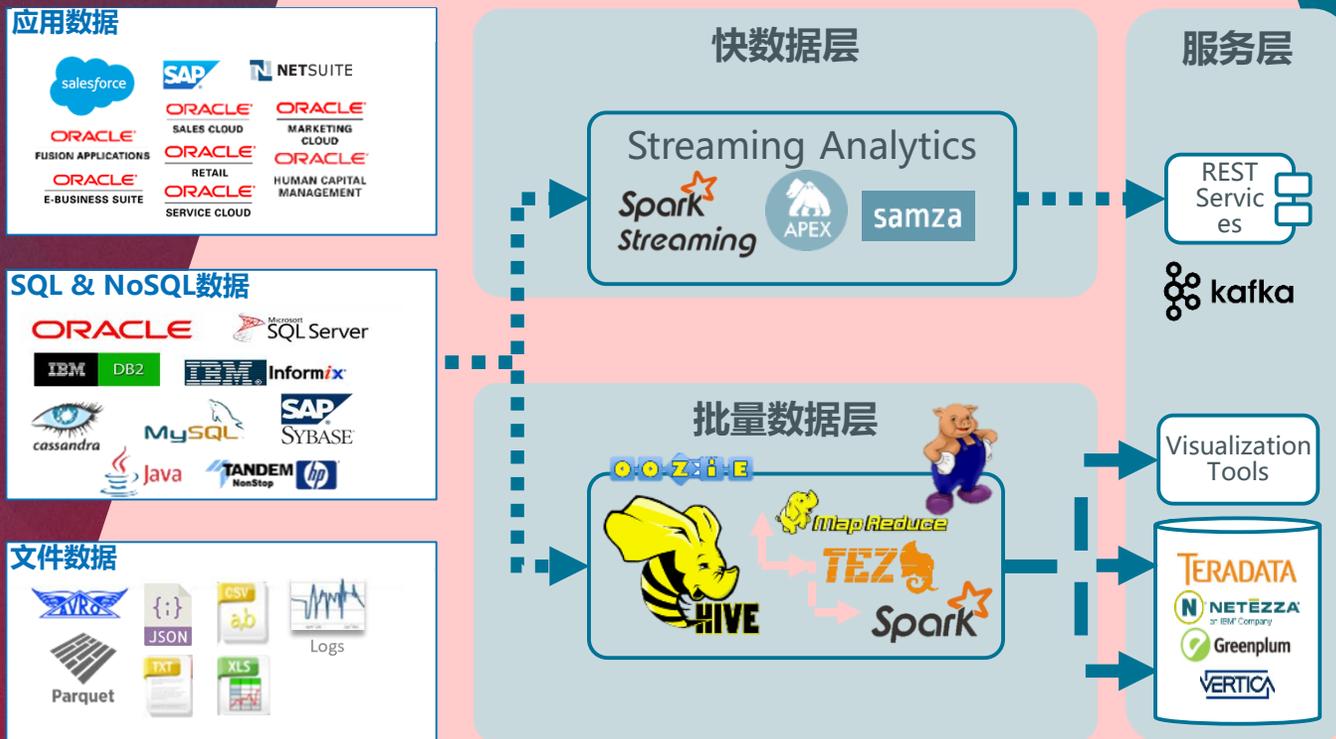
- 计算引擎：Spark、Hive
- 存储：对象存储、HDFS

### 交互式数据访问

- 直接查询超大型数据集
  - Hive SQL
  - Spark SQL
- 机器学习与图计算
  - Spark MLlib
  - Spark GraphX

# Oracle Data Integrator – 企业级批量数据处理解决方案

## Oracle Data Integrator



简化代码开发

集成异构数据源和目标

不同种类数据统一处理

优化数据回流

逻辑物理设计分离



# 实时数据查询

完全自治的数据  
仓库云服务



自治数据仓库云服务

集成了机器学习算  
法，专为智能分析  
设计

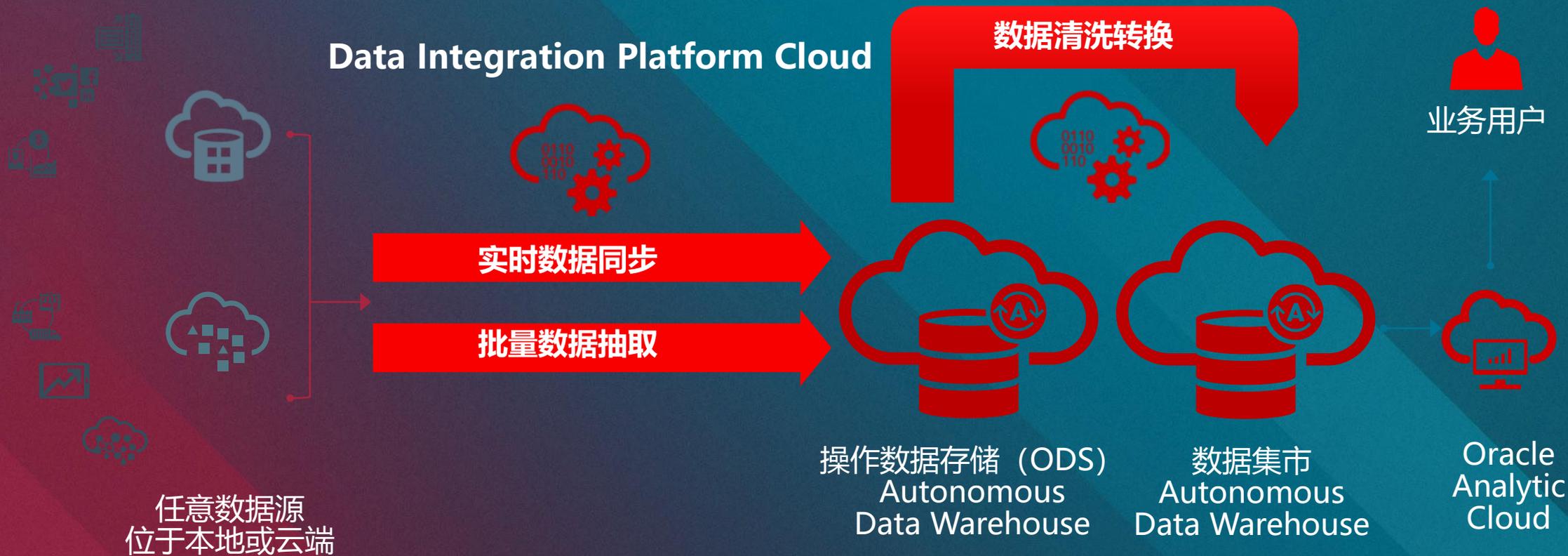
快速

简单

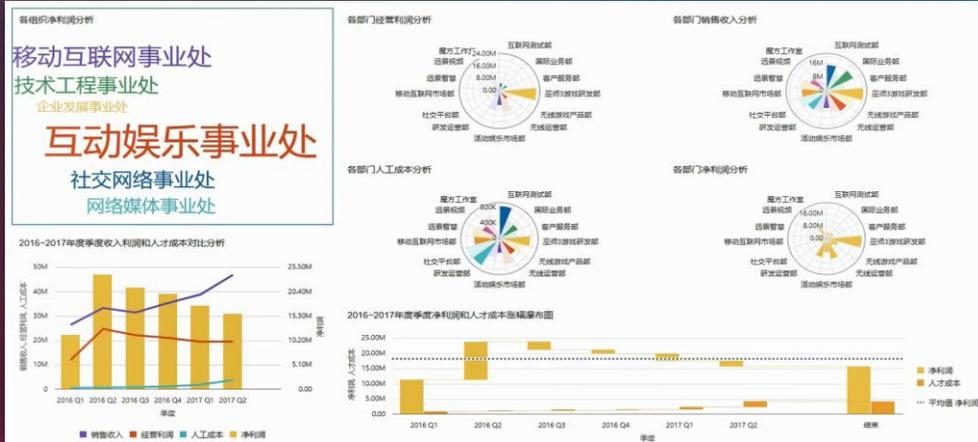
弹性

# Oracle自治数据仓库实时数据采集

## Data Integration Platform Cloud



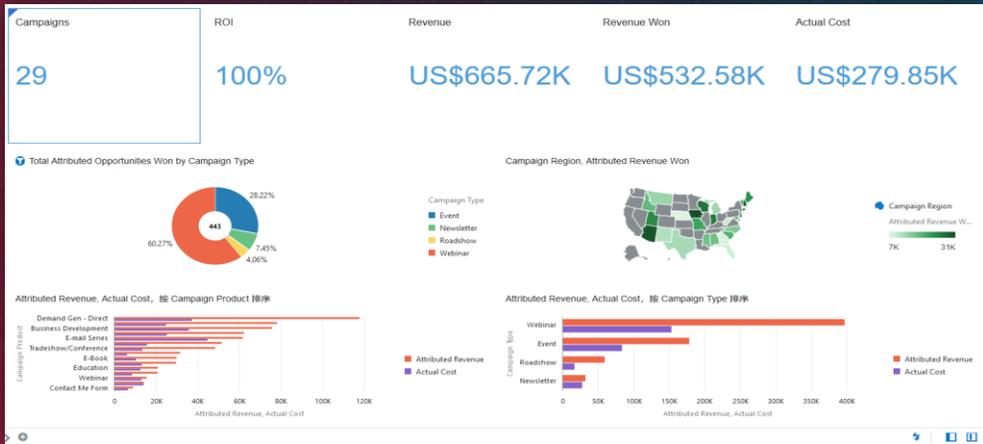
# 实时洞察



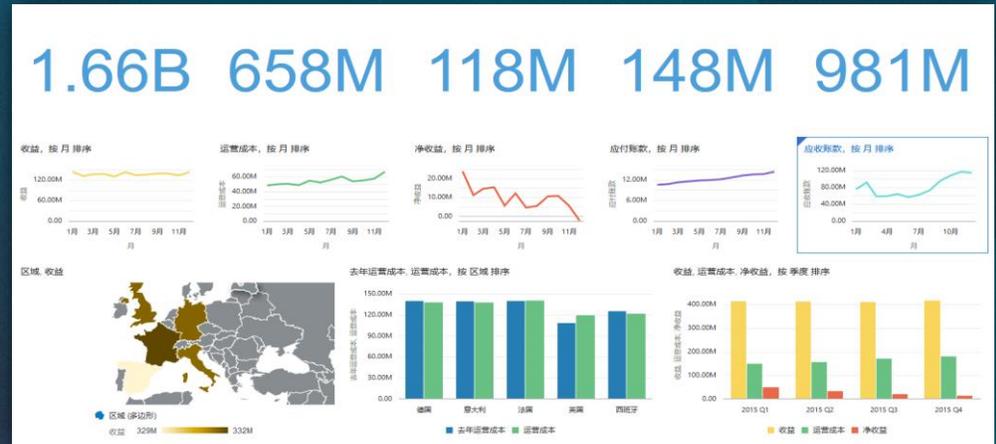
## 人事分析



## 销售分析

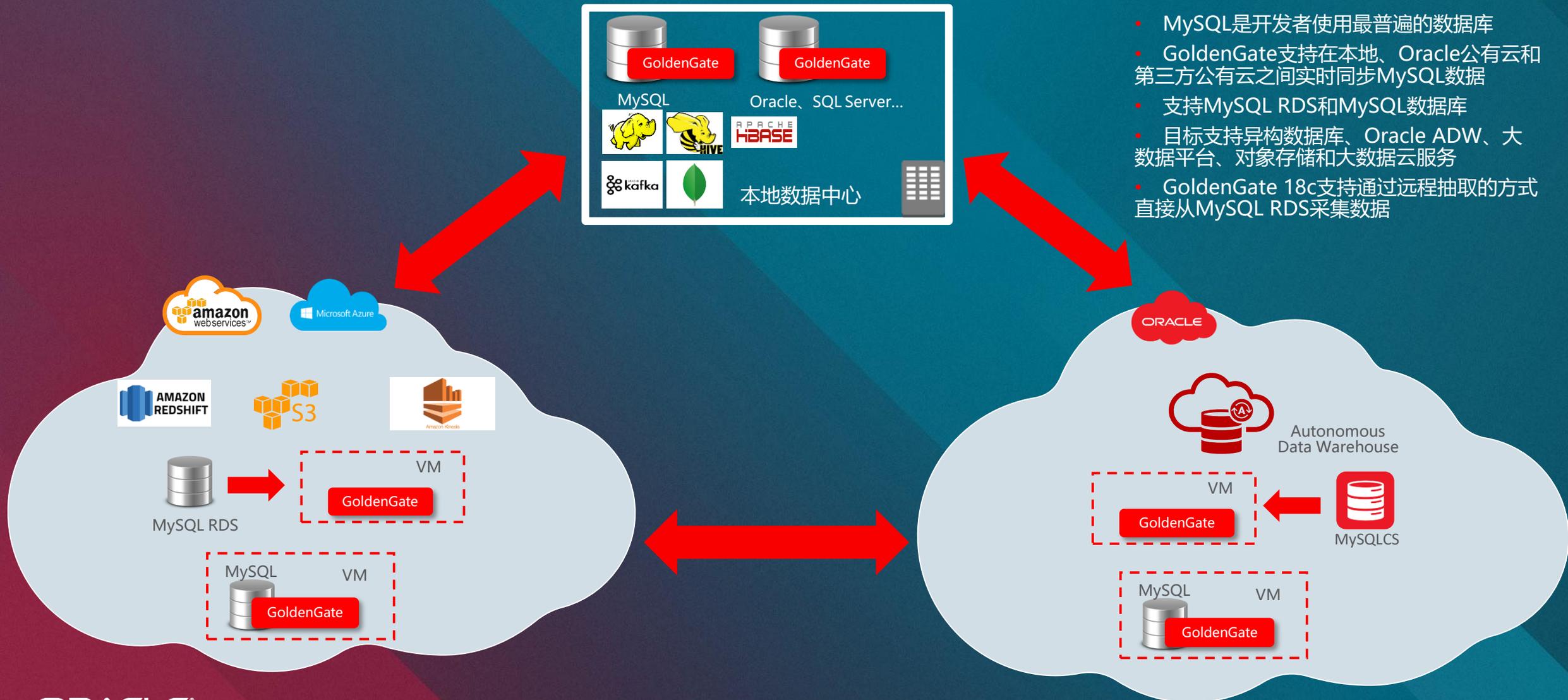


## 市场分析



## 财务分析

# 实时数据同步



- MySQL是开发者使用最普遍的数据库
- GoldenGate支持在本地、Oracle公有云和第三方公有云之间实时同步MySQL数据
- 支持MySQL RDS和MySQL数据库
- 目标支持异构数据库、Oracle ADW、大数据平台、对象存储和大数据云服务
- GoldenGate 18c支持通过远程抽取的方式直接从MySQL RDS采集数据

# GoldenGate 18c 对MySQL的支持

## MySQL数据库支持

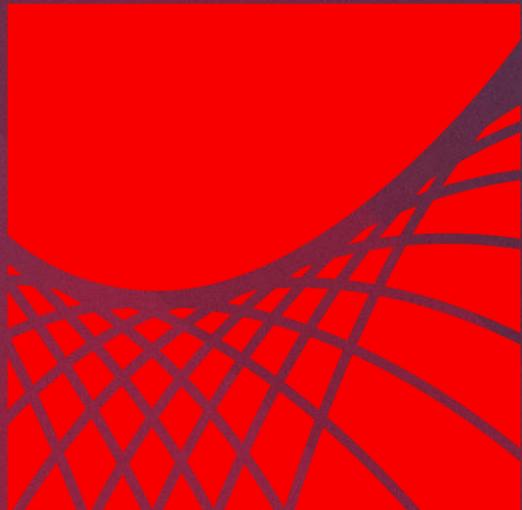
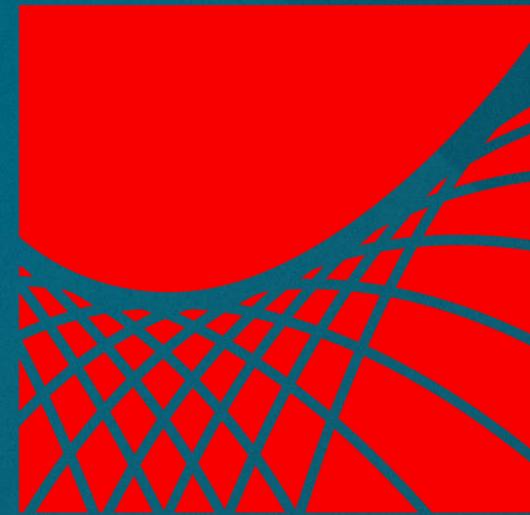
支持远程抽取的方式从MySQL RDS、MySQL实时采集数据

## DDL同步支持

支持MySQL之间DDL复制

## 支持MySQL Native Replication

支持从MySQL Native Replication的Slave节点采集数据



# GoldenGate for MySQL的优势

## Oracle GoldenGate

支持MySQL RDS、MySQL远程抽取

非侵入

低影响

支持8倍数据传输压缩，网络带宽占用少，  
确保低延迟数据同步

高性能

异构的源和目标

直接部署和运行在源端数据服务器时，  
CPU占用率低

在多云环境广泛支持异构数据库、操作系统、  
大数据平台、公有云服务

# 数据服务

## 基于SQL的实时查询

- 通过批量数据处理将分析汇总回流到数据集市
- 利用Oracle自治数据仓库极致的性能实现实时查询分析

## 在多云之间实时同步数据

- 支持在本地和多云之间实时同步数据
- 支持各种异构的数据源和目标

## 数据服务API

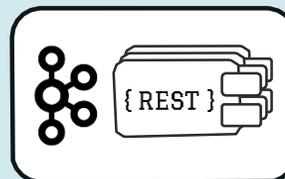
- 发布和订阅 (Kafka) , 提供基于REST API的数据服务
- 基于REST意味着API网关可用于安全访问控制列表

## 直接访问原始数据

- 访问对象存储、HDFS文件系统的数据
- 特别适用于机器学习等离线计算任务

## 多种方式消费数据

### 数据服务



### 数据消费

#### 数据服务API



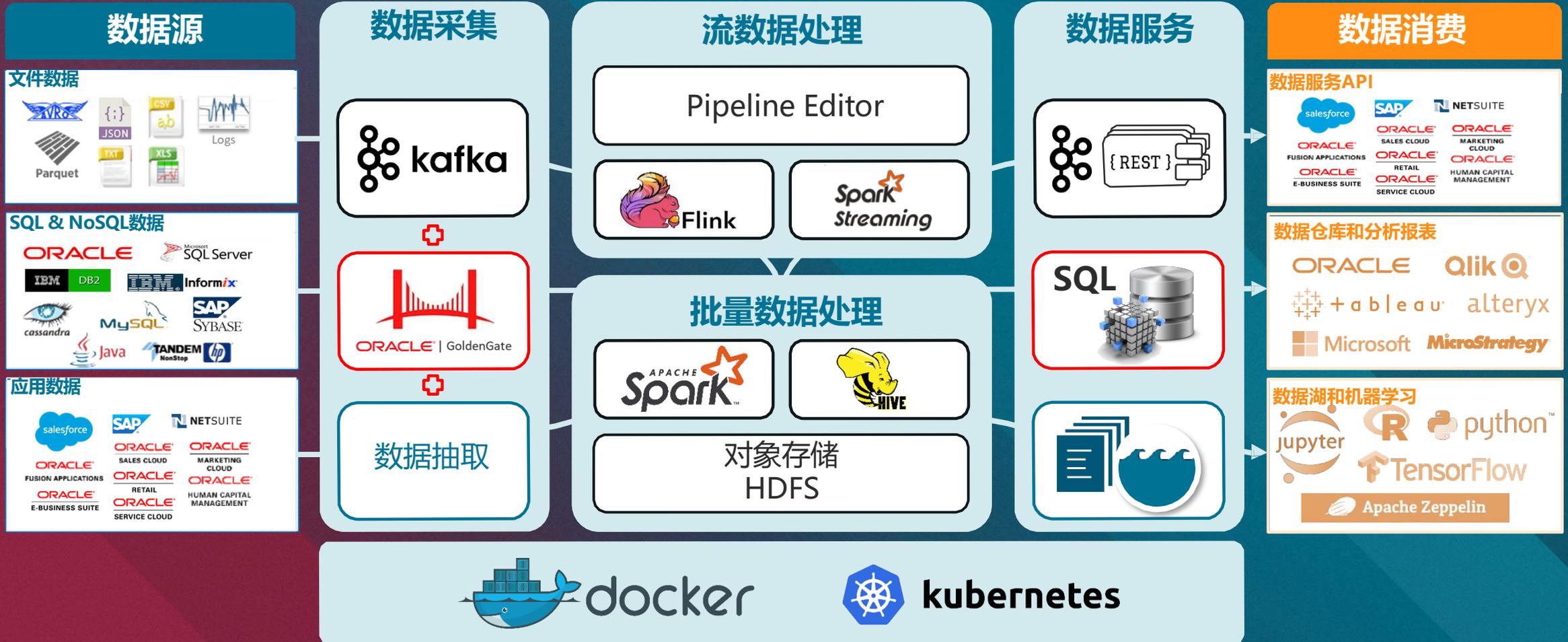
#### 数据仓库和分析报表



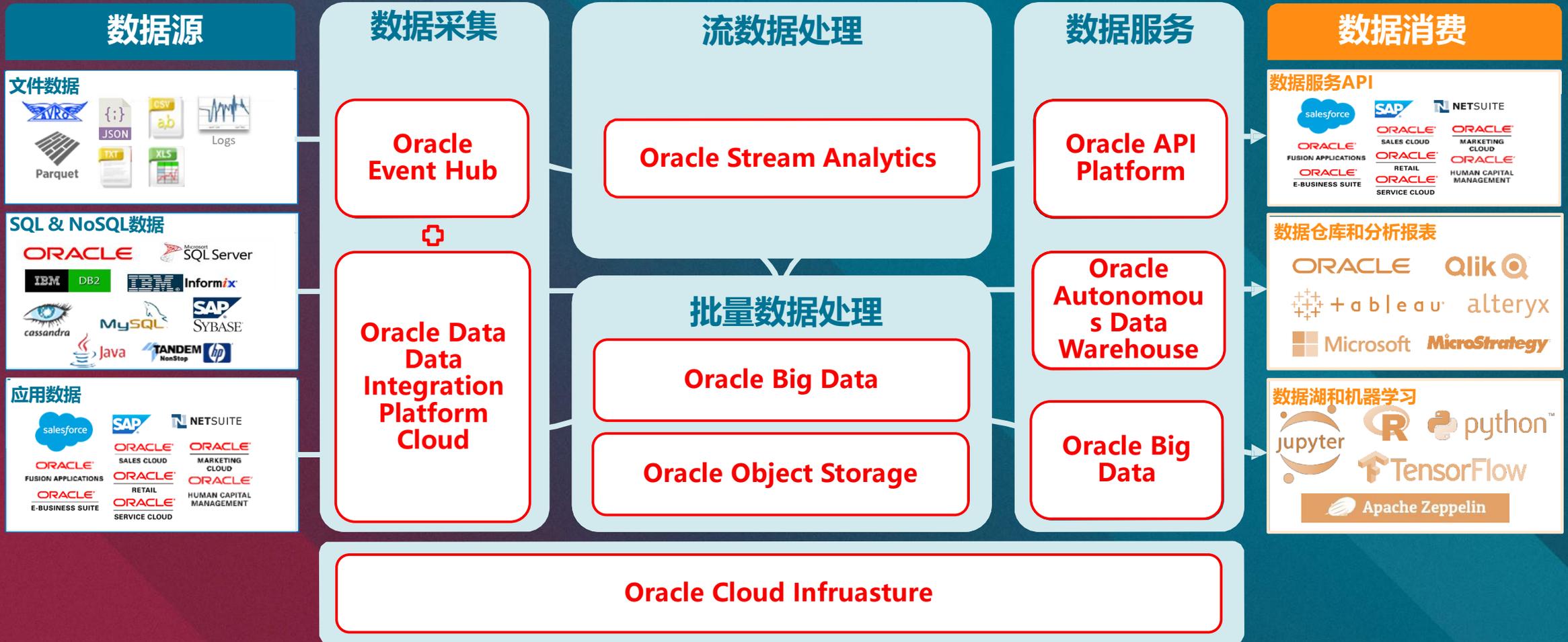
#### 数据湖和机器学习



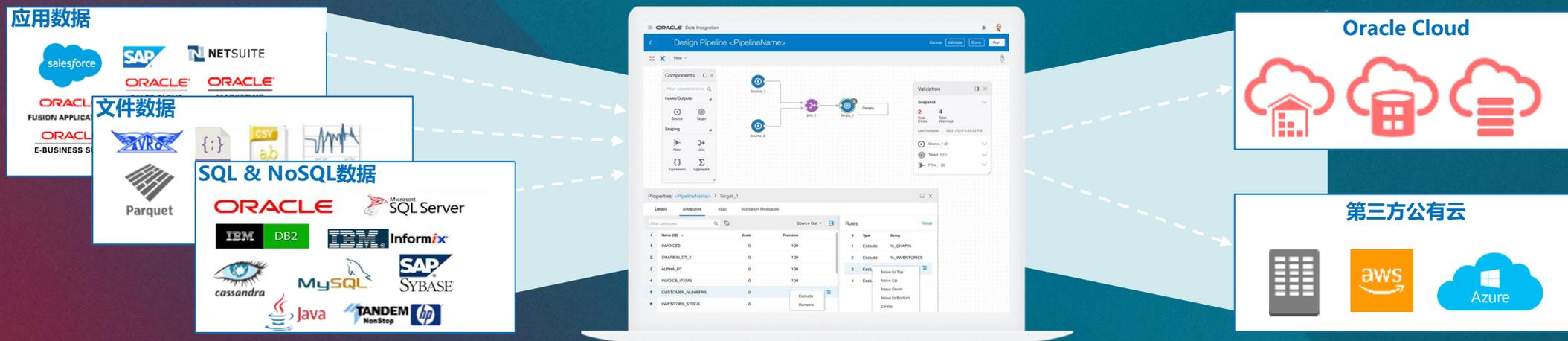
# 新一代数据集成平台技术选型



# 新一代数据集成平台技术选型



# 数据集成平台云 - Data Integration Platform Cloud (DIPC)



连接数据库、文件和应用

实时数据采集

实时数据同步

数据仓库  
自动化数据处理

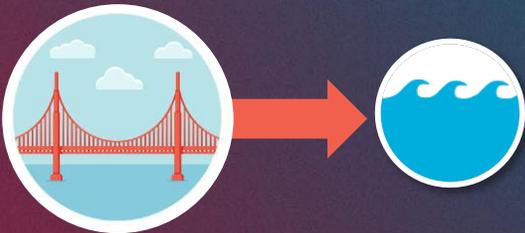
数据湖  
自动化数据处理

数据治理

# 数据集成平台云全方位覆盖数据集成需求

## 数据实时采集

业界领先的实时数据采集



## 数据湖加载

快速加载数据到数据湖



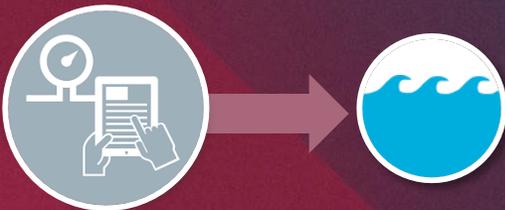
## 流数据处理

实时处理海量流数据



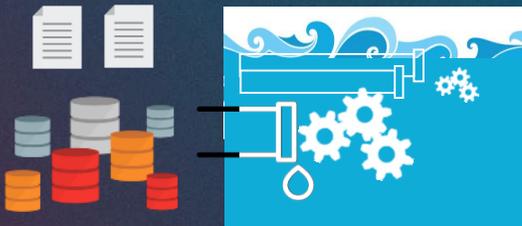
## 数据预处理

数据标准化、数据质量、关联汇总、特征工程



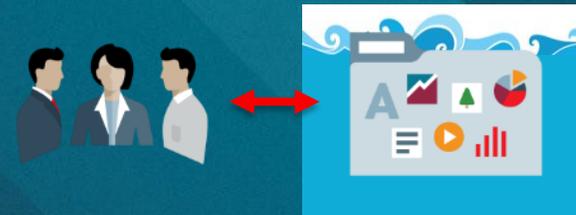
## 数据治理

全面数据质量管理

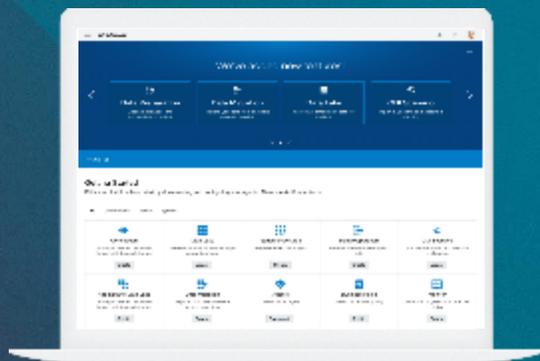


## 数据目录

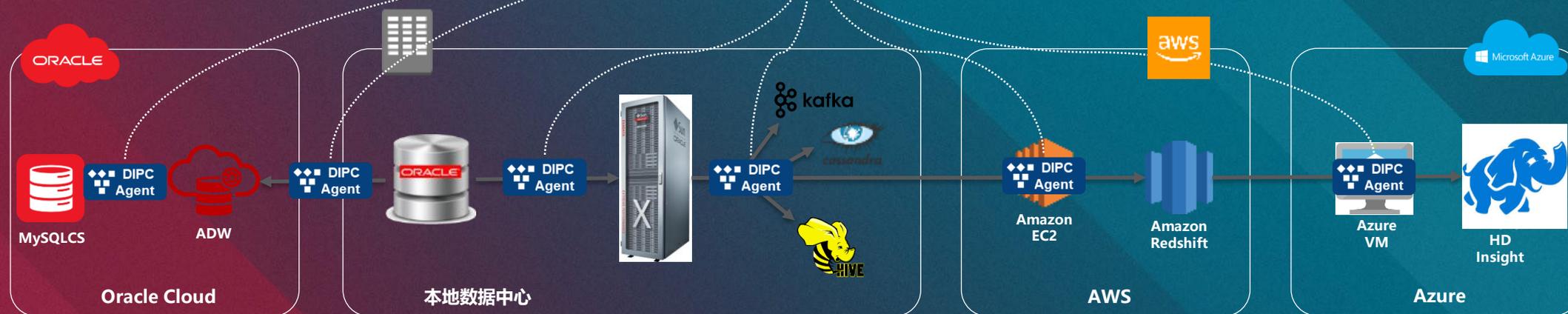
企业级数据资产地图



# 数据集成平台云实现Integration Anywhere



- DIPC支持在本地和云端同步、清洗转换数据，无论数据位于本地、Oracle公有云或者是第三方公有云
- DIPC提供图形化的配置向导和集中式的监控管理平台，帮助用户降低学习成本、提高运维效率



# 数据平台集成云提升开发者效率



DBA

- 开发数据采集、批量数据处理任务
- 使用SQL、存储过程、Spark SQL、Shell等



大数据开发者

- 开发数据湖加载、流数据处理、批量数据处理任务
- 使用Python、Java、Scala、SQL等



数据仓库开发者

- 开发数据仓库/数据集市加载、批量数据处理任务
- 使用SQL、存储过程、Shell、Java等



数据管家

- 管理数据清洗规则、数据质量管理、数据标准管理
- 一般使用工具



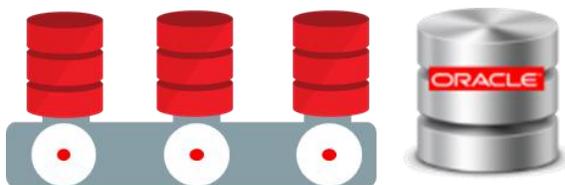
数据科学家

- 业务分析、特征工程、开发机器学习任务
- 使用R、Python、工具

DIPC提升数据平台开发者的研发效率

## Data Integration Platform Cloud

Oracle数据库



非Oracle数据库&文件数据



企业管理套件 & SaaS



# 数据集成平台云激活实时洞察

实时数据采集



DIPC实现高吞吐、低延迟的实时数据采集

实时数据处理



Stream Analytics提供企业级流数据处理能力

实时数据查询



DIPC实时采集数据到ADW里提供实时交互查询

实时数据同步



DIPC实现在多云环境数据库之间实时同步数据

实时洞察

谢谢

*ORACLE  
CODE*