基于数加的大数据仓库解决方案

署名：宁海元
职称：袋鼠云 CTO

主办单位：杭州 Alibaba Group 阿里巴巴集团 战略合作伙伴：intel

公司简介

袋鼠云由多名前阿里云资深技术专家创立，核心员工来自阿里巴巴、神州数码等

做为阿里云战略级合作伙伴，专注于为企业客户提供云计算和大数据技术服务及产品

袋鼠云是数加平台首个金牌合作伙伴，也是阿里云认证的区域服务商和云市场供应商

袋鼠云总部位于**杭州**，在**北京、贵阳、苏州**等地设有分公司（办事处）

¥1000 万人民币

2015年11月20日注册成立，获得天使投资

$450 万美元

2016年6月获得元璟资本领投盈动资本跟投的Pre-A轮投资

使命

让企业数据产生商业价值

愿景

做DT时代最好的云计算和大数据服务提供商

目标

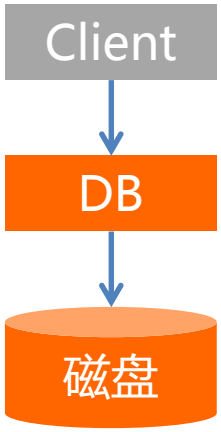
为10万+企业客户提供技术服务

扫码观看大会视频

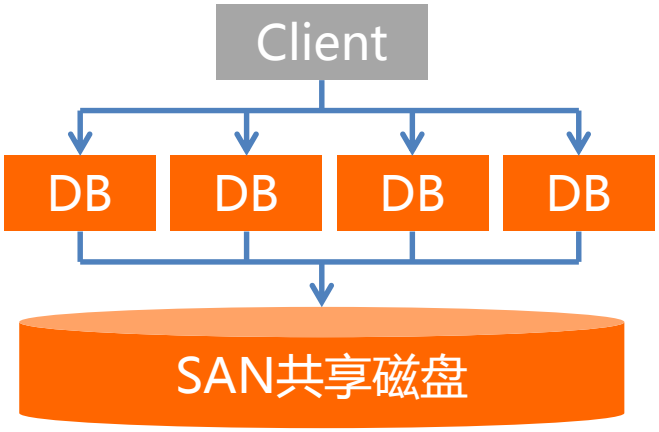
目 录
content

扫码观看大会视频

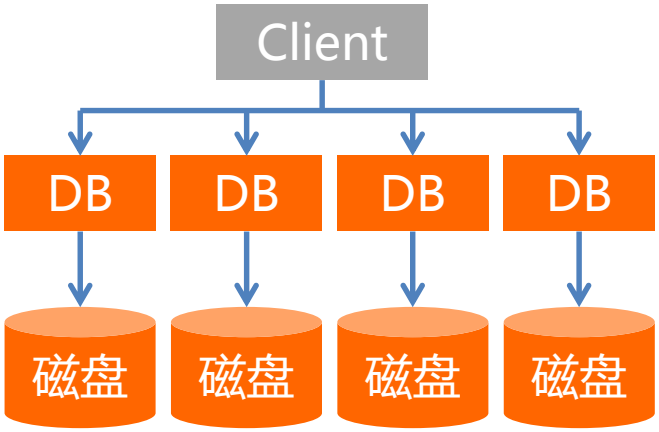
企业数据仓库现状



SMP
代表：小型机



共享磁盘
代表：Oracle RAC



MPP
代表：Greenplum,Teradata

- 存储能力受到制约，在数据日益增长的今天，集中式架构已不满足大数据的存储需求。
- OLTP与OLAP混合，导致平台管理已无法兼顾大数据的特殊要求(任务管理，元数据管理等）。
- 针对日益增长的大数据业务需求，需要提供大数据算法、组件支持。

扫码观看大会视频

存储	计算	安全
• 单机/共享存储扩展存在明显的天花板，大数据时代容易撞墙 • 价格高昂，无法保存全量数据	• 单机、RAC、MPP的计算能力存在天花板，ETL时间越来越长，早上九点前无法完成全部计算任务 • 任务调度方案复杂，需要很高的学习成本	• 数据仓库容灾成本高，很多企业的数据仓库没有做容灾，数据存在丢失的风险 • Hadoop的安全粒度较粗，无法满足多租户的隔离

数加平台的优势


PC APP WI-FI 基站 工业传感器 手环手表
大规模计算 实时计算 流式计算
SQL MR Shell 图形化
数据采集
计算引擎
数据加工
数加：打通数据生产全链条
数据应用
机器学习
数据分析
广告 推荐 定向 可视化大屏 智能语音
决策树 聚类 神经网络 支持向量机
拖拽 Report Dill down

扫码观看大会视频

- 拖拽式的操作界面
- 工作流调度
- 元数据管理
- 数据质量管理
- 一键式任务发布
- 可视化任务监控
- 完善的权限管控
- 弹性扩容

扫码观看大会视频

实时流计算

功能强大

强大流计算引擎；丰富数据采集工具；深度整合各类云存储

性能优越

关键指标超越Storm的性能6到8倍，秒级乃至毫秒级延迟，单个作业吞吐量可做到百万级别

简单易用

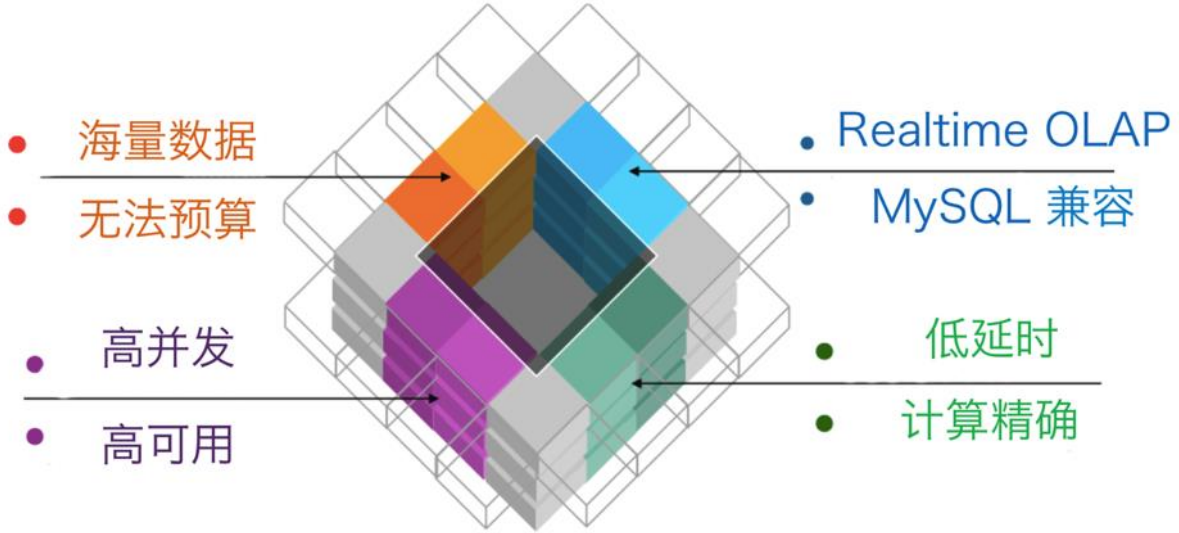
支持流式SQL处理(StreamSQL)；提供全流程流计算开发套件

成本低价

开发流程化、机器规模化；比Storm更省人力和机器成本

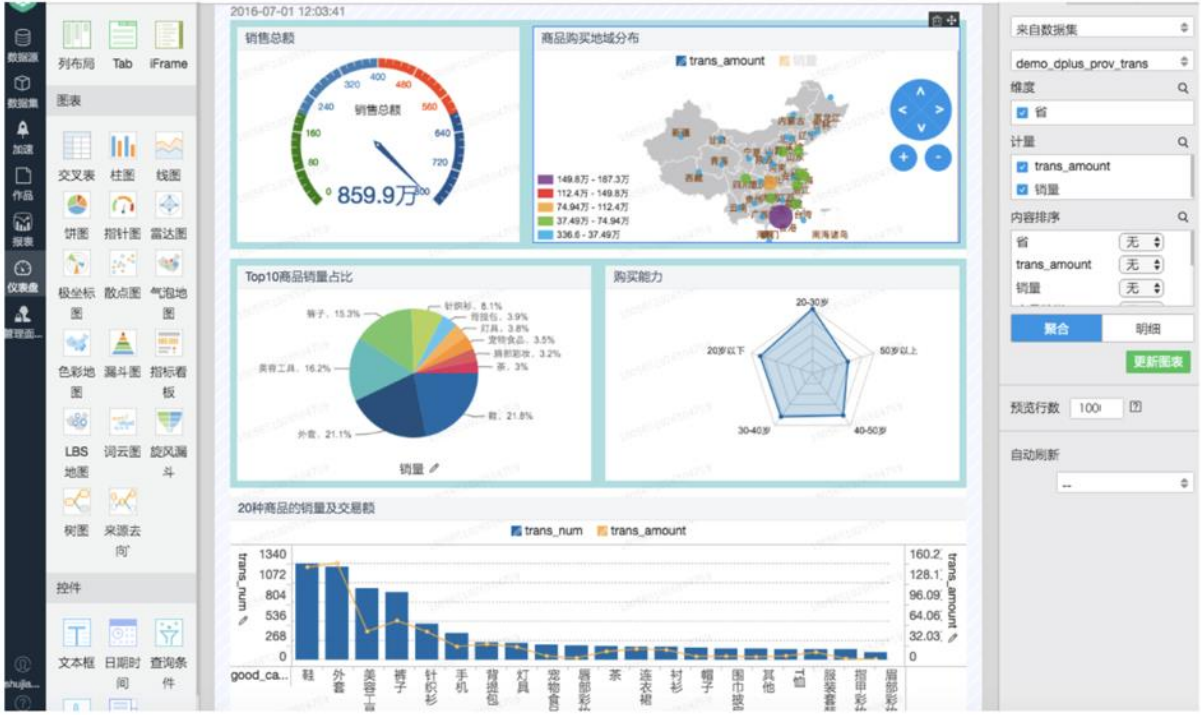
分析型数据库

海量数据实时高并发在线分析（OLAP）云计算服务，百亿数据，毫秒级计算。

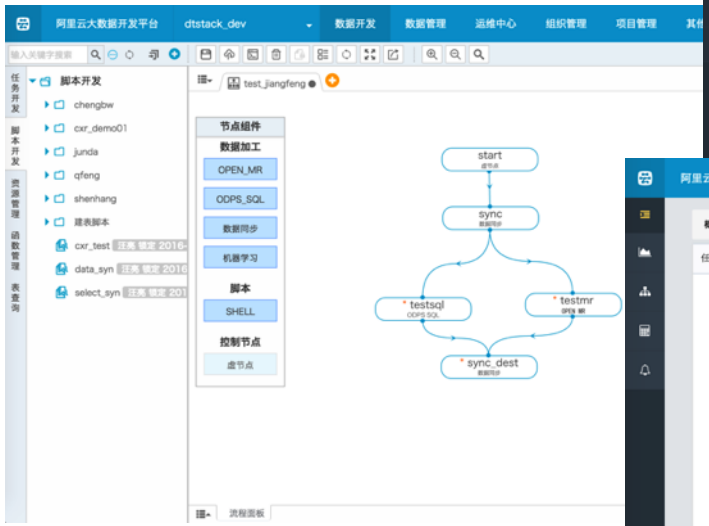


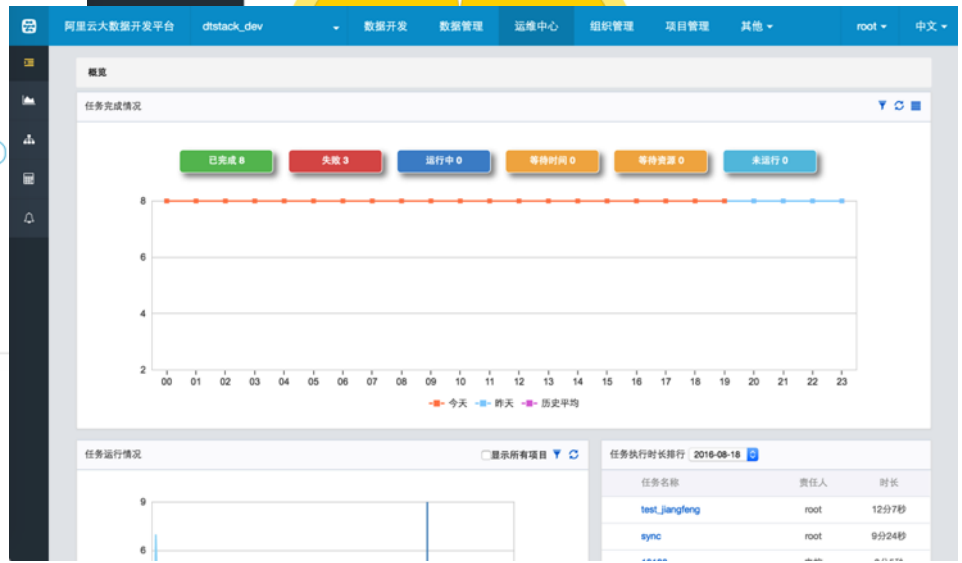
扫码观看大会视频

敏捷BI


2016-07-01 12:03:41
销售总额
859.9万
商品购买地域分布
trans_amount
Top10商品销量占比
销量
购买能力
20-30岁
30-40岁
40-50岁
50岁以上
20岁以下
20种商品的销量及交易额
trans_num
trans_amount
good_ca...
鞋
外套
美容工具
裤子
针织衫
手机
背提包
灯具
宠物食品
眉部彩妆
茶
连衣裙
衬衫
帽子
围巾/披肩
其他
T恤
服装套装
指甲彩妆
眉部彩妆
来自数据集
demo_dplus_prov_trans
维度
省
计量
trans_amount
销量
内容排序
聚合
明细
更新图表
预览行数
自动刷新
列布局
Tab
iFrame
图表
交叉表
柱图
线图
饼图
指针图
雷达图
极坐标图
散点图
气泡地图
色彩地图
漏斗图
指标看板
LBS地图
词云图
旋风漏斗
树图
来源去向
控件
文本框
日期时间
查询条件

Data IDE





扫码观看大会视频

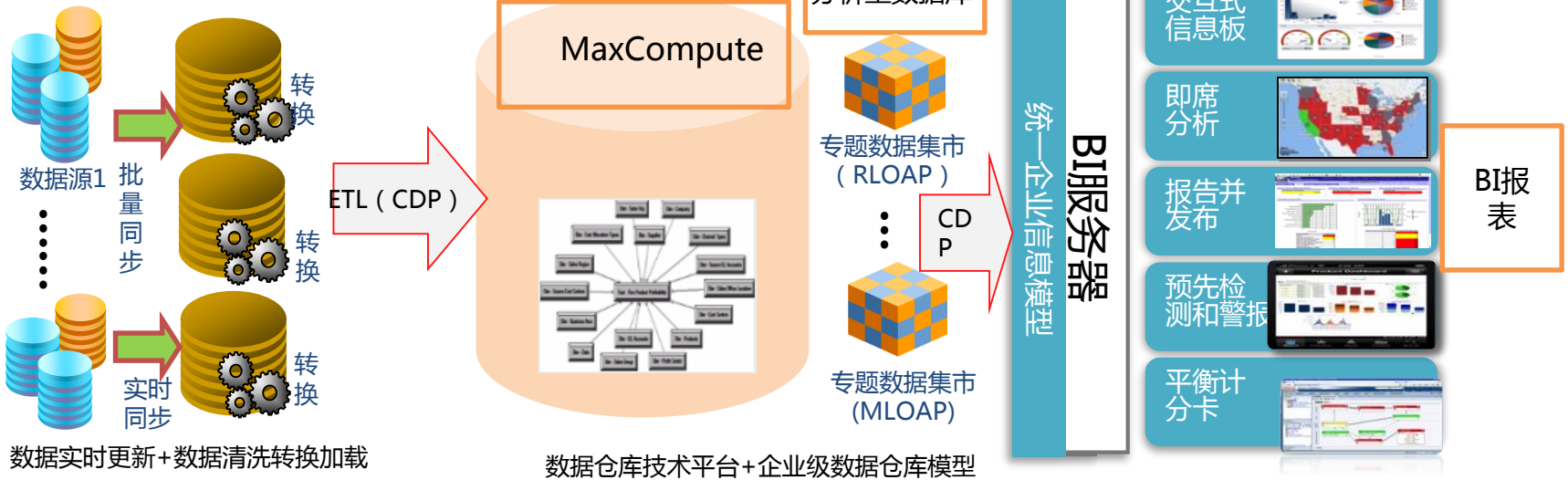
数加平台：完整的数据仓库平台

- 数据如何进入数据仓库
- 数据如何管理,性能和空间
- 用户如何使用数据提升管理思想

实时/批量数据同步+数据清洗转换 → 数据同步 / 数据转换

数据仓库+主题数据集市 → 中央数据仓库

商务智能分析平台+分析主题 → 信息展现平台



数据实时更新+数据清洗转换加载

数据仓库技术平台+企业级数据仓库模型

扫码观看大会视频

存储费用　1TB < 4000/年

基础价格	大于100GB部分	大于1TB部分	大于10TB部分	大于100TB部分	1PB以上部分
0.0192元/GB/天	0.0096元/GB/天	0.0084元/GB/天	0.0072元/GB/天	0.006元/GB/天	请通过工单联系我们

计算费用　10CU 15000/年起卖

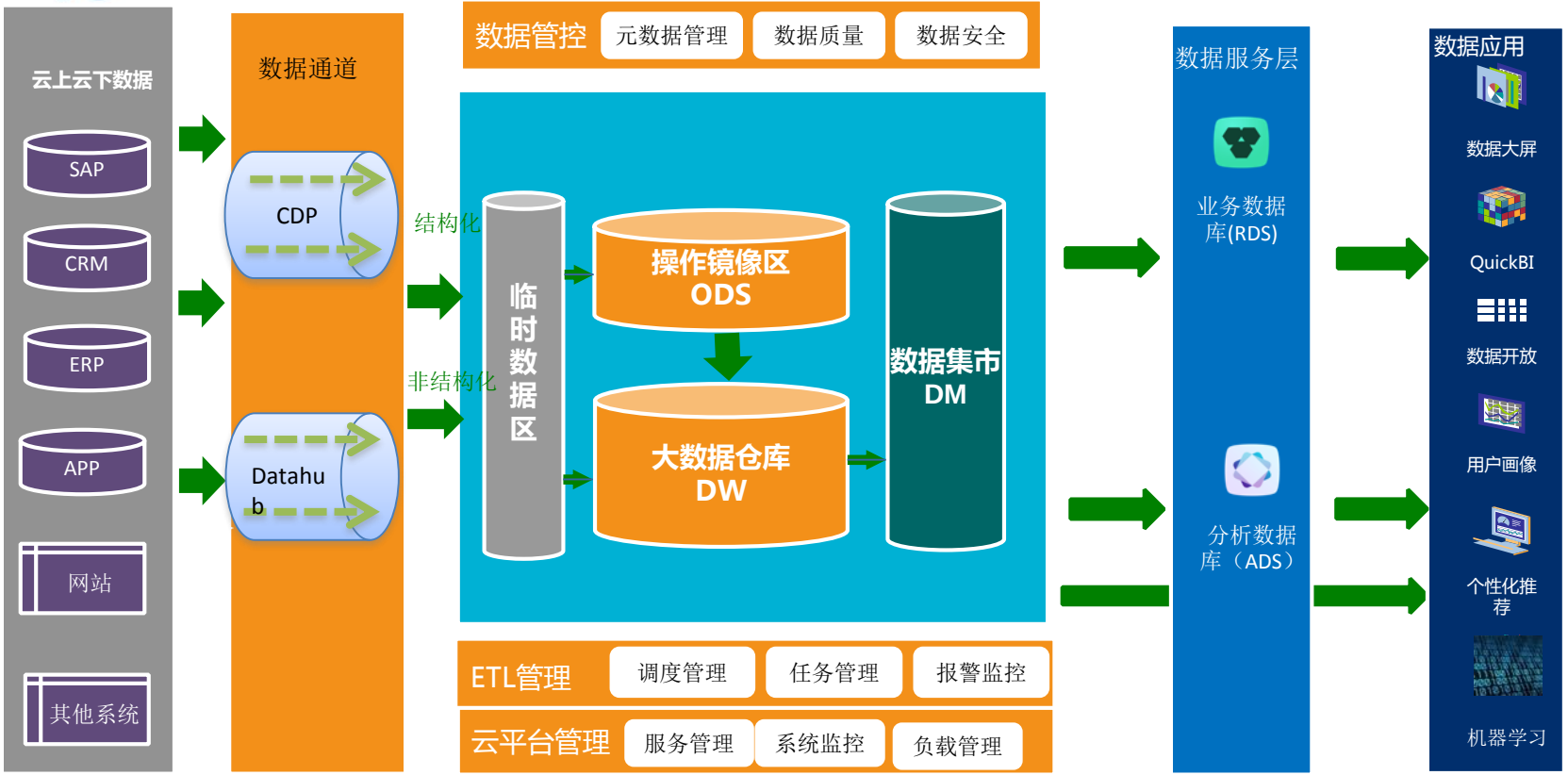
资源定义	内存	CPU	售价
1 CU	4GB	1 CPU	150元/月

扫码观看大会视频

袋鼠云解决方案

2016.6.20 袋鼠云成为数加首个签约的金牌合作伙伴

扫码观看大会视频


云上云下数据
SAP
CRM
ERP
APP
网站
其他系统
数据通道
CDP
Datahub
结构化
非结构化
数据管控
元数据管理
数据质量
数据安全
临时数据区
操作镜像区 ODS
大数据仓库 DW
数据集市 DM
ETL管理
调度管理
任务管理
报警监控
云平台管理
服务管理
系统监控
负载管理
数据服务层
业务数据库(RDS)
分析数据库（ADS）
数据应用
数据大屏
QuickBI
数据开放
用户画像
个性化推荐
机器学习

扫码观看大会视频


用户画像
精准营销
推荐引擎
可视化大屏
智能语音
图像识别
公众趋势
基于数加的大数据仓库(MaxCompute/StreamCompte/Analytic DB)

袋鼠云解决方案

某知名服装企业

客户介绍

某知名品牌在亚太地区全权运营机构

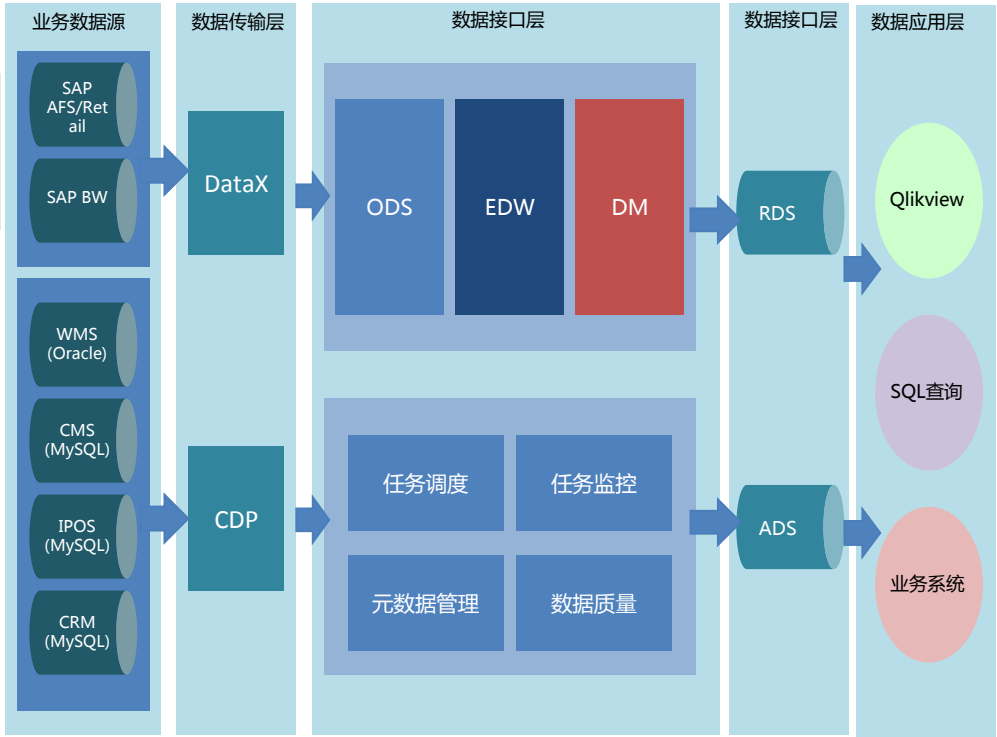
负责在亚太地区的品牌运作业务,主要经营品牌女装

客户需求

1. 随着数据量增长，即将达到单机数仓的内存天花板
2. 不支持弹性扩容，ETL计算任务执行超过9小时
3. 系统、工具较多，运维管理成本高

袋鼠云服务

1. 基于DataX二次开发服务，支持SAP数据源
2. 基于阿里云大数据仓库架构，ETL时间优化到2小时内
3. 传统数仓无缝升级到云平台，开发运维可视化统一管理



扫码观看大会视频

浙报传媒 ZHEJIANG DAILY MEDIA
百联集团
GQY
BYD
天弘基金 TIANHONG ASSET MANAGEMENT
亿享金服 ESHARE.CN
JIAGUWEN 甲骨文超级码
轻松筹
Leqee 乐视
美购 全球精品超市
街蜜 jiemi
UC express 优速快递

扫码观看大会视频

2016 The Computing Conference

THANKS