



华为云EI大数据服务及部分解决方案概述

华为云EI：让企业更智能



物流



电商



IOT



制造



金融



智慧城市



运营商



Enterprise Intelligence

行业场景

车联网

智能物流

智能制造

智能水务

.....

数据分析

Data Lake Factory(workflow编排/调度/DataIDE)

MRS

一站式大数据集群

CloudStream

实时流计算

DLI

数据湖探索

AI基础平台

MLS/DLS/GES/Batch

AI通用服务

视觉/语音

数据计算

CloudTable

表格存储

CSS

云搜索服务

DWS

数据仓库

数据存储

OBS、MRFS

数据集成

DIS(数据接入)

CDM(数据迁移)

使能

- 基于实践提供丰富的行业算法和预置模型
- 规模的AI实践经验，使能企业快速步入智能时代

敏捷

- 强大的SQL能力，业务移植便捷
- 丰富的工具支持，开发运维高效
- 完善的云服务，平滑迁移

可信

- 全组件HA、异地容灾、金融等保
- 开放共赢，可信赖的合作伙伴
- 企业级的大数据云服务，开放接口

华为EI云大数据：围绕数据开发流程，构筑全生命周期全栈服务



托管服务：
1.完全可控
2.平滑迁移

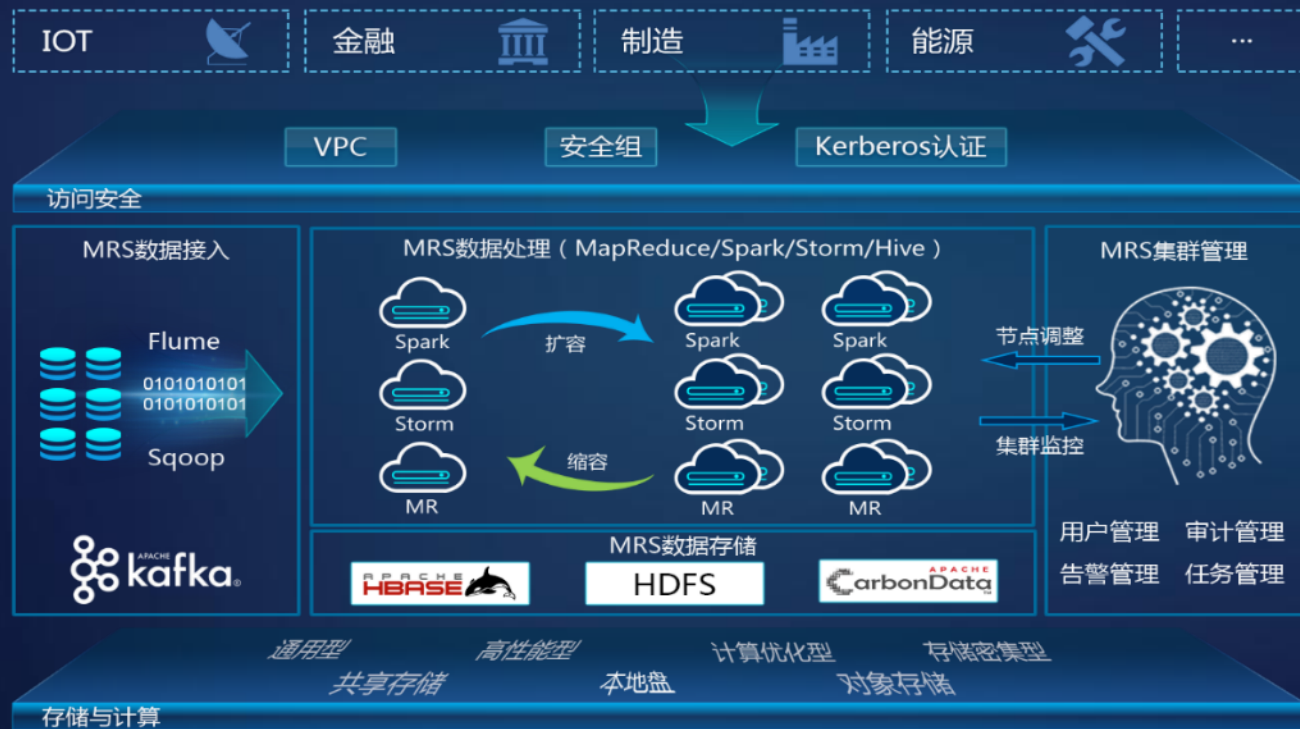
云原生服务：
1.按需付费
2.免维护



MRS：一站式大数据平台，Hadoop全家桶



MapReduce服务（MapReduce Service）：开源开放、租户完全可控的企业级Hadoop大数据集群云服务。更优地利用华为云服务资源，轻松部署和运行HDFS、MapReduce、Spark、HBase、Kafka等开源组件，为客户提供安全、稳定、低等本，高性能的大数据平台



易用：**自动弹性伸缩**



安全：获得欧洲**PSA认证**，支持**Kerberos安全认证、多级用户管理**



快速：内嵌Carbondata，**万亿数据秒级查询**



可靠：**全节点HA，实时**短信/邮件通知

成功案例

MRS服务，助力新再灵开创智能梯联网新纪元



实时监控：将处理后的数据导入电梯实时监控系统，对故障电梯及时上报告警。

极速查询：借助MRS HBase极速点查及批量scan能力，配合ElasticSearch向业务层提供电梯数据极速检索查询服务。

多维分析：通过使用MRS Spark组件，轻松构建批量计算应用，定时从HBase获取数据进行多维数据分析，生成云梯运行报表供上层业务部门使用

云搜索：云上Elasticsearch，全托管搜索和分析平台



云搜索服务（Cloud Search Service）是将软件(Elasticsearch、Kibana)及硬件(计算、存储、网络)云化增强后，为用户提供的全托管搜索和分析平台。

数据来源

云搜索服务

搜索及分析场景

搜索功能

- 支持Poisson、IK、拼音、简繁体等多种分词器
- 支持近义词、同义词
- 支持向量数据类型
- 支持搜索结果高亮显示
- 支持精确检索、模糊检索、组合条件检索、地理位置检索等

增强功能

- 支持用户自定义快照策略
- 支持用户自定义词库
- 支持汉明距离打分算法
- 支持乘积量化打分算法
- 支持欧式距离打分算法

处理能力

- 支持PB级数据
- 支持上百节点集群
- 支持在线扩容和词库更新无缝对接
- OBS/HBase/Kafka等多种数据源
- 近实时数据高性能数据索引



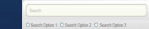
轨迹分析



以图搜图



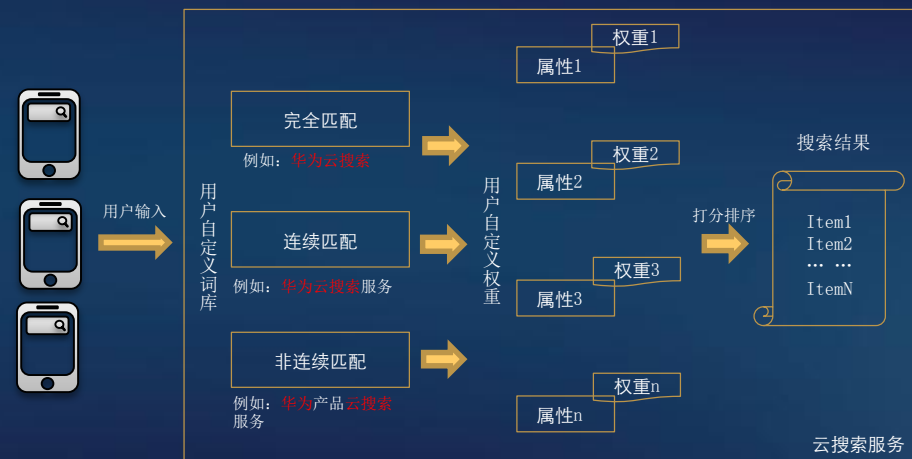
声纹检索



全文检索

成功案例

云搜索服务，帮助某互联网企业实现3亿用户社交APP的高性能、高精度搜索框



多类型检索：支持普通文本数据、AI类特征数据快速检索，满足用户文本、图片、音频、视频等多种数据检索诉求。



自研分词器：支持索引级词库，分词更精准，粒度更可控，辅助用户实现词库分离，提高搜索精准度。



零业务中断：扩容资源、自定义词库修改无需中断业务，轻松应对业务变化，提升业务可用性。



全托管运维：集群运行状态实时监测，专家团队7*24小时贴身看护，资源故障无感知替换，免除用户运维烦恼。

业务场景：手机社交APP的搜索框，对注册用户、发表内容进行搜索。

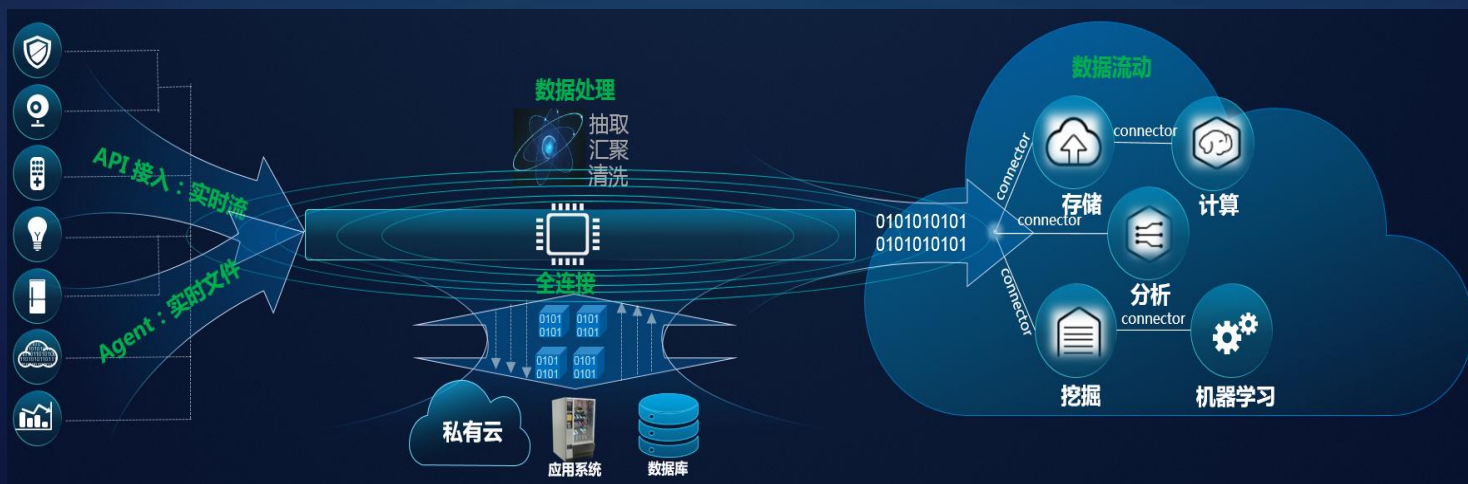
业务痛点：传统数据库搜索性能差，开源搜索引擎的精准度不高。

实现效果：搜索10亿条数据500ms内响应。根据自身业务特点，自定义词库、自定义权重，有效提升搜索精准度。

DIS：实时数据上云全连接管道



DIS (Data Ingestion Service) 是华为云提供的完全托管的实时数据接入服务。DIS提供了灵活数据采集、高效数据传输、实时数据分发能力，让您可以轻松构建基于实时数据的分析和应用。



高效：单通道**百万并发**，**毫秒级**时延，每日百TB级实时数据接入



安全可靠：支持数据**加密传输**，租户个资源和操作隔离，数据跨可用区存储，数据可用性达到**99.9%**



易用：提供多语言**SDK/Agent**，**对接**主流开源采集工具，轻松完成数据采集、传输和分发，预置数据**抽取、格式转换、压缩**等处理算子



低成本：按需使用，相比基于Flume和Kafka自建系统成本**降低5倍**

成功案例

助力某燃气集团打造IoT平台，支持实时数据业务创新



- 支撑从巡线业务**数千个**终端实时接入数据
- 实时数据入库端到端**秒级**延迟
- 数据接入相关人力成本**降低50%**

CloudTable：基于HBase、时序、时空的NoSQL数据库



基于Apache **HBase**提供的分布式、可伸缩、全托管的**NoSQL**数据存储服务，它提供了毫秒级的随机读写能力，适用于海量结构化数据、半结构化数据存储和查询应用。同时基于OpenTSDB和GeoMesa提供**时序**数据库能力和**时空**大数据查询、分析能力。



高性能：响应**时延毫秒级**，**TPS支持千万级**，横向扩展



生态开放：**兼容原生接口**
HBase/OpenTSDB/ GeoMesa



时序数据库：读写性能提升**30%-60%**，支持插值、降精度、聚合强大分析能力，10:1**高压比**，成本更低



时空大数据：帮助物联网存储和分析海量时空(spatio-temporal)数据，超越传统GIS

成功案例

助力某能源企业构建售电平台大数据服务



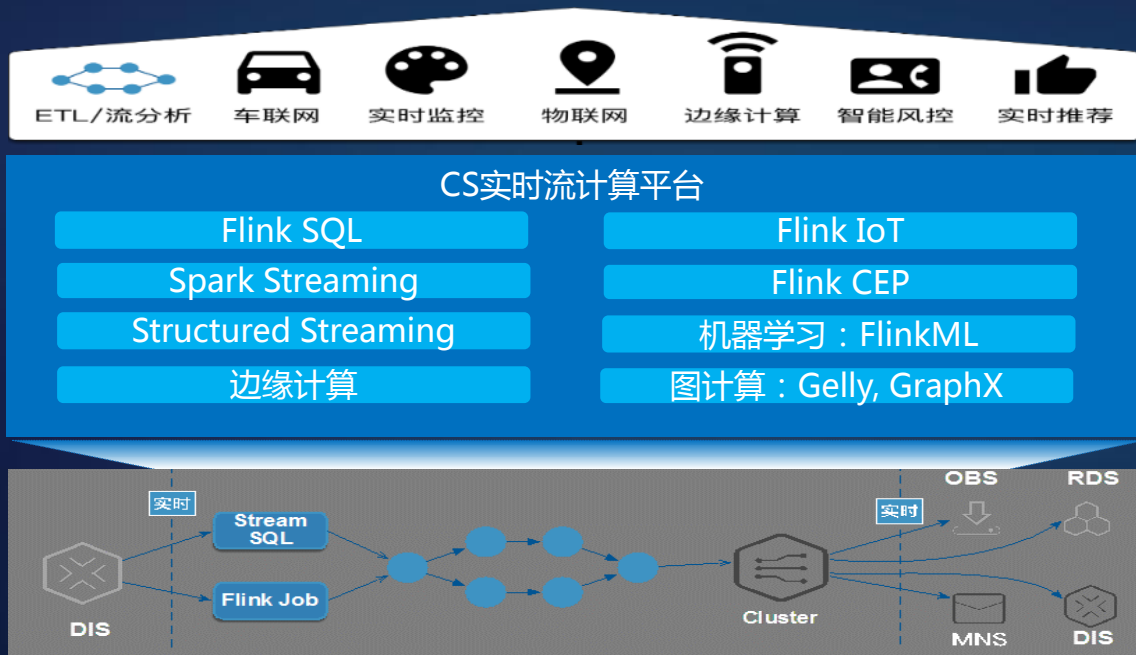
主要场景：用户需要使用Cloudtable存储电表数据，再对存储的电表数据进行点查、范围查询、联合查询和实时查询。


毫秒级响应效果：从一年207亿的数据中点查一条用时6ms，查询一天96条数据用时12.8ms，一个星期数据672条20.4ms，一个月数据2880条60ms，从一年45亿条数据中实时查出最新的一条数据用时1.1ms。

CloudStream : IoT实时流计算核心引擎



实时流计算服务 (Cloud Stream Service , 简称CS) 是实时流式大数据分析服务, 完全托管计算资源和Serverless体验, 即时执行Stream SQL、Spark和Flink作业, 提供低延时高吞吐的智能流计算平台。



 功能强: Flink + Spark **双引擎**, 业界领先

 安全: 物理隔离-**全托管**计算资源

 易用: 丰富的SQL函数, 地理位置算子, **在线编辑和调试**

 按需计费: 按SPU数+精确到**秒级**的使用时间计费

成功案例

CloudStream服务, 助力华为消费者BG精准营销

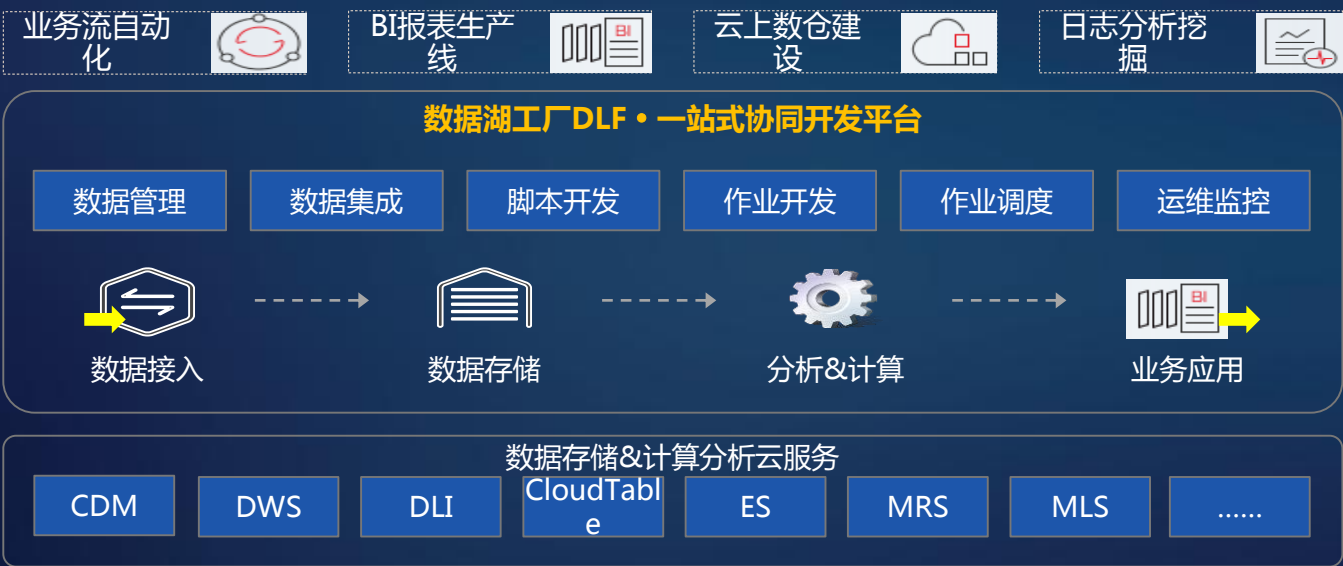


- 精准营销, 用户点击率提升**10倍**, 预约率提升**5.6倍**
- 覆盖2亿+用户, 15大类用户标签, 标签总数达到**80亿+**
- 平均每单负向舆情遏制从**10分钟**到**3分钟**

DLF：数据湖工厂



数据湖工厂服务（Data Lake Factory，简称DLF）提供一站式的大数据协同开发平台，帮助用户轻松完成数据建模、数据集成、脚本开发、作业调度、运维监控等多项任务，可以极大降低用户使用大数据的门槛，帮助用户快速构建大数据处理中心。



一站式IDE平台：**一站式**建设云上数仓，无需切换多个工具



数据湖开发：管理**多种**大数据服务，可实现**跨服务**作业编排和调度



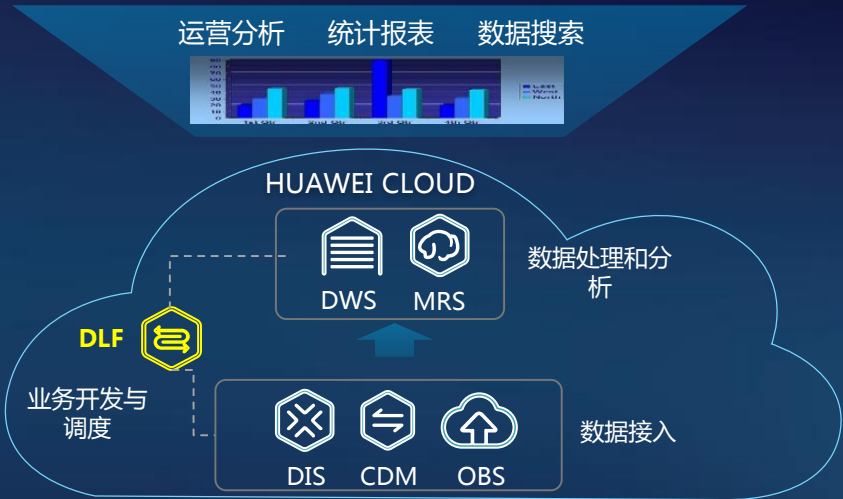
简单易用：在线SQL/Shell脚本编辑调试；预设**10多种**任务类型，拖拽式 workflow 编排



调度稳定高效：丰富的调度配置策略，**百万级别**的作业调度能力

成功案例

数据湖工厂助力某数字建筑平台服务商快速建设云上数据仓库



降低成本：即开即用，免运维，整体建设成本降低70%

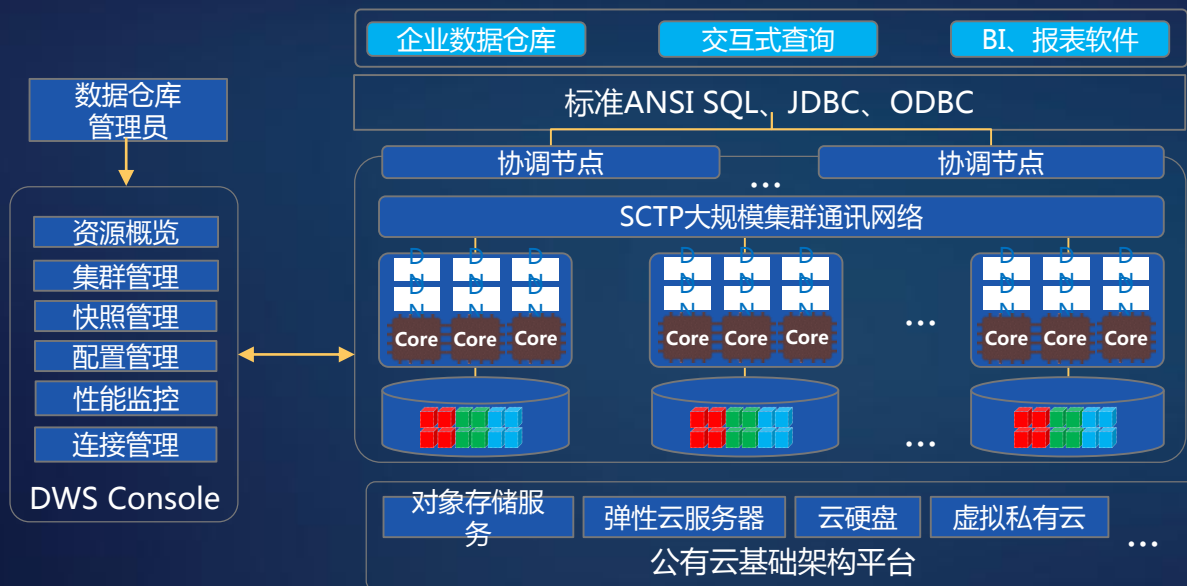
快速上线：一站式开发平台，简单易用，新的统计分析报表开发周期大幅降低，从1~2周缩短至1~2天

高效处理：快速整合多种异构数据源，数据统一管理和分析，节约50%数据处理时间

DWS：企业级数据仓库



DWS (Data Warehouse Service) 是完全托管的企业级数据仓库服务，具备免运维、在线扩展、多源高效的数据加载能力，兼容PostgreSQL生态。助力企业经济高效地对海量数据进行在线分析，实现数据快速变现。



可靠：**全节点HA**，系统无单点故障



安全：获得欧洲**PSA认证**，租户隔离保证数据隐私安全



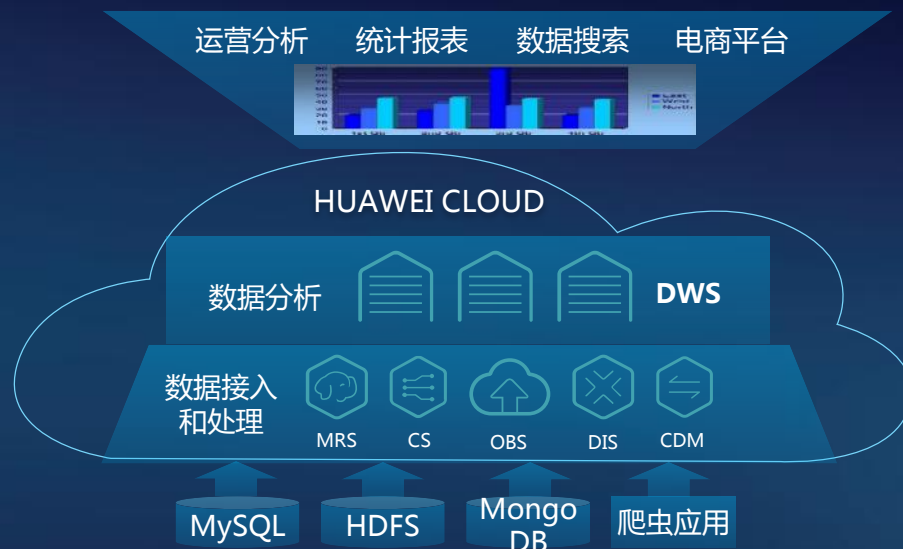
生态丰富：**兼容PG生态**，对接近20种ETL和BI工具，支持多源数据**高效入库**



低成本：按需租用，成本只是传统数仓的**十分之一**

成功案例

助力某建筑领域软件公司打造集团统一数据仓库



统一架构：原有多技术架构简化为DWS，降低数据流转成本和运维复杂度。

极速查询：汇聚全量多维度数据，在线查询分析使得存量数据最大化变现。

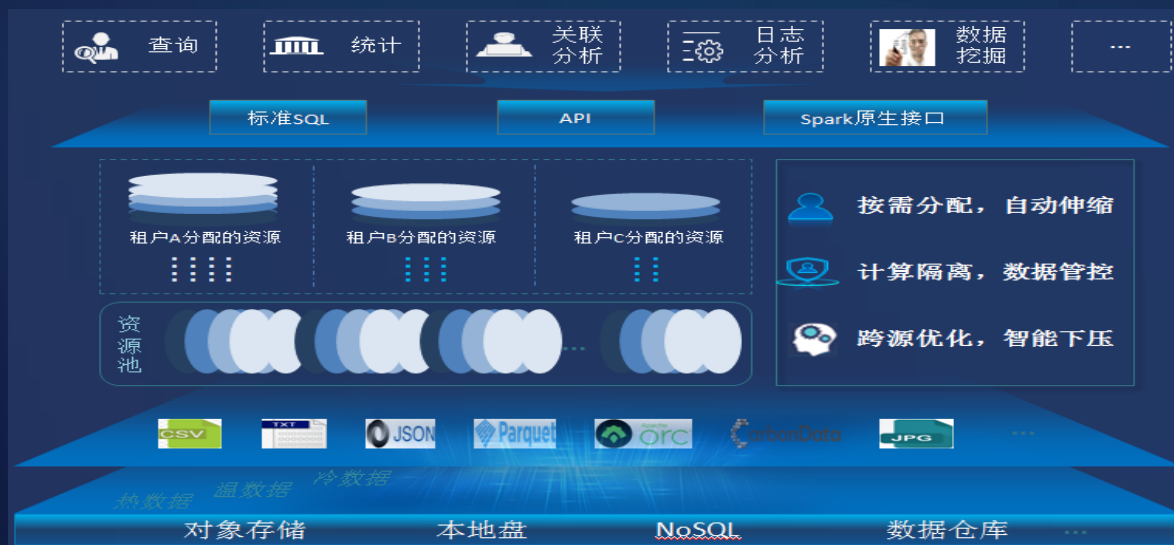
快速上线：兼容SQL标准，原有业务平滑迁移，新业务开发成本降低，上线周期从1~2周缩短至1~2天。

弹性扩展：资源按需弹性扩展，快速响应业务量增长。

DLI：数据湖时代的挖矿机



完全托管的数据湖分析服务，用户无需管理任何服务器，即开即用；服务支持标准SQL，完全兼容Spark生态接口，提供云上多态数据的统一分析能力

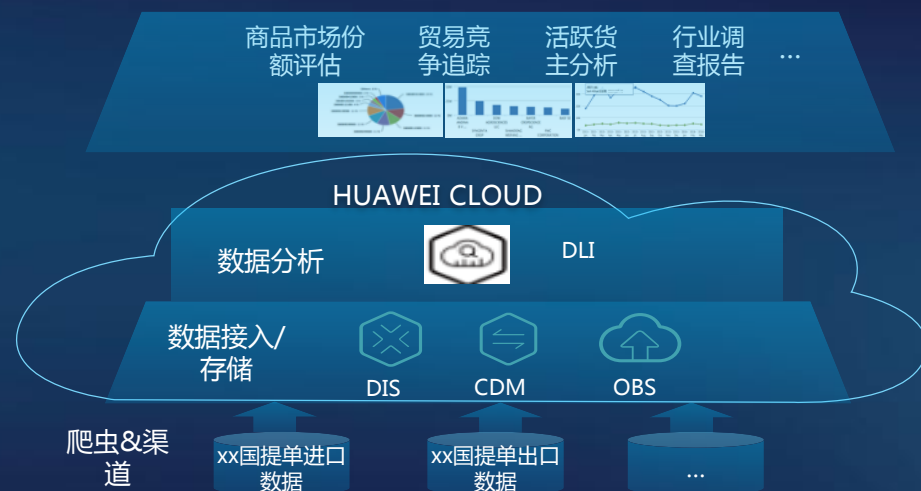


- 数据“0”搬迁：**业内首创跨源联邦分析**，无需额外的数据搬迁
- 数据易分享：**表/列/视图级**数据权限控制与授权，实现企业各部门或子公司间数据共享和变现
- 按需付费：扫描**1G数据3分钱**，TB~EB数据量任意扩展，无需自建集群
- 生态开放：业务**迁移零改动**，完全兼容SQL，Streaming，R，Python，ML等Spark接口

成功案例

助力瀚闻资讯高效大数据分析，实现业务快速创新

瀚讯是**国内首家**收集世界货物贸易统计及全球买家数据的商业机构，经过大数据分析，形成资讯报告，供会员查询或下载，平均**3万次访问/天**



建设成本降低70%：线下自建集群，每年耗费60万，云上零运维，即来即用

2小时全部迁移上云：.net开发的portal界面，通过ODBC无缝对接，已有3000多张表，300多impala SQL代码，平滑迁移

新业务上线周期缩短60%：原有平台人力80%投入运维，100T存量数据只有4%产生了价值

MLS：一站式数据挖掘分析平台

数据挖掘分析平台服务，帮助用户通过机器学习技术发现已有数据中的规律，从而创建机器学习模型，并基于机器学习模型处理新的数据，为业务应用生成预测结果



- 易用：通过**可视化的拖拽式** workflow，实现数据建模、分析、预测、可视化
- 丰富：**预置丰富的**机器学习算法，满足从数据导入和处理，到模型训练和评估、导出，覆盖预测分析**端到端**业务
- 开放：**交互式** Notebook，支持多种开源机器学习库（Python等）
- 一站式：提供特征工程、机器学习算法、建模、预测、模型全生命周期管理的**机器学习一站式**机器学习应用



典型案例

预测性维护



设备工作状态预测：如设备是否会发生故障、设备剩余寿命

推荐



广告推荐：根据客户喜好提供个性化推荐

客户挽留



分析客户属性及行为，预测客户流失率，指导企业制定挽留方案

零售商分群



根据零售商的进出货数据对零售商进行分类，进行针对性的管理

异常轨迹分析



分析车联网车辆行驶数据，发掘车辆异常轨迹，减少企业损失。

话题发现



舆情监控，从大量文本数据中识别热点话题，了解当前舆论趋势

驾驶行为分析



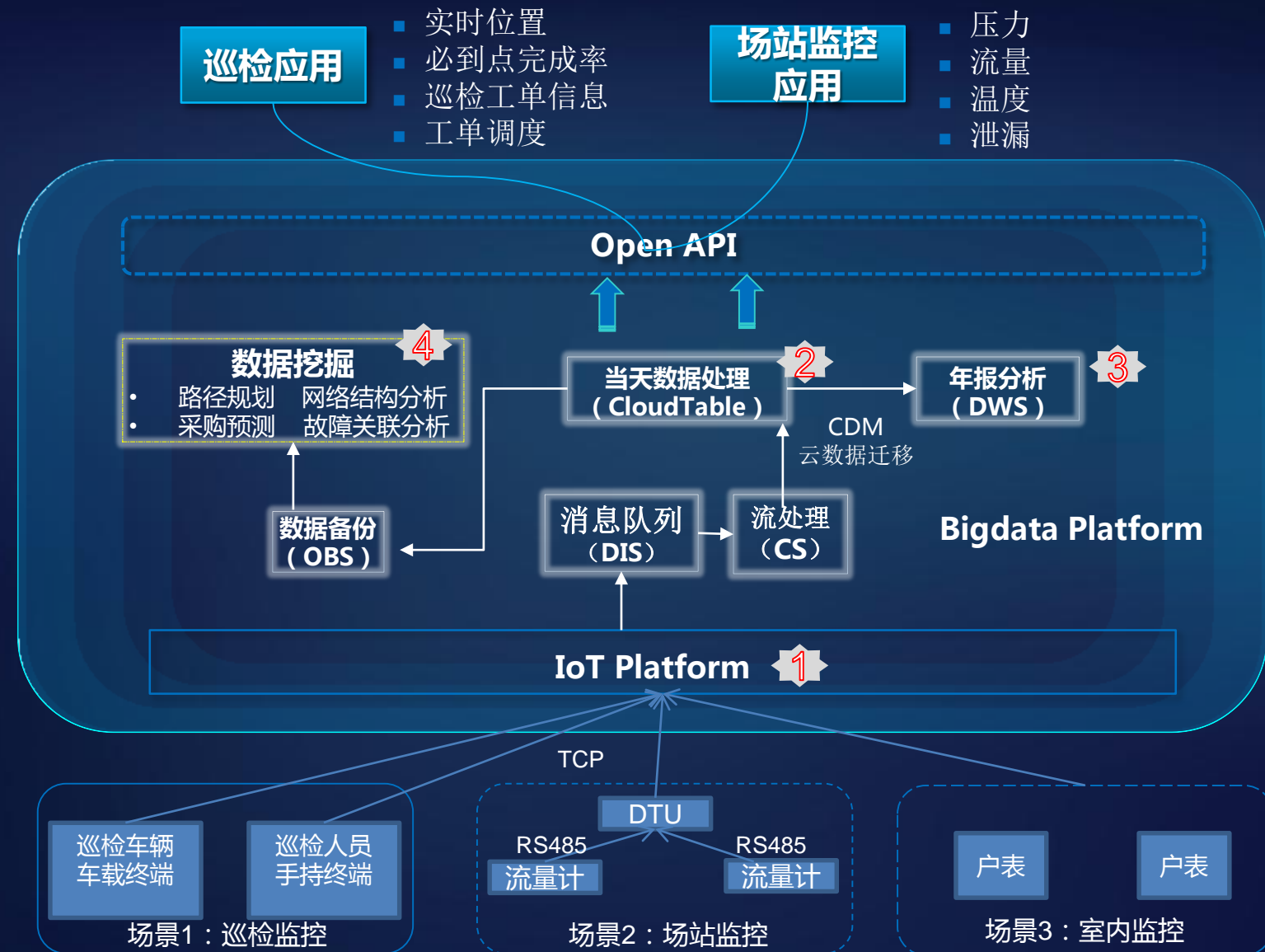
分析车辆行驶数据，识别用户驾驶习惯，保险创新带来二次收益

文本分类



新闻网站中文本的自动分类，如体育、娱乐、政治等

应用方案一：IoT实时监控大数据方案



业务特点

多数据源：车辆、手机上报信息，智能气表信息

数据分析：OLAP分析，统计分析，多维分析，数据挖掘，自助分析，关联查询，拉链表更新

海量数据：每个终端一秒生成一条记录，30秒上报一次；全年预计超过12TB。

高性能：TB数据关联分析秒级响应，在线交互式查询，支持快速业务决策

生态兼容：标准SQL、完备的事务处理能力，存储过程、兼容PostgreSQL生态；对接第三方BI工具

应用方案二：车联网平台解决方案



企业应用

驾驶司机	运输企业	租车公司	整车厂	4S店	金融保险
<ul style="list-style-type: none">实时车况监控环保驾驶比拼加油站、充电桩流量引导车辆维保提醒引导	<ul style="list-style-type: none">实时车况监控油耗成本控制行驶路线监控及优化物流优化	<ul style="list-style-type: none">车辆定位车辆实时报警监控车辆违停、越界报警监控	<ul style="list-style-type: none">新车目标人群定位新车测试采集产品设计优化	<ul style="list-style-type: none">客户维系新车推荐试乘试驾监控新车重购	<ul style="list-style-type: none">车辆保险UBI车辆贷款二手车估值

ELB、API Management

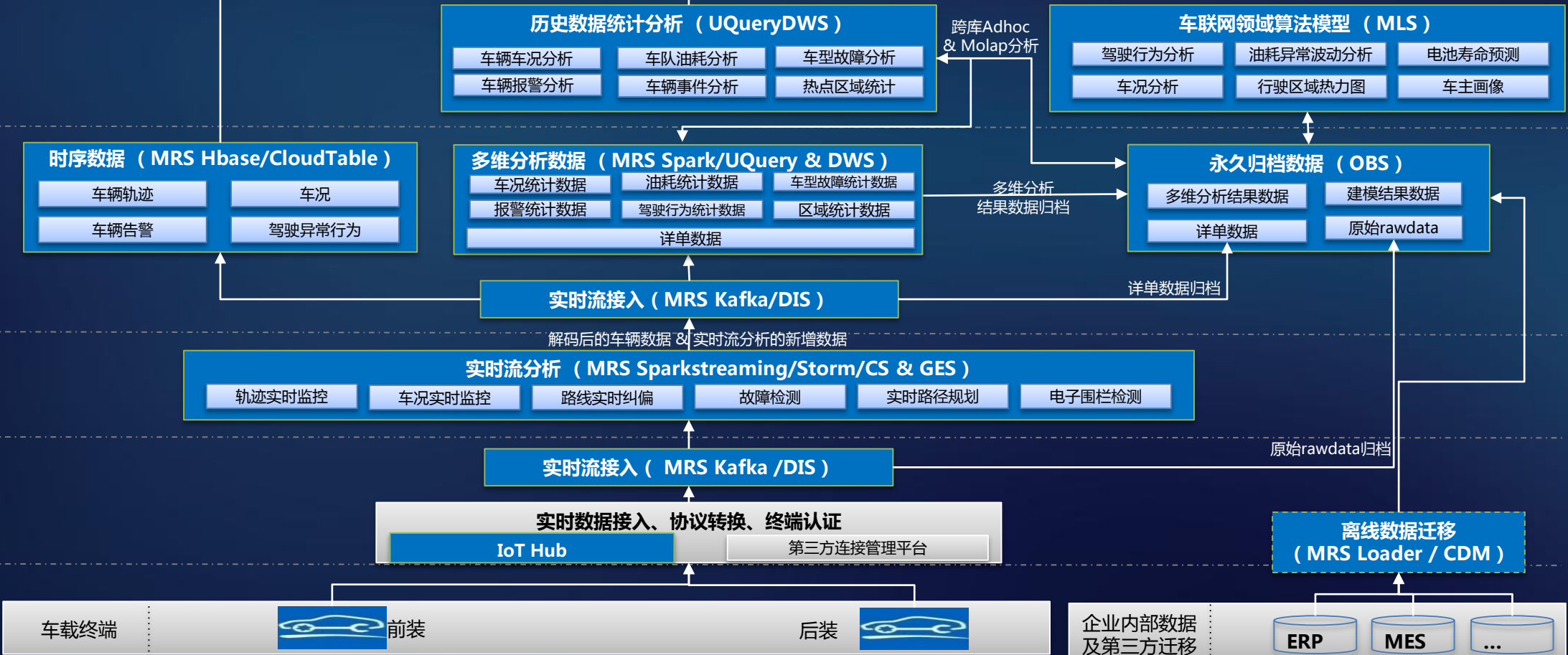
批量分析

数据存储

实时分析

数据集成

数据源



华为云EI：联接智慧，创造未来



THANK YOU

