

数据运营系统实现方案浅析

主讲人：纪应忠 阿里云专家

目录

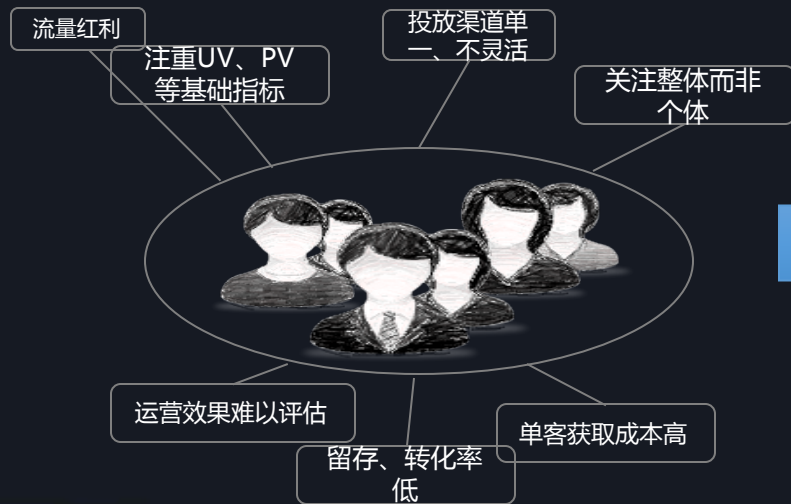
一、行业背景

二、阿里云数据运营解决方案

三、典型应用场景&案例



流量为王时代终结，数据化运营对业务发展变得日益重要



基于数据的精细化运营



80%的流量，只创造了20%的价值

20%的流量，创造80%的价值

精细化数据运营的目的不是为了省钱，而是为了更有效率的花钱！

按行业细分，数据运营需求及实现方式也各不相同

行业细分

业务需求场景

数据运营实现方式

游戏研发商

基础指标集

- DUV/MAU
- Retention
- ARPU/...

运营分析

- OLAP分析
- 报表、专题

游戏发行商

用户画像

- 游戏偏好
- 行为偏好
- 消费偏好/...

游戏渠道商

渠道分析

- 渠道用户质量分
- 拆
- ROI分析

智能算法

- 用户健康度分析
- 流失预测模型
- 用户画像/...

无数据运营

数据分析需求驱动

关系型数据库

接入成本低
通用性

存储容量限制
查询效率限制

降低技术门槛
缩短平台建设周期

借助三方平台

分析规则单一
无数据掌控权

开源方案自建

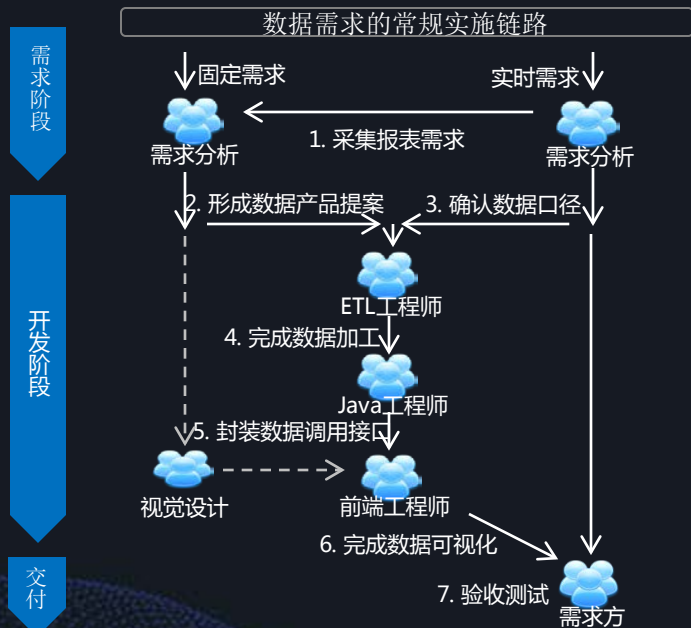
运维压力
集群扩容
数据源单一
智能化需求

阿里云平台

不同类型公司对数据化运营的业务需求各有侧重，实现的方式也多种多样



运营从粗放→精细化，高效、合理的数据架构是关键



技术平台复杂，运维成本高

数据→业务价值转化效率低

应用场景单调，需求响应速度慢

传统按需加工、被动响应的数据处理架构，不能支撑数据化运营的分析需求

目录

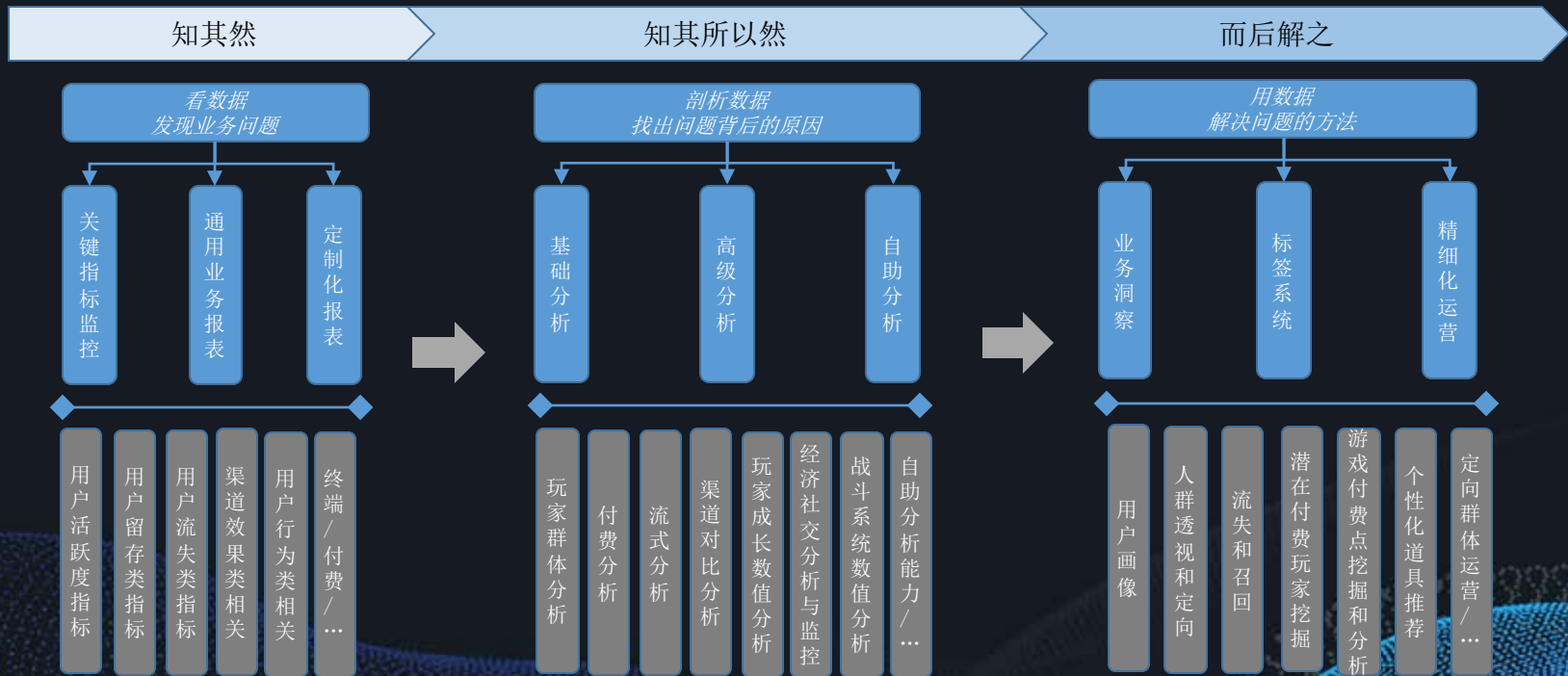
一、行业背景

二、阿里云数据运营解决方案

三、典型应用场景&案例



一个数据化运营系统需要解决的几个核心业务问题

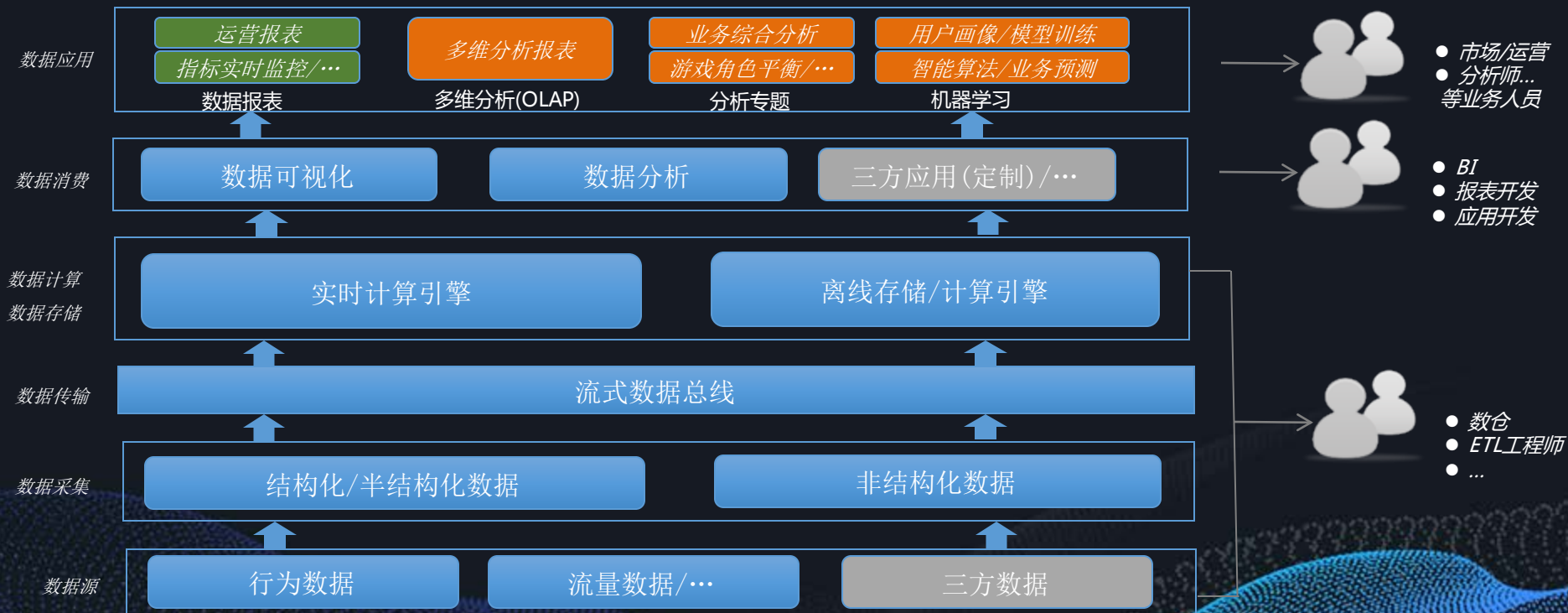




2017云栖大会·成都峰会
THE COMPUTING CONFERENCE



阿里云大数据工具，搭建一站式数据运营支撑平台





2017云栖大会·成都峰会
THE COMPUTING CONFERENCE



阿里云大数据工具，搭建一站式数据运营支撑平台

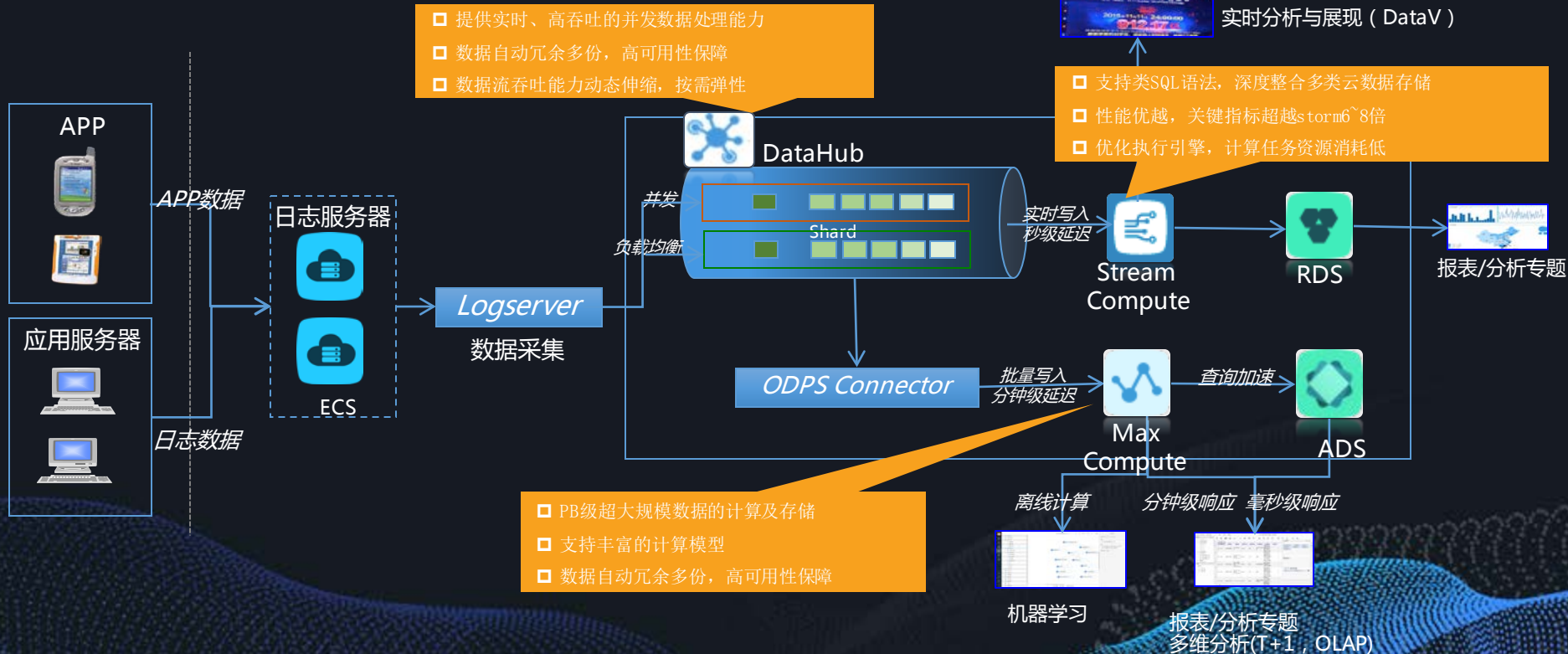




2017云栖大会·成都峰会
THE COMPUTING CONFERENCE

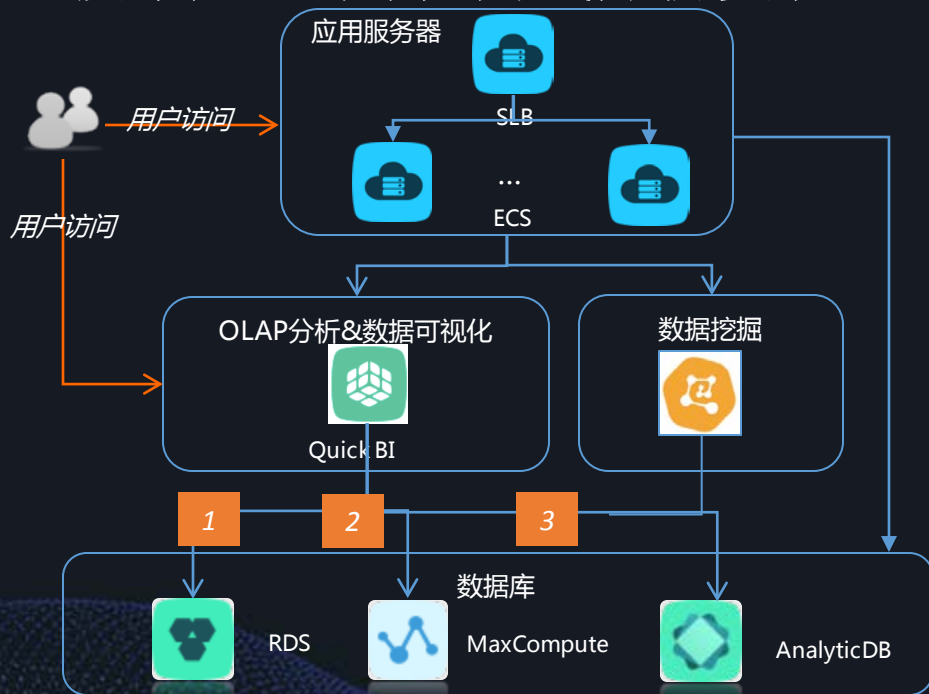


数据运营平台基础数据框架





技术产品架构-离线分析实现

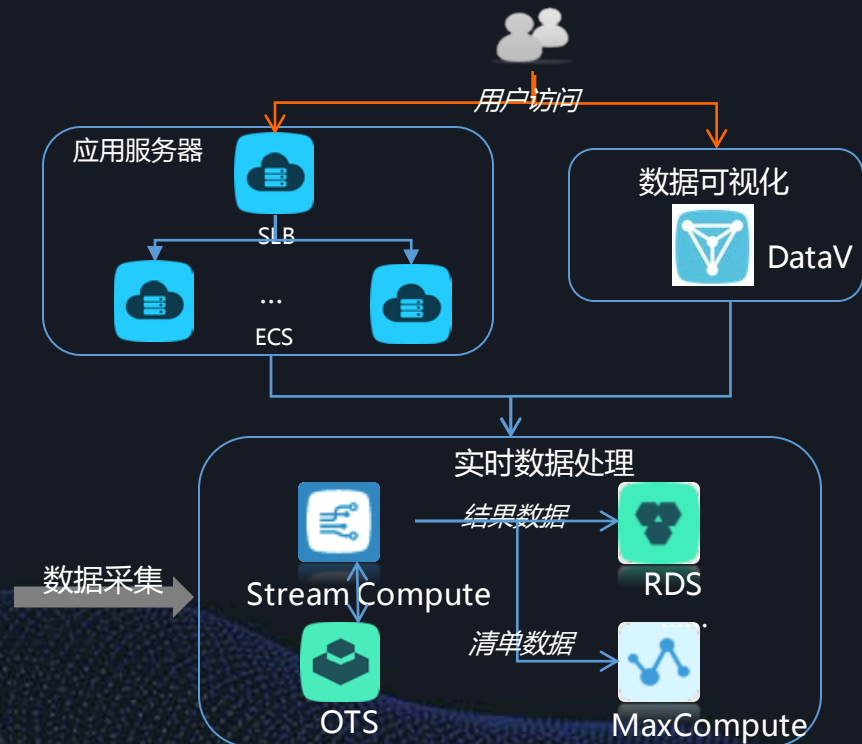


架构重点简介:

- ❑ **数据库:** 按应用场景、数据规模选用合适的计算/存储引擎;
- ❑ **OLAP分析&数据可视化:** 拖拽式、所见即多得的OLAP分析、报表配置能力;
- ❑ **机器学习:** 可视化方式的算法配置平台, 快速实现模型训练、智能算法等高复杂度的数据应用场景;



技术产品架构-实时数据分析实现



架构重点简介:

- 实时数据处理：数据总线（DataHub）+流计算平台（Stream Compute），构建高可用、高并发、可弹性扩容的实时数据处理模块；
- 实时数据分析：Data V平台提供的多种可视化控件，可以快速实现大屏类的实时分析场景

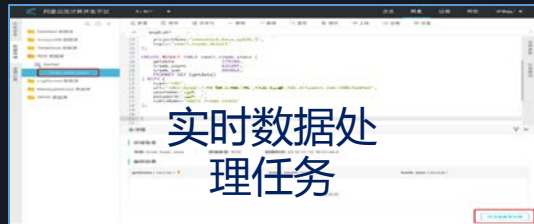


2017云栖大会·成都峰会
THE COMPUTING CONFERENCE

阿里云



大数据存储/计算平台，满足不同场景的数据应用要求



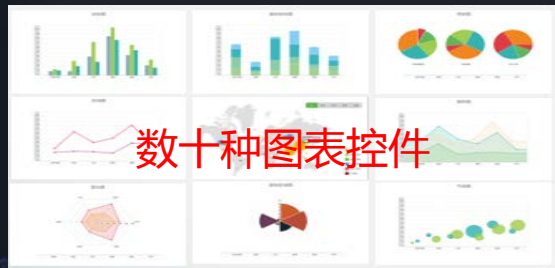
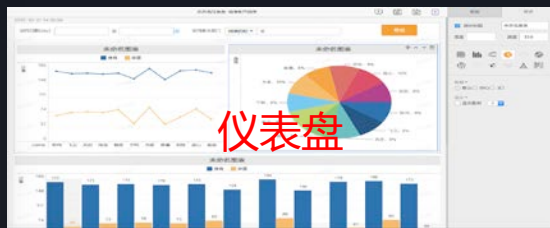
- 可视化开发界面、支持多种代码类型
- 离线调度、超百万级任务管理
- 在线运维、监控报警
- 快速数据集成
- 涵盖全链路数据血缘、业务元数据分析
- 数据生命周期、数据资产、数据权限等管理
- 多人协同工作



2017云栖大会·成都峰会
THE COMPUTING CONFERENCE



海量数据实时在线分析，拖拽式操作、丰富的可视化效果



- 海量数据实时分析
- 多维数据分析
- 智能加速引擎
- 丰富页面控件支撑、可自定义扩展
- 拖拽式操作，所见即所得
- 智能图表，比你更懂数据
- 灵活的报表集成方案



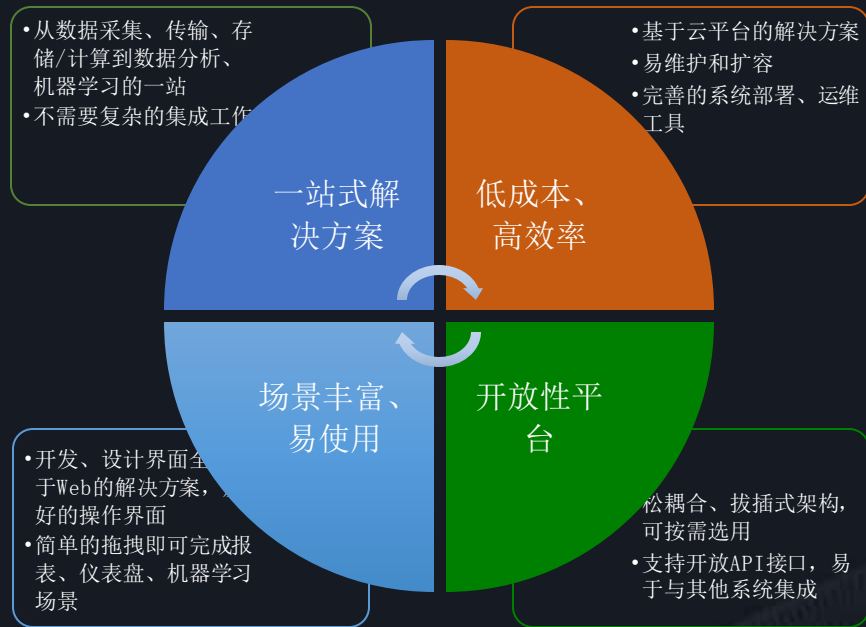
□可视化算法配置页面，通过拖、拉、拽的方式来操作实验

提供数据预处理、特征工程、机器学习算法、预测和评估、在线预测等内置模板

平台提供逻辑回归、随机森林、GBDT、K means、文本分析、图计算等100余种常规机器学习算法



方案优点总结



通过平台/工具的支撑，降低数据使用门槛，赋予人人使用大数据的能力，是实现数据化运营的架构基础！

目录

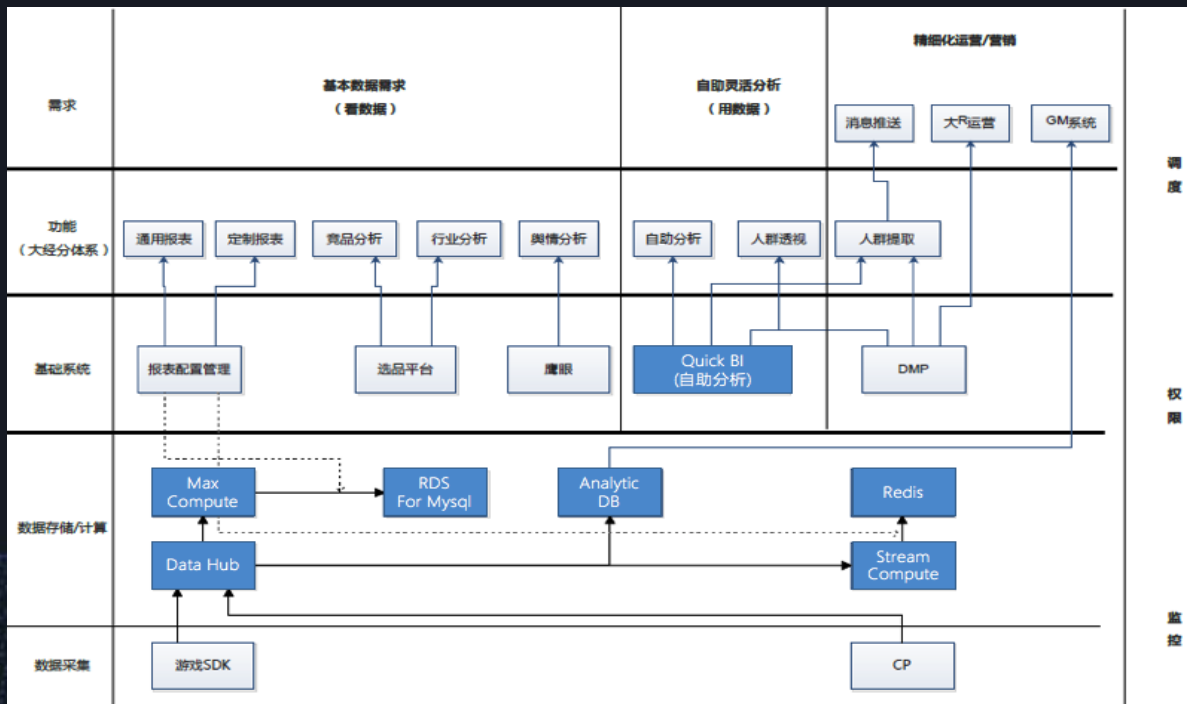
一、行业背景

二、阿里云数据运营解决方案

三、典型应用场景&案例



游戏数据运营系统-业务视图

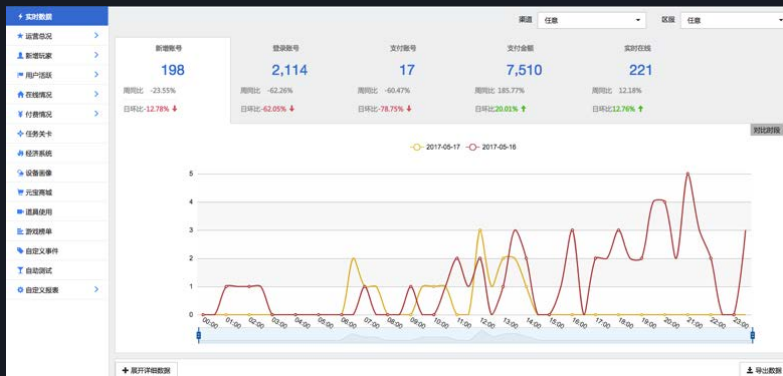


■ 基于Max Compute、RDS、Analytic DB

等构建面向不同应用场景的存储/计算引擎

■ 除了场景化的数据分析之外，还基于

Quick BI提供更灵活的自助式分析能力



基础操作

新增内容

配置管理

数据管理

基础信息

* 搜索名称

搜索名称

* 搜索范围

* 搜索范围

cube_sls_games_event_day

* 时间范围

日 4

* 搜索类型

自定义搜索

* 游戏

王侯将相

* 选择模板

选择模板

* 描述

描述

配置设置

模板code	名称	展示方式	操作
yndf	日期	日期	5
gid	游戏	多选题	5
os	平台	多选题	0
ch	渠道	多选题	5
server	区域	多选题	5

新增模板

编辑模板

删除模板

[illegible]



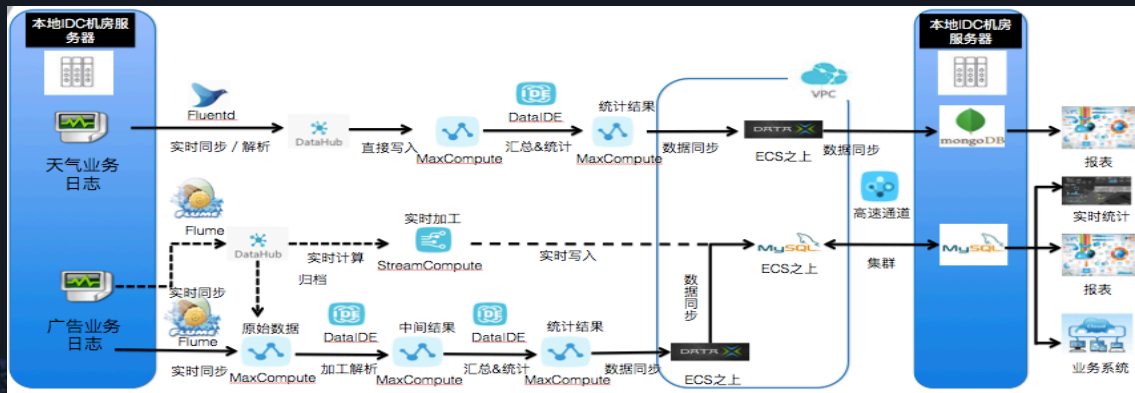
2017云栖大会·成都峰会
THE COMPUTING CONFERENCE



案例：墨迹天气 大数据平台

墨迹天气 成立到现在6年多，已经积累了5亿用户，每天有超过5亿次的天气查询请求和将近20亿次的广告请求，API每天产生的日志量大约在2TB左右。

运营团队需要将这些海量的数据导入云端，然后分天、分小时的展开数据分析作业，分析结果再导入数据库和报表系统，最终展示在运营人员面前。整个过程中数据量庞大，且计算复杂，这对云平台的大数据能力、生态完整性和开放性提出了很高的要求。



注：虚线部分功能在建设



Max Computer



高速通道



Stream Computer

value

5倍以上效率提升

节省70%存储计算

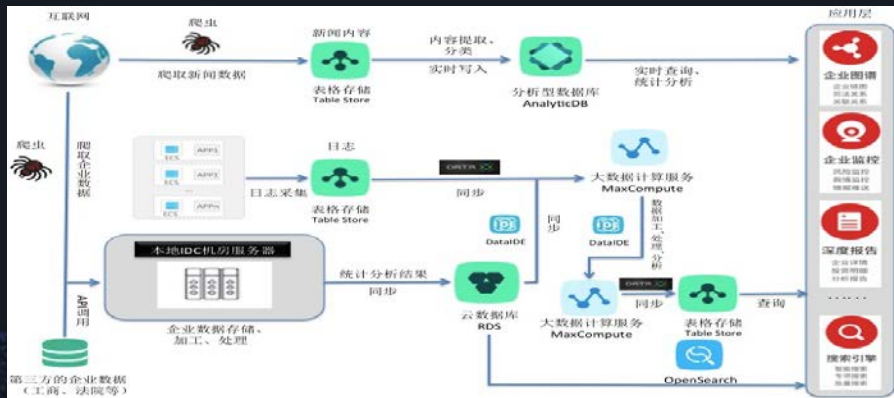
洞察力提升

业务优化

案例：企信宝 大数据平台

启信宝作为一家以大数据为核心竞争力的公司，借助大数据为客户降低执业风险、经营风险、投资风险和借贷风险，为客户提供数据解决方案以进行精准营销、拓客展业，为客户借贷管理、投资理财提供决策依据。

启信宝的企业数据范围广、维度多，覆盖8000万以上的企业，19个产业链，95个细分行业，100个以上企业数据维度，企业覆盖率达98%以上，数据融合复杂度高，任务处理量巨大、且对数据变化的实时性要求高（内容多维查询和统计分析）。



- 主要数据来源：互联网采集的数据以及API调用的第三方的企业数据；
- 数据处理分析：应用数据采集到阿里云存储容器，通过数加平台进行二次加工处理，最终同步到阿里云存储容器中进行数据展现分析
- 应用层：支撑企业图谱、企业监控、深度报告、搜索引擎等不同产品形态。



Max Computer



Analytic DB



Open Search



OTS

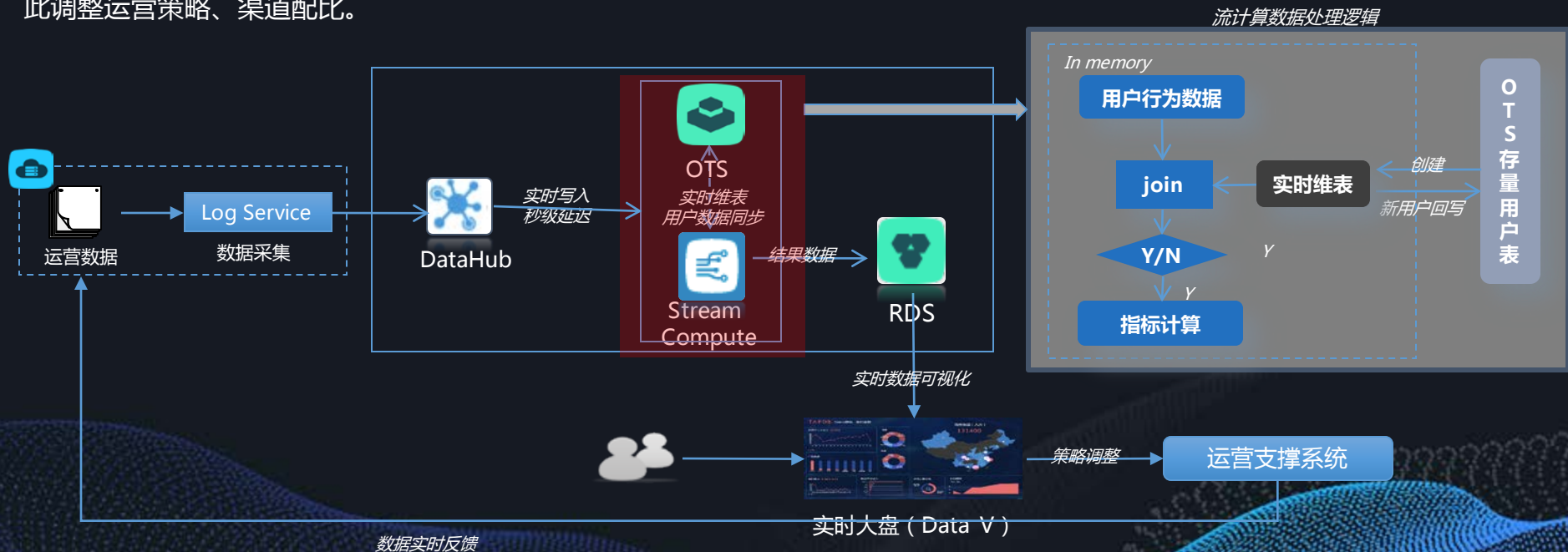


2017云栖大会·成都峰会
THE COMPUTING CONFERENCE



应用场景：运营实时监控大盘

游戏运营推广过程中，业务人员需要实时获取到关键的指标数据——比如，各渠道新增UV、PV、消费转换等，评估效果，并依此调整运营策略、渠道配比。





运营实时监控大盘（脚本示例）

流计算数据处理逻辑及脚本示例：

1、数据输入：创建来自Datahub的流式表

```
channel_id STRING,
platform_id STRING,
cookie_id STRING
) WITH (
  type='datahub',
  endpoint='http://dh-cn-hangzhou-internal.aliyuncs.com',
  roleArn='acs:ram::1251530108570787:role/aliyunstreamdefaultrole',
  projectName='haima',
  topic='haima_cookie_data'
);
```

2、创建维表：来自OTS的存量用户表

```
count_date STRING,
PRIMARY KEY (cookie)
) WITH (
  type='ots',
  endpoint='http://haima-ots-cn-hangzhou.ots-internal.aliyuncs.com',
  roleArn='acs:ram::1251530108570787:role/aliyunstreamdefaultrole',
  instance='haima-ots',
  tableName='dim_cookie'
);
```

3、创建结果表：RDS表，计算后的数据写入

```
platform_id STRING,
pv INT,
uv INT
) WITH (
  type='rds',
  url='jdbc:mysql://rm-uf6mh45h74izqrb39.mysql.rds.aliyuncs.com:3306/test',
  username='rdstest',
  password='Haimawan2016',
  tableName='result_table'
);
```



4、创建中间表

```
tmp_cookie_table (
  channel_id STRING,
  platform_id STRING,
  dt STRING
);
```

5、通过新日志数据与存量用户join，获取新用户，并写入中间表

```
select t1.dt, t1.cookie_id, t1.channel_id, t1.platform_id
from haima_cookie_data t1 left outer join dim_cookie t2
on t1.cookie_id = t2.cookie
where t2.cookie is not null;
```

6、结果数据写入RDS，比对逻辑是否正确

```
t.dt,
t.channel_id,
t.platform_id,
count(1) as pv,
count(distinct(cookie_id)) as uv
from tmp_cookie_table t
group by t.dt, t.channel_id, t.platform_id;
```

join

pv/uv/count_date

insert



2017云栖大会·成都峰会
THE COMPUTING CONFERENCE



更多案例...



全域大数据服务商
服务3000家品牌触达5亿消费者



2017云栖大会·成都峰会
THE COMPUTING CONFERENCE

阿里云

云栖社区

yyq.aliyun.com

飞天·智能

APSARA INTELLIGENCE

2017云栖大会·成都峰会

5月23日 成都世纪城天堂洲际大酒店