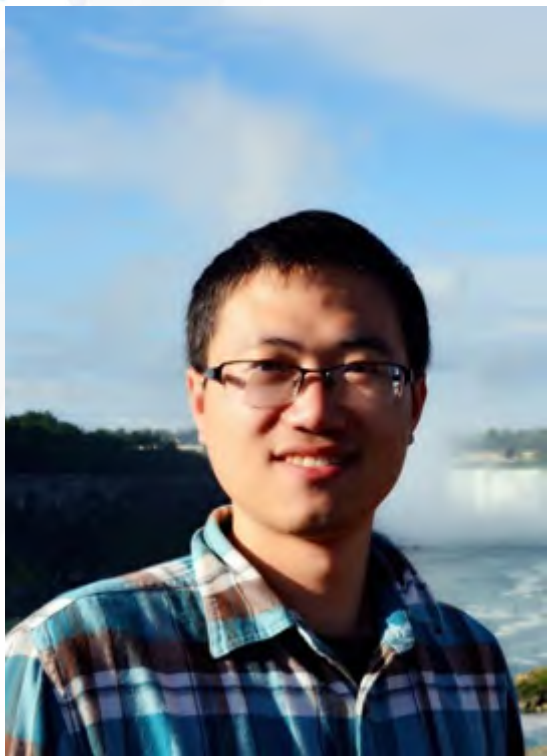


数据平台的实践及思考

搜狗 杨剑飞

关于我



- 硕士研究生
- 10年行业从业经验
- 2013年加入搜狗
- 资深高级开发工程师
- 负责海量数据存储、统计分析 & 治理优化

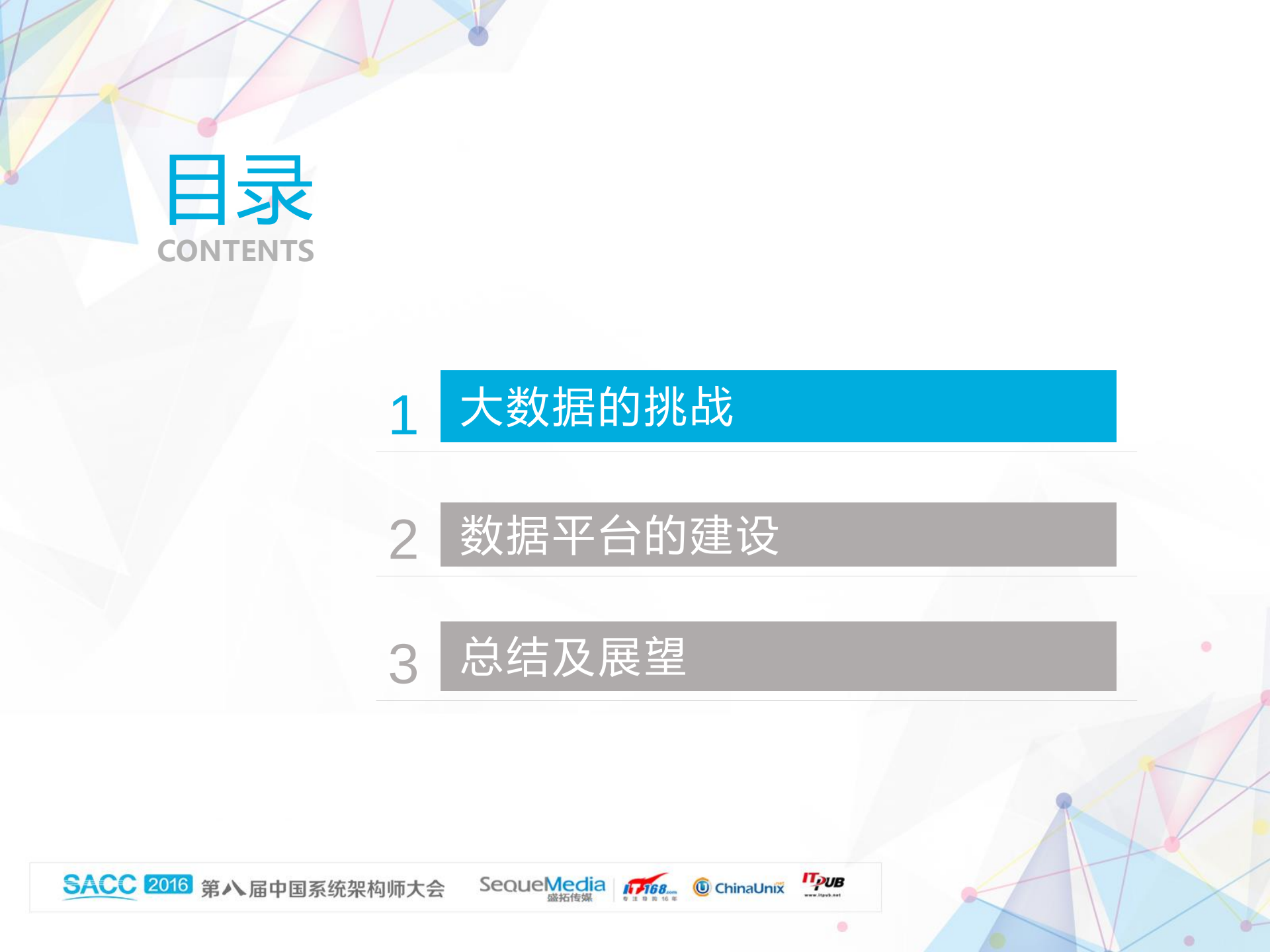
目录

CONTENTS

1 大数据的挑战

2 数据平台的建设

3 总结及展望



目录

CONTENTS

1 大数据的挑战

2 数据平台的建设

3 总结及展望

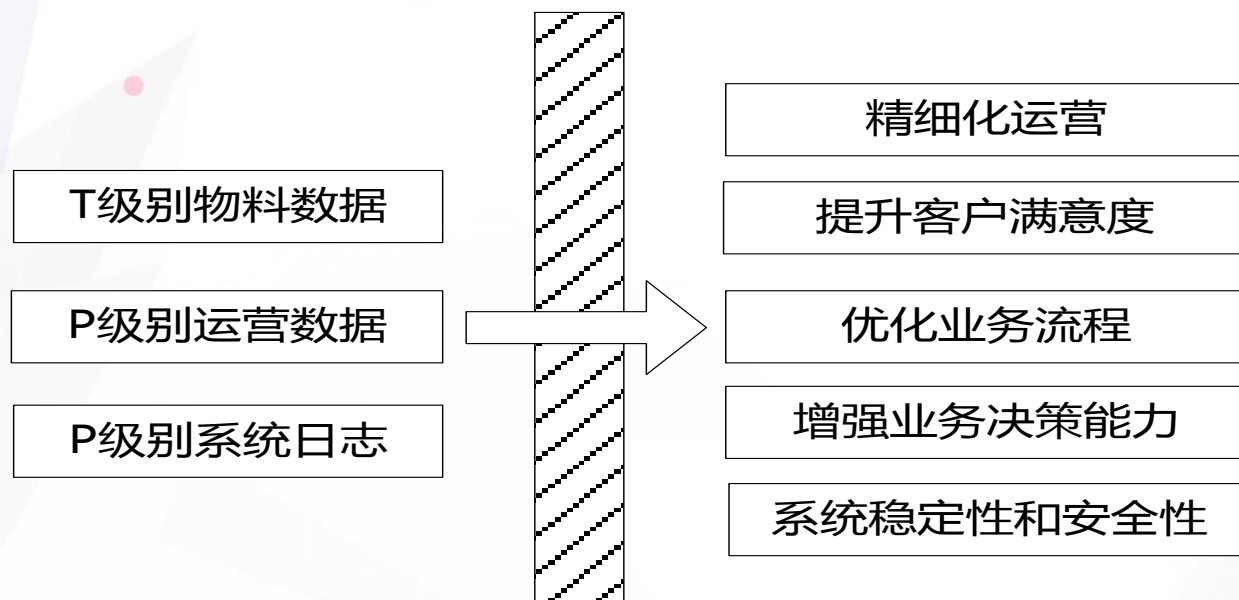
大数据的挑战



大数据带来的信息风暴正在变革我们的生活、工作和思维，大数据开启了一次重大的时代转型

——维克托·迈尔·舍恩伯格

大数据的挑战



■ Volume

PB级别数据量、T级别日增长量、庞大的集群节点

■ Velocity

毫秒级别实时查询、秒级别离线处理、时效性强

■ Variety

用户数据、运营数据、系统日志

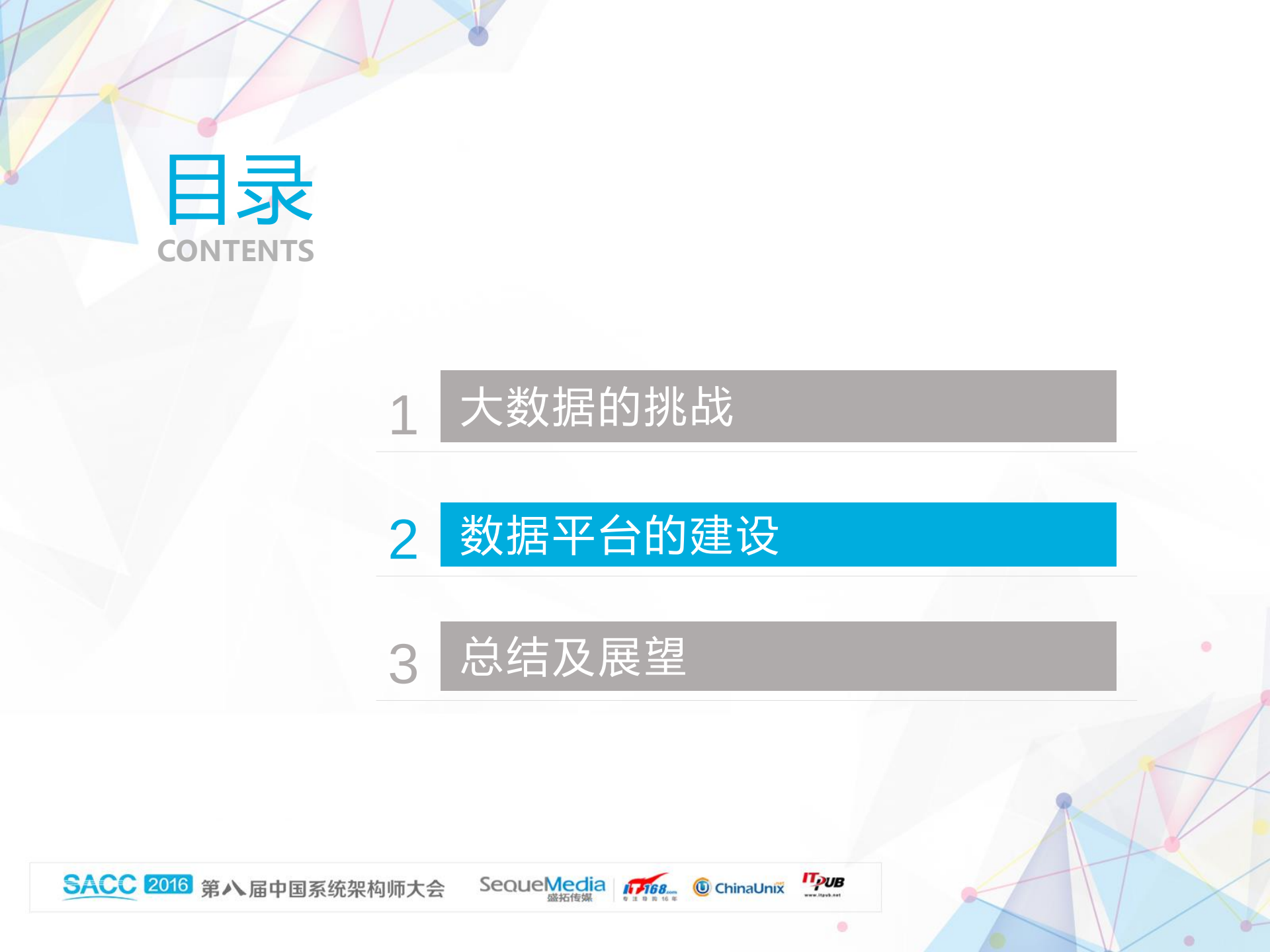
■ Value

精细化运营、精准化决策、规范化运维、提升客户满意度和质量

大数据的挑战

痛点

- 数据孤岛现象严重
- 数据格式参差不齐
- 数据分析时效性弱
- 数据安全无法保证
- 数据使用成本过高



目录

CONTENTS

1 大数据的挑战

2 数据平台的建设

3 总结及展望

平台架构

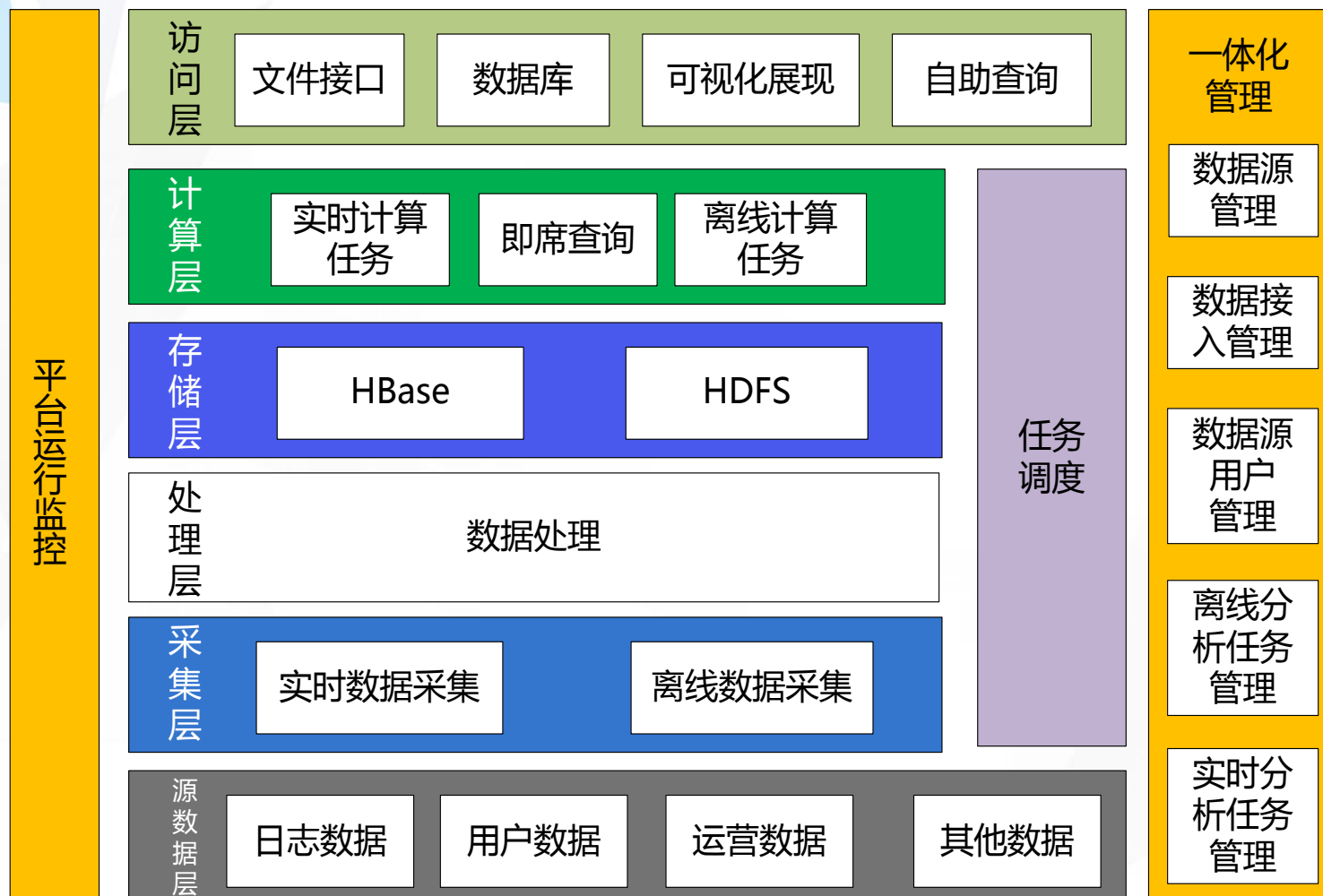
痛点

- 数据孤岛现象严重
- 数据质量参差不齐
- 数据分析时效性不强
- 数据安全无法保证
- 数据使用成本过高

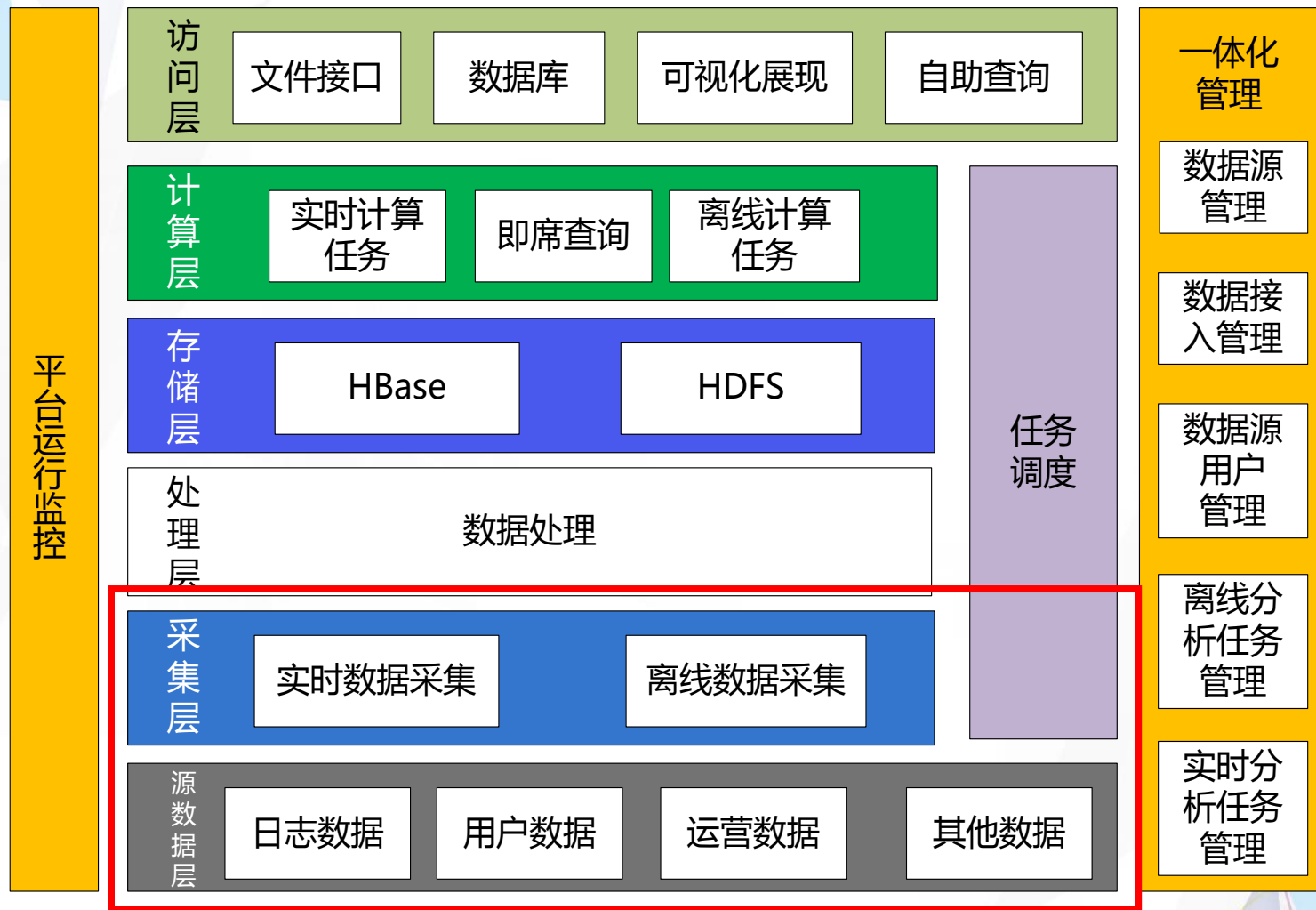
解决思路

- 统一数据接入存储
- 数据校验&格式转换
- 实时分析&离线分析
- 数据治理&权限控制
- OLAP查询

平台架构



数据接入



数据接入

- 面临的挑战
- 实时数据接入
- 离线数据接入
- 数据接入优化

数据接入--面临的挑战

■ 数据源类型多种多样

RDBMS、传统rsync数据源、mongo数据源等

■ 大流量

业务高峰期每小时产生G级别增量数据

■ 拥抱变化

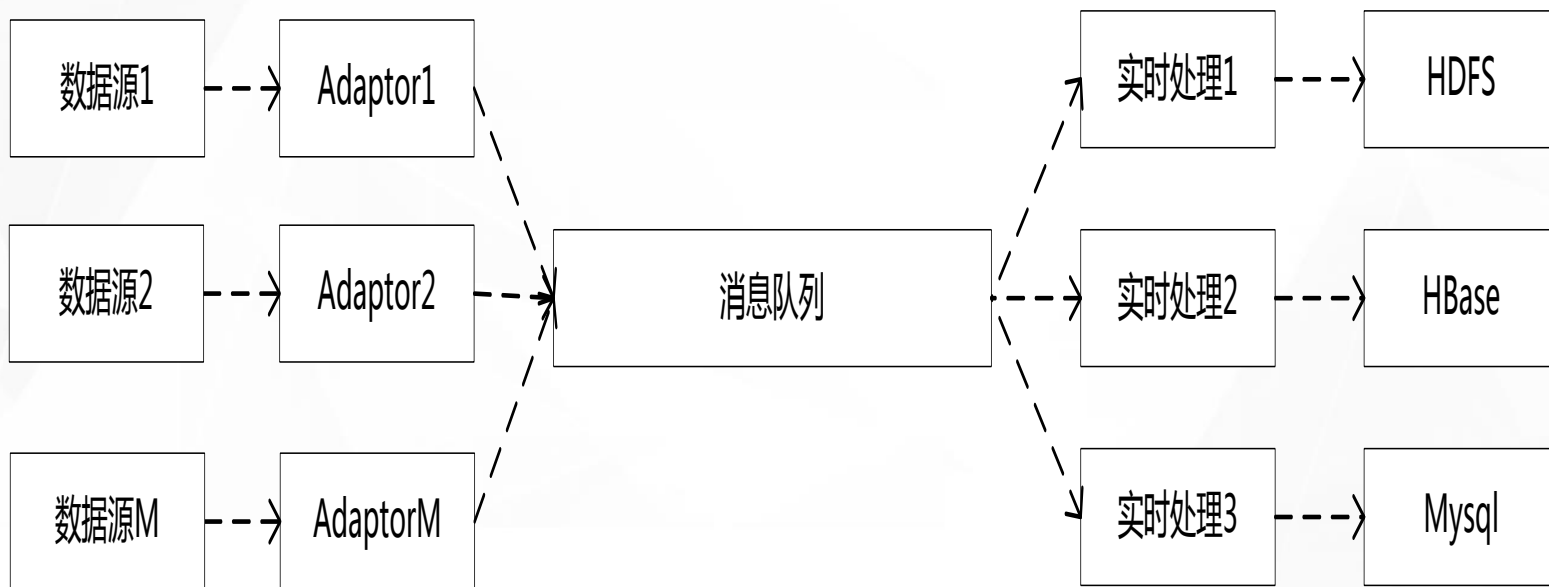
接入数据源越多，数据源变化概率就越大

■ 离线接入及实时接入

业务场景对时效性的要求不同，不同类型数据有不同计算时效性

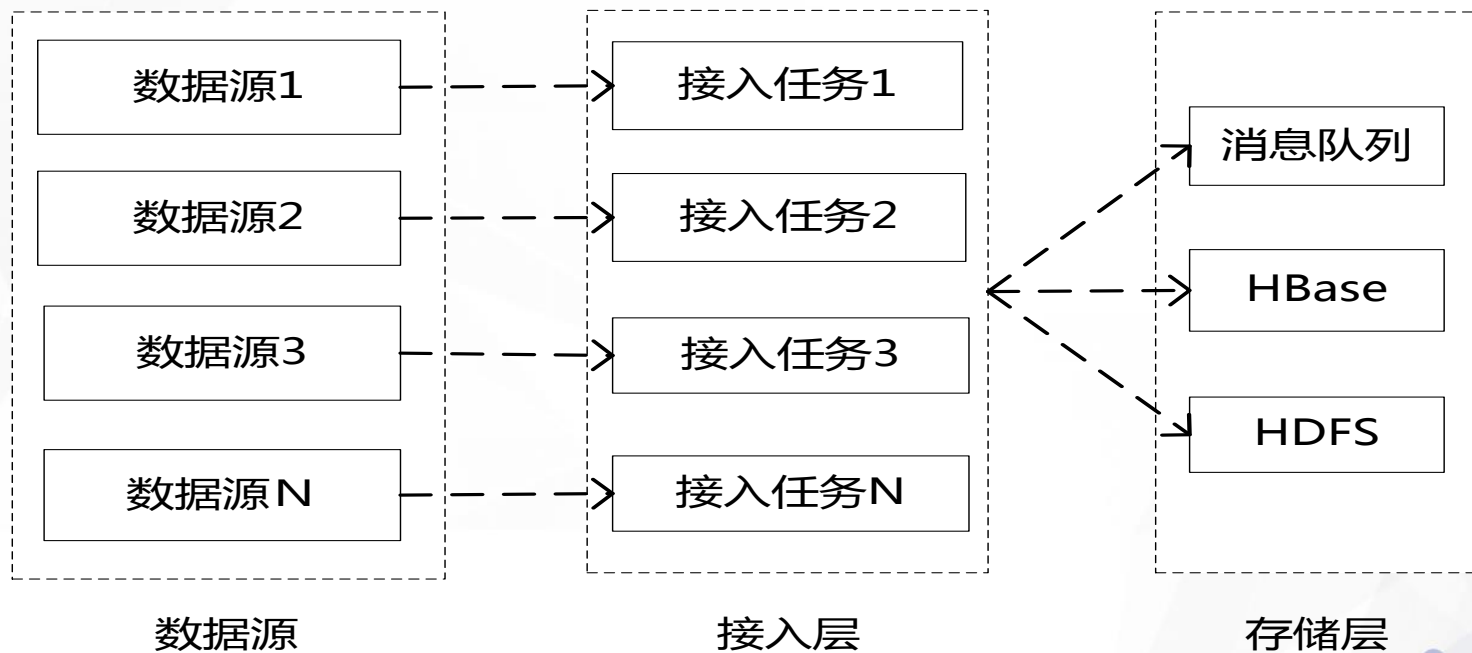
数据接入--实时数据接入

- 采集器实时采集数据源数据，通过消息队列提供给不同消费者
- 针对不同的实时数据源提供不同Adaptor



数据接入--离线数据接入

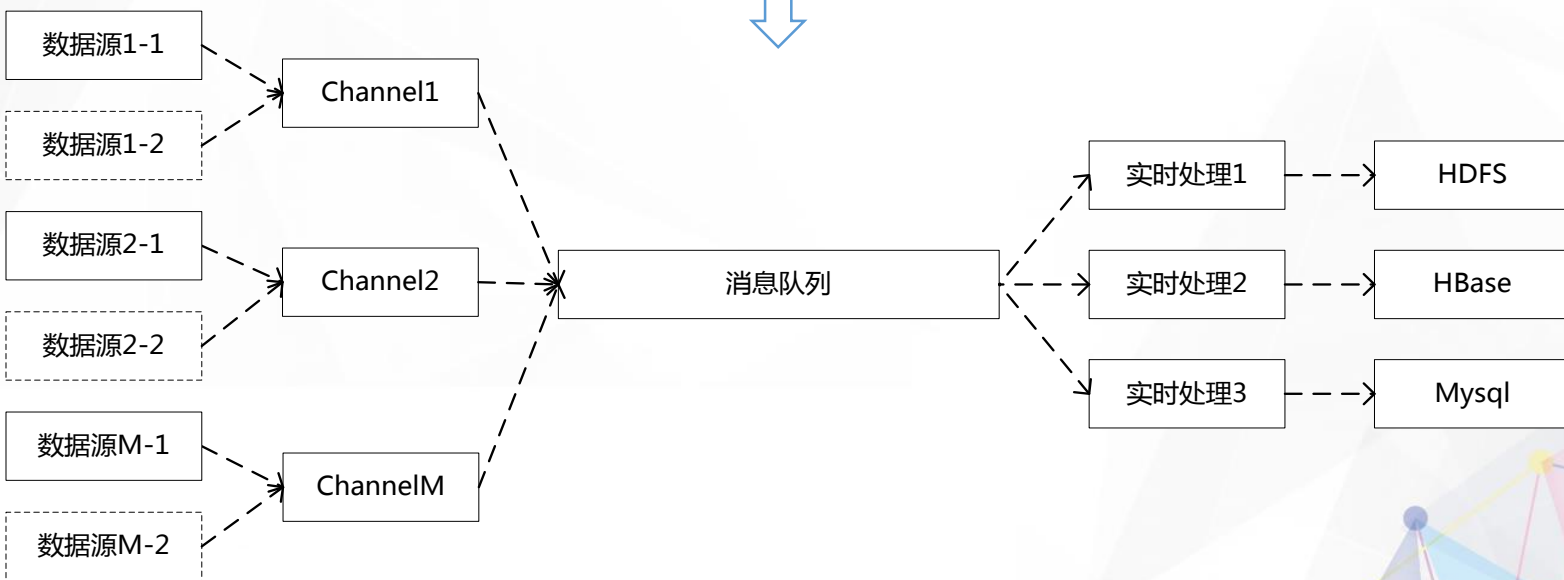
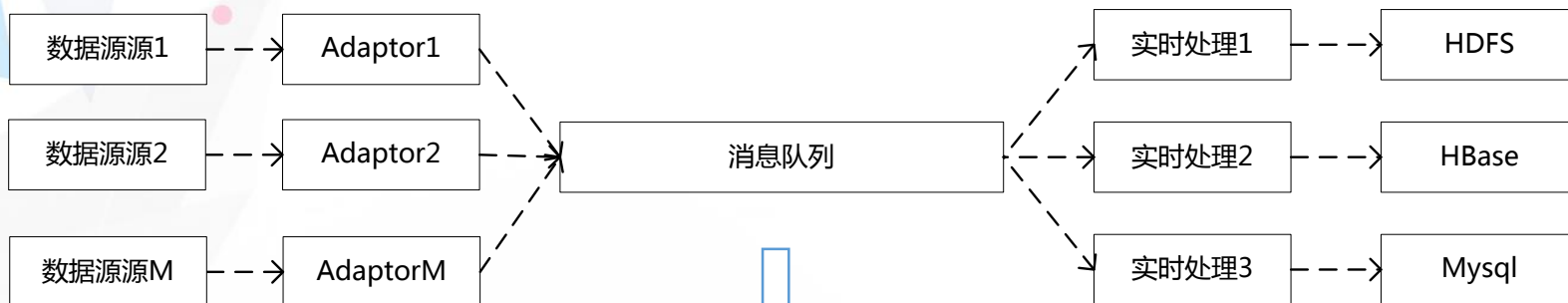
- 大批量且无实时分析场景的数据通过离线接入方式接入平台
- 离线接入通过定时任务调度将数据源中数据汇总到存储层



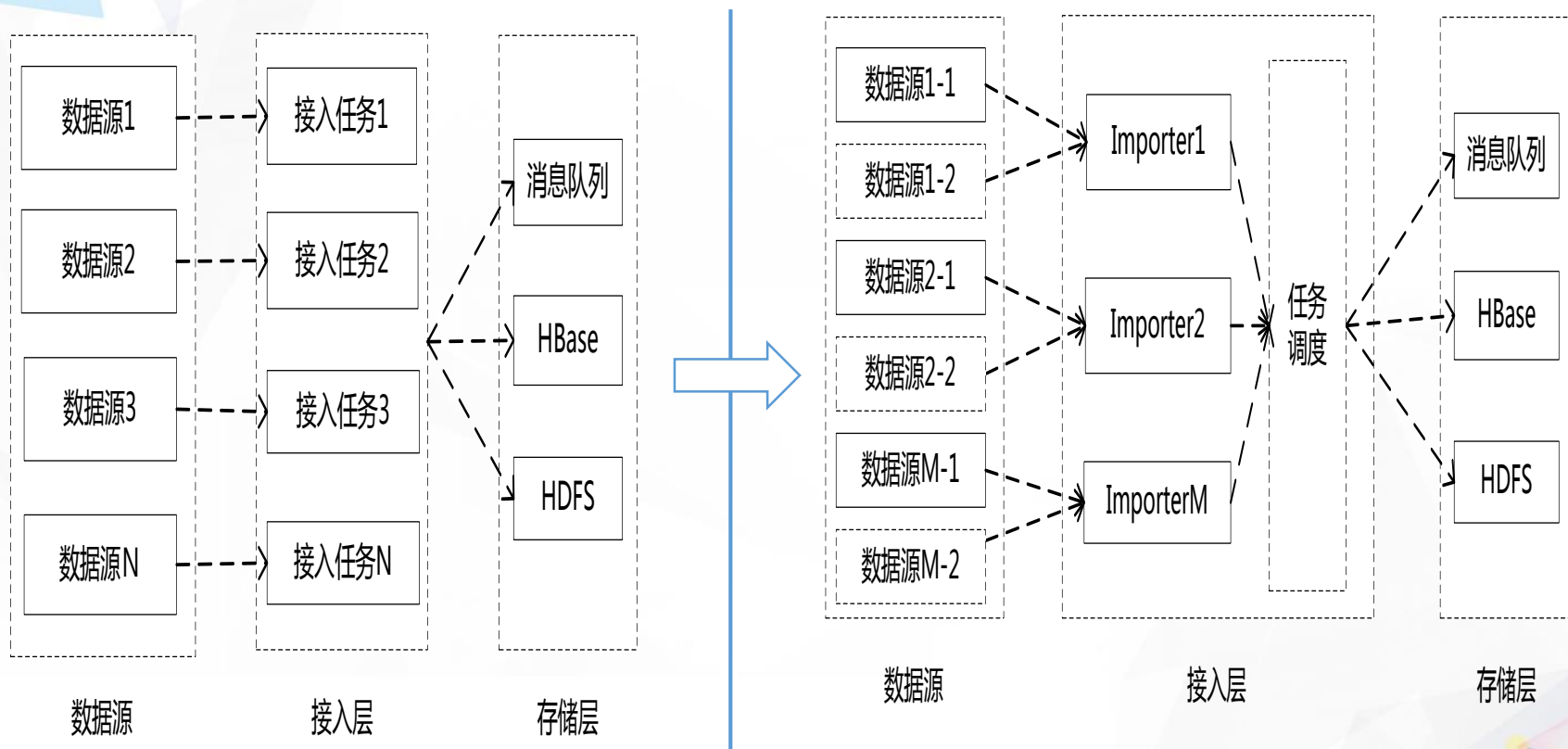
数据接入--数据接入优化

- 数据源及接入任务增加，带来的开发和维护工作量陡增，且效率不高
- 新增数据源需要从采集端到存储端进行定制开发，成本高
- 数据源格式或参数发生变化需要修改一系列相对应的采集器，不易维护
- 为此我们设计出数据接入流程，抽取接入配置模板

数据接入--实时接入优化



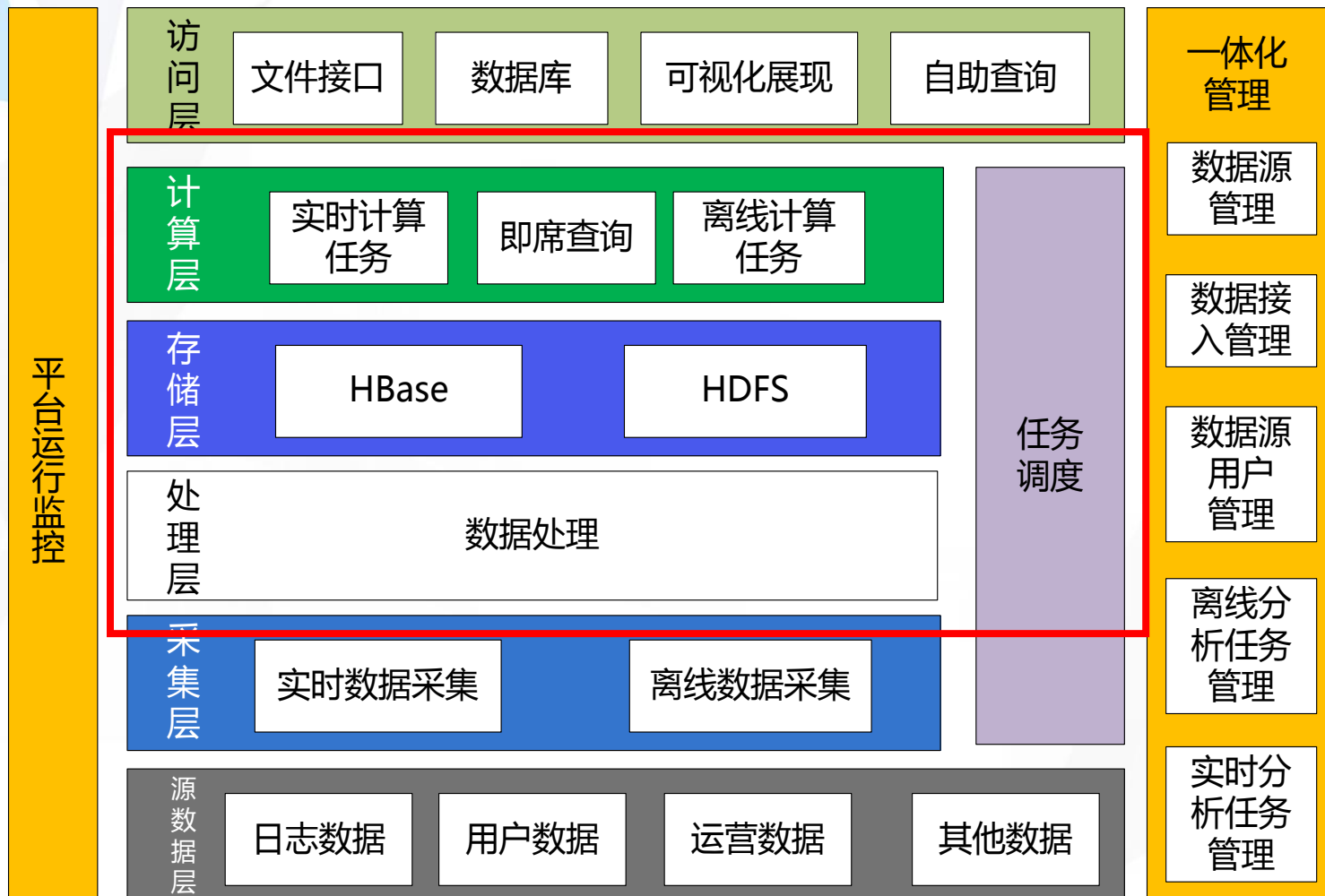
数据接入--离线接入优化



数据接入--优化收益

- 新增接入需求只需通过界面配置相关数据源参数及目标存储参数即可
- 数据源参数发生修改只需要调整数据源参数，无需修改代码，一处修改，所有生效
- 数据源维护交给运维负责，RD专注于业务逻辑开发
- 平台已接入数据一目了然，避免重复接入

存储计算



存储计算

- 面临的挑战
- 数据仓库
- ETL计算
- 任务调度

存储计算--面临的挑战

■ 数据治理

不同用户、不同用途的数据进行隔离，防止多个用户冗余存储，数据安全等

■ 计算任务多

业务快速发展导致计算任务也快速增加，并发任务多造成资源利用不均匀

■ 故障评估与故障恢复

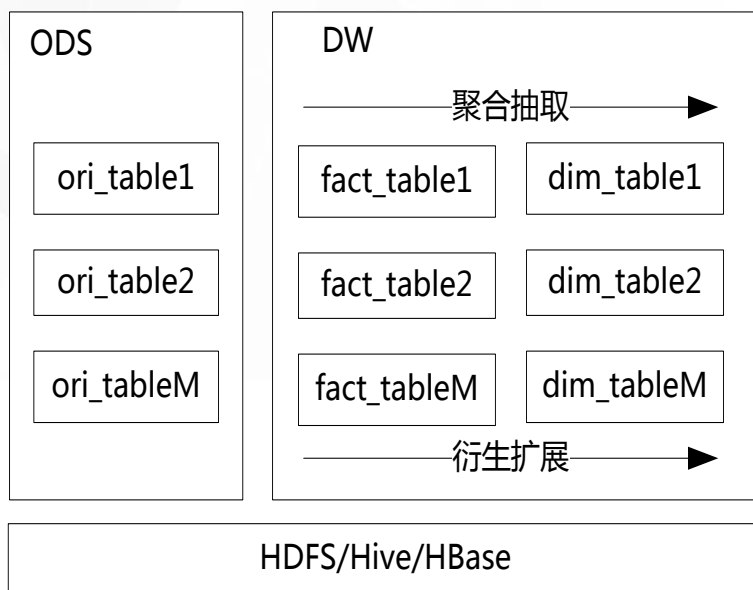
一旦某个数据出现问题，需要快速评估影响面及快速恢复

■ 数据依赖及计算任务依赖

由于计算任务所使用的数据存在依赖关系，导致了计算任务之间存在复杂的依赖关系

存储计算--数据仓库

- 使用数据仓库理念梳理数据
- 不同应用场景采用不同存储方式
- 控制数据读写权限保护数据安全
- 提供多种元数据查询界面



使用场景	存储策略
离线统计	Hive分区表
详情查询	HBase表
文件存储	HDFS存储

存储计算--数据仓库

数据平台

杨剑飞

数据源管理

数据查询

Storm管理

用户管理

Hive数据源

基本信息

Hive属性

字段信息

样本数据

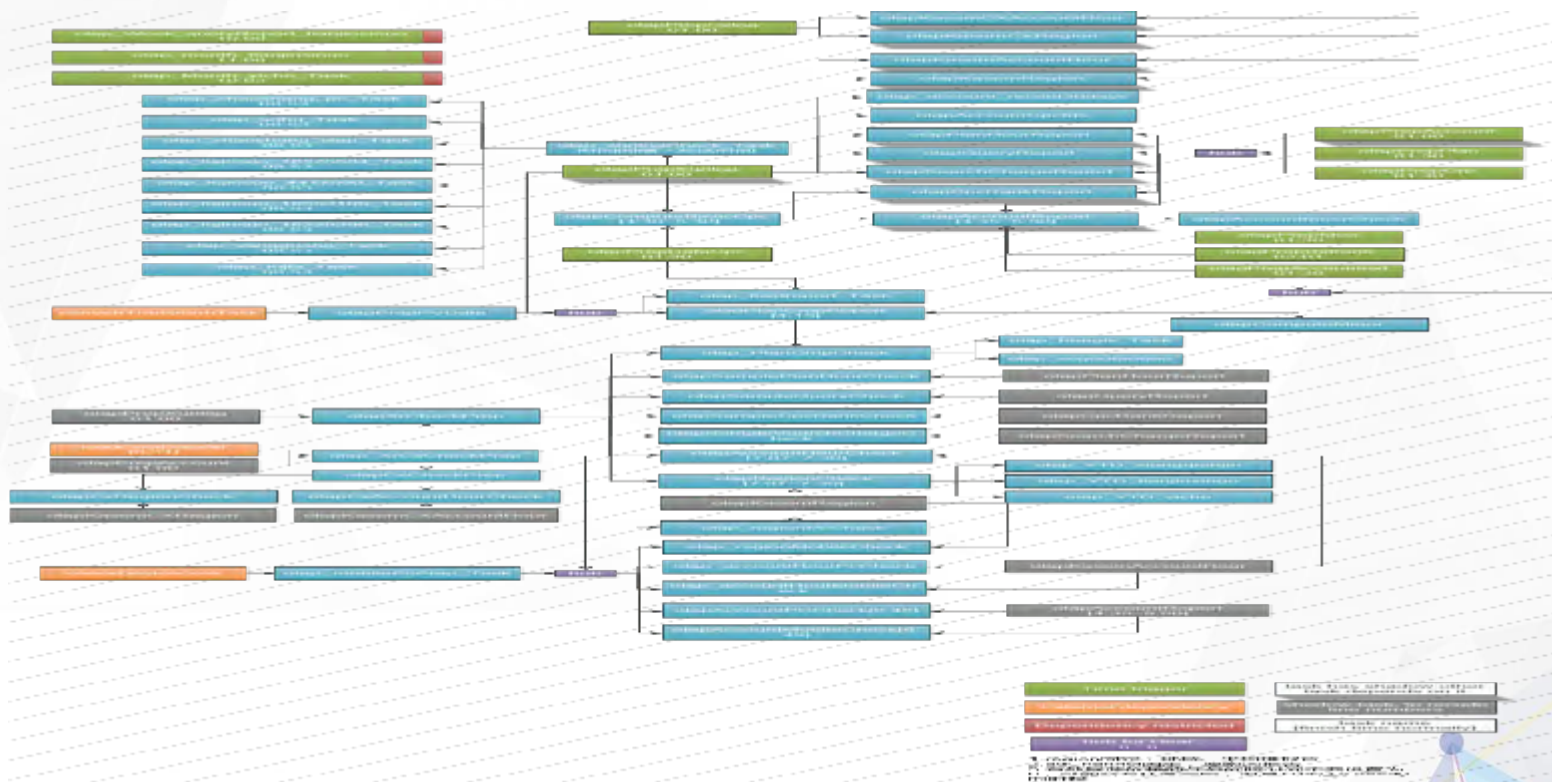
任务信息

返回

字段	类型	说明
account_id	bigint	account ID
	string	
	bigint	
	bigint	
	string	
	string	
	bigint	
	string	
	string	
	string	
comment_time	string	comment time

存储计算--血缘关系

- 快速定位数据影响，减少依赖路径
- 基于ETL任务抽取血缘关系



存储计算--ETL计算

- 涉及到数据接入后及业务统计等多个场景
- 包括抽取、转换、加载等多个动作
- 公共ETL计算

统一编码方式、数据格式转换

- 业务ETL计算

产生维度表、计算业务指标

- ETL结果一致性保证

统计口径规范化、核对同一口径下产出数据

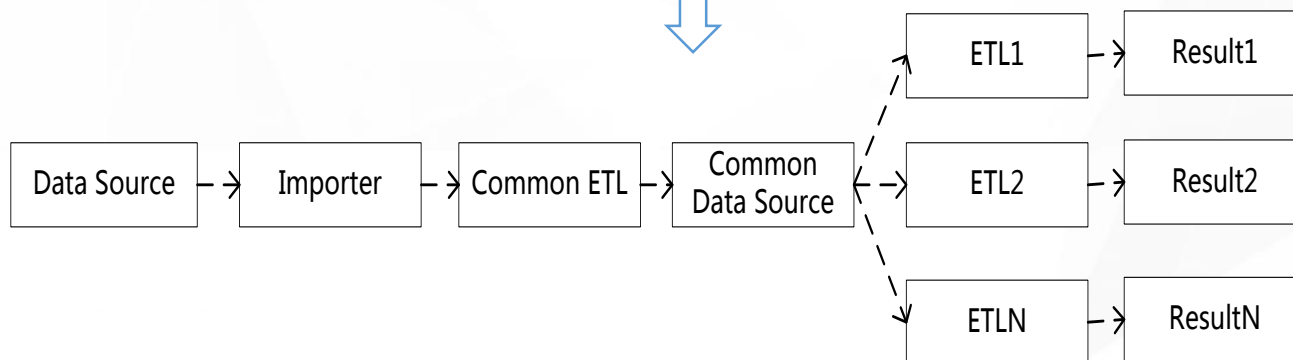
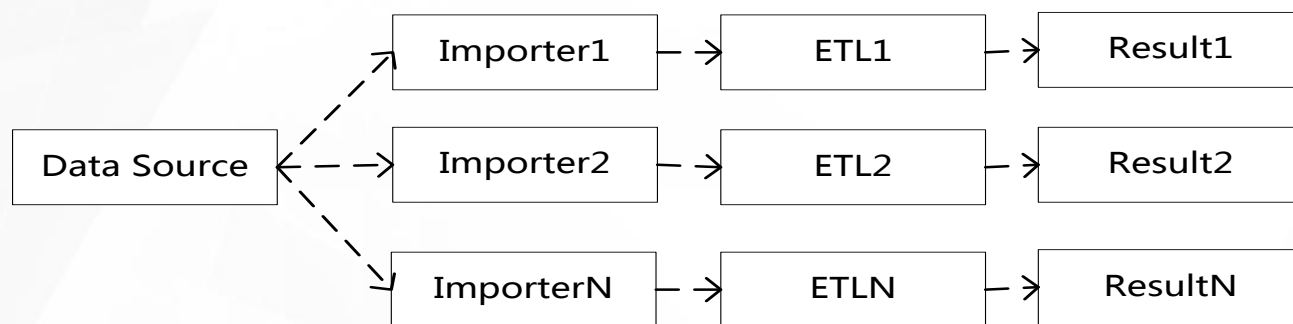
- ETL计算优化

存储计算--ETL计算

■ ETL计算优化

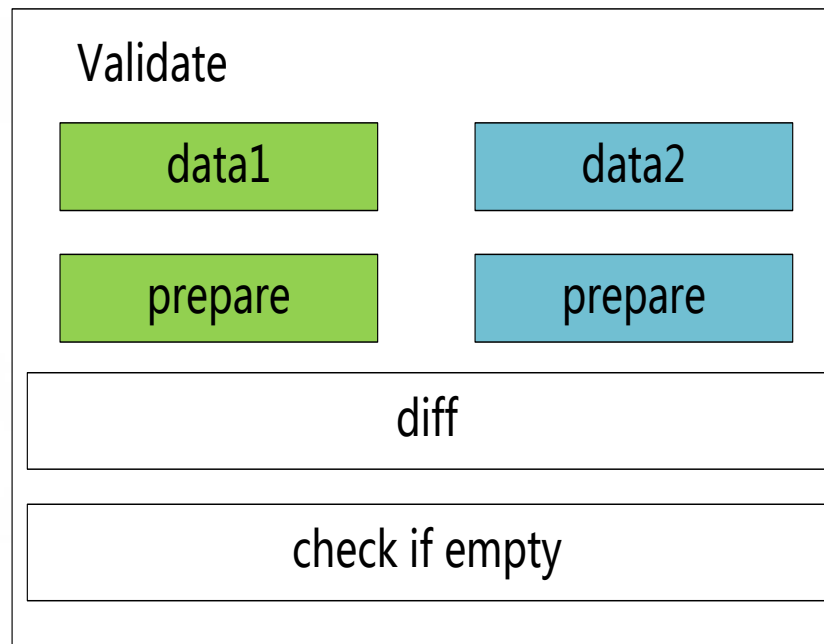
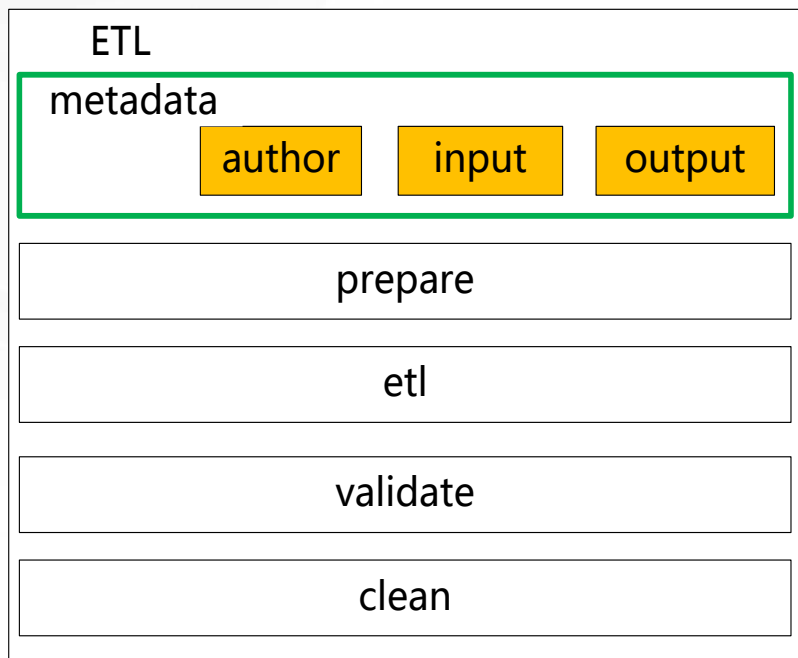
建立数据处理工作流，抽取公共计算逻辑

配置ETL元数据及任务依赖关系，抽取数据血缘关系



存储计算--ETL计算

- 元数据配置
- 结果校验



存储计算--任务调度

平台化

- 支持多业务线任务统一管理
与调度
- 提升任务开发效率、运维效率

多任务依赖

- 支持多个前驱任务或后继任务
- 支持任务依赖执行及跟踪

可扩展性

- 管理节点、调度节点、执行节点可水平扩展
- 支持任务运行类型可扩展
- 弹性扩展

稳定性

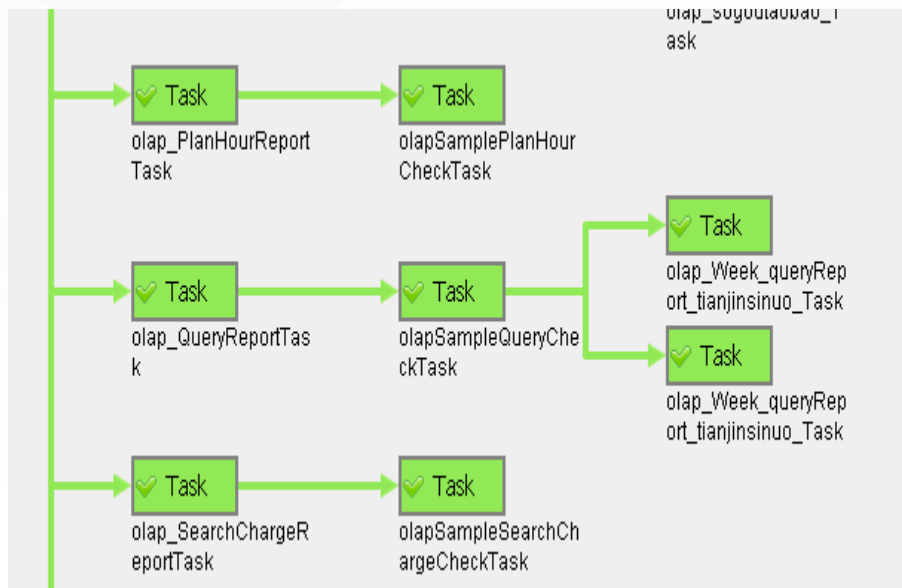
- 管理节点、调度节点、执行节点支持热备
- 节点宕机时不影响任务正常调度

存储计算--任务调度

- 任务按组划分
- 每组可有多个任务节点
- 任务并发限制
- 多个前驱任务或后继任务
- 任务运行结束后回调调度器
- 调度器根据任务报警配置发出失败报警

存储计算--任务调度

- 实时查看任务依赖链节点执行状态
- 实时查看任务依赖配置规则
- 通过任务依赖联调反推数据血缘关系

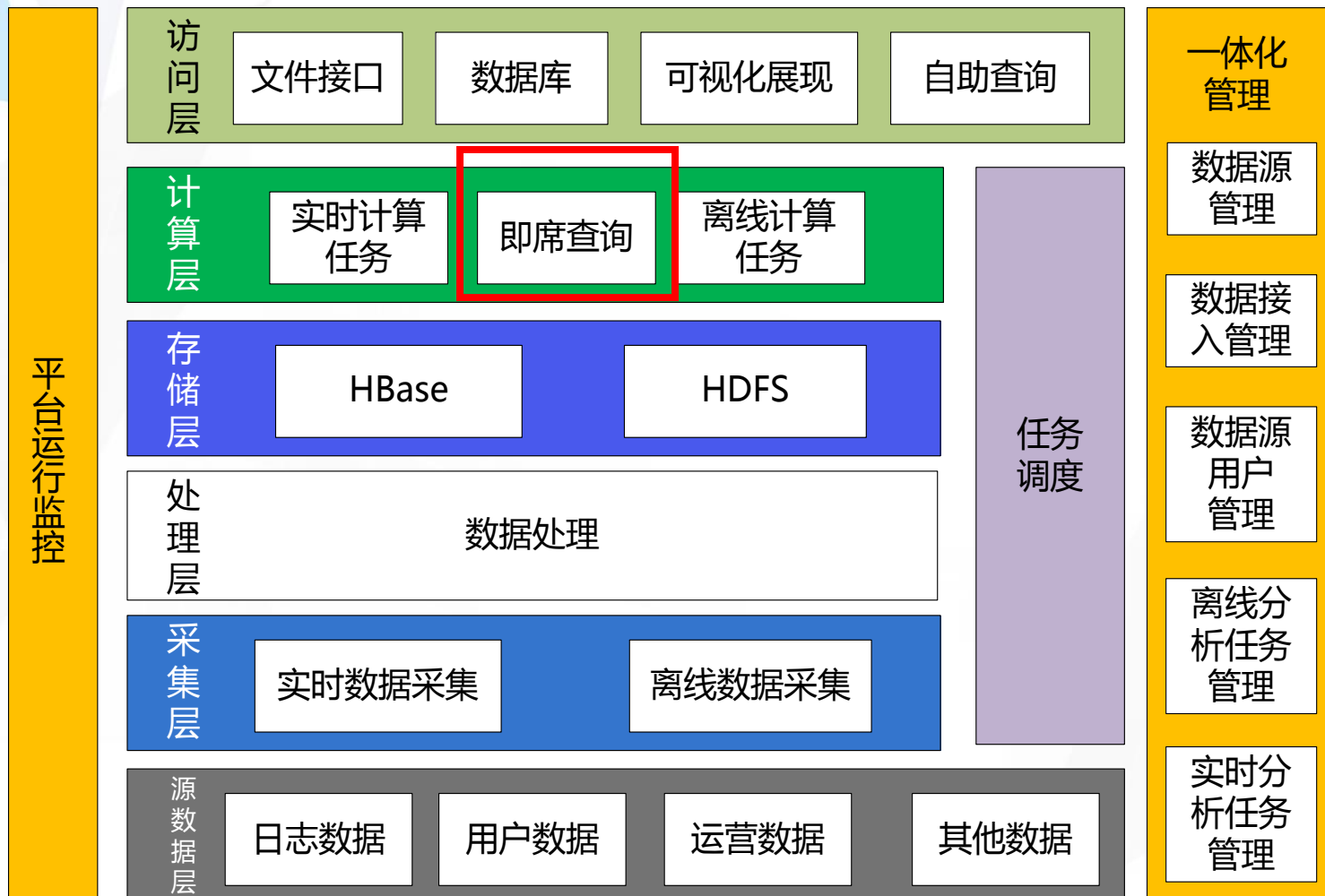


存储计算--任务调度

- 任务并发情况监控，方便合理分配资源
- 任务依赖链路执行情况跟踪，方便日志跟踪和问题快速排查



即席查询



即席查询

- 面临的挑战
- 固定指标查询
- 随机指标查询

即席查询--面临的挑战

快

- 查询响应越快越好
- 临时统计指标紧急支持

准

- 同一统计口径数据保持一致
- 多个指标互相完备

易

- 业务使用门槛低
- 界面操作一站式

即席查询--固定指标查询

- 报表数据大多数统计维度固定
- 空间换时间，基础表按固定维度定时汇总
- 统计结果校验

Cube Designer



Model Name

Cube Name

Fact Table

Lookup Table

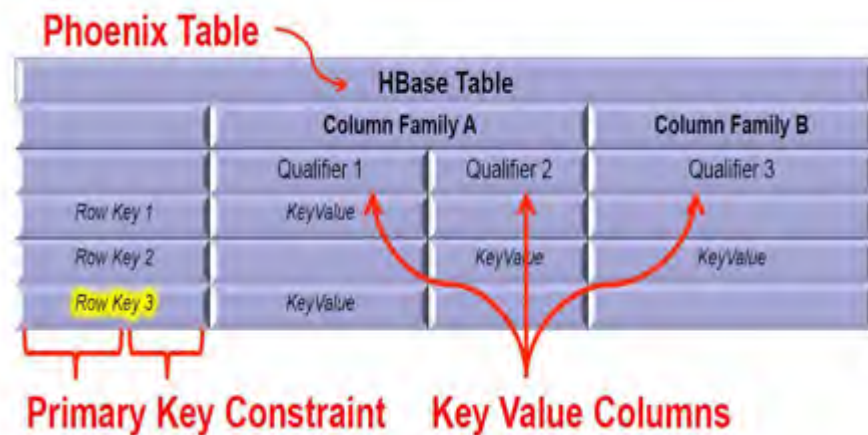
Dimensions

Measures

Description

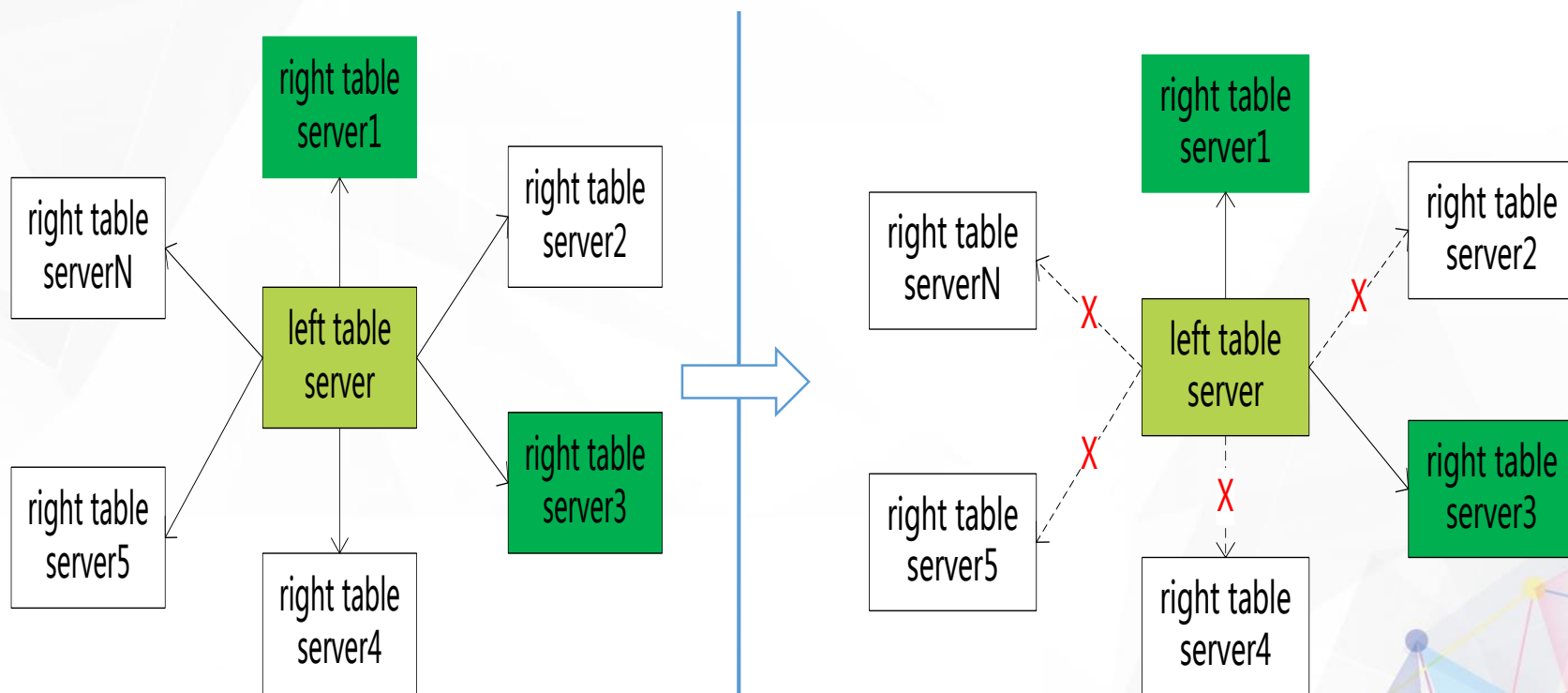
即席查询--随机指标查询

- 采用Phoenix优化HBase查询
- 支持JDBC方式访问和SQL标准
- 支持视图和支持索引
- 协处理器加速
- 定制优化，贡献多个patch



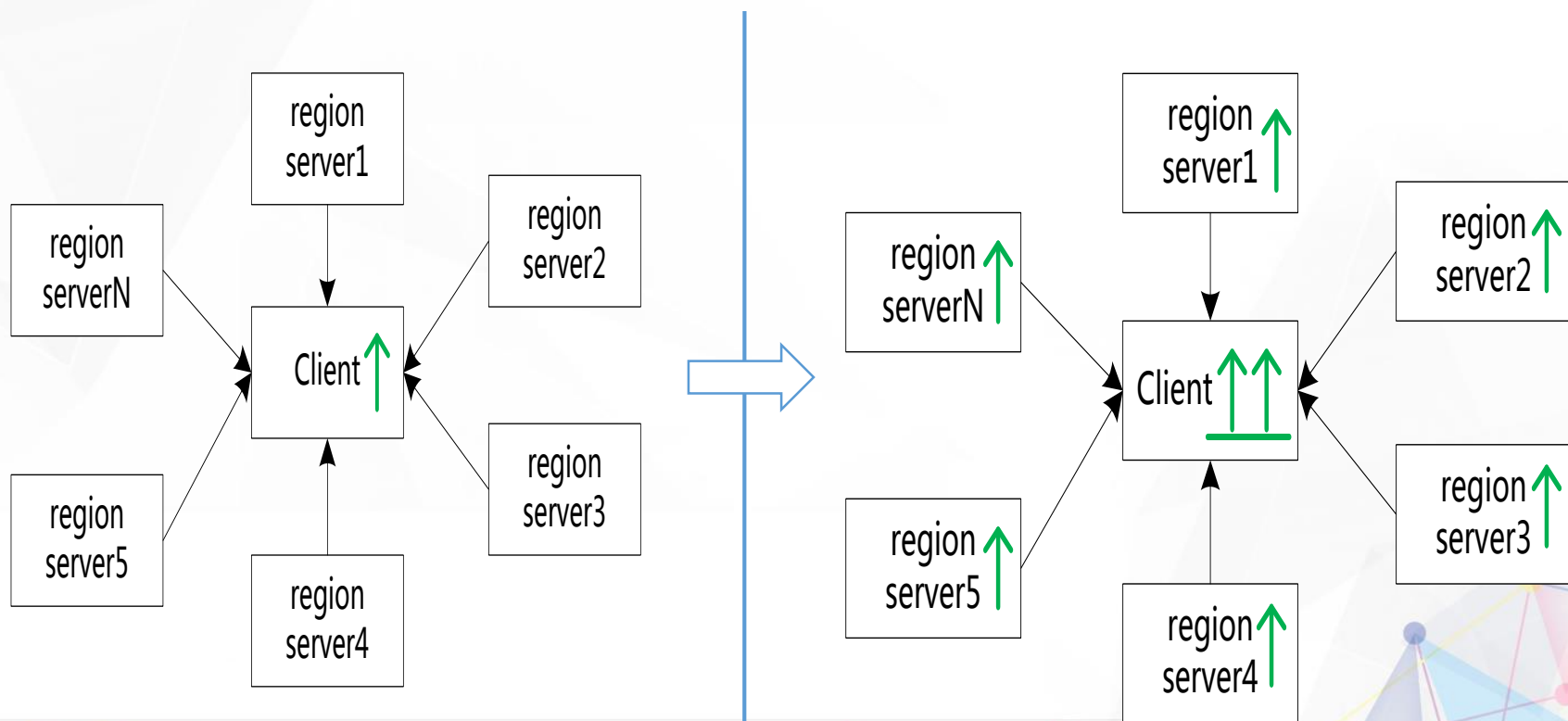
即席查询--phoenix优化(Join)

- 将原有的Join全局广播优化为只通知Rowkey所在RegionServer
- 减少通信IO时间，提高Join性能

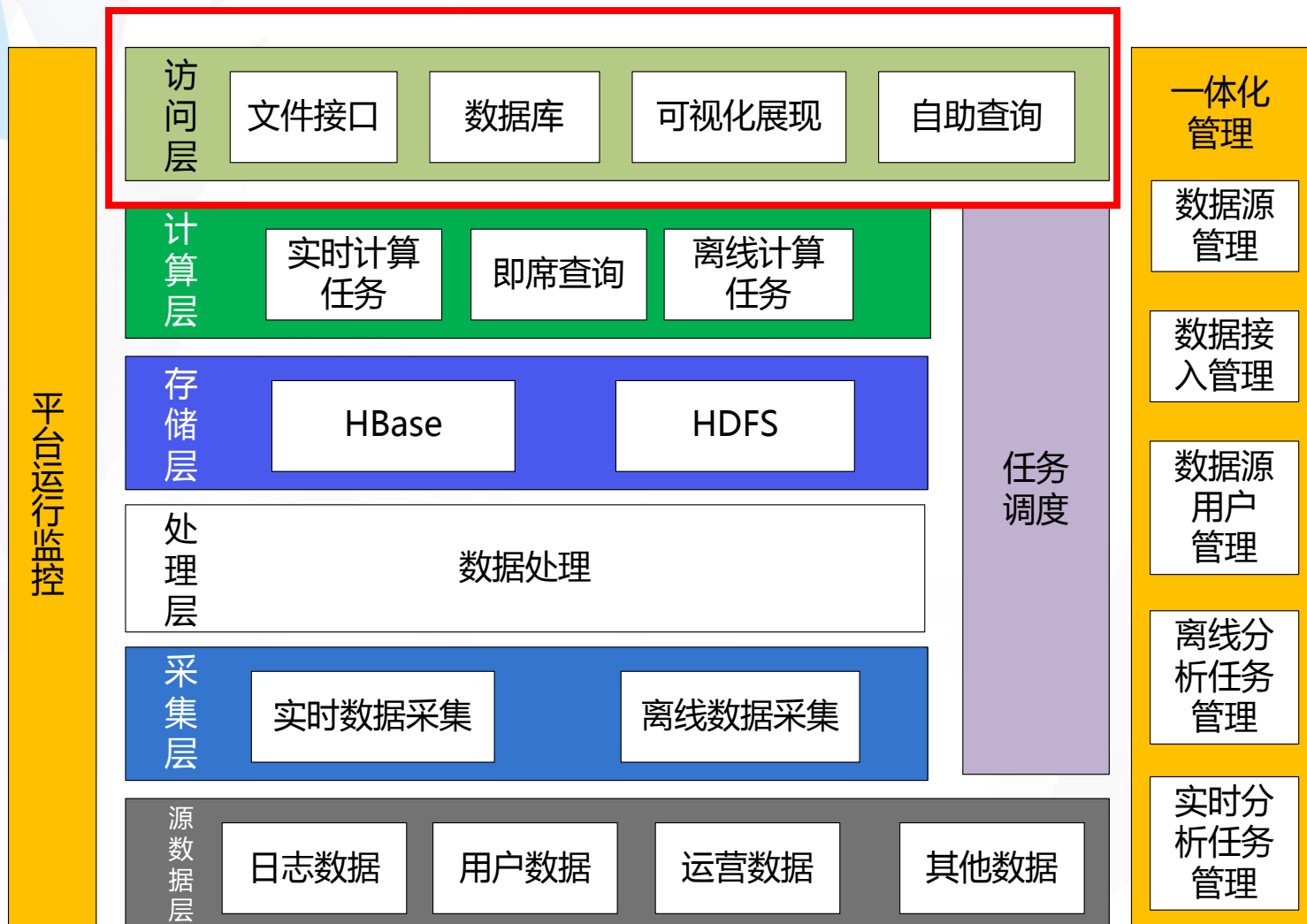


即席查询--phoenix优化(聚合)

- 原有的聚合排序全部在客户端完成，增加客户端内存和CPU负担
- 优化为RegionServer先内部聚合排序，后客户端进行二次排序



数据访问

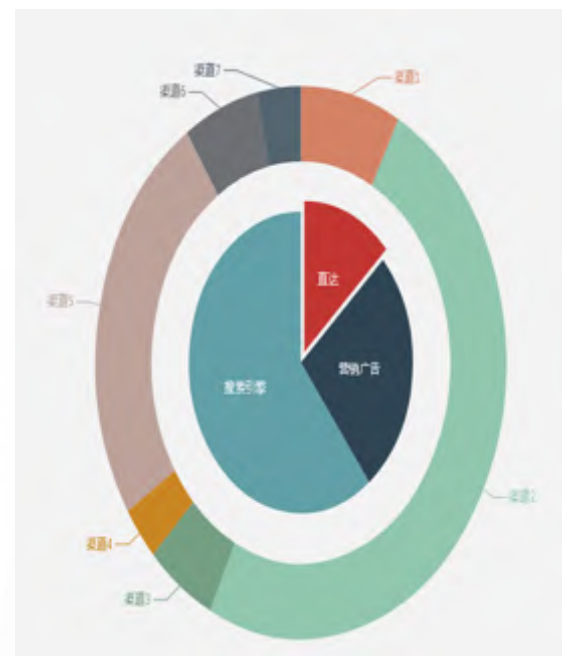
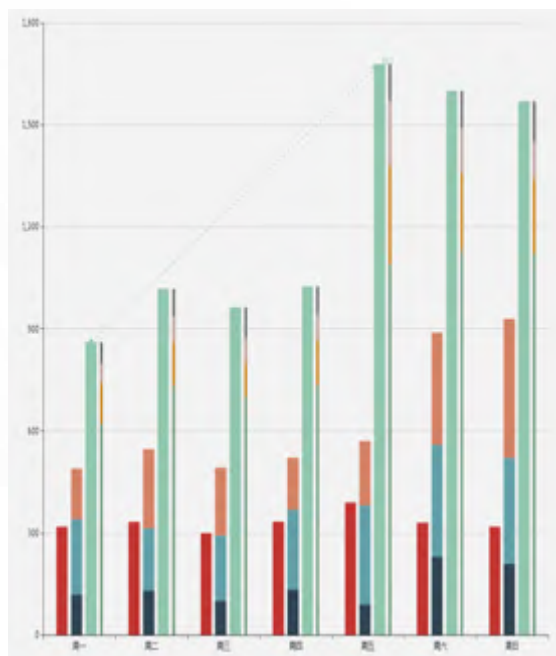
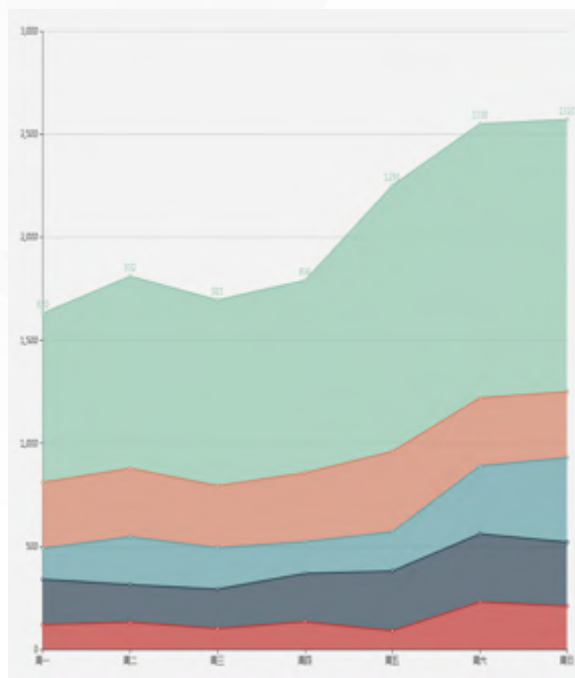


数据访问

- 可视化展现
- 自助查询

数据访问--可视化展现

- 提供柱图、饼图、曲线图等多种富有表现力展示形式
- 方便平台使用者将统计结果一站式转换
- 所见即所得



数据访问--自助

[illegible]

数据访问--自助查询

作业详情

作业名称:

作业描述:

查询模式:

脚本:

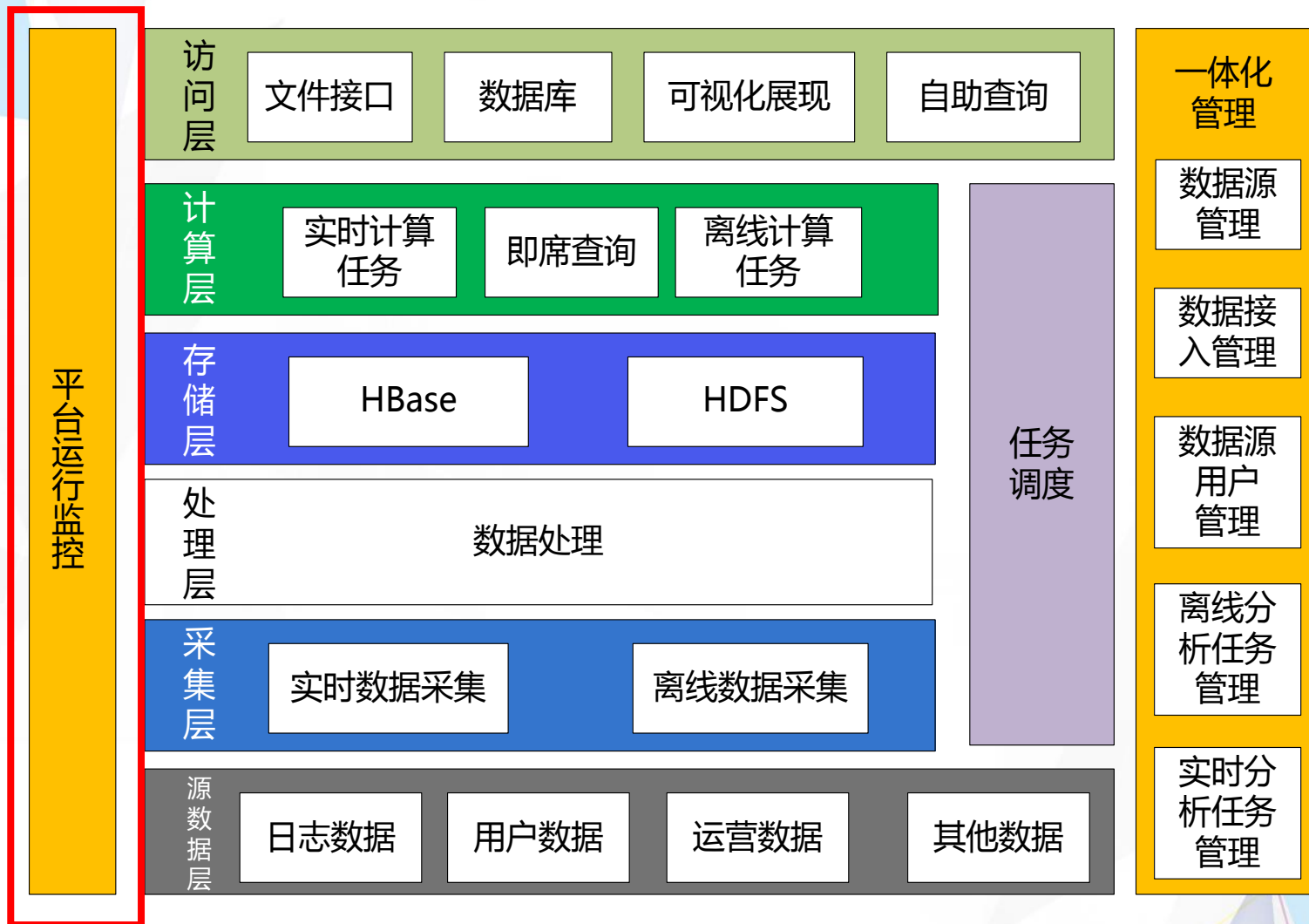
```
select * from (select * from schema1.table1 where id < 1000000) as t1
select * from (select * from schema2.table2 where id < 1000000) as t2
select * from (select * from schema3.table3 where id < 1000000) as t3
select * from (select * from schema4.table4 where id < 1000000) as t4
select * from (select * from schema5.table5 where id < 1000000) as t5
select * from (select * from schema6.table6 where id < 1000000) as t6
select * from (select * from schema7.table7 where id < 1000000) as t7
select * from (select * from schema8.table8 where id < 1000000) as t8
select * from (select * from schema9.table9 where id < 1000000) as t9
select * from (select * from schema10.table10 where id < 1000000) as t10
```

添加

返回

任务ID	提交时间	耗时	任务状态	日志	查询模式	执行脚本	结果下载	删除
41	2016-09-26 10:55:27		执行完成	下载日志	Hive	查看	下载结果	删除

平台监控



平台监控

■ 集群状态监控

硬盘、内存、CPU、网络IO、....

■ 基础设施监控

Hadoop集群、HBase集群、消息队列、实时计算引擎、....

■ 数据接入监控

数据源状态、数据接入状态、接入数据波动、.....

■ 业务计算监控

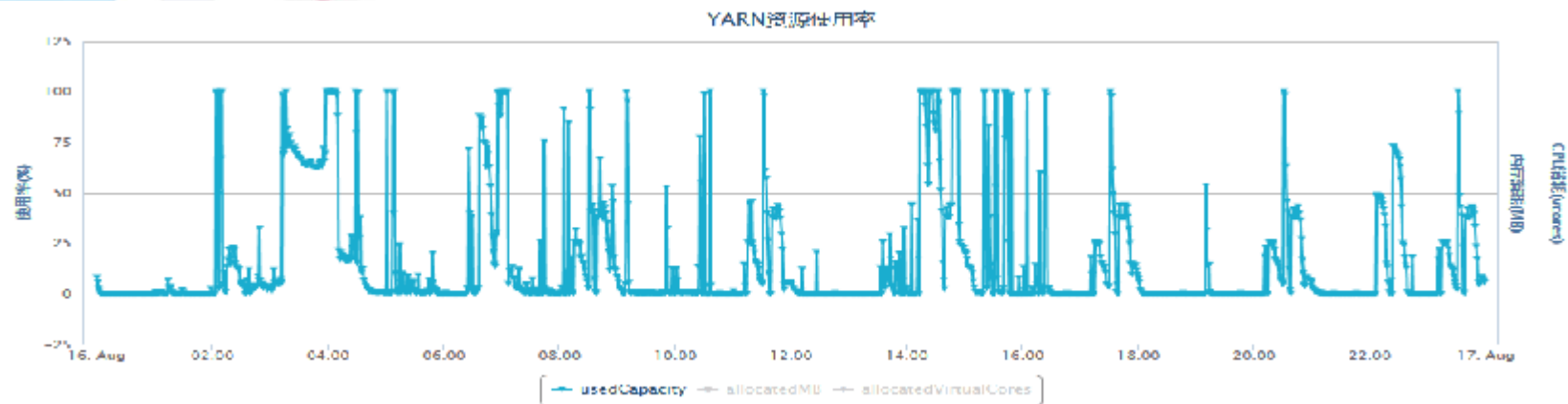
业务计算任务状态、计算结果一致性、.....

■ 平台使用监控

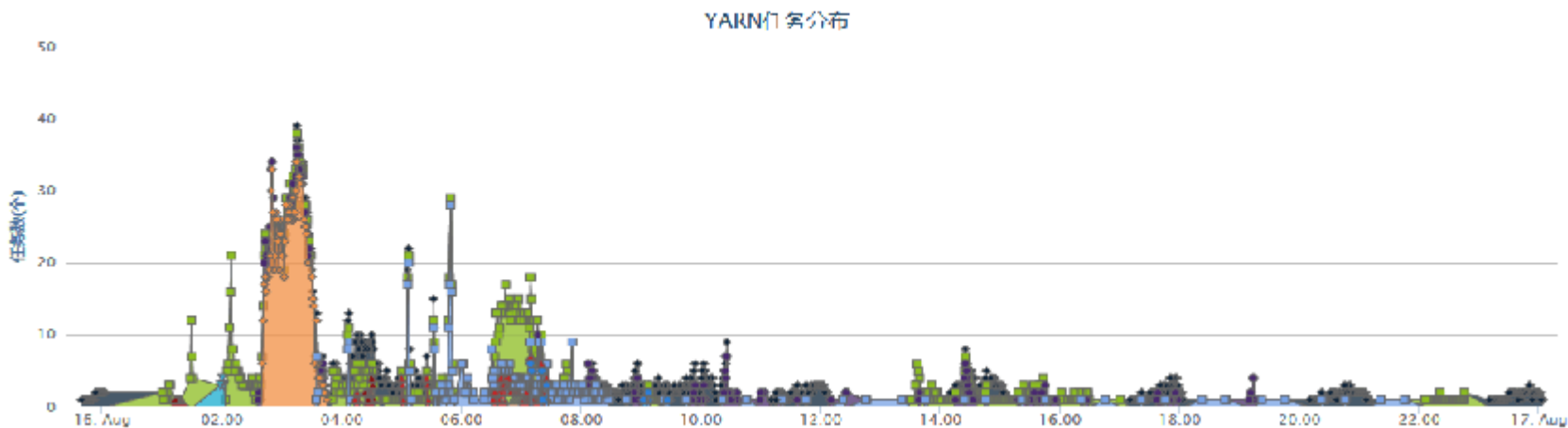
基础数据使用频度、业务数据使用频度、功能PV、.....

平台监控

■ 监控Yarn资源在各个时间段的使用情况



■ 监控YARN任务在各个业务线的分布情况

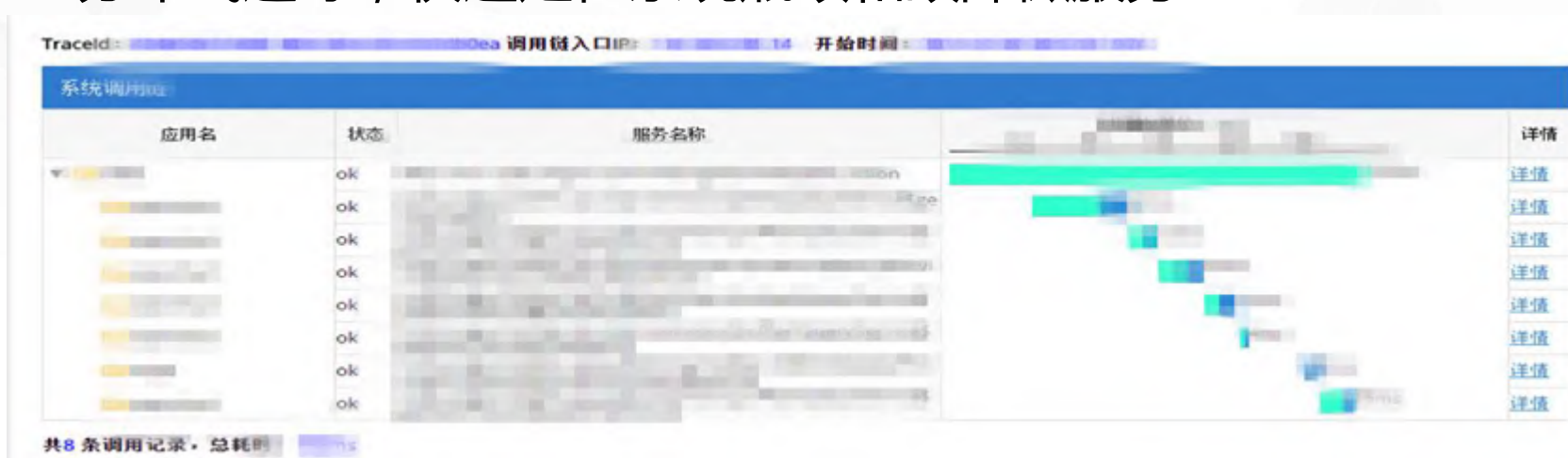


应用案例

■ 安全监控，非法入侵即刻感知及时处理



■ 分布式追踪，快速定位系统瓶颈和故障微服务



目录

CONTENTS

1 大数据的挑战

2 数据平台的建设

3 总结及展望

总结

- 数据接入--实时数据接入，离线数据接入
- 存储计算--数据仓库，ETL配置，调度平台
- 即席查询--固定指标空间换时间，随机指标借助于phoenix
- 数据使用--可视化展现，自助查询
- 平台监控--基础监控，业务监控，使用情况监控

体会

■ 从原始需求出发

为业务服务、为运维服务、为开发服务....

■ 从自身情况出发

别人的不一定适合你

■ 从基本规律出发

CAP理论、空间时间互换、.....

■ 从投入产出出发

大不一定好、多不一定好、.....

■ 从拥抱变化出发

唯一不变的是变化、什么最可能变化、什么变化的慢

展望

■ 业务规则引擎

抽取业务规则形成规则引擎，方便数据ETL及校验智能化

■ 多机房协作

跨机房高延迟带来不小挑战

■ 多用户协作

中间数据共享粒度、减少重复指标建设、保证对外输出数据准确

■ 即席查询优化

即席查询需要继续优化探索

■ 从社区中来到社区中去

关注开源、积极参与

THANKS

SequeMedia
盛拓传媒

IT168
中国IT168网

ChinaUnix

ITPUB
www.itpub.net