携程大数据平台与服务化实践

携程---------杨晓青

xq.yang@Ctrip.com

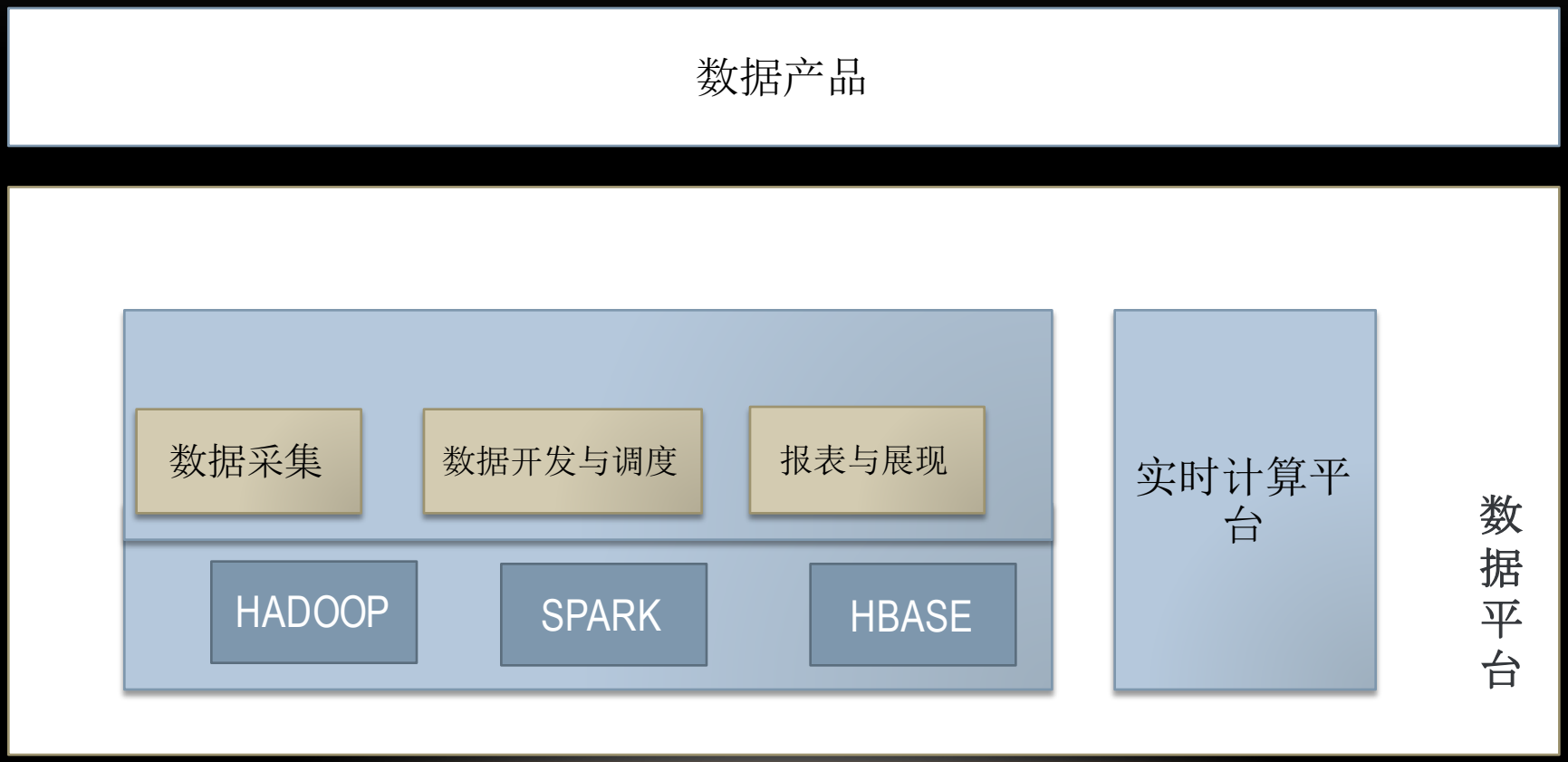
自我介绍

- 2014.8 ~ 携程-大数据平台，负责携程数据基础架构。
- 2012.4~2014.8 阿里巴巴-CDO-数据平台部，负责从无到有搭建galaxy流计算平台，集群规模到2K台。
- 2009.2~2012.4 腾讯-数据平台部，腾讯TDW平台从无到有建设。

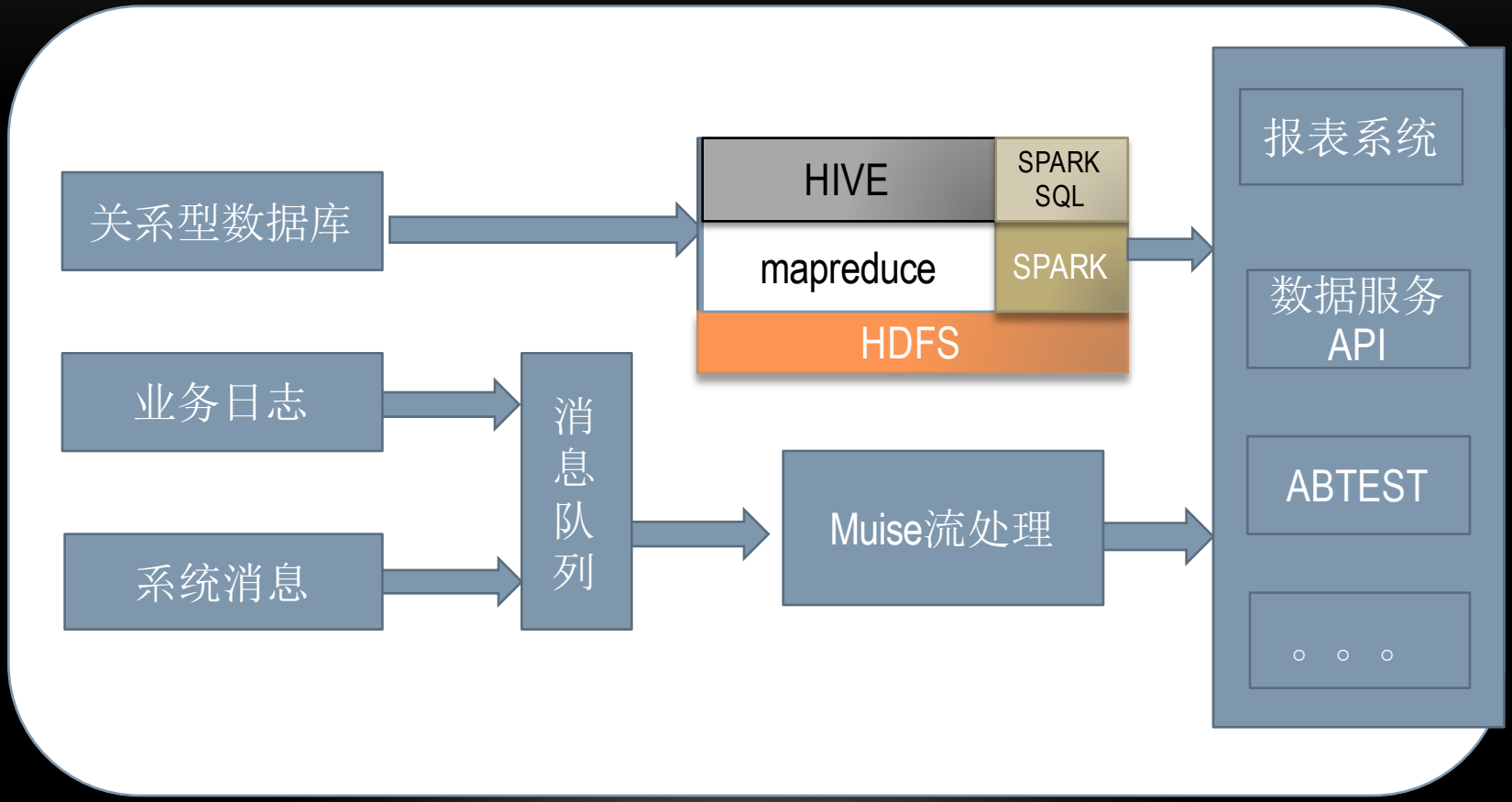
目录

目录

数据平台



数据处理流程


关系型数据库
业务日志
系统消息
消息队列
HIVE
SPARK SQL
mapreduce
SPARK
HDFS
Muise流处理
报表系统
数据服务API
ABTEST
。 。 。

大数据架构

数据采集与同步

数据调度平台

数据分析平台

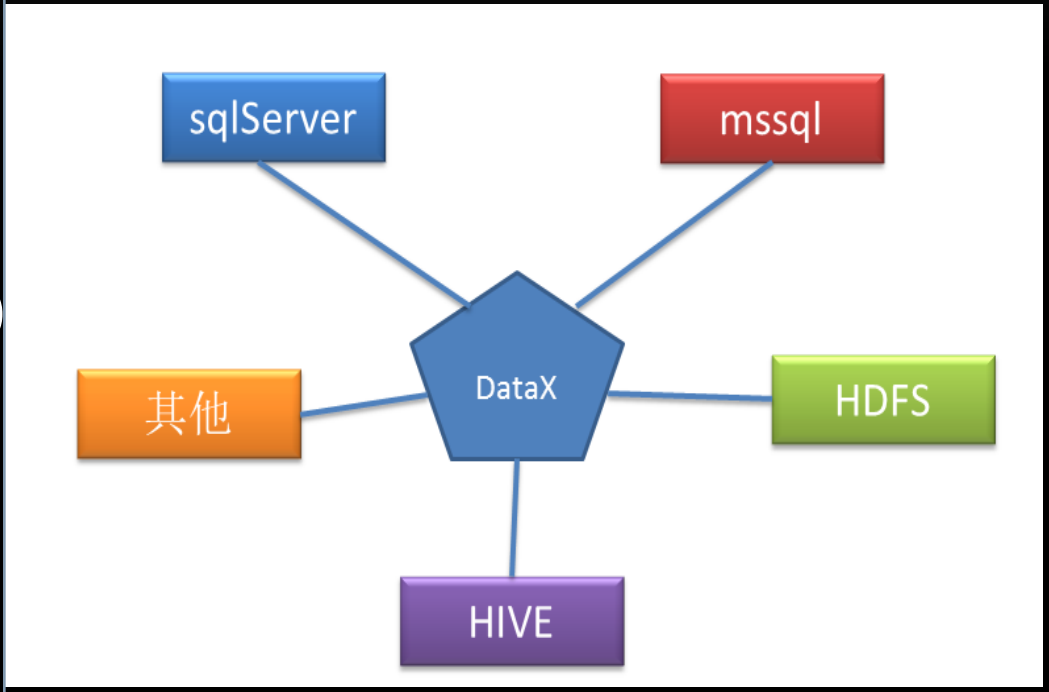
数据同步服务

1支持SQL Server，My SQL和Hive等携程主流数据库间任意同步（解决多种同步工具混用，效率难以保证，维护负责问题）

2.支持全量与增量数据同步模式

3.可线性扩展

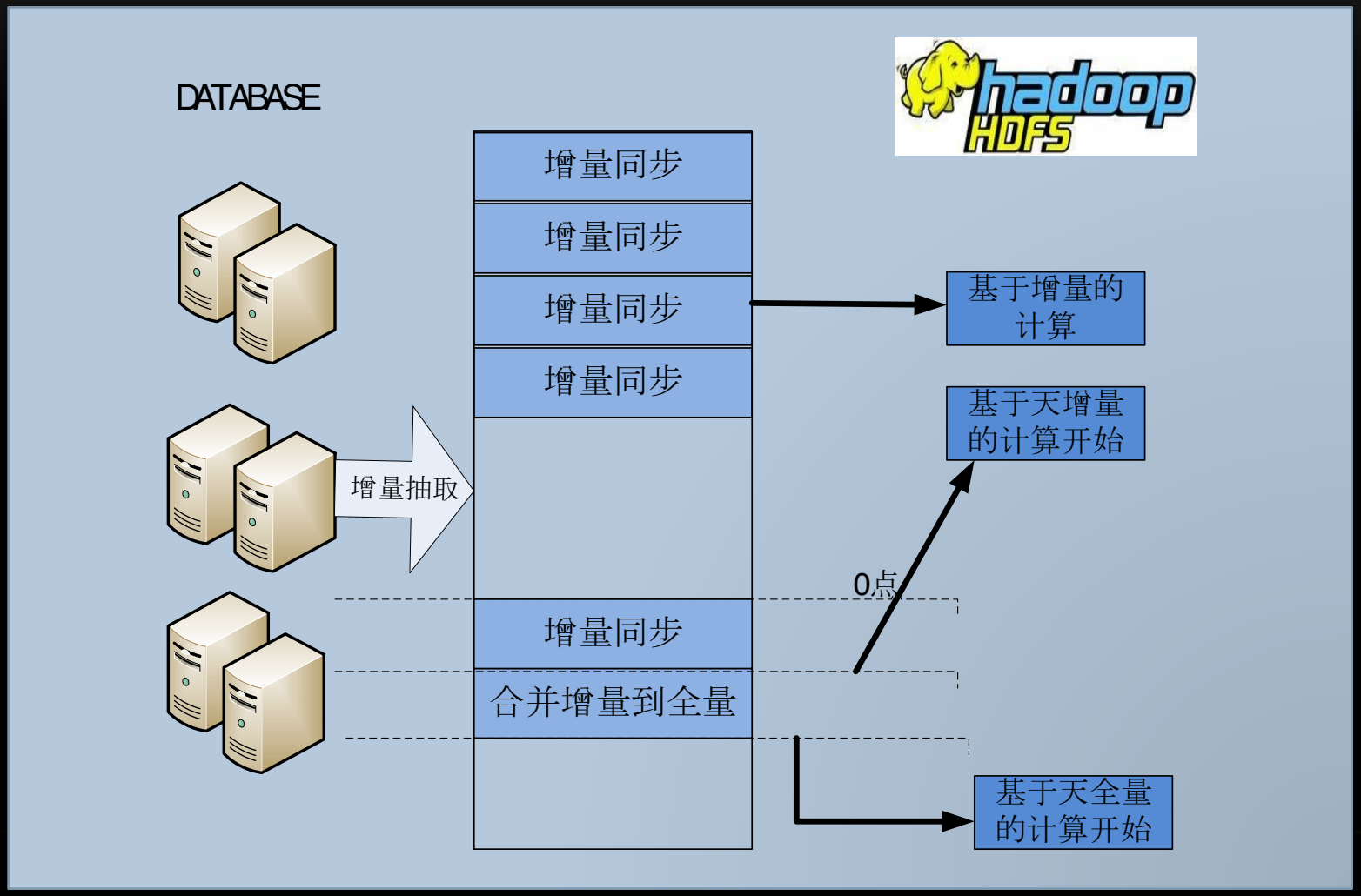
4.以SASS（SOFTWARE as a service）提供服务



交易数据同步配置


datax配置工具
源
源（【sqlserver|mysql】）
sqlserver
源数据库域名*
请填写域名，不要使用IP
源数据库端口*
源数据库名*
源数据库用户名*
源数据库密码*
源数据库表名（逗号分隔）*
源sql查询
querysplit
目标
目标
hive
目标数据库名*
目标数据库表名（逗号分隔）*
hive导入模式
全量不分区
hive文件类型
TXT
任务并发数
3
生成脚本
提示：需要拷贝生成的代码到编辑器中

准实时数据采集


DATABASE
hadoop
HDFS
增量同步
增量同步
增量同步
增量同步
增量抽取
增量同步
合并增量到全量
基于增量的计算
基于天增量的计算开始
0点
基于天全量的计算开始

运营数据

业务指标（日）	8月
同步表数据	8000+张
数据量	5T+
可用度	100%
同步行数	80亿条

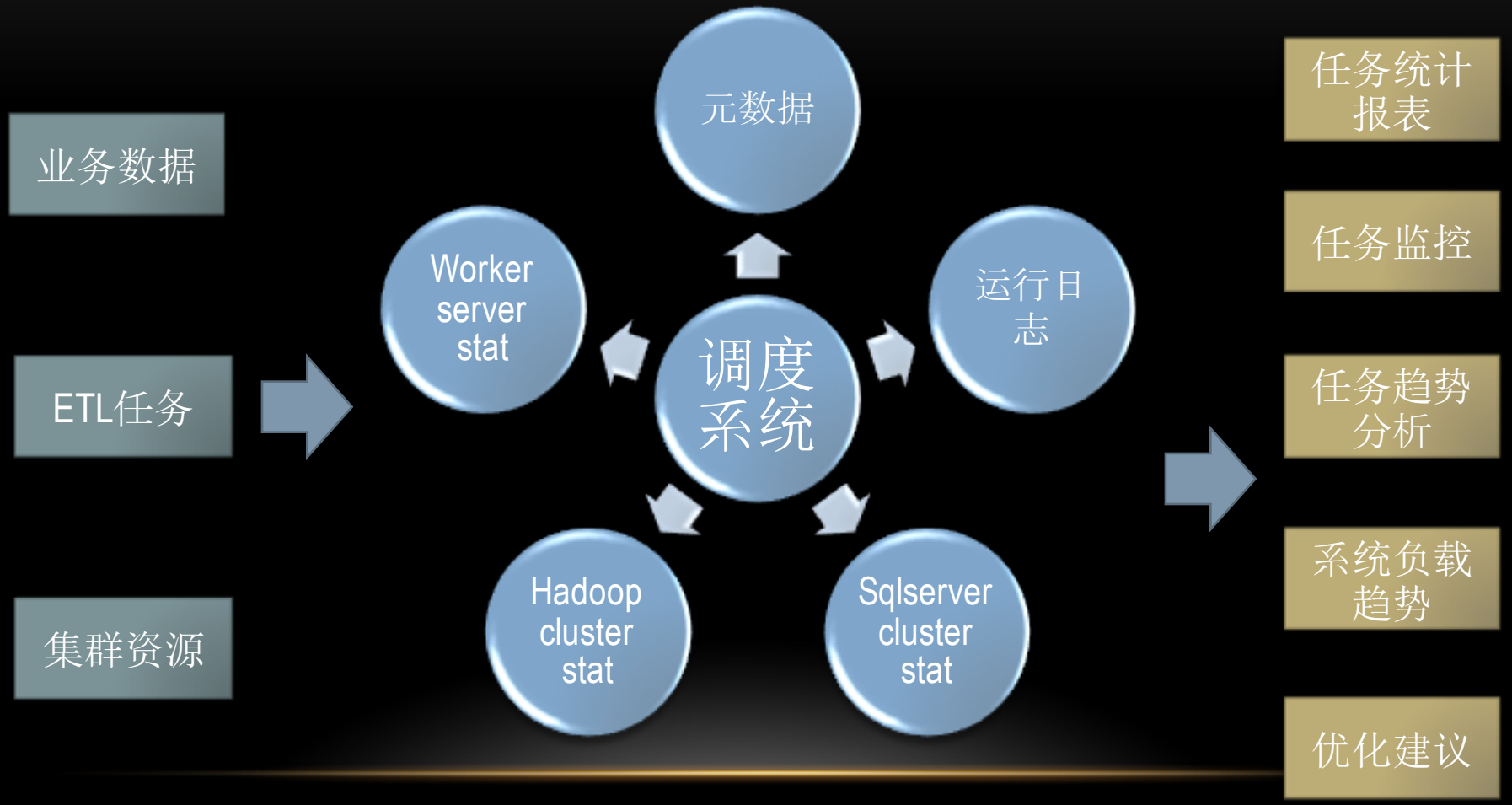
大数据架构

数据采集与同步

数据调度平台

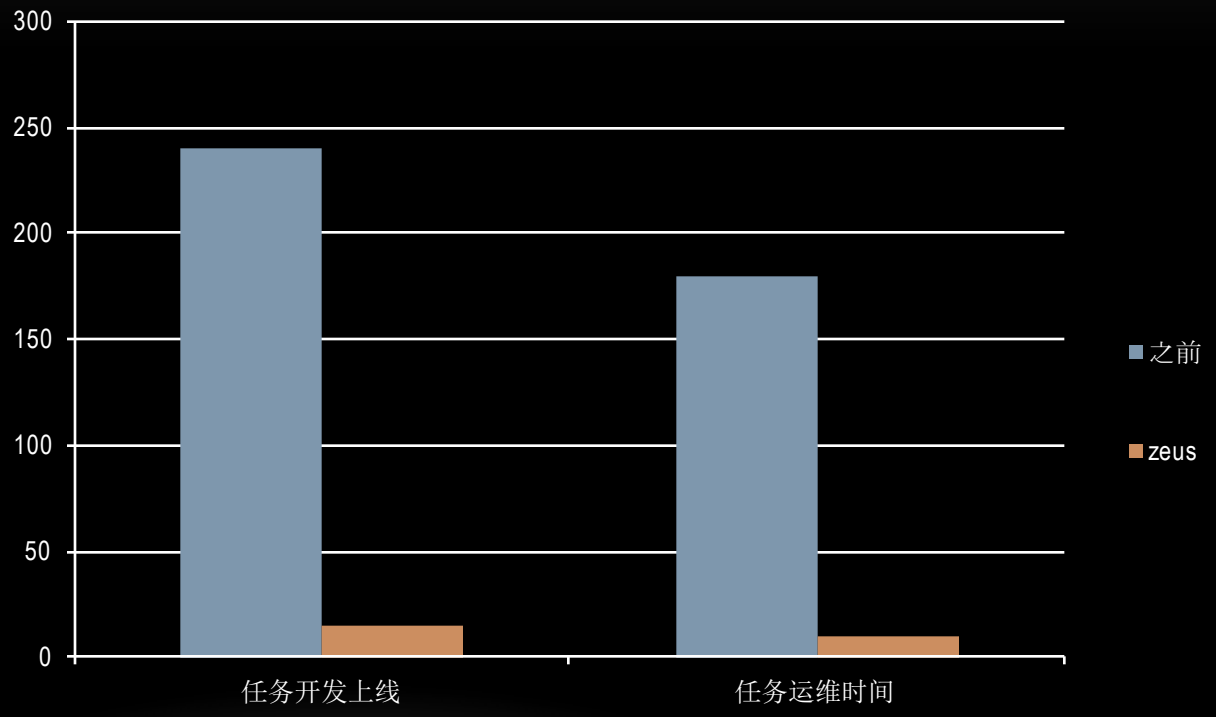
数据分析平台

调度系统


业务数据
ETL任务
集群资源
元数据
Worker server stat
运行日志
调度系统
Hadoop cluster stat
Sqlserver cluster stat
任务统计报表
任务监控
任务趋势分析
系统负载趋势
优化建议

生产率银弹

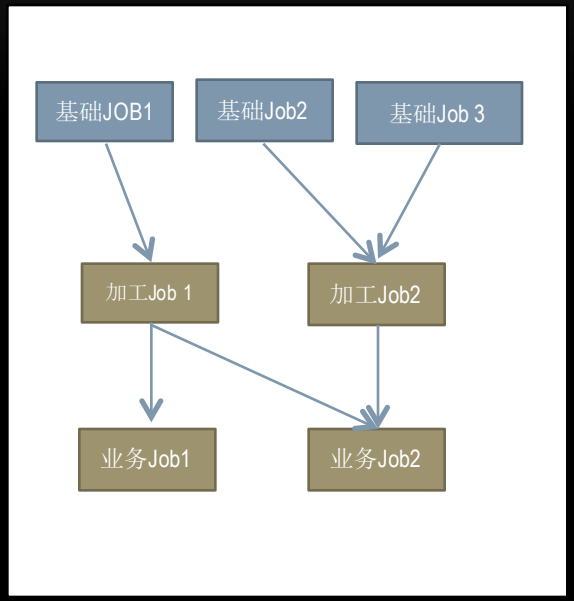
- 自动部署
- 智能调度
- 运维平台
- 监控告警
- 异构平台支持
- Web UI



系统模块

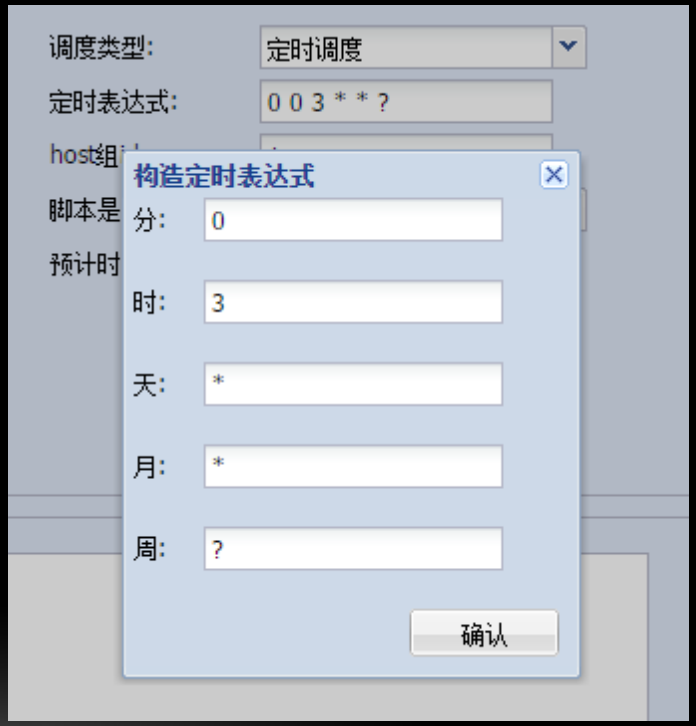
开发IDE	工作流	运维管理	智能管理
• Shell，hivesql语法高亮 • 调试运行 • 元数据查看 • 数据预览	• 依赖管理 • 时间触发 • 事件触发	• 手动重跑 • 依赖重跑 • 查看任务状 • 查看错误日志 • 失败，超时告警 • 邮件，短信，电话告警 • 每日任务报告	• 优先级 • 负载均衡 • 过载保护 • 分组调度

调度系统---任务触发方式



依赖触发

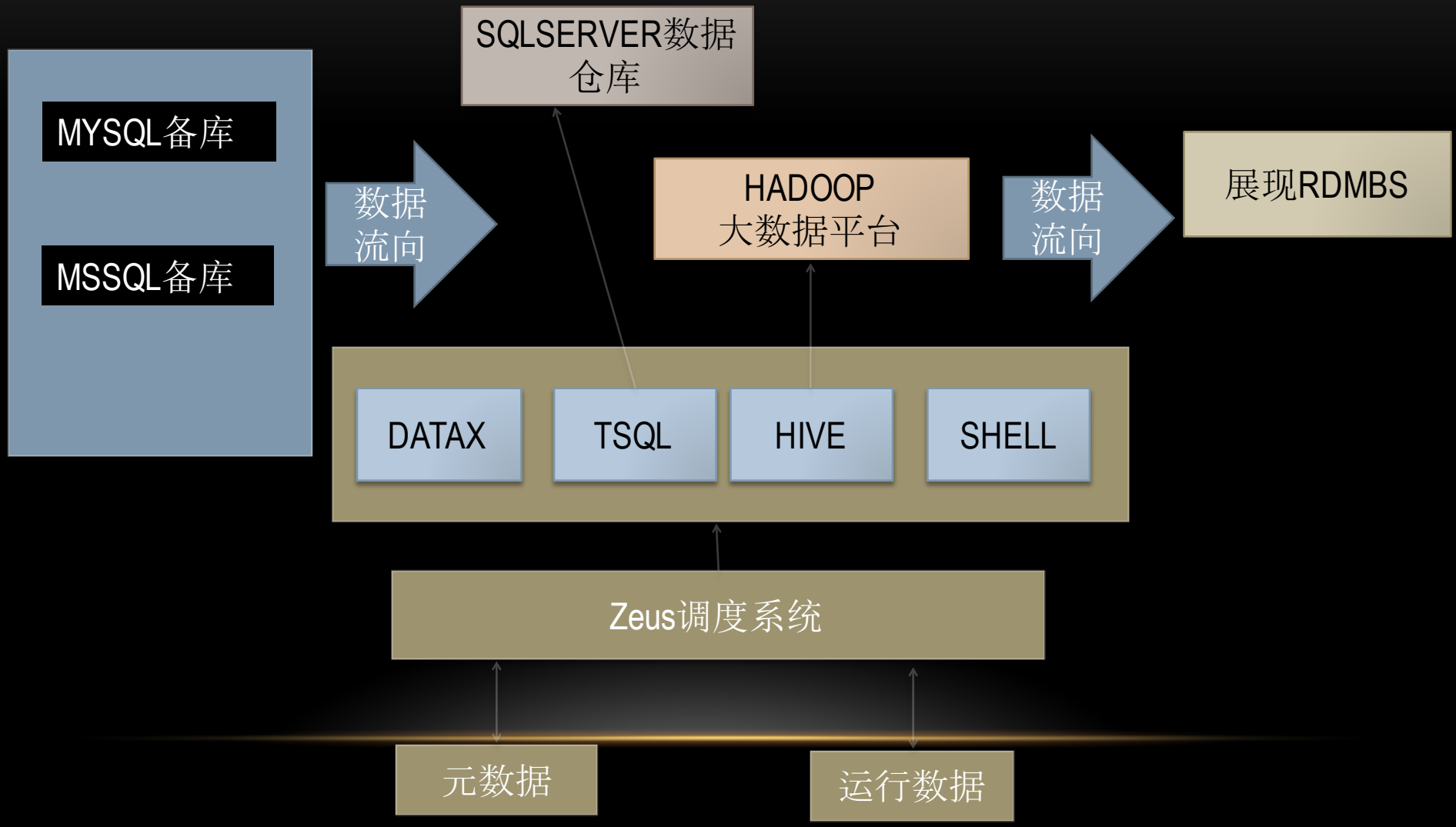
定时调度



循环触发

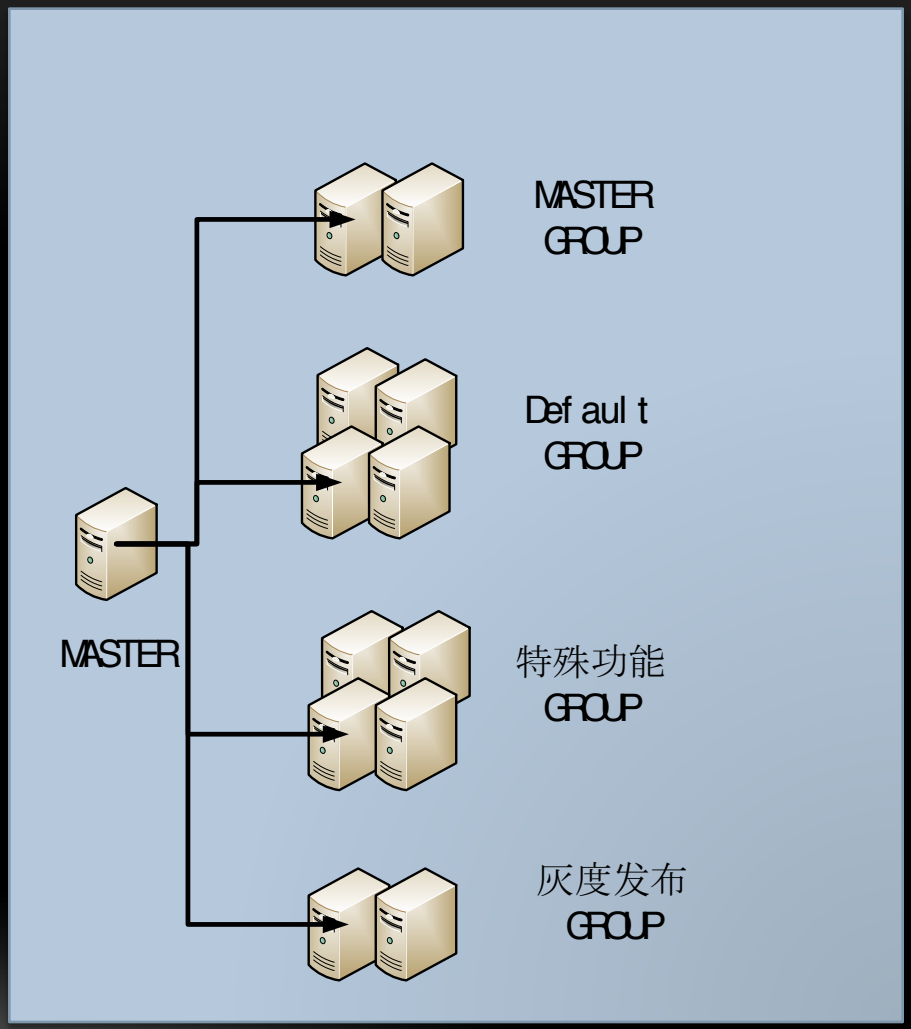
（尽请期待）

调度---驱动数据系统方式


SQLSERVER数据仓库
MYSQL备库
MSSQL备库
数据流向
HADOOP大数据平台
数据流向
展现RDMBS
DATAX
TSQL
HIVE
SHELL
Zeus调度系统
元数据
运行数据

调度---资源分组

- 基础功能隔离
- R集群化
- Python集群化
- Zeus HA
- Zeus灰度发布
- 特殊业务线隔离



ZEUS---优先级

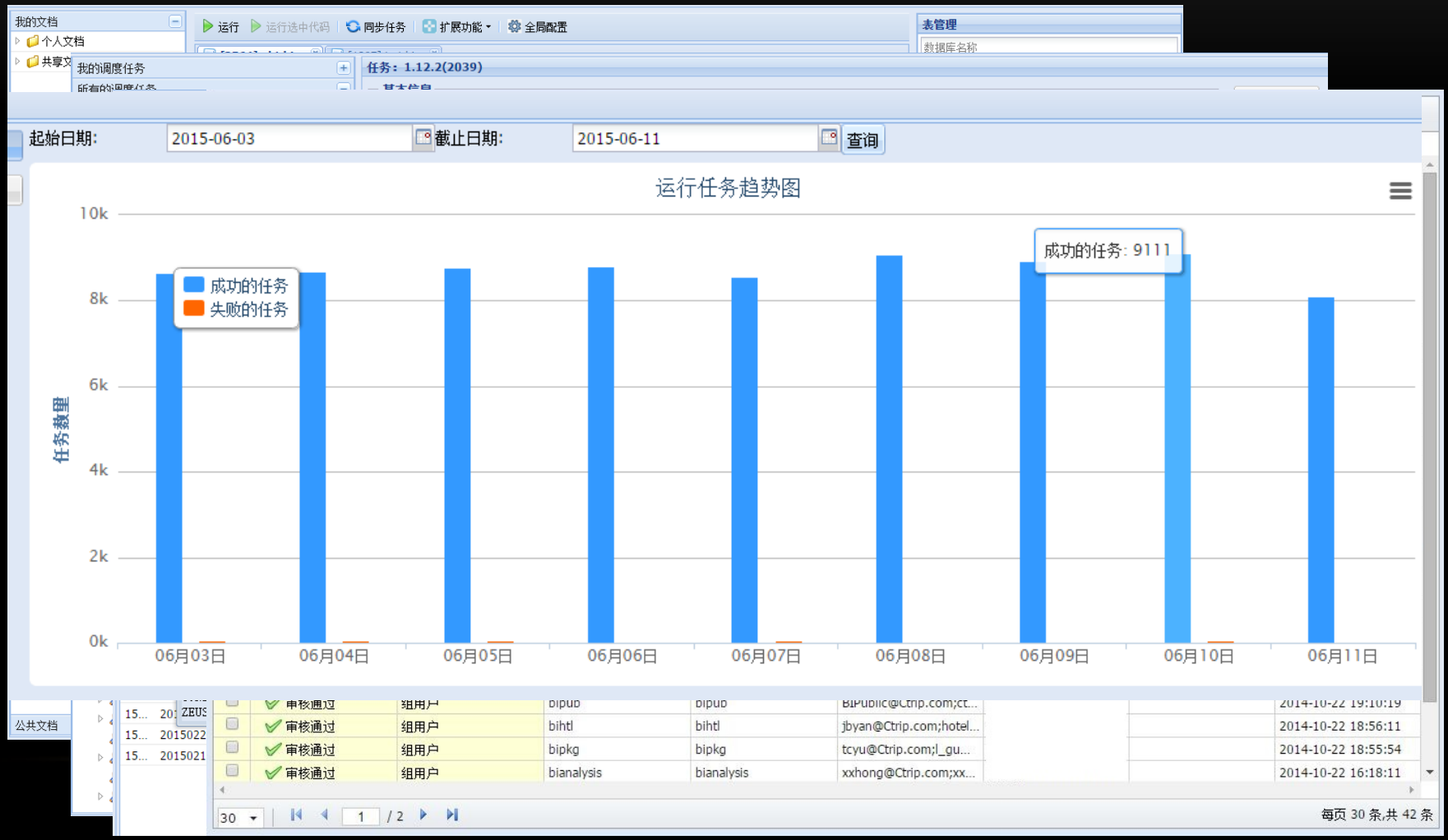

没有优先级
业务重要程度
ETL进度
客户反应
高
低
ETL JOBS
ETL JOBS
ETL JOBS
ETL JOBS
ETL JOBS
有优先级
业务重要程度
ETL进度
客户反应
高
低
ETL JOBS
ETL JOBS
ETL JOBS
ETL JOBS
ETL JOBS

应用情况

指标	
Zeus JOB数（日）	2w
集群规模	33台

开源地址：
https://github.com/ctripcorp/dataworks-zeus

功能演示


我的文档
个人文档
共享文
运行
运行选中代码
同步任务
扩展功能
全局配置
表管理
数据库名称
我的调度任务
任务：1.12.2(2039)
起始日期:
2015-06-03
截止日期:
2015-06-11
查询
运行任务趋势图
成功的任务: 9111
成功的任务
失败的任务
任务数量
10k
8k
6k
4k
2k
0k
06月03日
06月04日
06月05日
06月06日
06月07日
06月08日
06月09日
06月10日
06月11日
公共文档
审核通过
组用户
bihtl
jbyan@Ctrip.com;hotel...
2014-10-22 18:56:11
bipkg
tcyu@Ctrip.com;l_gu...
2014-10-22 18:55:54
bianalysis
xxhong@Ctrip.com;xx...
2014-10-22 16:18:11
30
1
/ 2
每页 30 条,共 42 条

系

JobId:201507280
JobId:201507290
JobId:201507300
JobId:201507310
JobId:201507180
JobId:201507190
JobId:201507200
JobId:201507220
JobId:201507230
JobId:201507230
JobId:201507240
JobId:201507250
JobId:201507260
JobId:201507270
JobId:201507280
JobId:201507290
JobId:201507300
JobId:201507310
JobId:201507310
JobId:201507180
JobId:201507180
JobId:201507180
JobId:201507231
JobId:201507231
JobId:201507300
JobId:201507231
JobId:201507181
JobId:201507181
JobId:201507181
JobId:201507181
JobId:201507181
JobId:201507181
JobId:201507181
JobId:201507181
JobId:201507180

host组信息：

组id	名称	描述	host	CurrentPosition
1	default	默认组	[redacted]	8
2	master	master抢占组		0
3	[redacted]	单个节点		0
4	[redacted]	单个节点		0
5	[redacted]	单个节点		0
6	[redacted]	单个节点		0
7	[redacted]	R节点		0
8	[redacted]	Python节点（地面专用）		0

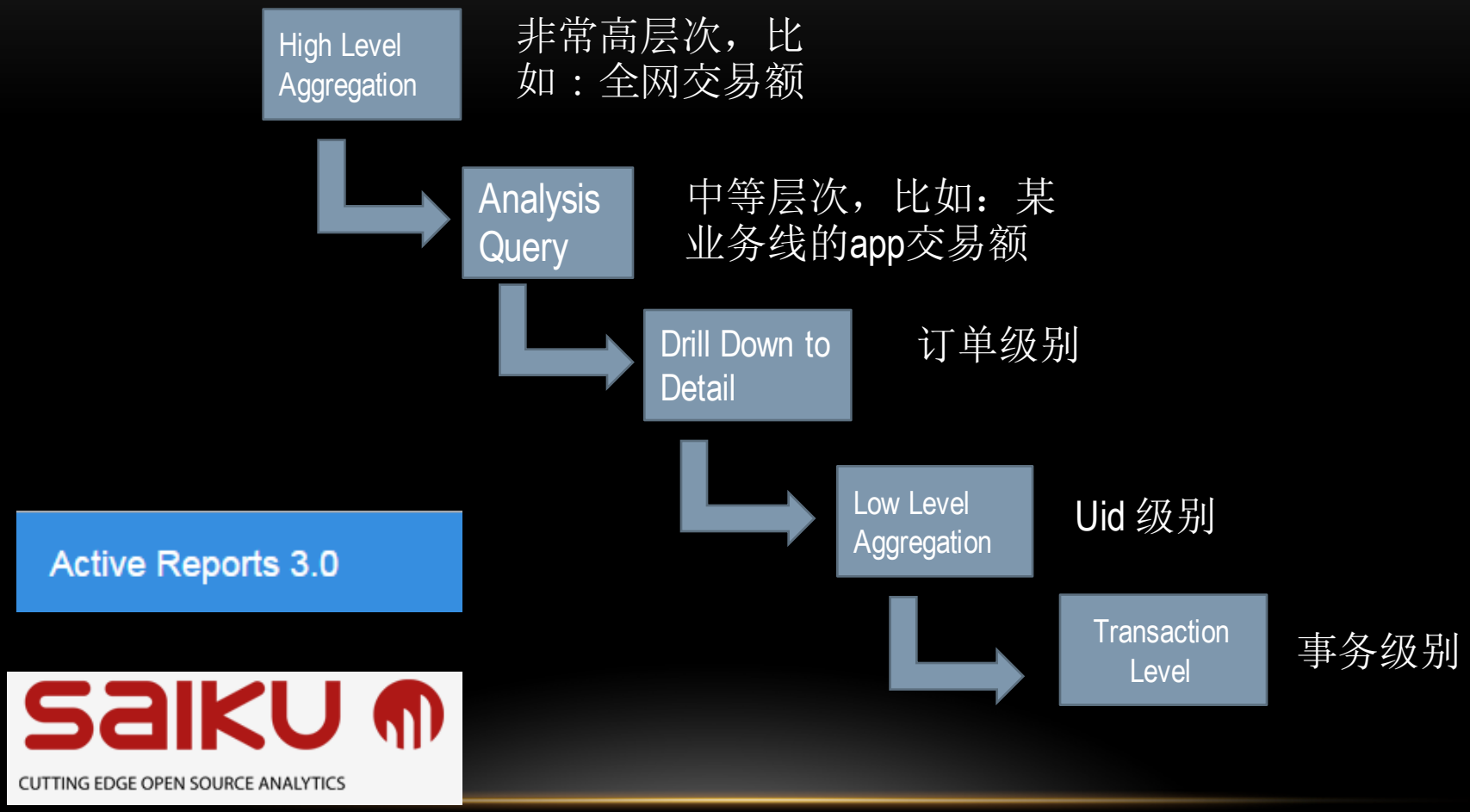
大数据架构

数据采集与同步

数据调度平台

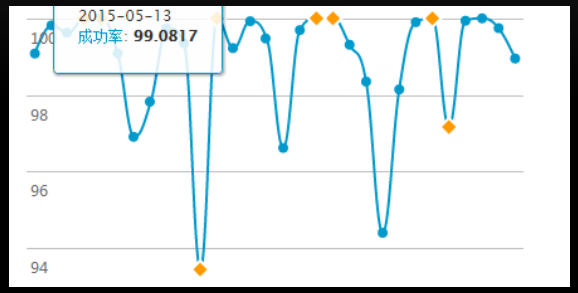
数据分析平台

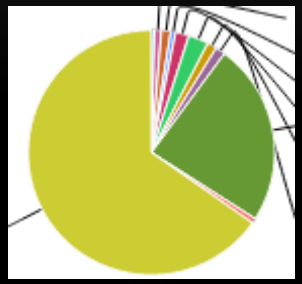
分析数据粒度层次

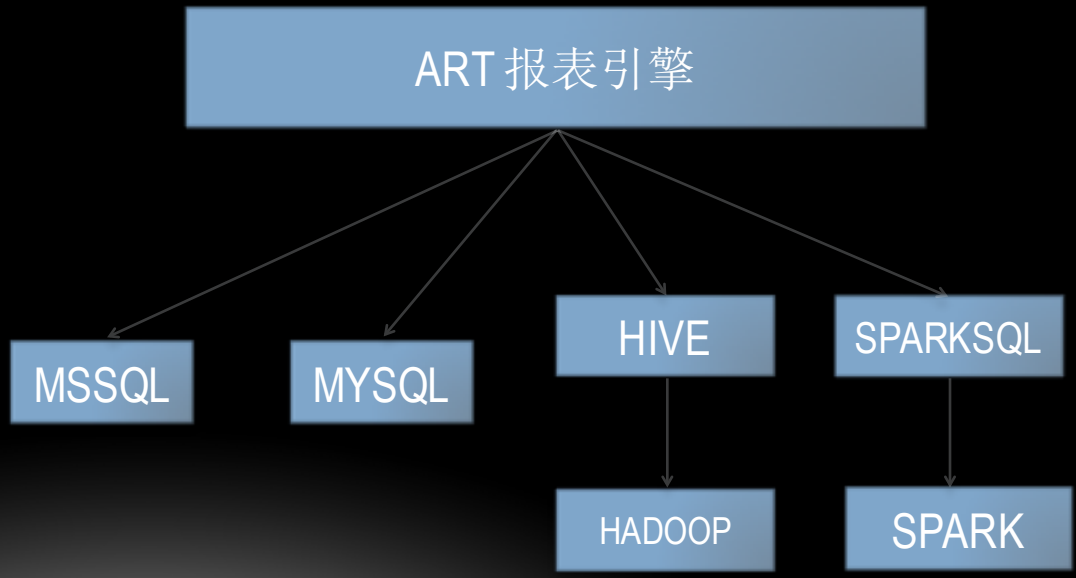

High Level Aggregation
非常高层次，比如：全网交易额
Analysis Query
中等层次，比如：某业务线的app交易额
Drill Down to Detail
订单级别
Low Level Aggregation
Uid 级别
Transaction Level
事务级别
Active Reports 3.0
saiku
CUTTING EDGE OPEN SOURCE ANALYTICS

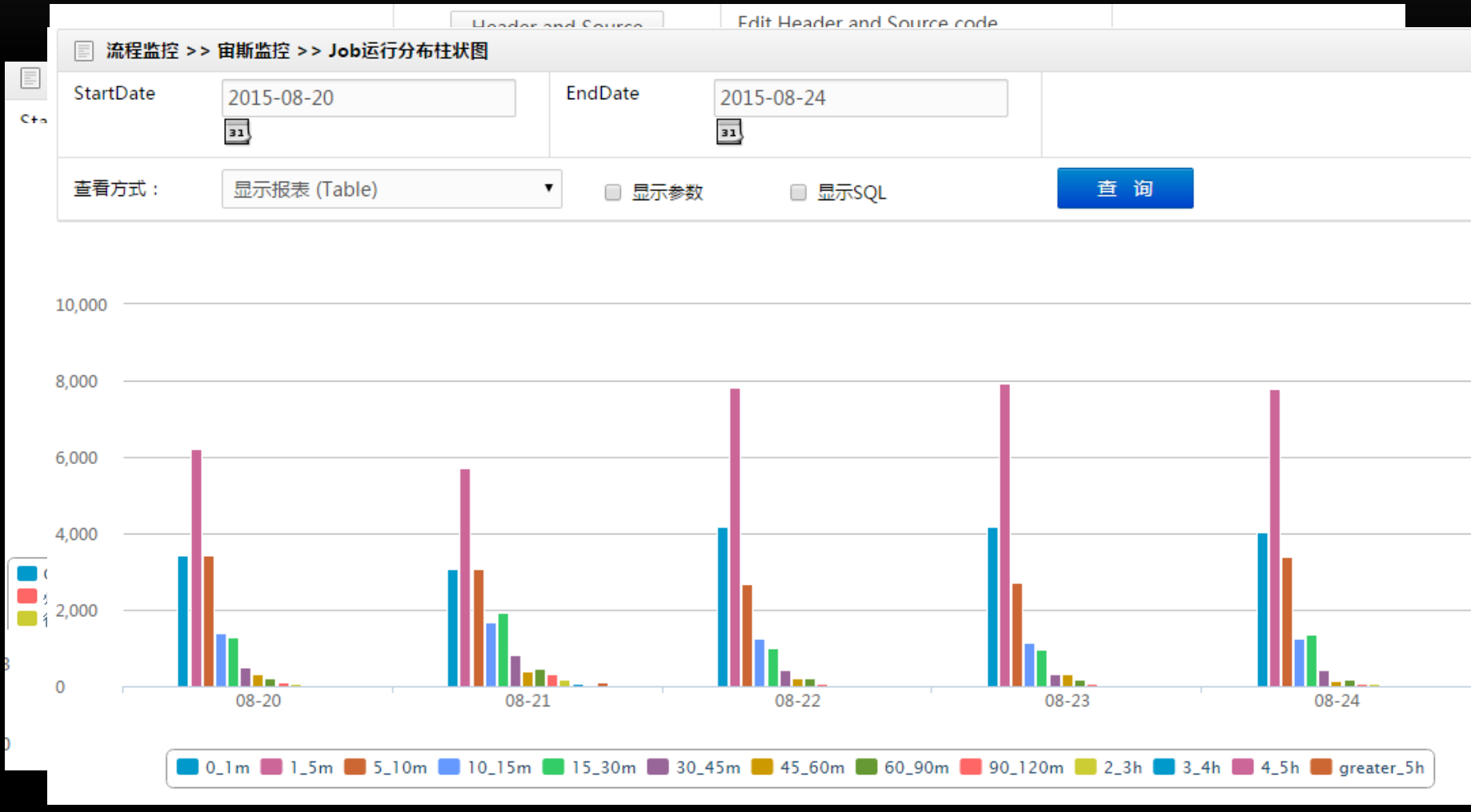
简单报表引擎ART

- 多种数据源
- 多种图例（饼图，柱图，线图等）
- 定时发送邮件，附带pdf/png附件
- 简单易用
- 分级权限管理



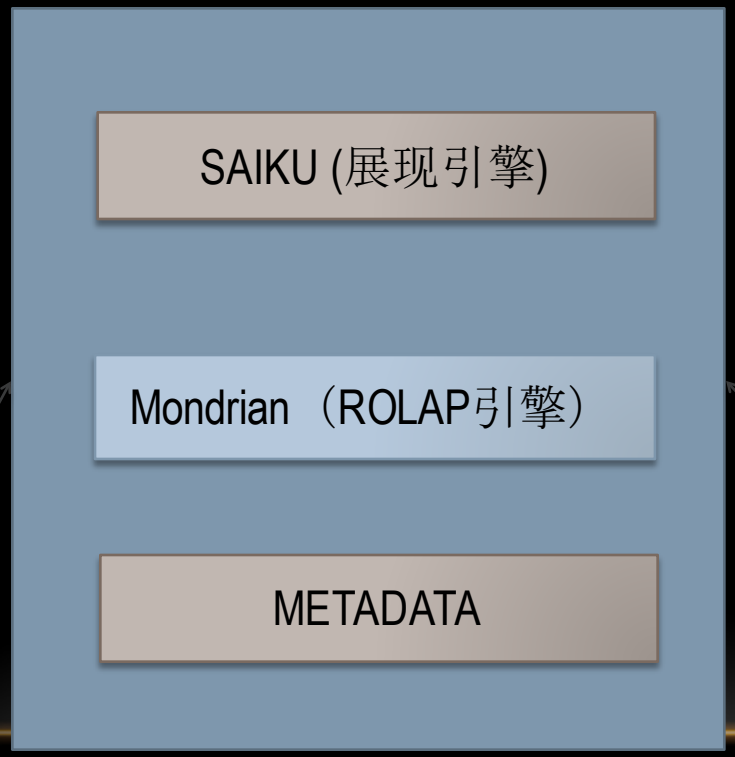





流程监控 >> 宙斯监控 >> Job运行分布柱状图
StartDate
2015-08-20
EndDate
2015-08-24
查看方式：
显示报表 (Table)
显示参数
显示SQL
查 询
10,000
8,000
6,000
4,000
2,000
0
08-20
08-21
08-22
08-23
08-24
0_1m
1_5m
5_10m
10_15m
15_30m
30_45m
45_60m
60_90m
90_120m
2_3h
3_4h
4_5h
greater_5h

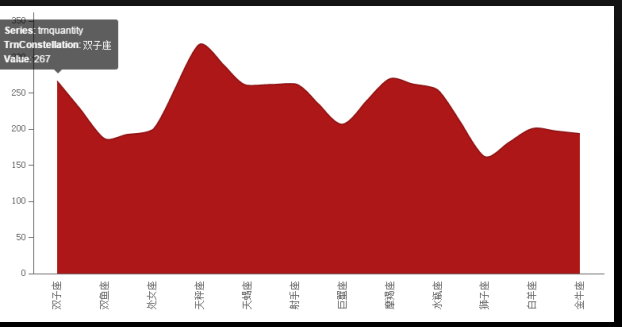
多维分析引擎SAIKU

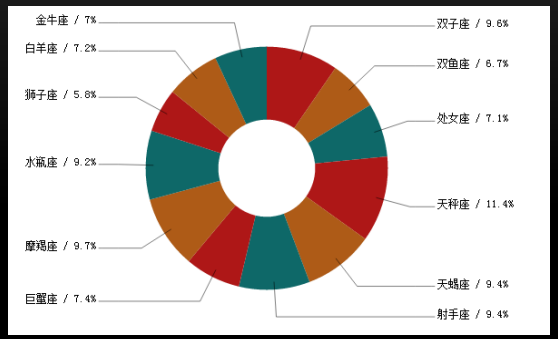
- 上钻，下钻
- 拖拽多维分析
- 多种图例
- 支持多种数据源

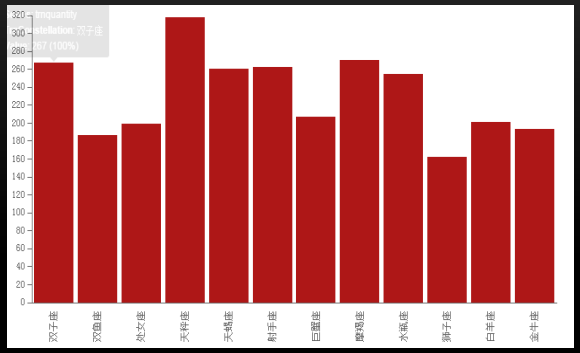


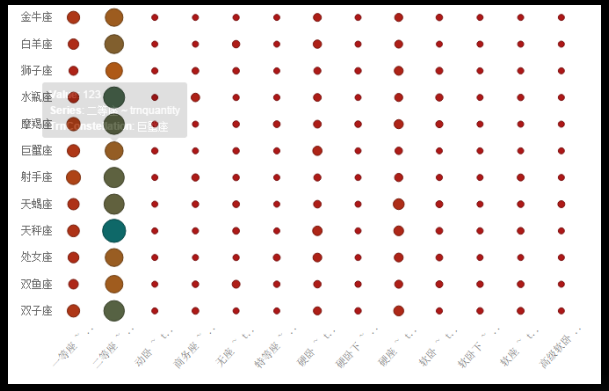
小数据量

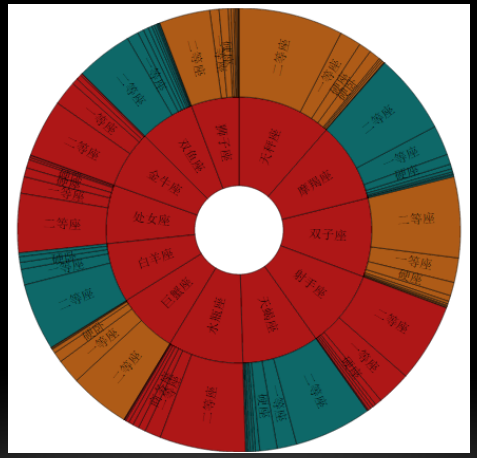
大数据量

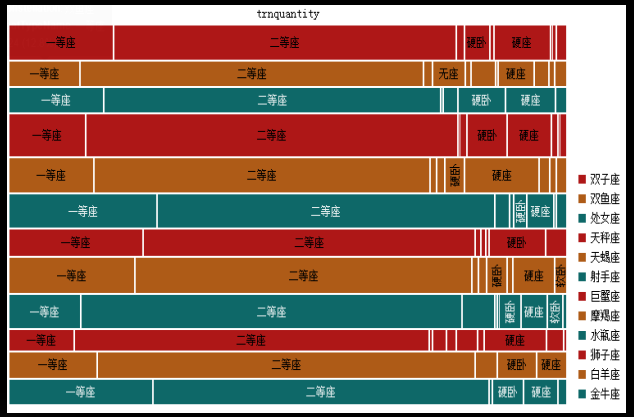

Series: trnquantity
TrnConstellation: 双子座
Value: 267
双子座
双鱼座
处女座
天秤座
天蝎座
射手座
巨蟹座
摩羯座
水瓶座
狮子座
白羊座
金牛座


金牛座 / 7%
白羊座 / 7.2%
狮子座 / 5.6%
水瓶座 / 9.2%
摩羯座 / 9.7%
巨蟹座 / 7.4%
双子座 / 9.6%
双鱼座 / 6.7%
处女座 / 7.1%
天秤座 / 11.4%
天蝎座 / 9.4%
射手座 / 9.4%


双子座
双鱼座
处女座
天秤座
天蝎座
射手座
巨蟹座
摩羯座
水瓶座
狮子座
白羊座
金牛座


金牛座
白羊座
狮子座
水瓶座
摩羯座
巨蟹座
射手座
天蝎座
天秤座
处女座
双鱼座
双子座




trnquantity
双子座
双鱼座
处女座
天秤座
天蝎座
射手座
巨蟹座
摩羯座
水瓶座
狮子座
白羊座
金牛座

谢谢大家