

—— World Of Tech 2017 ——

全球架构与运维技术峰会

2017年4月14日-15日 北京富力万丽酒店

ARCHITECTURE



出品人及主持人：

李鹏涛 京东商城
青龙研发部高级总监

大数据应用创新



苏宁易购

多维报表平台探索之路



王富平

苏宁

大数据中心高级架构师

分享主题：
多维分析平台实践

CONTENT

为什么我们要自研多维报表平台？ 01

苏宁多维报表平台方案概述 02

olap引擎技术选型 03

报表平台性能优化篇 04

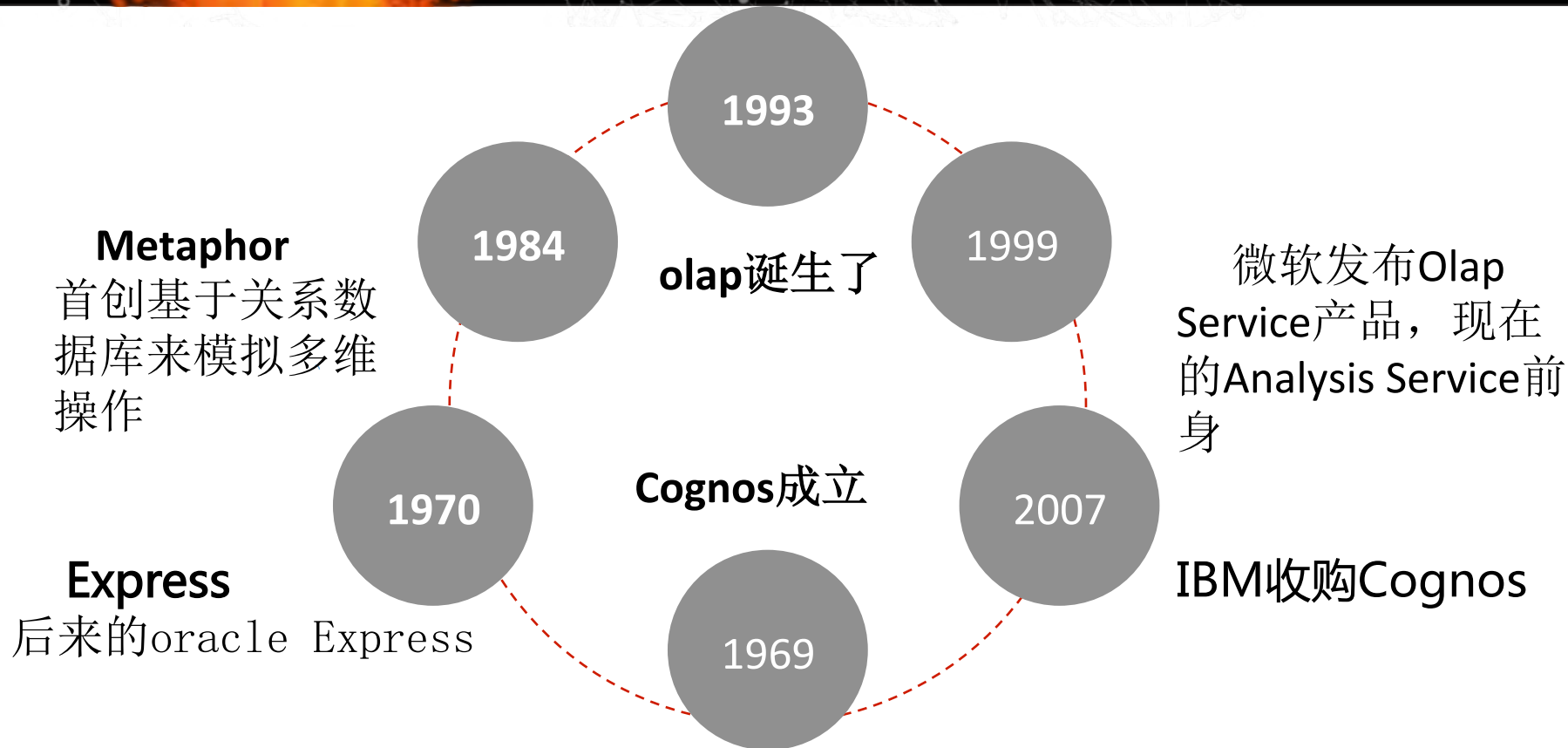
业务策略和架构选择 05

未来展望 06

01

为什么我们要自研多维报表平台

多维报表分析的发展史



苏宁多维分析发展史



苏宁早期采购WEBI系统，
来配合SAP系统使用，
WEBI系统已于16年停止
维护。

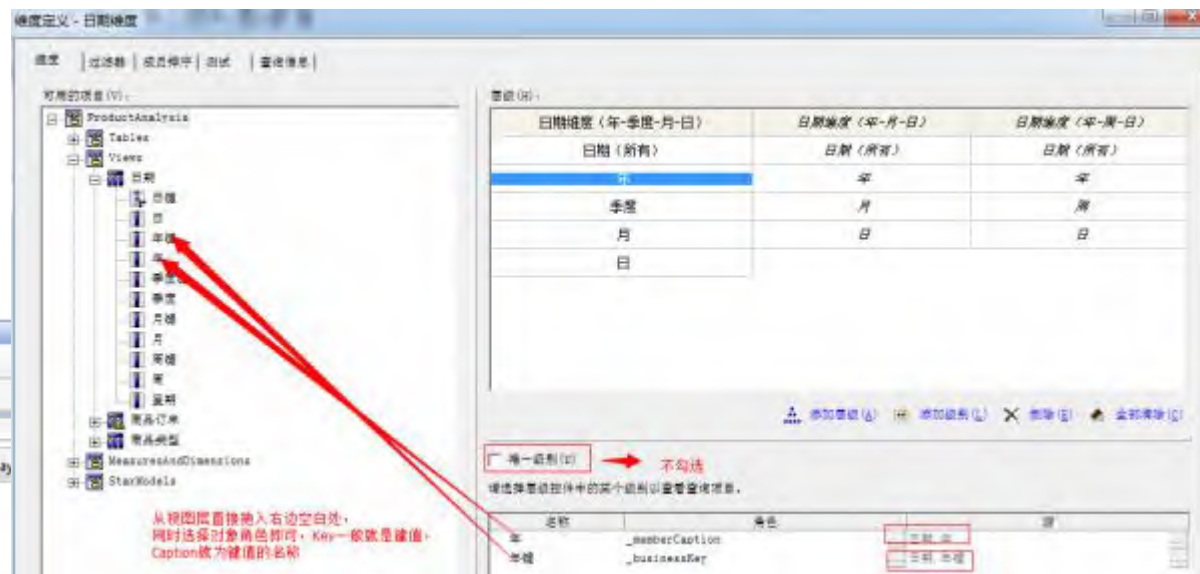
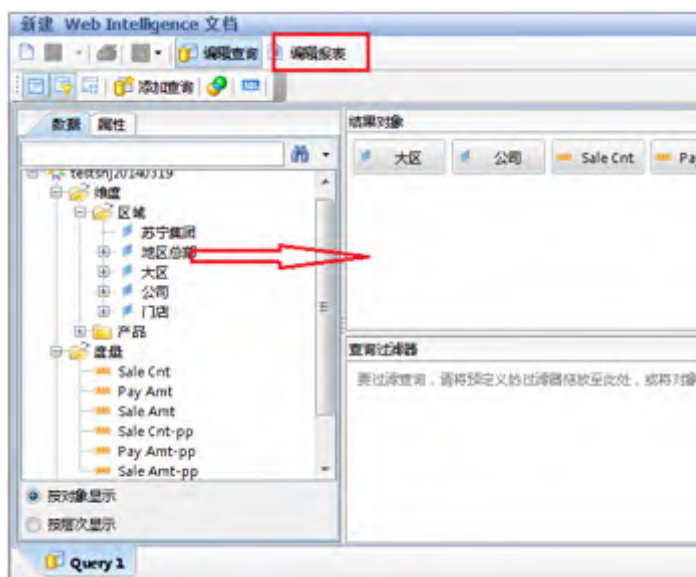


Cognos BI

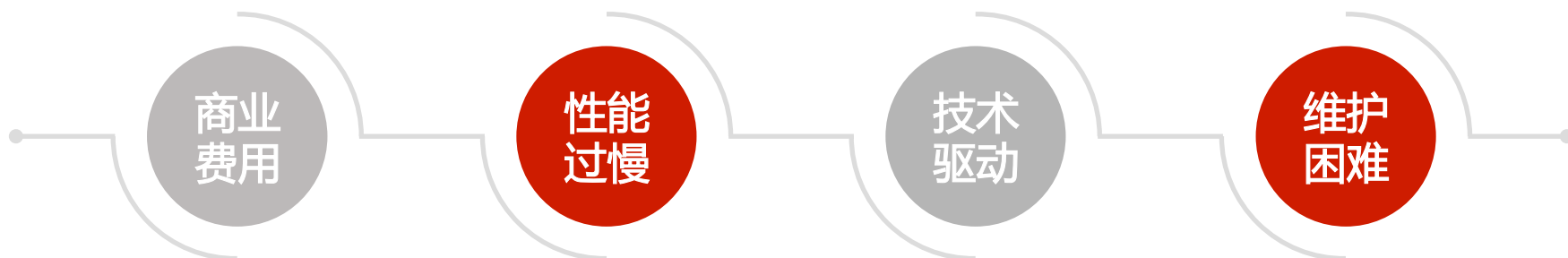
苏宁在13年采购Cognos8
版本，目前cognos上面
还有2-3千张报表，每天
定时生成。



Cognos BI



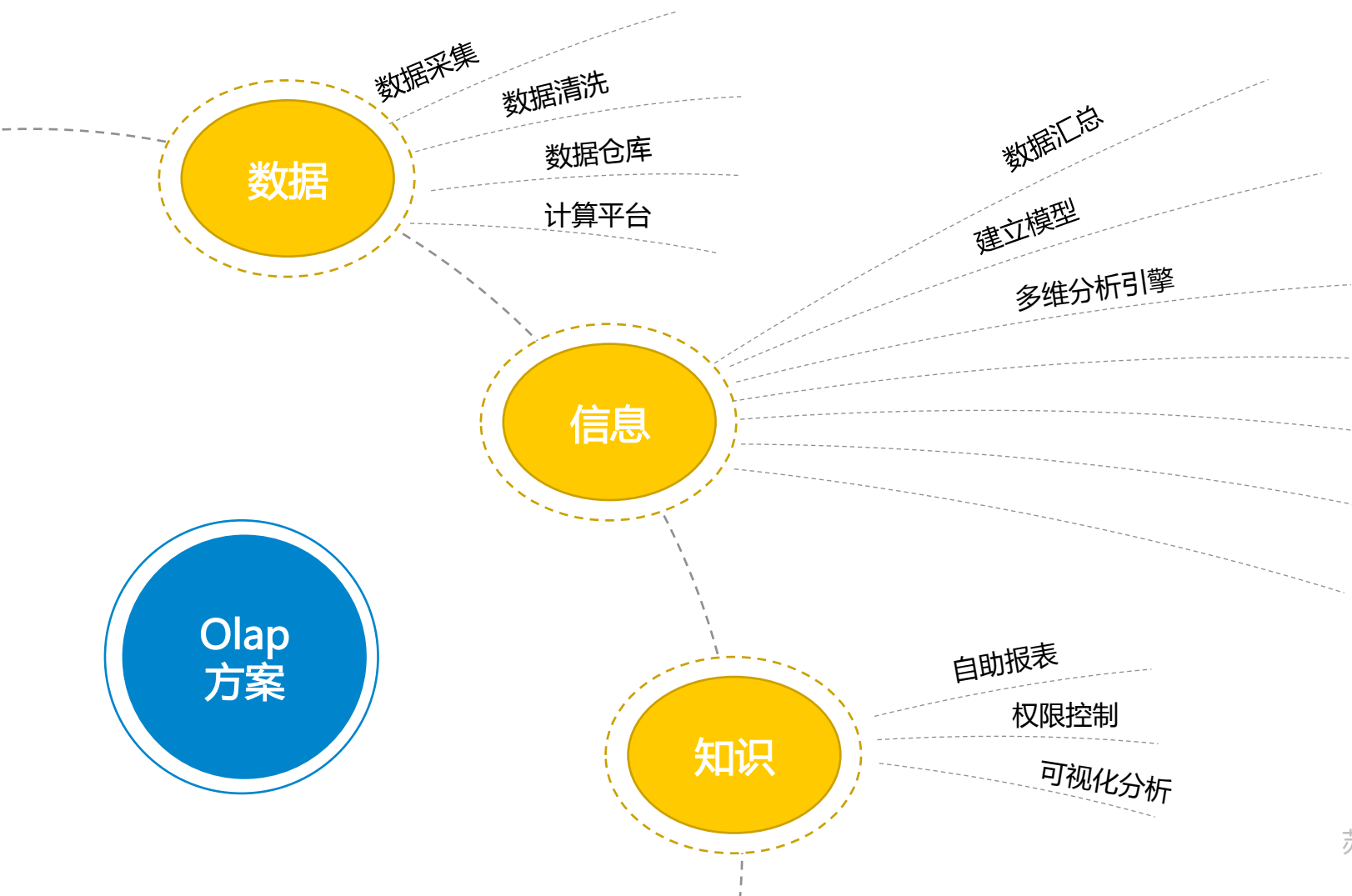
为什么决定自研多维报表平台



02

苏宁多维报表平台方案概述

苏宁多维报表平台设计理念



产品展示层

诸葛门户-PC
大数据中心诸葛门户-APP
大数据中心大屏 ?
分工待确认 ?APP(店+)等
BI

报表工具层

自助报表工具-天工

数据服务层

统一数据服务平台-百川
大数据中心(平台)/BI(指标)

数据集市层

消费者/会员集市-BI

广告集市-BI

物流集市-BI

中台集市-BI

搜索集市-BI

DB2/
MySQL/
Hbase
云计算

智能引擎集市-大数据中心

供应商集市-BI

员工集市-BI

文创集市-BI

数据仓库层

DM应用层--大数据中心/BI

DPA汇总层--大数据中心

SOR基础层--大数据中心

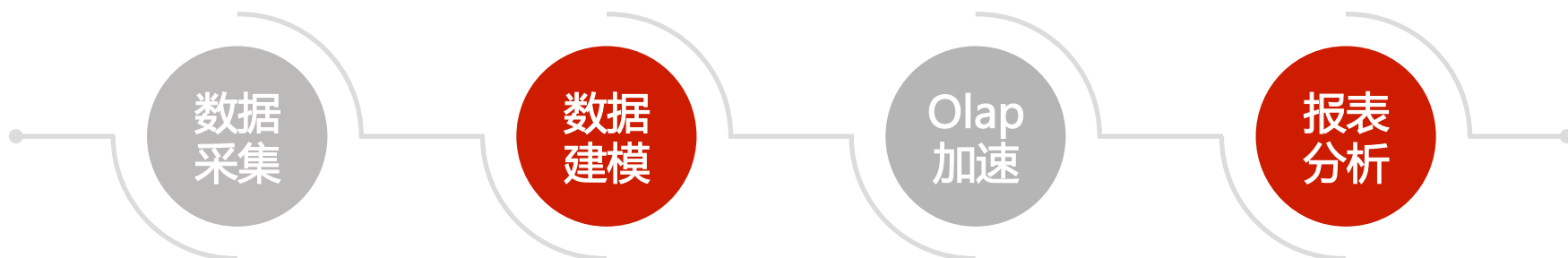
SSA缓冲层--大数据中心/BI

Hadoop计
算平台
云计算CBT调度平
台
云计算

业务系统

SSA(采集)
大数据中心SDK(采集)
大数据中心订单/库存/会员...
中台研中心网站/门店...
消费者研发中心

多维报表平台整体流程



03

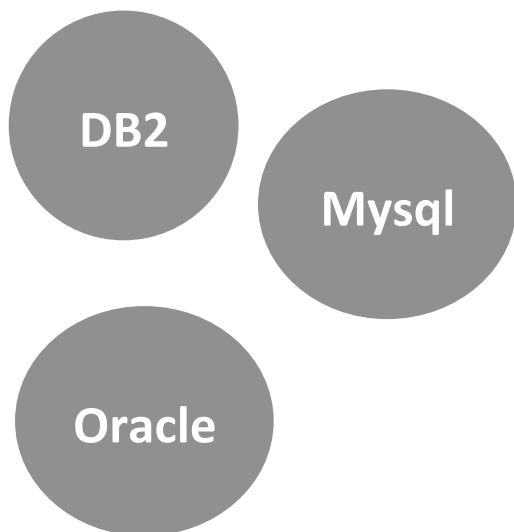
Olap引擎技术选型



Olap引擎选择理念

- 高性能：可扩展、支撑大数据场景
- 实时性：满足实时分析需求
- Sql接口：支持sql语法查询

传统数据库



Mpp解决方案



新技术方案





Olap引擎选择理念

适配所有

传统数据库

DB2

Mysql

Oracle

Mpp解决方案

Green
Plum

新技术方案

ES

Druid

Kylin



Druid介绍

- 实时计算
- 时间分区
- 列式存储
- 聚合计算 : count、distinct Count、max、min、sum

Timestamp		Dimensions			Metrics	
Timestamp	Page	Username	Gender	City	Characters Added	Characters Removed
2011-01-01T01:00:00Z	Justin Bieber	Boxer	Male	San Francisco	1800	25
2011-01-01T01:00:00Z	Justin Bieber	Reach	Male	Waterloo	2912	42
2011-01-01T02:00:00Z	Ke\$ha	Helz	Male	Calgary	1953	17
2011-01-01T02:00:00Z	Ke\$ha	Xeno	Male	Taiyuan	3194	170



Druid高效的原因

- 列式存储，快速维度过滤
- Rollup压缩
- 高效的聚合算法

1: Dictionary that encodes column values

```
{  
  "Justin Bieber": 0,  
  "Ke$ha":        1  
}
```

2: Column data

```
[0,  
 0,  
 1,  
 1]
```

3: Bitmaps – one for each unique value of the column

```
value="Justin Bieber": [1,1,0,0]  
value="Ke$ha":        [0,0,1,1]
```



Druid的sql支持

- Druid本身没有官方sql接口

mea1	有效订单数	count(valid_order_no_cnt)
mea2	短信发送率	count(cis_order_no_cnt) * 100.0 / (count(valid_order_no_cnt) + 0.1)



CALCITE



ES介绍

- 数据分片
- 高效索引
- 聚合计算 : count、distinct Count、 max、 min、 sum

Lucence

分布式
高扩展



ES-Sql支持

- ES本身无官方sql接口

维度	描述	表达式
dim0	登录类型	bi_appaction.logtype

度量	描述		表达式
	描述	单位	
mea0	次数		count(*)



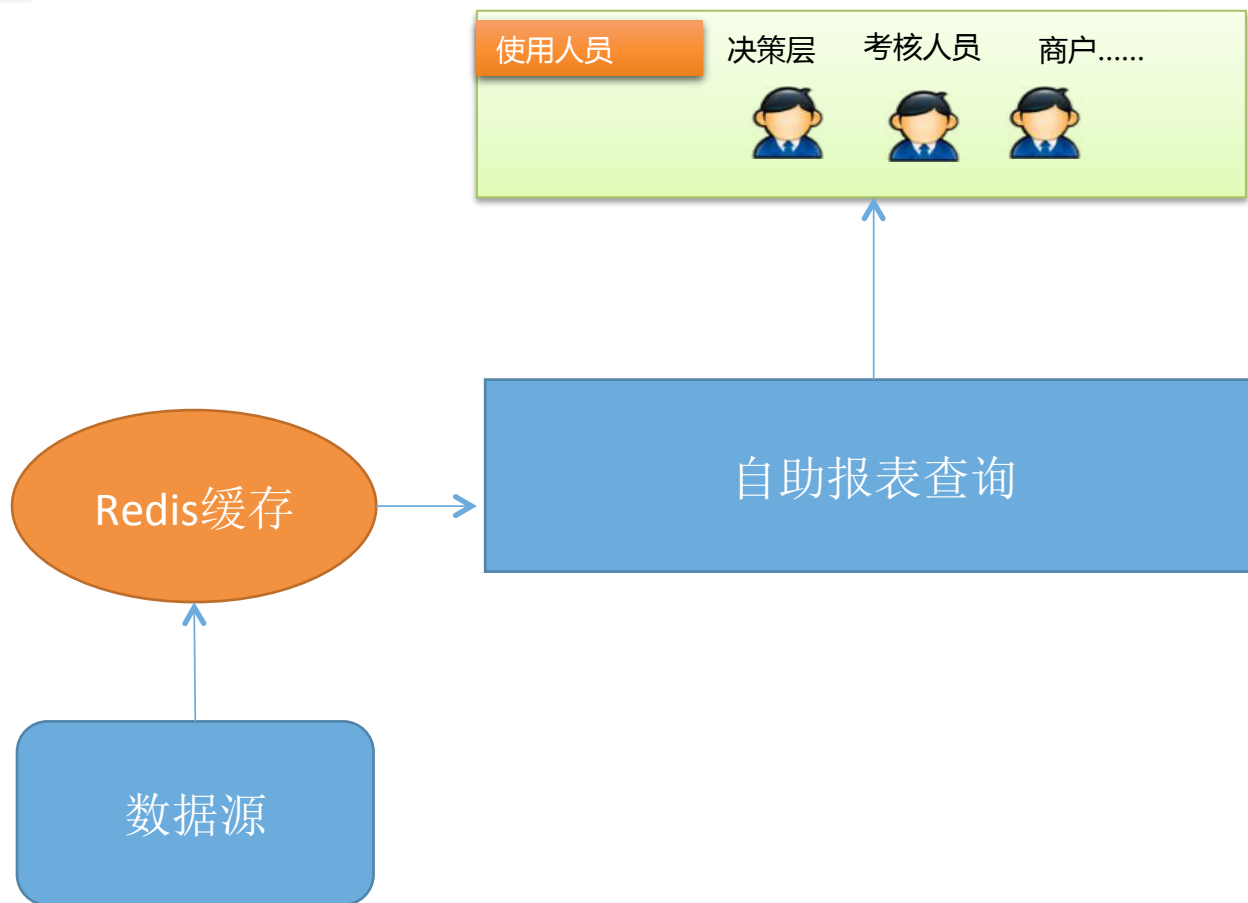
ElasticSearch-sql

后续使用
CALCITE

04

报表平台性能优化篇

缓存策略-定时更新

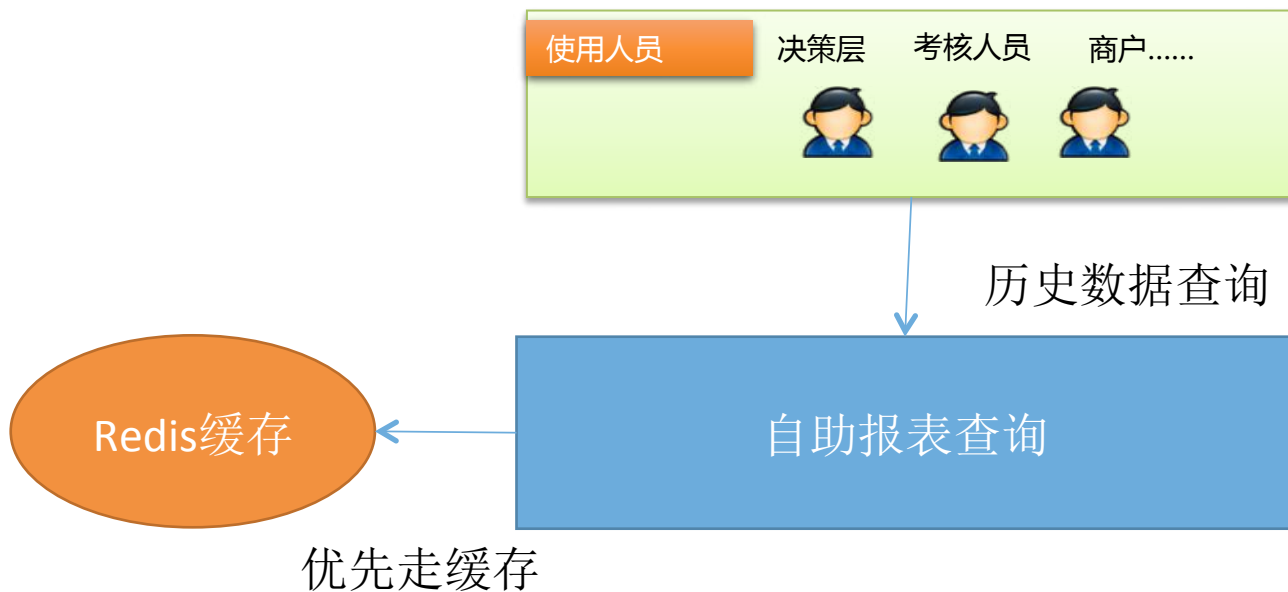


5分钟定时更新



缓存策略-历史缓存

- 数据Append only：日志等
- 一旦请求发生，会进行长时间缓存





缓存策略-主动缓存

- 统计数据使用习惯，对常用维度组合提前计算

Timestamp	Dimensions				Metrics	
Timestamp	Page	Username	Gender	City	Characters Added	Characters Removed
2011-01-01T01:00:00Z	Justin Bieber	Boxer	Male	San Francisco	1800	25
2011-01-01T01:00:00Z	Justin Bieber	Reach	Male	Waterloo	2912	42
2011-01-01T02:00:00Z	Ke\$ha	Helz	Male	Calgary	1953	17
2011-01-01T02:00:00Z	Ke\$ha	Xeno	Male	Taiyuan	3194	170

05

业务策略与架构选择



权限控制

- 如何给8000人开一张报表的权限？

30+大区

500+城市公司

1500+门店

1300+易购服务站



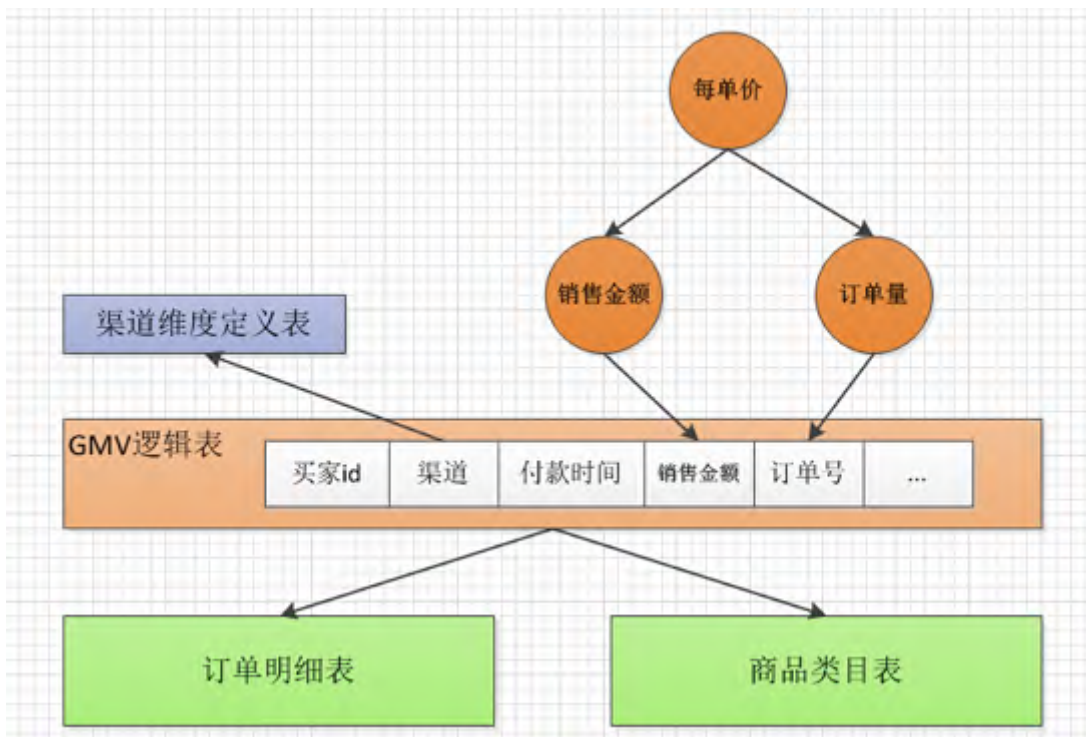
权限控制

- 维护报表权限维度：大区、城市公司、门店等
- 抽取ERP中人员身份信息，进行权限自动匹配

维度	描述	表达式
dim0	数据日期	__time
dim1	业务渠道	serv_unit_cd
dim2	明细渠道	distChannel
dim3	地区编码	region_cd
dim4	大区级别	gj_area_lvl_cd
dim5	大区编码	gj_areaCode
dim6	公司编码	gj_orgCode

衍生指标计算-同比、环比、占比

- 基础指标：由当前数据直接计算得出，无需其它数据
- 衍生指标：一般由多个基础指标数据（不同指标或不同维度）计算得出，甚至需要第三方数据确定逻辑：如判断哪一天是同比天需要第三方逻辑数据





衍生指标计算

- 复购率
- 财务指标： $(xx\text{产品结构实际进销差率} - \text{其他产品结构进销差额} / \text{其他产品结构付款金额}) * \text{变动比例} (\text{销售占比平均值} - \text{销售占比实际值}) * \text{产品总付款金额}$

预计算

06

未来展望



功能完善

- sql统一由calcite支撑
- 多模型组合



可视化建设

- 可视化能力

物流首页1	切换模型	未命名的工作表	分析面板
维度	表格 图表		行
Str. 大小件	大小件 地区名称 当日应送达作业单量		大小件 地区名称
Str. 地区级别			当日已完成量
Str. 地区名称			
Str. 物流排名			列
Str. 用户编码			当日应送达作业单量 当日已完成量
Date. 交货日期(旧)			
度量			
No. 当日应送达作业单量			
No. 当日已完成量			
No. 仓储			
No. 分拨			
No. 运输			

Thank you !