

MEIZU

大数据

可视化平台建设之路

- 赵天烁 2018. 03. 26

CONTENTS

01

现状&问题

02

当前目标

03

实现方案

04

未来规划

目录

CONTENTS

01

现状&问题

数据接入

- 渠道（形式）多样化
- 格式多样
- 脏数据
- 量级巨大
- 建模繁琐

数据质量

- 指标、维度一致性
- 数据延迟
- 血缘关系
- 通知机制

数据查询

- 性能差
- 实时性
- 大规模计算
- 数据挖掘
- 易用性
- 数据更新

数据安全

- 指标、维度权限
- 行级权限
- 参数化鉴权
- 数据交换脱敏
- 日志审计

组件

- 类型扩展
- 视觉定制
- 条件样式
- 联动
- 自定义组件

多终端

- 移动端适配
- 大屏适配
- PPT式数据报告
- 自定义主题
- 多端分享互动

功能特性

- 拖拽操作
- 自由排版
- 上卷下钻
- 区域分组联动
- 数据定义说明

通知预警

- 底层数据异常通知
- 底层数据恢复通知
- 业务波动通知
- 定时邮件

目录

02

当前目标

1. 基础功能
2. 灵活扩展
3. 体验优化
4. 平台集成
5. 场景封装

组件？工具？平台？解决方案？生态？

结论

CONCLUSION

01

基础功能

02

可扩展性

03

工具->平台

04

上下游集成

目录

CONTENTS

03

实现方案

为什么不直接购买商业产品？

PART ONE

01

定制

- 视觉定制
- 交互细节定制
- 可视化类型
- 功能定制



PART TWO

02

产品集成

- 安全集成
- 分析引擎集成
- 数据化运营产品集成



PART THREE

03

复用

- 数据集模型复用
- 数据接口复用
- 可视化组件复用
- 页面复用



PART FOUR

04

快速响应

- 功能需求
- 内容需求
- 环境运维
- 查询性能



数据接入

- 渠道（形式）多样化
- 格式多样
- 脏数据
- 量级巨大
- 建模繁琐

数据访问

- 性能差
- 易用性
- 实时性
- 数据更新
- 大规模计算
- 数据挖掘

数据质量

- 指标、维度一致性
- 数据延迟
- 血缘关系
- 通知机制

数据安全

- 指标、维度权限
- 行级权限
- 参数化鉴权
- 数据交换脱敏
- 日志审计

数据访问

分析引擎&数据交换&模型集市

PART ONE

实时计算

- Spark Streaming
- Storm
- Tindex

PART TWO

OLAP

- Kylin
- Vertica
- Tindex

PART THREE

即系查询

- Kylin
- Vertica
- mysql分库分表
- CarbonData

PART FOUR

文本检索

- Elastic Search
- Tindex

是否引入统一SQL访问层？

统一
SQL
访问层

权限过滤

日志审计

请求路由

性能优化

标准化

VS

成本高

周期长

兼容性

迁移成本

必要性

PART ONE

实时计算

- Spark Streaming
- Storm
- Tindex

PART TWO

OLAP

- Kylin
- Vertica
- Tindex

PART THREE

即系查询

- Kylin
- Vertica
- mysql分库分表
- CarbonData

PART FOUR

文本检索

- Elastic Search
- Tindex

No



统一
概念
模型

字
段

指
标

维
度

参
数

动
态
条
件

分
页

数
据
导
出

脚
本
语
言

My
Bati
s

条
件
标
签

PART ONE

实时计算

- Spark Streaming
- Storm
- Tindex

PART TWO

OLAP

- Kylin
- Vertica
- Tindex

PART THREE

即系查询

- Kylin
- Vertica
- mysql分库分表
- CarbonData

PART FOUR

文本检索

- Elastic Search
- Tindex

数据交换&模型集市？

可视化



Dashboards



Dashboard	Creator	Modified
Misc Charts		a minute ago
Births		a minute ago
World's Bank Data		a minute ago

[< Previous](#) | 1 | [Next >](#)



List Slice






Search

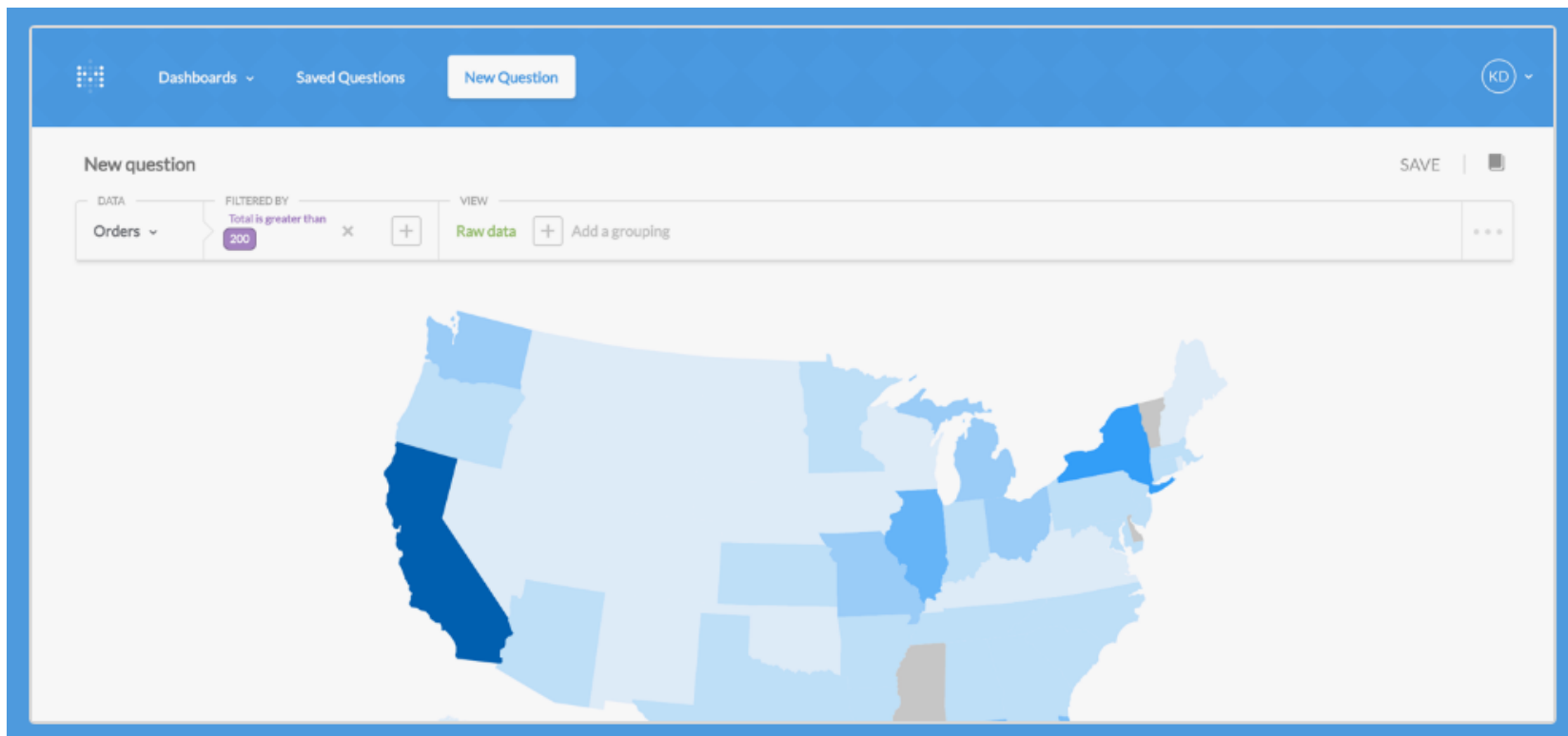


Actions



Record Count: 34

<input type="checkbox"/>		Slice ↑	Visualization Type ↑	Datasource ↑	Creator ↑	Last Modified ↑
<input type="checkbox"/>	  	Calendar Heatmap multiformat 7	cal_heatmap	multiformat_time_series		2 hours ago
<input type="checkbox"/>	  	Calendar Heatmap multiformat 6	cal_heatmap	multiformat_time_series		2 hours ago
<input type="checkbox"/>	  	Calendar Heatmap multiformat 5	cal_heatmap	multiformat_time_series		2 hours ago
<input type="checkbox"/>	  	Calendar Heatmap multiformat 4	cal_heatmap	multiformat_time_series		2 hours ago
<input type="checkbox"/>	  	Calendar Heatmap multiformat 3	cal_heatmap	multiformat_time_series		2 hours ago
<input type="checkbox"/>	  	Calendar Heatmap multiformat 2	cal_heatmap	multiformat_time_series		2 hours ago
<input type="checkbox"/>	  	Calendar Heatmap multiformat 1	cal_heatmap	multiformat_time_series		2 hours ago
<input type="checkbox"/>	  	Calendar Heatmap multiformat 0	cal_heatmap	multiformat_time_series		2 hours ago
<input type="checkbox"/>	  	Mapbox Long/Lat	mapbox	long_lat		2 hours ago
<input type="checkbox"/>	  	Calendar Heatmap	cal_heatmap	random_time_series		2 hours ago
<input type="checkbox"/>	  	Number of Girls	big_number_total	birth_names		2 hours ago



Metabase interface showing a "New question" configuration screen. The screen displays a table of user data with columns: Id, Name, Birth Date, City, Created At, Email, Latitude, Longitude, Source, and State. A modal window is open for filtering the "Created At" column, showing options for "Relative date" and "Specific date".

Table Data:

Id	Name	Birth Date	City	Created At	Email	Latitude	Longitude	Source	State	
1	Stanton	1977-02-05	North Raphaelle	2015-08-14T16:55:02.157Z	welch-stanton@gmail.com	63.95800697	77.97429579	Twitter	FL	
2	Otha	1977-12-20	Lednerland	2015-03-12T00:53:50.013Z	wehner.otha@yahoo.com	82.13585075	-167.74042751	Affiliate	AE	
3	Aglae	1963-08-01	Nikolasmouth	2015-02-02T22:12:04.825Z	aglae-koeppe@yahoo.com	-36.63332139	-121.32349590	Organic	AK	
4	Brandt	1964-04-18	South Justine	2015-07-30T05:54:36.234Z	brandt.brueen@yahoo.com	-65.03265230	-94.17265055	Twitter	NE	
5	Armando	1997-04-26	Mathildeshire	2015-04-24T04:17:20.467Z	armando.franecki@gmail.com	35.07821712	88.77050394	Google	NJ	
6	Antonina	1994-03-28	Naderstad	2015-01-22T08:51:41.618Z	antonina-bartell@gmail.com	12.44448463	-84.70599745	Organic	WY	
7	Mitchel	1995-06-29	North Green	2015-03-17T01:26:52.207Z	mitchel-kautzer@gmail.com	56.10835567	-176.27781282	Organic	NY	
8	Kassandra Hettinger	8849 Lemke Unions	1993-03-17	New Melvina	2015-04-18T02:53:41.427Z	hettinger.kassandra@yahoo.com	-88.75585261	-132.42751504	Facebook	AP
9	Angelina Mayert	44287 Darwin Plaza	1961-11-07	West Alexys	2015-03-21T03:21:26.275Z	mayert.angelina@gmail.com	-67.49979883	-146.46847766	Affiliate	WA

Filter Modal: User - Created At

- Relative date (selected)
- Specific date
- Today
- Yesterday
- Past 7 days
- Past 30 days
- LAST
- Week
- Month
- Year
- THIS
- Week
- Month
- Year
- Add filter



Superset

- 可视化类型丰富
- 探索式分析+可视化
- Dashboard构建
- SQL编辑器
- 权限控制完善
- 开箱即用
- 支持主流数据源

- 自定义SQL的语义层封装不完善
- 交互体验略复杂
- 可视化视觉效果一般

Metabase

- 探索式分析+可视化
- Dashboard构建
- 开箱即用
- 交互体验简洁，实用
- 视觉设计完成度高
- 支持主流数据源
- SQL编辑器

- 可视化类型较少
- 探索式分析自由度有限
- 权限控制简单

价值
VS
成本

标准化封装
VS
定制化需求



PART ONE

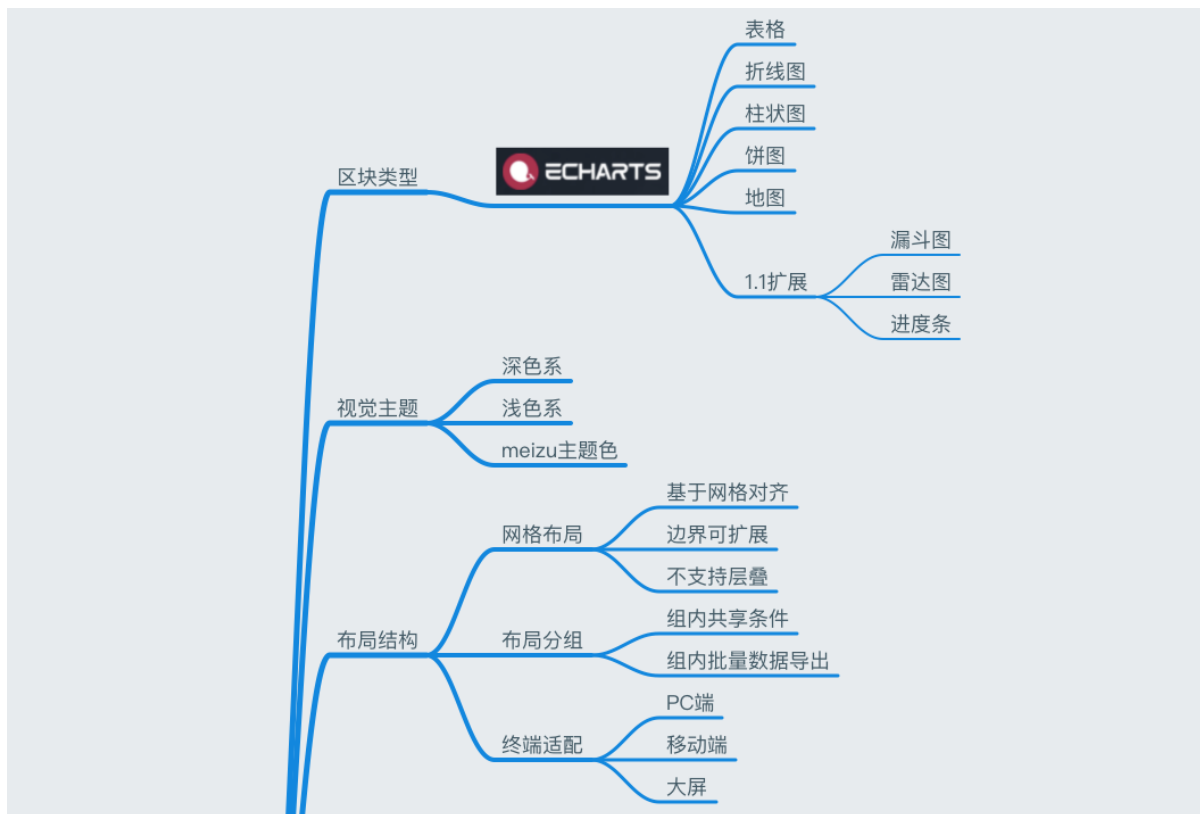
常用

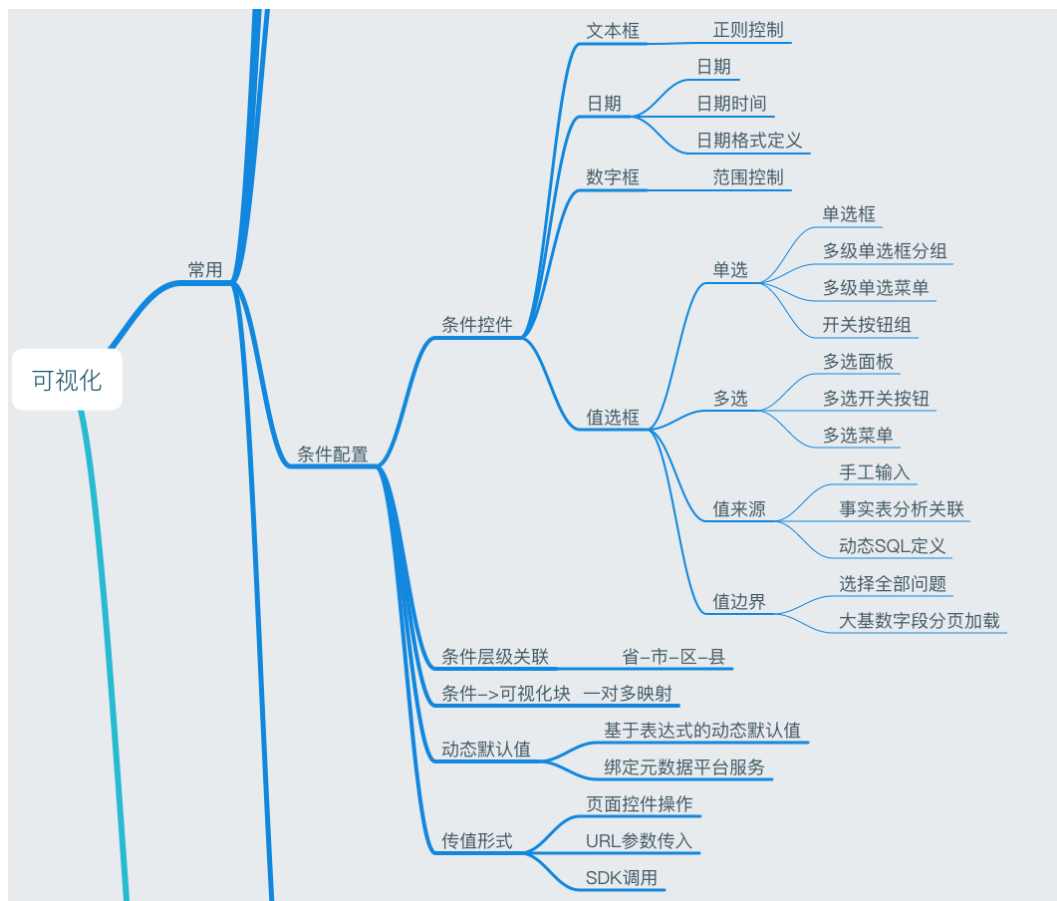
- 只覆盖常见、典型的：
 - ✓ 区块类型
 - ✓ 视觉主题
 - ✓ 布局结构
 - ✓ 功能特性

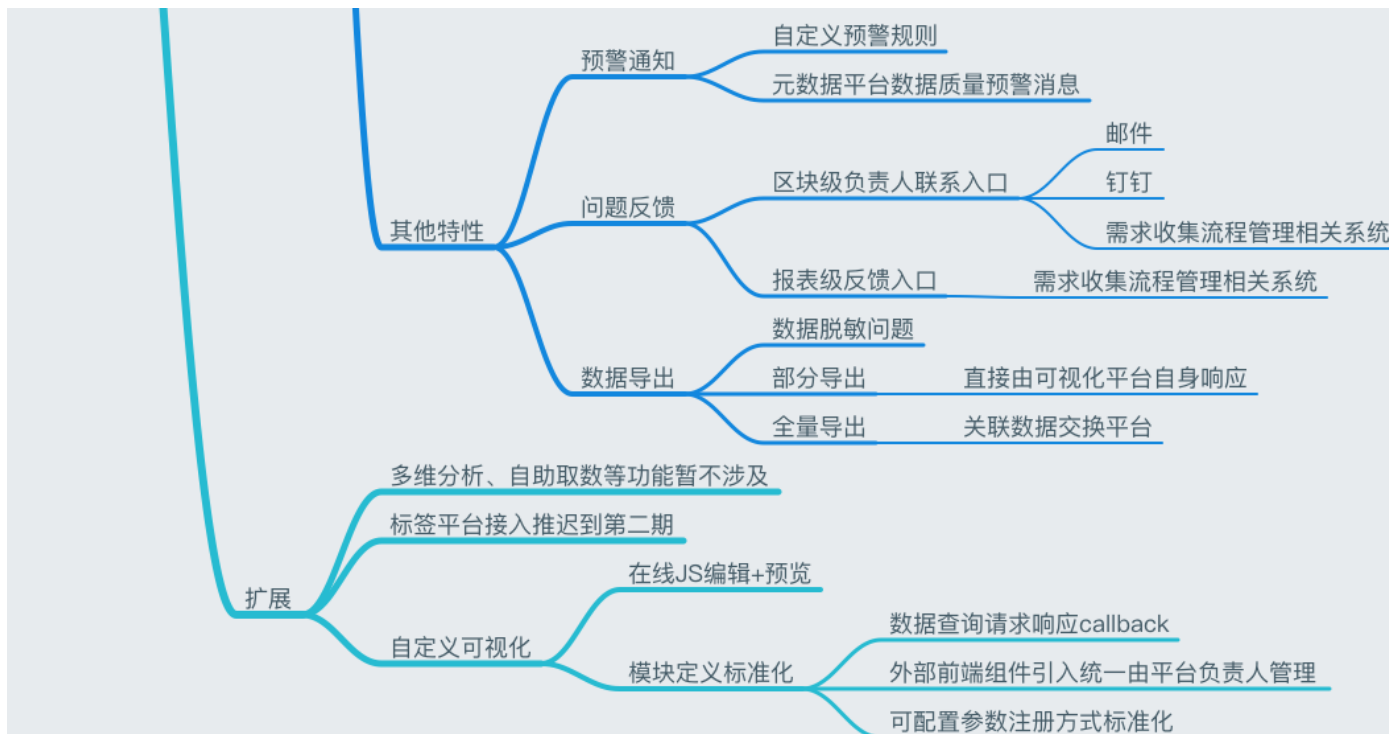
PART TWO

扩展

- 定制化需求支持自定义
- 自定义入口模板化
- 不涉及多维分析等高级特性的封装





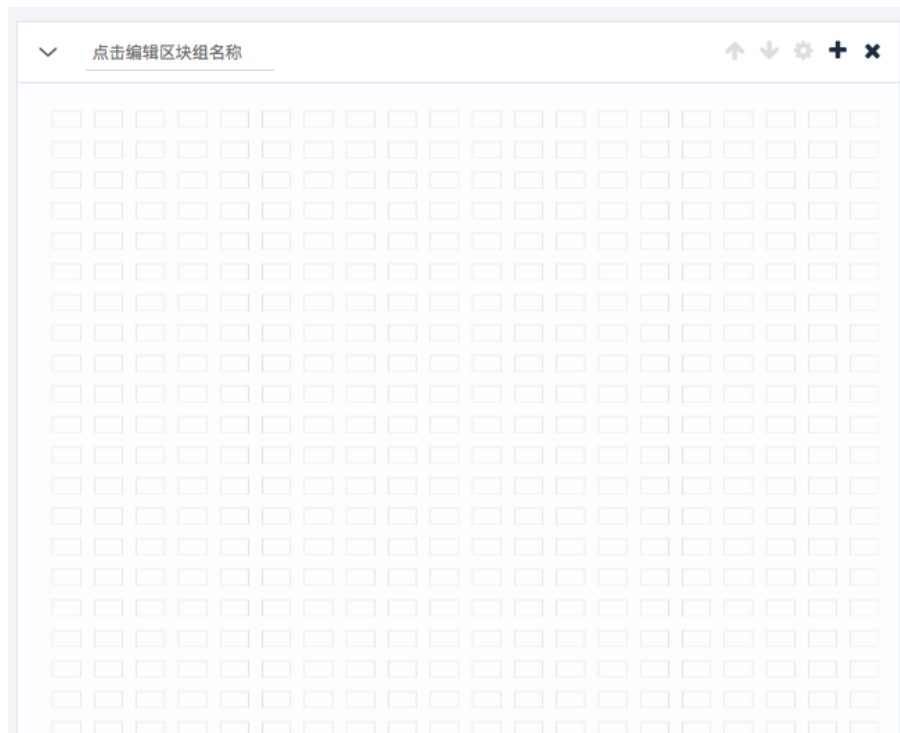


走过的弯路 踩过的坑

- 默认hbox模式
 - ✓ 灵活性太差
- 布局管理器+多套布局策略实现
 - ✓ 兼容性问题
 - ✓ 维护成本高
- 类PPT的绝对位置自由布局
 - ✓ 复杂度过高
 - ✓ 用户体验欠佳

基于网格的 自定义布局

PC、移动端、大屏
独立模板化



集成&服务化



分析引擎 服务化

PART ONE

统一概念模型

- ✓ 降低维护成本
- ✓ 提升复用性
- ✓ 数据质量统一管理
- ✓ 性能可用性统一优化
- ✓ 安全日志审计

PART TWO

Client封装

- 服务ID+条件参数结构
- 数据获取细节在平台中定义
- 支持同步、异步方式
- 异步方式支持消息通知

```
Result<List<Map<String, Object>> result=XXXService.query(816, params)  
Result<List<VisitorDetail> result2=XXXService.query(819, params)
```

可视化块 SDK

PART ONE

统一概念模型

- ✓ 避免重复开发
- ✓ 统一安全测试
- ✓ 统一公共组件
- ✓ 促进交流/共享
- ✓ 推进DPL落地

PART TWO

SDK封装

- ✓ 组件ID+条件参数结构
- ✓ 可视化块实现在平台中配置/开发实现
- ✓ 支持定时刷新
- ✓ 支持参数化主题切换

```
Charts.load(917,condition,'html_element')
```

元数据平台 集成

PART ONE

模型共享

- ✓ 各层模型元数据可选择性接入
- ✓ 指标、维度定义标注
- ✓ 基于血缘关系的定义继承
- ✓ 沉淀高质量数据集市

PART TWO

质量管理

- ✓ 数据延迟通知
- ✓ 数据异常波动通知
- ✓ 基于血缘关系通知
- ✓ 基于血缘关系的通知升级

权限平台 集成

PART ONE

结合元数据

- ✓ 对数据源访问直接对接元数据
- ✓ 同数据源相同表在各平台中权限一致

PART TWO

访问端透明

- ✓ 支持JDBC代理驱动鉴权
- ✓ 支持URL+参数鉴权

The End