

数字转型 架构演进

SACC

2019 中国系统架构师大会

SYSTEM ARCHITECT CONFERENCE CHINA 2019



2019年10月31-11月2日

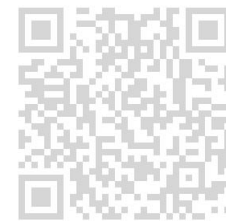


北京海淀永泰福朋喜来登酒店



全新IT技术私域交流平台

贝壳数据可视化分析平台实践



01



自我介绍

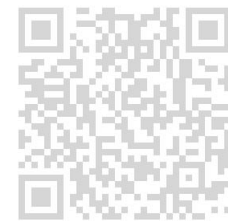


全新IT技术私域交流平台

自我介绍

姓名：彭建

公司/事业部：贝壳找房（北京）科技有限公司/数据智能中心



全新IT技术私域交流平台

02



贝壳可视化分析平台历程



全新IT技术私域交流平台

贝壳可视化分析平台建设历程

问题

- 效率低
- 不利于平台化



城市站及
各个业务方

指标平台（基于Kylin）

Adhoc（基于Spark）

各种API

大数据

优点

- 查询效率高
- 计算口径管理

缺点

- 学习和使用成本高
- 重量级，不利于探索性数据分析

优点

- 支持探索性数据分析

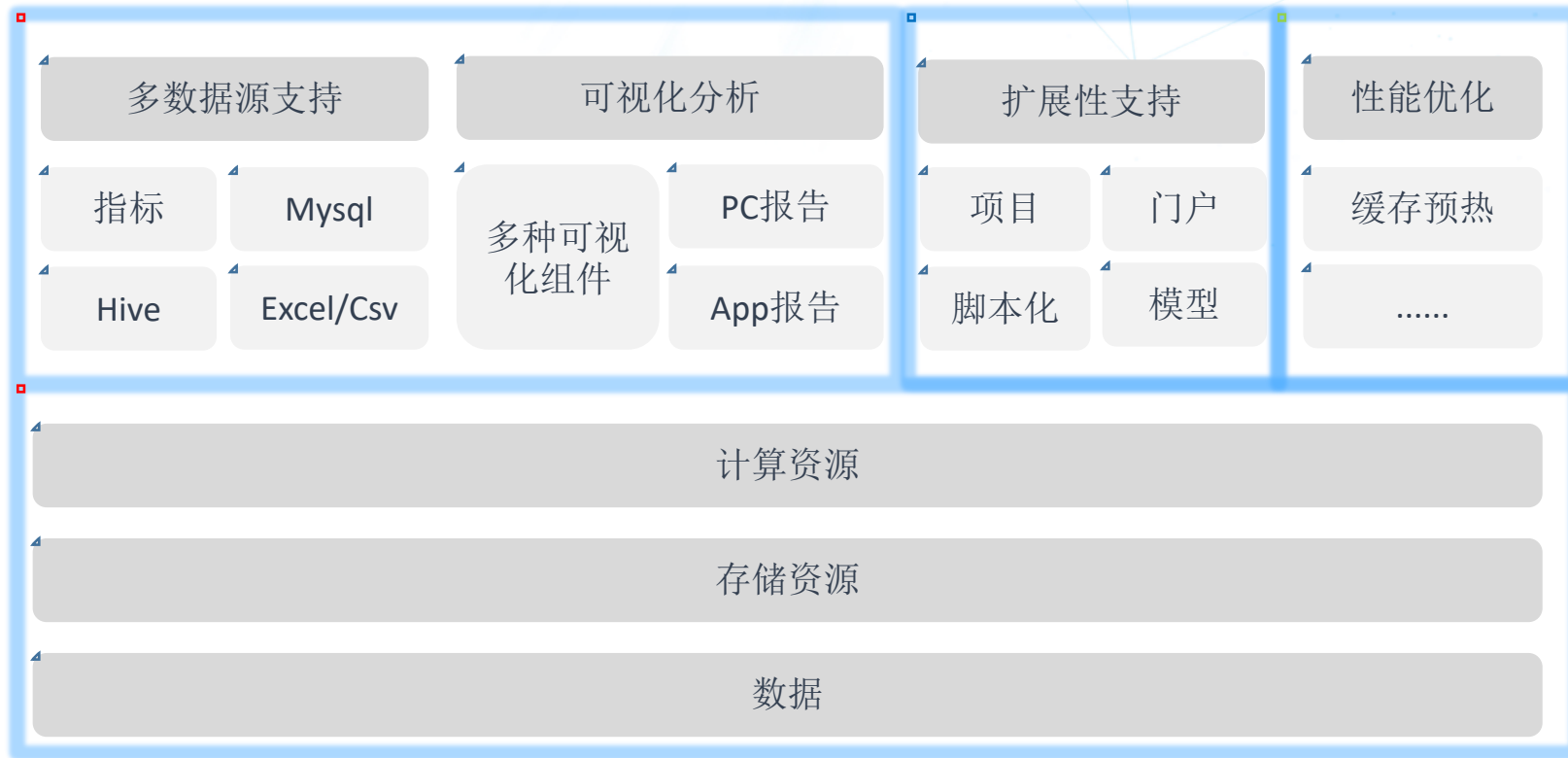
缺点

- 查询效率低
- 口径无法管理



全新IT技术私域交流平台

奥丁可视化分析平台的创建



奥丁

大数据资源

总结奥丁的作用：

- 覆盖了大数据分析的所有功能，实现一站式分析
- 替代以前一个个分析产品，实现降本增效



全新IT技术私域交流平台

03



奥丁的设计与优化

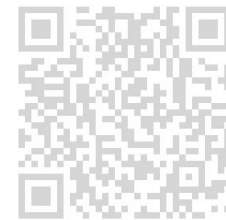


全新IT技术私域交流平台

奥丁的设计与优化

大数据分析系统面临的挑战

- 性能：提升大数据量查询的性能；
- 稳定性：大数据查询消耗资源较多，如何保证并发情况下的资源合理竞争；
- 可扩展性：如何满足各业务线的个性化数据需求；



全新IT技术私域交流平台

整体性能优化

大数据查询特点:

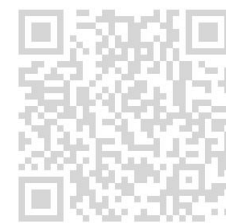
- 实时查询, 上亿数据量、多维度查询。
- 多个大表之间关联查询

性能优化目标:

- 10秒内的查询效率

优化方案和效果:

优化对象	优化方案	优化效果 (10秒内查询效率)
查询引擎Presto	<ul style="list-style-type: none"> • 参数调整 • Alluxio缓存 	qps: 22 --> 33, 10s内响应: 80% -> 93%
应用层	<ul style="list-style-type: none"> • 预热 	<ul style="list-style-type: none"> • 前一天慢日志回放: 97% • 前一周慢日志回放+分阶段回放: 99%
路由层	<ul style="list-style-type: none"> • 独立集群 	效果明显, 在小范围使用



全新IT技术私域交流平台

维度筛选性能优化

大数据维度筛选难点:

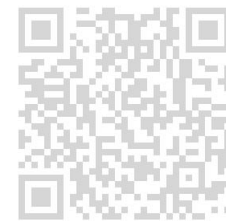
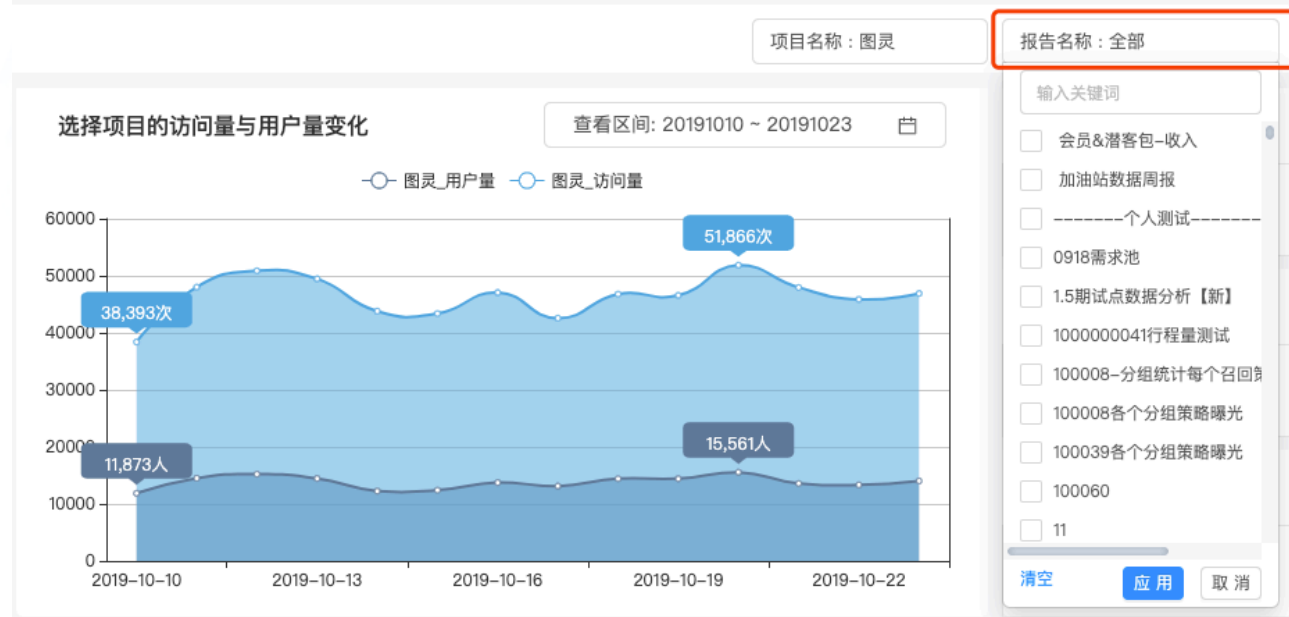
- 对任意表的任意字段支持模糊搜索
- 每列数据量大, 并且每天会变化

方案参考:

- 地动仪: 直接使用like查询
- 指标平台: 为每个公共维度字段建立维度表
- 元数据平台: 将要筛选的列导入到ES中

优化方案:

- select distinct 1000条, 前端模糊搜索
- 加上精确搜索



全新IT技术私域交流平台

稳定性优化

影响稳定性的原因:

- 资源有限

优化思路:

- 控制每个请求消耗的资源

优化方案和效果:

优化对象	优化方案	优化效果
查询范围	自动pt（分区）处理	解决 大多数 情况
模型和用户sql	Explain检查	拦截评估不通过的sql
大数据量下载	资源隔离和自动限流	避免了 少量 大数据量下载拖垮系统

方案总结:

- 通过自动分配、资源检查，避免不合理的资源使用情况发生
- 不增加用户成本，尽量采用自动化处理



全新IT技术私域交流平台

可扩展性优化

可扩展需求:

- 数据分析需求: 对可视化后的数据进一步解读
- 业务动作: 例如, 汇总、预警、颜色高亮等

优化方案: 提供编程接口, 以扩展服务端能力

出于性能、安全考虑, 提供沙箱环境 (可配置):

- 一次请求中, 调用rpc请求远程资源的方法, 不能超过10次
- 对于db操作, 只允许select, 不允许insert等
- db的select操作, 必须包含limit ***限制返回的记录条数
- 等等

月目标 月累计 月完成率

1,745 1,956 112.1%

360 347 96.4%

GMV日增1.1亿, 中部较高1.2亿, 西部较低1.0亿;
当月累计GMV 1.2亿, 北部较高为1.3亿,
东部较低为1.1亿;
当月累计佣金 0.7亿, 西部较高0.8亿, 东部较低0.6亿。

391 486 124.3%

278 501 180.2%

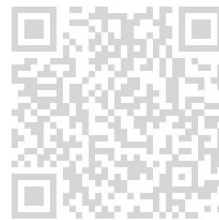
一、业绩概览

用户脚本

沙箱

业务扩展

系统扩展



全新IT技术私域交流平台

成果总结

性能： 总体10s内响应**99%**；
稳定性：**99.9%**，线上没有发生过系统被某个查询耗尽资源的情况；
扩展性： 支持业务方**自定义**扩展；



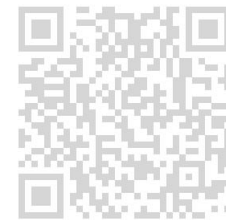
全新IT技术私域交流平台

技术总结

一、用简单有效方案解决99%的问题 + 定制化的方案解决剩下1%的问题

- Hive查询的自动pt处理 + 特殊处理
- 模型的Explain检查 + 物化
- 性能：Presto查询+独立资源查询
- 性能：维度筛选器的前端模糊搜索 + 精确搜索

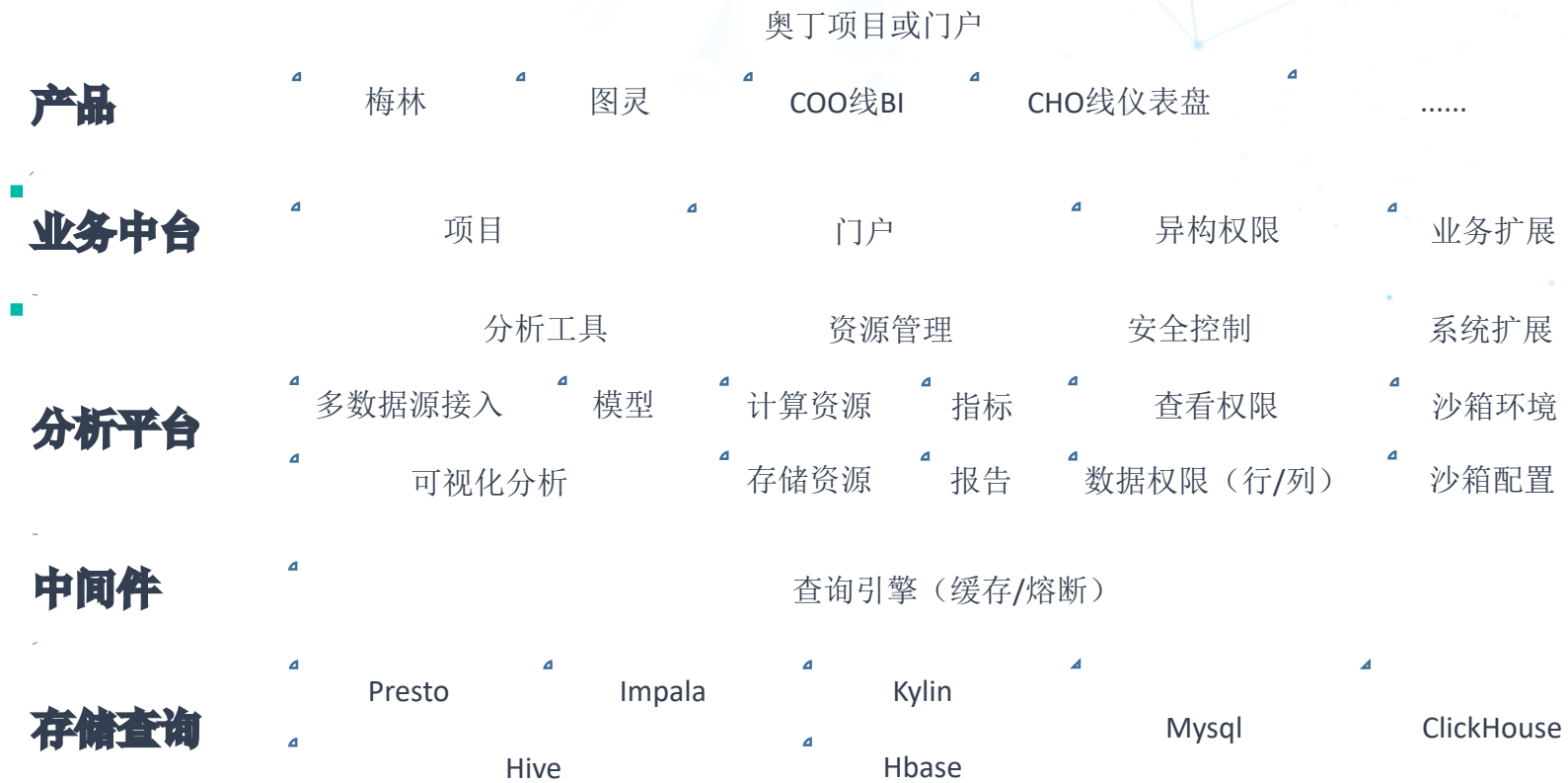
效果：成本低；快速实践，迭代中改进



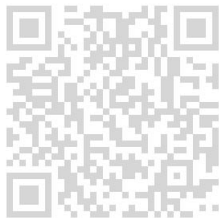
全新IT技术私域交流平台

技术总结

二、贝壳大数据分析平台设计方法



效果：平台+中台设计思想，能更清楚的将平台核心功能和业务对接分离来

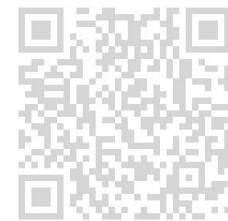


全新IT技术私域交流平台

04



未来规划



全新IT技术私域交流平台

未来规划

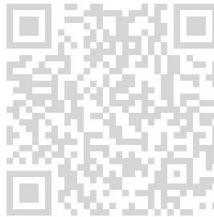
产品

丰富功能

提供更多的可视化组件：雷达图、地图等
支持更多定制化样式和交互

数据分析

帮助用户对数据进一步解读



全新IT技术私域交流平台

未来规划

技术

性能

独立资源加速自动化。
精细化预热

中台建设

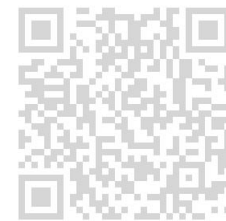
让更多的业务方以低成本接入
在对接业务中，逐步积累、搭建



全新IT技术私域交流平台



Thanks



全新IT技术私域交流平台