携程集中式日志及其周边生态系统

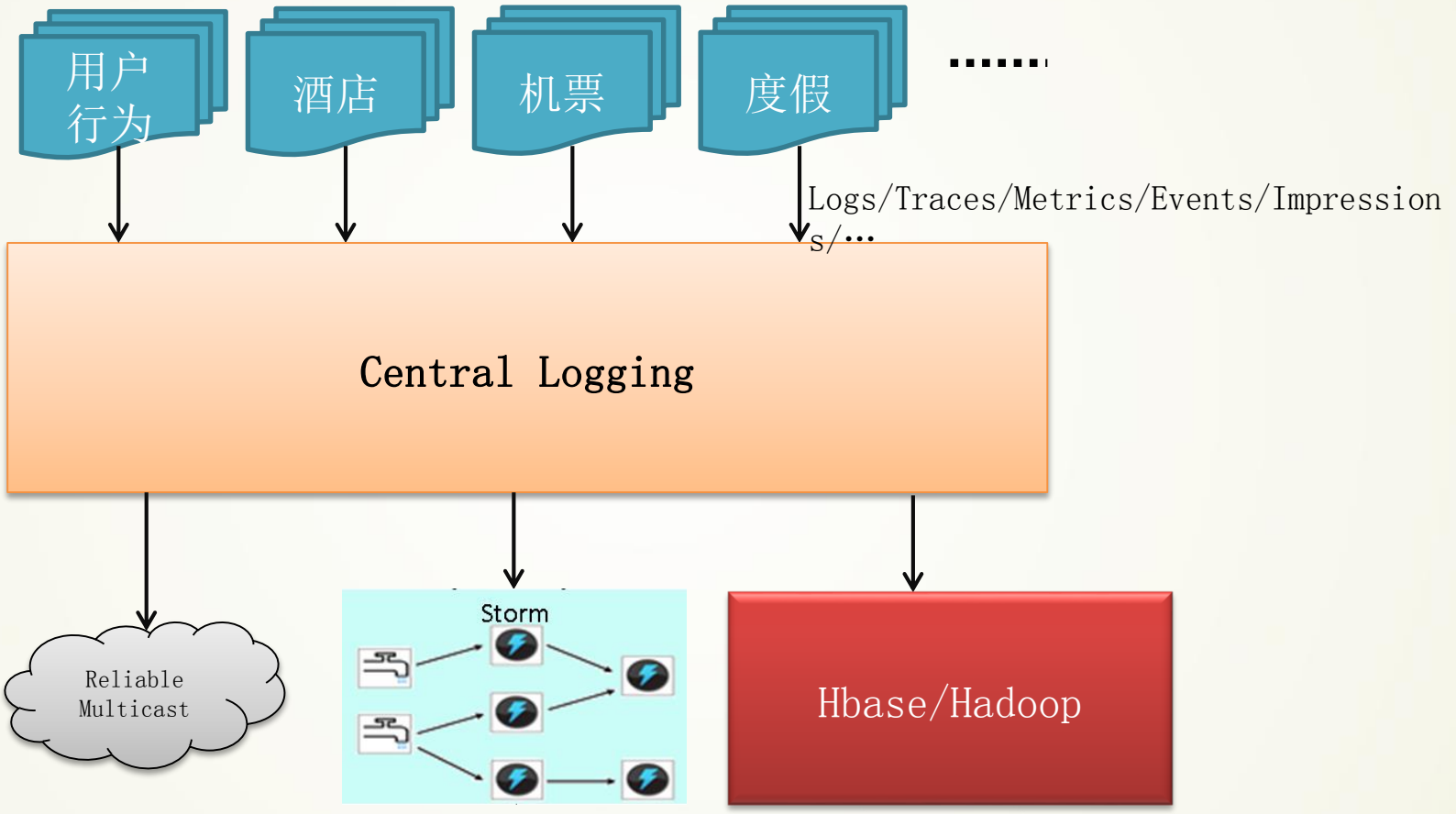
Ctrip R&D Framework

2013-04

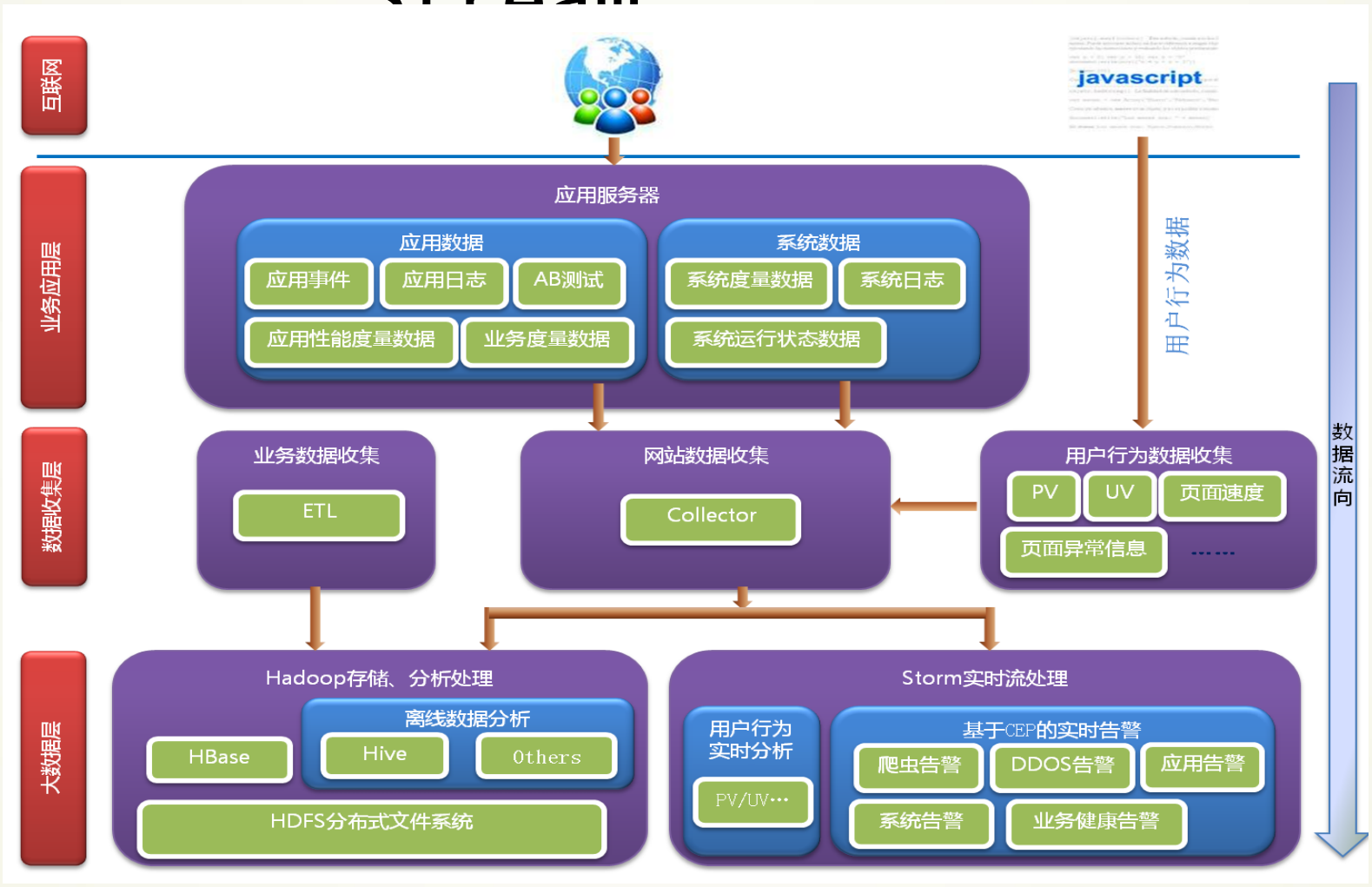
Agenda

1. Brief Introduction
2. Central Logging
3. Log View
4. Dashboard
5. User Behavior Tracking
6. Alerting
7. Data Farm

Central Logging



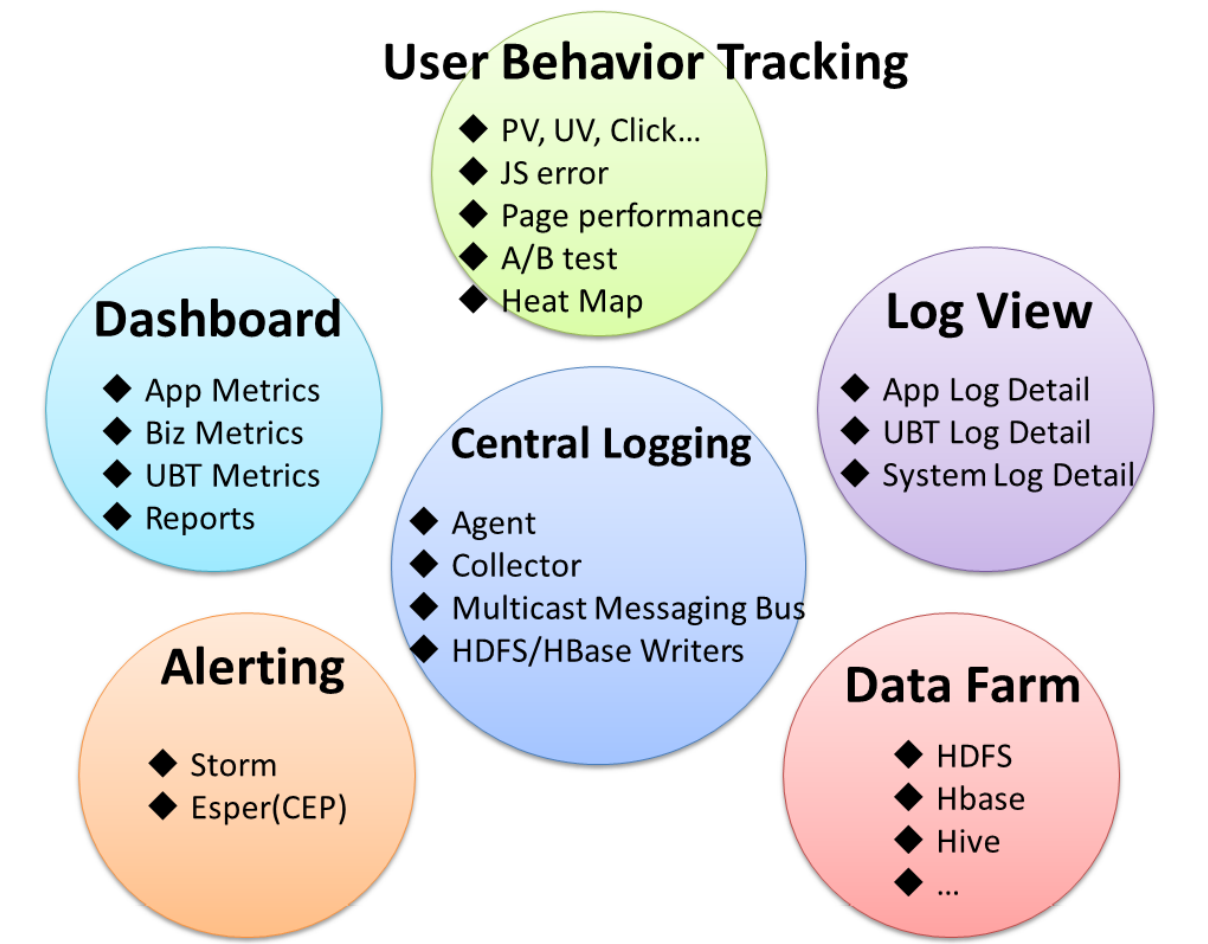
Central Logging - Data Stream



Central Logging Goals

- Dev：
  - 诊断URL/WS/DB/APP异常，帮助调试
  - 发现性能瓶颈（profiling）
  - 关联前、后端的调用
- PM/BU：
  - 通过用户查询理解用户需求（query understanding）
  - 基于用户行为调整搜索结果和推荐（impressions/clicks/orders）
  - 实时收集、分析A/B测试结果，以提高转换率
- Ops：
  - 网站Capacity分析，更合理地规划硬件资源
  - 系统、应用、业务的实时告警
  - 了解不同应用之间的依赖耦合关系

Central Logging & Eco-system

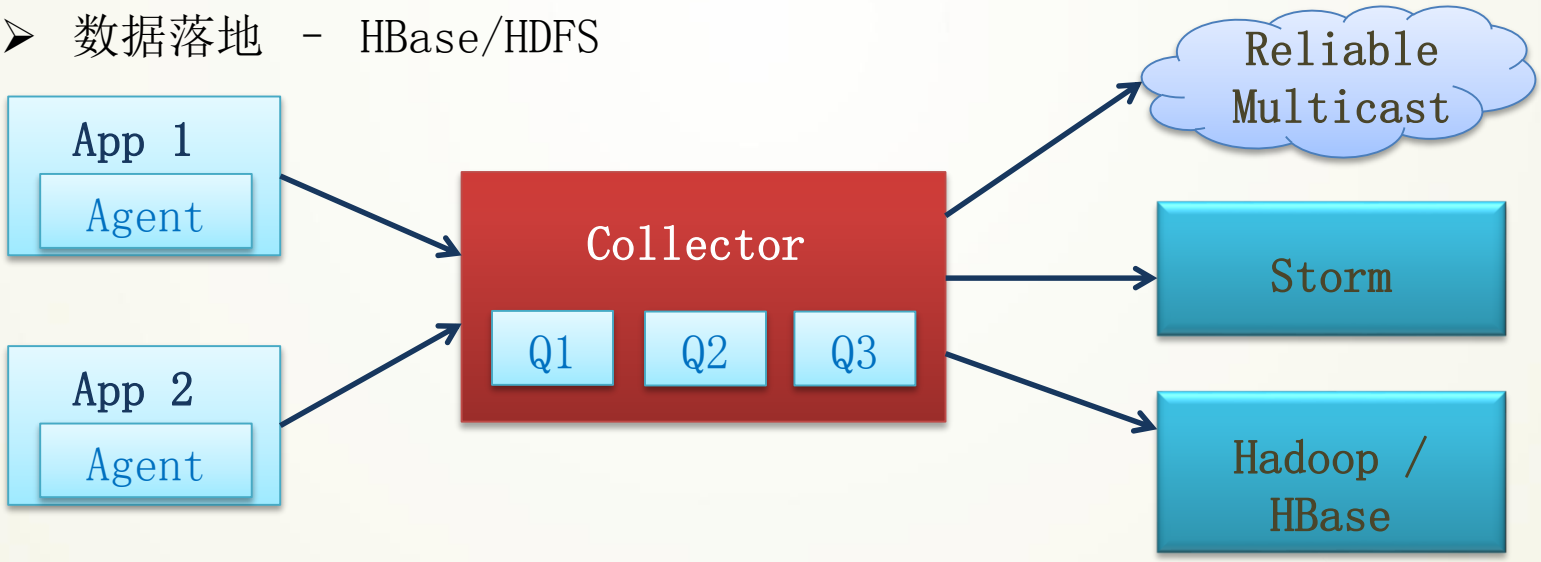


Agenda

1. Brief Introduction
2. Central Logging
3. Log View
4. Dashboard
5. User Behavior Tracking
6. Alerting
7. Data Farm

Central Logging

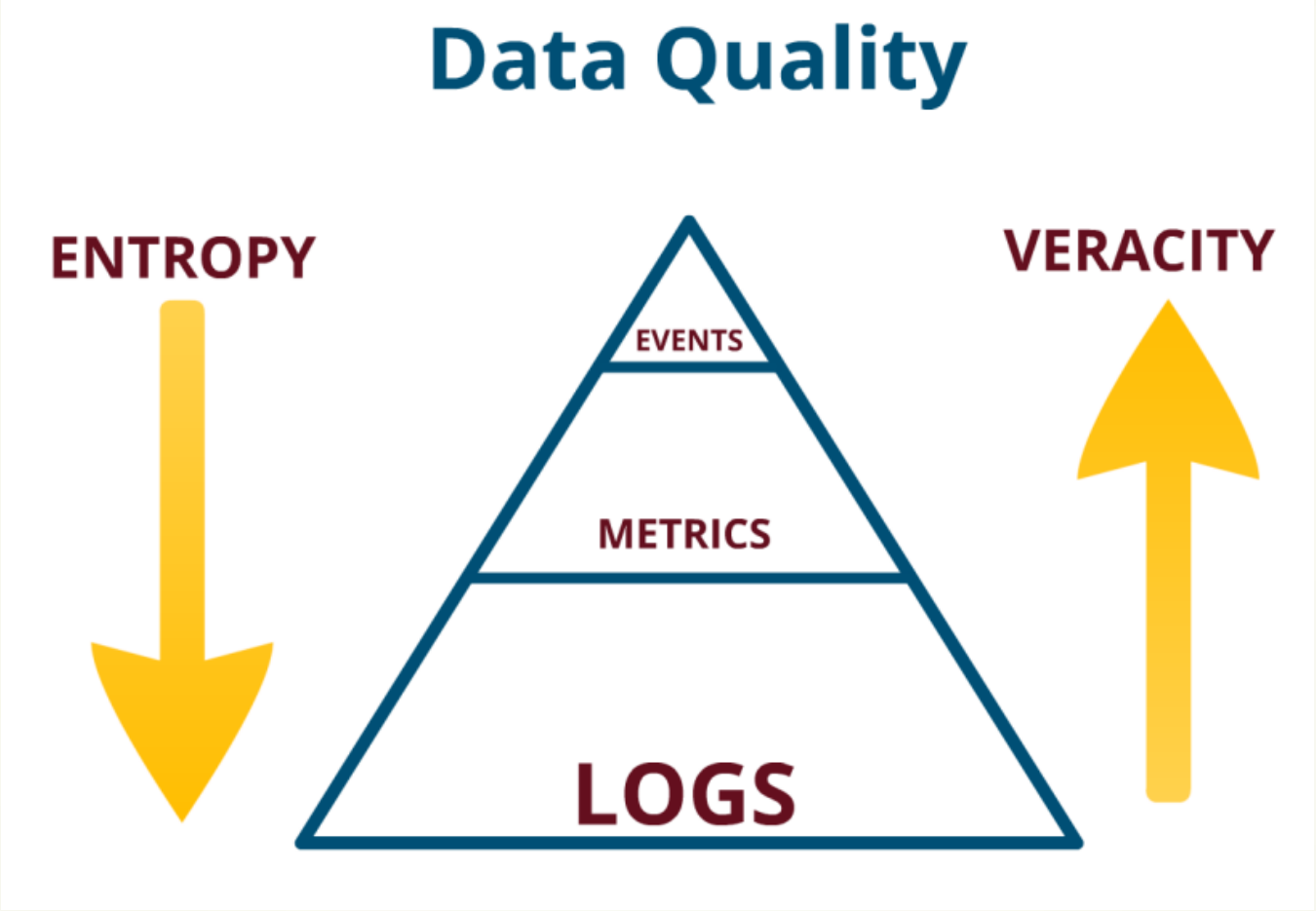
- 客户端 - Logging Agent
- 数据汇聚层 - Collector
- 数据消费者:
  - 实时消息处理 - Reliable Multicast
  - 实时流式处理 -Storm
  - 数据落地 - HBase/HDFS



Central Logging - Agent API

- Logging API
  - 提供类logback/log4j/log4net接口
  - 基于tags扩展，便于搜索查询
- Trace API
  - 度量单次URL/WS/DB调用
  - 通过TraceID关联前、后端调用，厘清依赖关系
- Metrics API
  - 分钟级聚合统计
  - 支持avg, sum, min, max等聚合操作
  - 支持用户自定义tags(维度)
- Event API
  - 指明event类型、发生时间和其他相关属性
  - 作为CEP实时处理的数据源

Central Logging - Data Quality



Reference: http://www.infoq.com/presentations/Big-Data-Monitoring-eBay

Central Logging - Agent Data Collection

- SOA框架
  - 应用启动、关闭、心跳事件
  - 错误、异常日志度量统计
  - URL/WS/DB/Cache等Trace调用
    - 度量统计
    - TraceID生成和传递
    - SQL hash、statement采样
- 应用
  - 应用日志
  - 业务、应用度量数据
  - 应用自定义事件

Central Logging - Logging Collector

- 类似消息队列的broker
  - At-most-once (not reliable)
  - 日志无排序需求
  - 支持Topic语义
  - 内存 + mmap队列，屏蔽后端故障
- 吞吐量优先
  - 支持长、短链接通信
  - 支持以chunk为单位打包、压缩传输
- 可水平扩展
- 高可用

Agenda

1. Brief Introduction
2. Central Logging
3. Log View
4. Dashboard
5. User Behavior Tracking
6. Alerting
7. Data Farm

Log View

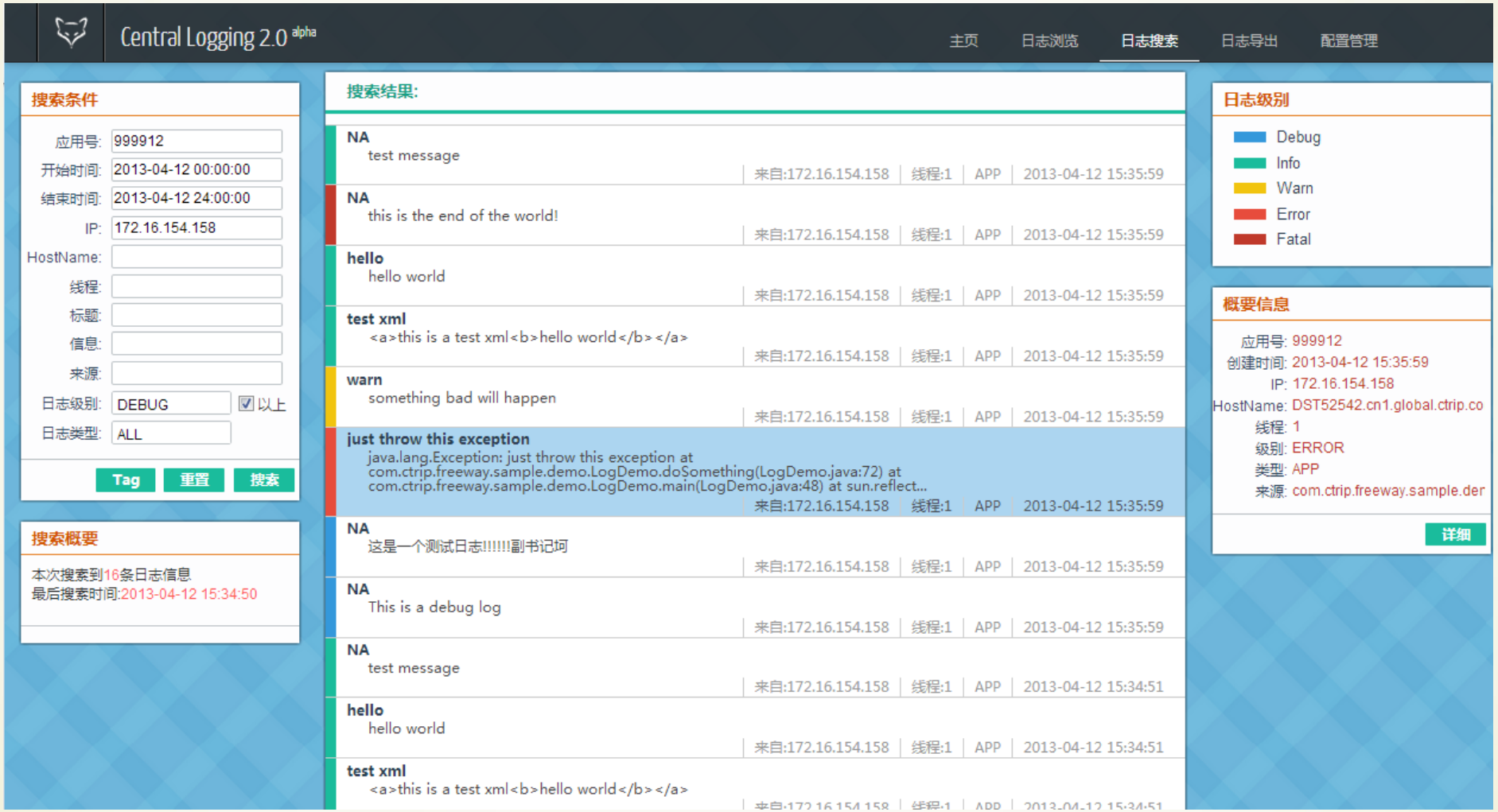
- 查看Log、Trace信息
  - 支持浏览或搜索方式
- 关联前、后端调用链日志
- 性能分析、Profiling
- 关键技术：Hbase Data Schema Design

Log View - Log Navigation Demo

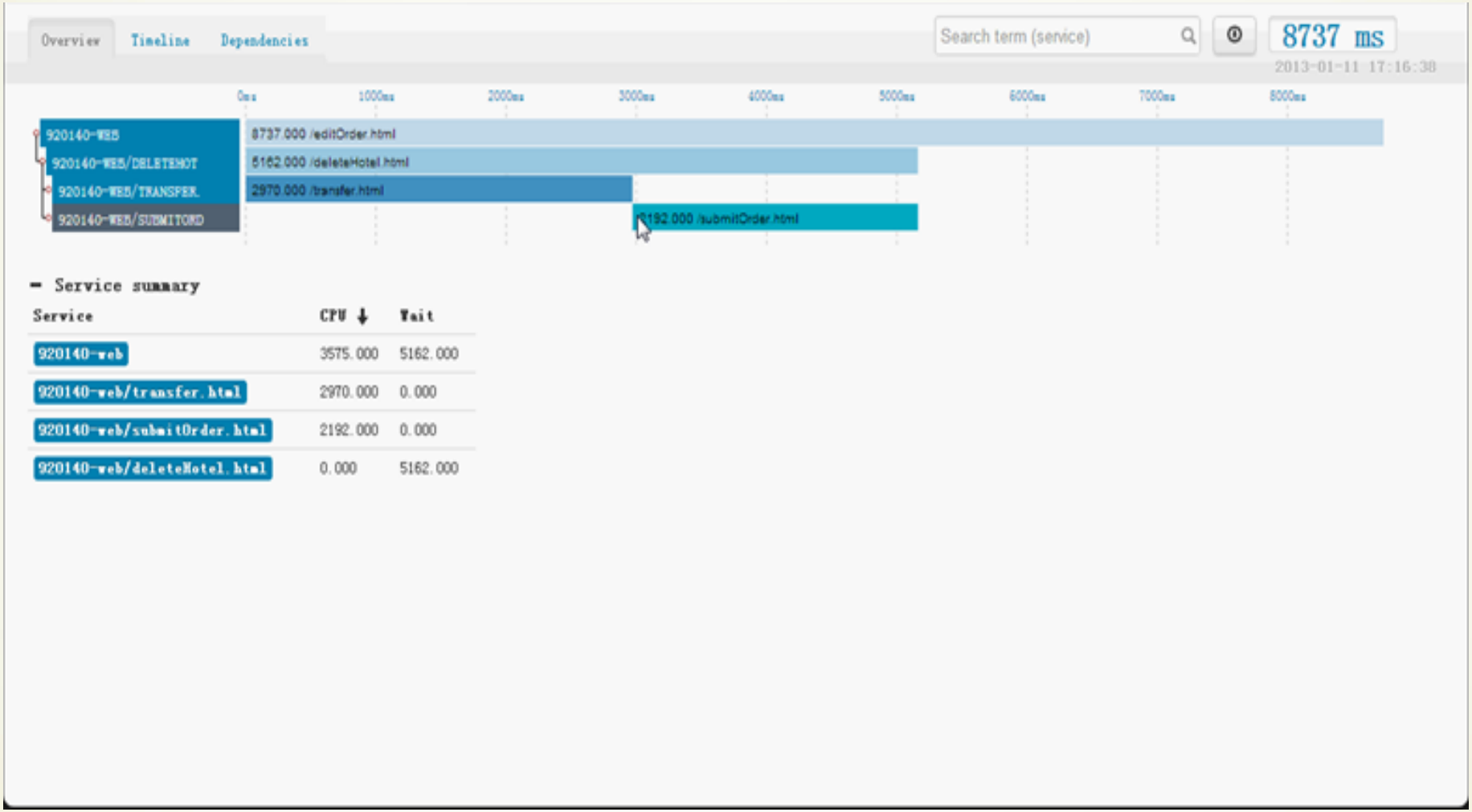
Log View - Log Navigation Demo (Cont.)

Log View - Log Search Demo (Old)

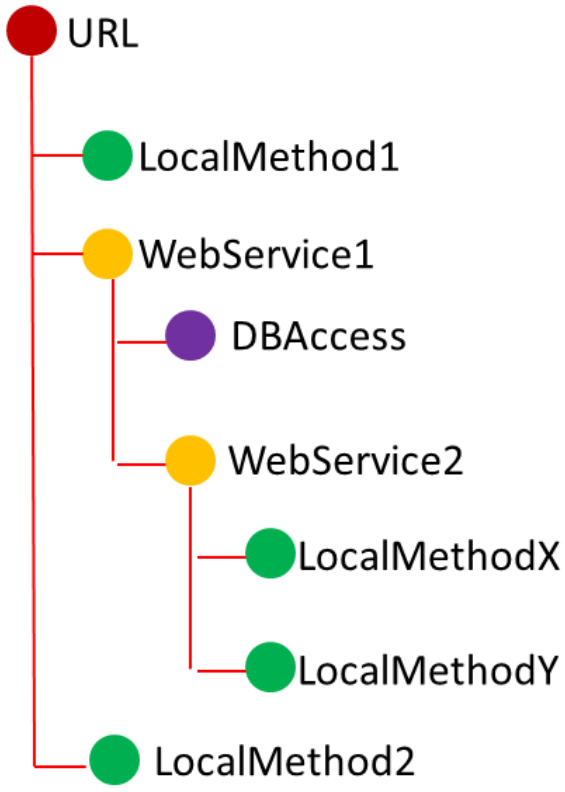
Log View - Log Search Demo (New)

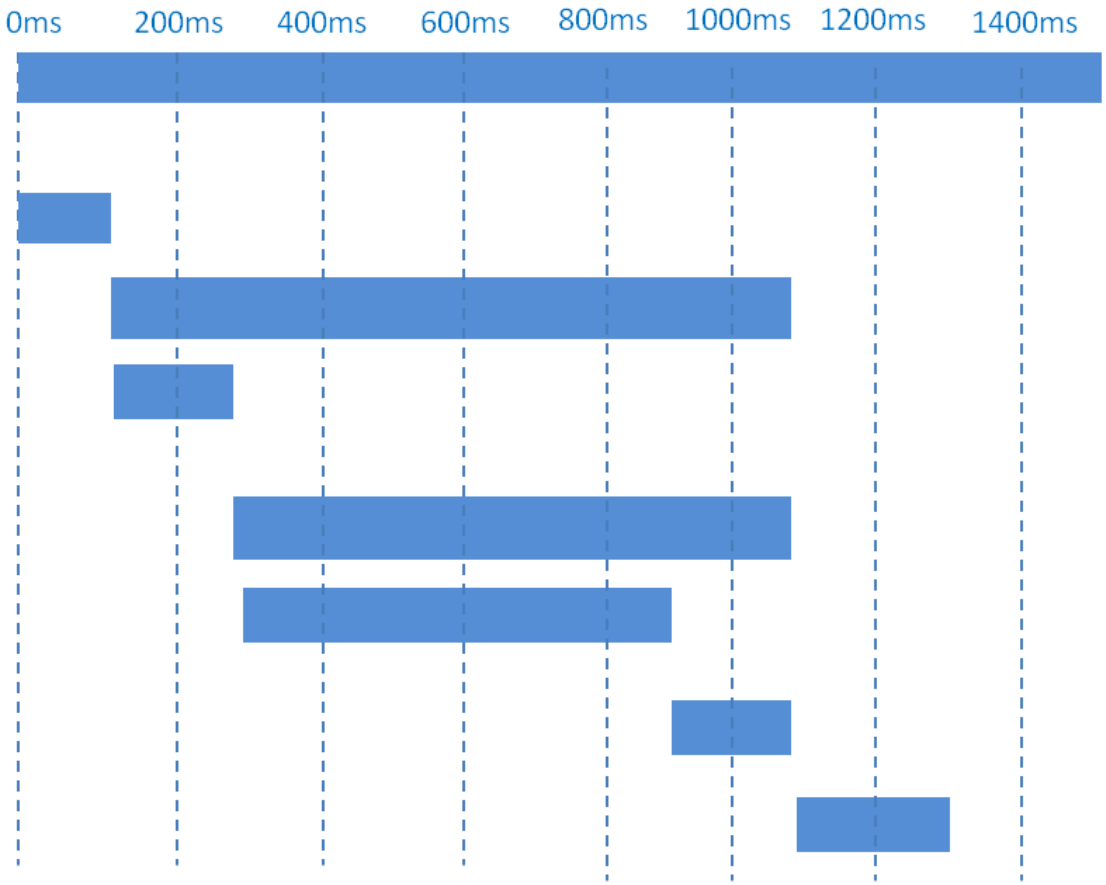


Log View - Trace Demo



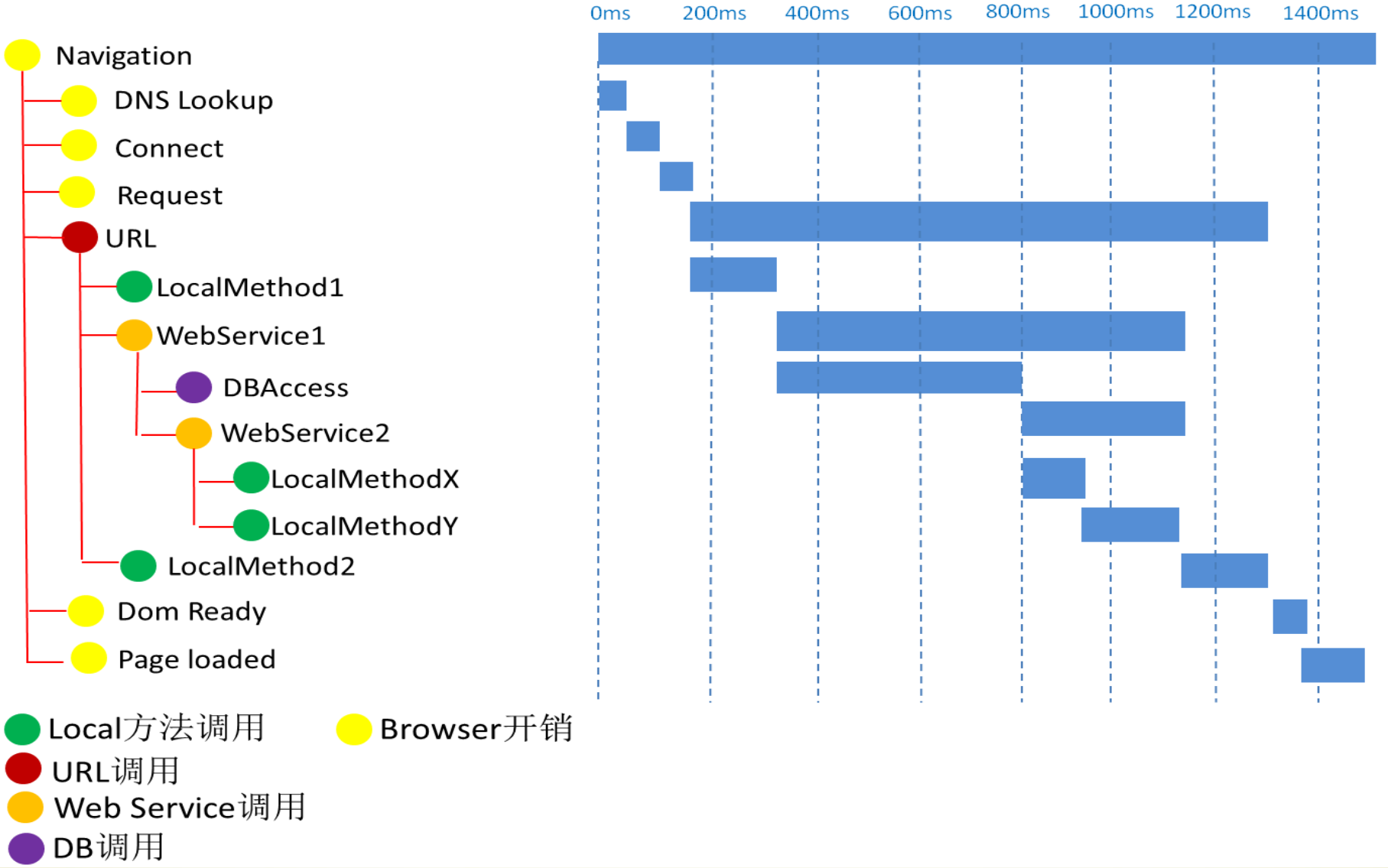
Log View - Trace Today



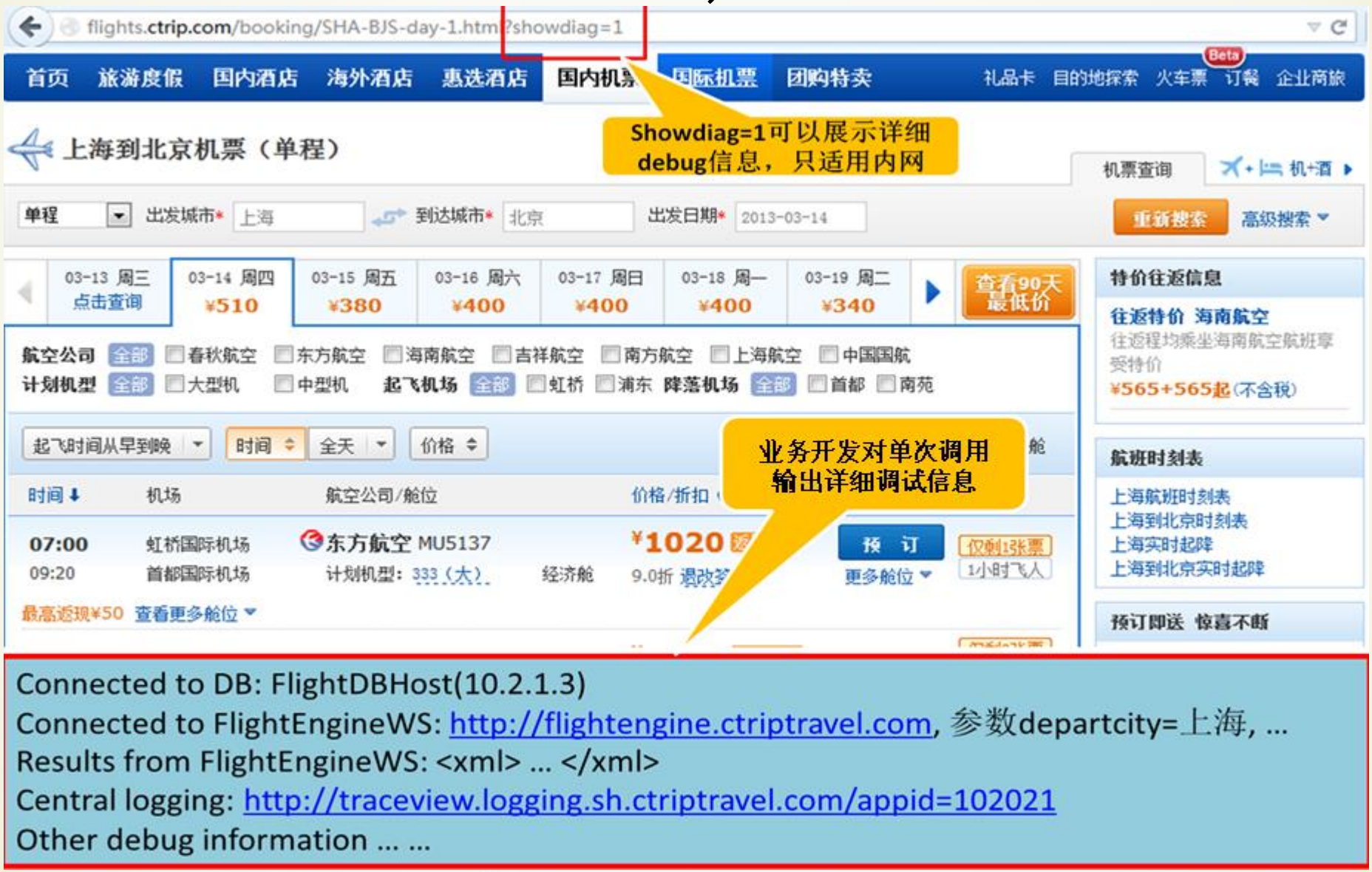


Local方法调用
URL调用
Web Service调用

Log View - Trace Tomorrow

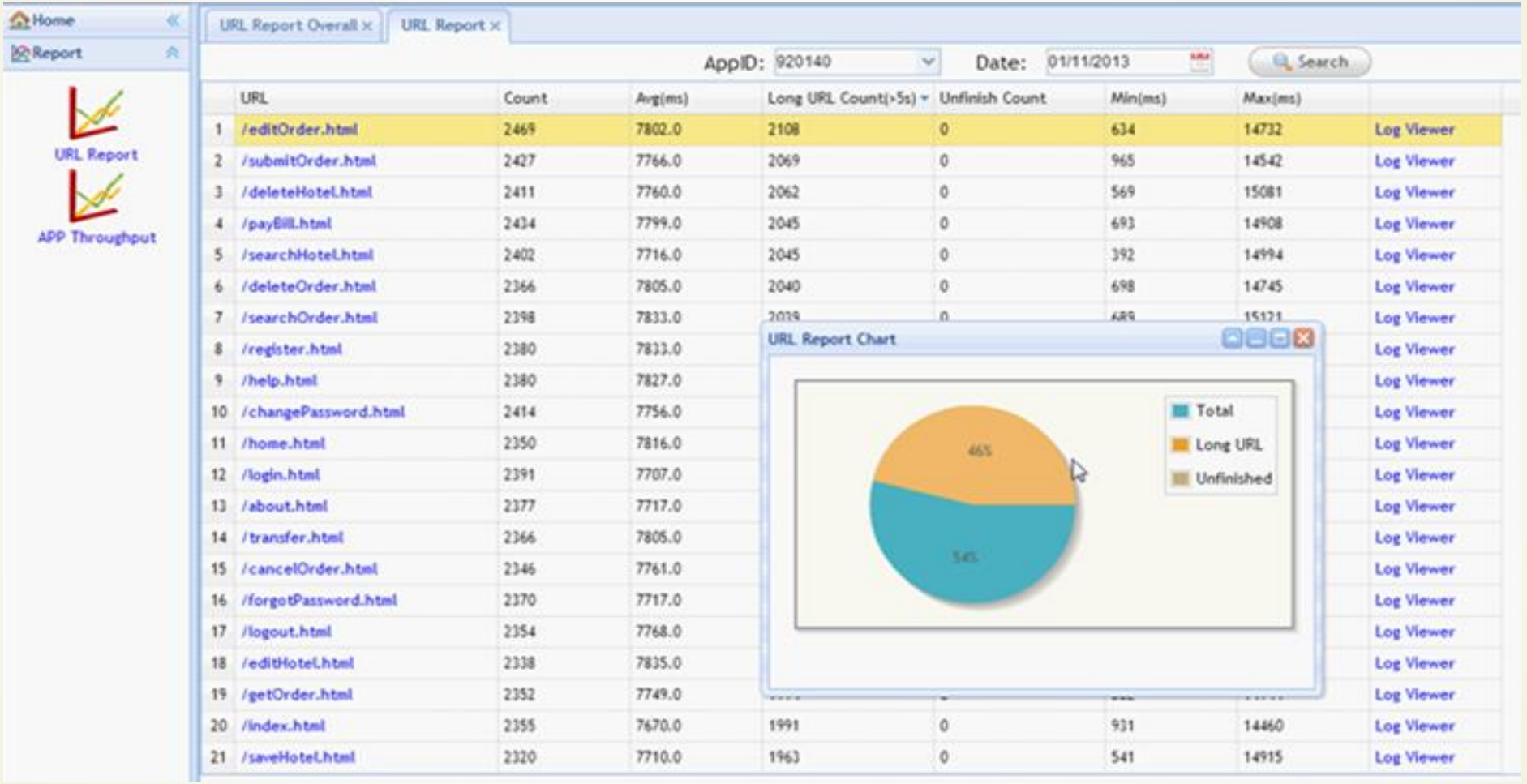


Log View - FE & BE Correlation (Coming soon)



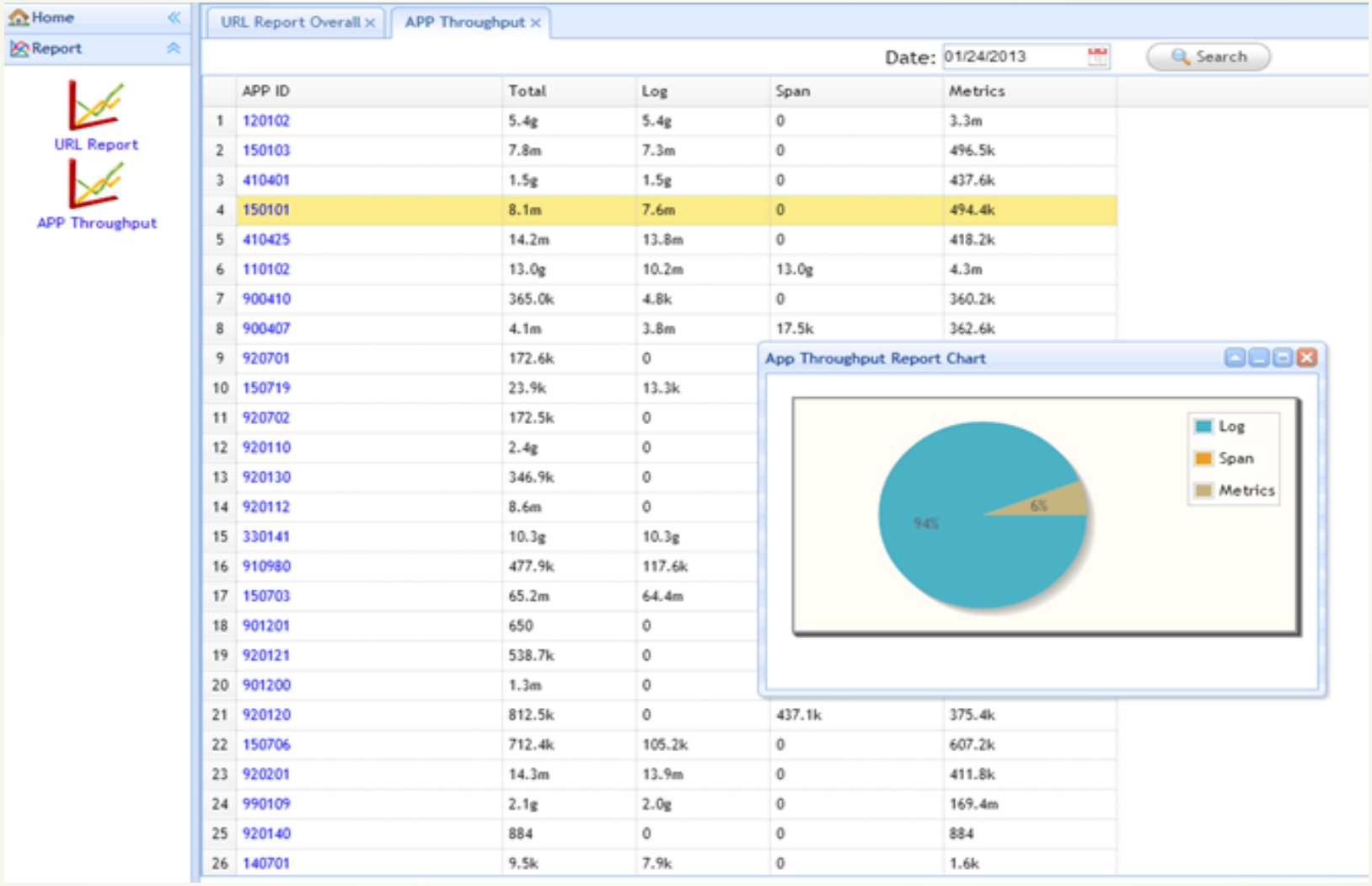
Connected to DB: FlightDBHost(10.2.1.3)
Connected to FlightEngineWS: http://flightengine.ctriptravel.com, 参数departcity=上海, ...
Results from FlightEngineWS: <xml> ... </xml>
Central logging: http://traceview.logging.sh.ctriptravel.com/appid=102021
Other debug information ... ...

Log View - Long URL Requests Demo



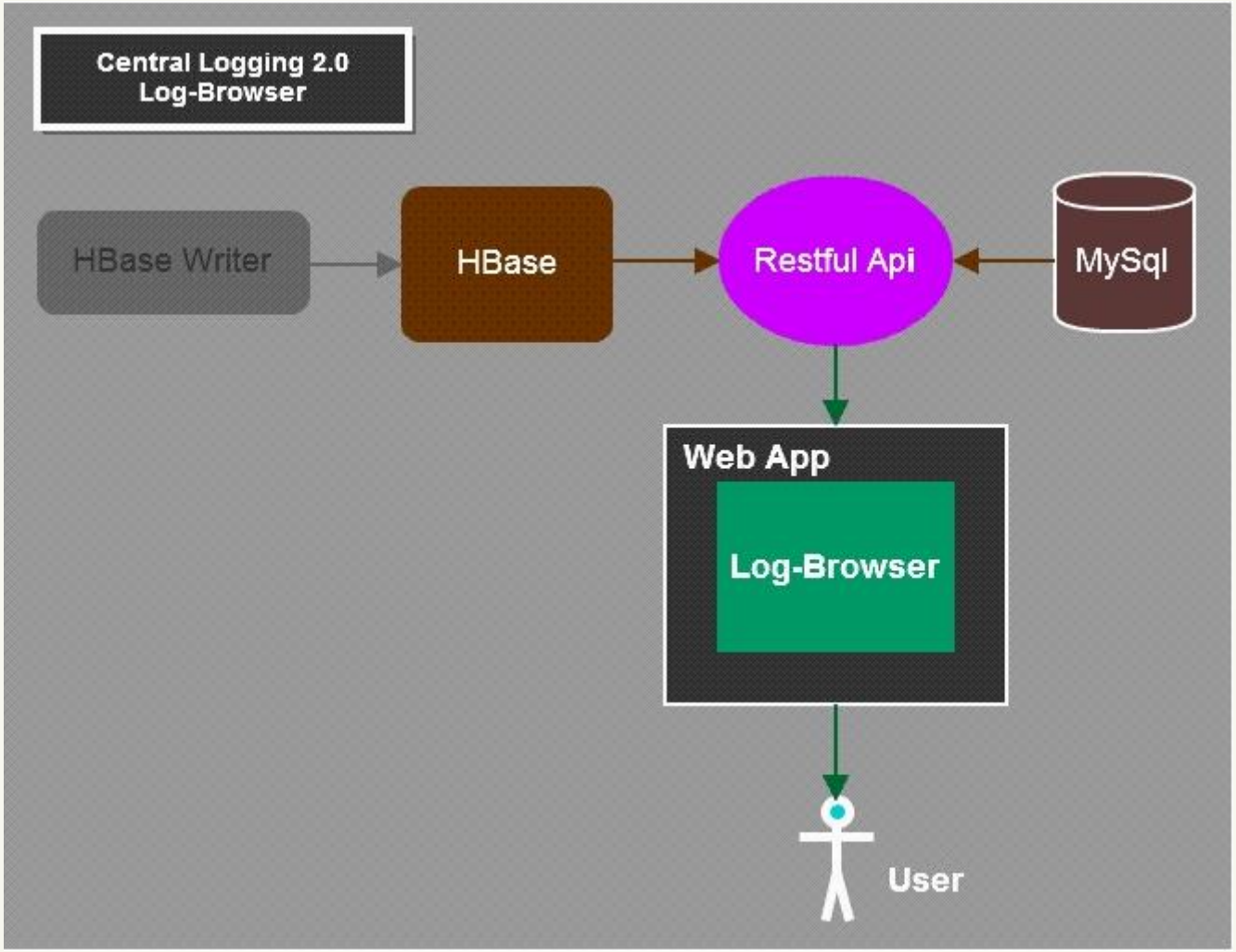
	URL	Count	Avg(ms)	Long URL Count(>5s)	Unfinish Count	Min(ms)	Max(ms)	
1	/editOrder.html	2469	7802.0	2108	0	634	14732	Log Viewer
2	/submitOrder.html	2427	7766.0	2069	0	965	14542	Log Viewer
3	/deleteHotel.html	2411	7760.0	2062	0	569	15081	Log Viewer
4	/payBill.html	2434	7799.0	2045	0	693	14908	Log Viewer
5	/searchHotel.html	2402	7716.0	2045	0	392	14994	Log Viewer
6	/deleteOrder.html	2366	7805.0	2040	0	698	14745	Log Viewer
7	/searchOrder.html	2398	7833.0	2039	0	689	15121	Log Viewer
8	/register.html	2380	7833.0					Log Viewer
9	/help.html	2380	7827.0					Log Viewer
10	/changePassword.html	2414	7756.0					Log Viewer
11	/home.html	2350	7816.0					Log Viewer
12	/login.html	2391	7707.0					Log Viewer
13	/about.html	2377	7717.0					Log Viewer
14	/transfer.html	2366	7805.0					Log Viewer
15	/cancelOrder.html	2346	7761.0					Log Viewer
16	/forgotPassword.html	2370	7717.0					Log Viewer
17	/logout.html	2354	7768.0					Log Viewer
18	/editHotel.html	2338	7835.0					Log Viewer
19	/getOrder.html	2352	7749.0					Log Viewer
20	/index.html	2355	7670.0	1991	0	931	14460	Log Viewer
21	/saveHotel.html	2320	7710.0	1963	0	541	14915	Log Viewer

Log View - Application Log Throughput Demo

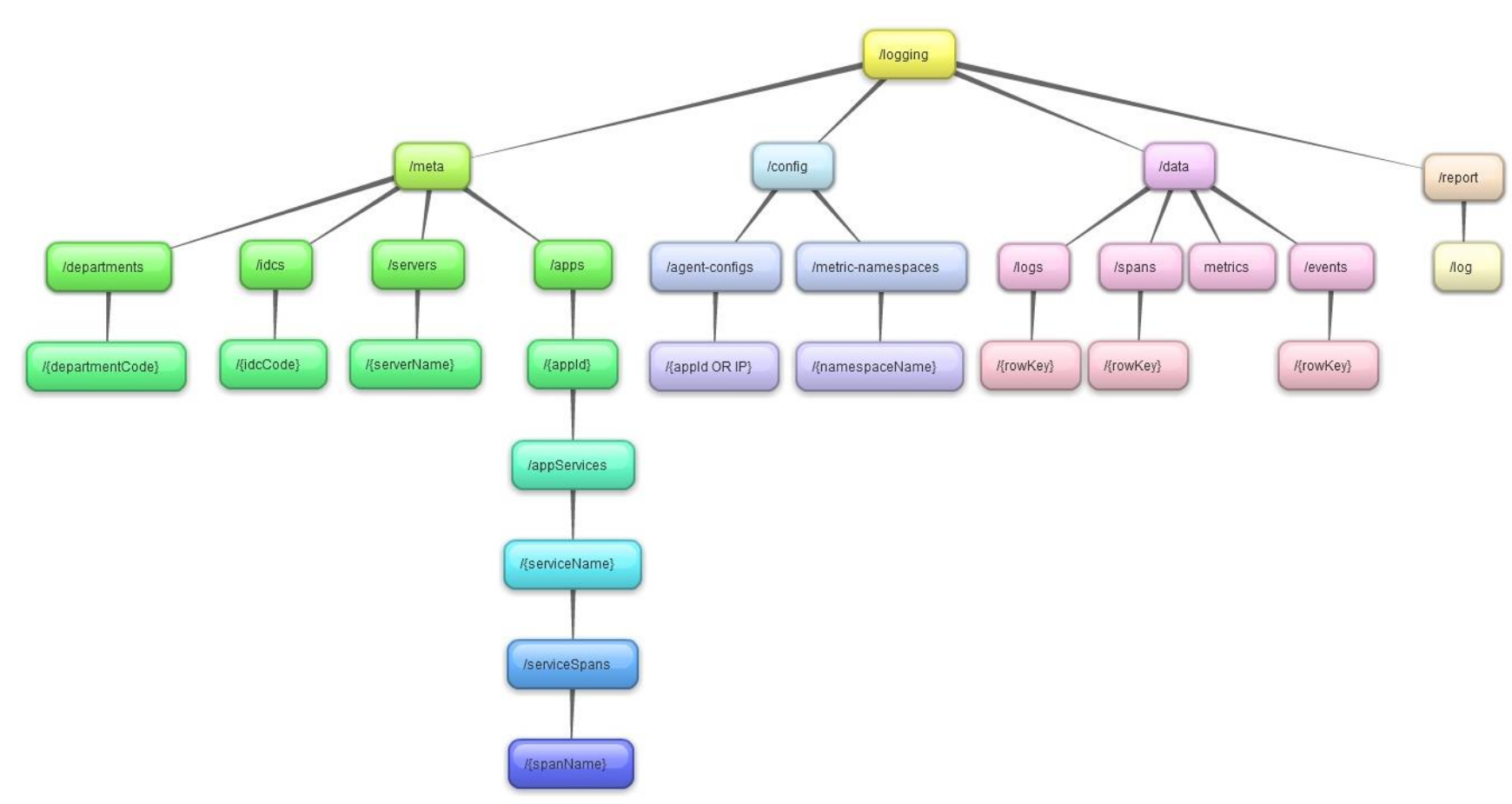


	APP ID	Total	Log	Span	Metrics
1	120102	5.4g	5.4g	0	3.3m
2	150103	7.8m	7.3m	0	496.5k
3	410401	1.5g	1.5g	0	437.6k
4	150101	8.1m	7.6m	0	494.4k
5	410425	14.2m	13.8m	0	418.2k
6	110102	13.0g	10.2m	13.0g	4.3m
7	900410	365.0k	4.8k	0	360.2k
8	900407	4.1m	3.8m	17.5k	362.6k
9	920701	172.6k	0		
10	150719	23.9k	13.3k		
11	920702	172.5k	0		
12	920110	2.4g	0		
13	920130	346.9k	0		
14	920112	8.6m	0		
15	330141	10.3g	10.3g		
16	910980	477.9k	117.6k		
17	150703	65.2m	64.4m		
18	901201	650	0		
19	920121	538.7k	0		
20	901200	1.3m	0		
21	920120	812.5k	0	437.1k	375.4k
22	150706	712.4k	105.2k	0	607.2k
23	920201	14.3m	13.9m	0	411.8k
24	990109	2.1g	2.0g	0	169.4m
25	920140	884	0	0	884
26	140701	9.5k	7.9k	0	1.6k

Log View - Architecture



Log View - RESTful API Design



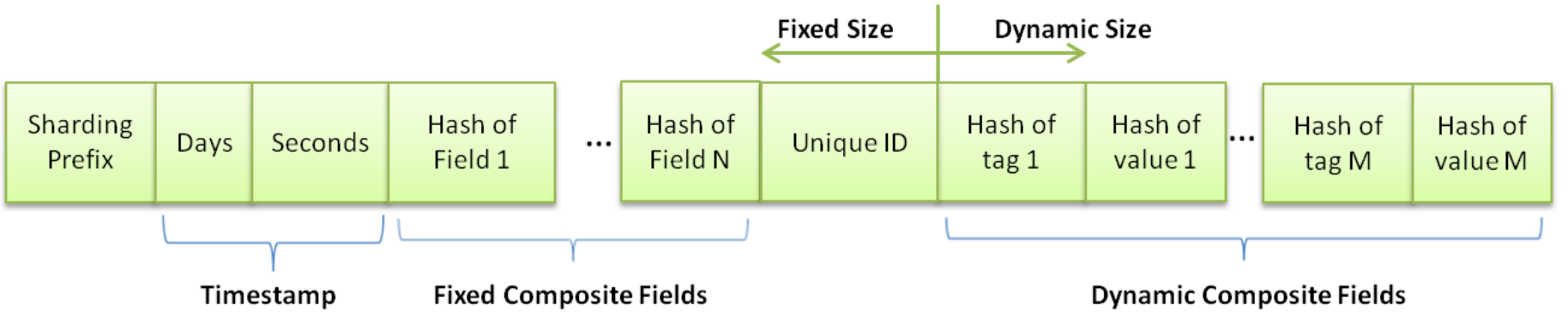
Log View - Optimizations

- 索引表Row Key设计
- 两段式查询优化
- 基于Atomic Append的列更新、插入(HBASE-4102)

Log View - Index Row Key Design

- 分片前缀（sharding prefix）
  - 使得数据平衡分布
- 时间戳（Timestamp）
  - Region重用
  - 按时间顺序倒排
- 组合索引优化
  - 固定字段索引优化
  - 用户自定义标签索引优化

Log View - Index Row Key Design (Cont.)



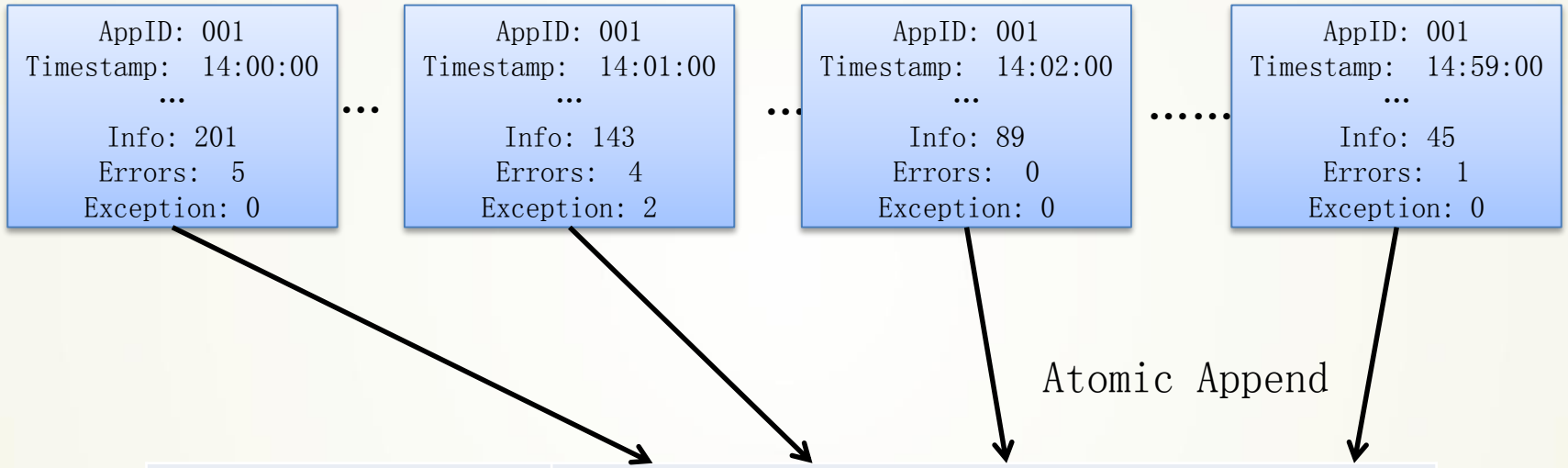
- 分片前缀（Sharding prefix）
  - 数值类型，e.g. application id
  - Bit/byte reversal
- 时间戳
  - 天：(System.currentTimeMillis() / (86400 * 1000)) % 30, （TTL=30天）
  - 秒：86400 - ((System.currentTimeMillis() /1000) % 86400)
- 索引字段
  - 字段按值分布进行hash压缩（1-2字节）
  - 支持静态、动态索引字段

Log View - Two-Path Query Design

- 两段式查询优化
  1. 对索引表scan查询
     - 目前采用 基于Key filter非回溯正则
     - 将来考虑使用FuzzyRowFilter (fast-forwarding优化)
  2. 对原表multiget查询 (+column filters)
- 解决hash冲突：Skew factor
- 性能提升 10X-100X

Log View - Atomic Append Optimization (HBASE-4102)

应用日志统计消息 - 时间线



Row Key	Column Family = stat				
{appid=001}+{TS=14:00}+……	Qualifer: 1	Qualifer: 2	Qualifer: 3	…	Qualifer: 60
	Info: 201 Errors: 5 Exception: 0	Info: 143 Errors: 4 Exception: 2	Info: 89 Errors: 0 Exception: 0	…	Info: 45 Errors: 1 Exception: 0

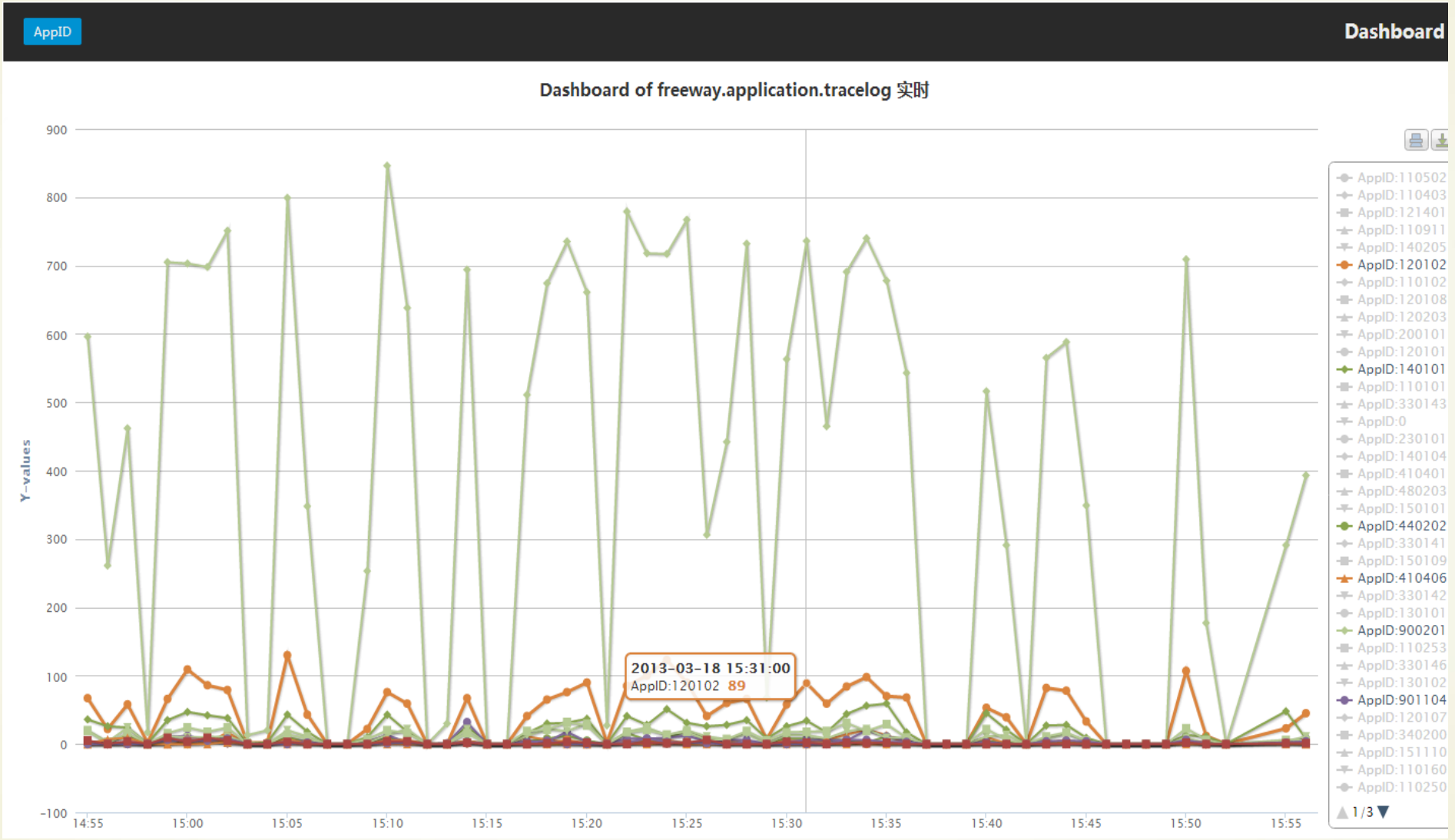
Agenda

1. Brief Introduction
2. Central Logging
3. Log View
4. Dashboard
5. User Behavior Tracking
6. Alerting
7. Data Farm

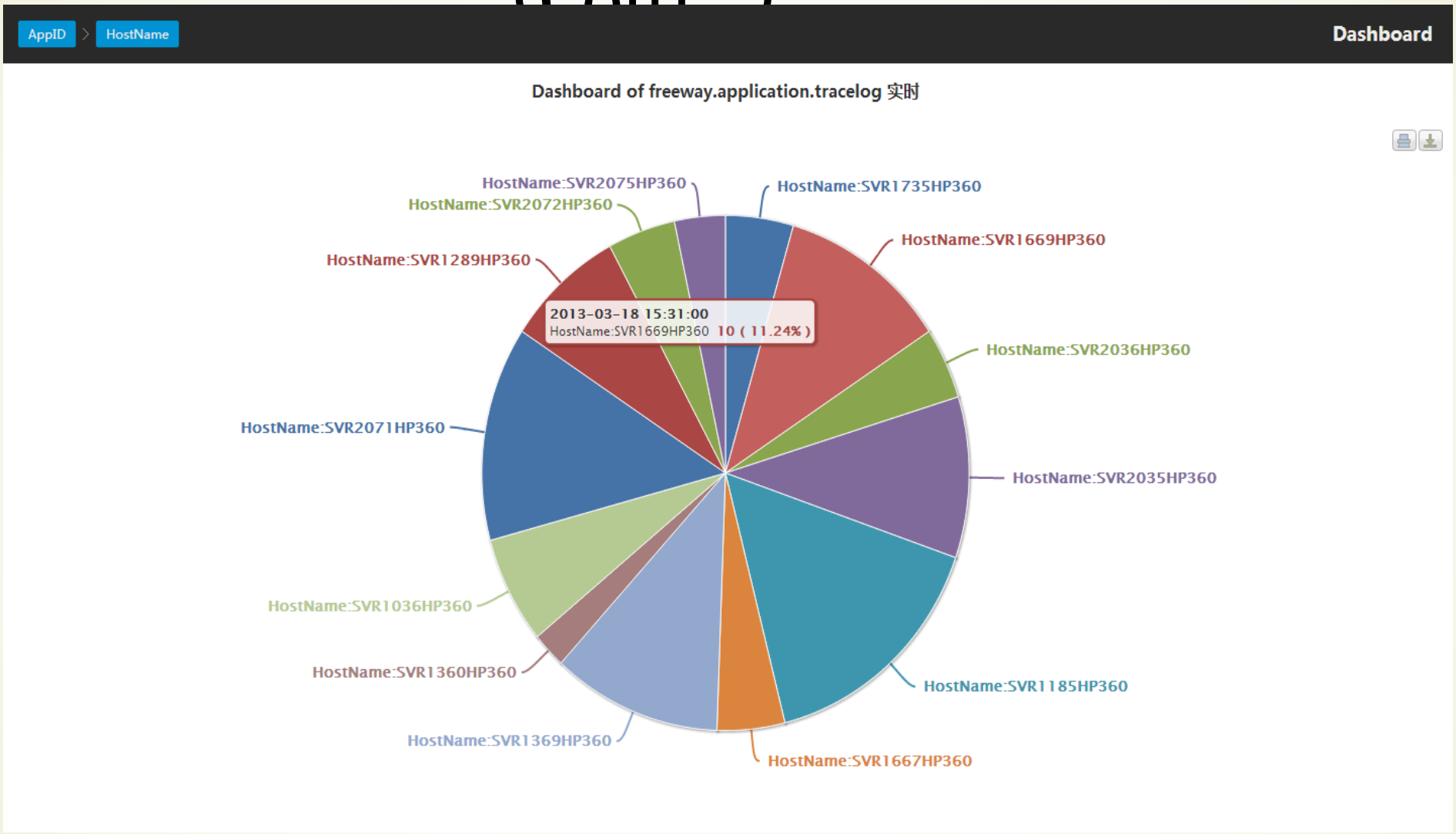
Dashboard

- 度量数据的展示
  - 应用、业务、系统度量数据
  - 多维度实时钻取
- 可水平扩展的度量查询引擎
- JS框架无缝整合
- 后端使用定制的OpenTSDB

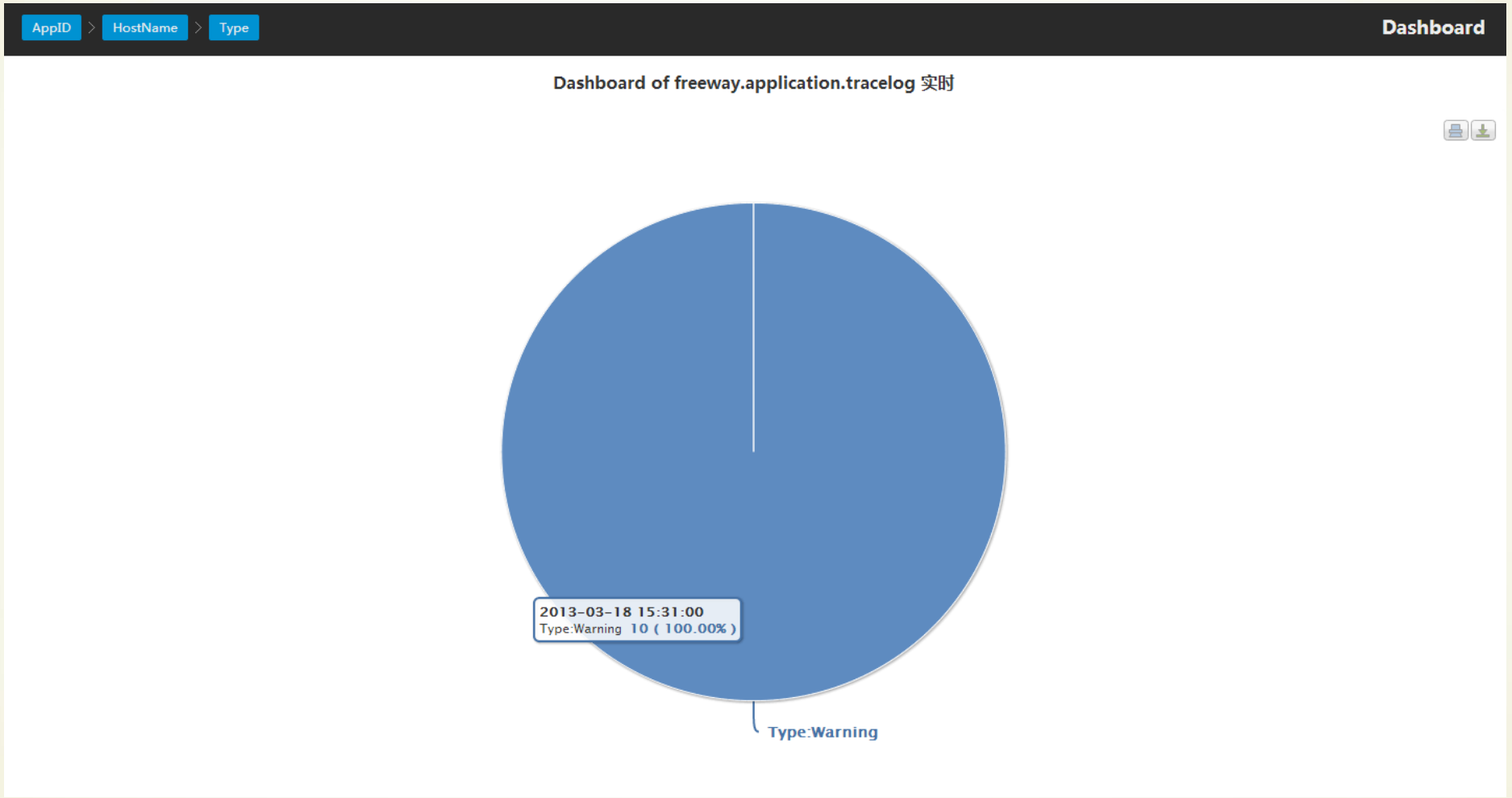
Dashboard - Trace Log Demo



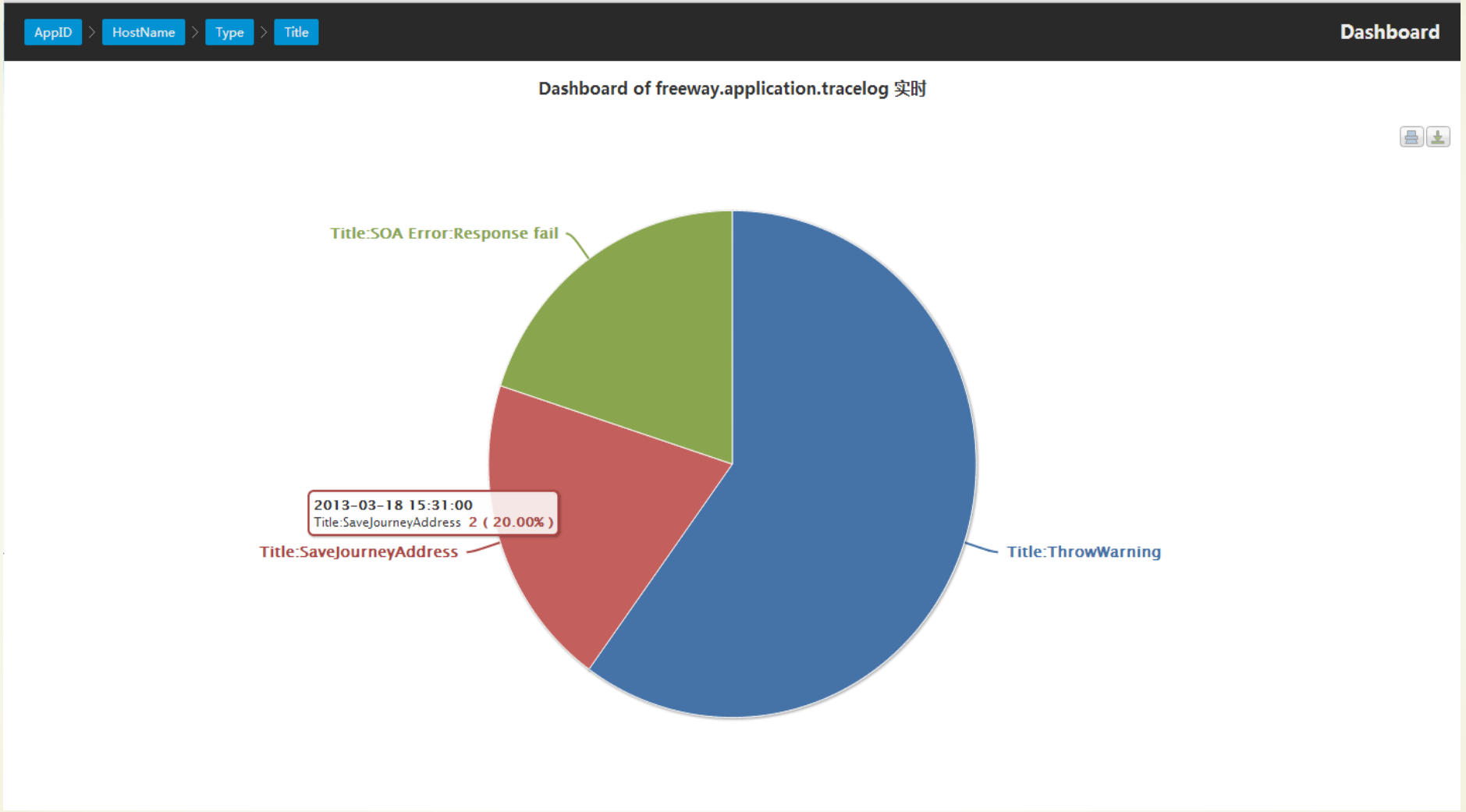
Dashboard - Trace Log Demo (Cont.)



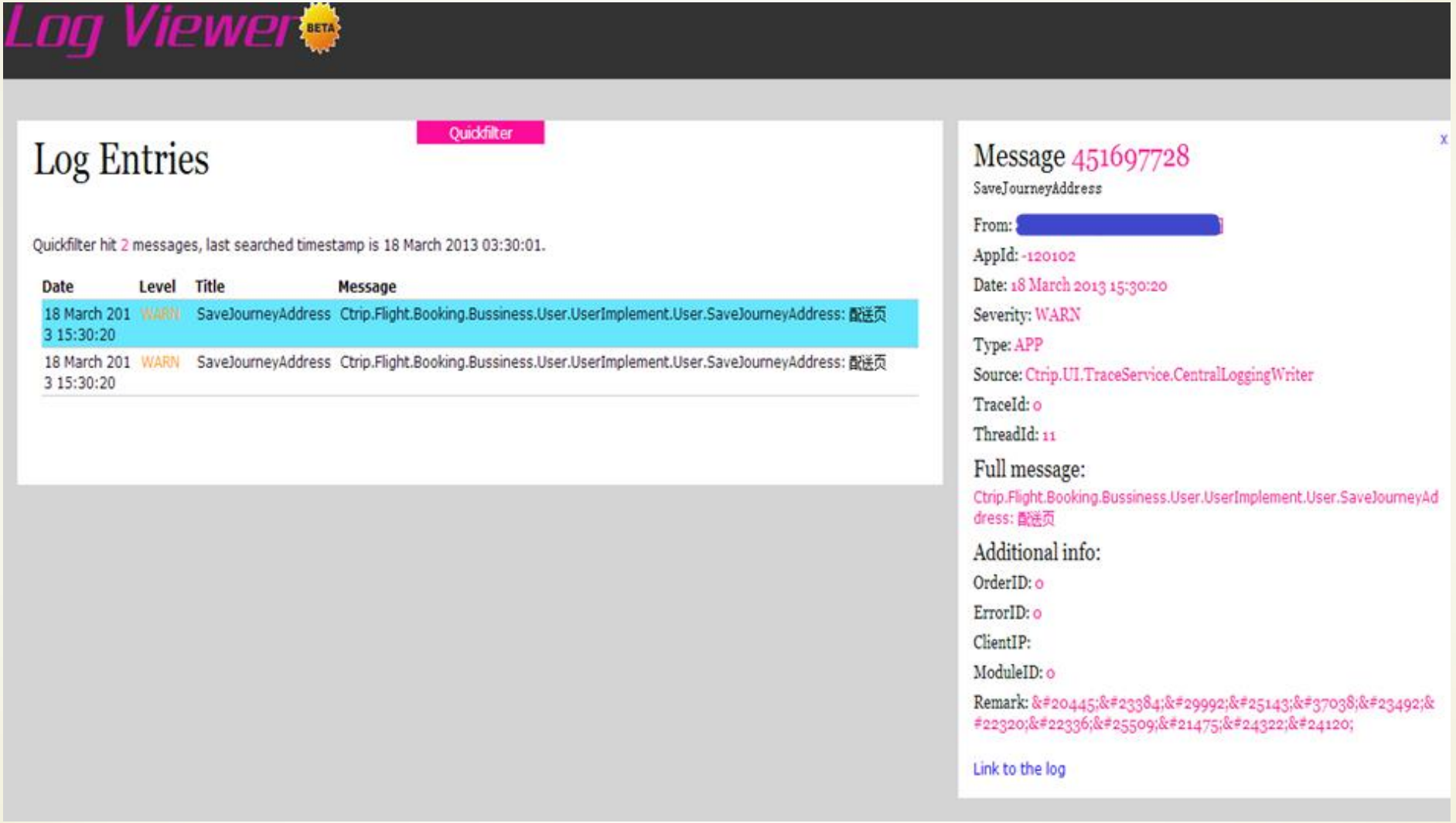
Dashboard - Trace Log Demo (Cont.)



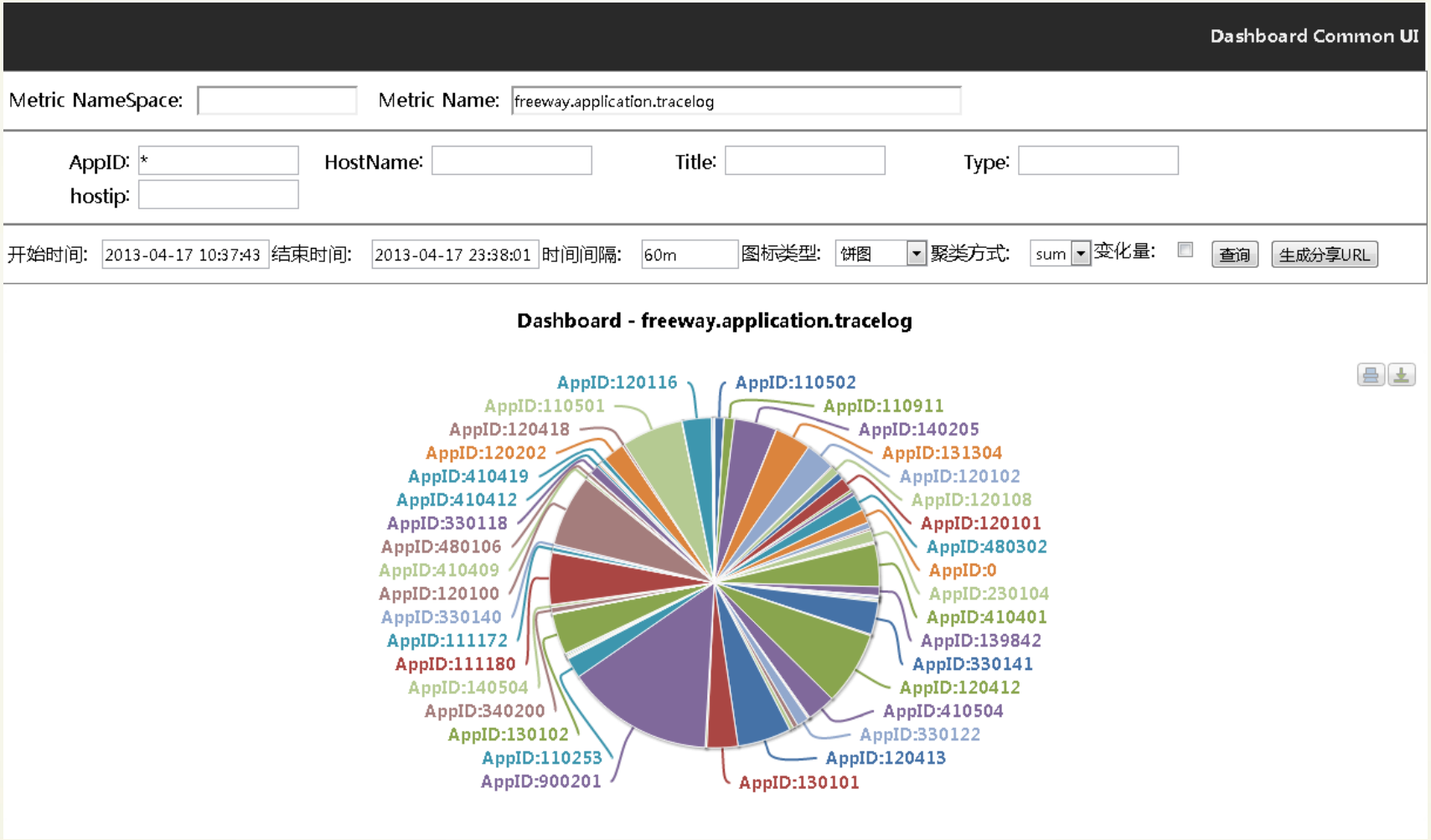
Dashboard - Trace Log Demo (Cont.)



Dashboard - Trace Log Demo (Cont.)

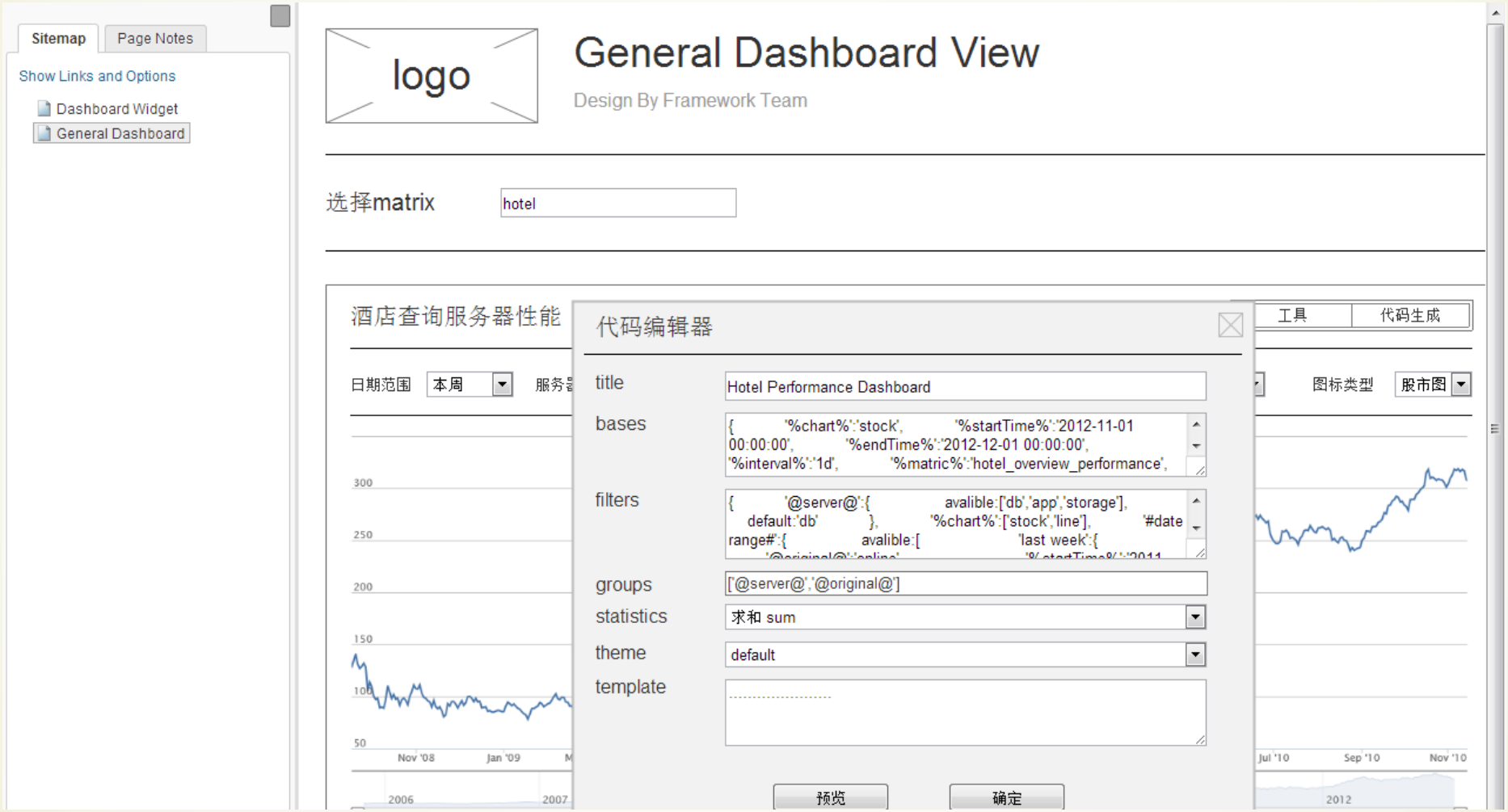


Dashboard - Common UI Demo

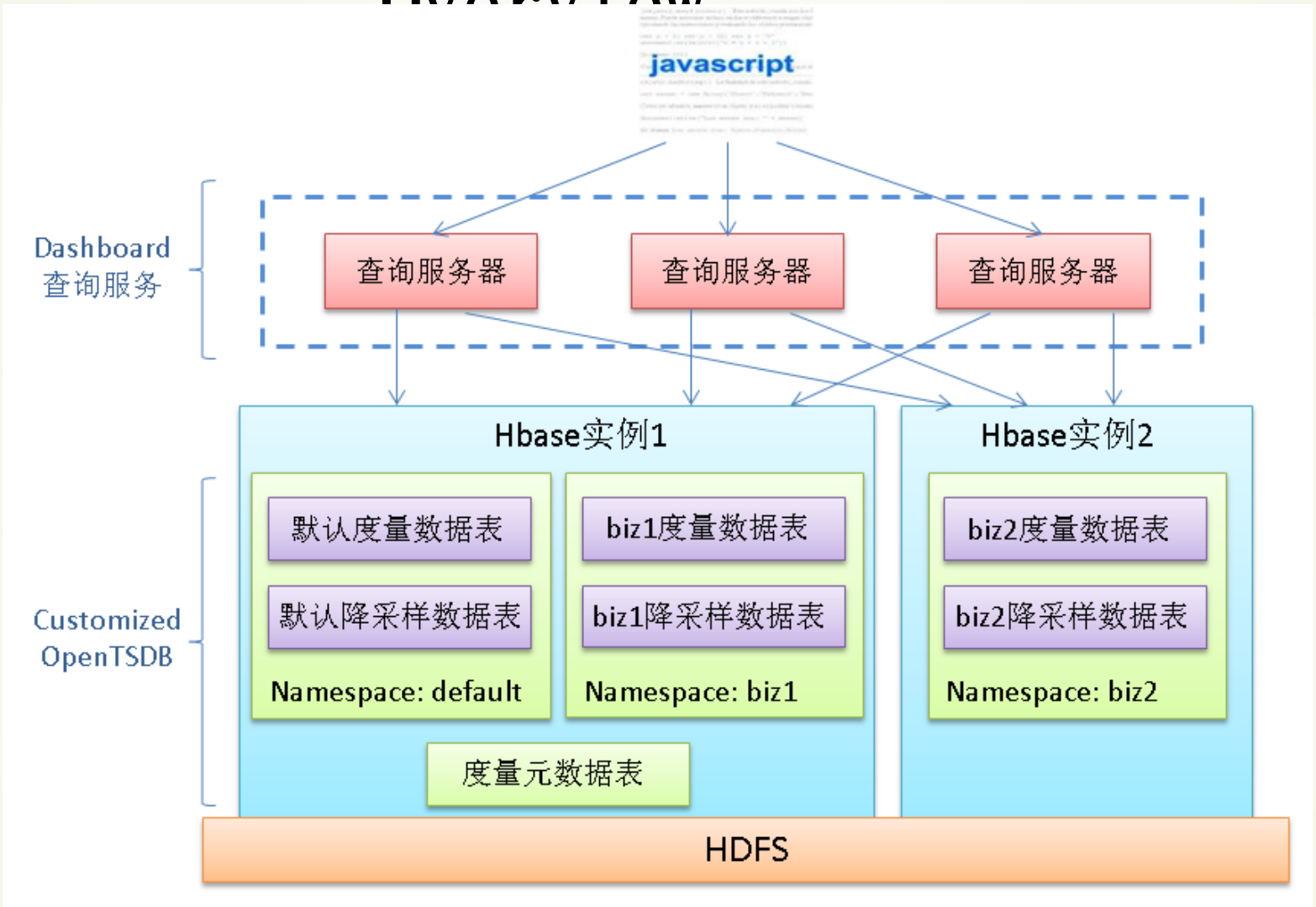


Dashboard - Dashboard JS Designer

Dashboard - Dashboard JS Designer (Cont.)



Dashboard - Architecture Overview



Dashboard - Customized OpenTSDB

- 新增basic time series(BTS)元数据(申请专利)
- 查询服务的cache优化(申请专利)
- 基于不同业务类型(namespace)分片
- 查询服务可水平扩展
- 定时对原数据进行降采样（map/reduce）
- 放宽原来最多8个tags的限制(至32个)
- 支持中文encoding
- 扩展复合数值类型，支持avg/dev降采样聚合

Dashboard - JS Framework

- 整合到携程JS框架
- 支持各类图表展示（基于highcharts）
- 所见即所得的JS设计
- 支持OpenTSDB的各类聚合方式
- 支持自定义查询条件、group by方式
- 按查询时间范围自适应调整时间间隔

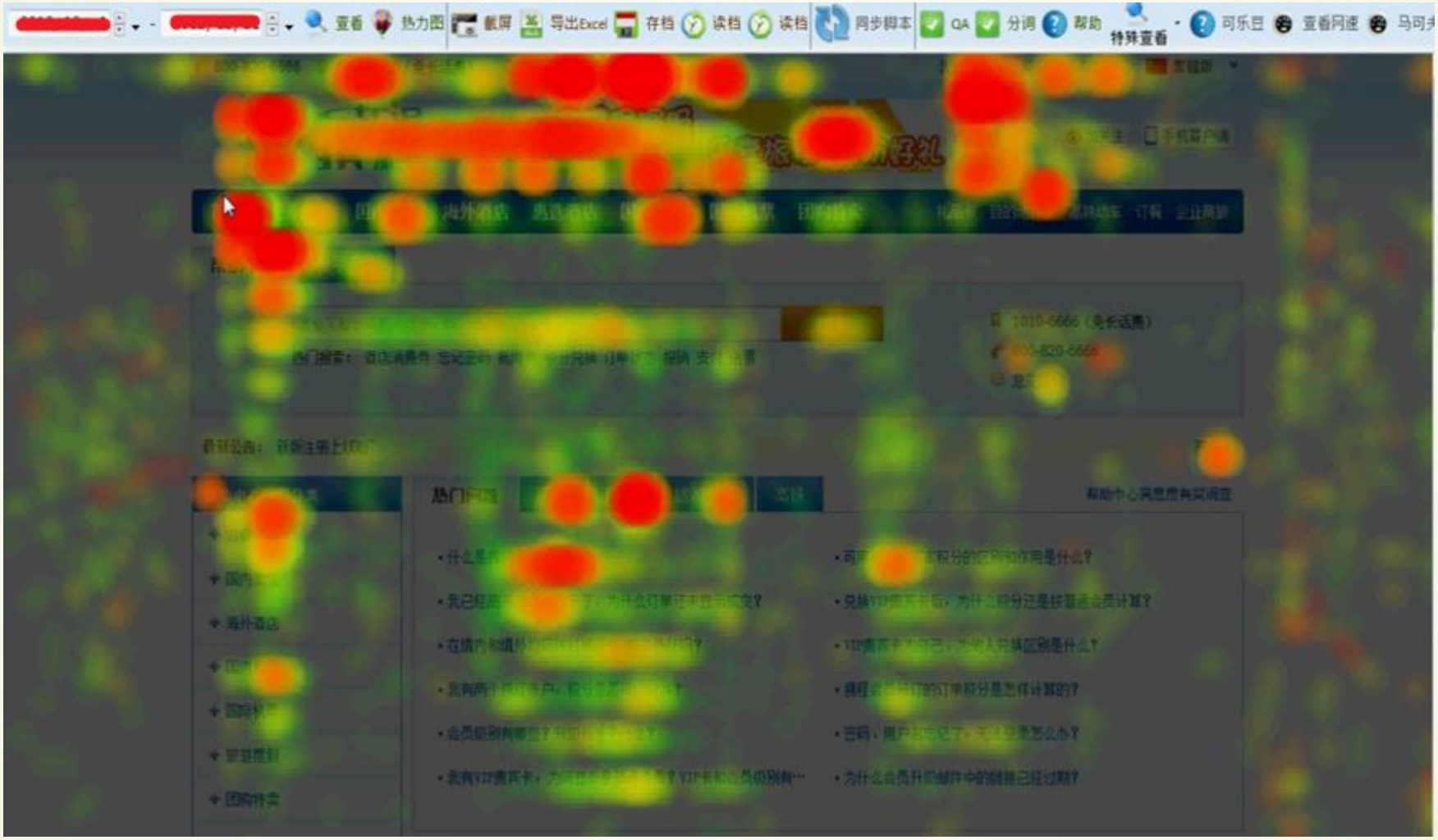
Agenda

1. Brief Introduction
2. Central Logging
3. Log View
4. Dashboard
5. User Behavior Tracking
6. Alerting
7. Data Farm

User Behavior Tracker (UBT)

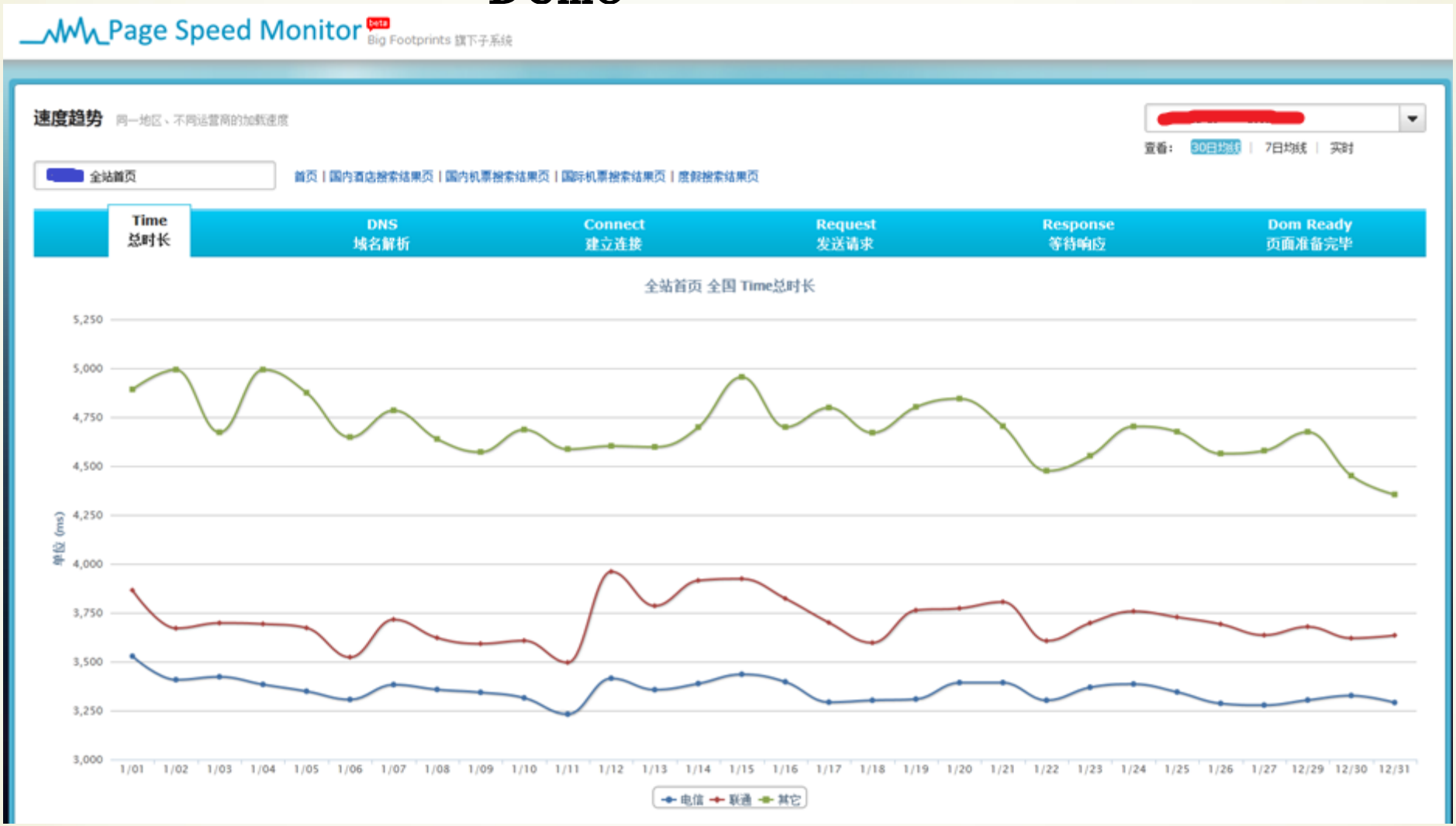
- BI分析重要数据源
  - 用户行为数据
    - PV, UV, Click, JS error, page performance…
  - 前端业务自定义数据回传
  - AB测试数据收集
  - 页面、流程转化率
  - 用户体验分析
    - Query understanding
    - Impression/Click-Through
    - 订单完成费力度

UBT - Heat Map Demo

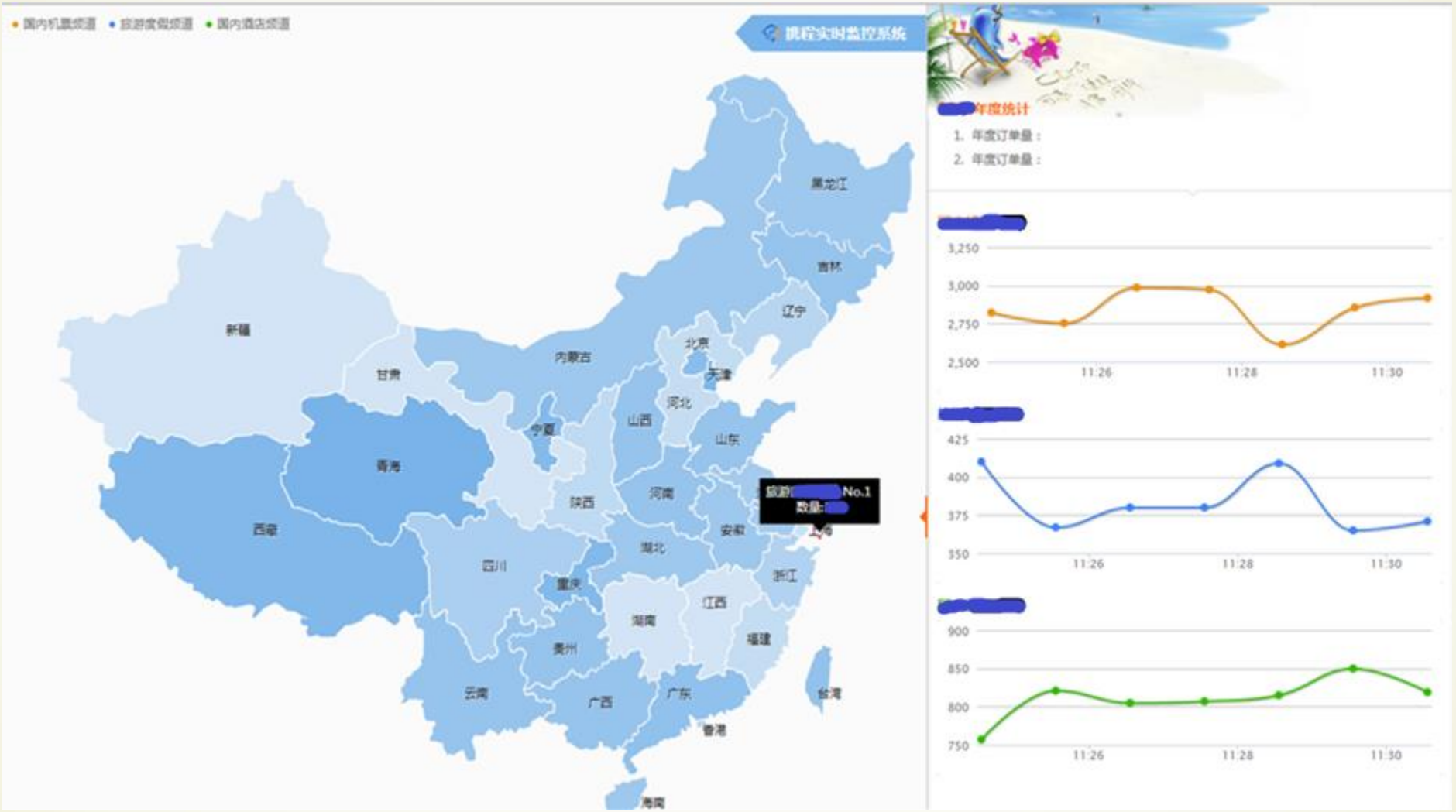


UBT - Page UBT Visualization Demo

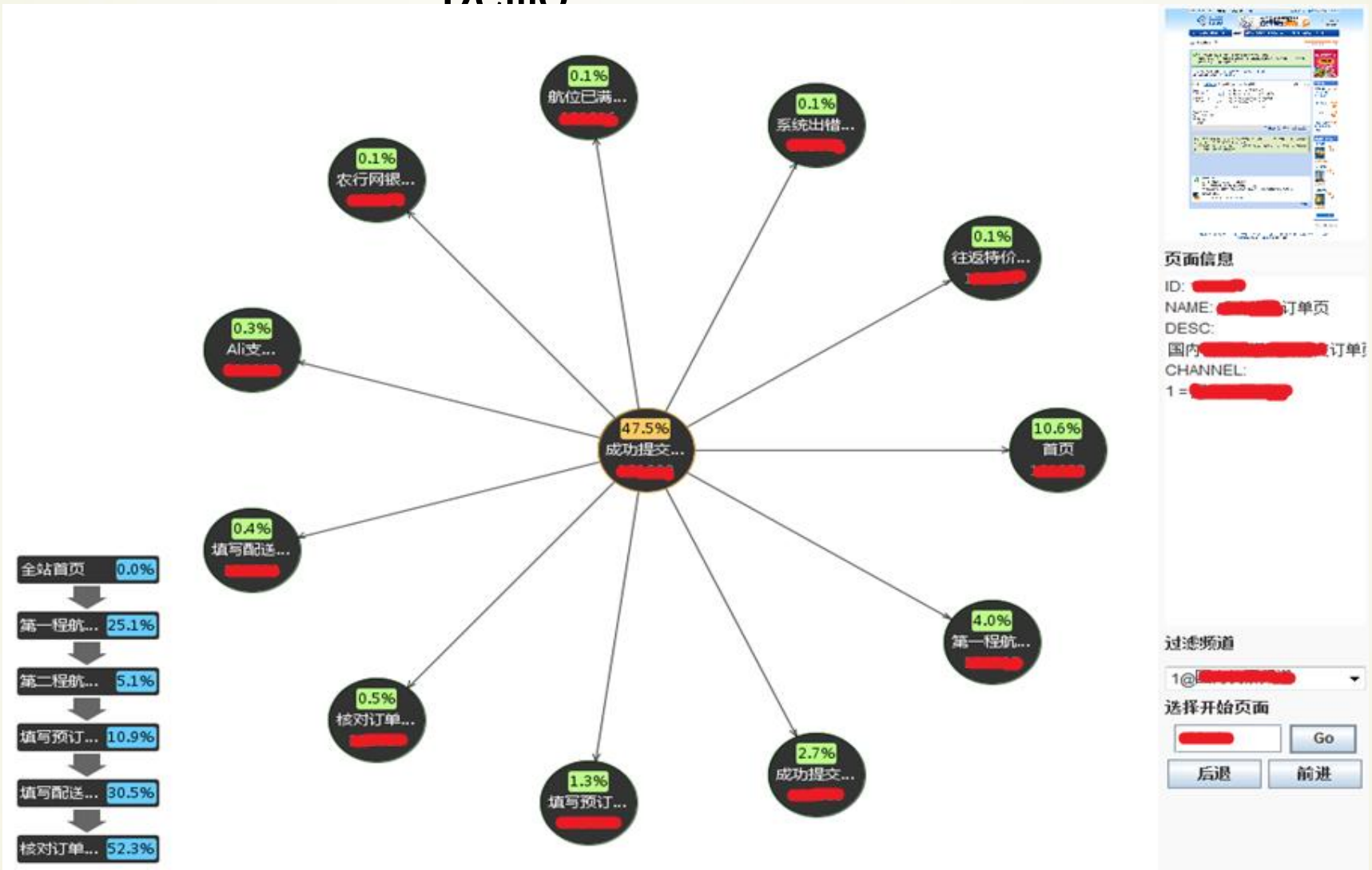
UBT - Page Speed Monitoring Demo



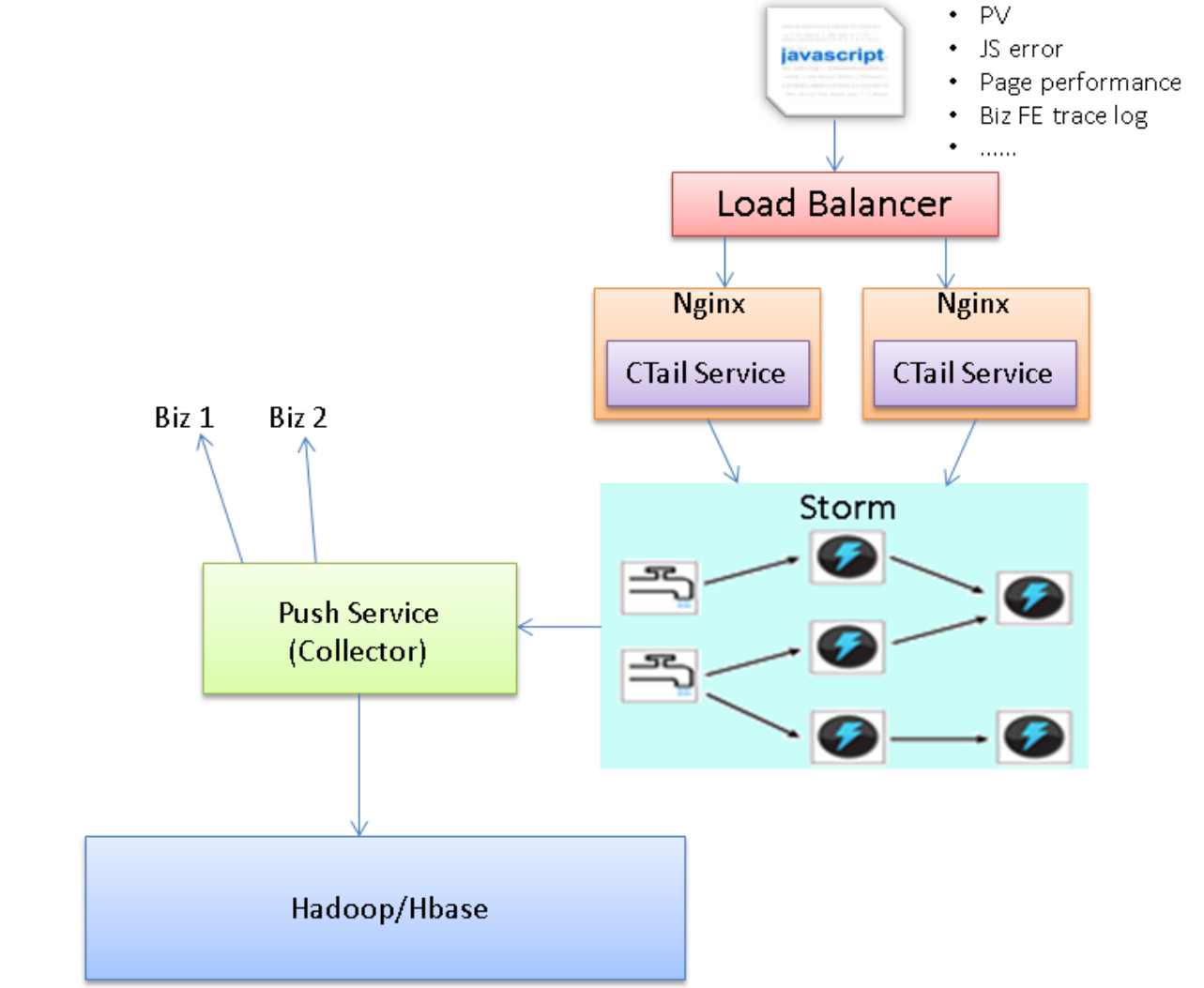
UBT - Real-Time Order Monitoring Demo



UBT - Real-Time Order Monitoring Demo



UBT - Architecture Overview



UBT - Data Collection

- 前端页面JS埋点（JS框架支持）
- Nginx接收回传数据，输出到syslog
- Ctail服务读取syslog，作为spout数据源
  - 支持游标定位
  - 支持数据重放
  - 支持日志轮转
  - 支持压缩传输

UBT - Storm Stream Processing

- 采用transaction方式(Trident Storm)
- 故障自动回复
- 实时统计，metrics写入dashboard
  - 每个session内所有感兴趣的事件
  - 外站来源统计
  - 主要页面的转换率
  - A/B测试结果
  - ……
- 基础UBT数据实时推送到UBT Push Service
  - 保存到Hadoop/Hbase
  - 推送给某些实时业务服务

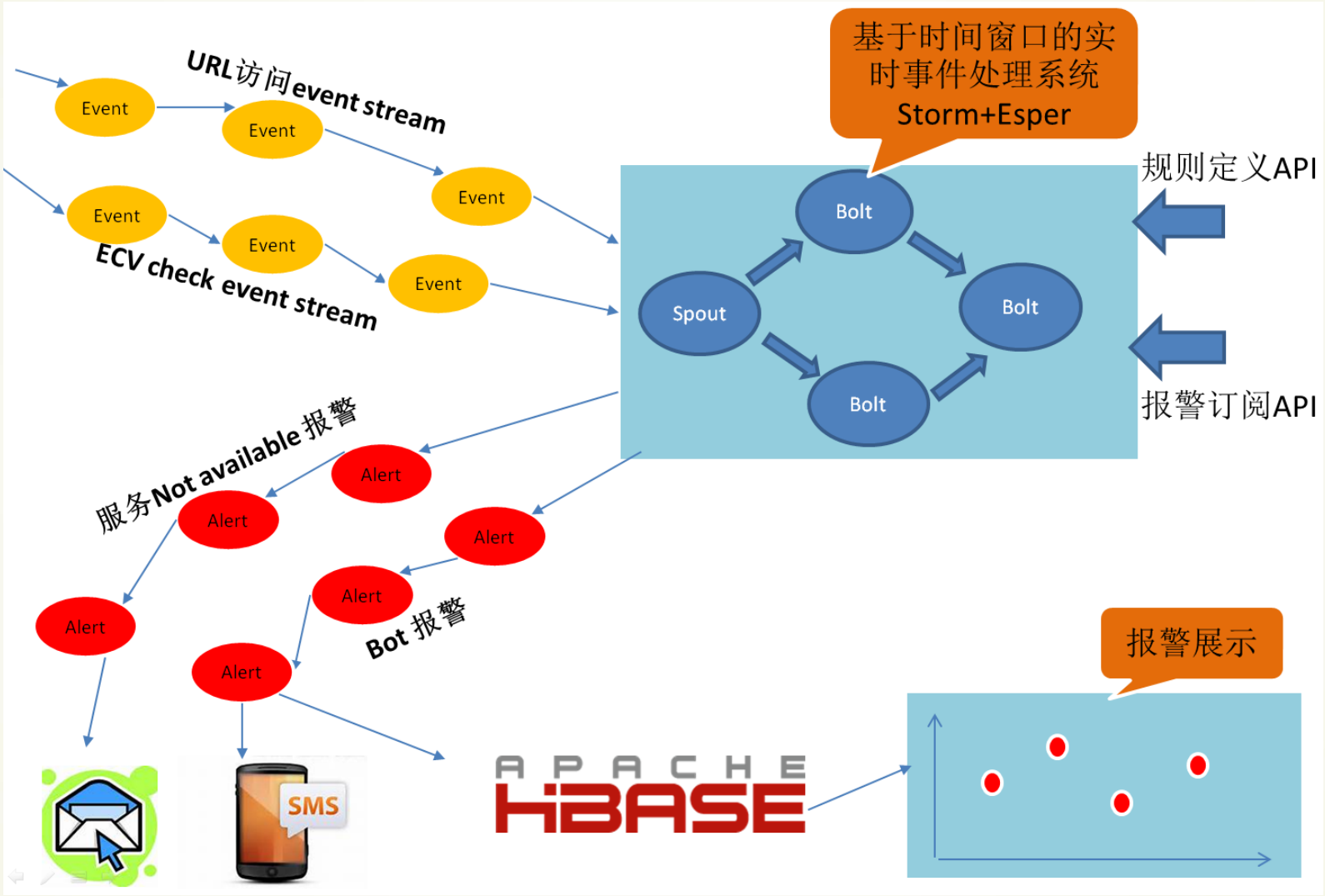
Agenda

1. Brief Introduction
2. Central Logging
3. Log View
4. Dashboard
5. User Behavior Tracking
6. Alerting
7. Data Farm

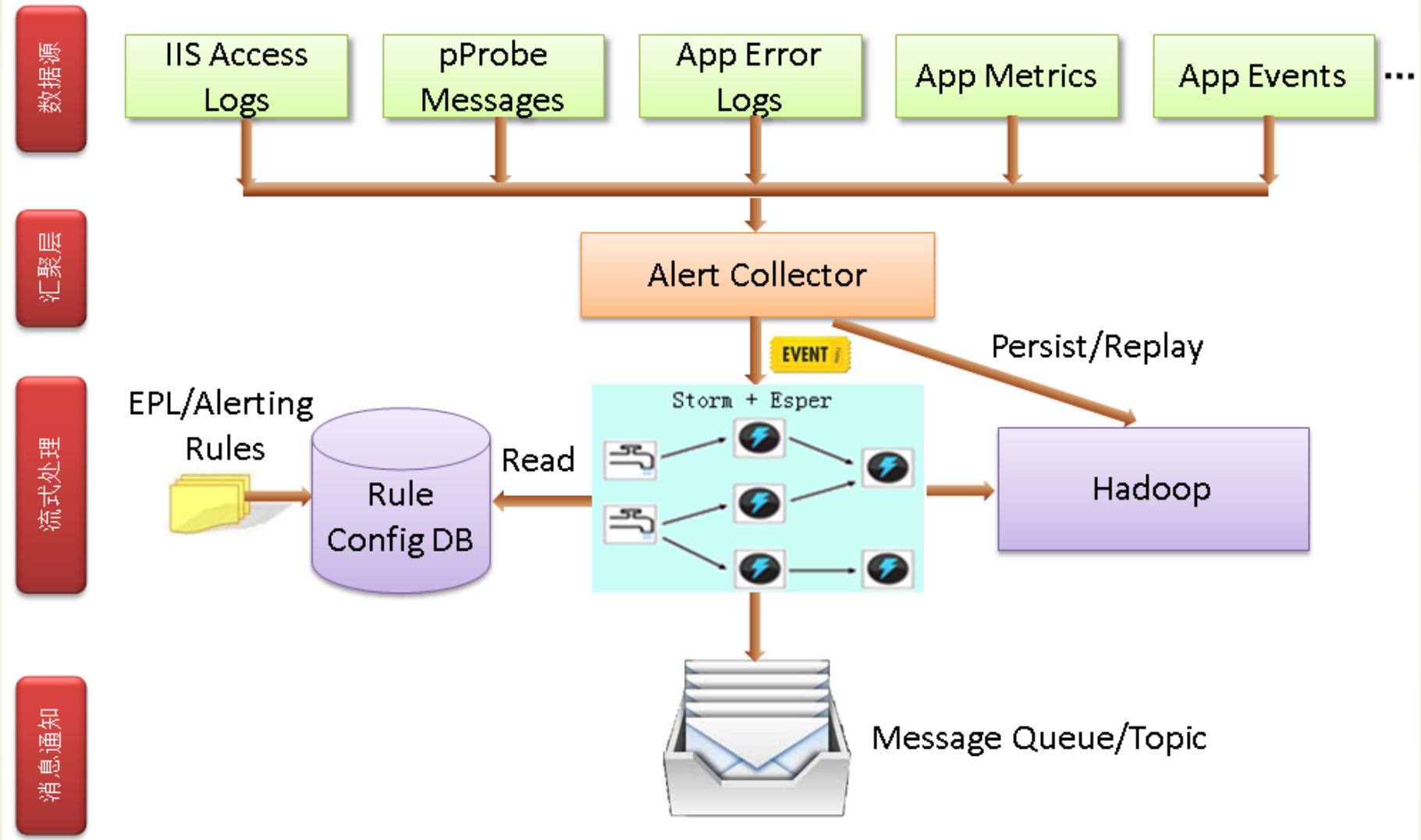
Alerting

- 业务异常报警
- 应用健康报警
- 系统异常报警
- DDoS攻击检测报警
- 恶意爬虫检测报警
- ……

Alerting - Framework



Alerting - Framework



Alerting - Event Processing Design

- 通用Event定义，支持namespace
- 通用Storm拓扑设计，支持分片逻辑
- 旁路Event数据至Hadoop（replay、dry run）
- EPL规则管理，规则跟新无需重启Storm
- 基于Event时间戳实时排序

Alerting - Rule & Alerts

规则

```
insert into _ipAlarm select ip, urlid, sum(count) as count, min(time) as starttime, max(time) as endtime,
"severe" as level, "ip" as type from Ips.win:time([redacted]) group by ip, urlid having sum(count) >= [redacted]
insert into _ipAlarm select ip, urlid, sum(count) as count, min(time) as starttime, max(time) as endtime,
"medium" as level, "ip" as type from Ips.win:time([redacted]) group by ip, urlid having sum(count) >= [redacted] and
sum(count) < [redacted]
insert into _ipAlarm select ip, urlid, sum(count) as count, min(time) as starttime, max(time) as endtime,
"small" as level, "ip" as type from Ips.win:time([redacted]) group by ip, urlid having sum(count) >= [redacted] and
sum(count) < [redacted]
```

报警

ip	level	count	url_count	type	urlid	starttime	endtime	urlcategory
[redacted]	severe	294	147	ip_category	0	0320-17:42:00	0320-18:11:00	[redacted]/international
[redacted]	severe	286	5	ip_category	0	0320-17:27:00	0320-17:50:30	[redacted]/international
[redacted]	severe	253	127	ip_category	0	0320-17:12:00	0320-17:41:00	[redacted]/international
[redacted]	severe	239	22	ip_category	0	0320-17:28:00	0320-17:56:30	[redacted]/international
[redacted]	severe	226	8	ip_category	0	0320-17:19:00	0320-17:35:30	[redacted]/international
[redacted]	severe	221	17	ip_category	0	0320-17:31:00	0320-17:56:00	[redacted]/international
[redacted]	severe	215	18	ip_category	0	0320-17:24:30	0320-17:52:00	[redacted]/international
[redacted]	severe	209	5	ip_category	0	0320-17:42:30	0320-18:11:00	[redacted]/international
[redacted]	severe	197	13	ip_category	0	0320-17:29:00	0320-17:58:00	[redacted]/international
[redacted]	severe	187	13	ip_category	0	0320-17:27:30	0320-17:54:00	[redacted]/international
[redacted]	severe	178	20	ip_category	0	0320-17:14:00	0320-17:42:00	[redacted]/international
[redacted]	severe	176	24	ip_category	0	0320-17:39:30	0320-18:08:30	[redacted]/international
[redacted]	severe	174	11	ip_category	0	0320-17:27:30	0320-17:56:30	[redacted]/international
[redacted]	severe	174	16	ip_category	0	0320-17:37:30	0320-17:56:30	[redacted]/international
[redacted]	severe	168	25	ip_category	0	0320-17:30:00	0320-17:54:00	[redacted]/international
[redacted]	severe	166	83	ip_category	0	0320-17:16:00	0320-17:45:00	[redacted]
[redacted]	severe	164	5	ip_category	0	0320-17:12:00	0320-17:41:00	[redacted]/international
[redacted]	severe	150	1	ip_category	0	0320-17:15:30	0320-17:44:00	[redacted]/international
[redacted]	severe	148	14	ip_category	0	0320-17:35:00	0320-17:56:00	[redacted]/international

Agenda

1. Brief Introduction
2. Central Logging
3. Log View
4. Dashboard
5. User Behavior Tracking
6. Alerting
7. Data Farm

Data Farm - Building

- 数据仓库
- 各类数据整合
  - 应用日志数据
  - 业务数据
  - 用户行为数据
  - 系统日志数据
  - 各类度量数据
  - ……
- 数据挖掘、分析

Data Farm - Planning

- 各类数据的ETL
- 全面启动ACL，保护数据安全性
- 基于任务依赖调度（Oozie?）
- 执行引擎选择（Hive, Impala, Shark, Phoenix？）
- 优化数据处理

Thanks
Q&A

欢迎莅临

2013中国数据库技术大会