

携程风控

实时大数据处理经验分享

郑佰云

1

风控需要哪些数据

2

风控处理数据中的挑战

3

聊聊风控的3个核心服务

只要和旅游相关的，携程都做。。。。

机票

酒店

旅游

火车票

汽车票

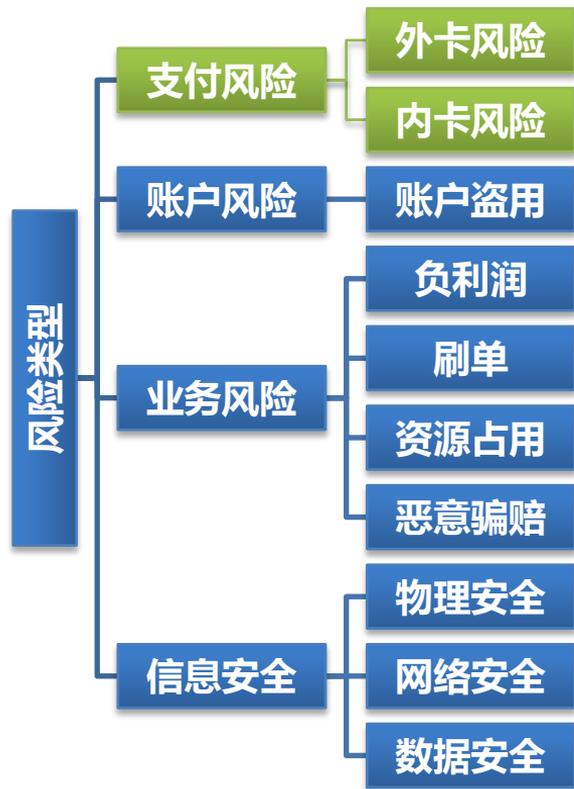
门票

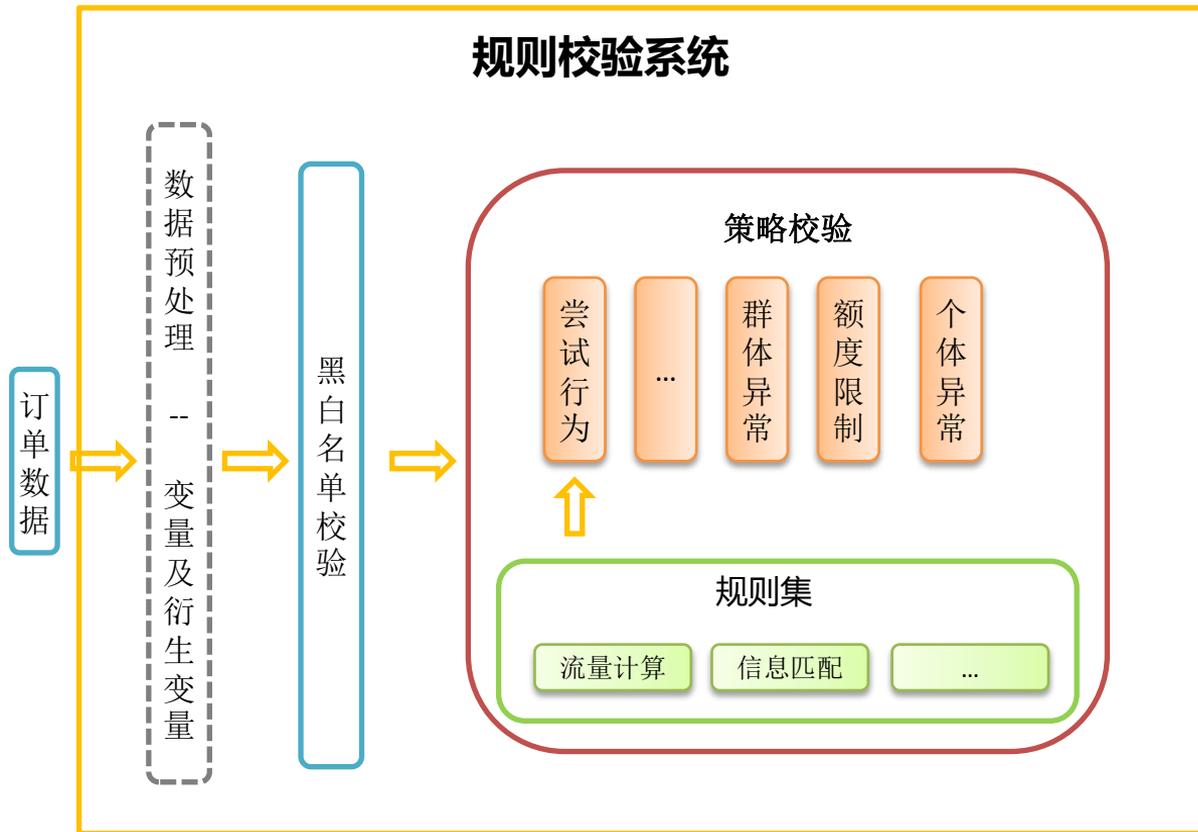
用车

团购

邮轮

礼品卡





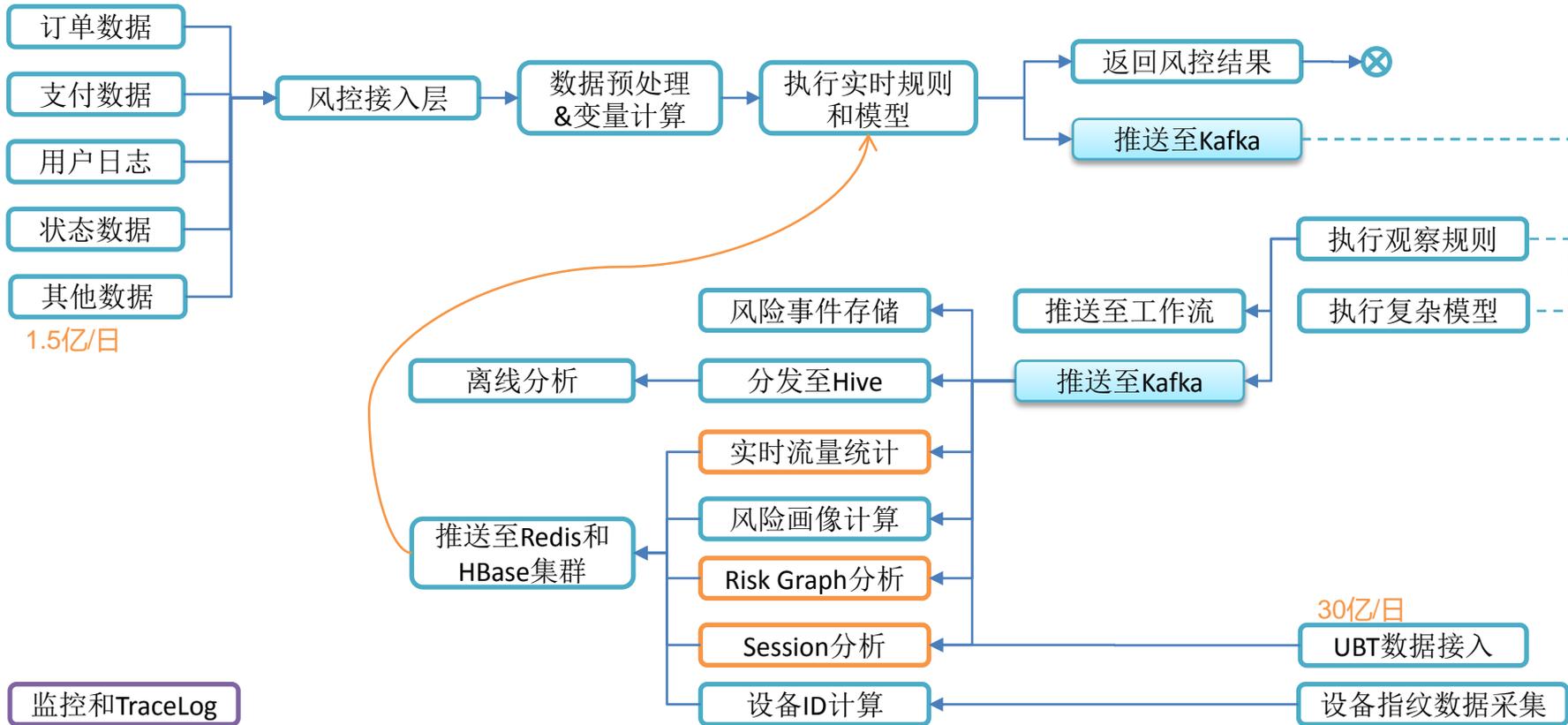
我们的挑战：

规则复杂：规则执行需要依赖非常多的外部数据、以及大量的实时统计数据；

数据复杂/分散：接入全站各个业务线的订单数据、支付数据、状态变更事件、登录/注册事件、操作日志、营销活动、用户行为数据等，且数据量大；数据之间需要共享；

以**实时风控**为主，有极高的响应要求；

数据流



一些例子：

一个uid，24小时（精确到分钟级）内的订单数量（根据订单号去重、包含当笔交易）

同一个设备7天内、登录的UID数量

酒店产线，90天内，同一ip，订单数量

同一UID，10分钟内密码错误次数

特性：

统计维度灵活、复杂；

支持不同精度；

数据量大、访问量大；

滑动窗口跨度大；

有去重的需求；

规则灵活多变、参数/阈值调整频率高；

需要快速响应需求的变化；

规模：

目前，我们的流量服务日查询量300亿次以上

我们是怎么做的：

1. 使用Redis存储流量数据
2. 多实例、分布式部署，提升吞吐量
3. 充分利用Redis的原子操作，提升性能

举例：

同一uid一天内（近1440分钟）的订单量（按orderId去重）

实现：

Redis里保存分钟精度、小时精度的统计值、以及统计值对应的orderId的集合（Set）

查询流量时先把近一天（近1440分钟）折算出最少窗口数

例如： 2016-06-25 18:07 - 2016-06-26 18:06

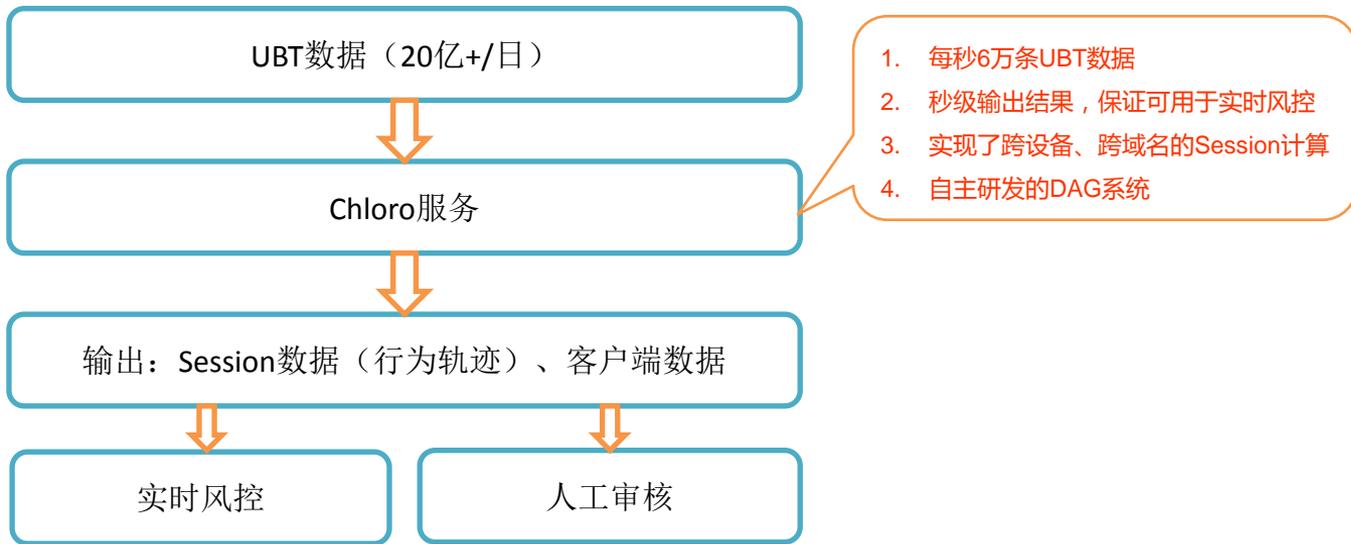
合并后：

2016-06-25 18:07 - 2016-03-25 18:59 : 53分钟

2016-06-25 17:00 - 2016-03-26 17:00 : 23小时

2016-06-26 18:00 - 2016-03-26 18:06 : 7分钟

最后对上面的83个KEY做sunion（并集）操作



vid	sid	uid	订单号	开始结束时间	访问时长	APP版本	分辨率	设备	城市/IP	网络
2A395980C5CC11E58C3598E7F54F6CF8	86	1183510747	2495056923	2016-07-26 19:25:43 2016-07-26 19:33:37	7分钟55秒	6.18.2	1080*1776	HUAWEI RIO-AL00 Android 5.1	苏州 121.238.245.84	中国电信 wifi

Session id:86 访问34页

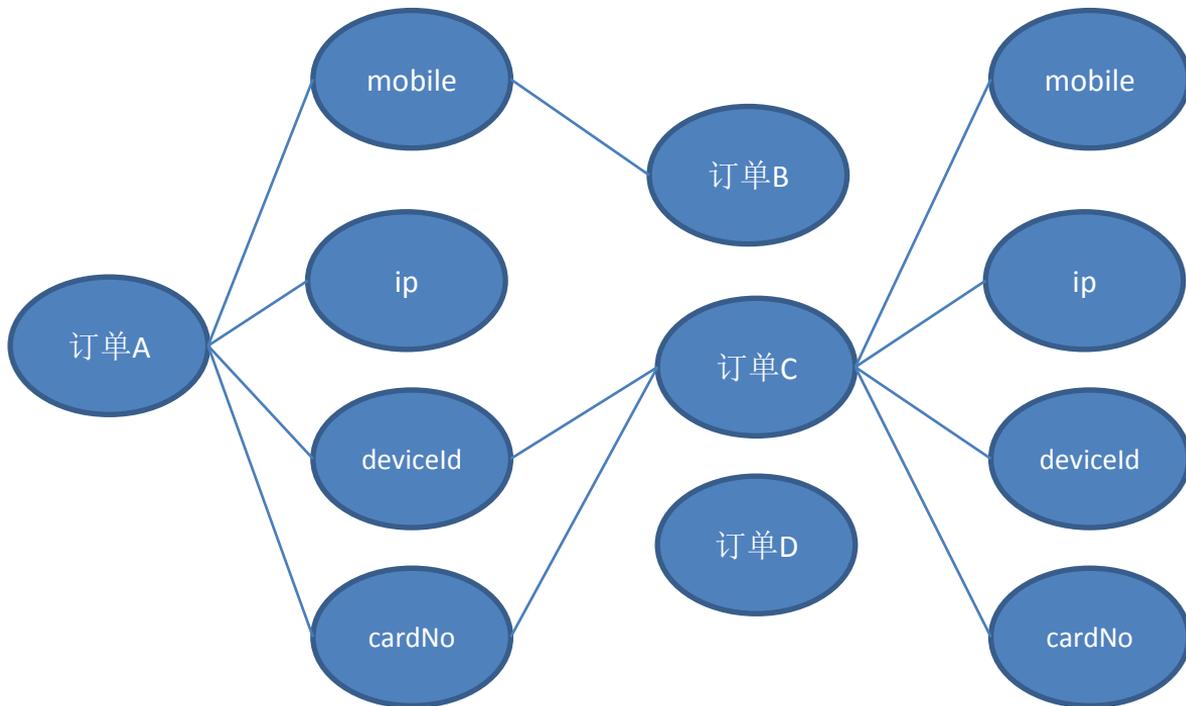
2126	2016-07-26 19:25:43	5"	native 基础业务Native 主页
2127	2016-07-26 19:25:48	3"	native 机票Native 国内机票查询页
2128	2016-07-26 19:25:51	2"	native 基础业务Native 城市列表页
2129	2016-07-26 19:25:52	2"	native 机票Native 国内机票查询页
2130	2016-07-26 19:25:54	1"	native 基础业务Native 城市列表页
2131	2016-07-26 19:25:55	32"	native 机票Native 国内机票查询页
2132	2016-07-26 19:26:27	6"	native 机票Native 国内机票查询页
2133	2016-07-26 19:26:33	31"	native 机票Native 国内机票单程列表页
2134	2016-07-26 19:27:04	1"	native 机票Native 国内机票中间页
2135	2016-07-26 19:27:05	5"	native 机票Native 国内机票单程列表页

用户行为分析



```
"geo" : {  
  "region" : "云南",  
  "countryIsoCode" : "CN",  
  "gpsLatitude" : 22.010317,  
  "isp" : "移动",  
  "gpsCity" : "景洪",  
  "longitude" : 102.702011,  
  "timeZone" : "Asia/Chongqing",  
  "latitude" : 25.051001,  
  "gpsLongitude" : 100.80134,  
  "country" : "中国",  
  "city" : "昆明"  
},
```

```
"attrs" : {  
  "meta_timezone" : "Asia/Shanghai",  
  "env_carrier" : "移动",  
  "env_preSourceId" : "4061",  
  "env_androidID" : "ac3d7cb68cb5bfd4",  
  "package" : "ctrip.android.view",  
  "app_version" : "6.21.0",  
  "env_deviceType" : "GN8003",  
  "env_appVersion" : "621.000",  
  "env_extendsourceid" : "",  
  "androidid" : "ac3d7cb68cb5bfd4",  
  "env_allianceid" : "67695",  
  "env_logtime" : "2016-11-26 11:14:46",  
  "env_deviceName" : "gncompiler",  
  "env_PushSwitch" : "1",  
  "env_isWechatWakeUp" : "false",  
  "env_screenWidth" : "1080",  
  "country_no" : "中国",  
  "orderId" : "282951...",  
  "meta_imei" : "8617...",  
  "env_sid" : "509726",  
  "requestId" : "S_APP_6166A0077410AE998169C3E2F8C5770F",  
  "env_imsi" : "460027...",  
  "businessType" : "101",  
  "env_screenHeight" : "1920",  
  "env_networkType" : "4G",  
  "view_time" : "2016-11-26 11:18:23",  
  "meta_sdkver" : "2.6.6",  
  "env_serialNum" : "OFFQTGG...",  
  "env_os" : "Android",  
  "env_mac" : "020000000000",  
  "seq" : "10468",  
  "country_language" : "中文",  
  "env_countryType" : "0",  
  "env_ouid" : "",  
  "env_clientcode" : "32001169110030346316",  
  "env_osVersion" : "6.0",
```



技术的演变：

2013年前，基于数据库查询，千万级

尝试过Neo4J等多个开源Graph方案

目前，需要回溯近1年的50亿订单数据...

基于HBase实现的Graph存储

核心思想是把Graph存储简化成KV存储，充分利用了HBase支持1:N数据存储特性

案件侦测&关联排查



订单编号 2529073195 命中5个:1 命中2个:2 命中1个:196 全部:199 查看所选:0

订单号	订单日期	订单类型	录黑 UID	联系人手机	邮箱	卡号	证件号	IP	DID	ClientID	提现卡号	提现证件号	操作
2529073195	2016-07-25	高铁	F	_WeChat403168883	18600000000		2300000000	14.136.203.107					
2435999878	2016-07-25	酒店	F	E14011920	13500000000			*		1202102041			关联
2433511525	2016-07-24	酒店	F	M413128776	15900000000			*		1200109431			关联
2469422620	2016-07-17	机票	F	*	13300000000		2300000000	113.0.223.16		1200108061			关联
2412163703	2016-07-15	酒店	F	M289167080	13900000000	64007987		*		1202114021			关联
2412153645	2016-07-15	酒店	F	M289167080	13900000000	64007987		*		1202114021			关联
2409787416	2016-07-15	酒店	F	_WeChat227024822	13400000000			*		1202116471			关联
2409481138	2016-07-14	酒店	F	M289167080	13900000000	64007987		*		1202114021			关联
2407634340	2016-07-13	酒店	F	_WeChat403422737	13000000000			*		1200106311			关联
36811210	2016-07-11	消费券	F	13827797667	13820000000			*		1200117781			关联
2398338132	2016-07-09	酒店	F	3000696130	18500000000		63854384	*		1200101341			关联
2264859130	2016-07-09	用车	F	1102803029	13600000000		59169486	*		0903112041			关联
2395956203	2016-07-08	酒店	F	ywsfj800720	13900000000		1069655	*		1200109941			关联
2395603351	2016-07-08	酒店	F	E33650902	18600000000			*		1200113941			关联
2394651946	2016-07-07	酒店	F	2056923792	18600000000			*	82e34d2e-c	1200109621			关联
36648993	2016-07-06	消费券	F	D117876249	15900000000			*		1202106981			关联
2391312371	2016-07-05	酒店	F	D117876249	15900000000	46553239		*		1202106981			关联
2442197254	2016-07-04	机票	F	M268797542	*		2300000000	117.40.225.183					关联
2442197240	2016-07-04	机票	F	M268797542	*		*	117.40.225.183	e3125fff-2				关联
2383107138	2016-07-01	酒店	F	3009781965	17700000000			*		1202113341			关联

1. 交易数据50亿+；
2. 关联速度小于1S；

谢谢大家