

Prism数据杂货铺

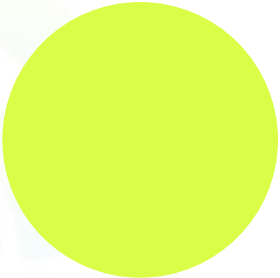
自我介绍



- 姓名: 徐磊
- 部门: 平台事业部
- 简要介绍: 负责日志平台设计和运维, Python/Java工程师。

目录

- 1 ▶ 为什么是杂货铺
- 2 ▶ 先定个小目标
- 3 ▶ 扩大后的经营问题
- 4 ▶ 覆盖更广的用户



一、为什么是杂货铺

Prism ?

- 以**数据可视化**为出发点
- 以**降低**数据和数据分析软件获取**成本**为目标的**实时**数据平台

如果Hadoop是Walmart



那我们.....是杂货铺



小店面

员工少

营业长

低库存，快销售

没保安

没导购

没仓库

没配送

没售后

杂货铺有什么？

进货



选购



上架



仓储



配送



结算



想象一下！



logstash



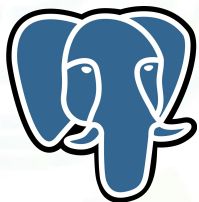
APACHE
STORM™



APACHE
kafka™
A distributed streaming platform



Flink



elasticsearch



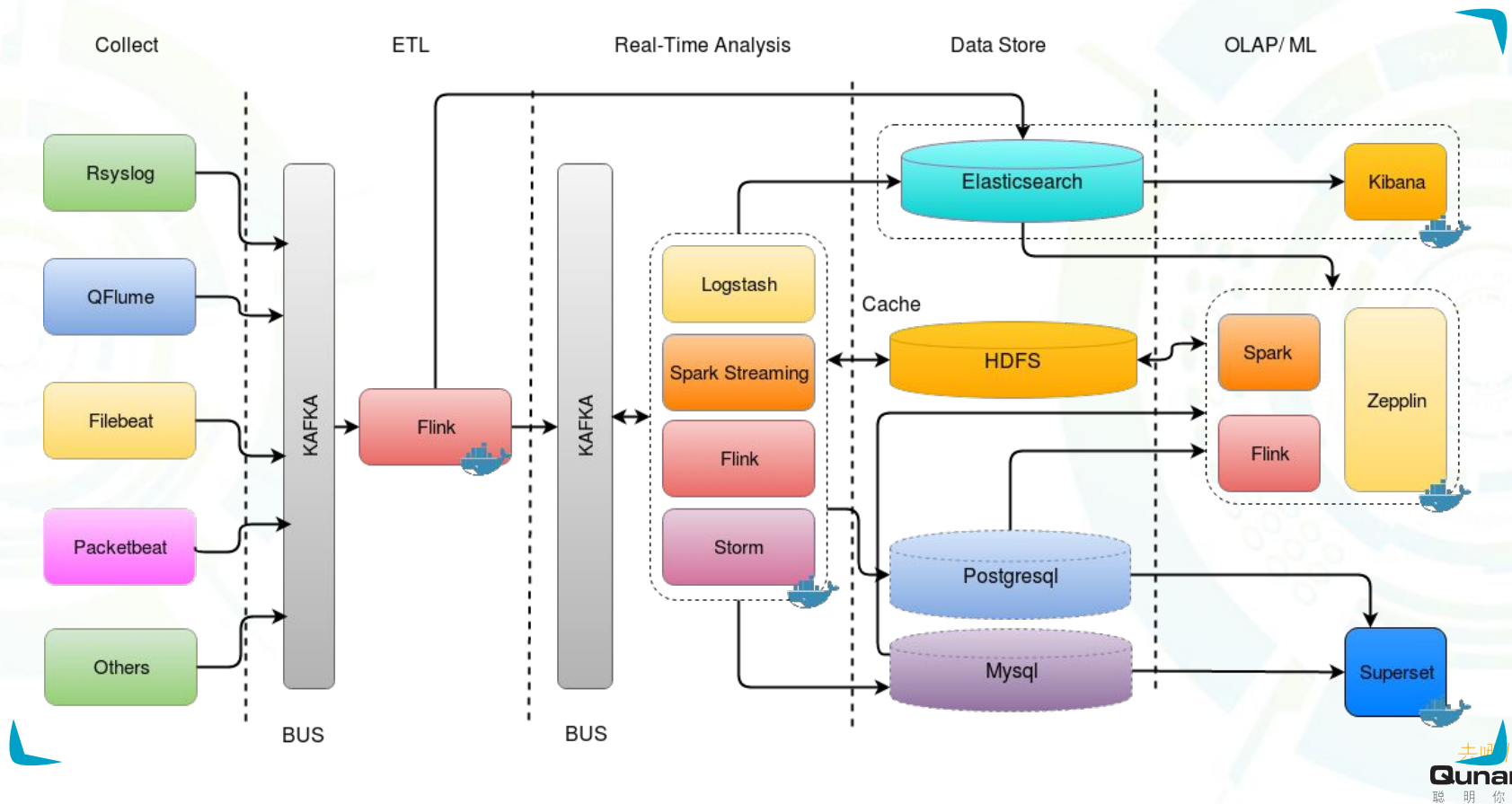
Superset

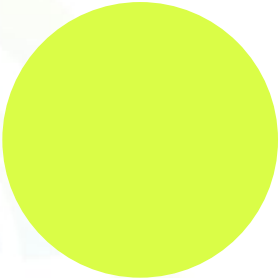
kibana

杂货铺的一天



我们的一天





二、先定个小目标

我们希望的



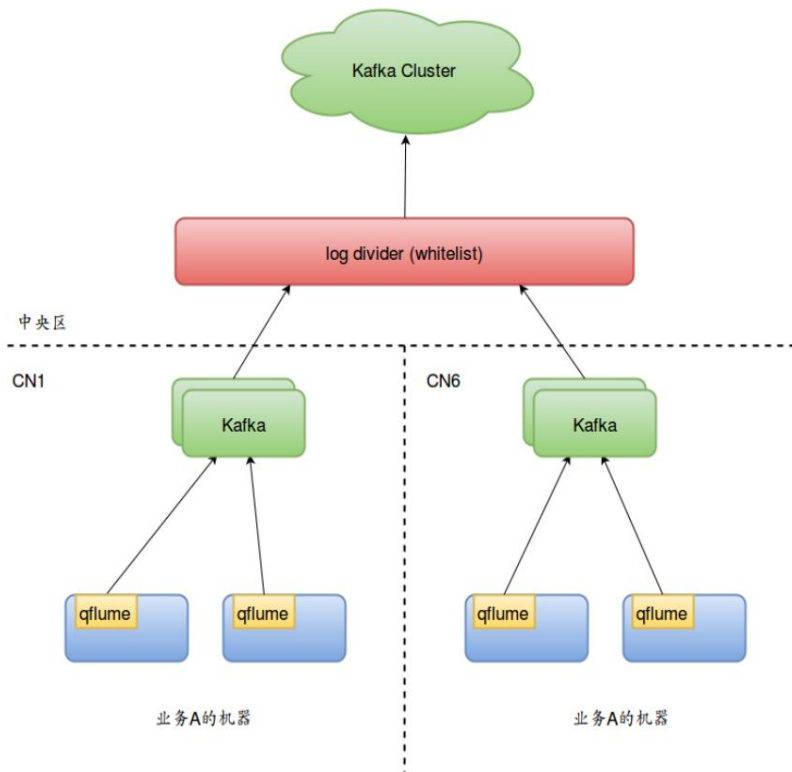
然而现实.....



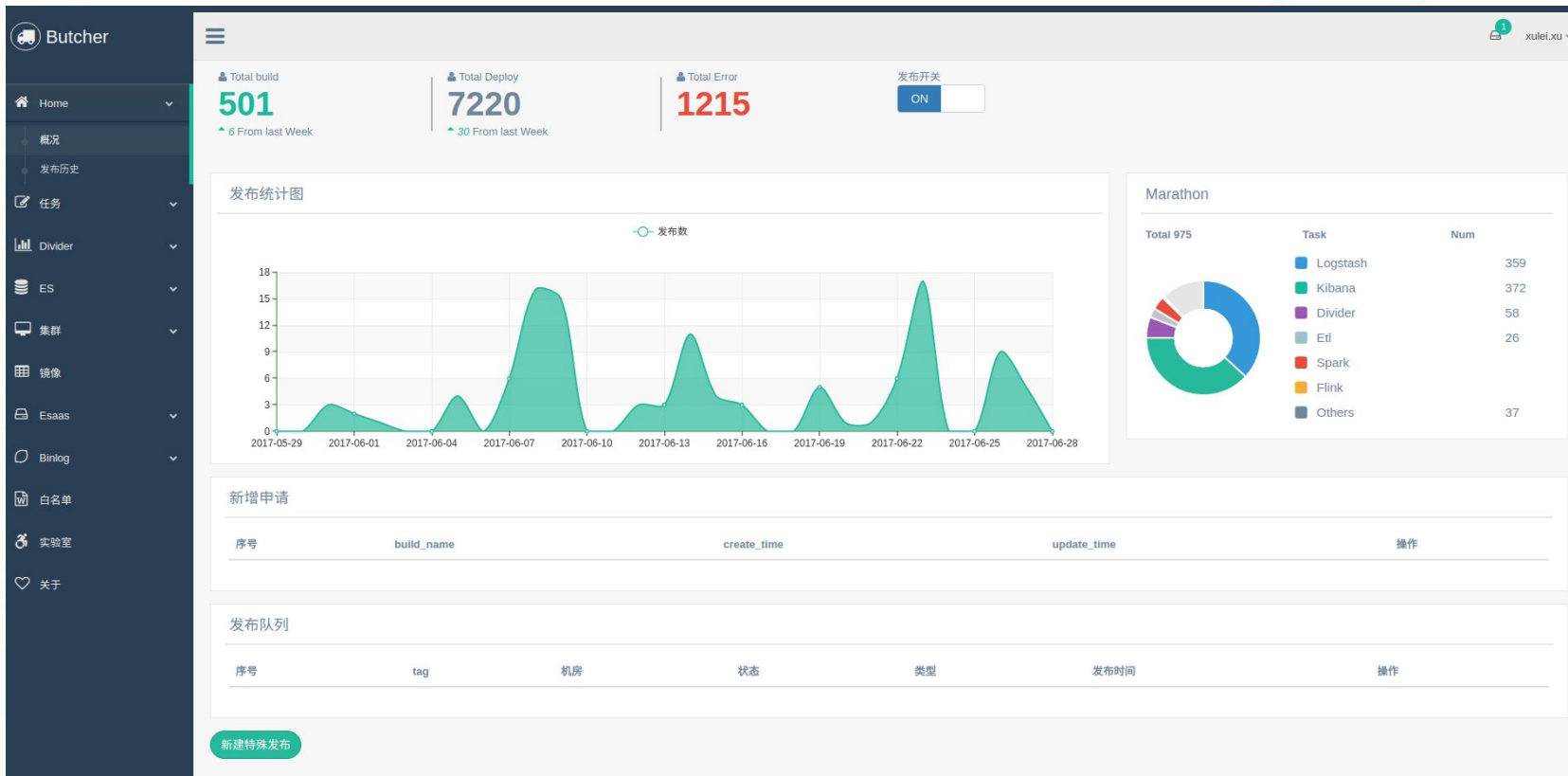
嗯.....先装修下！

买几个好货架

- Kafka Cluster
 - SSD卡
 - 万兆网卡
 - 每个机房一套集群
 - 24小时数据回放



买套ERP和扫描枪



摆几盆花草

Config 277

Init

Debug history

Filter by name

- el.digo.m_innovation_pitcher_poster >
- el.hotel.m_sitemap_report >
- el.train.m_train_maholding >
- el.train.m_web_hylin >
- el.usercenter.m_holding >
- el.wireless.m_innovation_pitcher_poster >
- el.wireless.m_invocation_flight >
- el.wireless.m_invocation_flight_smk >
- el.wireless.m_invocation_holding >
- el.wireless.m_invocation_hylin >
- logger.beta.car >
- logger.beta.flight >
- logger.beta.hotel >
- logger.beta.piao >
- logger.beta.sp >
- logger.beta.usercenter >
- logger.beta.wireless >
- logger.beta >
- logger.car.m_car_sawordtime >

Pattern -> [logs.logger.flight.m_car_data](#) [GitLab](#) [Watcher](#) [Kibana](#)

```
HTTPDATE %{MONTHDAY}/%{MONTH}/%{YEAR}:%{TIME} \+{%{INT}}
LOGTIME %{YEAR}\-%{MONTHNUM}\-%{MONTHDAY} %{HOUR}:%{MINUTE}:%{SECOND}
ERROR_KEYWORD \b(ERROR!.*?[eE]xception!.*?[Tt]imeout!Trying to connect to!WARN)
CLIENT ((unknown!%{IPV4}),[ ]?)*(%{IPV4})!-)
```

Input

☐ 自定义input字段

1

Filter

```
1 filter {
2   # ruby {
3   #   code => "event['log_duration'] = (Time.parse(event['@timestamp']).to_s).to_f * 1000).to_i - event['send_time'].to_i"
4   # }
5 }
```

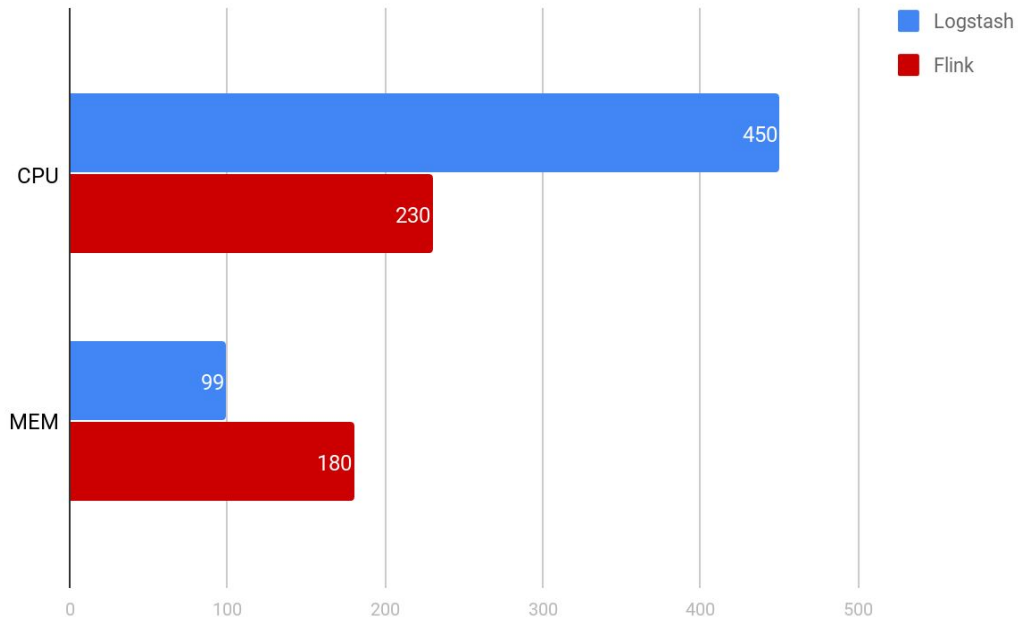
Debug

Commit

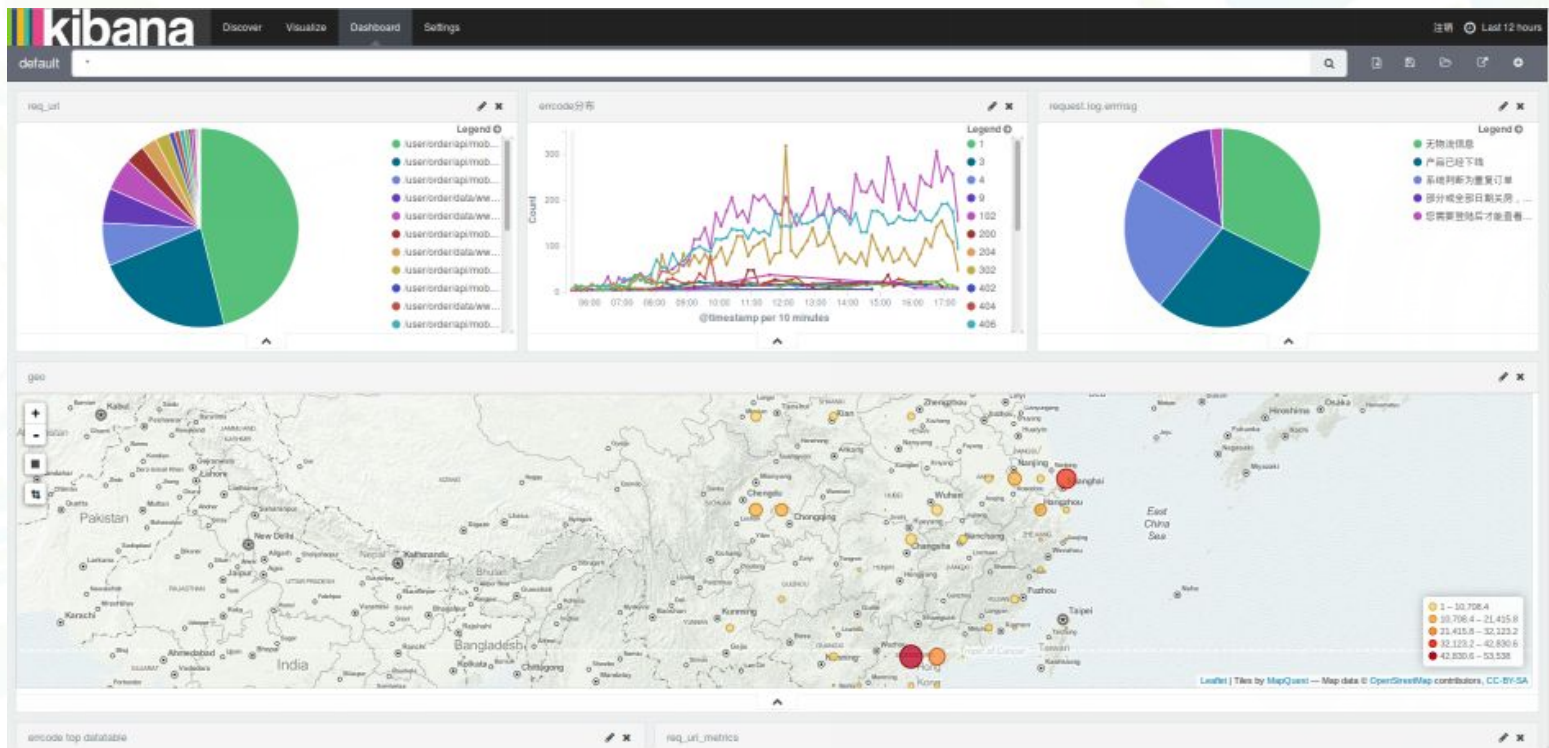
Deployment

```
scrot -s /tmp/s.png
> ~ scrot -s /tmp/s.png
> ~ scrot -s /tmp/s.png
```

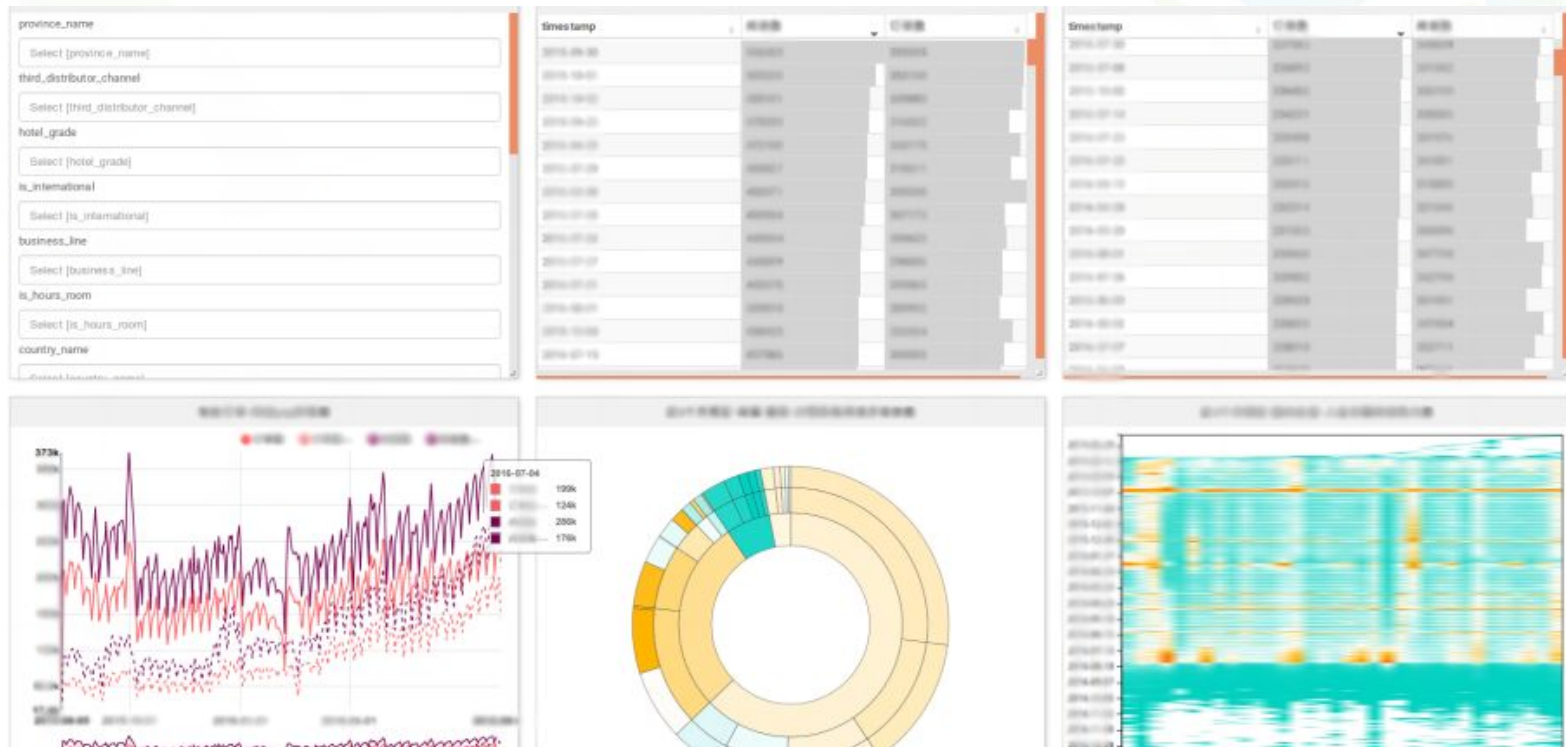
雇个强壮的工人



微信支付——Kibana



支付宝——Superset



刷脸——Zeppelin

```
import org.apache.spark.ml.Pipeline
import org.apache.spark.ml.classification.RandomForestClassificationModel, RandomForestClassifier
import org.apache.spark.ml.evaluation.MulticlassClassificationEvaluator
import org.apache.spark.ml.feature.IndexToString, StringIndexer, VectorIndexer
import org.apache.spark.sql.DataFrame

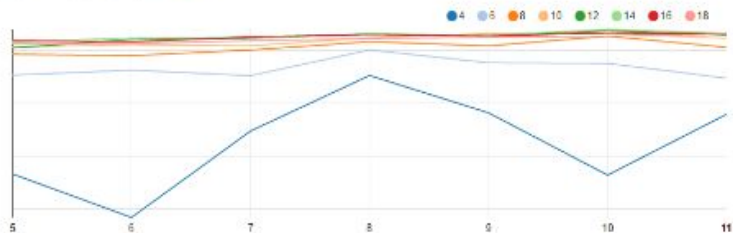
// def learnByRandomForestCls(data:DataFrame, features:Array[String]){
//   val model = buildRandomForest(data, features)
//   printModelSummary(model)
// }

// learnByRandomForestCls(inst_conf_ds, inst_conf_fs)
// numTrees: array
// maxDepth: array

val numTreesArr = Array(5,6,7,8,9,10,11)
val maxDepthArr = Array(4,6,8,10,12,14,16,18)
val (modelsOf_models, featureImportances) = createRModels(inst_conf_ds, inst_conf_fs, numTreesArr, maxDepthArr)
models.registerTempTable("RModels")
```

```
%sql
select maxDepth, numTrees, F1
from RModels
group by numTrees, maxDepth, F1
order by maxDepth
```

settings



Took 3 seconds (outdated)

```
%sql
select *
from RModels
order by F1 desc
```

numTrees	maxDepth	P	R	F1
10	14	0.8717451859028703	0.8717451859028703	0.870850190543188
10	12	0.8710185297323493	0.8710185297323493	0.869024252003352
10	16	0.8692330144120141	0.8692330144120141	0.8669011752970884
11	14	0.868909670582535	0.868909670582535	0.8656107860962191
9	10	0.8692018893060434	0.8692018893060434	0.866254981244555
10	10	0.8684752381350213	0.8684752381350213	0.8650605166480937
8	14	0.867748570945994	0.867748570945994	0.8659777723374367
8	12	0.8678605863267031	0.8678605863267031	0.8658673730056035

Took 0 seconds (outdated)

顾客越来越多
经营也跟得上

62

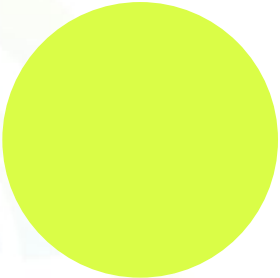
bil/day

700

TB

350

apps



三、扩大后的经营问题

我要买的货物在哪儿



服务台/指引牌

元数据管理

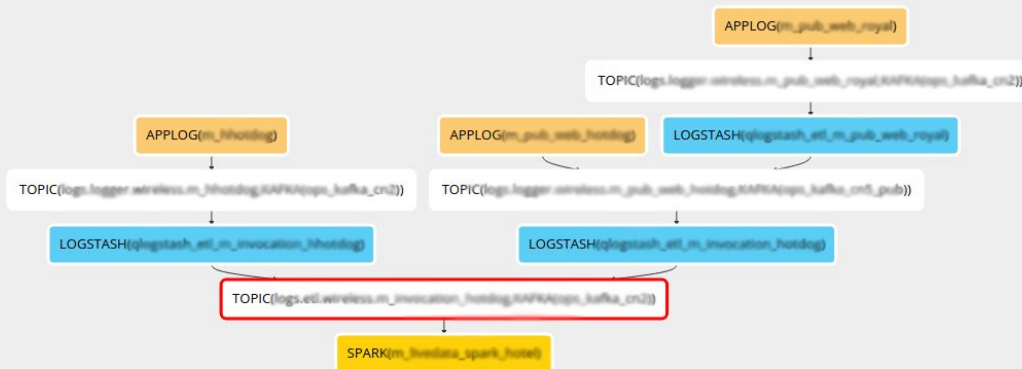
Q META查询

data:logs.etl.wireless.m_invocation_hotdog

Q 查询

查询完毕!

APPLOG LOGSTASH SPARK FLINK EXTERNAL NONE DATA



节点类型

TOPIC(logs.etl.wireless.m_invocation_hotdog)

TOPIC(logs.etl.wireless.m_invocation_hotdog)

★ 属性 ■ 值

★ host_app_code ■ qps_kafka_cn2

★ 名称 ■ logs.etl.wireless.m_invocation_hotdog

★ 附加信息
belong_app_code m_livedata_spark_hotel
topic_level app_code

加强安保工作

- Kafka
 - 处理时延
 - 消息积压
 - JVM
- Elasticsearch
 - Segment
 - Cache
 - JVM
- Spark
 - Schedule delay
 - Process delay



售前与售后

- 配合业务线
 - 验收测试
 - 数据迁移
- WIKI与使用手册
 - Elasticsearch Admin手册
 - Prism使用手册
 - ESaaS使用手册
- 机器人
 - 内网小Q
 - 小拿

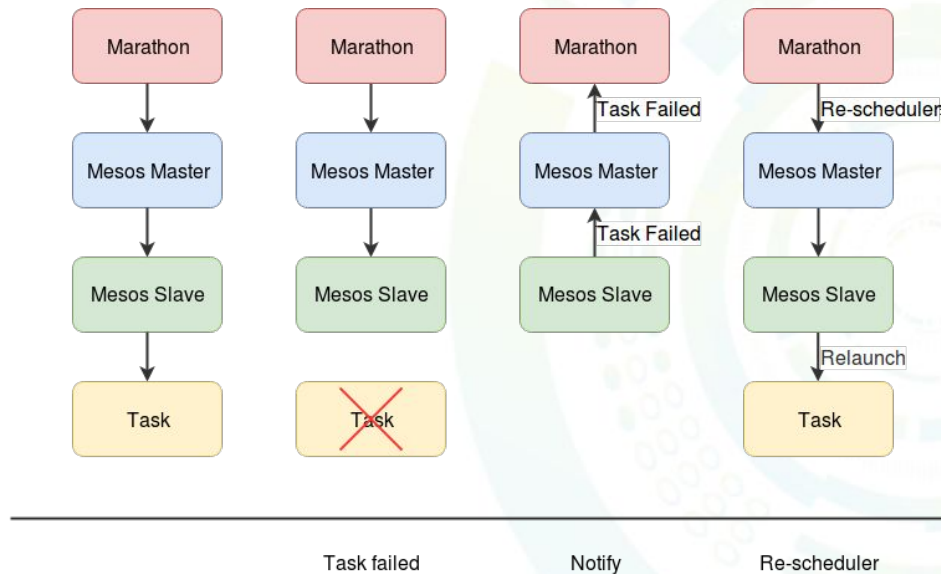


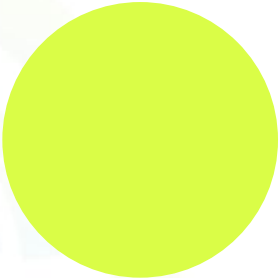
突然宕机了.....



Auto failover

- systemd/runit/supervisor
- task re-scheduler
- health check
- dynamic reservation



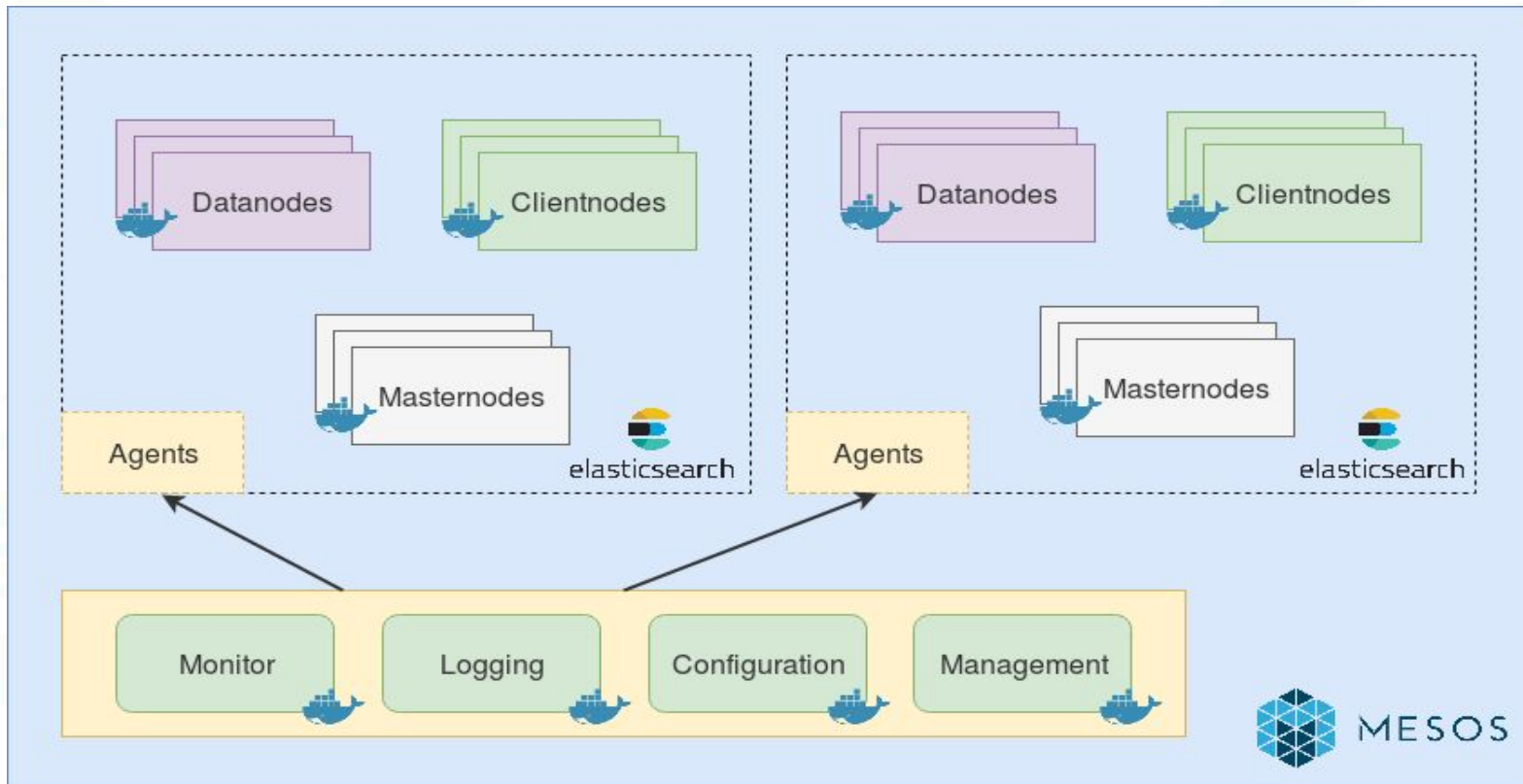


四、覆盖更广的客户

成为数据界的7-11



Elasticsearch as a Service



细分顾客喜好

- 大搜
 - 首页卡片
 - 猜你喜欢
- 风控
 - 反欺诈
 - 事件库
- 更重要的服务——施工中
 - QSS
 - 用户订单



去哪儿攻略-旅游景点.
机票酒店门票. 周边...

获取

Beijing Qunar Informa...

★★★★★ (109)



TOP旅行

众包到家

- Spark/Storm/Flink服务
 - 多版本任意选
 - 集群动态扩容
 - 集成Jenkins
- Logstash
 - 新业务不推荐使用



订单跟踪

- Nginx入站Trace
 - 已经接入大部分日志
 - 域名统计
 - 响应时间
 - 状态码
 - Cookies
- 流量拓扑
 - coming soon !



更多细节请移步

- [Prism 使用手册](#)
- [ESAAS 使用手册](#)
- [Spark on Mesos 使用手册](#)

祝大家购物愉快

Q&A