

ES在云上APM中的应用实践

腾讯云-高斌龙

背景

- **通用APM平台**

- App Crash分析**

- 1. crash时用户操作的路径
 - 2. crash时内存趋势
 - 3. crash是否与特定机型、系统版本相关

- App 内用户行为分析**

- 1. 如何量化新特性被用户喜爱

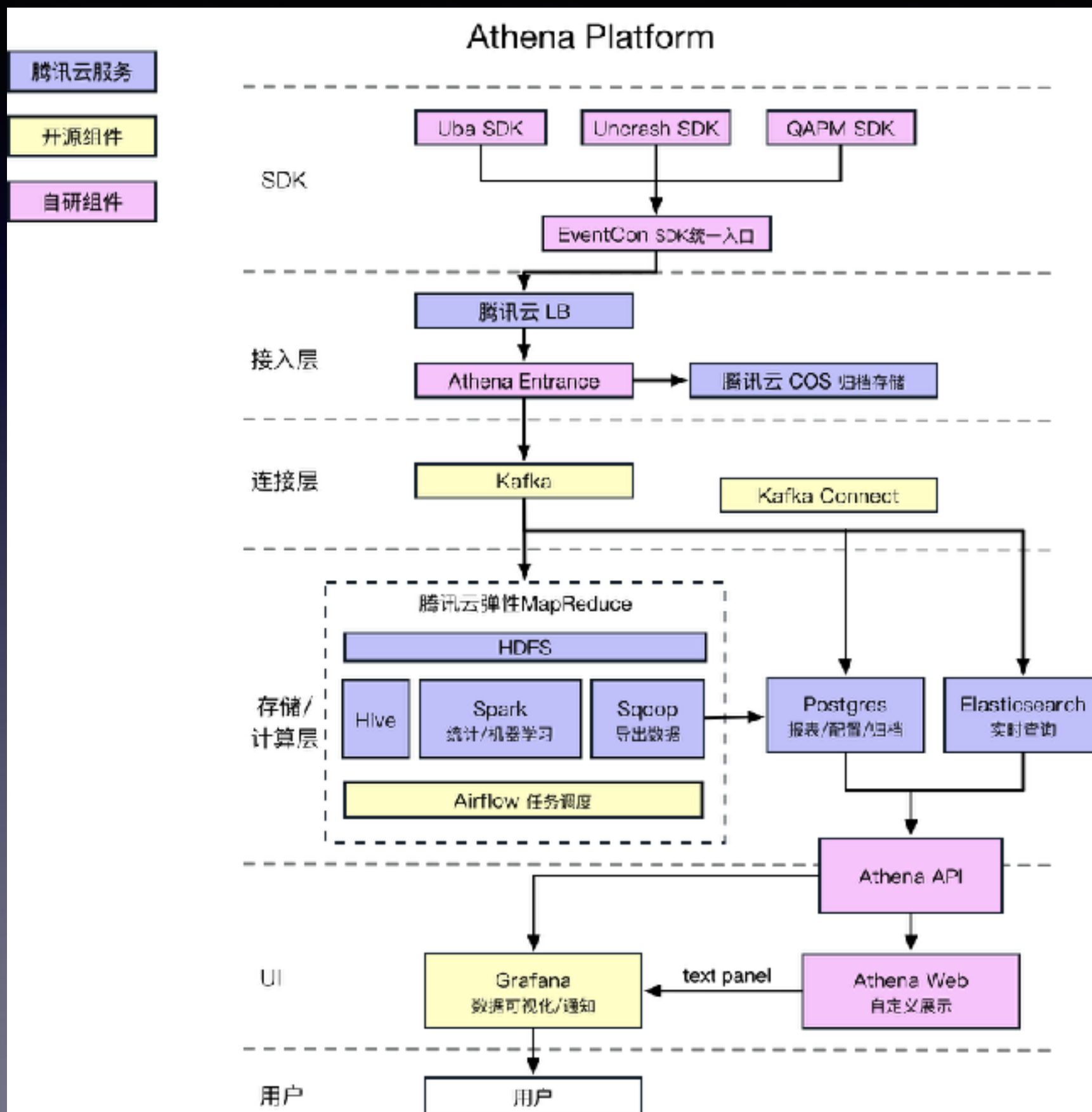
- App内测试**

- 1. 哪些路径容易crash
 - 2. 哪些路径用户最常访问

- **特点**

- 1. 多个异构数据源
 - 2. 服务于各个产品线(手机QQ,微视,QQ空间等)
 - 3. 实时监控

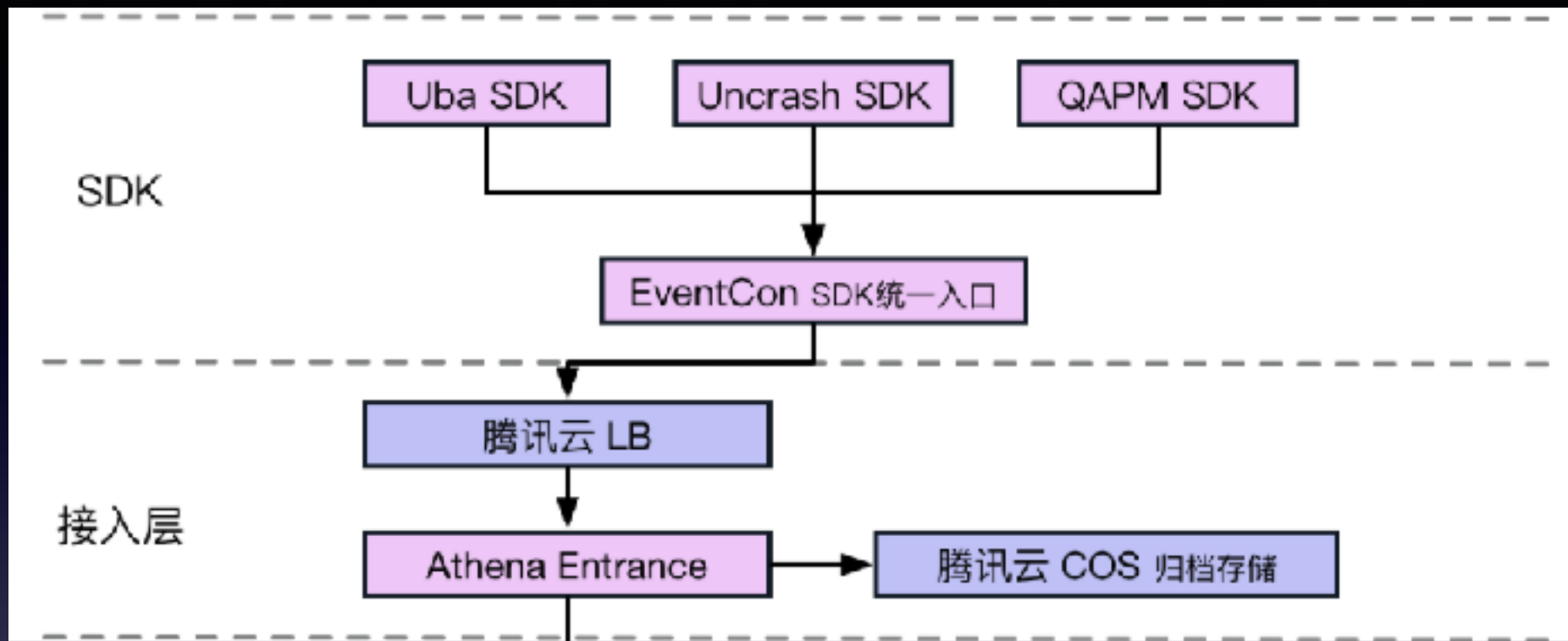
架构



云服务

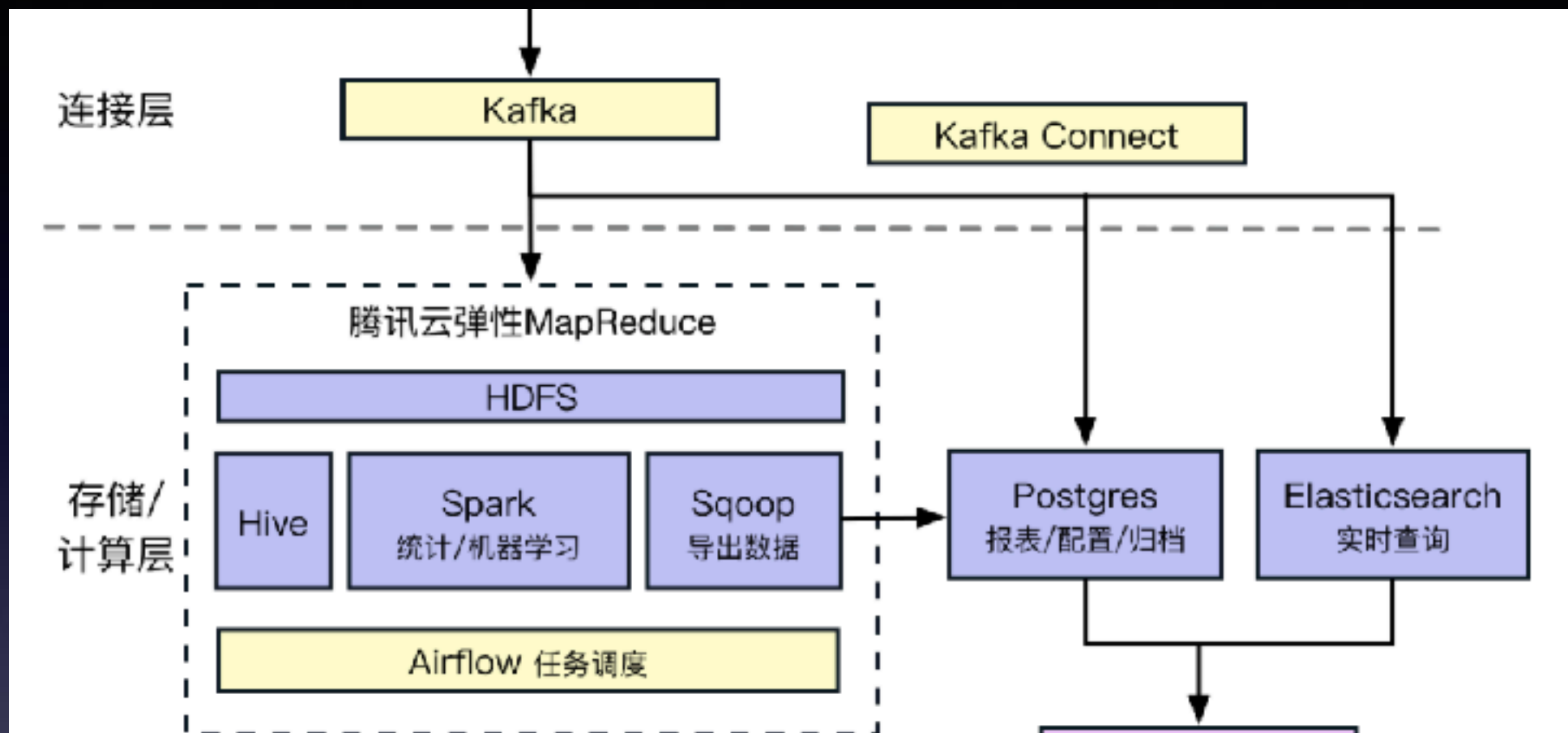
- **负载均衡CLB:** 数据接入服务的负载均衡
- **对象存储COS:** 源数据归档、ES数据快照
- **容器服务CKE:** 部署AirFlow, Grafana等服务
- **云数据库PostgreSQL:** 报表、归档
- **弹性mapreduce:** 计算引擎
- **Kafka:** 消息队列
- **Elasticsearch:** 实时查询

实现过程1:数据接入层



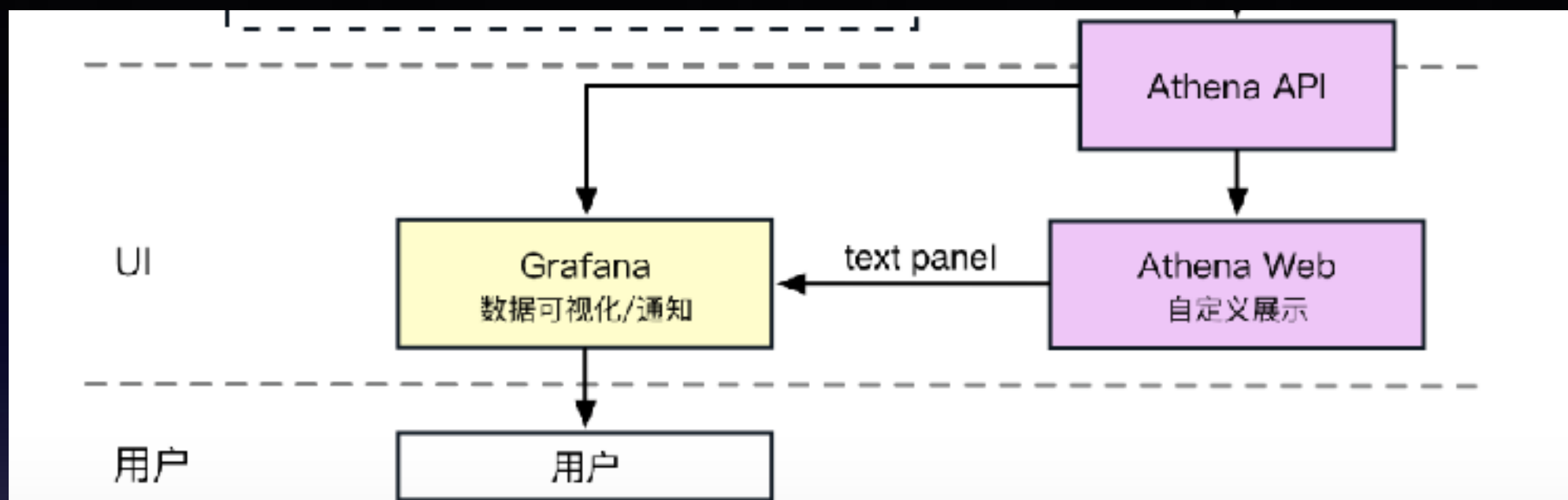
- 定义数据模型，通过SDK上报
 - Uba SDK: 采集用户行为数据
 - Uncrash SDK: 采集crash数据
 - QAPM SDK: 采集性能数据
- EventCon SDK统一上报入口
- Athena Entrance数据上报后台服务, 容器化部署, CLB提供容器公网访问能力
- 原始数据统一归档到对象存储COS中

实现过程2:存储/计算层



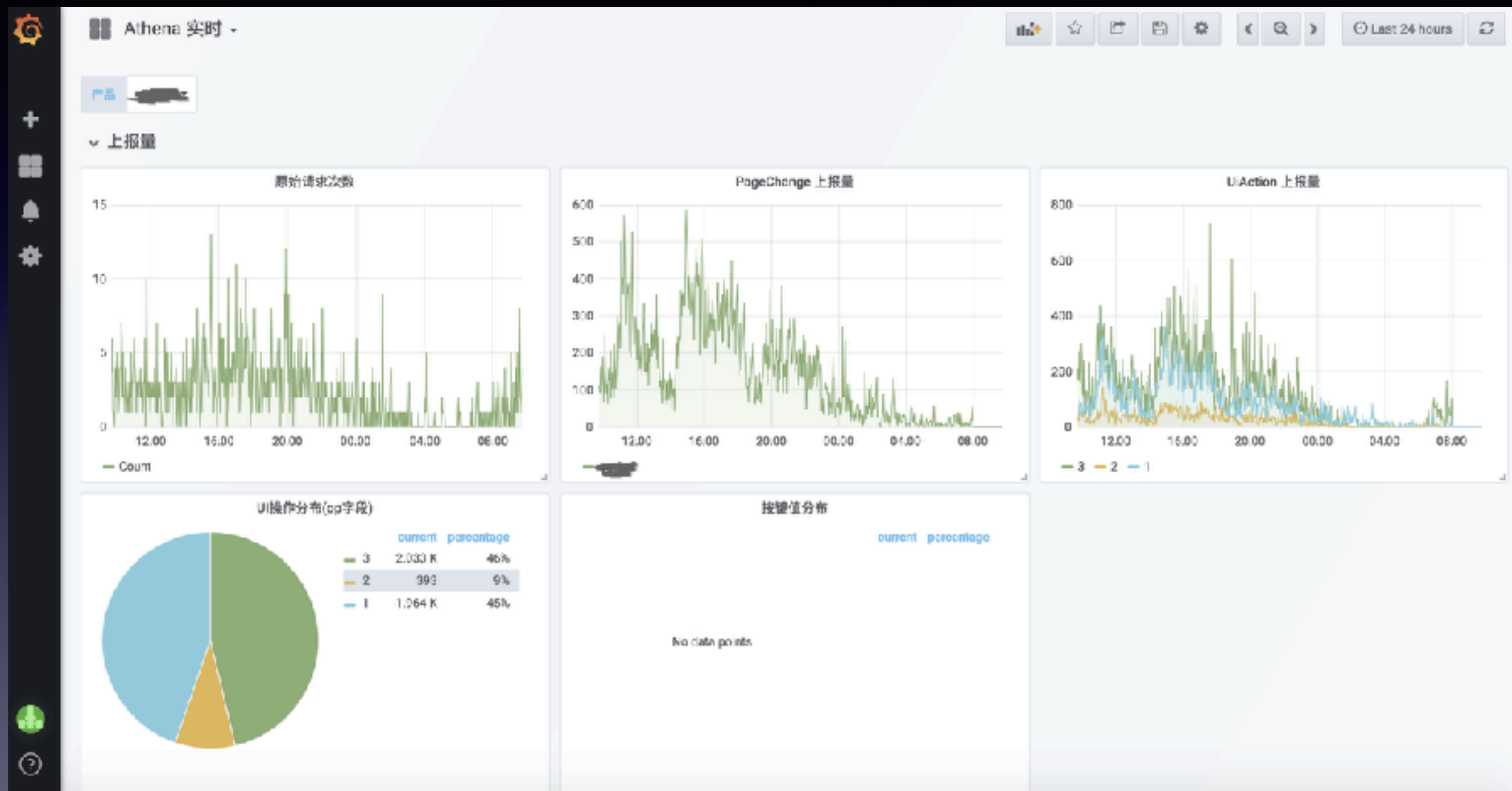
- 原始数据统一推送至Kafka
- Kafka中的数据同步至EMR中的HDFS, 同步表结构到Hive
- 编写Spark计算任务, 从Hive中加载数据, 计算结果写回Hive
- Sqoop导出Hive结果到PostgreSql
- Kafka中的数据同步到Elasticsearch, 提供实时查询能力

实现过程3:可视化



- 提供对外REST API, 从Elasticsearch和PostgreSql中查询数据
- Grafana进行数据展示与报警
- Athena Web: 自定义的Web展示组件

实时数据展示



ES实践

业务需求

- 按天建索引，保留一个月数据
- 数据量TB级别
- 期望廉价存储，节省成本
- 期望弹性扩缩

传统运维

- 在服务器中执行定时自动清理过期索引的脚本
- 在服务器中执行定期snapshot的脚本
- 手动划分冷热节点、定期转存
- 提前规划集群规模，进行扩容

函数计算+Elasticsearch

无服务器云函数SCF

- 无需关心计算资源
- 编写简单、目的单一的函数
- 事件触发代码执行
- 弹性伸缩
- 按函数执行次数付费

实现功能

- 索引生命周期管理
- 数据备份
- 数据冷热分离
- 数据同步

SCF+Curator实现ES索引生命周期管理

1.根据curator模板创建函数

创建方式

空白函数
使用helloworld示例创建空白函数

模板函数
使用示例代码模板创建函数

函数名称

test_curator

最长60个字符，字母开头，支持 a-z, A-Z, 0-9, -, _，且需要以数字或字母结尾

模板搜索

关键字: curator

请选择您要进行搜索的标签，多个过滤标签用回车键: 

Python2.7-ES_curator_clear
此示例通过使用 Elasticsearch curator 库，实现了 ES 索引的删除清理。可为其配置定时触发器后，定时执行并清理 ES 内索引。

开发语言: Python2.7
标签: elasticsearch, curator, python2.7, timer

共1项

下一步

SCF+Curator实现ES索引生命周期管理

2.编辑函数参数

函数名

最长60个字符，字母开头，支持a-z, A-Z, 0-9

运行环境 Python2.7

执行方法 Index.main_handler

代码详情

```
1 # -*- coding: utf8 -*-
2 from datetime import datetime
3 from elasticsearch import Elasticsearch
4 import curator
5
6 esServer = "10.16.16.137:9200" # 修改为 es server 地址+端口
7 esPrefix = "logstash-" # 查找的 index 索引前缀
8 esCuratorTimeStr = "%Y-%m-%d" # 索引中的时间格式
9 esCuratorTimeUnit = "days" # 过滤清理的时间单位
10 esCuratorTimeCount = 8 # 时间间隔
11
12 # 如上示例可以清理 索引格式类似为 logstash-2018-10-11 , 当前时间的8个天前的索引
13
14 ESServer = Elasticsearch(esServer)
15
16 def clean_index():
17     ilo = curator.IndexList(ESServer)
18     ilo.filter_by_regex(kind='prefix', value=esPrefix)
19     ilo.filter_by_age(source='name', direction='older', timestring=esCuratorTimeStr, unit=esCuratorTimeUnit, count=esCuratorTimeCount)
20     try:
21         ilo.empty_list_check()
22     except Exception, e:
23         pass
```

上一步

完成

SCF+Curator实现ES索引生命周期管理

3.配置定时触发

添加触发方式

触发方式  定时触发

定时任务名称  delete_cron

最长60个字符，字母开头，支持 a-z, A-Z, 0-9, -, _，且需以数字或字母结尾

触发周期 每天

立即启用 ☒

保存取消

更多云函数...

数据备份

- 实现ES插件: repository-cos
- 定期进行snapshot, 备份ES数据到对象存储COS中

更多云函数...

冷热分离

- 划分冷热节点，定期迁移旧索引到冷节点，实现数据冷热分离

更多云函数...

数据同步

- 从消息队列(CMQ, CKafka)中同步数据, 事件触发
- 从对象存储COS中同步数据, 事件触发

添加触发方式

触发方式 ?

请选择...

- ✓ 定时触发
- COS触发
- CMQ主题订阅触发
- Ckafka触发
- API网关触发器

定时任务名称 i

0-9, -, _, 且需要以数字或字母结尾

触发周期

每5分钟

立即启用

☒

Thanks!



专业、垂直、纯粹的 Elastic 开源技术交流社区
<https://elasticsearch.cn/>