



GOPS2018
Shenzhen

GOPS

全球运维大会 2018

2018.4.13-4.14

中国·广东·深圳·南山区 圣淘沙大酒店（翡翠店）





GOPS2018
Shenzhen

亿级用户百TB级数据的AIOps实践之路 (一)

周荣 华为消费者BG云运维部



GOPS2018
Shenzhen

目录

- ➔ **1** 华为消费者业务介绍
- 2** 云服务运维面临的挑战
- 3** AIOps 实践之路：数据价值（业务监控）
- 4** AIOps 实践之路：数据平台
- 5** AIOps 实践之路：数据智能



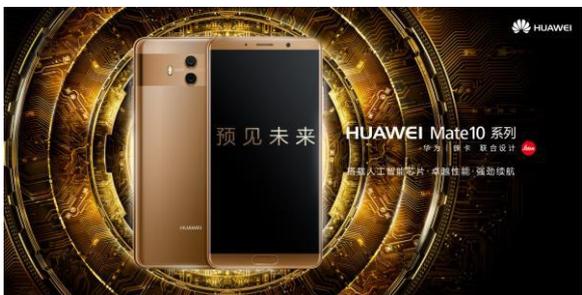
GOPS2018
Shenzhen

华为消费者BG业务介绍



消费者 BG 业务持续增长

- 17年，消费者BG（CBG）收入较16年增幅57%+，占华为总收入43%+；
- 17年，华为与荣耀品牌手机发货 1.53亿台，全球份额突破10%；



全球 TOP100 品牌

★ BRANDZ	#49	★ Interbrand	#70
★ Forbes	#88	★ Brand Finance	#25

华为&荣耀手机



笔记本&平板



穿戴设备



智能家居



软件应用





GOPS2018
Shenzhen

华为消费者BG云服务业务介绍

华为消费者云服务业务（17年，较16年）

- ★ 华为帐号，注册用户 3.3亿，增长 57%
- ★ 云存储，照片&视频，195 PB，增长 108%
- ★ PUSH在线，2.5亿并发，增长 221%
- ★ 应用市场，下载 5.2亿/日，增长 61%
- ★ 华为主题，用户数2.7亿，增长 99%
- ★ 运动健康，用户数4600万，增长 113%





GOPS2018
Shenzhen

目录

1 华为消费者业务介绍

➔ 2 云服务运维面临的挑战

3 AIOps 实践之路：数据价值（业务监控）

4 AIOps 实践之路：数据平台

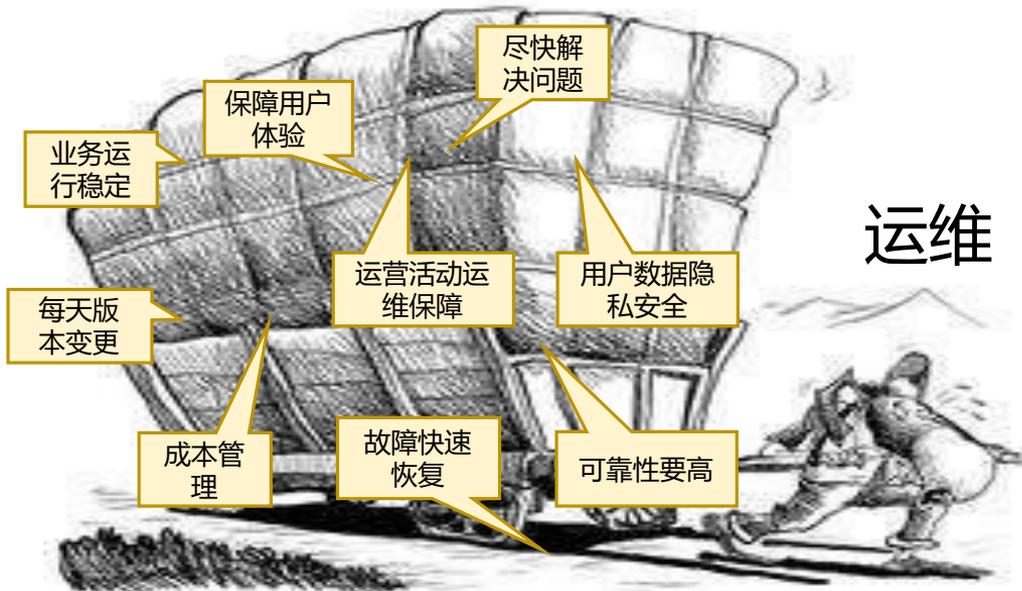
5 AIOps 实践之路：数据智能



GOPS2018
Shenzhen

快速发展下给运维带来的挑战

产品 客户体验



1. 业务多、形式多样

- 存在**100+存量业务**（外部、内部）；
- 业务场景不同，存量**大**，**数据形式和格式多种多样**；

2. 数据价值与数据成本间的平衡

- **监控指标“熵”减严重**，问题难以察觉；
- 采集端统计指标，**数据变更成本高**；

3. 数据规模大，增长速度快

- **数万+主机**、**千亿/天**数据量
- 17年**实时处理**数据从4T/天增到**120TB/天**；
- 随着数据增加，**人找数据的成本**越来越大；

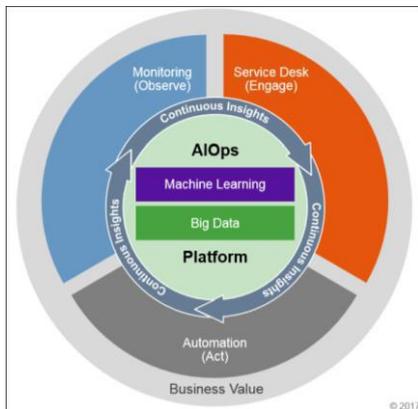


GOPS2018
Shenzhen

业界的趋势 —— 数据驱动运维 (AIOps)

AIOps : Gartner在16年发布的报告中首先提出了基于大数据及算法 (Algorithmic IT Operations) 的 IT运维概念。随着人工智能的快速兴起, Gartner 在17年将 AIOps 概念从基于大数据及算法, 扩充为基于人工智能 (Artificial Intelligence for IT Operations , AIOps), 认为通过大数据、机器学习及高级分析技术, 提供具备主动性、人性化及动态可视化的能力, 直接或间接地提升目前传统IT运维 (监控、自动化、服务台) 的能力。

Figure 1. Hype Cycle for Artificial Intelligence, 2017



知识
↑
信息
↑
数据



GOPS2018
Shenzhen

我们的对策 —— 数据驱动运维（三板斧）

1. 业务多、形式多样

- 存在**大量存量业务**；
- 业务场景不同，**数据形式和格式多种多样**；

2. 数据价值与成本不成正比

- **监控数据“熵”减严重**，问题难以察觉；
- 采集端统计指标，**数据变更成本高**；

3. 数据规模大，增长速度快

- **数万+**主机、**千亿/天**数据量
- 17年**实时处理**数据从4T/天增到**120TB/天**；
- **人找数据的成本**随着数据的增加，越来越

大；

3. 随着运维数据的应用场景，**分步**
构建**智能化运维服务能力（学件）**

2. 围绕数据价值，**分步**打造高性
能、低开发成本的**运维数据平台**

1. 降低数据接入成本，做好**业务监**
控，体现业务数据价值

知识



信息



数据



GOPS2018
Shenzhen

目录

1 华为消费者业务介绍

2 云服务运维面临的挑战



3 AIOps 实践之路：数据价值（业务监控）

4 AIOps 实践之路：数据平台

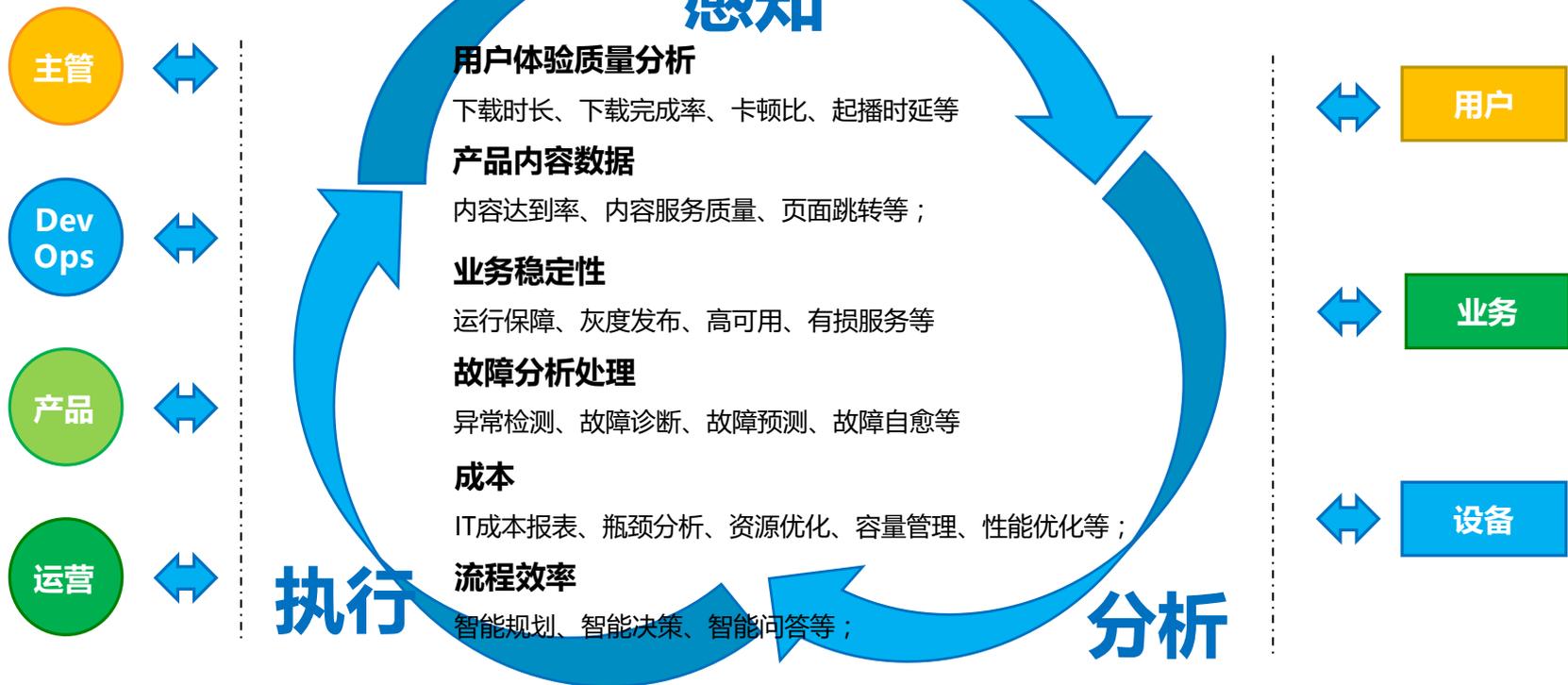
5 AIOps 实践之路：数据智能



GOPS2018
Shenzhen

AI Ops 实践之路：从哪来，到哪去

1. 数据到价值





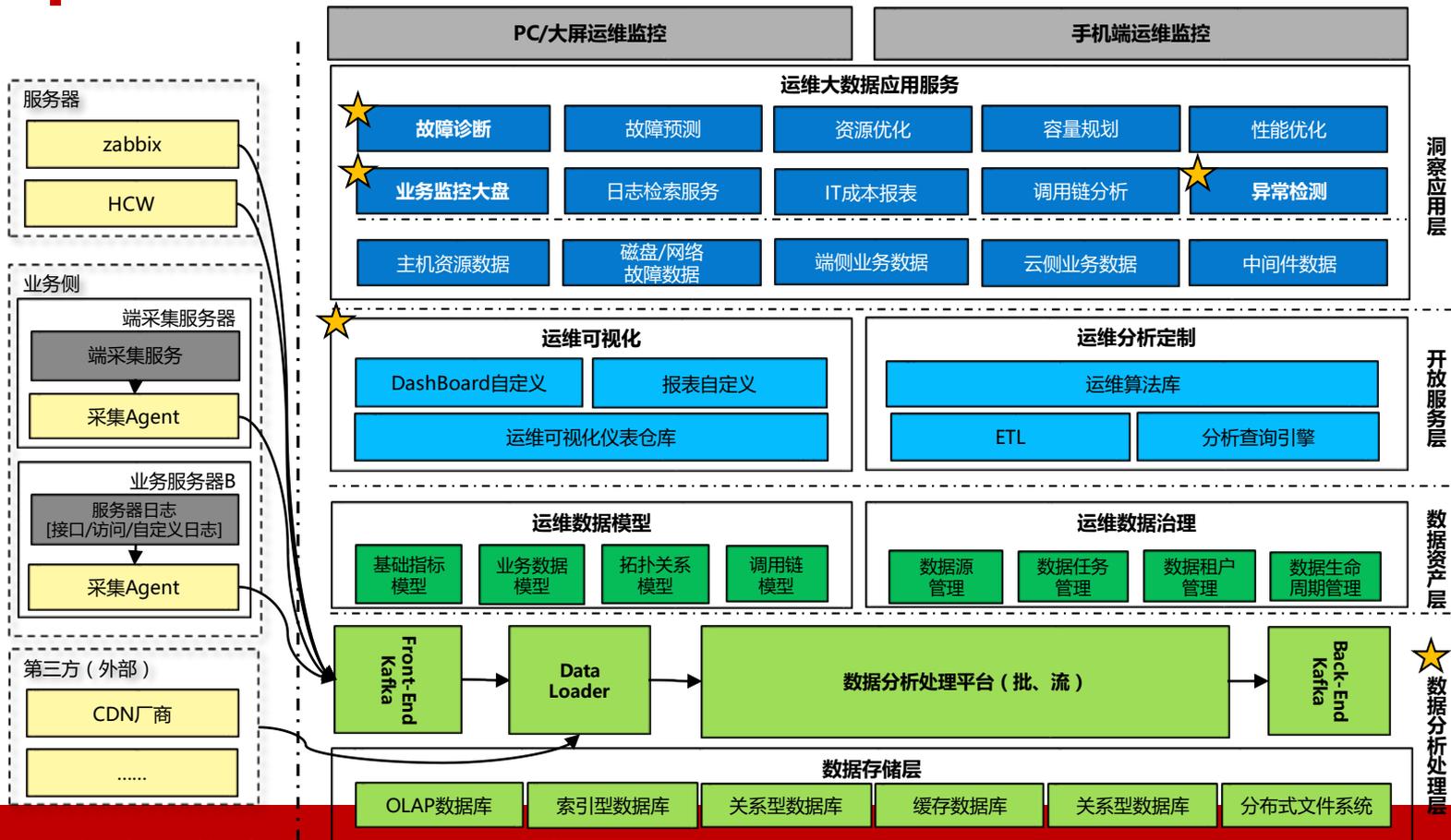
GOPS2018
Shenzhen

AIOps 实践之路：核心价值与演进

	数据感知		智能分析	智能执行	
	v1.0	v2.0	v3.0	v4.0	v5.0
质量保障	业务指标监控	异常检测	故障诊断	故障预测	故障自愈
成本管理	成本指标监控	异常检测	资源优化	容量规划	性能优化
流程效率	流程数据可视	异常检测	智能问答	智能决策	智能变更



AIOps 实践之路：架构总览



洞察应用层

开放服务层

数据资产层

★ 数据分析处理层



GOPS2018
Shenzhen

AIOps 实践之路：业务监控



业务系统的运行情况了如指掌



GOPS2018
Shenzhen

目录

1 华为消费者业务介绍

2 云服务运维面临的挑战

3 AIOps 实践之路：数据价值（业务监控）

➔ 4 AIOps 实践之路：数据平台

5 AIOps 实践之路：数据智能



GOPS2018
Shenzhen

AIOps 实践之路：数据平台 - 对策

1. 存量业务大，业务场景多

- 低成本、灵活数据接入



1. 数据后端清洗、入库

- SparkStreaming (Kettle ? Talend ?)

2. 亿级用户，数据量大

- 1年，4TB->120TB



2. 支撑海量数据处理框架

- Kafka、Spark、Hive、MPP DB ;

3. 原始数据信息查询要求

- 多维分析，即席查询，实时

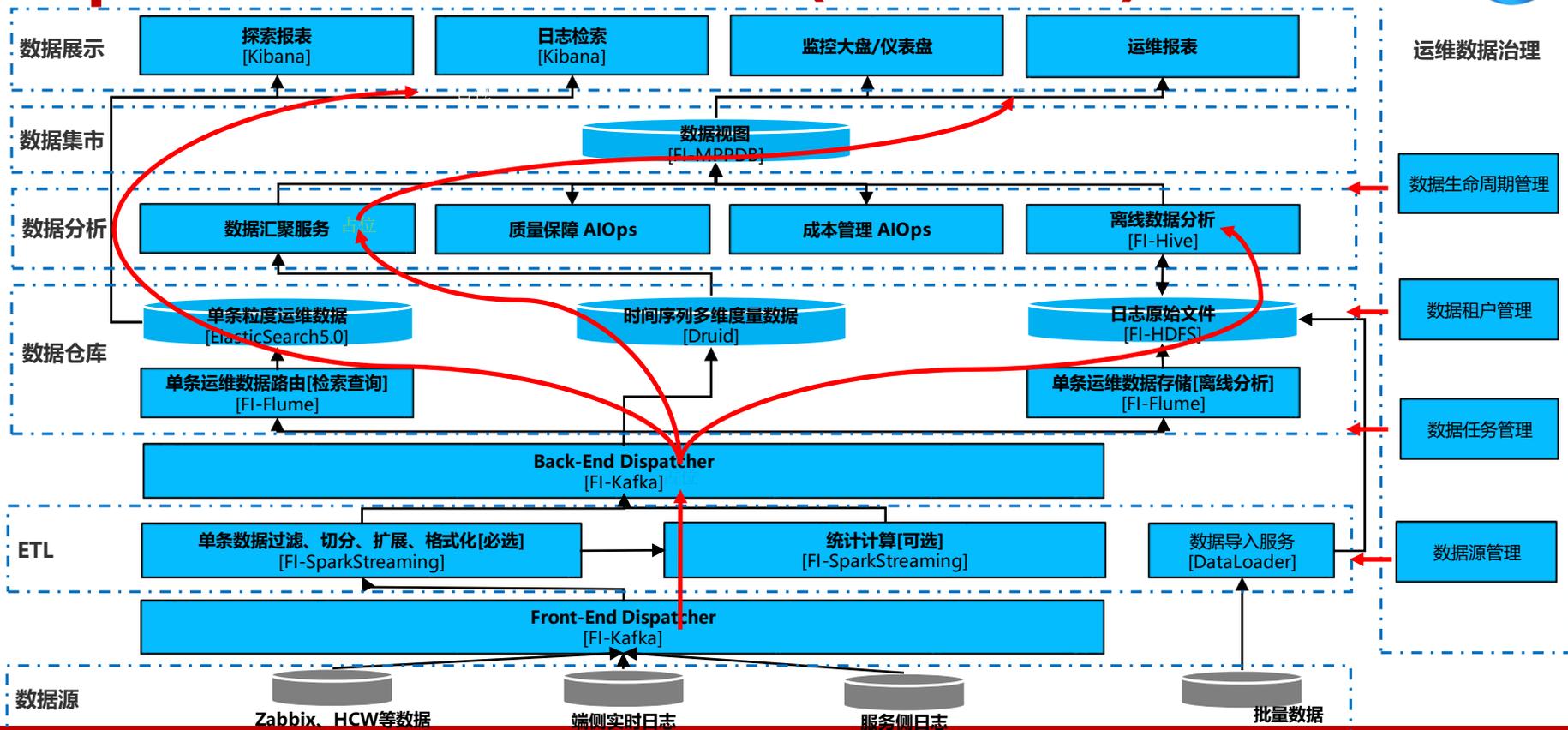


3. 实时可视化、OLAP数据引擎

- Vue (Angular)、Druid、Prometheus



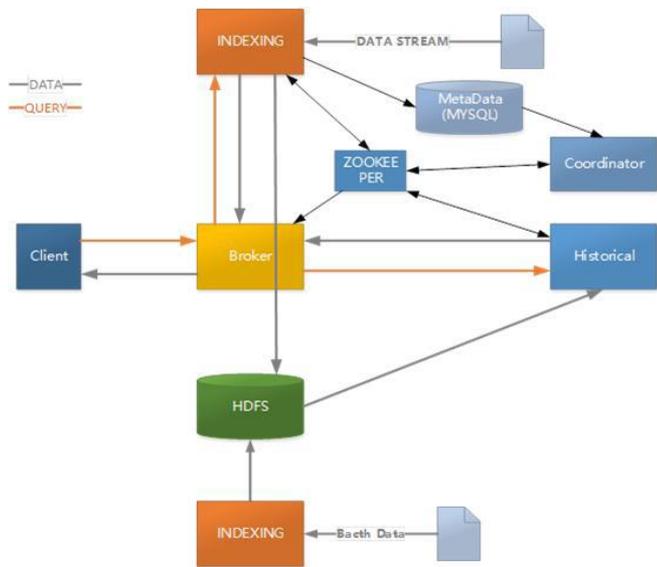
AIOps 实践之路：数据平台（数据分析）





AI Ops 实践之路：数据平台 - OLAP数据引擎

- Druid 开源实时大数据分析引擎：面向列存储、shared-nothing架构、高效索引结构，来达成在秒级对十亿行级别的表进行多维探索分析查询。



部分来源：<http://druid.io/druid.html>

Timestamp	Dimensions				Metrics	
Timestamp	Page	Username	Gender	City	Characters Added	Characters Removed
2011-01-01T01:00:00Z	Justin Bieber	Boxer	Male	San Francisco	1800	25
2011-01-01T01:00:00Z	Justin Bieber	Reach	Male	Waterloo	2912	42
2011-01-01T02:00:00Z	Ke\$ha	Helz	Male	Calgary	1953	17
2011-01-01T02:00:00Z	Ke\$ha	Xeno	Male	Taiyuan	3194	170

```
timestamp      page      language  city      country  ...  added  deleted
2011-01-01T00:01:35Z  Justin Bieber  en      SF      USA      ...  10     65
2011-01-01T00:01:63Z  Justin Bieber  en      SF      USA      ...  15     62
2011-01-01T01:02:51Z  Justin Bieber  en      SF      USA      ...  32     45
2011-01-01T01:01:11Z  Ke$ha        en      Calgary CA      ...  17     87
2011-01-01T01:02:24Z  Ke$ha        en      Calgary CA      ...  43     99
2011-01-01T02:03:12Z  Ke$ha        en      Calgary CA      ...  12     53
...
```



```
timestamp      page      language  city      country  ...  added  deleted
2011-01-01T00:00:00Z  Justin Bieber  en      SF      USA      ...  25     127
2011-01-01T01:00:00Z  Justin Bieber  en      SF      USA      ...  32     45
2011-01-01T01:00:00Z  Ke$ha        en      Calgary CA      ...  60     186
2011-01-01T02:00:00Z  Ke$ha        en      Calgary CA      ...  12     53
...
```



GOPS2018
Shenzhen

目录

- 1 华为消费者业务介绍
- 2 云服务运维面临的挑战
- 3 AIOps 实践之路：数据价值（业务监控）
- 4 AIOps 实践之路：数据平台
- ➔ 5 AIOps 实践之路：数据智能



AI Ops 实践之路：数据智能（业务背景 - 质量）

云服务当前的**主机规模**，**业务数**，**数据量**，**复杂度**都达到了新的高度，随之给业务团队带来了显著挑战：传统以人工为主的运维手段，面对**异常检测**、**故障定界定位**、**故障预测**以及**故障自愈**的**质量保障**方面，不仅力不从心，而且难以做好。

业务规模增长带来运维数据的爆发式增长



说明：截止17年底

运维数据的爆发式增长背景下，传统人工运维方式的痛点与问题



别

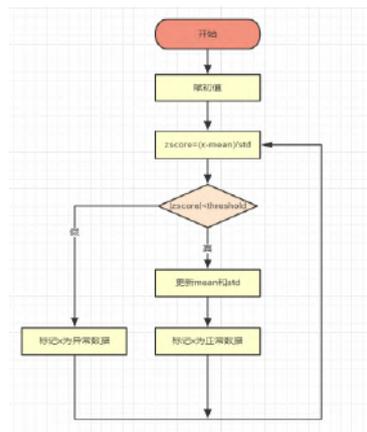


异常检测 — 数据源干扰 (算法v0.8)

z-score 算法

算法原理： $zscore = (x - mean) / std$

算法流程：



算法表现：



优点：流式计算效率很高，异常发生时能突显异常。

缺点：异常数据会改变阈值，本质是因为zscore计算用到的mean和std对异常值敏感，即使使用的抑制方法，治标不治本，具体表现为对不同序列，都需要人工调参，增加了推广成本。

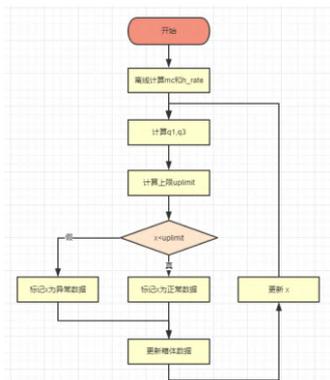
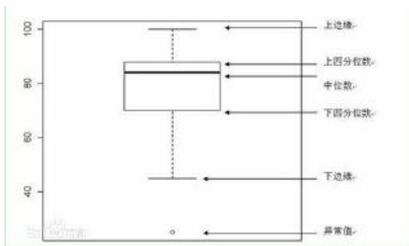


异常检测 - 数据源干扰 (算法v1.0)

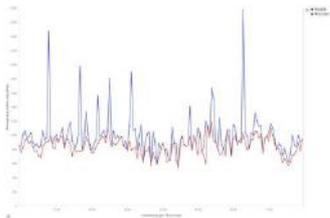
Boxplot法

算法原理： 考虑到百分位数对异常不敏感，即新来一部分异常数据，百分位数不会明显改变，使用箱线图来去除数据源异常；

算法流程：



算法表现：



优点：无需人工调参，对异常数据容忍性高。

缺点：该算法需要保存序列中最近的N个数据，相比z-score计算量大。

此外，使用多种指标推广发现，该算法得出的下限偏高，因此推荐在上限场景中使用。



GOPS2018
Shenzhen

异常检测 — 指标异常检测（算法1）

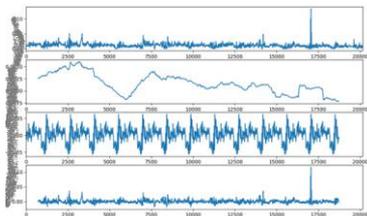
序列分解算法

算法原理： 时间序列分解为季节项（周期项）、趋势项、残差项。使用历史数据分解的趋势项和季节项之和作为预测值的均值，使用残差项计算标准差，得到预测的波动范围，进行异常检测。

算法流程：



算法表现：



优点： 离线计算，效率很高，无需较多调参，适用范围广，推广成本低。

缺点： 存在数值波动时，算法生成的阈值不能适应，容易误告警。对于在时间维度周期性不明显的指标，效果不佳。



GOPS2018
Shenzhen

异常检测 — 指标异常检测（算法2）

多工况检测算法

算法原理： 算法的核心是如何合理的划分工况；采用聚类的方式，基于历史数据特征的贡献度进行聚类，不同的类便是不同的工况；比如：某个时间段的请求量范围下是一个具有相同业务特征的工况。

算法流程：



算法表现：



优点：有效消除告警虚报；

缺点：检测结果受训练数据影响，在业务变化剧烈时检测结果会受到影响。



智能运维，我们在路上





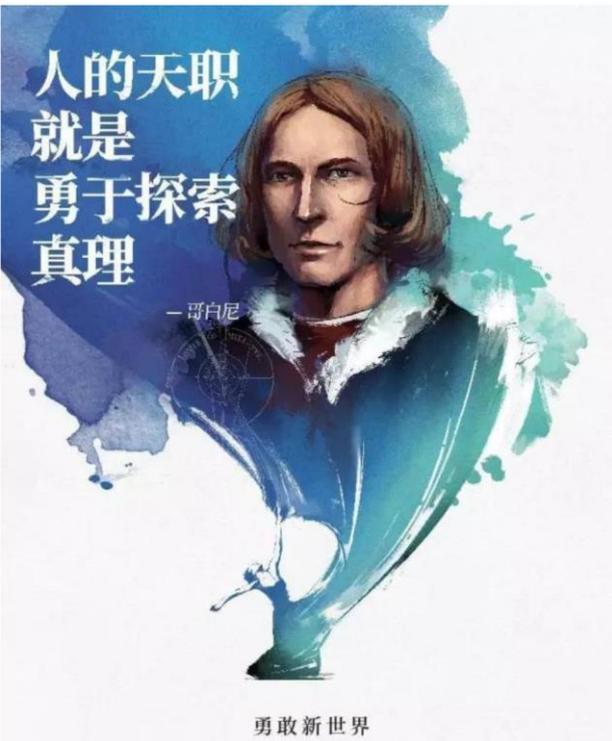
GOPS2018
Shenzhen

Join us, Make a difference!

运维工程师：网络运维、CDN运维、安全运维、业务领域

运维开发工程师：自动化工具开发、大数据平台开发

运维AI工程师：数据建模、数据算法、机器学习、智能化工程落地





GOPS2018
Shenzhen

Q&A





GOPS2018
Shenzhen



Thanks

高效运维社区
开放运维联盟

荣誉出品



GOPS2018
Shenzhen

想第一时间看到高效运维社区
的新动态吗？

