



十年架构 成长之路

# SACC 第十届中国系统架构师大会

SYSTEM ARCHITECT CONFERENCE CHINA 2018

2018年10月17-10月21日 北京海淀永泰福朋喜来登酒店



# 阿里巴巴大数据智能运维实践

柯旻(大舞) 阿里巴巴大数据运维总监



**SACC**

第十届中国系统架构师大会  
SYSTEM ARCHITECT CONFERENCE CHINA 2018



# 个人介绍



柯旻（大舞） 阿里巴巴大数据运维总监

2007-2010 (中国yahoo) 负责中国yahoo基础运维相关工作  
(IDC / 网络 / 硬件 / SYS)

2011-至今主要负责整个阿里集团所有离线，流&批计算，分析型数据库，算法平台等大数据平台公有云以及内部平台的运维，运维产品开发，架构以及对外输出。支持10W+服务器的平台日常运营维护工作。

(hadoop/hbase/jstorm/odps/ads/blink/数据通道/GPU...等大数据平台产品)



十年架构 成长之路

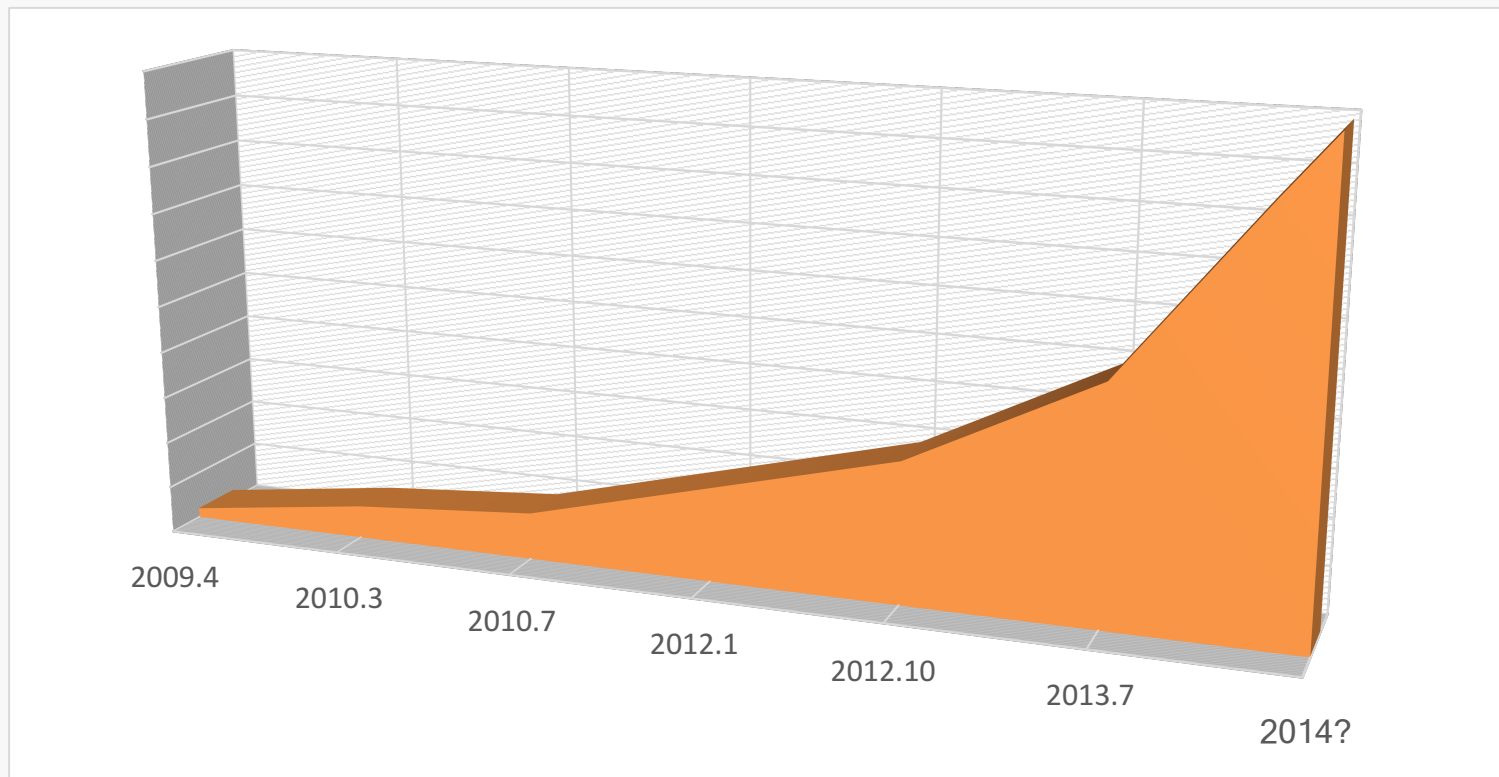


# 阿里巴巴运维演进史

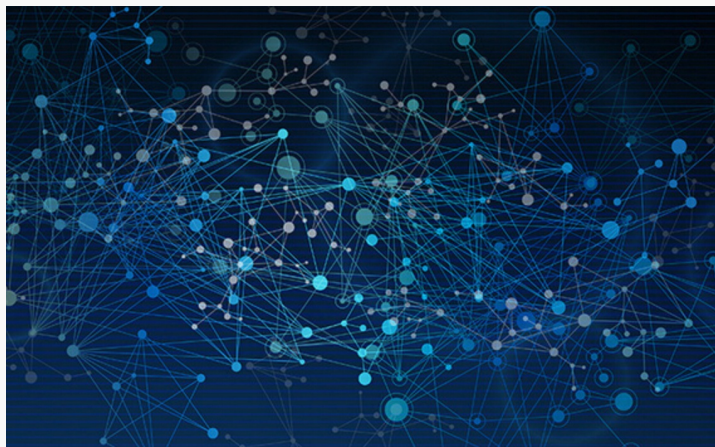
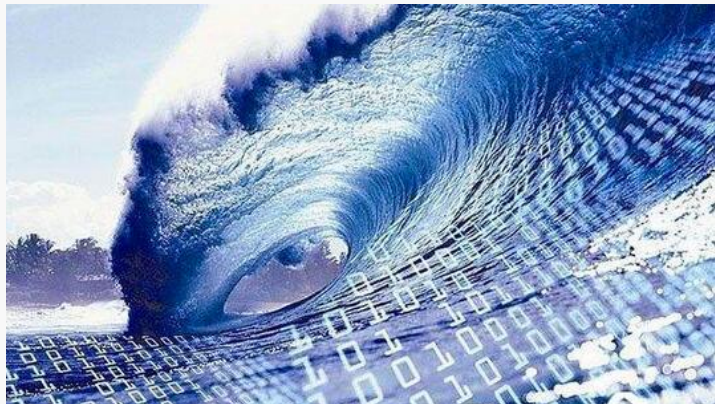




# 规模增长



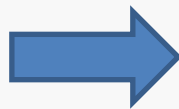
# 多维度海量数据



# 建立运维世界的数字镜像

## 运维数据仓库

Commands/Tools Based Operation -> SQL Based Operation



# 常见的运维数据

## 元数据



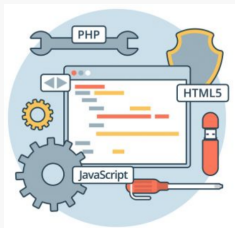
服务器



集群



网络

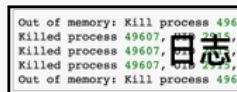
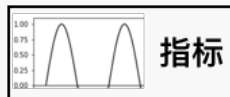


应用



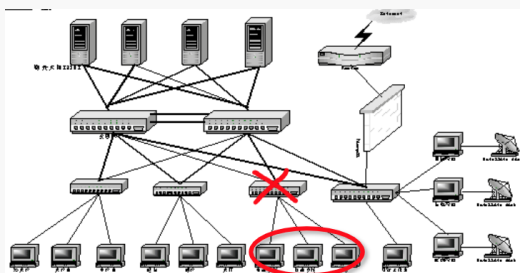
服务

## 运行时



- **元数据**描述了对对象的静态属性
- **运行时**描述了对对象的动态状态
- 两者组合描述了运维世界



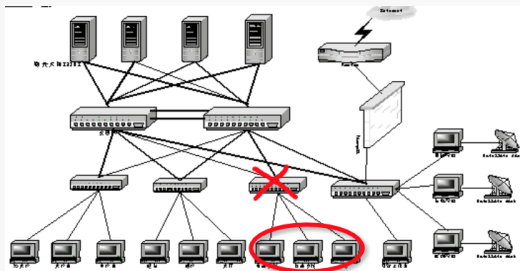


CMD

```
~ pssh -i -h hosts 'get_uplink '
~ cat up_switch | uniq -c
```



1号交换机故障导致丢包



运维世界->DB



SQL

```
SELECT up_switch, avg(tcp_retran) as tcp_retran_avg
FROM server_info GROUP BY up_switch;
```



1号交换机故障导致丢包

# 数据分层设计

解决：

数据加工难

重复劳动

加工链路长，出错概率高

DWS/ADS层（汇总/应用级数据）

DWD层（明细层）

ODS层(事实)

DIM层(维度)

- 运维场景对数据实时性的要求永远是贪婪的
- 需要根据真实的场景和需求，选择合适的时效性

- **ODS（Operational Data Store）**  
如从DB、API、日志等系统抽取过来的数据
- **DIM（Dimension）**  
如机器维度表、监控部署维度表等
- **DWD(Data Warehouse Detail)**  
基于ODS层，按照业务领域进行整合的明细数据层
- **DWS(Data Warehouse Summary)**  
基于DWD的统计层，一般根据DWD的某个维度来统计
- **ADS（Application Data Service）**  
基于DWS、DWD的应用展示层；直接输出给应用使用

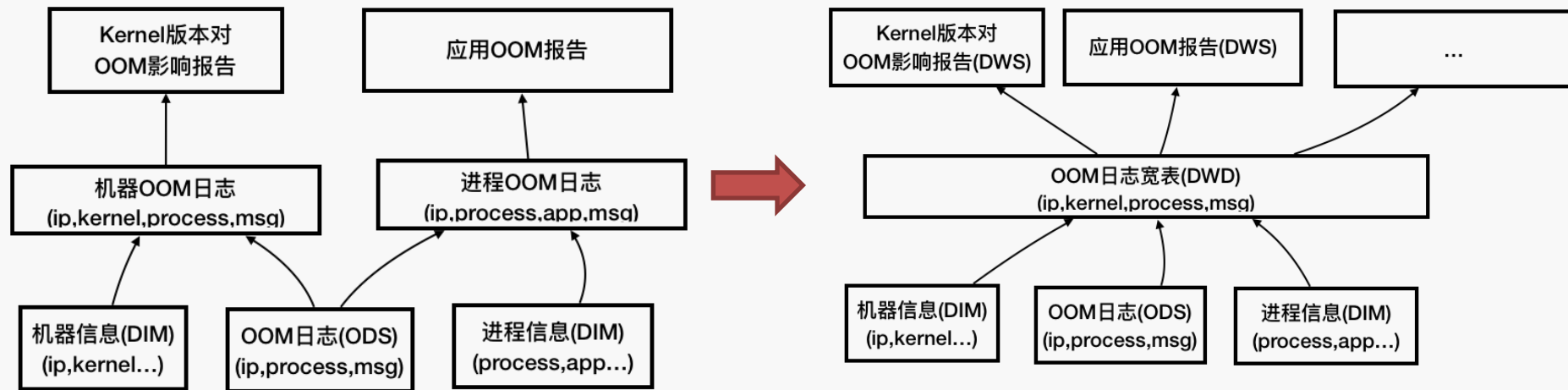
书同文（语言）



车同轨（方法）



行同伦（模式）





# 异常检测 (Anomaly Detection)

涵盖从业务层到硬件底层几乎所有垂直领域

各种设备的管控，如空调、服务器、硬盘

各种软件应用的监控

各种业务指标的监控

一切需要监控且有监控数据的对象，  
都可以使用异常检测来进行智能管控

数据

Streaming data

Log data

Data in other formats

广义的异常检测包括：

异常检测

异常预测

根因分析

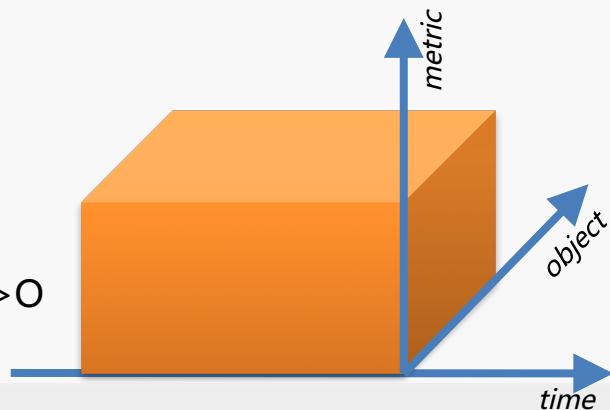


# 异常检测的类型

- Fix 2 dims, and only focus on data in 1 dim
  - T**: only consider time dim, 单一对象单一metric (即单个时间序列): spikes & dips、趋势变化、范围变化
  - M**: only consider metric, 找出不符合metric之间相互关系的数据
  - O**: only consider object, 找出与众不同的对象
- Fix 1 dim, and focus anomalies in 2 dims
  - MT**: 固定对象, 考虑多个时间序列 (每个对应一个metric), 并找出其相互变化方式不同的作为异常
  - MO**: 不考虑时间特性, 考虑多个对象且每个对象都可以用多个metric表示, 如何从中找出不同的对象
  - TO**: 多个对象单一metric, 找出变化趋势不同的对象
- Find out anomalies in the (metric, object, time) 3d space
  - MTO**: 多个对象多个metric在不同时间的数据

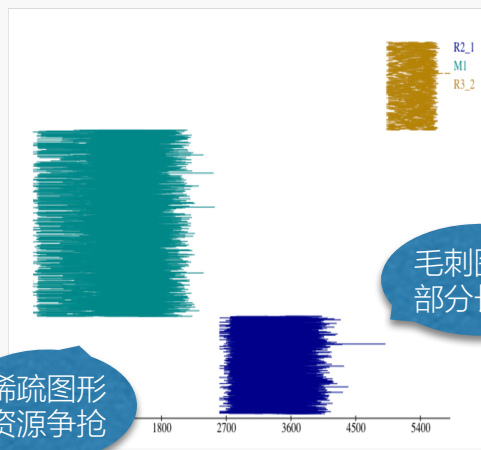
我们重点考虑M, T和MT

三个维度按照优先级排列为M>T>O

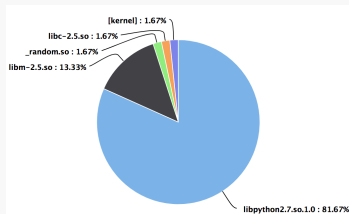


# 数据化运维案例——全链路分析诊断

- 提供自助式全链路的诊断视角
- 通过一个作业的入口，关联延伸到整个系统的上下游
- 资源满足/配置检查/数据依赖/历史对比/长尾倾斜....



任意机器进程  
CPU消耗分析

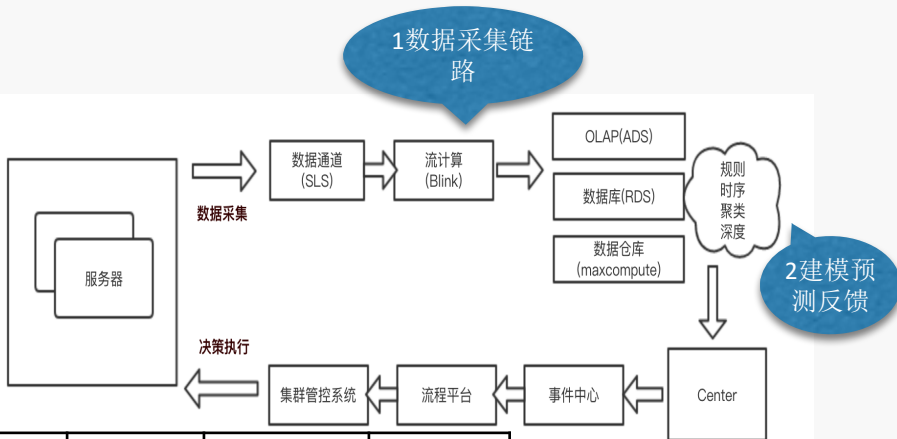


# 数据化运维案例—硬件自愈

- 10w级的服务器上部署硬件相关信息采集插件，打印硬件状态日志
- 经过数据通道、流计算、OLAP系统，建立起了—份实时硬件状态表
- Center程序通过定时运行SQL任务，使用规则判断硬件异常

决策定义

决策	进程影响	进程影响时间	数据影响	数据影响时间	影响范围
重启服务	1	5	*	5	服务组件
重新部署	1	5	*	0	服务组件
重载配置	0	0	*	0	服务组件
内核升级	1	5400	*	0	整机
重启机器	1	1200	*	1200	整机
进入无盘	1	1200	*	1200	整机
重新克隆	1	5400	*	永久	整机
整机维修	1	345600	*	永久	整机
机器下线	1	永久	*	永久	整机
数据盘维修	0	0	disk: sdx	永久	部件
系统盘维修	1	345600	disk: sda	永久	部件



3决策执行链路

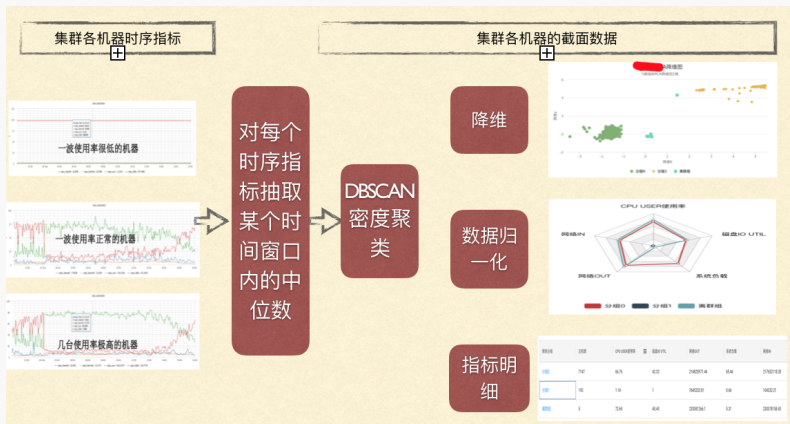
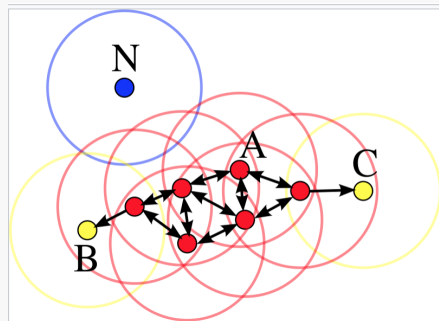
1. 数据采集
  2. 算法分析
  3. 决策执行
- 一年处理20万次自愈事件
  - 服务器可用率 99%

# 数据化运维案例—聚类异常检测

## • 关于算法选择，Why DBScan？

基于密度的聚类算法，能够将具有足够高密度的区域划分为簇，并在具有噪声的数据中发现任意形状的簇

- 无需指定聚类组的数量
- 支持离群点的自动发现

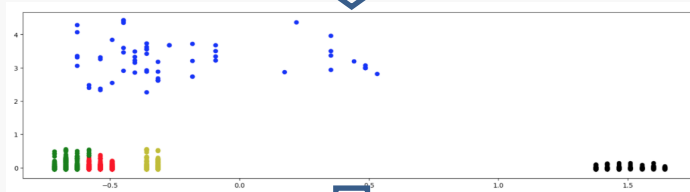
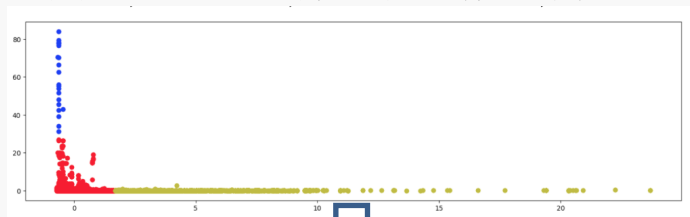


# 数据化运维案例—聚类寻优

数据传输用户任务数量大，参数配置困难如何优化？

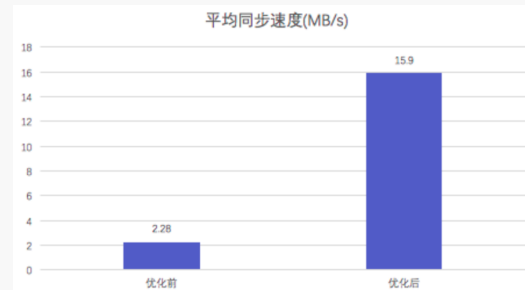
20W  30W

固定属性k-means聚类，找出每类的最佳配置



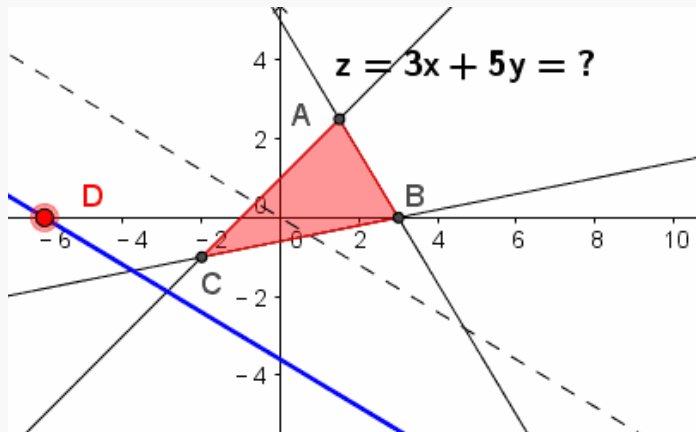
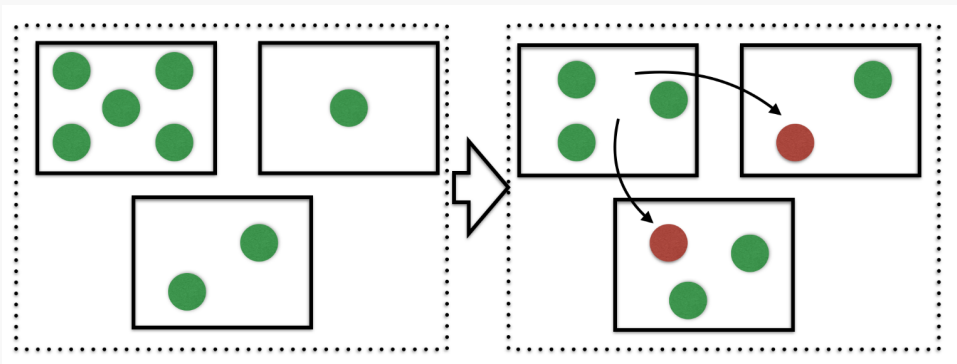
调优表

优化前大部分任务集中在5MB/s的低效同步速度区间内，优化后大都在在5~10Mb/s、10MB/s~50MB/s、50MB~100MB/s的高速区间。  
优化前平均速度2.28MB/s，优化后15.9MB/s，均速直接提升7倍



# 数据化运维案例—运筹优化

- 运维 - “运” -> 运筹
- 运维领域有较多的规划问题可以用运筹优化的方法来解决
  - ✓ 多集群容量均衡
  - ✓ 动态调整用户配额
  - ✓ 如何优化带宽
  - ✓ 任务参数智能优化
  - ✓ .....



线性规划

$$\text{Min } std(\sum_{i=1}^m (traffic_i * X_{ij})) \quad (j=1\dots n)$$

$$\text{s.t. } \sum_{j=1}^n x_{ij} == 1 (i = 1\dots m)$$

$$x_{ij} = 0 \text{ 或 } 1$$



# 运维数据价值提取



数据化一切对象【全域数据】

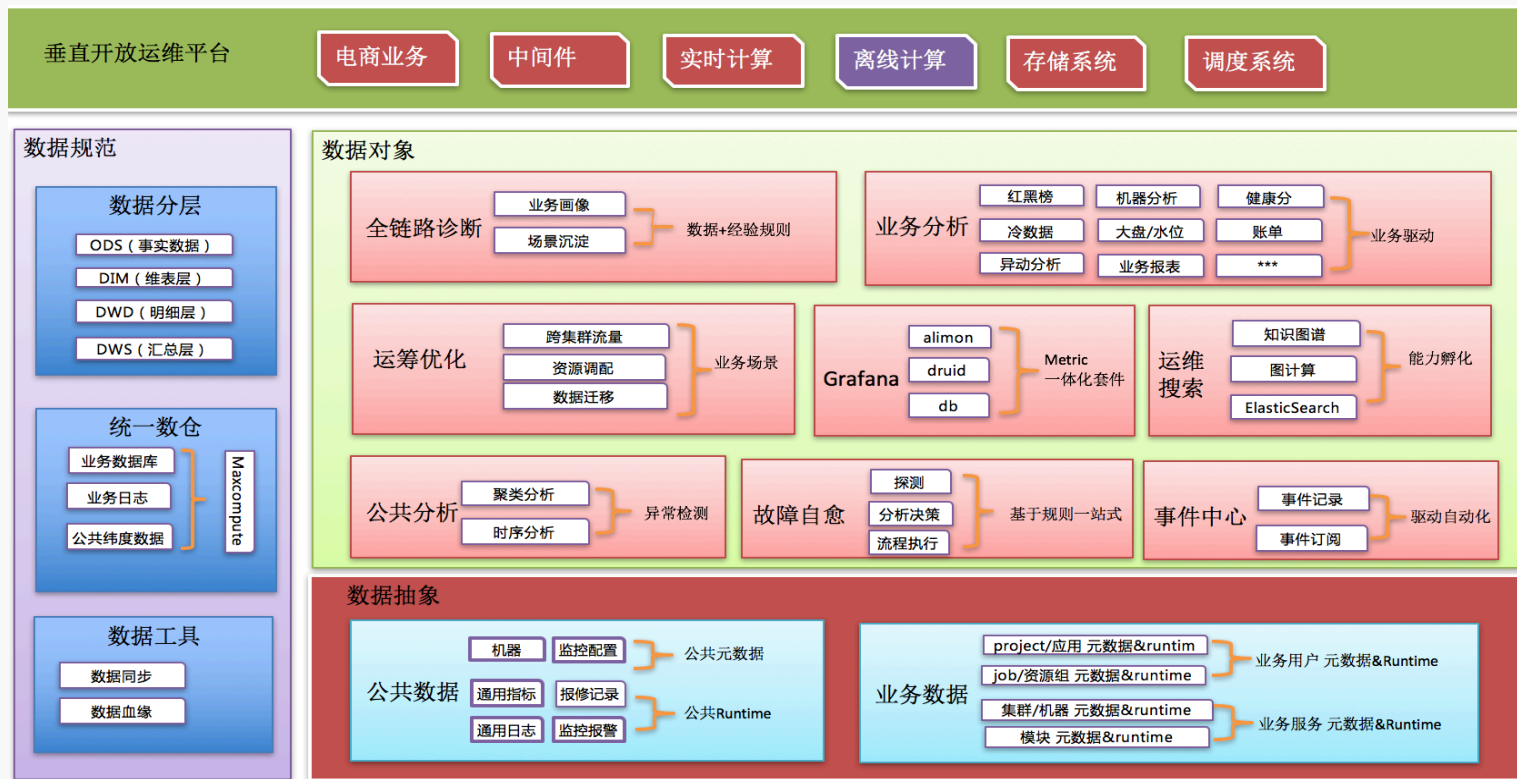
结构化一切数据【知识图谱】

连接一切数据【运维搜索】

让数据说话【数据即视图】

让数据流动【数据驱动业务】

# 数据化运维中台



# DataOps

- 定义：

把运维数据采集起来，深度挖掘运维数据的价值，从系统“**稳定性、成本、效率、安全**”多个维度，为运维提供**数据决策基础和依据**。

- 挑战：

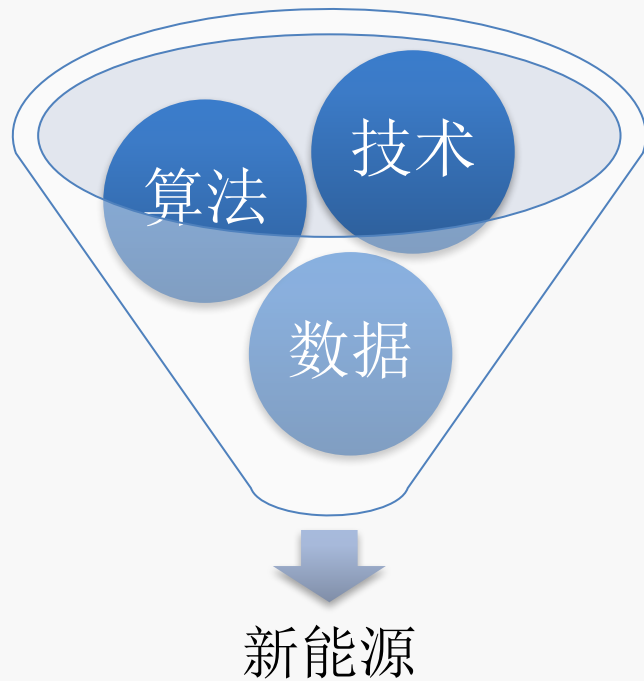
有效地**收集、清洗**

数据实效性、准确性

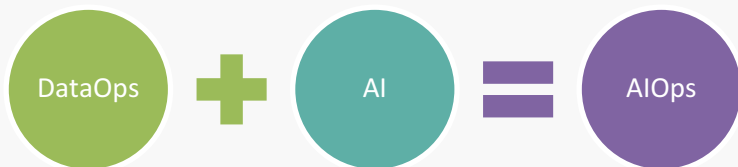
有序、有结构地分类组织和存储**管理**

用算法打通、**连接**，**分析**这些数据

有效、智能地将价值数据**提炼**、展示出来



# 智能化运维



## 智能感知

- 事件关联/事件流图
- 知识图谱
- 智能预测

## 智能决策

- 无监督机器学习
- 深度学习神经网络
- 规则&算法结合

## 智能执行

- ChatOps + NLP
- 无人值守
- 高度自动化流程

## 专业领域 Domain Knowledge

Supply Chain  
Power & Electric  
Monitoring  
Scheduling  
Network Operation & Maintenance

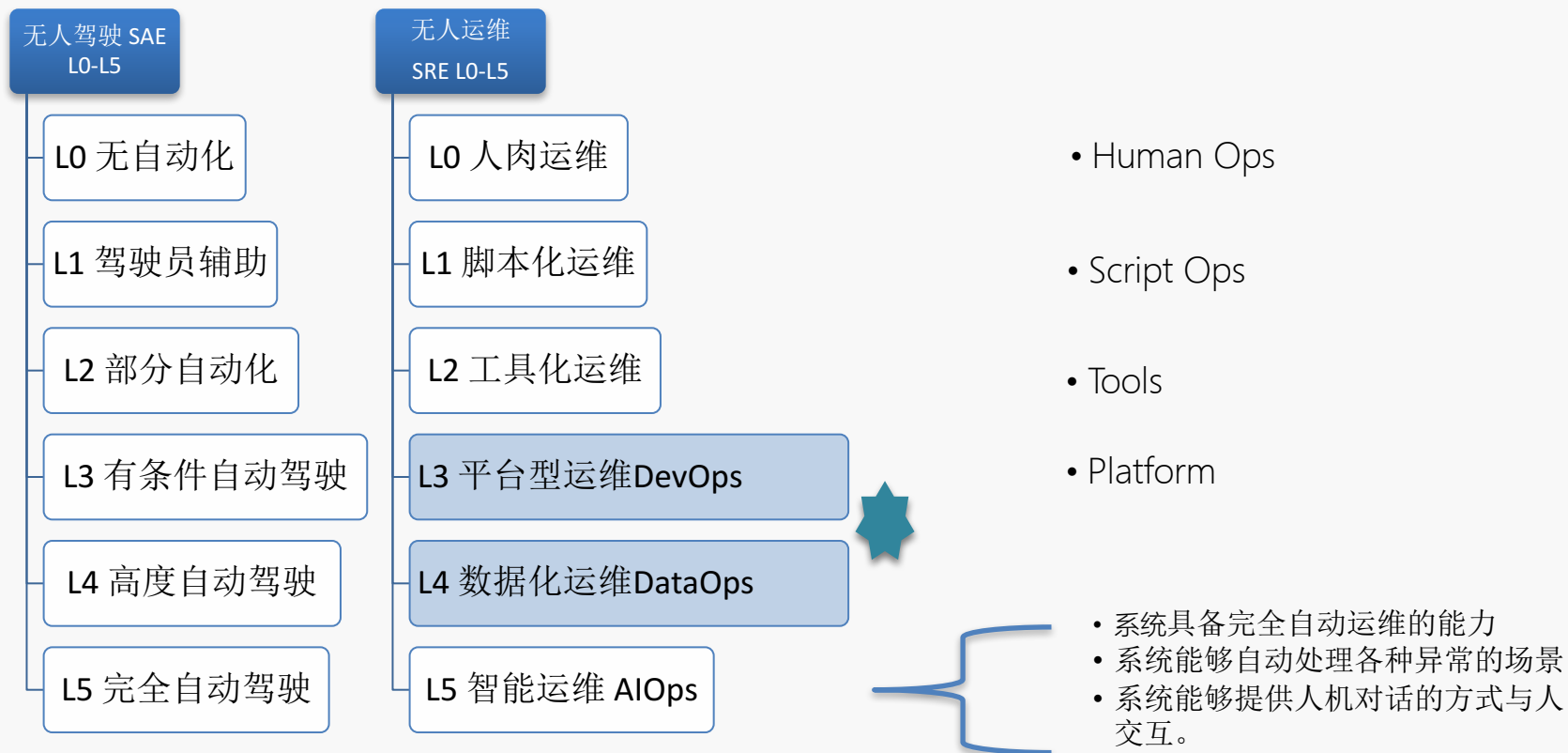
## 机器学习 Machine Learning

Anomaly Detection  
Regression  
Classification  
Association Rule Analysis  
Root Cause Analysis

## 优化算法 Optimization

Dynamic Programming  
Inventory Control  
Linear Programming  
Nonlinear Programming  
Simulation-Monte Carlo  
Analysis

# 自动驾驶 V AIOps运维



# 智能化不等于无人化

## 黑天鹅事件 /

形容小概率、难预测的突发风险



## 灰犀牛事件 /

形容大概率、可预测、波及范围大的风险



# 欢迎加入我们！！！！

- 成熟的阿里大数据体系
- 我们所运维的对象就是阿里大数据
- 足够大的舞台



关注【AliDataOps】  
数据智能运维时代与你同行

- 运维的对象足够复杂
- 数据的量级超乎想象
- 我们需要更高效、更智能



扫一扫上面的二维码图案，加我微信





THANKS