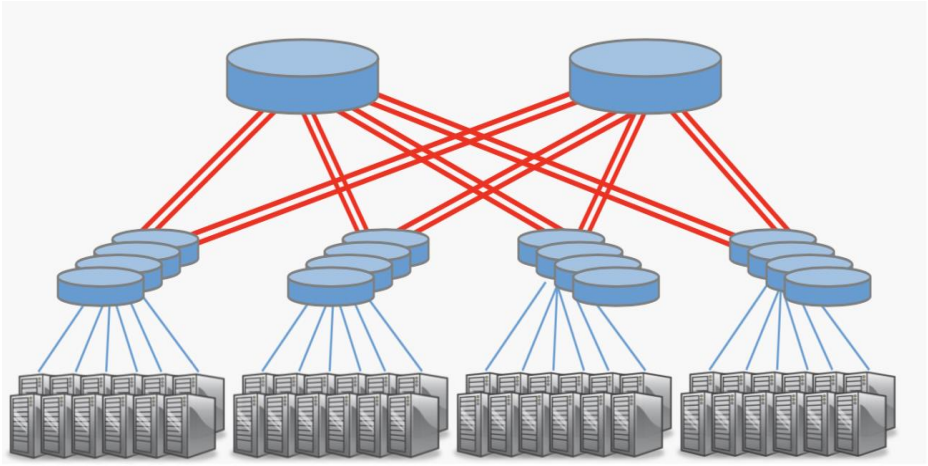
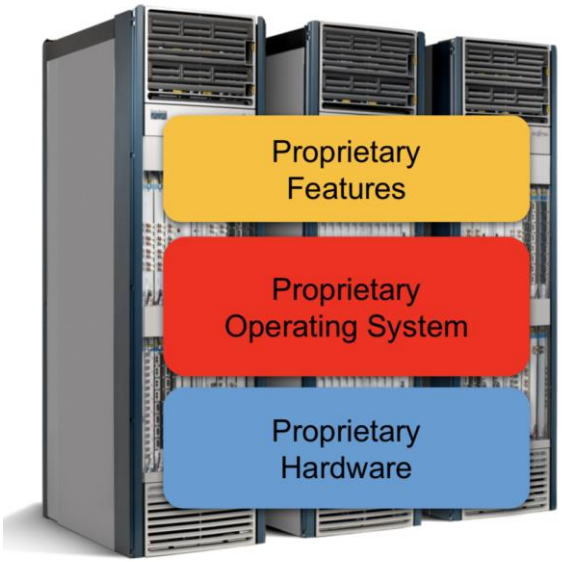
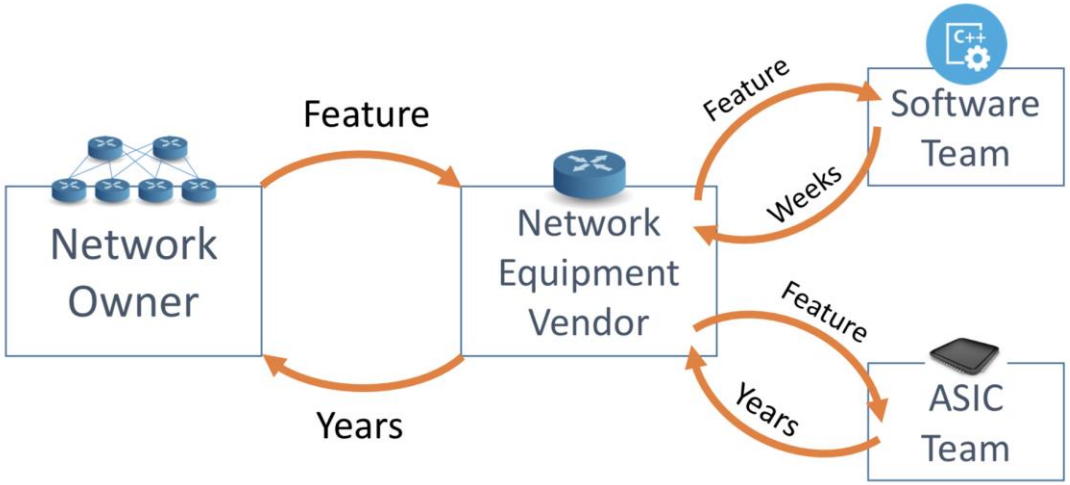
大规模数据中心对网络的需求



服务器：100,000
交换机：5,000

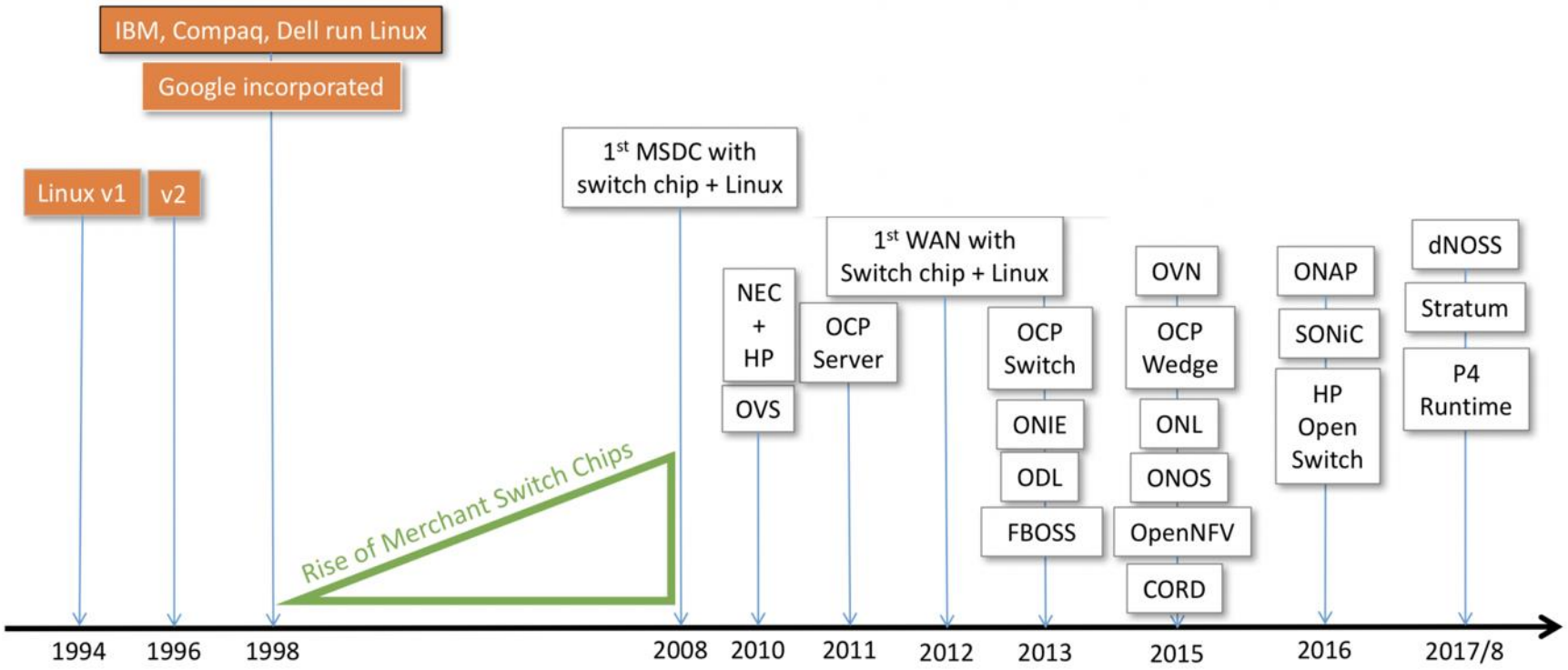
大规模数据中心对网络的挑战



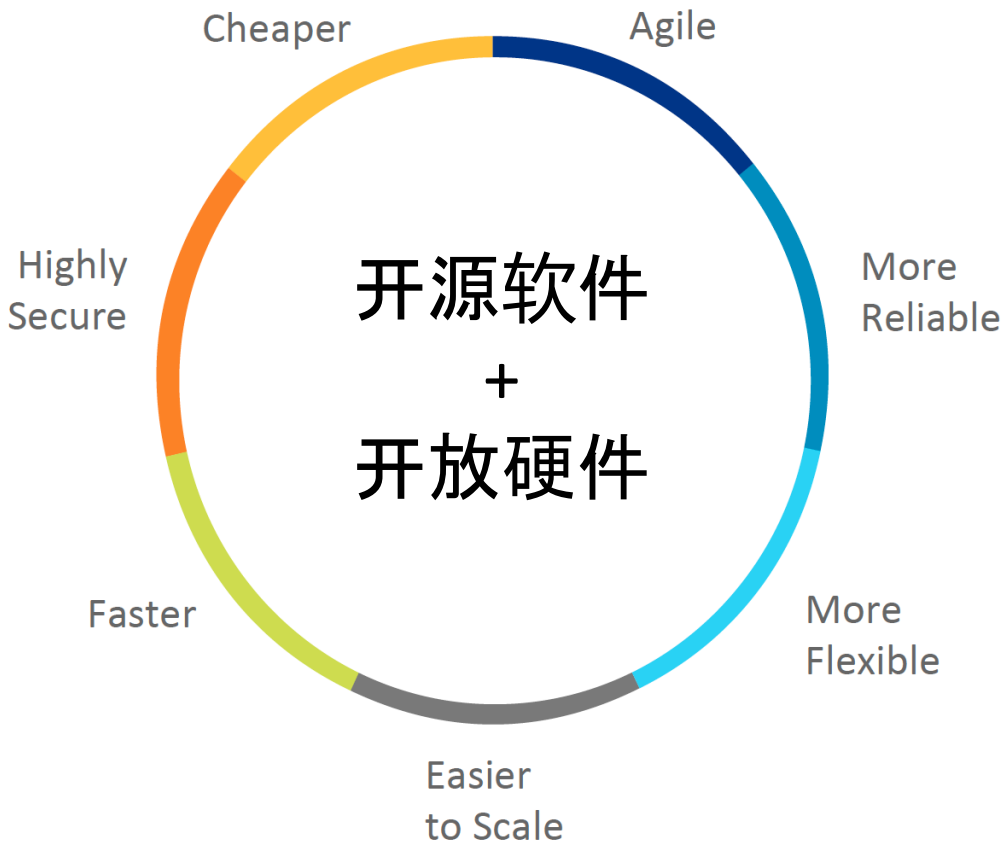


- 厂商设备多厂家、多型号
- 系统封闭、功能复杂
- 问题解决、特性引入节奏慢、周期长

数据中心计算和网络的开放趋势



数据中心软硬件开放带来的收益



数据中心白盒交换机生态

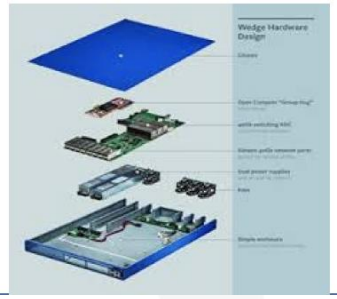
ONIE

OpenNSL

SONiC

OpenSwitch

P4.org



BROADCOM

BAREFOOT NETWORKS

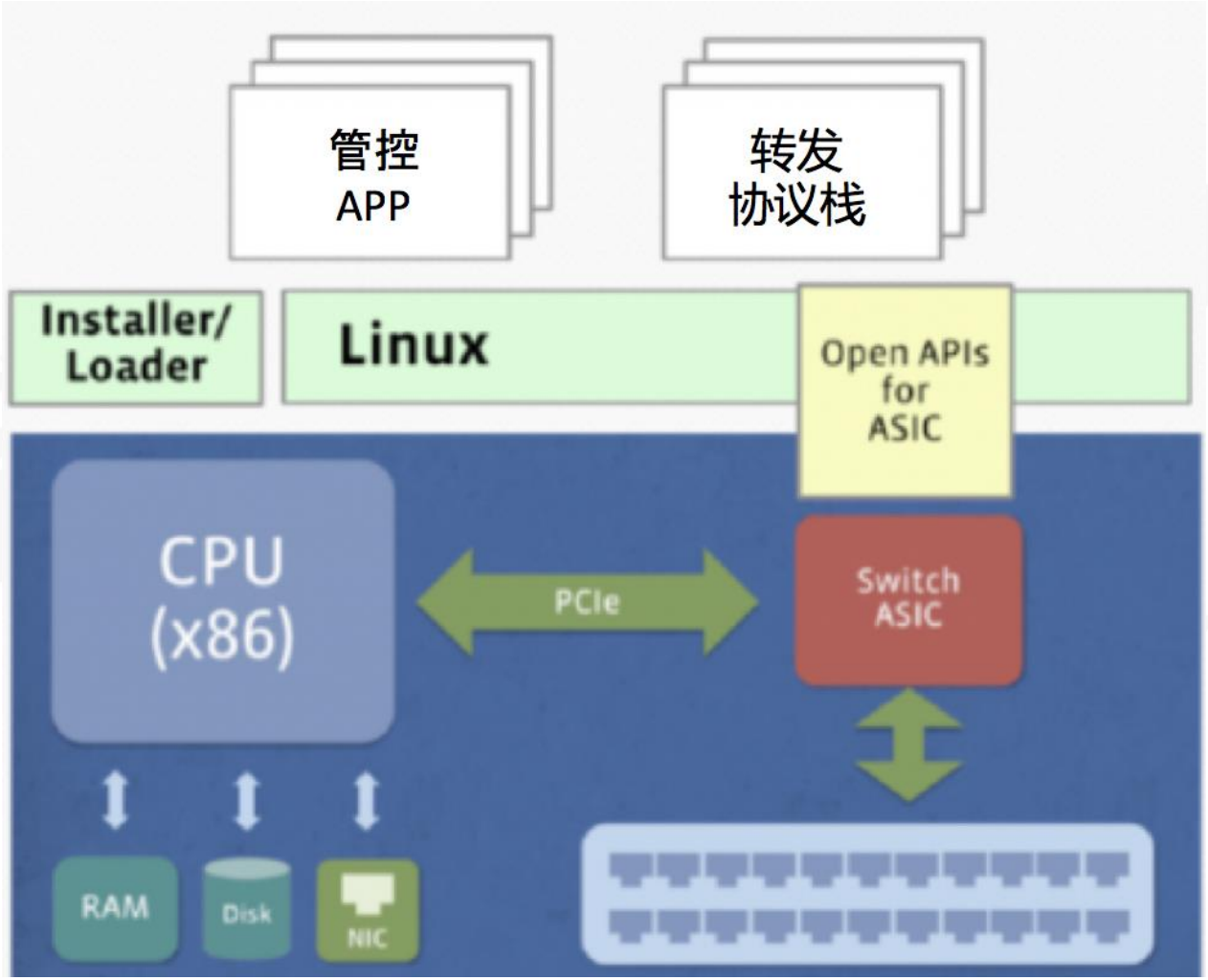
CAVIUM

MARVELL

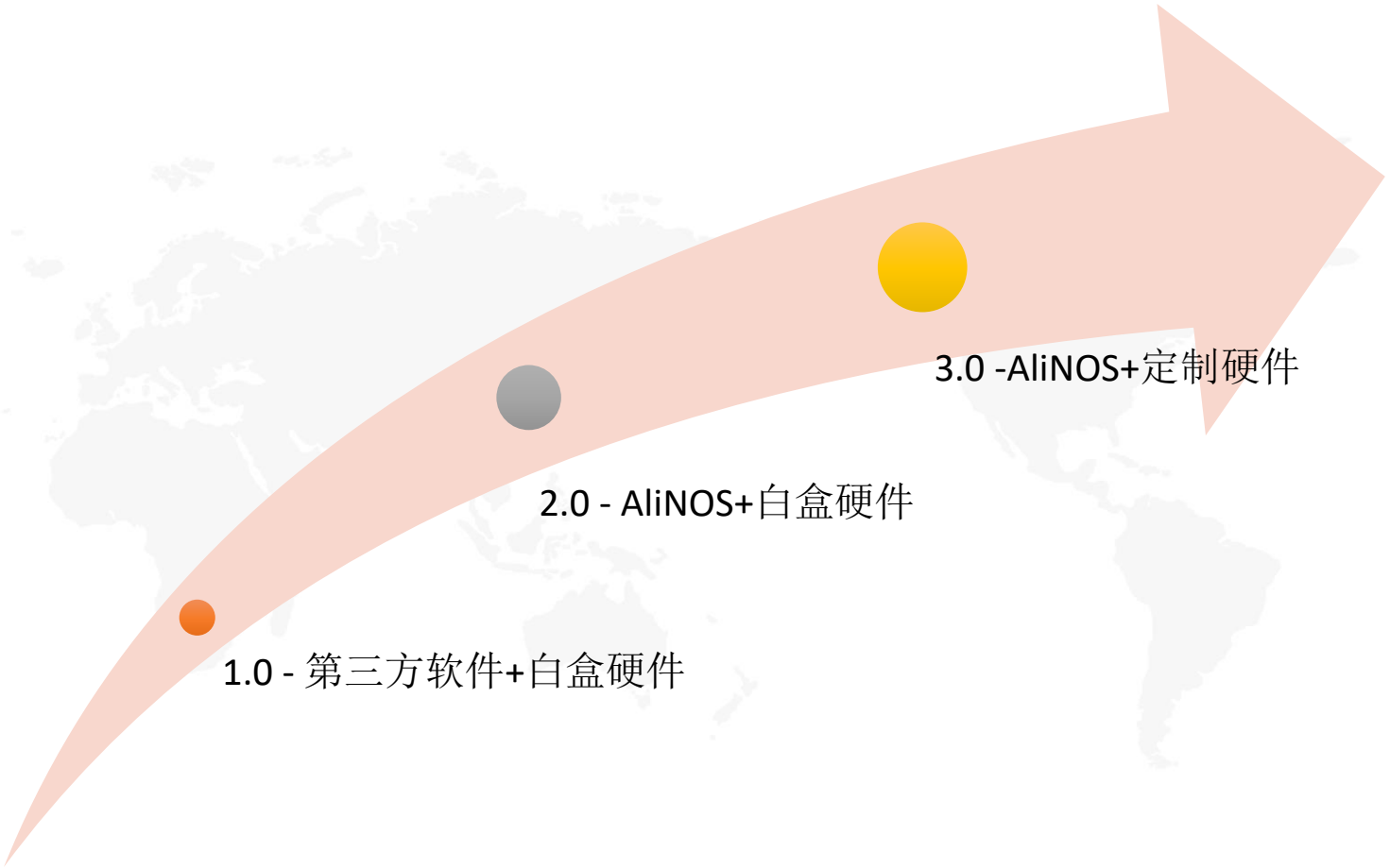
Mellanox TECHNOLOGIES

centec networks

Nephos


管控
APP
转发
协议栈
Installer/
Loader
Linux
Open APIs
for
ASIC
CPU
(x86)
PCIe
Switch
ASIC
RAM
Disk
NIC

阿里自研交换机演进历程




----- Alibaba Network Operating System -----
Unauthorized access and/or use are prohibited.
All access and/or use are subject to monitoring.
Help: https://workitem.aone.alibaba-inc.com/project/604087

AliNOS从开发到运维

需求评估		研发测试				生产灰度		规模运维			
NOS选择	网络架构	特性评估	软件架构	特性开发	测试迭代	社区合作	对接管控	故障演练	规模交付	日常巡检	故障处理	日常变更

基于Ali需求对SONiC的评估

结构化配置

1.0

- 通过中心化的minigraph文件实现设备的结构化配置

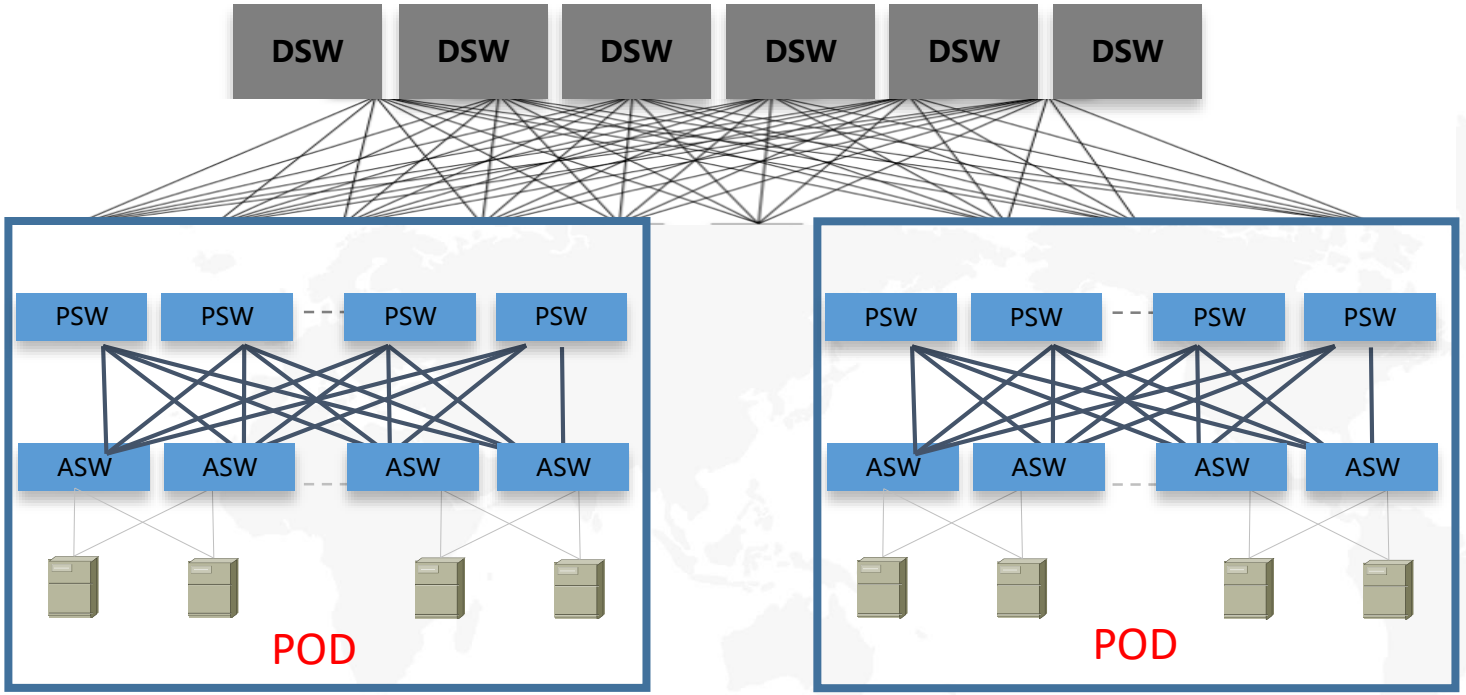
2.0

- 引入ConfigDB，支持增量配置的变更和保存
- 开发CLI framework支持运维日常巡检和变更

3.0

- 结合SONiC DB定义了基于gNMI模型驱动的结构化数据
- 支持完整的gRPC管控方案，支持dial-in/dial-out多种模式

自研AliNOS研发 – 定制开发



25G / 100G

3-tier CLOS
50K+ servers

double uplink to ASW

RDMA

- 基础特性
  - TACACS+, LACP fallback, Vlan trunk, gRPC
- 阿里定制需求
  - 支持双上联
  - 支持RDMA
  - 支持INT

- 场景
  - 日常迭代
  - 版本更新
  - 上线发布
  - 平台引入
  - 问题复现
- 目标
  - 全覆盖
  - 自动化
- 挑战
  - 无法也不可能搭建和线上规模1:1的测试环境
  - 测试对象包括复杂的硬件版本和硬件拓扑
  - 测试覆盖率除了复杂的软件、芯片功能，还包括基于实际硬件的性能和稳定性

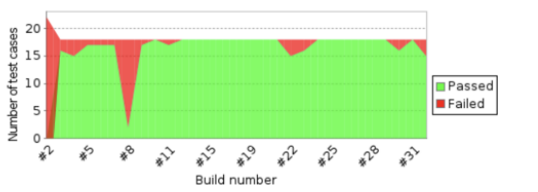
自研AliNOS研发 – 测试自动化

- 多维度测试环境
  - 软件测试环境
    - UT，VS
  - 混合测试环境
    - 社区Ansible，阿里Ansible
  - 硬件测试环境
    - 单机、组网
- 自动化测试框架
  - 多环境复用
  - 跨设备操作自动化
  - 设备原子操作抽象
  - 重用测试用例
  - 自动生成测试报告

Robot Framework Test Results

Executed: 20171207 08:12:28.227
Duration: 2:26:25.822 (-0:25:07.217)
Status: 18 critical test, 15 passed, 3 failed
18 test total (±0), 15 passed, 3 failed
Results: report.html
log.html
Original result files

Test Result Trend



Zoom to changes Show only failed Show only critical all Max builds

Show bigger image

Failed Test Cases

Name	Crit.	Duration	Age
Dev-Master.T0 Test.Test bgp_fact	yes	0:00:08.279	1
Dev-Master.T0 Test.Test bgp_speaker	yes	0:01:09.918	1
Dev-Master.T0 Test.Test lag_2	yes	0:01:38.535	1

Test Suites

Name	Failed tests (diff)	Total tests (diff)	Duration (diff)
Dev-Master	3 (+3)	18 (±0)	2:26:25.822 (-0:25:07.217)
Dev-Master.T1-Lag Test	0 (±0)	2 (±0)	0:37:22.103 (-0:01:35.333)
Dev-Master.T1 Test	0 (±0)	2 (±0)	0:29:32.859 (+0:01:21.112)
Dev-Master.T0 Test	3 (+3)	12 (±0)	0:46:45.752 (-0:24:55.333)
Dev-Master.Ptf32 Test	0 (±0)	2 (±0)	0:25:26.860 (+0:00:19.875)

自研AliNOS生产灰度

- 灰度阶段
  - 实验室
  - 小批量
  - 大规模
- 对接管控
  - 配置生成
  - 版本升级
  - 模版开发
- 故障演练
  - Failover场景
  - 设备隔离操作
  - 业务压力测试

自研AliNOS规模运维

- 规模交付
  - POAP/ZTP
- 日常运营
  - 版本管理
  - 升级策略
  - 配置变更
- 故障处理
  - 设备巡检
  - 故障告警
  - 故障定位
  - 故障恢复

总结：

- 自研的目的是为了更好的满足大规模数据中心对网络的要求
- 开放的数据中心软硬件生态为自研提供了条件
- 自研NOS的评估和开发要结合自身的业务和网络需求
- 完备的测试和灰度策略是自研设备规模部署的关键
- 从厂商到自研的引入意味着运维模式的转变
  - 保守升级 vs 快速迭代
  - 命令行操作 vs 结构化管控
  - 依赖厂商 vs 自主开发
  - 关注厂商差异 vs 关注系统软硬件

THANKS!