



全球运维大会

2016

DevOps 2.0: 重塑运维价值



北京站

会议时间：12月16日 - 12月17日

会议地点：北京国际会议中心

主办单位：



阿里大数据计算平台运维实践

范伦挺（萧一）
阿里巴巴集团 技术专家



目录

- ➔ **1** 大规模计算平台运维挑战
- 2** 自动化平台建设
- 3** 数据驱动精细化运维
- 4** 运维转型思考



阿里大数据架构演进

ODPS平台上线

单集群超5K

异地多活/离在线混布



异地多集群/异地灾备

单集群超10K



大规模计算平台运维挑战

1. 规模大、小概率事件常态化

- 各类硬件故障
- 网络链路不稳定
- 工具容易出问题

2. 多机房多地域

- 延时增加
- 资源不均衡



目录

- 1 大规模计算平台运维挑战
- 2 自动化平台建设
- 3 数据驱动精细化运维
- 4 运维转型思考



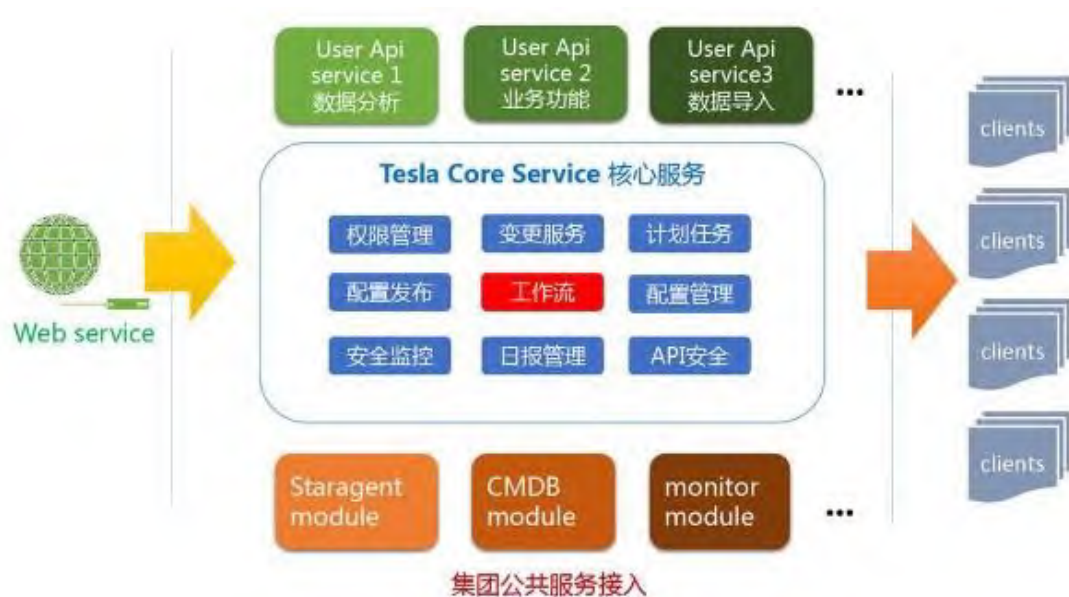
自动化平台建设

1. 自动化意义

- 稳定
- 提高效率

2. 自动化方向

- 变更
- 问题排查
- 硬件维修
- 交付检查



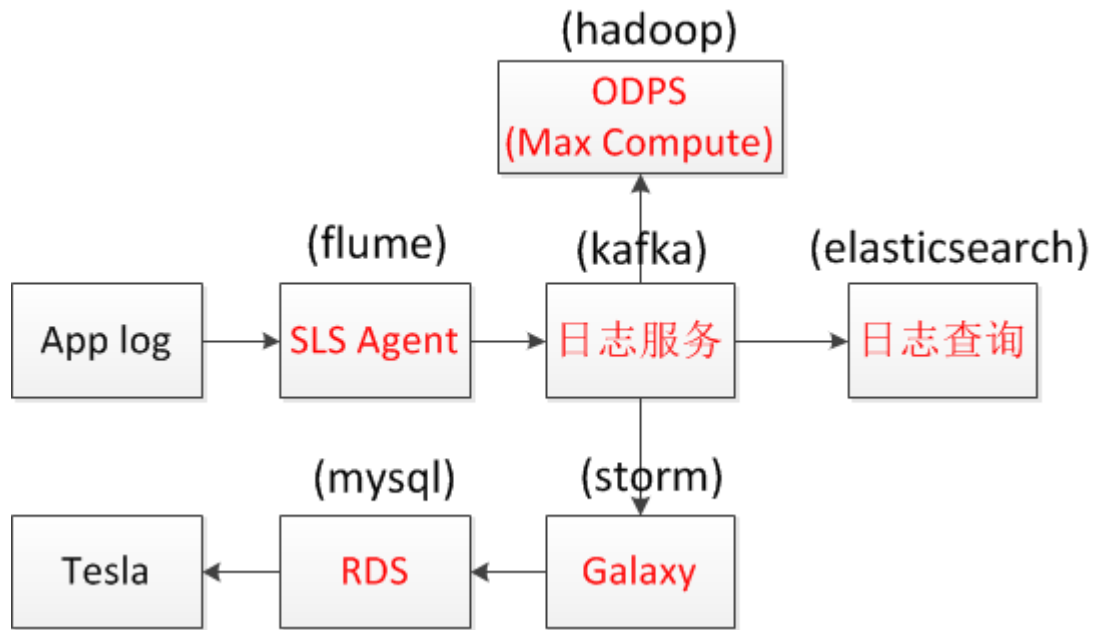
自动化平台建设

1. 变更自动化

The screenshot displays the Tesla ODPs (Operational Data Processing) interface for managing automation. At the top, it shows the account '28210 (CF编 号: 219974)' and '当前进度 100%'. A progress bar at the top right indicates 20 steps, with step 12 highlighted in orange. The main interface is divided into two sections: '准备' (Preparation) and '工作流: ODPS 前端front' (Workflow: ODPS Frontend). The '准备' section shows '已完成' (Completed) with a '提交change free' button. The '工作流' section shows '只可查看' (View Only) and a '查看执行详情' (View Execution Details) button, which is highlighted with a red box and a red arrow. The right side of the interface shows the execution logs for the workflow, including '执行成功' (Execution Successful) and '报警方式' (Alert Method) options (旺旺, 短信, 电话). The logs also show the '超时时间(秒): 900' and '剩余自动重试次数: 0'.

自动化平台建设

2. 问题排查



阿里实时日志分析架构



自动化平台建设

2. 问题排查

时间段快速选择:

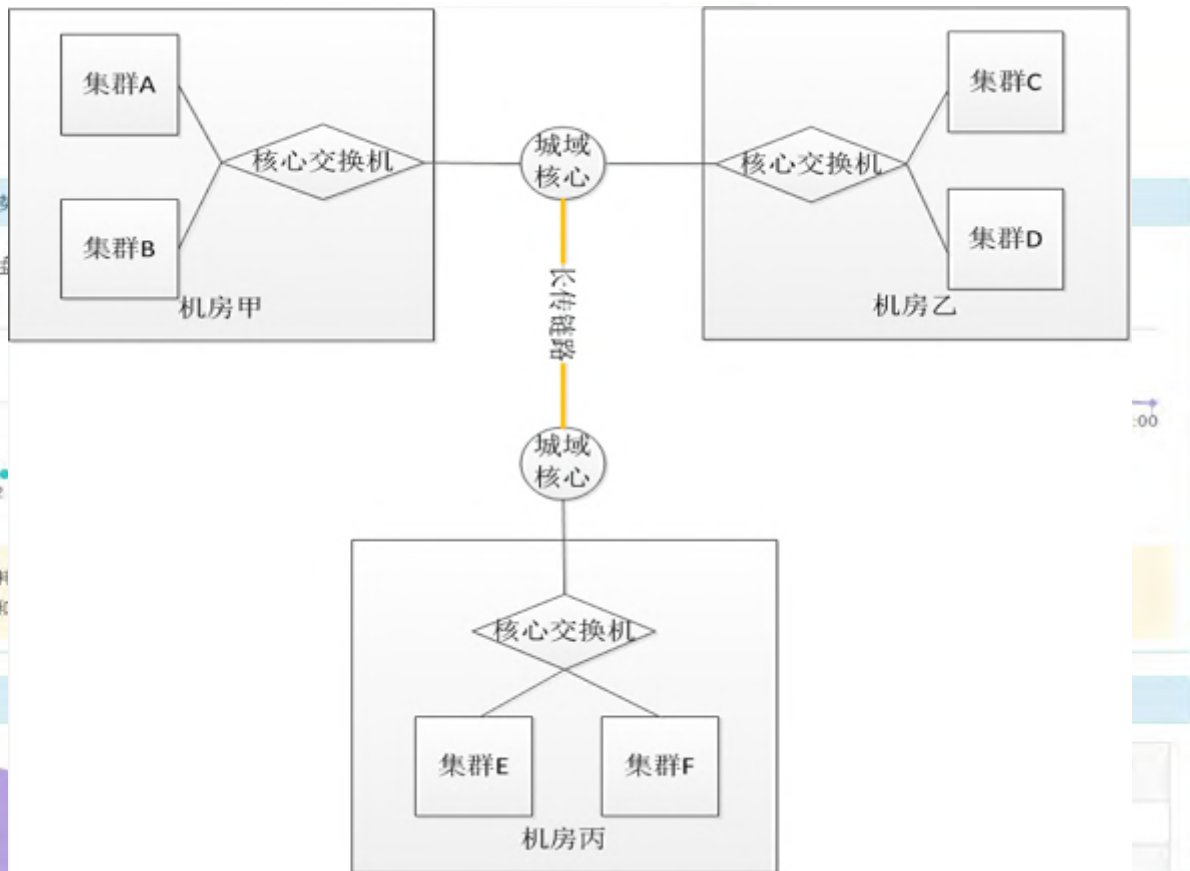
盘古Top直读流量趋势

最近N分钟的趋势



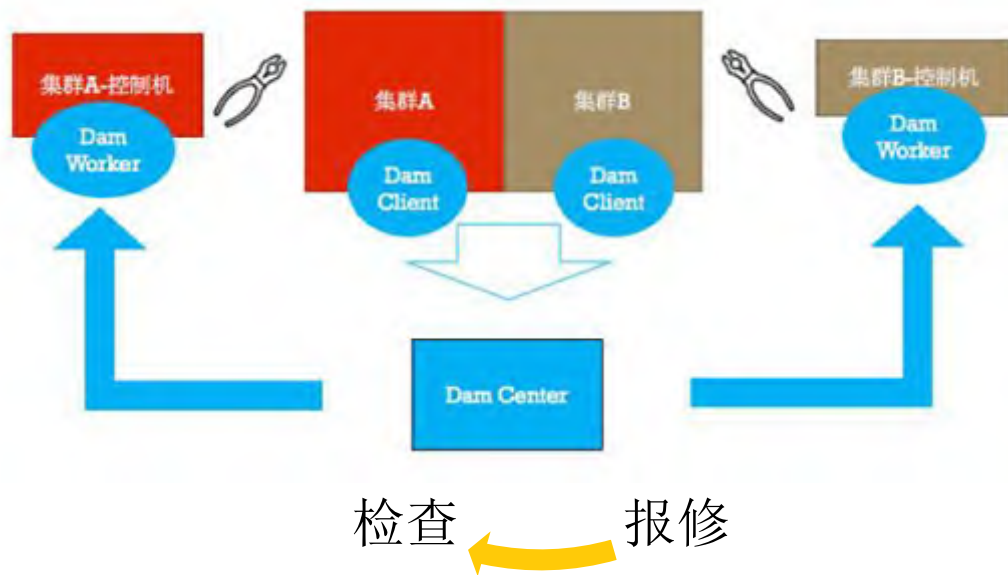
- 最近一段时间内盘古消耗
- 点击查看instance趋势

TopN分布



自动化平台建设

3. 硬件维修-DAM



自动化平台建设

3. 硬件维修-DAM

Dam-checker信息来源：

- 硬盘/板卡：kernel log/smartctl/tsar
- 内存：ipmitool/mcelog/stream
- CPU/风扇：mcelog/cpu频率/ipmitool
- 网络/网卡/交换机端口：tsar/kernel log
- 主板：排除以上可能



自动化平台建设

3. 硬件维修-效果

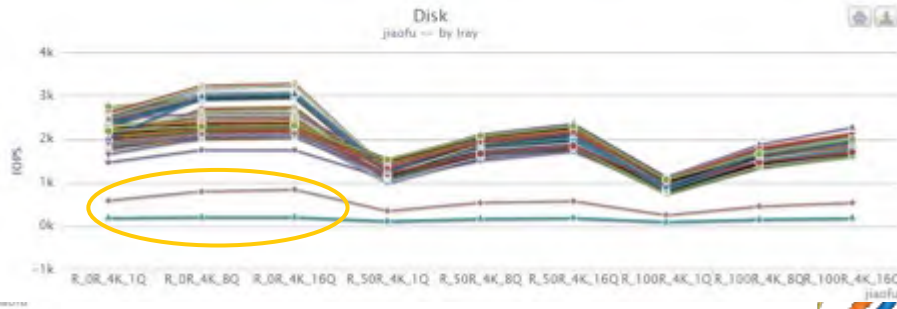
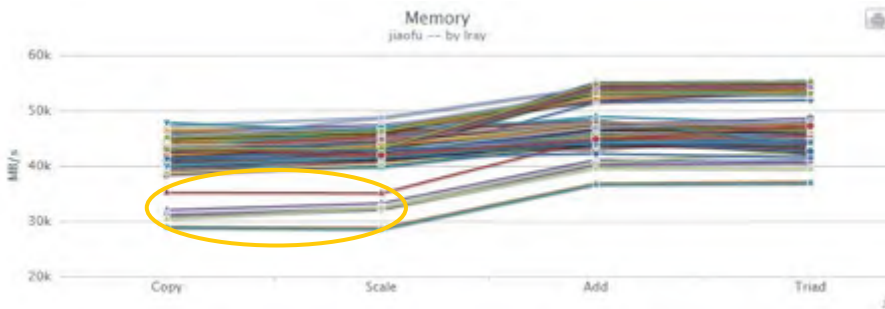
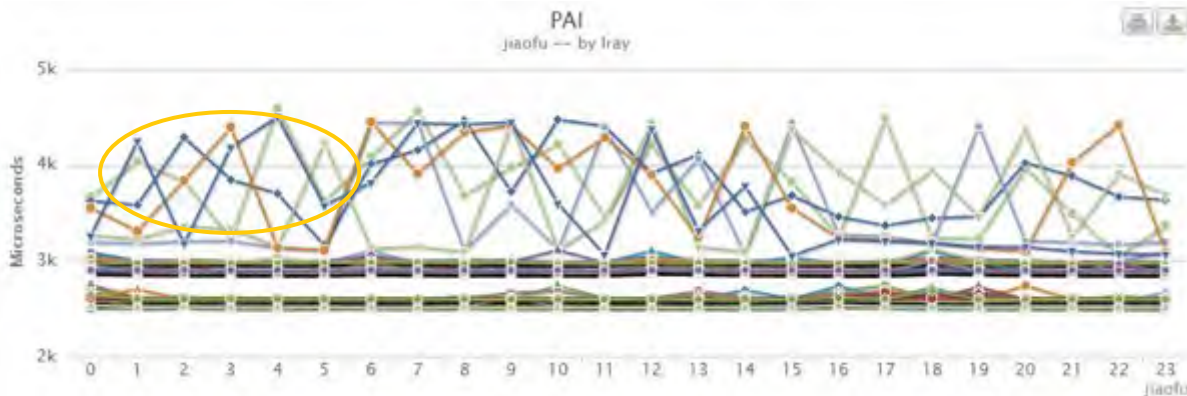
集群	机器数(dam覆盖)	异常工单	7天工单趋势	今天工单	24小时新增工单	1小时内新增	稳定分趋势	稳定分
		2		31		0		97.92
		8		37		0		95.76
		6		37		1		98.46
		8		82		0		97.81
		8		90		2		95.69
		5		115		1		93.43
		76		241		0		95.50



自动化平台建设

4. 交付检查

- 软件交付检查
- 硬件交付检查



目录

1 大规模计算平台运维挑战

2 自动化平台建设

→ 3 数据驱动精细化运维

4 运维转型思考



数据驱动精细化运维



知己知彼，方能百战不殆！
数据能通晓过去未来之事！



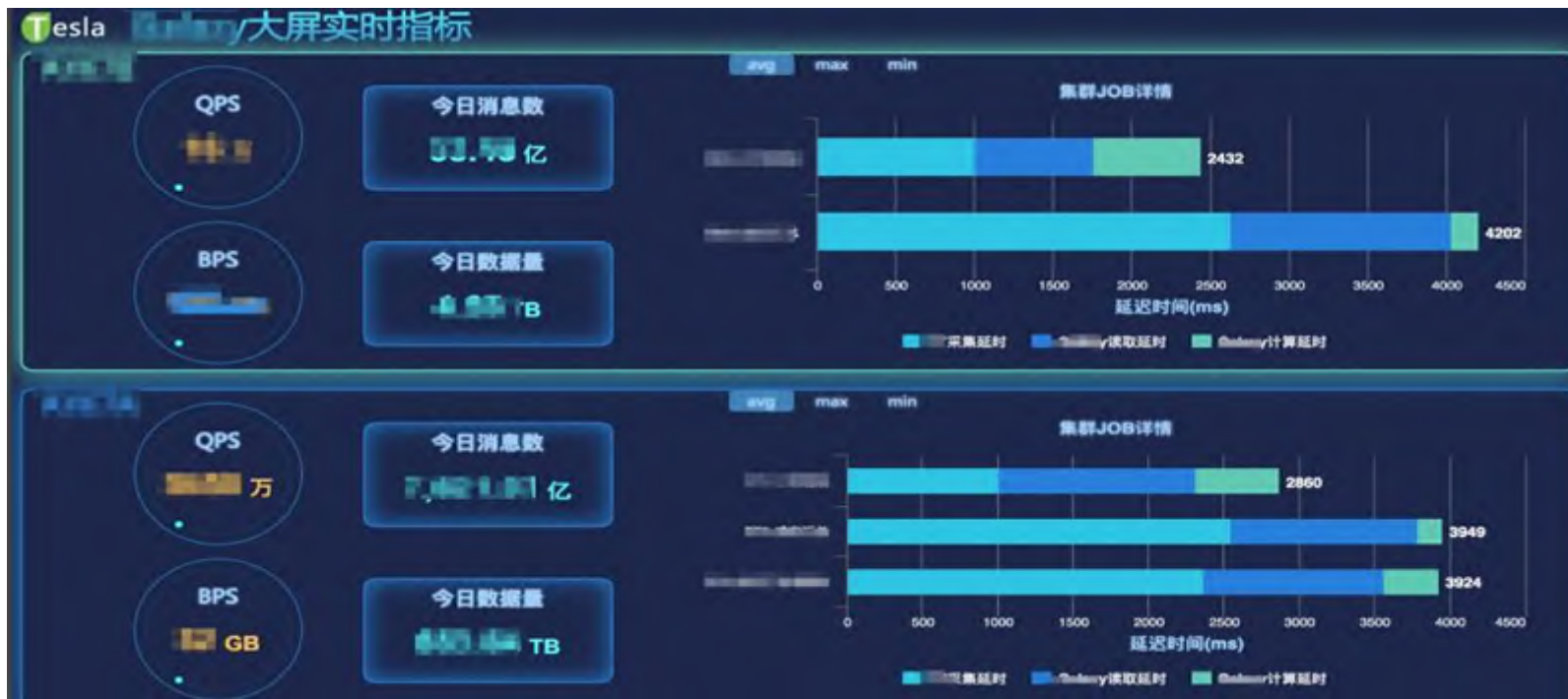
数据驱动精细化运维

1. 实时大屏



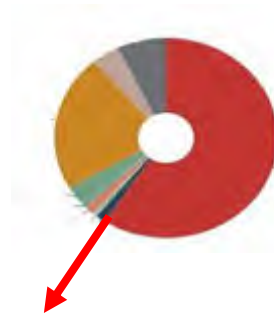
数据驱动精细化运维

1. 实时大屏



数据驱动精细化运维

2. 存储分析



以4T单盘为例，ext4默认格式化inode数为2.4亿，inode的大小是256 Bytes。

单台12块盘机器Inode占用= $2.4 \times 10^8 \times 256 \times 12 / 1024 / 1024 / 1024 = 686.6\text{GB}$

万台集群= $686.6 \times 10000 / 1024 = 6.54\text{PB}$

按平均文件大小10M计算，单盘所需inode仅为420000，即使按照10倍计算inode仅需
 $42 \times 10^5 / (2.4 \times 10^8) = 1.75\%$ ，至少可以**节省6PB**

注：可优化量根据系统平均文件大小定。



数据驱动精细化运维

3. 资源优化



数据驱动精细化运维

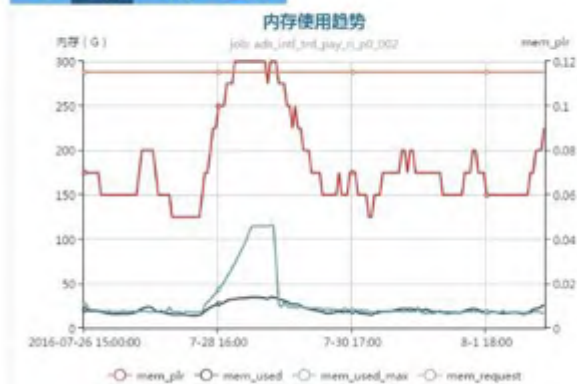
3. 资源优化

job总体信息

集群	用户	jobid	logview	delay		
		ads_intl_trd_pay_ri_p0_002-73-1469189062	logview	0		
Worker总数	内存申请(LGV) (M)	内存使用(PUV) (M)	内存使用率(PLR)	cpu申请(LGV) (core)	cpu使用(PUV) (core)	cpu使用率(PLR)
48	294912	26987	0.09	48	2.95	0.06

job使用趋势

一天 一周 一个月 三个月



目录

1 大规模计算平台运维挑战

2 自动化平台建设

3 数据驱动精细化运维

→ 4 运维转型思考



运维转型思考-运维向运营转型

运维：稳定、安全 --活着（“眼前的苟且”）

运营：服务、效益 --发展（“诗和远方”）



运维转型思考-自动化进程



人肉时代（人是一切）



工具时代（人为主，器为辅）



智慧时代（突破人的能力极限）



平台时代（器为主，人为辅）



运维转型思考-效率向价值转型



喝咖啡？



数据分析



数据化/可视化

运维转型思考



DevOpsDays 即将首次登陆中国



DevOps 之父 Patrick Debois 与您相约
DevOpsDays 北京站 2017年3月18日



门票早鸟价仅限前100名，请从速哟

<http://2017-beijing.devopsdayschina.org/>



GOPS2016
Beijing



想第一时间看到
高效运维社区公众号
的好文章吗？



请打开高效运维社区公众号，点击右上角小人，如右侧所示设置就好



Thanks

高效运维社区
开放运维联盟

荣誉出品



GOPS2016
Beijing