

自动化与数据化运维

-从硬盘开始

刘毅





“大数据” 时代的变化

- 服务器硬件架构
- 存储架构
- 人员配比
- 带电操作

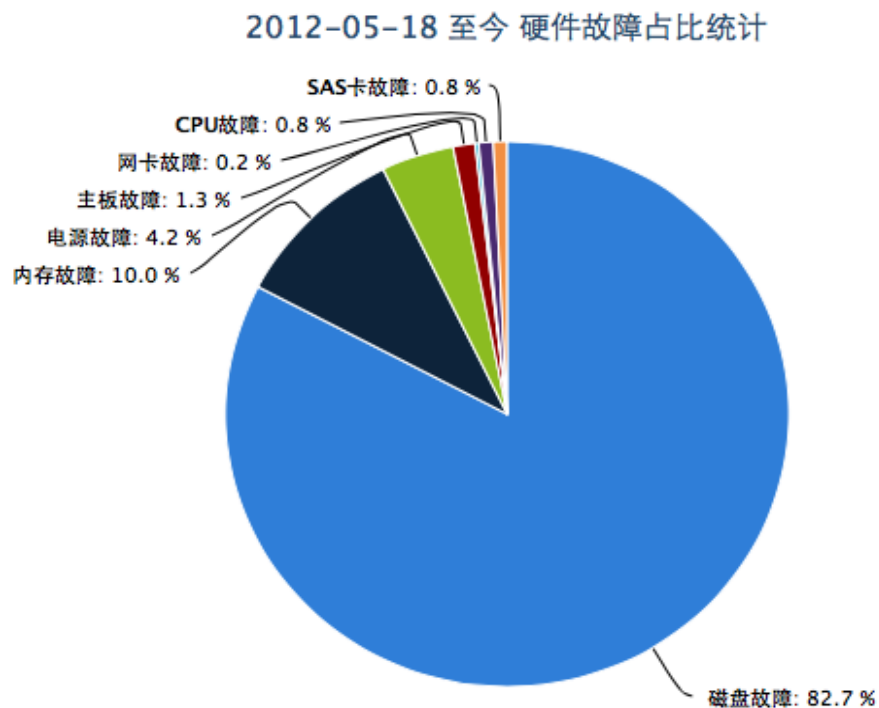
96%

3/Day





运维面临诸多挑战 - 以硬盘为例：



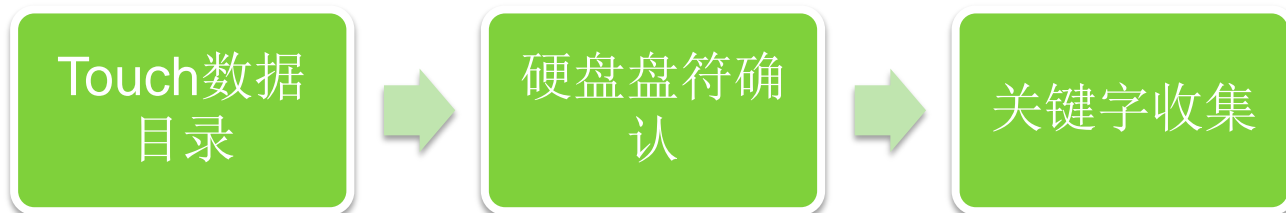


- 文件系统易进入只读模式
- 各种修复流程复杂，操作时间长
- 基数大、无时无刻存在

繁琐复杂，劳动密集



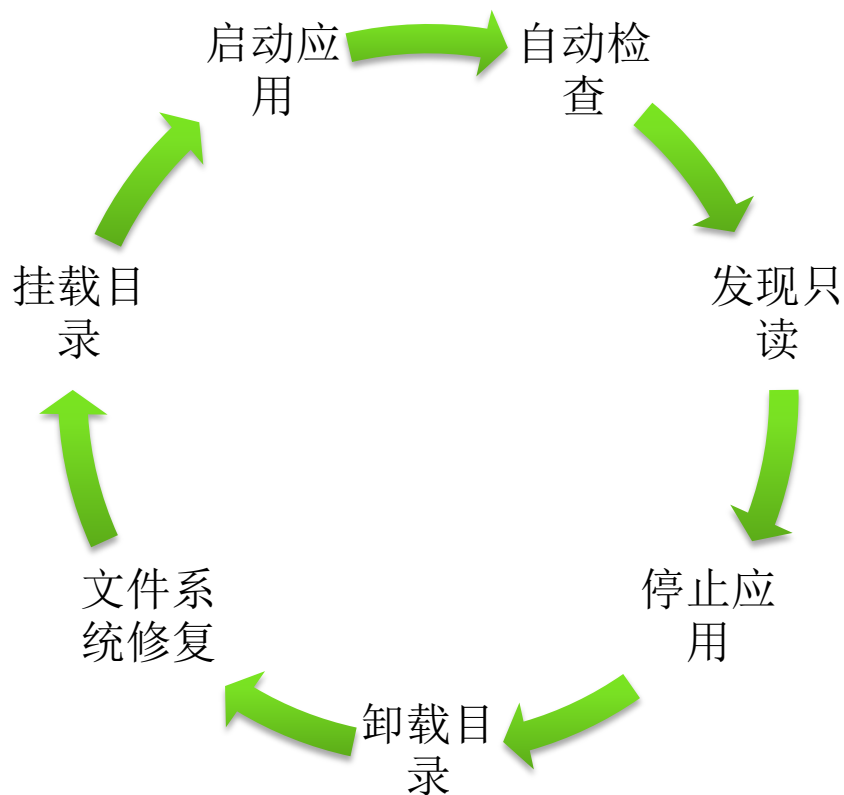
一、文件系统、坏盘





修复自动化:

- 报警多，处理繁琐，耗时长
- 每次修复需要停掉一台节点的应用





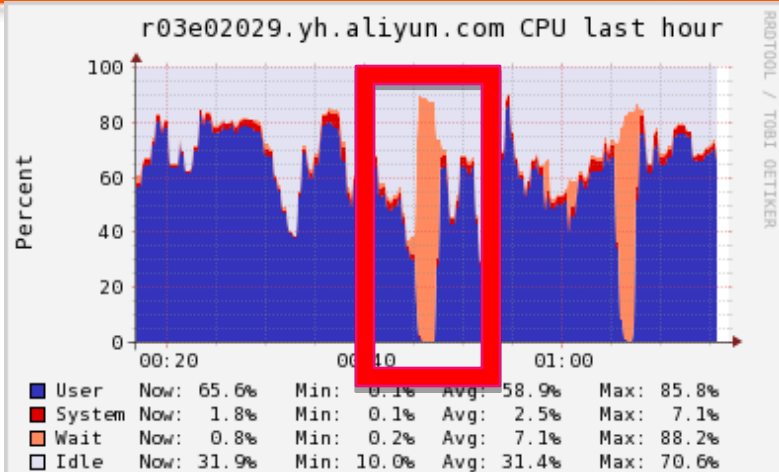
- 紧急疑难问题
- 影响生产稳定

危险系数高，紧急救火



二、慢盘、僵尸盘

- 事后分析问题场景
- 不断的补充新的场景



sdi 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00 0.00 0.00 0.00 0.00 100.00

689202 transaction, each up to 8192 blocks
average:

0ms waiting for transaction

5339ms running transaction

0ms transaction was being locked

0ms flushing data (in ordered mode)

47ms logging transaction

21160085us average transaction commit time

```

root      7791  0.0
root      24270 0.0
root      24296 0.0
root      24420 0.0
root      24440 0.0
admin     29609 0.0
admin     11411 0.0
admin     13028 0.0
admin     14440 0.0
admin     15034 0.0
admin     15444 0.0
admin     15826 0.0
admin     20968 0.6

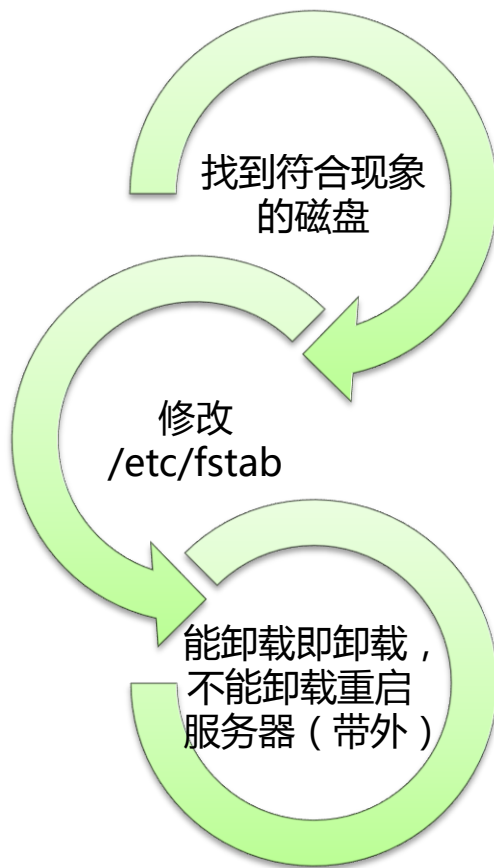
D 13:45 0:00 python /tmp/dayu_app_me_run/check_file_system.py
D 14:04 0:00 python /tmp/dayu_app_me_run/check_file_system.py
D 14:12 0:00 python /tmp/dayu_app_me_run/check_file_system.py
D 14:17 0:00 python /tmp/dayu_app_me_run/check_file_system.py
D 14:19 0:00 python /tmp/dayu_app_me_run/check_file_system.py
S 14:22 0:00 /usr/local/eagleeye/bin/agent_eagleeye -d
D 14:36 0:10 /apsara/tubo/TempRoot/sys/pangu/ChunkServerRole@r62o12

gu/disk7/check_filesystem.08150806
-o all -x dd -g 10.206.48.11
ome/admin/jiaofu/main/subs/fio_ran.sh
ne=R_OR_4K_1Q --rw=randrw --direct=1 -
ne=R_OR_4K_1Q --rw=randrw --direct
--name=R_OR_4K_1Q --rw=randrw --direct
/nome/admin/aresolve/aresolve --dayuclient
python /tmp/dayu_app_me_run/check_file_system.py
python /tmp/dayu_app_me_run/check_file_system.py
python /tmp/dayu_app_me_run/check_file_system.py
python /tmp/dayu_app_me_run/check_file_system.py
python /tmp/dayu_app_me_run/check_file_system.py
0:00 /usr/local/eagleeye/bin/agent_eagleeye -d
/apsara/tubo/TempRoot/sys/pangu/ChunkServerRole@r62o12

```




二、慢盘、僵尸盘自动化解决:





- 足够多的样本
- 拼概率、拼人品
- 事后补救可能于事无补

防患于未然



三、提前预判硬盘健康

- SMART & SMARTMON TOOLS

```
ID# ATTRIBUTE_NAME
1 Raw_Read_Error_Rate
3 Spin_Up_Time
4 Start_Stop_Count
5 Reallocated_Sector_Ct
7 Seek_Error_Rate
9 Power_On_Hours
10 Spin_Retry_Count
12 Power_Cycle_Count
184 End-to-End_Error
187 Reported_Uncorrect
188 Command_Timeout
189 High_Fly_Writes
190 Airflow_Temperature_Cel
191 G-Sense_Error_Rate
192 Power-Off_Retract_Count
193 Load_Cycle_Count
194 Temperature_Celsius
195 Hardware_ECC_Recovered
```

ID#	ATTRIBUTE_NAME	FLAG	VALUE	WORST	THRESH	...	UPDATED	WHEN_FAILED	RAW_VALUE
1	Raw_Read_Error_Rate	0x000f	084	063	044	Pre-fail	Always	-	241720330
3	Spin_Up_Time	0x0003	100	100	000	Pre-fail	Always	-	0
4	Start_Stop_Count	0x0032	100	100	020	Old_age	Always	-	9
5	Reallocated_Sector_Ct	0x0033	001	001	036	Pre-fail	Always	FAILING_NOW	4096
7	Seek_Error_Rate	0x000f	088	060	030	Pre-fail	Always	-	774500633
9	Power_On_Hours	0x0032	073	073	000	Old_age	Always	-	24392
10	Spin_Retry_Count	0x0013	100	100	097	Pre-fail	Always	-	0
12	Power_Cycle_Count	0x0032	100	100	020	Old_age	Always	-	9
184	End-to-End_Error	0x0032	100	100	099	Old_age	Always	-	0



三、提前预判硬盘健康

- 无效的Thresholds值
- 规则依赖单一参数

改进优化

- 参数联动
- 增加其他预判纬度

```
==== START OF READ SMART DATA SECTION ====
SMART Self-test log structure revision number 1
Num Test_Description Status报查询 Remaining LifeTime(hours) LBA_of_first_error
# 1 Short offline Completed: read failure 90% 15334 1216353448
# 2 Short offline Completed: read failure 60% 15310 1216351855
# 3 Short offline Completed: read failure 10% 15238 1216351855
# 4 Short offline Completed: read failure 50% 15214 1216351855
# 5 Short offline Completed: read failure 50% 15142 1216353448

DH = Device/Head Register [HEX]
DC = Device Command Register [HEX]
ER = Error register [HEX]
ST = Status register [HEX]
```

学习互联网与自
• 重要的参数
• ATA Error



三、提前预判硬盘健康

磁盘健康检查日报

建议处理：

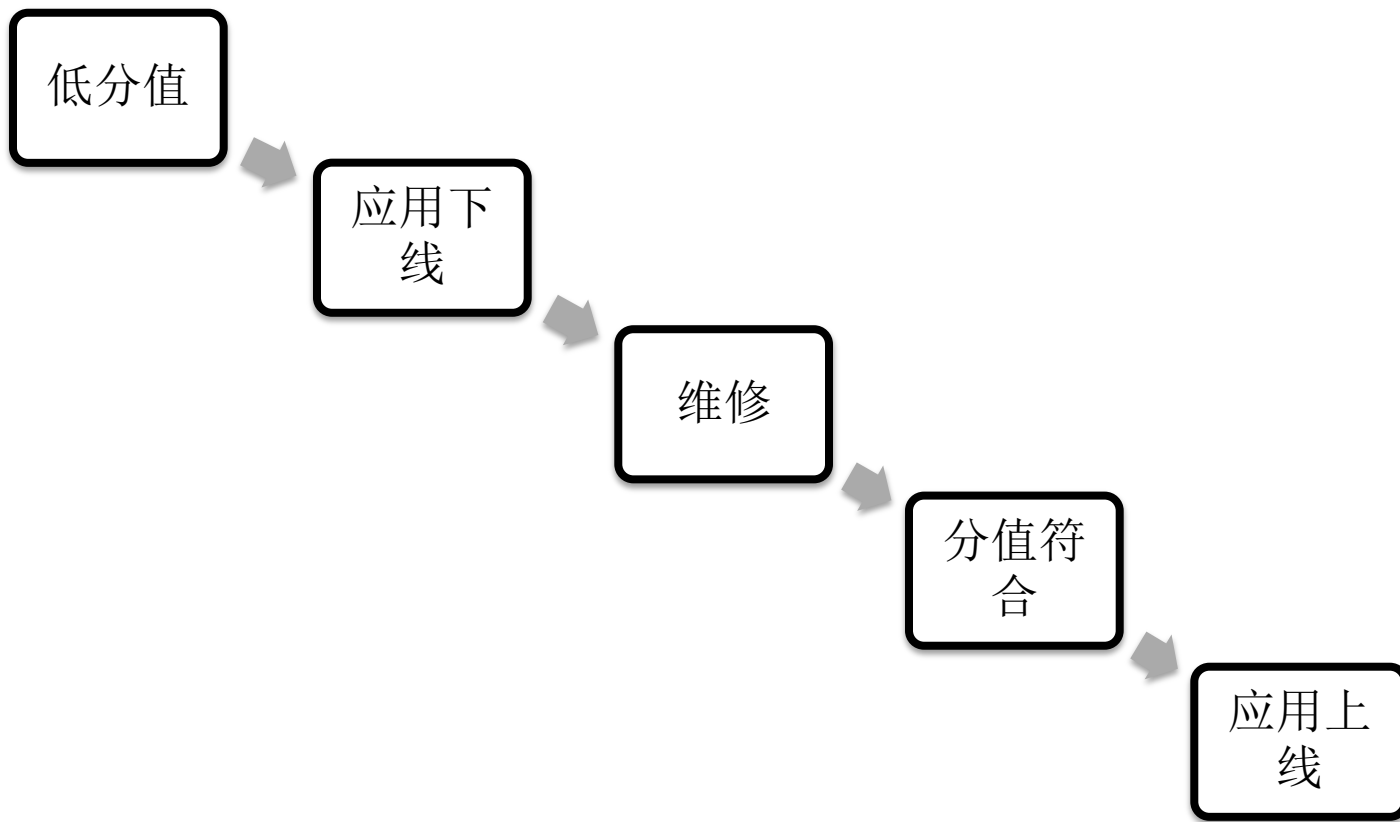
/dev/sdg	[reformat_disk]
/dev/sdm	[reformat_disk]
/dev/sdd	[reformat_disk]
/dev/sdi	[reformat_disk]
/dev/sde	[reformat_disk]
/dev/sdg	[replace_disk]

554 , 以下只列出异常

	磁盘分数											
	sdd:100	sde:100	sdf:100	sdg:100	sda:100	sdb:100	sdc:100	sdl:100	sdh:100	sdi:10	sdj:100	sdk:100
	sdd:100	sde:100	sdf:100	sdg:100	sda:100	sdb:100	sdc:25	sdl:100	sdh:100	sdi:100	sdj:100	sdk:100
	sdd:100	sde:100	sdf:26	sdg:100	sda:100	sdb:100	sdc:100	sdl:100	sdh:100	sdi:100	sdj:100	sdk:100
PE待确认	sdd:100	sde:100	sdf:100	sdg:100	sda:100	sdb:100	sdc:100	sdl:100	sdh:100	sdi:100	sdj:100	sdk:23
PE待确认	sdd:100	sde:100	sdf:100	sdg:100	sda:100	sdb:100	sdc:100	sdl:100	sdi:5	sdh:100	sdi:100	sdj:100
	sdd:M->M	sde:M->M	sdf:M->M	sdg:M->M	sda:M->M	sdb:M->M	sdc:M->M	sdl:M->M	sdh:M->M	sdi:M->M	sdj:M->M	sdk:M->M
	sdd:M->M	sde:M->M	sdf:M->M	sdg:M->M	sda:M->M	sdb:M->M	sdc:M->M	sdl:M->M	sdh:M->M	sdi:M->M	sdj:M->M	sdk:M->M
	sdd:M->M	sde:M->M	sdf:M->M	sdg:M->M	sda:M->M	sdb:M->M	sdc:M->M	sdl:M->M	sdh:M->M	sdi:M->M	sdj:M->M	sdk:M->M
	sdd:M->M	sde:M->M	sdf:M->M	sdg:M->M	sda:M->M	sdb:M->M	sdc:M->M	sdl:M->M	sdh:M->M	sdi:M->M	sdj:M->M	sdk:M->M
	sdd:100	sde:100	sdf:100	sdi:5	sda:100	sdb:100	sdc:100	sdl:100	sdh:100	sdi:100	sdj:100	sdk:100
	sdd:100	sde:100	sdf:100	sdg:100	sda:100	sdb:100	sdc:11	sdl:100	sdh:100	sdi:100	sdj:100	sdk:100
	sdd:100	sde:100	sdf:24	sdg:100	sda:100	sdb:100	sdc:100	sdl:70->100	sdh:100	sdi:100	sdj:100	sdk:100
	sdd:M->M	sde:M->M	sdf:M->M	sdg:M->M	sda:M->M	sdb:M->M	sdc:M->M	sdl:M->M	sdh:M->M	sdi:M->M	sdj:M->M	sdk:M->M
	sdd:100	sde:100	sdf:100	sdi:0	sda:100	sdb:100	sdc:100->70	sdl:70->100	sdh:100	sdi:100	sdj:100	sdk:100

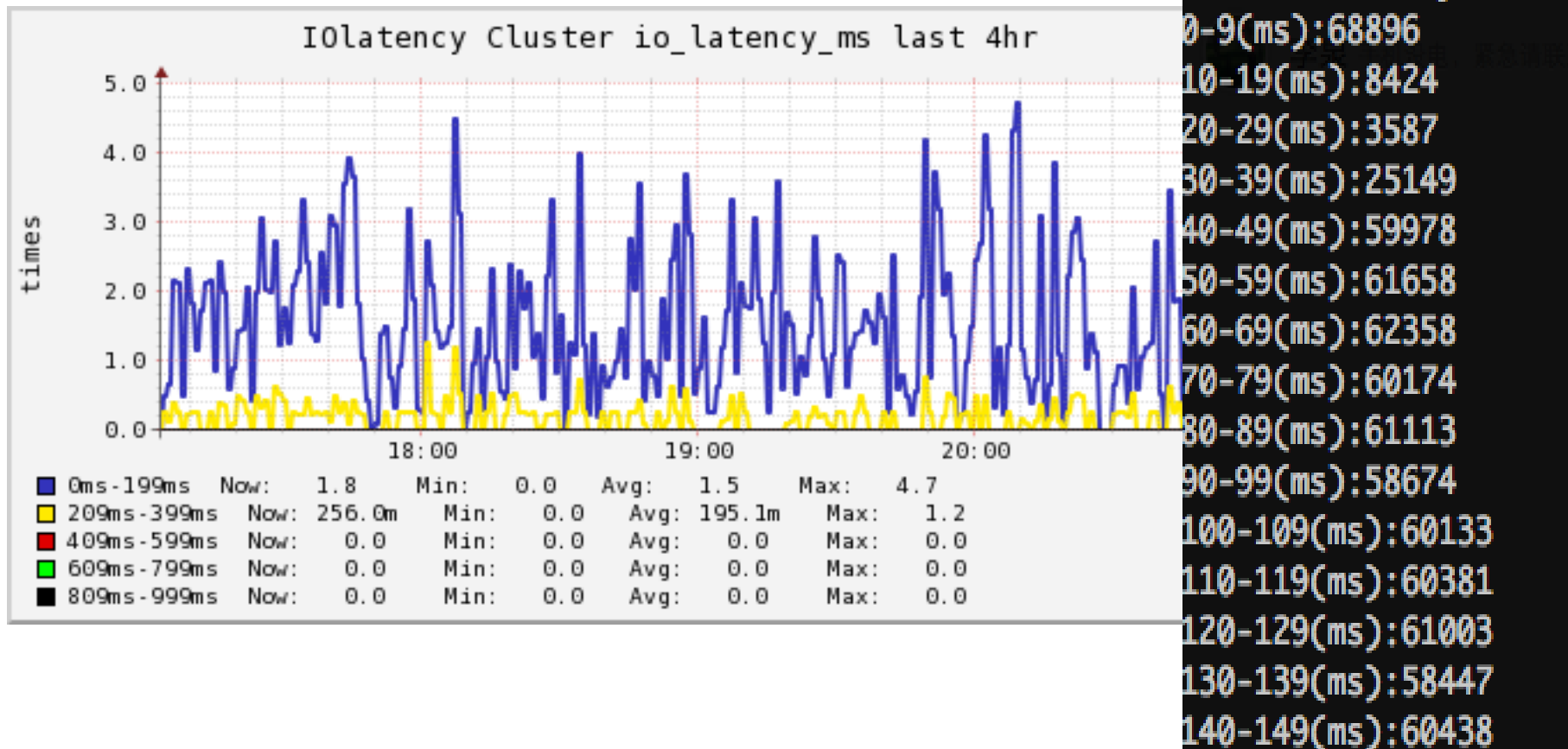


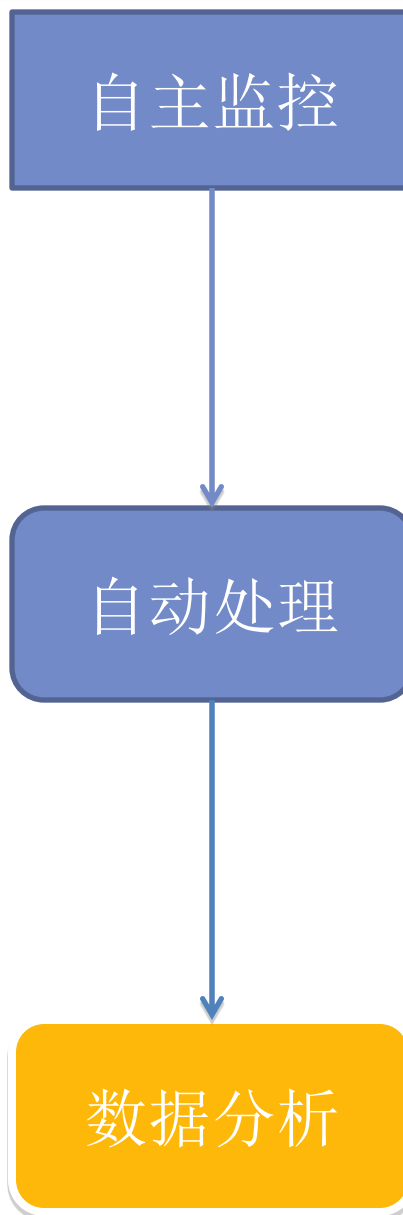
三、预警与平台相配合：





四、与内核团队合作







一、日志数据分析：

云计算平台的133天中：

- 625次文件系统的fsck。
- 平均每天需进行4.7次的修复。
- 保守按耗费半小时来计算，造成每天有141分钟额外单台不可用时间（需要停掉应用再卸载目录）

应用更加智能：

- 应用程序能允许在线卸载目录
- 在线挂载目录后应用程序能够识别
- 选用更可靠的EXT4文件系统



141分钟
downtime

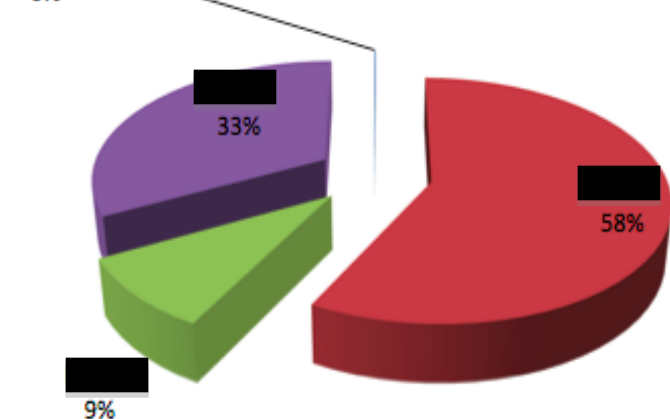
减少恢复时间



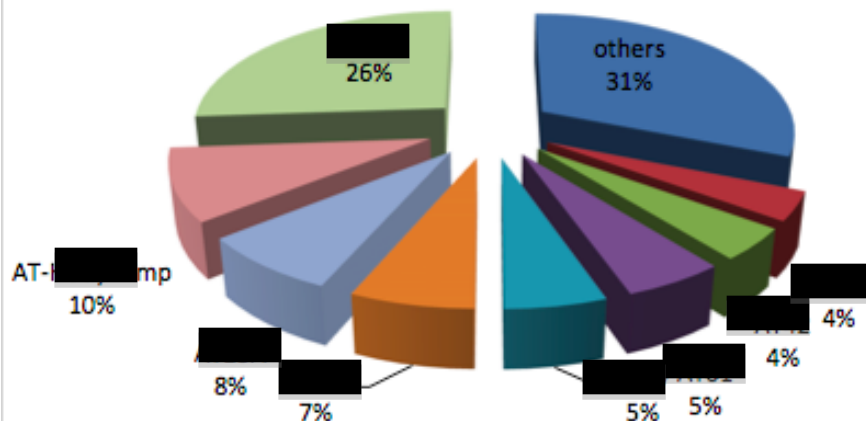


二、运维数据分析：

硬盘厂商各自占有的百分率



按集群分布



改进现有运维方式：

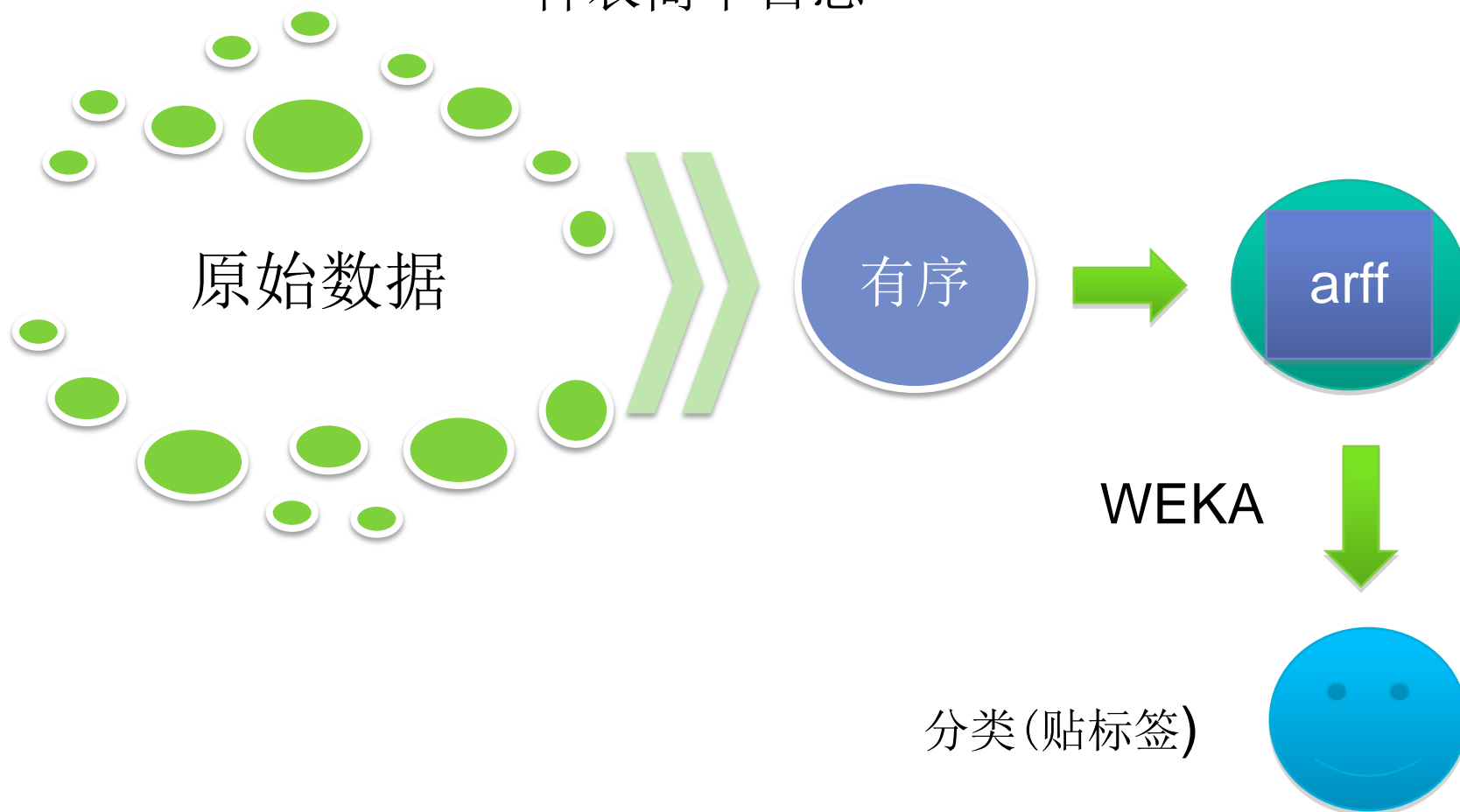
- 减少扩容带来的影响
- 感觉有了数据的支撑
- 提高运维效率

排名\指标	wio	
	机器名	昨天值(次数)
1	r01e14034	4.87(1)
2	r05c08041	4.76(1)
3	r02a08001	4.66(21)
4	r02b05041	4.26(18)
5	r02a08009	4.19(0)



三、海量运维数据挖掘初探：

神农简单日志

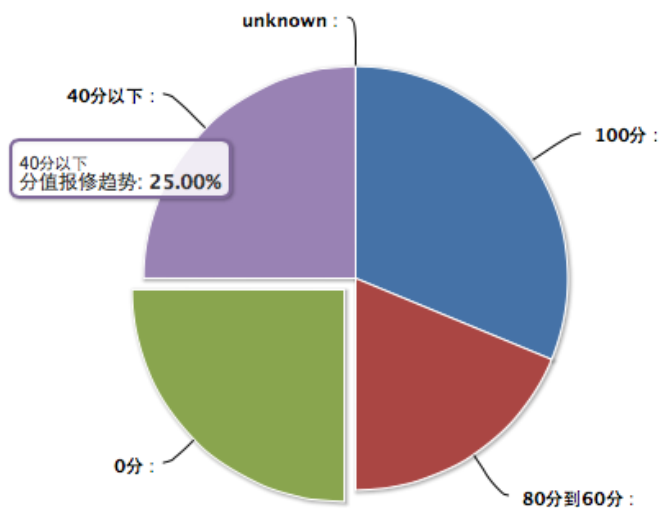




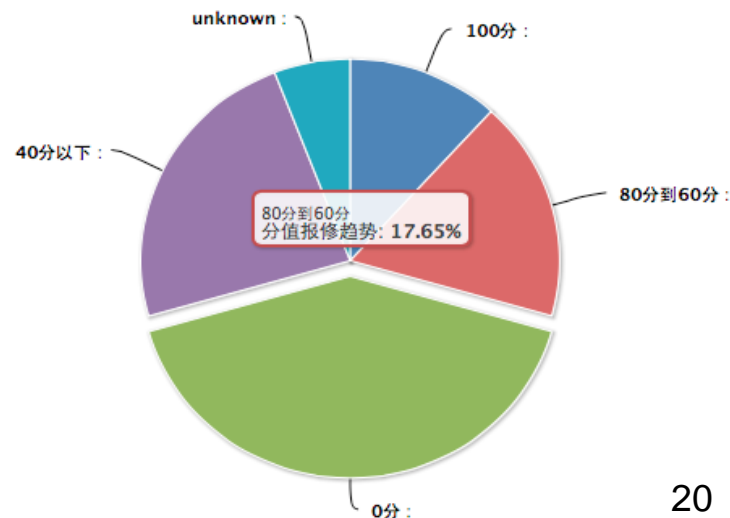
三、海量运维数据挖掘初探：

- 验证规则
- 新的发现

报修趋势



报修趋势





- 自动高效的收集数据
- 对数据敏感、借助平台的力量
- 提高运维自动化的效率，预防故障发生



感谢：

- 云计算PE团队

谢谢

