阿里云
专有云Agility版

告警参考

产品版本：V1.1.0

文档版本：20180416

法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的”现状“、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含”阿里云”、Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

表 -1: 格式约定

格式	说明	样例
	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	 **禁止**： 重置操作将丢失用户配置数据。
	该类警示信息可能导致系统重大变更甚至故障，或者导致人身伤害等结果。	 **警告**： 重启操作将导致业务中断，恢复业务所需时间约10分钟。
	用于警示信息、补充说明等，是用户必须了解的内容。	 **说明**： 导出的数据中包含敏感信息，请妥善保存。
	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	 **说明**： 您也可以通过按**Ctrl** + **A**选中全部文件。
>	多级菜单递进。	**设置** > **网络** > **设置网络类型**
粗体	表示按键、菜单、页面名称等UI元素。	单击 **确定**。
courier字体	命令。	执行 cd /d C:/windows 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	bae log list --instanceid *Instance_ID*
[]或者[a|b]	表示可选项，至多选择一个。	ipconfig *[-all|-t]*
{}或者{a|b}	表示必选项，至多选择一个。	swich *{stand | slave}*

目录

1 基础告警

1.1 Alarm-01.100.0000.0000-硬盘使用率过高

当检测到任何一个分区磁盘使用率或者inode使用率超过80%时，产生P4告警，超过90%时产生P1告警。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	tianji	硬件,tianji.TianjiClient

处理方法

1. 登录天基控制台。
2. 选择**运维** > **机器运维**，查找报警的host或ip。
3. 选中需要的机器，打开terminal service。
4. 执行df和df -i命令，查看硬盘空间占用情况。
5. 执行du命令，查找空间占用大的目录。
6. 执行docker images命令，查找是否是过期docker images太多。
7. 如果该服务的硬盘空间虽然占用较大，但使用率恒定，可以在**服务运维** > **监控实例**下编辑报警规则，调大报警阈值。

1.2 Alarm-01.100.0005.0000-NTP时间出现偏移

${sync}!=0 ${offset}>500。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	tianji	硬件,tianji.TianjiClient

处理方法

1. 登录天基控制台。
2. 选择**运维** > **机器运维**，查找报警的host或IP。
3. 选中需要的机器，打开terminal service。
4. 执行date命令，查看该host机器时间是否正常。

5. 执行ntpdate time.ntp.org命令，找到ntp服务器。
6. 在crontab中添加"0 12 * * * * /usr/sbin/ntpdate"。

1.3 Alarm-01.100.0010.0000-内存使用率过高

$util_max>95。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	tianji	硬件,tianji.TianjiClient

处理方法

1. 登录天基控制台。
2. 选择**运维** > **机器运维**，查找报警的host或ip。
3. 选中需要的机器，打开terminal service。

   一般情况下，linux服务器内存占用率较高时，还要同时判断服务是否正常，服务正常可以不做任何处理。
4. 如果服务不正常，执行top命令，并按内存列排序，找出内存占用最大的进程，重启进程。
5. 同步报备阿里云技术支持。

1.4 Alarm-01.100.0015.0000-CPU负载过高

$load5_max>20。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	tianji	硬件,tianji.TianjiClient

处理方法

1. 登录天基控制台。
2. 选择**运维** > **机器运维**，查找报警的host或IP。
3. 选中需要的机器，打开terminal service。

   一般情况下，linux服务器cpu load较高时，同时判断服务是否正常，服务正常可以不做任何处理。

4. 如果服务不正常，执行top命令，按cpu占用率排序，找出cpu占用最大的进程，重启进程。
5. 同步报备阿里云技术支持。

1.5 Alarm-01.100.0020.0000-网络流量过高

$ifin_max>52428800||$ifout_max>52428800。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	tianji	硬件,tianji.TianjiClient

处理方法

1. 登录天基控制台。
2. 选择**运维** > **机器运维**，查找报警的host或IP。
3. 选中需要的机器，打开terminal service。

   一般情况下，linux服务器cpu load较高时，需要同时判断服务是否正常，服务正常可以不做任何处理。

1.6 Alarm-01.100.0025.0000-CPU使用率过高

$util_max>200。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	tianji	硬件,tianji.TianjiClient

处理方法

1. 登录天基控制台。
2. 选择**运维** > **机器运维**，查找报警的host或IP。
3. 选中需要的机器，打开terminal service。

   一般情况下，linux服务器cpu 使用率较高时，需要同时判断服务是否正常，服务正常可以不做任何处理。
4. 如果服务不正常，执行top命令，按CPU占用率排序，找出cpu占用最大的进程，重启进程。
5. 同步报备阿里云技术支持。

1.7 Alarm-01.100.0030.0000-TCP连接数据过多

$_ports_max>500||$_timewait_max>100000||$_closed_max>100000||$_estab_max>100000||$_TCP_max>100000。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	tianji	硬件,tianji.TianjiClient

处理方法

1. 登录天基控制台。
2. 选择**运维** > **机器运维**，查找报警的host或IP。
3. 选中需要的机器，打开terminal service。
4. 在TianjiPortal的机器视图中，查看流量监控图。
5. 如果是网络突发抖动，请观察是否还会出现类似情况。

   如果是持续上涨，请扩容该服务，或者增强硬件配置。
6. 其它情况请联系阿里云技术支持。

1.8 Alarm-01.100.0035.0005-主机Ping响应超时

$rta_avg>500||$loss_max>80。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1	tianji	远程,tianji.TianjiClient

处理方法

各业务引入时自定义。

1.9 Alarm-01.100.0035.0010-主机SSH无响应

$state!=0。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1	tianji	远程,tianji.TianjiClient

处理方法

各业务引入时自定义。

1.10 Alarm-01.100.0035.0015-HTTP探测无响应

$state!=0。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1		远程,tianji.TianjiClient

处理方法

各业务引入时自定义。

1.11 Alarm-01.100.0035.0020-页面探测无响应

$state!=0。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1	-	远程,tianji.TianjiClient

处理方法

各业务引入时自定义。

1.12 Alarm-01.100.0035.0025-主机Ping无响应

$state!=0。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1	-	远程,tianji.TianjiClient

处理方法

各业务引入时自定义。

1.13 Alarm-01.100.0035.0030-端口探测失败

$state!=0。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1	-	远程,tianji.TianjiClient

处理方法

各业务引入时自定义。

1.14 Alarm-01.100.0035.0035-vip探测失败

$state!=0。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1		远程,tianji.TianjiClient

处理方法

各业务引入时自定义。

1.15 Alarm-01.100.0035.0040-URL检查失败

$state!=0。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1		远程,tianji.TianjiClient

处理方法

各业务引入时自定义。

1.16 Alarm-01.100.0080.0000-JVM线程状态异常

$count_max>2000||$deadlock_count_max>0。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	tianji	硬件,tianji.TianjiClient

处理方法

1. 登录天基控制台。
2. 选择**运维** > **机器运维**，查找报警的host或IP。
3. 选中需要的机器，打开terminal service。
4. 执行docker ps命令，查找相应的容器。
5. 查看jvm进程，是否可以自动恢复。

   如果可以自动恢复，建议修改报警规则，增加出现异常次数。

   如果规律的多次出现，建议对该服务进行扩容。

1.17 Alarm-01.100.0090.0000-JVM GC次数过多

$count_max>2000。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	tianji	硬件，tianji.TianjiClient

处理方法

1. 进入TianjiPortal，选择**运维** > **机器运维**，查找报警的host或IP。
2. 选中你要的机器，打开terminal service，执行docker ps命令，查找到相应的容器。
3. 查看jvm进程，看是否可以自动恢复，如果可以自动恢复，建议修改报警规则，增加出现异常次数。

   如果规律的多次出现，建议对该服务进行扩容。

1.18 Alarm-01.100.0115.0000-JVM堆内存使用率过高

$usage_max>93。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	tianji	硬件,tianji.TianjiClient

处理方法

1. 登录天基控制台。
2. 选择**运维** > **机器运维**，查找报警的host或IP。
3. 选中需要的机器，打开terminal service。
4. 执行docker ps命令，查找相应的容器。
5. 查看jvm进程，是否可以自动恢复，如果可以自动恢复，建议修改报警规则，增加出现异常次数
6. 如果规律的多次出现，建议对服务进行扩容。

1.19 Alarm-01.100.0125.0000-容器硬盘使用率过高

当检测到任何一个分区磁盘使用率或者inode使用率超过80%时，产生P4告警，超过90%时，产生P1告警。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	tianji	硬件,tianji.TianjiClient

处理方法

1. 登录天基控制台。
2. 选择**运维** > **机器运维**，查找报警的host或IP。
3. 选中需要的机器，打开terminal service。
4. 执行docker ps命令，查找到相应的容器。
5. 执行df和df -i命令，查看硬盘空间占用情况。
6. 执行du命令，查看空间占用大的目录。
7. 执行docker images命令，查找是否是过期docker images太多。

如果该服务的硬盘空间虽然占用较大，但使用率恒定，可以在**服务运维** > **监控实例**下编辑报警规则，调大报警阈值。

1.20 Alarm-01.100.0130.0000-容器NTP时间出现偏移

${sync}!=0 ${offset}>500。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	tianji	硬件,tianji.TianjiClient

处理方法

1. 登录天基控制台。
2. 选择**运维** > **机器运维**，查找报警的host或IP。
3. 选中需要的机器，打开terminal service。

   执行docker ps命令，查找到相应的容器。
4. 执行date命令，查看该host机器时间是否正常。
5. 找到ntp服务器，执行ntpdate time.ntp.org命令。

   在crontab中添加"0 12 * * * * /usr/sbin/ntpdate"。

1.21 Alarm-01.100.0135.0000-容器内存使用率过高

$util_max>95。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	tianji	硬件,tianji.TianjiClient

处理方法

1. 登录天基控制台。
2. 选择**运维** > **机器运维**，查找报警的host或IP。
3. 选中你要的机器，打开terminal service。
4. 执行docker ps命令，查找到相应的容器。

一般情况下，linux服务器内存占用率较高时，需要同时判断服务是否正常，服务正常可以不做任何处理。

5. 如果服务不正常，执行top命令，按内存列排序，找出内存占用最大的进程，重启进程。
6. 同步报备阿里云技术支持。

1.22 Alarm-01.100.0140.0000-容器CPU负载过高

$load5_max>20。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	tianji	硬件,tianji.TianjiClient

处理方法

1. 登录天基控制台。
2. 选择**运维** > **机器运维**，查找报警的host或IP。
3. 选中需要的机器，打开terminal service。
4. 执行docker ps命令，查找到相应的容器。

   一般情况下，linux服务器cpu load较高时，需要同时判断服务是否正常，服务正常可以不做任何处理。
5. 如果服务不正常，执行top命令，按CPU占用率排序，找出CPU占用最大的进程，重启进程。
6. 同步报备阿里云技术支持。

1.23 Alarm-01.100.0145.0000-容器网络流量过高

$ifin_max>52428800||$ifout_max>52428800。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	tianji	硬件、tianji.TianjiClient

处理方法

1. 登录天基控制台。
2. 选择**运维** > **机器运维**，查找报警的host或IP。

3. 选中需要的机器，打开terminal service。
4. 执行docker ps命令，查找到相应的容器。

   一般情况下，linux服务器cpu load较高时，还要同时判断服务是否正常，服务正常可以不做任何处理。

1.24 Alarm-01.100.0150.0000-容器CPU使用率过高

$util_max>200。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	tianji	硬件、tianji.TianjiClient

处理方法

1. 登录天基控制台。
2. 选择**运维** > **机器运维**，查找报警的host或IP。
3. 选中需要的机器，打开terminal service。
4. 执行docker ps命令，查找到相应的容器。

   一般情况下，linux服务器cpu 使用率较高时，需要同时判断服务是否正常，服务正常可以不做任何处理。
5. 如果服务不正常，执行top命令，按cpu占用率排序，找出cpu占用最大的进程，重启进程。
6. 同步报备阿里云技术支持。

1.25 Alarm-01.100.0155.0000-容器TCP连接数据过多

$_ports_max>500||$_timewait_max>100000||$_closed_max>100000||$_estab_max>100000||$_TCP_max>100000。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	tianji	硬件,tianji.TianjiClient

处理方法

1. 登录天基控制台。

2. 选择**运维** > **机器运维**，查找报警的host或IP。
3. 选中需要的机器，打开terminal service。
4. 执行docker ps命令，查找到相应的容器。
5. 在TianjiPortal的机器视图中，查看流量监控图。

   如果是网络突发抖动，请观察是否还会出现类似情况。
6. 如果是持续上涨，请扩容该服务，或者增强硬件配置。
7. 其它情况，请报备阿里云技术支持。

2 对象存储OSS

2.1 Alarm-02.305.0001.00001-check_nginx_port

当ocm所在机器的80端口无法连接时，会产生该告警。 当80端口可以连接时，该告警会自动清除。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1	oss-ocm-server	oss-ocm-server. TEngine#

可能原因

- 机器坏了。
- 进程未启动。

影响范围

如果所有OCM都不能连接，OSS就不能正常工作。如果还有OCM能工作，可能不会影响OSS正常服务。

处理方法

1. 执行curl IP命令，检查是否有响应。
   - 如果有，表示误报。
   - 如果没有，请跳转至下一步。
2. 执行ps auxf|grep -e "tengine\|nginx"|grep -v grep命令，检查是否有nginx相关的进程。
   - 如果有，收集信息联系技术支持。
   - 如果没有，查看天基没有启动tengine或者nginx进程的原因，查看*apsara/apache/logs/*路径下是否有当天日志，收集信息联系技术支持

2.2 Alarm-02.305.0001.00002-check_ocm_server_process_fix

当ocm_server进程重启的时候，会产生该告警。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1	oss-ocm-server	oss-ocm-server.OcmServer#

可能原因

- 资源不够。
- 配置不对导致重启程序bug。
- 升级后，重启了进程忘记关报警。

影响范围

OCM不能服务可能会影响正常的OSS读写。

处理方法

1. 执行ps auxf|grep ocm_server|grep -v grep命令，查看日志，找到ocm_server启动的进程所在的目录。
2. 进入目录，查看apsara_log_conf.json，查找日志打印的位置。
3. 收集相关日志和配置文件ocm_conf.json。
4. 收集本机的系统日志dmsg。
5. 联系技术支持。

2.3 Alarm-02.305.0001.00003-oss_check_net_error_drop

当网卡有丢包的时候，会产生该告警，当不丢包时会恢复。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P4	oss-ocm-server	oss-ocm-server.TEngine#

可能原因

- 网卡不好。

- 网络有问题。

影响范围

OCM不能服务可能会影响正常的OSS读写。

处理方法

需要配合OSS整体服务来看，如果OSS服务正常，不影响OSS的正确请求可暂时不处理。

检查网卡流量是否超过上限。

- 如果是，表示正常，可能需要扩容，以减少网络的流量。
- 如果否，联系硬件工程师，查看网卡是否正常。

2.4 Alarm-02.305.0001.00004-check_tengine_ssl_cert_expire_stat

检查tengine的ssl证书文件是否快要过期，证书有效期小于30天则报警。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	oss-ocm-server	oss-ocm-server.TEngine#

可能原因

有证书，且证书过期。

处理方法

1. 申请新的证书。
2. 替换证书。

专有云一般情况下没有证书。

2.5 Alarm-02.305.0001.00005-check_2ethstatus_oss

检查网卡是否工作正常，使用自己的脚本可以灵活指定需要的网络监测规则，比如10000M网卡，包括是否在正常设置的speed，是否双工。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	oss-ocm-server	oss-ocm-server.TEngine#

可能原因

- 网卡坏了。
- 配置不对。

处理方法

联系驻场工程师替换网卡。

2.6 Alarm-02.305.0001.00006-check_nginx_process

当nginx进程重启的时候，会产生该告警。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1	oss-ocm-server	oss-ocm-server.TEngine#

可能原因

- 资源不够。
- 配置不对导致重启程序bug。
- 升级后，重启了进程忘记关报警。

影响范围

OCM不能服务可能会影响正常的OSS读写。

处理方法

1. 查看*/apsara/apache/logs*目录access log中是否正常。

2. 执行curl http://IP地址/systermoperation/checkocmstatus -i命令，查看是否为200，OSS服务是否恢复正常。
   - 如果正常，记录进程重启时间，如果不影响服务，可以暂时不用处理。
   - 如果不正常，服务不可用，联系技术支持。

2.7 Alarm-02.305.0001.00007-check_tsar_nginx

检查tsar模块配置是否正确. 检查tsar是否采集nginx的相关数据。如果tsar未采集nginx的QPS，则无法获取应用的QPS情况。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	oss-ocm-server	oss-ocm-server.TEngine#

可能原因

没有进行相关配置。

影响范围

影响后续调查问题。

处理方法

开启 tsar 的 nginx 采集功能。

1. 将*/etc/tsar/tsar.conf*文件中**mod_nginx off** 修改为**mod_nginx on**。
2. 在**output_studio_mod**的末尾加上**mod_nginx**。
3. 对 nginx 进行设置，添加如下配置至默认主机：

```
location /nginx_status {
stub_status on;
access_log off;
allow 127.0.0.1;
deny all;
}
```

4. 重新载入，设置生效后等候片刻。

   tsar 的采集间隔是5分钟，修改crontab文件，即可在 tsar 输出中看到 nginx 的相关数据，对于 tengine 的设置与此相同。
5. 如果采集不正常，检查http://localhost/nginx_status。

执行tsar --nginx 命令，查看是否有正确内容输出。

2.8 Alarm-02.305.0001.00008-check_toa_module

检测toa模块是否加载到内存，toa模块是为了让后端的realserver能够看到真实的clientip而不是lvs的dip。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	oss-ocm-server	oss-ocm-server.TEngine#

可能原因

没有进行相关配置。

影响范围

会影响后续调查问题。

处理方法

安装toa模块 slb toa：注意要跟内核版本匹配，下面的例子是1089内核。

sudo yum install slb-vtoa-ali1089 -b current -y

sudo /sbin/modprobe slb_vtoa

2.9 Alarm-02.305.0001.00009-check_kernel_param

检查内核参数是否符合需求，这些内核参数是OSS在运行过程中的一些经验积累。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	oss-ocm-server	oss-ocm-server.TEngine#

可能原因

没有进行相关配置。

影响范围

OCM不能服务可能会影响正常的OSS读写。

处理方法

执行如下脚本：

```
oss_init_kernel_param.sh
alios7_judge(){
local alios7_flag=false
release_output=`cat /etc/redhat-release 2>/dev/null`
echo "${release_output}"|grep -i "release 7" 1>/dev/null 2>&1 && alios7_flag=true
echo ${alios7_flag}
}
alios7_flag=`alios7_judge`
if [ "x${alios7_flag}" == "xtrue" ];then
echo "this is a alios7 machine,no need to init_kernel_param"
exit 0
fi
uuid=`cat /proc/sys/kernel/random/uuid`
date_str=`date +%Y%m%d%H%M`
working_dir="/tmp/kuorong_tmp/${date_str}/${uuid}"
mkdir -p ${working_dir}
echo "conf copy working dir is ${working_dir}"
#随机端口范围指定原因参考：http://wiki.aliyun-inc.com/projects/apsara/wiki/ApsaraPort ，所以最小从32768开始。
cat << EOF > ${working_dir}/kernel_parameter_output
#ADD BY OSS_INIT_KERNEL_PARAM BEGIN
vm.max_map_count = 8388608
net.ipv4.tcp_rmem = 4096 87380 4194304
net.ipv4.tcp_wmem = 4096 16384 4194304
net.core.wmem_default = 8388608
net.core.rmem_default = 8388608
net.core.rmem_max = 16777216
net.core.wmem_max = 16777216
net.core.netdev_max_backlog = 204800
net.core.somaxconn = 204800
net.ipv4.tcp_max_orphans = 3276800
net.ipv4.tcp_max_syn_backlog = 204800
net.ipv4.tcp_tw_recycle = 0
net.ipv4.tcp_tw_reuse = 1
net.ipv4.tcp_tw_timeout = 15
net.ipv4.tcp_fin_timeout = 15
net.ipv4.ip_local_port_range = 32768 61000
net.ipv4.tcp_syncookies = 0
#ADD BY OSS_INIT_KERNEL_PARAM END
EOF
cat /etc/sysctl.conf |grep "ADD BY OSS_INIT_KERNEL_PARAM" 2>/dev/null
init_kernel_param_flag=$?
if [ "x${init_kernel_param_flag}" == "x0" ];then
 #随机端口范围指定原因参考：http://wiki.aliyun-inc.com/projects/apsara/wiki/ApsaraPort ，所以最小从32768开始
cat /etc/sysctl.conf |grep "net.ipv4.ip_local_port_range = 32768 61000" > /dev/null
port_range_fix_flag=$?
 if [ "x${port_range_fix_flag}" != "x0" ];then
 sudo sed -i 's@net.ipv4.ip_local_port_range = .*@net.ipv4.ip_local_port_range = 32768 61000@' /etc/sysctl.conf
 sudo /sbin/sysctl -p
 echo "change port range to 32768 61000 done"
 fi
cat /etc/sysctl.conf |grep "net.ipv4.tcp_tw_timeout = 15"
tw_timeout_flag=$?
 if [ "x${tw_timeout_flag}" != "x0" ];then
```

```
sudo sed -i '/net.ipv4.tcp_tw_reuse = 1/a\net.ipv4.tcp_tw_timeout = 15' /etc/sysctl.conf > /dev/
null
sudo /sbin/sysctl -p
echo "add tw timeout to 15 done"
fi
exit 0
fi
sudo bash -c "sudo cat ${working_dir}/kernel_parameter_output >> /etc/sysctl.conf" || { echo "
add kernel param failed";exit 12; }
echo "init kernel param..."
sudo /sbin/sysctl -p #默认参数中有net.nf_conntrack_max会导致sysctl -p报错，所以我们不对其
做容错处理
```

2.10 Alarm-01.305.0001.00010-OCM_ACCESSLOG_WEBSERVER

OCM前端机使用的模板，目前只有对于5xx的监控，可以根据ocm需求定制。60秒采集一次。检查前端机的5XX错误比例，${code_5xxRa}>2&&${code_5xx}>30 。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	oss-ocm-server	oss-ocm-server.TEngine#

可能原因

- 请求失败。
- 压力过大。
- 数据库异常。

处理方法

1. 找到access log，一般在/apsara/apache/logs/，假如今天是20170301，则为 access_log.20170301。
2. 执行grep InternalError access_log.20170301命令，找到32位的RequestID。
3. 通过request id收集相关日志发送给技术支持。

2.11 Alarm-01.305.0002.00001-OSS_ACCESSLOG_WEBSERVER_ALL

OCM前端机使用的模板，目前只有对于5xx的监控，可以根据ocm需求定制。60秒采集一次。检查前端机的5XX错误比例，${code_5xxRa}>2&&${code_5xx}>30 。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	oss-server	oss-server.TEngine#

可能原因

- 请求失败
- 压力过大
- 数据库异常

处理方法

1. 找到access log，一般在*/apsara/apache/logs/*，假如今天是20170301，access_log.20170301。
2. 执行grep InternalError access_log.20170301命令，找到32位的RequestId。
3. 通过request id收集相关日志发送给技术支持。

2.12 Alarm-02.305.0002.00003-oss_check_net_error_drop

当网卡有丢包的时候，会产生该告警，当不丢包时会恢复。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P4	oss-server	oss-server.TEngine#

可能原因

- 网卡不好。
- 网络有问题。

影响范围

影响正常的OSS读写。

处理方法

需要配合OSS整体服务查看，如果OSS服务正常不影响，OSS的正确请求可暂时不处理。

检查网卡流量是否超过上限。

- 如果超过上限，表示正常，可能需要扩容，以减少网络的流量。
- 如果没有超过，联系硬件工程师，查看网卡是否正常。

2.13 Alarm-02.305.0002.00004-check_tengine_ssl_cert_expire_stat

检查tengine的ssl证书文件是否快要过期，证书有效期小于30天则报警。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	oss-server	oss-server.TEngine#

可能原因

有证书，且证书过期。

处理方法

- 申请新的证书。
- 替换证书。

专有云一般情况下都没有证书。

2.14 Alarm-02.305.0002.00005-check_2ethstatus_oss

检查网卡是否工作正常，使用自己的脚本可以灵活指定需要的网络监测规则，比如10000M网卡，包括是否在正常设置的speed，是否双工。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	oss-server	oss-server.TEngine#

可能原因

- 网卡坏了。
- 配置不对。

处理方法

联系驻场工程师替换网卡。

2.15 Alarm-02.305.0002.00006-check_nginx_process

当nginx进程重启的时候，会产生该告警。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1	oss-server	oss-server.TEngine#

可能原因

- 配置不对，导致重启。
- 程序bug，导致重启。
- 升级重启进程，忘记关报警。

影响范围

影响正常的OSS读写。

处理方法

1. 查看*/apsara/apache/logs*路径下access log日志是否正常。
2. 执行curl http://IP地址/systermoperation/checkossstatus -i 命令，查看OSS服务是否恢复，是否返回200。
   - 如果恢复，记录进程重启时间，如果不影响服务，可以暂时不用处理。
   - 如果没有恢复，服务不可用，联系技术支持。

2.16 Alarm-02.305.0002.00007-check_tsar_nginx

检查tsar模块配置是否正确. 检查tsar是否采集nginx的相关数据。如果tsar未采集nginx的QPS，这样一方面无法获取应用的QPS情况。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	oss-server	oss-server.TEngine#

可能原因

没有进行相关配置。

影响范围

影响后续调查问题。

处理方法

开启 tsar 的 nginx 采集功能。

1. 将*/etc/tsar/tsar.conf*文件中**mod_nginx off**修改为**mod_nginx on**。
2. 在**output_studio_mod**的末尾加上**mod_nginx**。
3. 对 nginx 进行设置，添加如下配置至默认主机：

```
location /nginx_status {
stub_status on;
access_log off;
allow 127.0.0.1;
deny all;
}
```

4. 重新载入设置生效后，等候片刻，即可在 tsar 输出中看到 nginx 的相关数据，对于 tengine 的设置与此相同。

   tsar 的采集间隔是5分钟，可以在crontab中修改。
5. 如果采集不正常，检查 http://localhost/nginx_status。
6. 执行tsar --nginx命令，查看是否有正确内容输出。

2.17 Alarm-02.305.0002.00008-check_toa_module

检测toa模块是否加载到内存，toa模块是为了让后端的realserver能够看到真实的clientip而不是lvs的dip。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	oss-server	oss-server.TEngine#

可能原因

没有进行相关配置。

影响范围

影响后续调查问题。

处理方法

安装toa模块&& slb toa：注意要跟内核版本匹配，下面的例子是1089内核。

sudo yum install slb-vtoa-ali1089 -b current -y

sudo /sbin/modprobe slb_vtoa

2.18 Alarm-02.305.0002.00009-check_kernel_param

检查内核参数是否符合需求，这些内核参数是OSS在运行过程中的一些经验积累。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	oss-server	oss-server.TEngine#

可能原因

没有进行相关配置。

影响范围

大压力下会影响正常的OSS读写。

处理方法

执行如下脚本：

```
oss_init_kernel_param.sh
alios7_judge(){
local alios7_flag=false
release_output=`cat /etc/redhat-release 2>/dev/null`
echo "${release_output}"|grep -i "release 7" 1>/dev/null 2>&1 && alios7_flag=true
echo ${alios7_flag}
}
alios7_flag=`alios7_judge`
if [ "x${alios7_flag}" == "xtrue" ];then
echo "this is a alios7 machine,no need to init_kernel_param"
exit 0
fi
uuid=`cat /proc/sys/kernel/random/uuid`
date_str=`date +%Y%m%d%H%M`
working_dir="/tmp/kuorong_tmp/${date_str}/${uuid}"
mkdir -p ${working_dir}
echo "conf copy working dir is ${working_dir}"
#随机端口范围指定原因参考：http://wiki.aliyun-inc.com/projects/apsara/wiki/ApsaraPort ，所以最小从32768开始
```

```
cat << EOF > ${working_dir}/kernel_parameter_output
#ADD BY OSS_INIT_KERNEL_PARAM BEGIN
vm.max_map_count = 8388608
net.ipv4.tcp_rmem = 4096 87380 4194304
net.ipv4.tcp_wmem = 4096 16384 4194304
net.core.wmem_default = 8388608
net.core.rmem_default = 8388608
net.core.rmem_max = 16777216
net.core.wmem_max = 16777216
net.core.netdev_max_backlog = 204800
net.core.somaxconn = 204800
net.ipv4.tcp_max_orphans = 3276800
net.ipv4.tcp_max_syn_backlog = 204800
net.ipv4.tcp_tw_recycle = 0
net.ipv4.tcp_tw_reuse = 1
net.ipv4.tcp_tw_timeout = 15
net.ipv4.tcp_fin_timeout = 15
net.ipv4.ip_local_port_range = 32768 61000
net.ipv4.tcp_syncookies = 0
#ADD BY OSS_INIT_KERNEL_PARAM END
EOF
cat /etc/sysctl.conf |grep "ADD BY OSS_INIT_KERNEL_PARAM" 2>/dev/null
init_kernel_param_flag=$?
if [ "x${init_kernel_param_flag}" == "x0" ];then
#随机端口范围指定原因参考：http://wiki.aliyun-inc.com/projects/apsara/wiki/ApsaraPort ，所以最小从32768开始
cat /etc/sysctl.conf |grep "net.ipv4.ip_local_port_range = 32768 61000" > /dev/null
port_range_fix_flag=$?
if [ "x${port_range_fix_flag}" != "x0" ];then
sudo sed -i 's@net.ipv4.ip_local_port_range = .*@net.ipv4.ip_local_port_range = 32768 61000@' /etc/sysctl.conf
sudo /sbin/sysctl -p
echo "change port range to 32768 61000 done"
fi
cat /etc/sysctl.conf |grep "net.ipv4.tcp_tw_timeout = 15"
tw_timeout_flag=$?
if [ "x${tw_timeout_flag}" != "x0" ];then
sudo sed -i '/net.ipv4.tcp_tw_reuse = 1/a\net.ipv4.tcp_tw_timeout = 15' /etc/sysctl.conf > /dev/null
sudo /sbin/sysctl -p
echo "add tw timeout to 15 done"
fi
exit 0
fi
sudo bash -c "sudo cat ${working_dir}/kernel_parameter_output >> /etc/sysctl.conf" || { echo "add kernel param failed";exit 12; }
echo "init kernel param..."
```

```
sudo /sbin/sysctl -p #默认参数中有net.nf_conntrack_max会导致sysctl -p报错，所以我们不对其做容错处理
```

2.19 Alarm-01.305.0002.00010-check_ossserver_openfilelimit_all

OSS前端机使用的监控 ，检查oss_server已经打开的fd个数和最大限制的limit个数。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值报警	P2	oss-server	oss-server.OssServer #

可能原因

没有进行相关配置。

影响范围

大压力下会影响正常的OSS读写。

处理方法

- 当fd个数大于20000时，需要重启oss_server进程。
- 当最大可以打开的fd个数小于1024时，需要查看启动的环境变量设置的fd最大是多少。

  工作时间不发送短信，非工作时间发送短信。

  - warning：oss_server fd限制为1024,需要修改bash环境变量。
  - critical：oss_server已经打开的fd>20000,需要重启进程。

2.20 Alarm-02.305.0002.00011-check_oss_server_process_restart_fix

当oss_server进程重启的时候，会产生该告警。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1	oss-server	oss-server.OssServer #

可能原因

- 资源不够。
- 配置不对，导致重启。
- 程序bug，导致重启。
- 升级重启进程，忘记关报警。

影响范围

影响正常的OSS读写。

处理方法

1. 执行ps auxf|grep oss_server|grep -v grep命令，查看日志，找到oss_server启动的进程所在的目录。
2. 进入目录，查看*apsara_log_conf.json*文件，获取日志打印的位置。
3. 收集相关日志和配置文件ocm_conf.json。
4. 收集本机的系统日志dmsg。
5. 联系技术支持。

2.21 Alarm-02.305.0002.00012-check_ossserver_mem

采集指定进程名的虚机内存，物理内存，当前状态大于45%报警。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	oss-server	oss-server.OssServer #

可能原因

内存没有控制好，流量过大。

影响范围

影响正常的OSS读写。

处理方法

收集配置文件，查看是否设置了*"DataCacheCapacity":"8589934592"*。

- 如果已经设置，查看是否设置为本机内存的30%以下。

如果是，可能发生了内存泄露，联系技术支持，收集相关信息后，重启进程。

- 如果没有设置，设置为本机内存30%以下，重启进程。

2.22 Alarm-02.305.0002.00013-check_nginx_port

当oss所在机器的80端口无法连接时，会产生该告警. 当80端口可以连接时，该告警会自动清除。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1	oss-server	oss-server.OssServer #

可能原因

- 机器损坏。
- 进程未启动。

影响范围

- 如果所有OSS都不能连接，OSS就不能正常工作。
- 如果还有OSS能工作，可能不会影响OSS正常服务。

处理方法

1. 执行curl IP命令，检查是否有响应。
   - 如果有响应，表示误报。
   - 如果没响应，请跳转至2。
2. 执行ps auxf|grep -e "tengine\|nginx"|grep -v grep命令，检查是否有nginx相关的进程。
   - 如果有，收集信息联系技术支持。
   - 如果没有，查看天基没有启动tengine或者nginx进程的原因，查看*/apsara/apache/logs/*目录下是否有当天的日志，收集信息联系技术支持。

2.23 Alarm-02.305.0002.00014-working_online_me_alarm

检查机器me不正常的时候，机器是否是working online，或者机器是否是天基API有问题。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1	oss-server	oss-server.OssServer #

可能原因

没有安装me或者是天基的基础组件。

影响范围

影响后续调查问题。

处理方法

联系技术支持。

2.24 Alarm-02.305.0003.00001-check_quota_client

OSS产品系列的quota check，check quota 同步情况，没有同步会报警。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1	oss-server	oss-server.OssServer #

可能原因

- 没有日志。
- 日志所在磁盘损坏。

影响范围

影响计量数据。

处理方法

1. 执行df -lh命令，查看access log所在磁盘是否满了。
   - 如果满了，清除日志，保留空间。

日志文件一般在*/apsara/apache/logs*路径下。

- 如果没满，请跳转至*2*。

2. 在磁盘上创建一个临时文件，查看是否能够创建成功。
   - 如果能，确认磁盘没有问题。

     收集quota agent日志所在目录 logs下的所有日志，联系技术支持工程师。
   - 如果不能，磁盘坏了，联系硬件工程师，维修磁盘。

2.25 Alarm-02.305.0003.00002-_quota_agent_process_fix

当quota agent进程重启的时候，产生该告警。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	oss-server	oss-server.OssServer #

可能原因

- 程序bug，导致重启。
- 升级重启进程，忘记关报警。

影响范围

对OSS应用几乎无影响，可能会影响计量数据。

处理方法

1. 执行ps auxf|grep quota_agent|grep -v grep命令，查看日志，找到启动的进程所在的目录。
2. 进入目录，在*apsara_log_conf.json*文件中，查看日志打印的位置。
3. 收集相关日志和配置文件。
4. 收集本机的系统日志dmsg。
5. 联系技术支持。

2.26 Alarm-02.305.0003.00003-check_quota_agent_mem

采集指定进程名的虚机内存和 物理内存，当前状态大于10GB报警。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	oss-server	oss-server.OssServer#

可能原因

没有控制好内存。

影响范围

对OSS应用几乎无影响，可能会影响计量数据。

处理方法

1. 执行ps auxf|grep quota_agent|grep -v grep命令，查看日志，找到启动进程所在的目录。
2. 进入目录，查看*apsara_log_conf.json*文件下日志打印的位置。
3. 收集相关日志和配置文件。
4. 收集本机的系统日志dmsg。
5. 联系技术支持。

2.27 Alarm-02.305.0004.00001-check_quota_data_to_sls

检查quota数据推送到云监控是否正常，不正常就报警。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	oss-quota-master	oss-quota-master.QuotaMaster#

可能原因

连接到SLS网络不通。

影响范围

对OSS应用几乎无影响，可能会影响计量数据。

处理方法

1. 执行ps auxf|grep quotamaster_main| grep -v grep命令，查看日志，找到启动进程所在的目录。
2. 进入目录，在*apsara_log_conf.json*文件中查看日志打印的位置。
3. 收集相关日志和配置文件。
4. 收集本机的系统日志dmsg。
5. 联系技术支持。

2.28 Alarm-02.305.0004.00002-check_oss_quota_master

当quotamaster_main进程重启的时候，会产生该告警。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	oss-quota-master	oss-quota-master. QuotaMaster#

可能原因

- 程序bug，导致重启。
- 升级重启了进程，忘记关报警。

影响范围

对OSS应用几乎无影响，可能会影响计量数据。

处理方法

1. 执行ps auxf|grep quotamaster_main| grep -v grep命令，查看日志，找到启动进程所在的目录。
2. 进入目录，在*apsara_log_conf.json*文件下，查看日志打印的位置。
3. 收集相关日志和配置文件。
4. 收集本机的系统日志dmsg。
5. 联系技术支持。

2.29 Alarm-02.305.0004.00003-check_oss_quota_master_syncpoint

oss产品系列的quota check，check quota 同步情况，没有同步会报警。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1	oss-quota-master	oss-quota-master.QuotaMaster#

可能原因

- quota service的数据没有收集成功。
- 推送到OMS失败，可能是OMS有问题，也有可能是到OMS的网络有问题。

影响范围

影响计量数据。

处理方法

1. 执行df -lh命令，查看access log所在磁盘是否满了。
   - 如果满了，清除日志，保留空间。

     日志文件一般在*/apsara/apache/logs*路径下。
   - 如果没满，请跳转至*2*。
2. 在磁盘上创建一个临时文件，查看是否能够创建成功。
   - 如果能，确认磁盘没有问题。

     收集quota agent日志所在目录 logs下的所有日志，联系技术支持工程师。
   - 如果不能，磁盘坏了，联系硬件工程师，维修磁盘。

3 表格存储Table Store

3.1 Alarm-02.310.0010.00020-表格存储sqlonline_master进程发生重启

当sqlonline_master发生重启时，产生该告警。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	-	表格存储	TableStore

可能原因

- sqlonline_master所在机器有硬件故障，导致重启。
- 触发sqlonline_master未知bug，导致重启。

影响范围

- 在sqlonline_master重启期间删建表以及修改、获取表meta操作会失败。
- 在sqlonline_master重启期间，数据读写可能出现失败。

处理方法

1. 登录到sqlonline_master所在机器：在OTS ag上执行r wl命令，找到sys/sqlonline-OTS对应replyAddress，登录到对应的机器。
2. 执行cd /apsara/tubo/TempRoot/sys/sqlonline-OTS/sqlonline/命令，打开sqlonline_master所在的目录。
3. 收集该目录下的所有日志，联系技术支持。

3.2 Alarm-02.310.0100.10101-表格存储前端机出现5XX报警

当表格存储前端机检测，检测周期为1分钟，到5xx错误请求数超过50时，产生该告警。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	-	表格存储	TableStore

可能原因

常见的5xx错误原因：

- 分区在自动分裂过程中会出现短时间服务不可用。
- 用户瞬时压力过大超过分区性能极限或机器性能极限时，出现服务忙错误。

影响范围

用户访问出错，如果持续发生该告警，说明系统出现异常，需要调查原因。

处理方法

1. 根据报警信息，登录到对应的机器上。
2. 执行cd /apsara/ots_server/logs命令，打开ots_server日志所在目录。
3. 收集该目录下的所有日志，联系技术支持。

3.3 Alarm-02.310.0004.00001-表格存储前端机http连接数超限

当表格存储前端机检测到http连接数超过10000时，产生该告警。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	-	表格存储	TableStore

可能原因

ots_server机器压力过大。

影响范围

ots_server不稳定，访问会出现超时、出错等问题。

处理方法

对ots_server机器组进行扩容。

3.4 Alarm-02.310.0010.00001-表格存储ots_server进程发生重启

当ots_server发生重启时，产生该告警。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	-	表格存储	TableStore

可能原因

ots_server由于未知bug导致crash。

影响范围

访问该前端机的部分请求失败。

处理方法

1. 根据报警信息登录到对应的机器。
2. 执行cd /apsara/ots_server/logs命令，打开ots_server日志所在目录。
3. 收集该目录下的所有日志，联系技术支持。

3.5 Alarm-02.310.0010.00010-表格存储sqlonline_worker进程发生重启

当sqlonline_worker发生重启时，产生该告警。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	-	表格存储	TableStore

可能原因

- sqlonline_worker所在机器有硬件故障，导致重启。
- 触发sqlonline_worker未知bug，导致重启。

影响范围

访问到该sqlonline_worker上分区的部分请求失败。

处理方法

1. 根据报警信息登录到对应机器上。
2. 执行ps aux | grep sqlonline_worker命令，查看sqlonline_worker进程。
3. 执行如下命令，打开sqlonline_worker所在的目录。

   cd /apsara/tubo/TempRoot/sys/sqlonline-OTS/SqlWorkerRole@*/sqlonline/
4. 收集该目录下的所有日志，联系技术支持。

3.6 Alarm-02.310.0020.00020-sqlonline_master coredump

当sqlonline_master发生coredump时，产生该告警。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	-	表格存储	TableStore

可能原因

sqlonline_master由于未知原因导致coredump。

影响范围

- 在sqlonline_master重启期间，删建表以及修改、获取表meta操作会失败。
- 在sqlonline_master重启期间，数据读写可能出现失败。

处理方法

1. 登录到sqlonline_master所在机器。
2. 在OTS ag上执行r wl命令，找到*sys/sqlonline-OTS*对应replyAddress，登录到对应的机器。
3. 执行如下命令，打开sqlonline_master所在的目录，找到sqlonline_master binary。

   cd /apsara/tubo/TempRoot/sys/sqlonline-OTS/sqlonline/
4. 找到sqlonline_master生成的core文件。
5. 请收集sqlonline_master binary和对应的core文件，联系技术支持。

3.7 Alarm-02.310.0020.00010-sqlonline_worker coredump

当sqlonline_worker发生coredump时，产生该告警。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	-	表格存储	TableStore

可能原因

sqlonline_worker由于未知原因导致coredump。

影响范围

访问到该sqlonline_worker上分区的部分请求失败。

处理方法

1. 根据报警信息登录到到对应机器上。
2. 执行ps aux | grep sqlonline_worker命令，查看sqlonline_worker进程。
3. 执行如下命令，打开sqlonline_worker所在的目录并找到sqlonline_worker binary。

   cd /apsara/tubo/TempRoot/sys/sqlonline-OTS/SqlWorkerRole@*/sqlonline/
4. 找到sqlonline_worker生成的core文件。
5. 请收集sqlonline_worker binary和对应的core文件，联系技术支持。

3.8 Alarm-02.310.0010.00002-ots_tengine进程发生重启

当ots_tengine进程发生重启时，发生该告警。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	-	表格存储	TableStore

可能原因

ots_tengine由于未知bug导致crash。

影响范围

访问该前端机的部分请求失败。

处理方法

1. 根据报警信息登录到对应的机器上。
2. 执行cd /apsara/ots_tengine/logs命令，打开ots_tengine日志所在目录。
3. 收集该目录下的所有日志，联系技术支持。

3.9 Alarm-02.310.0010.00004-表格存储replication_server进程发生重启

当replication_server发生重启时，产生该告警。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	-	表格存储	TableStore

可能原因

- replication_server所在机器有硬件故障，导致重启。
- 触发replication_server未知bug，导致重启。

影响范围

数据同步到备集群可能出现不及时。

处理方法

1. 根据报警信息登录到对应的机器上。
2. 执行如下命令，打开replication_server日志所在目录。

   cd /apsara/sqlonline_replication_server/logs
3. 收集该目录下的所有日志，联系技术支持。

3.10 Alarm-02.310.0100.10000-表格存储出现warning日志

当表格存储出现warning日志时，发生该告警。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	-	表格存储	TableStore

可能原因

- worker_alert_log失败，sqlonline_worker出现alert日志。
- check_replicate_status失败，双集群数据同步速度过慢。
- ots_check_cgroup失败，机器或相关进程cpu使用超限。
- check_ots_server_health失败，有ots_server机器无法服务。
- ots_server_alert失败，ots_server出现alert日志。
- replication_server_alert_log失败，sqlonline_replication_server出现alert日志。
- check_sqlonline_master_process失败，sqlonline_master发生进程重启。
- master_alert_log失败，sqlonline_master出现alert日志。
- check_sqlservice失败，部分机器的sqlonline_worker进程没有启动。

影响范围

表格存储服务不稳定。

处理方法

- worker_alert_log失败，登录到对应机器，并打开sqlonline_worker目录，查看sqlonline_alert.LOG日志，联系技术支持。
- check_replicate_status失败，联系技术支持。
- ots_check_cgroup失败，联系技术支持。
- check_ots_server_health失败，登录到对应机器上，查看*/apsara/OTSAdmin/alert/warning/monitor_result*.log*文件，找到check_ots_server_health失败的机器并登录。

  查看ots_server、ots_tengine进程是否存在，*/var/www/html/ots_server.heartbeat*是否存在。
- ots_server_alert失败，登录到对应机器，并打开*/apsara/ots_server/logs*目录，查看sqlonline_alert.LOG日志，联系技术支持。
- replication_server_alert_log失败，登录到对应机器，并打开*/apsara/sqlonline_replication_server/logs*目录，查看sqlonline_alert.LOG日志，联系技术支持。
- check_sqlonline_master_process失败，登录到对应机器上，并打开sqlonline_master目录，查看sqlonline_alert.LOG日志，联系技术支持。
- master_alert_log失败，登录到对应机器上，并打开sqlonline_master目录，查看sqlonline_alert.LOG日志，联系技术支持。
- check_sqlservice失败，登录到对应机器上，查看*/apsara/OTSAdmin/alert/warning/monitor_result*.log*，找到check_sqlservice失败的日志，联系技术支持。

3.11 Alarm-02.310.0100.20000-表格存储出现critical日志

当表格存储出现critical日志时，发生该告警。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	-	表格存储	TableStore

可能原因

- check_cpls失败，partition load不起来。
- worker_alert_log失败，sqlonline_worker出现alert日志。
- check_replicate_status失败，双集群数据同步速度过慢。
- ots_check_cgroup失败，机器或相关进程cpu使用超限。
- check_ots_server_health失败，有ots_server机器无法服务。
- ots_server_alert失败，ots_server出现alert日志。
- replication_server_alert_log失败，sqlonline_replication_server出现alert日志。
- check_sqlonline_master_process失败，sqlonline_master发生进程重启。
- master_alert_log失败，sqlonline_master出现alert日志。
- check_sqlservice失败，部分机器的sqlonline_worker进程没有启动。

影响范围

表格存储服务出现严重异常，需要及时处理，否则服务受到严重影响。

处理方法

- check_cpls失败，在ots ag上执行`sql cpls`，并挑选一个partition登录到对应的worker上。

  打开sqlonline_worker所在目录，收集*sqlonline_alert.LOG**和*sqlonline_error.LOG**，联系技术支持。
- worker_alert_log失败，登录到对应机器，并打开sqlonline_worker目录，查看*sqlonline_alert.LOG*日志，联系技术支持。
- check_replicate_status失败，联系技术支持。
- ots_check_cgroup失败，联系技术支持。
- check_ots_server_health失败，登录到对应机器上，查看*/apsara/OTSAdmin/alert/warning/monitor_result*.log*日志，找到check_ots_server_health失败的机器并登录。

查看ots_server、ots_tengine进程是否存在，*/var/www/html/ots_server.heartbeat*是否存在。

- ots_server_alert失败，登录到对应机器，打开*/apsara/ots_server/logs*目录，查看*sqlonline_alert.LOG*日志，联系技术支持。
- replication_server_alert_log失败，登录到对应机器，打开*/apsara/sqlonline_replication_server/logs*目录，查看*sqlonline_alert.LOG*日志，联系技术支持。
- check_sqlonline_master_process失败，登录到对应机器上，打开*sqlonline_master*目录，查看*sqlonline_alert.LOG*日志，联系技术支持。
- master_alert_log失败，登录到对应机器上，打开*sqlonline_master*目录，查看*sqlonline_alert.LOG*日志，联系技术支持。
- check_sqlservice失败，登录到对应机器上，查看*/apsara/OTSAdmin/alert/warning/monitor_result*.log*日志，找到check_sqlservice失败的日志，联系技术支持。

3.12 Alarm-02.310.0001.00001-表格存储前端机cpu过高

当表格存储前端机所在cpu使用过高时，发生该告警。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警		表格存储	TableStore

可能原因

前端机所在机器压力过大。

影响范围

ots_server不稳定，长时间会出现访问超时、出错等问题。

处理方法

对ots_server机器组进行扩容。

3.13 Alarm-02.310.0001.00002-表格存储后端机cpu过高

当表格存储后端机所在cpu使用过高时，发生该告警。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	-	表格存储	TableStore

可能原因

后端机所在机器压力过大。

影响范围

sqlonline_worker服务不稳定，长时间会出现访问超时、出错等问题。

处理方法

查看所有ots后端机的CPU使用情况，如果所有机器都CPU使用率很高，需要对sqlonline_worker机器组进行扩容，如果不是请联系技术支持。

3.14 Alarm-02.310.0200.00001-PostCheck检查不通过

当对应的ServerRole运行异常时，发生该告警。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警		表格存储	TableStore

可能原因

ServerRole异常。

影响范围

TableStore无法正常工作。

处理方法

1. 通过天基查看对应的SR的PostCheck检查信息，如果信息中没有详细的错误信息，那么需要参见2。
2. 登录报错的机器，查看monitor的运行日志，日志路径为*/cloud/log/TableStore*/service-role#/{app}_monitor/*。

3. 基于日志判断修复方案或者联系技术支持。

3.15 Alarm-02.310.0200.00002-测试镜像运行不通过

当对应的Service运行异常时，发生该告警。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	-	表格存储	TableStore

可能原因

自身的ServerRole异常或者依赖的服务异常。

影响范围

TableStore无法正常工作。

处理方法

1. 查看错误报告。
2. 获取错误报告中的RequestID。
3. 登录OTS的大数据管家，域名一般是bigdata-ots.aliyun.com
4. 选择**监控中心** > **日志分析** > **请求日志搜索**，将RequestID填入搜索框中搜索。
5. 结果中有详细的错误信息，如果无法处理，请联系技术支持。

4 女娲（nvwa）

4.1 Alarm-02.005.0001.00001-check_nuwa_config

查看nuwa config。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P1)	nuwa	nuwa

可能原因

配置文件丢失。

影响范围

影响集群服务。

处理方法

联系nuwa开发排查问题。

4.2 Alarm-02.005.0003.00001-check_nuwa_election_event

这个脚本检查女娲后端zk是否发生了选举，以及zxid低32位是否快耗尽。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P3)	Nuwa	Nuwa系统中的NuwaZK serverrole

可能原因

nuwa zk发生了重新选举，后端zk的zxid低32位已经快接近耗尽的边缘。

影响范围

nuwa发生选举影响不大。

处理方法

登录到集群中看下发生选举原因即可，不需要做什么操作。

4.3 Alarm-02.005.0001.00002-check_nuwa_proxy_log

通过从集群ag上扫描女娲proxy的log，检查是否存在长时间没有log输出的情形。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P1)	Nuwa	Nuwa系统中的NuwaProxy serverrole

可能原因

该机器上的nuwa proxy停止工作了，磁盘是否只读了，或者其他什么原因。

影响范围

影响nuwa proxy进程。

处理方法

登录到集群中看这台机器出现了什么问题；

- 进程依赖的挂载磁盘可能只读了导致日志的缺失，需要重启机器。
- nuwa proxy进程有问题，需要重启进程进行检查。
- 如果是其他相关问题，需要特殊情况特殊处理。

4.4 Alarm-02.005.0002.00001-check_nuwa_zookeeper_log

通过从集群ag上扫描女娲zk的log，检查是否存在长时间没有log输出的情形，或者是依赖的磁盘变成只读模式。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P1)	Nuwa	Nuwa系统中的NuwaZK serverrole

可能原因

该机器上的nuwa ZK停止工作了，磁盘是否只读了，或者其他什么原因。

影响范围

影响nuwa zk进程。

处理方法

登录到集群中看这台机器出现了什么问题：

- 机器的磁盘可能只读了导致日志的缺失，这个时候应该需要重启机器。
- nuwa zk进程有问题，需要重启进程进行检查。
- 如果是其他相关问题，需要特殊情况特殊处理。

4.5 Alarm-02.005.0001.00003-check_nuwa_proxy_service

针对女娲的proxy，逐台Server检查服务是否可用。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P1)	Nuwa	Nuwa系统中的NuwaProxy serverrole

可能原因

该nuwa proxy有异常，可能是机器挂掉，有可能ssh不通，有可能nuwa proxy进程有问题。

影响范围

影响nuwa proxy进程。

处理方法

登录到集群中看这台机器出现了什么问题；

- 当机器ping不通或者ssh不通的话，需要重启机器，如果机器重启不成功，替换机器。
- 如果是其他相关问题，需要特殊情况特殊处理。

4.6 Alarm-02.005.0002.00002-check_nuwa_zk_service

针对女娲的ZK，逐台Server检查服务是否可用。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P1)	Nuwa	Nuwa系统中的NuwaZK serverrole

可能原因

该nuwa zk有异常，可能是机器挂掉，有可能ssh不通，有可能nuwa zk进程有问题。

影响范围

影响nuwa zk进程。

处理方法

登录到集群中看这台机器出现了什么问题；

- 当机器ping不通或者ssh不通的话，需要重启机器，如果机器重启不成功，就需要替换机器。
- 如果是其他相关问题，需要特殊情况特殊处理。

4.7 Alarm-01.005.0003.00002-check_nuwa_server_disk

针对女娲后端ZK的依赖/apsara盘，snapshot盘以及txnlog盘进行监控，并且针对proxy/zk的log输出情况进行监控。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P1)	Nuwa	Nuwa系统中的NuwaZK/NuwaProxy serverrole

可能原因

/apsara盘使用率超过85%，snapshot盘以及txnlog盘使用率超过70%，或者是proxy/zk的log超过5min没有任何输出。

影响范围

影响proxy和zk进程的磁盘读写，进而影响进程。

处理方法

登录到集群中看这台机器的磁盘与目录空间；

- 磁盘空间满了，需要清理。
- 目录空间满了，需要清理。
- 如果是其他相关问题，需要特殊情况特殊处理。

4.8 Alarm-02.005.0004.00001-check_nuwa_config_in_tianji

检查所有的机器的配置文件是否和tianji中的终态一致(适用于跨集群配置打通)。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P2)	Nuwa	Nuwa系统中的NuwaConfig serverrole

可能原因

跨集群打通配置遇到了问题，会造成跨集群访问的失败。

影响范围

影响nuwa跨集群访问的功能。

处理方法

需要调查tianji下打通集群为什么失败，失败的原因有很多，模板配置的出错，tianji api的访问失败等。

5 MiniLVS

5.1 Alarm-01.211.0001.00001-VIP库存

可分配 VIP 资源过少 (默认20) 时告警。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	miniLVS	minilvs.API#

可能原因

该环境VIP消耗过多。

影响范围

如果库存耗尽时影响新 VIP 申请，不影响在用 VIP。

处理方法

1. 评估业务是否有新增 VIP 需求。
   - 如果否，请调整阈值。
   - 如果是，请跳转至2.
2. 网络工程师分配新的 VIP 资源段，扩容到 minilvs。

5.2 Alarm-02.211.0002.00001-LVSNode KVM连通性

LVSNode KVM 连不通。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P2)	miniLVS	minilvs.LVSNode#

可能原因

KVM 异常。

影响范围

KVM 网络或者负载异常。

处理方法

在 minilvs 所在 ops 两台宿主机上: 执行`virsh list --all`，获取 kvm Name: minilvsNNN。

执行`virsh destroy $name; virsh start $name`命令重启。

5.3 Alarm-02.211.0001.00002-API 可用性

miniLVS 管控API 异常。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	紧急(P1)	miniLVS	minilvs.API#

可能原因

miniLVS 管控API 异常, 返回值不合预期.可能原因: DNS 异常、LVSNode 异常、DB 长时间连接失败。

影响范围

影响新 VIP 的生产，及已有 VIP 的变配，不影响已有 VIP 的流量转发。

处理方法

1. 执行`dig minilvs-api.${intranet-domain}`命令，确认解析正常。
2. 执行`ping minilvs-api.${intranet-domain}`命令，确认 VIP 可以ping 通。
3. 执行`curl minilvs-api.${intranet-domain}/slb/api?list_lvs_node`命令，确认是否返回正常。

   如果还是无法定位问题，请收集上述处理过程的输出，联系 minilvs 开发排查。

5.4 Alarm-02.211.0002.00001-LVSNode KVM连通性

LVSNode KVM 连不通。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P2)	miniLVS	minilvs.LVSNode#

可能原因

KVM 异常。

影响范围

KVM 网络或者负载异常。

处理方法

在 minilvs 所在 ops 两台宿主机上: 执行virsh list --all，获取 kvm Name: minilvsNNN。

执行virsh destroy $name; virsh start $name命令重启。

6 MiniRDS

6.1 Alarm-02.301.0001.0001-check slave(sql thread down)

当slave sql 线程 down。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P2)	minirds	minirds.db

可能原因

主从同步失败。

影响范围

如果不及时处理导致 master slave 数据不一致。

处理方法

执行curl "http://${api}/action=recover&url=${url}&port=${port}"命令，重搭备库。

6.2 Alarm-02.301.0001.0002-check alive

当mysql实例 down。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	紧急(P1)	minirds	minirds.db

可能原因

mysql实例故障。

影响范围

不及时处理可能服务不能保证。

处理方法

执行curl "http://${api}/action=recover&url=${url}&port=${port}"命令重搭。

6.3 Alarm-02.301.0001.0003-chk_thread_connected above 8000

mysql 连接数过高。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	minirds	minirds.db

可能原因

mysql 连接数过高。

影响范围

导致后续服务中断。

处理方法

执行kill process kill -9 ${mysqld_pid}命令。

6.4 Alarm-02.301.0001.0004-chk_slavelag behind 36000

slave 。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	minirds	minirds.db

可能原因

网络原因。

影响范围

可能导致数据不一致。

处理方法

执行tcpdump –i any命令，抓包检验，是否网络问题。

6.5 Alarm-02.301.0001.0005-chk_mysql_aborted_conn above 10

Aborted 连接数过多。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	minirds	minirds.db

可能原因

网络原因。

影响范围

服务不稳定。

处理方法

执行tcpdump –i any命令，抓包检验，是否网络问题。

6.6 Alarm-02.301.0001.0006-chk_slaveio

当slave io 线程 down。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P2)	minirds	minirds.db

可能原因

主从同步失败。

影响范围

如果不及时处理导致 master slave 数据不一致。

处理方法

执行curl "http://${api}/action=recover&url=${url}&port=${port}"命令，重搭备库。

7 云控制台

7.1 Alarm-01.105.0008.0001-内存使用率过高

内存使用率过高。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	ClusterOwner	软件

可能原因

服务接收metaq数据过多。

影响范围

DTCenter服务可能出现异常。

处理方法

1. 重启容器中的dtcenter服务，将*/alidata/www/logs/tomcat7/*，*/opt/tomcat7/logs/*的日志文件保存下来，发给数梦的同学。
2. 执行free，查看内存是否使用率仍然很高，如果很高需要做内存扩容。

7.2 Alarm-01.105.0008.0002-CPU负载过高

CPU负载过高。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	ClusterOwner	软件

可能原因

业务过于繁忙。

影响范围

DTCenter服务可能出现异常。

处理方法

重启容器中的dtcenter服务，将*/alidata/www/logs/tomcat7/*， */opt/tomcat7/logs/*的日志文件保存下来。

7.3 Alarm-01.105.0008.0003-网络流量过高

网络流量过高。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	ClusterOwner	软件

可能原因

可能受到攻击。

影响范围

DTCenter服务可能出现异常。

处理方法

查看是否受到攻击。

7.4 Alarm-01.105.0008.0004-CPU使用率过高

CPU使用率过高。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	ClusterOwner	软件

可能原因

业务过于繁忙。

影响范围

DTCenter服务可能出现异常。

处理方法

重启容器中的DTCenter服务，将*/alidata/www/logs/tomcat7/*， */opt/tomcat7/logs/*的日志文件保存下来。

7.5 Alarm-01.105.0008.0005-主机Ping响应超时

主机Ping响应超时。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1	ClusterOwner	软件

可能原因

业务过于繁忙。

影响范围

DTCenter服务可能出现异常。

处理方法

1. 重启容器中的dtcenter服务，将*/alidata/www/logs/tomcat7/*， */opt/tomcat7/logs/*的日志文件保存下来。
2. 执行df，查看磁盘空间是否满了，如果满了，查找*1*日志文件以及diamond目录下的日志文件。

7.6 Alarm-01.105.0008.0006-主机Ping无响应

主机Ping无响应。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1	ClusterOwner	软件

可能原因

业务过于繁忙。

影响范围

DTCenter服务可能出现异常。

处理方法

1. 重启容器中的dtcenter服务，将*/alidata/www/logs/tomcat7/*， */opt/tomcat7/logs/*的日志文件保存下来。
2. 执行df，查看磁盘空间是否满了，如果满了，查找*1*日志文件以及diamond目录下的日志文件。

7.7 Alarm-01.105.0008.0007-vip探测失败

vip探测失败。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1	ClusterOwner	软件

可能原因

业务过于繁忙。

影响范围

DTCenter服务可能出现异常。

处理方法

1. 重启容器中的dtcenter服务，将*/alidata/www/logs/tomcat7/*， */opt/tomcat7/logs/*的日志文件保存下来。
2. 执行df，查看磁盘空间是否满了，如果满了，查找*1*日志文件以及diamond目录下的日志文件。

7.8 Alarm-01.105.0008.0008-HTTP探测无响应

HTTP探测无响应。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1	ClusterOwner	软件

可能原因

可能服务启动不正常。

影响范围

DTCenter服务可能出现异常。

处理方法

1. 执行free，查看内存是否使用率很高，如果很高需要做内存扩容。
2. 执行df，查看磁盘空间是否满了，如果满了，查找日志文件以及diamond目录下的日志文件并删除，恢复磁盘空间。

3. 执行netstat -nao；ps aux| grep tomcat，看是否在监听18080端口，以及tomcat进程是否存在。如不存在需要执行4
4. 重启容器中的tomcat服务，将*/alidata/www/logs/tomcat7/*， */opt/tomcat7/logs/*的日志文件保存下来。

7.9 Alarm-01.105.0008.0009-页面探测无响应

页面探测无响应。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1	ClusterOwner	软件

可能原因

可能服务启动不正常。

影响范围

DTCenter服务可能出现异常。

处理方法

1. 执行free，查看内存是否使用率很高，如果很高需要做内存扩容。
2. 执行df，查看磁盘空间是否满了，如果满了，查找并1下面的日志文件以及diamond目录下的日志文件并删除，恢复磁盘空间。
3. 执行netstat -nao；ps aux| grep tomcat，看是否在监听18080端口，以及tomcat进程是否存在。如不存在需要执行4
4. 重启容器中的tomcat服务，将*/alidata/www/logs/tomcat7/*， */opt/tomcat7/logs/*的日志文件保存下来。

7.10 Alarm-01.105.0008.0010-Url检查失败

Url检查失败。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1	ClusterOwner	软件

可能原因

可能服务启动不正常。

影响范围

DTCenter服务可能出现异常。

处理方法

1. 执行free，查看内存是否使用率很高，如果很高需要做内存扩容。
2. 执行df，查看磁盘空间是否满了，如果满了，查找并1下面的日志文件以及diamond目录下的日志文件并删除，恢复磁盘空间。
3. 执行netstat -nao；ps aux| grep tomcat，看是否在监听18080端口，以及tomcat进程是否存在。如不存在需要执行4
4. 重启容器中的tomcat服务，将*/alidata/www/logs/tomcat7/*， */opt/tomcat7/logs/*的日志文件保存。

8 ODPS

8.1 Alarm-02.200.0001.00000-check_server_alive

服务器可能发生宕机。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P2)	单机	硬件

可能原因

服务器load、cpu等过高。

影响范围

该服务器不可用。

处理方法

联系驻场或者带外重启该服务器，重启时间大概5-60分钟，看能否ssh方式登陆，如果不能需要联系系统同学查看是硬件还是系统有问题。

8.2 Alarm-02.200.0001.00001-check_ssh

服务器可能发生宕机。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P2)	单机	硬件

可能原因

服务器load、cpu等过高，或者ssh服务不正常。

影响范围

服务器无法登录。

处理方法

联系驻场或者带外重启该服务器，重启时间大概5-60分钟，看能否ssh方式登录，如果不能需要联系系统同学查看是硬件还是系统有问题。

8.3 Alarm-01.200.0001.00002-check_disk_usage

磁盘空间满。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	单机	硬件

可能原因

没有及时清理数据或者将大量数据写入。

影响范围

可能会影响运行在该服务器上的进程存放临时数据。

处理方法

1. 登录该机器，执行df -h命令，查看哪个目录磁盘空间满。
2. cd 磁盘满的目录：执行sudo du -h . --max-depth=1或者ls -trlh查找大文件或者大目录，清理对应文件或者目录。

8.4 Alarm-02.200.0001.00003-check_eth_status

服务器网卡状态异常。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P2)	单机	硬件

可能原因

网卡或者对应交换机出现问题。

影响范围

服务器可能无法联网。

处理方法

1. 联系驻场查看服务器对应的交换机状态是否ok。
2. 如果交换机没有问题，查看对应服务器网口是否正常。

   如果正常，说明服务器网卡有问题，联系相关人员进行停机维修，不正常说明网口有问题，联系相关人员进行维修。

8.5 Alarm-02.000.0001.00000-check_pangu_master_switch

check pangu master是否发生切换。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P2)	pangu master	pangu

可能原因

心跳丢失。

影响范围

无影响。

处理方法

1. 登录集群ag。
2. puadmin gems查看盘古服务。
3. 根据报警master登录具体master查看进程启动时间和查看log(/apsara/pangu_master/log/pangu_master.LOG)并联系pangu开发排查问题。

8.6 Alarm-02.000.0001.00001-盘古不可读写

当维护工具检测到盘古不可读写的时候，产生该告警，检测周期为1分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1	盘古	盘古整体

可能原因

- 盘古没有足够的磁盘空间。

- 盘古chunkserver数目不足。
- 网络问题。

影响范围

盘古不可读写，影响上层所有往盘古写入数据的服务。

处理方法

1. 登录PanguTools机器，执行/apsara/deploy/puadmin lscs命令，查看处于NORMAL的chunkserver数以及DISK_OK的磁盘数目是否正常。
   - 如果正常，请跳转至2。
   - 如果盘古空间已满，请通知集群管理员。
2. 查看集群的网络状态，PanguTools所在机器与盘古Master机器之间的网络情况。

8.7 Alarm-01.000.0001.00002-盘古集群中temp file文件大小超过阈值

当维护工具检测到集群中temp file大小超过阈值的时候，产生该告警，检测周期为1小时。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古整体

可能原因

上层服务写入太大的temp file文件。

影响范围

temp file过大，会占用过多盘古的存储空间，影响其他服务使用盘古。

处理方法

登录PangtuTools所在的机器，执行/apsara/deploy/puadmin cs -tempfile -top 1命令，查找最大的temp file，找到temp file的写入者，跟写入者确认写入文件大小是否合理。

8.8 Alarm-01.000.0001.00003-盘古存在有0副本文件

维护工具检测到集群中存在0副本的文件时，产生该告警，检测周期为5分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	盘古	盘古整体

可能原因

同一时间大量机器或者磁盘损坏。

影响范围

影响数据安全性。

处理方法

1. 登录PanguTools所在机器，执行/apsara/deploy/puadmin lscs命令，查看状态非NORMAL的Chunkserver，并将机器或者进程重新启动。

8.9 Alarm-02.010.0001.00000-check_fuxi_master_hang

监控fuxi master运行状态。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P1)	fuxi master	fuxi master

可能原因

程序异常。

影响范围

影响集群作业调度运行。

处理方法

1. 登录集群，执行r al命令，确认fuxi master是否正常服务。
2. 联系fuxi开发，确认fuxi hang住的时间和影响。
3. 必要时刻可以kill掉主master(主master ip为：r primary fm)触发切换。

8.10 Alarm-01.000.0001.00004-盘古存在有1副本文件

维护工具检测到集群中存在1副本的文件时，产生该告警，检测周期为5分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古整体

可能原因

同一时间大量的机器或者磁盘损坏。

影响范围

影响数据安全性。

处理方法

调大集群的replication流量限制，使其尽快复制。

8.11 Alarm-02.000.0001.00005-check_pangu_file_replicate

查看文件副本数。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P1)	文件	文件

可能原因

文件副本所在机器宕机。

影响范围

影响对应文件副本数。

处理方法

1. 登录集群ag，查看当前集群是否有dis机器`puadmin lscs | grep DISCON`dis的chunkserver。
2. 如果有disconnect机器，通知驻场同学重启服务器以使副本恢复。
3. 如果还有问题，联系pangu开发进行排查。

8.12 Alarm-01.010.0001.00001-check_fuxi_job_num

集群作业数。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	集群	集群

可能原因

用户某一段时间提交作业过多。

影响范围

影响集群稳定，后续作业不会被调用。

处理方法

联系odps开发查找提交作业数过多的原因，找出作业并kill。

8.13 Alarm-02.010.0001.00002- odps_apsara_pm_ag-check_package_manager_alive

集群中的PackageManager进程是否存活。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	紧急(P1)	fuxi	PackageManager

可能原因

PackageManager进程无法启动，或者hang住。

影响范围

无法向fuxi上传package，fuxi无法正常拉起进程。

处理方法

登录PackageManager机器，查看PackageManager进程是否存在。

8.14 Alarm-02.010.0001.00003-check_fuxiservice_status

check fuxi master状元。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P2)	fuxi master	fuxi

可能原因

程序异常、心跳丢失。

影响范围

影响集群作业调度。

处理方法

1. 登录集群ag。
2. 执行r al命令，查看能否正常显示集群运行的作业。
3. 执行r primary fm命令，查看fuxi master ip。
4. 登录对应ip，查看master进程启动时间以及查看报警时间段log(/apsara/fuxi_master/fuxi_master.LOG)，将相关报错信息发给fuxi 开发人员。

8.15 Alarm-02.005.0001.00000-check_nuwa_zk

查看nuwa状态。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P1)	nuwa	nuwa

可能原因

心跳丢失、对应服务器状态不正常。

影响范围

影响集群稳定。

处理方法

1. 执行/apsara/deploy/nuwa_console --address=nuwa:///localcluster/ --console --admin=true命令，确定nuwa服务是否正常。
2. 登录报警机器确定nuwa进程是否存在。
3. 联系nuwa开发排查问题。

8.16 Alarm-02.010.0002.00000-check_package_manager

查看package manager状态。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P2)	package manager	package manager

可能原因

程序异常。

影响范围

影响集群包管理。

处理方法

1. 登录集群ag，执行/apsara/deploy/rpc_wrapper/rpc.sh pl 命令，确认是否能正常取到package list。
2. 如果没有出现package list，联系pacakge 开发排查。

8.17 Alarm-02.010.0001.00004-check_fuxi_master_alive

查看集群fuxi master是否alive。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P1)	fuxi master	fuxi

可能原因

程序异常退出、服务器宕机。

影响范围

影响集群稳定。

处理方法

1. 登录报警fuxi master，查看进程是否存在，存在进行下一步调查。
2. 确定*/apsara/cloud/data/corefile*是否有coredump出现。
3. 如果1和2检查都正常，该错误应该是误报，联系fuxi开发排查。

8.18 Alarm-01.010.0002.00001-check_package_manager_alive

查看集群pangu master进程是否alive。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	package manager	package manager

可能原因

程序异常退出、服务器宕机。

影响范围

影响集群包管理。

处理方法

1. 登录报警pacakge manager，查看进程是否存在。
   - 不存在，联系package 开发进行处理。
   - 存在，执行下一步。
2. 确定*/apsara/cloud/data/corefile*是否有coredump出现。
3. 如果1和2都正常，应该是误报，联系package开发排查问题。

8.19 Alarm-02.005.0001.00001-check_nuwa_config

查看nuwa config。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P1)	nuwa	nuwa

可能原因

配置文件丢失。

影响范围

影响集群服务。

处理方法

联系nuwa开发排查问题。

8.20 Alarm-01.000.0001.00006-盘古replication队列长度过长告警

维护工具检测到集群中Replication队列长度过长时，产生该告警，检测周期为一分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古整体

可能原因

大量的磁盘或者机器损坏。

影响范围

影响盘古的性能和数据安全。

处理方法

降低前端读写，减小盘古自身的压力。

8.21 Alarm-01.000.0001.00007-盘古工作模式告警

维护工具检测到集群中盘古工作模式不对的时候，产生该告警，检测周期为三分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	盘古	盘古整体

可能原因

大于一半的master机器挂了。

影响范围

影响盘古的可用性。

处理方法

修复没有运行的pangu master所在机器。

8.22 Alarm-01.000.0001.00008-盘古总文件数量过多告警

维护工具检测到集群中盘古文件数目过多时，产生该告警，检测周期为一分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古整体

可能原因

盘古中文件数量过大。

影响范围

影响盘古的可用性。

处理方法

删除一些不需要的文件，或者扩大盘古master，cs的内存。

8.23 Alarm-01.000.0001.00009-盘古空间使用超限告警

维护工具检测到集群中盘古使用量超限的时候，产生该告警，检测周期为一天。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古整体

可能原因

盘古中的文件容量过大。

影响范围

影响盘古的可用性。

处理方法

删除不需要使用的文件，或者扩容一些CS。

8.24 Alarm-01.000.0001.00010-盘古SECONDARY master数量不对告警

维护工具检测到集群中盘古SECONDARY master数量不足时，产生该告警，检测周期为半小时。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古整体

可能原因

存在master机器挂了。

影响范围

影响盘古的可用性。

处理方法

修复没有运行的pangu master所在机器。

8.25 Alarm-01.000.0001.00011-盘古binary文件不一致告警

维护工具检测到集群中盘古binary文件版本不一致时，产生该告警，检测周期为一小时。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古整体

可能原因

所有master的二进制文件md5不一致。

影响范围

影响盘古的可用性。

处理方法

将所有的second master 进程修复为一致。

如果是primary master不一致，先切换再修复。

8.26 Alarm-01.005.0001.00002-check_nw_zk_queue

zk请求队列。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	nuwa	nuwa

可能原因

nuwa进程处理变慢。

影响范围

影响集群飞天服务。

处理方法

该监控出现可能会影响现有整个集群服务，需要联系nuwa开发上线排查。

8.27 Alarm-01.000.0001.00010-盘古SECONDARY master数量不对告警

维护工具检测到集群中盘古SECONDARY master数量不足时，产生该告警，检测周期为半小时。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古整体

可能原因

存在master机器挂了。

影响范围

影响盘古的可用性。

处理方法

修复没有运行的pangu master所在机器。

8.28 Alarm-01.010.0001.00005-check_FuxiMaster_queue_size

fuxi master请求处理队列。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	fuxi master	fuxi

可能原因

请求过多或者fuxi进程处理变慢。

影响范围

整个集群fuxi处理能力。

处理方法

fuximaster连接数过多，可能是作业导致，也可能是fuxi master本身处理变慢引起，联系fuxi开发进行排查。

8.29 Alarm-01.000.0001.00011-盘古binary文件不一致告警

维护工具检测到集群中盘古binary文件版本不一致时，产生该告警，检测周期为一小时。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古整体

可能原因

所有master的二进制文件md5不一致。

影响范围

影响盘古的可用性。

处理方法

将所有的second master 进程修复为一致。

如果是primary master不一致，先切换再修复。

8.30 Alarm-01.000.0002.00000-check_cs_sendbuffer

查看cs sendbuffer是否打满。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	chunkserver	chunkserver

可能原因

对端cpu打满或者网络丢包。

影响范围

影响cs服务能力。

处理方法

联系pangu开发查看sendbuffer打满机器，查看网络是否正常、服务器正常服务。

8.31 Alarm-02.500.0001.00000-check_port_80

查看80端口是否被监听。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P1)	tunnel/frontend	tunnel/frontend

可能原因

没有nginx进程、服务器宕机。

影响范围

单台服务器无法服务。

处理方法

登录机器，执行`ps aux | grep nginx`命令，查看进程是否存在。

- 如果不存在但nginx已经安装，执行`sudo /home/admin/nginx/sbin/nginx -s start`启动。
- 如果nginx没有安装，判断是frontend还是tunnel，联系对应开发排查问题。

8.32 Alarm-02.500.0001.00001-check_coredump

进程是否发生core文件。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P2)	tunnel/frontend	tunnel/frontend

可能原因

对应进程出现异常。

影响范围

单台服务器可能无法正常服务。

处理方法

判断是frontend还是tunnel，联系对应开发进行排查，查看服务是否正常。

8.33 Alarm-02.500.0001.00002-check_status_file

status.html是否存在。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P2)	tunnel/frontend	tunnel/frontend

可能原因

tunnel/frontend进程异常。

影响范围

单台服务器无法服务。

处理方法

- 如果是升级过程，请忽略。
- 如果是机器重装,联系frontend/tunnel开发部署服务。

如果不是上述两种情况，联系frontend或者tunnel开发排查。

8.34 Alarm-02.500.0002.00000-check_frontend_process_exists

frontend进程是否存在。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P2)	frontend	frontend

可能原因

服务器异常、进程异常。

影响范围

单台服务器无法服务。

处理方法

联系frontend开发排查进程为何不存在。

8.35 Alarm-02.500.0001.00003-check_toa_odps

查看toa模块是否加载。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P2)	单台服务器	单机报警

可能原因

没有insmod toa。

影响范围

无法查看真实IP。

处理方法

执行如下命令：

```
sudo modprobe toa
```

```
lsmod | grep toa
```

8.36 Alarm-02.500.0003.00000-check_tunnel_service

tunnel进程是否存在。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P2)	tunnel	tunnel

可能原因

服务器异常、进程异常。

影响范围

单台服务器无法服务。

处理方法

联系tunnel开发排查服务不正常原因。

8.37 Alarm-01.500.0004.00000-Check_ExecutorWorker_sql_relative_task_default_QPS

sql_relative总请求QPS。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	executorworker	executorworker

可能原因

sql_relative类型请求数过多。

影响范围

影响sql_relative类型作业正常运行。

处理方法

联系executor work开发进行排查。

8.38 Alarm-01.500.0004.00001-Check_ExecutorWorker_sql_relative_task_default_Latency

tryrun的latency。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	executorworker	executorworker

可能原因

sql_relative类型操作重跑延迟增加负载过高，延迟增加。

影响范围

影响sql_relative类型作业正常运行。

处理方法

联系executor work开发进行排查。

8.39 Alarm-01.500.0004.00002-Check_ExecutorWorker_aggregate_task_default_QPS

aggregate总请求QPS。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	executorworker	executorworker

可能原因

aggregate类型请求数过多。

影响范围

影响aggregate类型作业运行。

处理方法

联系executor work开发进行排查。

8.40 Alarm-01.500.0004.00003-Check_ExecutorWorker_aggregate_task_default_Latency

aggregate tryrun的latency。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	executorworker	executorworker

可能原因

aggregate类型操作重跑延迟增加。

影响范围

影响aggregate类型作业正常运行。

处理方法

联系executor work开发进行排查。

8.41 Alarm-01.500.0004.00004-Check_ExecutorWorker_RunningTaskCount

executorwork运行的线程数。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	executor	executor

可能原因

job数过多。

影响范围

影响整个服务单位时间运行job的能力。

处理方法

如果报警时间超过报警间隔3-5倍，也就是联系收到3-5次报警。

联系executor work排查是否需要扩容，如果偶尔在业务高峰期出现一次，属于压力过大，符合预期。

8.42 Alarm-01.500.0004.00005-Check_ExecutorWorker_EasyRPC_Latency

executor父进程网络耗时。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	executor	executor

可能原因

服务器负载过高、网络不稳定。

影响范围

影响executor正常服务。

处理方法

联系executor work开发进行排查。

8.43 Alarm-01.500.0005.00000-check_odpsworker_requestpoolsize

odps worker 请求队列长度。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	odpswork	odpswork

可能原因

job数过多，odpswork数量不足。

影响范围

影响odps服务正常运行。

处理方法

联系odpswork开发排查odpswork是否正常工作。

- 如果是，则减少提交的job数或者增加odpswork数。
- 如果不是，则需要开发进一步排查。

8.44 Alarm-01.500.0005.00001-check_OdpsWorker_StoreEventLatecy

odpswork storeEvent方法的latency。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	odpswork	odpswork

可能原因

进程繁忙、ots延迟过大。

影响范围

影响storeEvent进程运行。

处理方法

查看对应机房网络流量打满和ots服务是否正常，联系scheduler worker开发上线排查，否则处理机房网络流量打满问题或者ots服务异常问题。

8.45 Alarm-01.500.0006.00000-check_SchedulerWorker_CreateInstanceQPS

异步instance请求的qps。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	schedulerwork	schedulerwork

可能原因

instance数增加、schedulerwork处理能力下降。

影响范围

影响schedulerwork正常运行。

处理方法

联系scheduler worker开发排查。

8.46 Alarm-01.500.0006.00001-Check_SchedulerWorker_RunningTaskCount

集群总的executorwork处理线程总数。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	schedulerwork	schedulerwork

可能原因

job数增加。

影响范围

影响整个odps服务处理能力。

处理方法

降低提交的job数或者联系odps executor work开发增加executor work线程数。

8.47 Alarm-01.500.0007.00000-check_QuotaWorkerRole_CPUUsage

quotaworker cpu使用率。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	quotawork	quotawork

可能原因

quotawork请求数过多、进程夯。

影响范围

影响tunnel服务的quota设置。

处理方法

联系quotawork开发上线排查，紧急情况可以对对应进程gcore $pid(对应进程pid）。

8.48 Alarm-01.500.0007.00001-check_QuotaWorkerRole_MEMUsage

quotaworker mem使用率。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	quotawork	quotawork

可能原因

quotawork请求数过多、进程夯。

影响范围

影响tunnel 服务的quota设置。

处理方法

联系quotawork开发上线排查，紧急情况可以对对应进程执行`gcore $pid`命令，pid对应进程pid。

8.49 Alarm-01.500.0008.00000-check_MessageServerRole_CPUUsage

messagework CPU使用率。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	messagework	messagework

可能原因

message过多、进程异常。

影响范围

影响ODPS各角色间消息传送。

处理方法

联系messagework开发上线排查,紧急情况可以对对应进程`gcore $pid`，pid表示对应进程pid。

8.50 Alarm-01.500.0008.00001-check_MessageServerRole_MEMUsage

messagework mem使用率。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	messagework	messagework

可能原因

message过多、进程异常。

影响范围

影响ODPS各角色间消息传送。

处理方法

联系messagework开发上线排查，紧急情况可以对对应进程执行gcore $pid，pid表示对应进程pid。

8.51 Alarm-01.500.0009.00000-check_hiveserver_fn_createPartition_latency

create partition的latency。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	hiveserver	hiveserver

可能原因

OTS异常。

影响范围

影响ODPS服务创建partition。

处理方法

查看对应机房网络流量是否打满和OTS服务是否正常，联系hiveserver开发上线排查，否则处理机房网络流量打满问题或者ots服务异常问题。

8.52 Alarm-01.500.0010.00000-check_ddl_server_thread_pool_state

正在运行的ddltask个数。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	hiveserver	hiveserver

可能原因

有ddl hang，ddlserver数量不够。

影响范围

影响ddl操作。

处理方法

联系hiverserver开发上线排查。

8.53 Alarm-01.500.0010.00001-check_ddl_server_request_qps

submit ddl task的qps。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	hiveserver	hiveserver

可能原因

ddl作业增加。

影响范围

ddl操作延迟增大。

处理方法

查看对应机房网络流量没有打和ots服务正常，联系hiveserver开发上线排查，否则处理机房网络流量打满问题或者ots服务异常问题。

8.54 Alarm-01.500.0010.00002-check_ddl_server_ots_operate_latency

ddlserver持久化meta的latency。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	hiveserver	hiveserver

可能原因

OTS异常。

影响范围

影响meta数据一致性。

处理方法

查看对应机房网络流量没有打和ots服务正常，联系hiveserver开发上线排查，否则处理机房网络流量打满问题或者ots服务异常问题。

8.55 Alarm-01.500.0010.00003-check_ddl_server_execute_latency

ddl执行的平均时间。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	hiveserver	hiveserver

可能原因

OTS异常。

影响范围

影响ddl服务吞吐量。

处理方法

查看对应机房网络流量没有打和ots服务正常，联系hiveserver开发上线排查，否则处理机房网络流量打满问题或者ots服务异常问题。

8.56 Alarm-01.500.0011.00000-Check_RecycleWorker_CPUUsage

recyclework CPU使用率。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	recyclework	recyclework

可能原因

recyclework进程异常、负载过高。

影响范围

影响recyclework正常运行。

处理方法

联系recycle开发上线排查，紧急情况可以对对应进程执行gcore $pid命令，pid为对应进程pid。

8.57 Alarm-01.500.0011.00001-Check_RecycleWorker_MEMUsage

recyclework mem使用率。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P2)	recyclework	recyclework

可能原因

recyclework进程异常、负载过高。

影响范围

影响recyclework正常运行。

处理方法

联系recycle开发上线排查，紧急情况可以对对应进程执行gcore $pid命令，pid为对应进程pid。

8.58 Alarm-01.500.0009.00001-check_hiveserver_ThreadsRunnable

hiveserver runnable运行数。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	重要(P1)	hiveserver	hiveserver

可能原因

线程没有及时释放。

影响范围

影响hiveserver服务稳定。

处理方法

联系hiveserver开发上线排查。

9 伏羲

9.1 Alarm-02.010.0001.00002- odps_apsara_pm_ag-check_package_manager_alive

集群中的PackageManager进程是否存活。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	紧急(P1)	fuxi	PackageManager

可能原因

PackageManager进程无法启动，或者hang住。

影响范围

无法向fuxi上传package，fuxi无法正常拉起进程。

处理方法

登录PackageManager机器，查看PackageManager进程是否存在。

9.2 Alarm-02.010.0002.00003-odps_apsara_fm_ag-check_fuxi_master_hang

集群中的FuxiMaster不服务。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	紧急(P1)	fuxi	FuxiMaster

可能原因

FuxiMaster进程无法响应请求。

影响范围

无法向伏羲提交任务。

处理方法

1. FuxiMaster会自动处理这种情况，收到报警后请登录集群使用/apsara/deploy/rpc_wrapper/rpc.sh al命令，检查是否恢复 。
2. 如果未恢复，请检查nuwa 服务是否正常，是否存在压力过大的情况。

9.3 Alarm-01.010.0002.00004-odps_apsara_fm_ag-check_fuxi_job_num

FuxiMaster提交的任务数过多。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	紧急(P1)	fuxi	FuxiMaster

可能原因

- 用户odps任务提交过多，或者提交了太多的merge task

影响范围

- fuxi性能下降，甚至完全无法响应

处理方法

停掉不低优先级的作业

9.4 Alarm-02.010.0003.00005-odps_apsara_fm_ag-check_fuxi service_status

集群中aos_fuxi包不存在。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	紧急(P1)	fuxi	Package

可能原因

- 误删除了aos_fuxi包

影响范围

- 无法启动job或者service

处理方法

检查Package这个serverrole进程是否存在，如果存在重启该进程。

9.5 Alarm-02.010.0002.00006-odps_apsara_fm_ag-check_fuxi_master_switch

FuxiMaster发生了主备切换。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	重要(P2)	fuxi	FuxiMaster

可能原因

FuxiMaster原机器无法连接到nuwa。

影响范围

FuxiMaster主机数目减少，服务能力降低。

处理方法

检查切换前的fuximaster机器是否正常，是否存在软件硬件故障。

9.6 Alarm-02.010.0002.00007-odps_apsara_fm_ag-check_fuxi_master_alive

FuxiMaster进程不存在或者FuxiMaster不工作。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	紧急(P1)	fuxi	FuxiMaster

可能原因

FuxiMaster主机无法联系、FuxiMaster进程不存在。

影响范围

FuxiMaster无法正常服务。

处理方法

检查FuxiMaster机器能否联通，如果能联通，检查Fuximaster进程是否存活。

10 盘古

10.1 Alarm-01.000.0002.00001-盘古Master checkpoint数量不足

当维护工具检测到盘古Master做的checkpoint数量不符合阈值的时候，产生该告警，检测周期为一小时。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Master

可能原因

- 盘古Master服务异常，长时间没有做checkpoint。
- 新集群。

影响范围

- 没有足够的checkpoint，会影响master进程启动时候的记载时间。
- 对数据安全性有一定影响。

处理方法

1. 检查集群是否是新创建集群。
   - 如果是，请忽略。
   - 如果否，请跳转至2。
2. 登录报警机器，查看/apsarapangu空间是否足够，如果空间不足，释放一下磁盘空间。

10.2 Alarm-02.000.0001.00001-盘古不可读写

当维护工具检测到盘古不可读写的时候，产生该告警，检测周期为1分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1	盘古	盘古整体

可能原因

- 盘古没有足够的磁盘空间。
- 盘古chunkserver数目不足。
- 网络问题。

影响范围

盘古不可读写，影响上层所有往盘古写入数据的服务。

处理方法

1. 登录PanguTools机器，执行/apsara/deploy/puadmin lscs命令，查看处于NORMAL的chunkserver数以及DISK_OK的磁盘数目是否正常。
   - 如果正常，请跳转至2。
   - 如果盘古空间已满，请通知集群管理员。
2. 查看集群的网络状态，PanguTools所在机器与盘古Master机器之间的网络情况。

10.3 Alarm-01.000.0001.00002-盘古集群中temp file文件大小超过阈值

当维护工具检测到集群中temp file大小超过阈值的时候，产生该告警，检测周期为1小时。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古整体

可能原因

上层服务写入太大的temp file文件。

影响范围

temp file过大，会占用过多盘古的存储空间，影响其他服务使用盘古。

处理方法

登录PangtuTools所在的机器，执行/apsara/deploy/puadmin cs -tempfile -top 1命令，查找最大的temp file，找到temp file的写入者，跟写入者确认写入文件大小是否合理。

10.4 Alarm-02.000.0003.00001-盘古chunkserver发生core dump

当维护工具检测到集群中chunkserver有core dump发生的时候，产生该告警，检测周期为15秒。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	盘古	盘古Chunkserver

可能原因

盘古Chunkserver程序运行异常，导致core dump。

影响范围

导致盘古Chunkserver进程重启，可能会造成上层服务发生core dump当时读写延迟高。

处理方法

保留*/cloud/data/corefile*目录下的core dump文件，联系盘古的开发人员，查看core dump的原因。

10.5 Alarm-02.000.0003.00002-盘古Chunkserver有特殊的事件发生

当维护工具检测到集群中chunkserver evet log中有Error级别的告警的时候，产生该告警，检测周期为一小时。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	盘古	盘古Chunkserver

可能原因

发生了诸如磁盘上下线，磁盘状态变ERROR，网络错误等错误事件。

影响范围

可能会影响发生问题的Chunkserver的状态。

处理方法

查看*/apsara/pangu_chunkserver/log*目录下的pangu_event.LOG，查看发生了哪些event，针对event采取相关的措施。

10.6 Alarm-01.000.0003.00003-盘古Chunkserver机器上的load过高

当维护工具检测到集群中chunkserver load高于指定阈值的时候，产生该告警，检测周期为15秒。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Chunkserver

可能原因

盘古Chunkserver上的进程数量过多。

影响范围

影响盘古Chunkserver的运行环境。

处理方法

登录报警的Chunkserver，查看哪些进程占用了大量CPU，确认该进程的运行状态是否符合预期。

10.7 Alarm-01.000.0003.00004-盘古Chunkserver map的so过多

当维护工具检测到集群中chunkserver map的so过多的时候，产生该告警，检测周期为1分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Chunkserver

可能原因

操作系统该异常。

影响范围

影响Chunkserver正常运行。

处理方法

1. 查看*/var/log/messages*是否存在异常。
2. 修复系统的异常。

10.8 Alarm-01.000.0003.00005-盘古Chunkserver内存使用过高

当维护工具检测到集群中chunkserver 内存使用高于指定阈值的时候，产生该告警，检测周期为10分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Chunkserver

可能原因

盘古Chunkserver上的chunk数量过多。

影响范围

影响该Chunkserver的运行状态，有OOM风险。

处理方法

登录PanguTools机器，执行/apsara/deploy/pu quota命令，查看是否创建的文件数量是否过多，联系应用确认写入文件量是否合理。

10.9 Alarm-01.000.0003.00006-盘古Chunkserver网络的recv流量过高

当维护工具检测到集群中chunkserver recv的网络流量过高时，产生该告警，检测周期为15秒。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Chunkserver

可能原因

盘古Chunkserver压力过大。

影响范围

影响该Chunkserver的服务质量。

处理方法

降低Client写压力。

10.10 Alarm-01.000.0003.00007-盘古Chunkserver网络的send流量过高

当维护工具检测到集群中chunkserver send的网络流量过高时，产生该告警，检测周期为15秒。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Chunkserver

可能原因

盘古Chunkserver压力过大。

影响范围

影响该Chunkserver的服务质量。

处理方法

降低Client读压力。

10.11 Alarm-01.000.0003.00008-盘古Chunkserver打开的文件句柄数目过多

当维护工具检测到集群中chunkserver 打开的文件句柄数过多的时候，产生该告警，检测周期为15秒。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Chunkserver

可能原因

盘古Chunkserver打开的chunks文件过多。

影响范围

影响该Chunkserver的服务质量。

处理方法

1. 执行ls -l /proc/<pid>/fd命令，查看Chunkserver进程打开的文件句柄。
2. 删除无用的文件。

10.12 Alarm-02.000.0003.00009-盘古Chunkserver进程有重启

当维护工具检测到集群中chunkserver 有重启的时候，产生该告警，检测周期为1分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	盘古	盘古Chunkserver

可能原因

盘古Chunkserer进程发生了重启，有可能是因为OOM。

影响范围

影响该Chunkserver的正常运行。

处理方法

1. 登录到该Chunkserver，查看dmesg信息，查看是否有OOM。
   - 如果有，确定是因为OOM，造成进程重启，查看Chunkserver内存增长原因。
   - 如果没有，请跳转至2。
2. 查看*/apsara/pangu_chunkserver/log*目录下pangu_chunkserver.LOG.1，搜索FATAL log，查看上一次进程死掉原因。

10.13 Alarm-02.000.0003.00010-盘古Chunkserver ulimit 设置错误告警

当维护工具检测到集群中chunkserver 机器ulimit设置不为unlimited的时候，产生该告警，检测周期为1分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	盘古	盘古Chunkserver

可能原因

ulimit设置错误。

影响范围

可能会影响Chunkserver正常运行。

处理方法

设置盘古Chunkserver机器**ulimit**为**unlimited**。

10.14 Alarm-01.000.0003.00011-盘古Chunkserver 机器/apsara目录空间不足

当维护工具检测到集群中chunkserver 机器/apsara空间不足的时候，产生该告警，检测周期为15秒。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Chunkserver

可能原因

/apsara目录被其他进程大量占用。

影响范围

影响Chunkserver正常运行。

处理方法

登录Chunkserver机器，查看/apsara目录使用情况，联系集群管理员进行清理。

10.15 Alarm-01.000.0003.00012-盘古Chunkserver 机器/apsarapangu目录空间不足

当维护工具检测到集群中chunkserver 机器/apsarapangu空间不足的时候，产生该告警，检测周期为15秒。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Chunkserver

可能原因

/apsarapangu目录被其他进程大量占用。

影响范围

影响Chunkserver正常运行。

处理方法

登录Chunkserver机器，查看/apsarapangu空间使用情况，联系集群管理员进行清理。

10.16 Alarm-01.000.0003.00013-盘古Chunkserver 机器根目录空间不足

当维护工具检测到集群中chunkserver 机器根目录空间不足的时候，产生该告警，检测周期为15秒。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Chunkserver

可能原因

根目录被其他进程大量占用。

影响范围

影响Chunkserver正常运行。

处理方法

登录Chunkserver机器，查看根目录空间使用情况，联系集群管理员清理。

10.17 Alarm-02.000.0002.00002-盘古master发生core dump

当维护工具检测到集群中master有core dump发生时，产生该告警，检测周期为15秒。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	盘古	盘古Master

可能原因

盘古Master程序运行异常，导致core dump。

影响范围

导致盘古Master进程重启，对服务安全性有风险。

处理方法

保留*/cloud/data/corefile*目录下的core dump文件，联系盘古开发人员，查看core dump的原因。

10.18 Alarm-02.000.0002.00003-盘古Master有特殊的事件发生

当维护工具检测到集群中Master evet log中有Error级别的告警时，产生该告警，检测周期为一小时。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	盘古	盘古Master

可能原因

发生了诸如磁盘上下线，磁盘状态变ERROR，网络错误等错误事件。

影响范围

影响发生问题的Master的状态。

处理方法

查看*/apsara/pangu_master/log*目录下的pangu_event.LOG，查看发生了哪些event，针对event采取相关的措施。

10.19 Alarm-01.000.0002.00004-盘古Master机器上的load过高

当维护工具检测到集群中Master load高于指定阈值时，产生该告警，检测周期为15秒。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Master

可能原因

盘古Master上的进程数量过多。

影响范围

影响盘古Master的运行环境。

处理方法

登录报警的Master，查看哪些进程占用了大量CPU，确认该进程的运行状态是否符合预期。

10.20 Alarm-01.000.0002.00005-盘古Master map的so过多

当维护工具检测到集群中master map的so过多时，产生该告警，检测周期为1分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Master

可能原因

操作系统异常。

影响范围

影响Master正常运行。

处理方法

1. 查看*/var/log/messages*是否存在异常。
2. 修复系统的异常。

10.21 Alarm-01.000.0002.00006-盘古Master内存使用过高

当维护工具检测到集群中Master 内存使用高于指定阈值的时候，产生该告警，检测周期为10分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Master

可能原因

盘古上保存的文件数目太多。

影响范围

影响该Master的运行状态，有OOM风险。

处理方法

登录PanguTools机器上执行*/apsara/deploy/pu quota*查看是否创建的文件数量是否过多，联系应用确认写入文件量是否合理

10.22 Alarm-02.000.0002.00007-盘古Master内存overcommit参数配置错误

当维护工具检测到集群中Master 内存overcommit参数配置错误的时候，产生该告警，检测周期为10分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	盘古	盘古Master

可能原因

内存overcommit 配置错误。

影响范围

对盘古Master正常运行有风险。

处理方法

登录该盘古Master机器，设置overcommit参数为0。

10.23 Alarm-01.000.0002.00008-盘古Master内存速度不符合预期告警

当维护工具检测到集群中Master内存速度不符合预期的时候，产生该告警，检测周期为1小时。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Master

可能原因

盘古Master内存硬件错误。

影响范围

对盘古Master运行有影响。

处理方法

更换内存。

10.24 Alarm-01.000.0002.00009-盘古Master网络的recv流量过高

当维护工具检测到集群中Master recv的网络流量过高的时候，产生该告警，检测周期为15秒。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Master

可能原因

盘古Master压力过大。

影响范围

影响该Master的服务质量。

处理方法

减少读写，删除，创建文件相关的操作。

10.25 Alarm-01.000.0002.00010-盘古Master网络的send流量过高

当维护工具检测到集群中Master send的网络流量过高的时候，产生该告警，检测周期为15秒。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Master

可能原因

盘古Master压力过大。

影响范围

影响该Master的服务质量。

处理方法

减少读写，删除，创建文件相关的操作。

10.26 Alarm-01.000.0002.00011-盘古Master打开的文件句柄数目过多

当维护工具检测到集群中Master 打开的文件句柄数过多时，产生该告警，检测周期为15秒。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Master

可能原因

盘古Master打开的chunks文件过多。

影响范围

影响该Master的服务质量。

处理方法

1. 执行ls -l /proc/<pid>/fd命令，查看Master进程打开的文件句柄。
2. 删除不需要的文件。

10.27 Alarm-02.000.0002.00012-盘古Master进程有重启

当维护工具检测到集群中Master 有重启时，产生该告警，检测周期为1分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	盘古	盘古Master

可能原因

盘古Master进程发生了重启，可能是因为OOM。

影响范围

Master进程重启比较严重，影响该Master的正常运行。

处理方法

1. 登录到该Master，看dmesg信息，查看是否有OOM。
   - 如果有，确定是由于OOM造成进程重启，查看Master内存增长原因。
   - 如果没有，请跳转至2。
2. 查看 */apsara/pangu_master/log*路径下pangu_master.LOG.1，搜索FATAL log，查看上一次进程坏死原因。

10.28 Alarm-02.000.0002.00013-盘古Master ulimit 设置错误告警

当维护工具检测到集群中master 机器**ulimit**没有设置为**unlimited**时，产生该告警，检测周期为1分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	盘古	盘古Master

可能原因

ulimit设置错误。

影响范围

可能会影响Master正常运行。

处理方法

设置盘古Master机器**ulimit**为**unlimited**。

10.29 Alarm-02.000.0004.00001-盘古Supervisor进程发生重启

当维护工具检测到集群中Supervisor 有重启时，产生该告警，检测周期为1分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P2	盘古	盘古Supervisor

可能原因

盘古Supervisor进程发生了重启，有可能是因为OOM。

影响范围

可能影响相关的运维操作。

处理方法

1. 登录到Supervisor，看dmesg信息，查看是否有OOM。
   - 如果有，确定是因为OOM造成进程重启，查看Supervisor内存增长原因。
   - 如果没有，请跳转至2。
2. 查看*/apsara/pangu_supervisor/log*路径下pangu_supervisor.LOG.1，搜FATAL log，看上一次进程坏死原因。

10.30 Alarm-01.000.0003.00014-检查混合存储机型有效文件在ssd盘的长度

维护工具检测对所有的CS的机器检查有效数据的容量是否超过指定的容量大小，如果超出则产生该警告，检测周期为1分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Chunkserver

可能原因

cs机器的压力太大，或者ssd盘的读写速度太慢。

影响范围

可能会影响数据安全。

处理方法

设置对应的机器为 READONLY:

```
puadmin cs -stat tcp://<cs ip>:10260 --set=READONLY
```

10.31 Alarm-01.000.0003.00015-检查混合存储机型ssd盘中数据失败的次数

维护工具检测对所有的CS的机器检查replay ssd盘失败的次数，如果存在失败则产生该警告，检测周期为1分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Chunkserver

可能原因

SSD盘可能坏了。

影响范围

可能会影响数据安全。

处理方法

设置对应的机器为 READONLY:

`puadmin cs -stat tcp://<cs ip>:10260 --set=READONLY`

10.32 Alarm-02.000.0002.00014-盘古Master发生切换告警

维护工具检测到集群中Master发生主备切换时，产生该告警，检测周期为1分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
事件告警	P1	盘古	盘古Master

可能原因

盘古Master因为进程本身或者网络原因造成主备切换。

影响范围

切换时，可能会影响盘古Master的服务质量。

处理方法

登录到PanguTools所在的机器，执行/apsara/deploy/puadmin gems命令，查看盘古Master状态，查看所有Master是否处于NORMAL状态：

- 如果是，盘古Master状态正常，可能是当时网络问题。
- 如果不是，查看错误Master的原因。

10.33 Alarm-01.000.0003.00016-盘古Chunkserver坏盘数量过多告警

维护工具检测到集群中Chunkserver上坏盘过多时，产生该告警，检测周期为1分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	盘古	盘古Chunkserver

可能原因

磁盘损坏。

影响范围

盘古存储容量下降。

处理方法

登录PanguTools所在机器，执行/apsara/deploy/puadmin lscs命令，查看状态非DISK_OK的磁盘，并将坏盘下线。

10.34 Alarm-01.000.0003.00017-盘古Chunkserver写满的磁盘数量过多告警

维护工具检测到集群中Chunkserver上被写满的磁盘数量过多时，产生该告警，检测周期为1分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	盘古	盘古Chunkserver

可能原因

写入盘古数据过多。

影响范围

盘古存储容量下降。

处理方法

登录PanguTools所在机器，执行/apsara/deploy/puadmin lscs命令，查看盘古使用情况，如果使用过多，请联系集群负责人扩容。

10.35 Alarm-01.000.0003.00018-盘古Chunkserver HANG盘数量过多告警

维护工具检测到集群中Chunkserver上HANG盘过多的时候，产生该告警，检测周期为1分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Chunkserver

可能原因

磁盘HANG住。

影响范围

造成过多僵尸进程。

处理方法

登录PanguTools所在机器，执行/apsara/deploy/puadmin lscs命令，查看状态DISK_HANG的磁盘，将HANG盘的机器重启。

10.36 Alarm-01.000.0001.00003-盘古存在有0副本文件

维护工具检测到集群中存在0副本的文件时，产生该告警，检测周期为5分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	盘古	盘古整体

可能原因

同一时间大量机器或者磁盘损坏。

影响范围

影响数据安全性。

处理方法

1. 登录PanguTools所在机器，执行/apsara/deploy/puadmin lscs命令，查看状态非NORMAL的Chunkserver，并将机器或者进程重新启动。

10.37 Alarm-01.000.0001.00004-盘古存在有1副本文件

维护工具检测到集群中存在1副本的文件时，产生该告警，检测周期为5分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古整体

可能原因

同一时间大量的机器或者磁盘损坏。

影响范围

影响数据安全性。

处理方法

调大集群的replication流量限制，使其尽快复制。

10.38 Alarm-01.000.0001.00005-盘古replication流量过大

维护工具检测到集群中replication流量过大时，产生该告警，检测周期为5分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古整体

可能原因

同一时间大量的机器或者磁盘损坏。

影响范围

存在数据安全性隐患。

处理方法

登录PanguTools机器，执行/apara/deploy/puadmin lscs命令，查看是否有过多的DISCONNECTED机器或者DISK_ERROR的盘。

10.39 Alarm-02.000.0002.00015-盘古Master主从之间log同步差距过大

维护工具检测到集群中Master主从log同步差距过大时，产生该告警，检测周期为5分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	盘古	盘古Master

可能原因

盘古Maste主从之间网络存在问题。

影响范围

影响盘古Master的服务安全性。

处理方法

查看盘古Master主从之间的网络情况。

10.40 Alarm-02.000.0002.00016-盘古Master工作队列过长

维护工具检测到集群中Master工作队列过长时，产生该告警，检测周期为半小时。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Master

可能原因

盘古Master压力过大。

影响范围

影响盘古运行状态。

处理方法

减少读写，删除，创建文件相关的操作。

10.41 Alarm-02.000.0002.00017-盘古Master状态告警

维护工具检测到集群中Master状态不对时，产生该告警，检测周期为一分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Master

可能原因

盘古Maste因为硬件，网络等原因导致状态不对。

影响范围

影响盘古Master的服务安全性。

处理方法

登录PanguTools机器，执行/apsara/deploy/puadmin gems命令，查看哪个master状态不对并调查原因。

10.42 Alarm-01.000.0001.00006-盘古replication队列长度过长告警

维护工具检测到集群中Replication队列长度过长时，产生该告警，检测周期为一分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古整体

可能原因

大量的磁盘或者机器损坏。

影响范围

影响盘古的性能和数据安全。

处理方法

降低前端读写，减小盘古自身的压力。

10.43 Alarm-01.000.0001.00007-盘古工作模式告警

维护工具检测到集群中盘古工作模式不对的时候，产生该告警，检测周期为三分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	盘古	盘古整体

可能原因

大于一半的master机器挂了。

影响范围

影响盘古的可用性。

处理方法

修复没有运行的pangu master所在机器。

10.44 Alarm-01.000.0001.00008-盘古总文件数量过多告警

维护工具检测到集群中盘古文件数目过多时，产生该告警，检测周期为一分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古整体

可能原因

盘古中文件数量过大。

影响范围

影响盘古的可用性。

处理方法

删除一些不需要的文件，或者扩大盘古master，cs的内存。

10.45 Alarm-01.000.0001.00009-盘古空间使用超限告警

维护工具检测到集群中盘古使用量超限的时候，产生该告警，检测周期为一天。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古整体

可能原因

盘古中的文件容量过大。

影响范围

影响盘古的可用性。

处理方法

删除不需要使用的文件，或者扩容一些CS。

10.46 Alarm-01.000.0001.00010-盘古SECONDARY master数量不对告警

维护工具检测到集群中盘古SECONDARY master数量不足时，产生该告警，检测周期为半小时。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古整体

可能原因

存在master机器挂了。

影响范围

影响盘古的可用性。

处理方法

修复没有运行的pangu master所在机器。

10.47 Alarm-01.000.0001.00011-盘古binary文件不一致告警

维护工具检测到集群中盘古binary文件版本不一致时，产生该告警，检测周期为一小时。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古整体

可能原因

所有master的二进制文件md5不一致。

影响范围

影响盘古的可用性。

处理方法

将所有的second master 进程修复为一致。

如果是primary master不一致，先切换再修复。

10.48 Alarm-02.000.0002.00018-盘古Normal file的操作队列过长告警

维护工具检测到集群中盘古Master Normal file操作队列过长的时候，产生该告警，检测周期为两分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Master

可能原因

Normal File相关的操作过多。

影响范围

影响盘古的可用性。

处理方法

减少Normal File 相关的：读写，删除，创建文件相关的操作。

10.49 Alarm-01.000.0003.00019-盘古Chunkserver sendbuffer过高报警

维护工具检测到集群中盘古Chunkserver sendbuffer过大时，产生该告警，检测周期为两分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Chunkserver

可能原因

读写压力过多。

影响范围

影响盘古的可用性。

处理方法

减少前端读写压力。

10.50 Alarm-02.000.0002.00019-盘古normal file的读操作队列过长告警

维护工具检测到集群中盘古Master 读操作队列过长时，产生该告警，检测周期为两分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Master

可能原因

Normal File读压力过大。

影响范围

影响盘古的可用性。

处理方法

减少前端Normal File 的读。

10.51 Alarm-02.000.0002.00020-盘古normal file的写操作队列过长告警

维护工具检测到集群中盘古Master 写操作队列过长时，产生该告警，检测周期为两分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Master

可能原因

Normal File写压力过大。

影响范围

影响盘古的可用性。

处理方法

减少前端Normal File 的写。

10.52 Alarm-02.000.0002.00021-盘古Master batch 操作队列过长告警

维护工具检测到集群中盘古Master batch操作队列过长时，产生该告警，检测周期为两分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Master

可能原因

Batch相关的操作过多。

影响范围

影响盘古的可用性。

处理方法

减少前端Batch相关的操作。

10.53 Alarm-02.000.0002.00022-盘古Master batch 读操作队列过长告警

维护工具检测到集群中盘古Master batch读操作队列过长时，产生该告警，检测周期为两分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Master

可能原因

Batch 读相关的操作过多。

影响范围

影响盘古的可用性。

处理方法

减少前端Batch读相关的操作。

10.54 Alarm-02.000.0002.00023-盘古Master batch 写操作队列过长告警

维护工具检测到集群中盘古Master batch写操作队列过长时，产生该告警，检测周期为两分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Master

可能原因

Batch 写相关的操作过多。

影响范围

影响盘古的可用性。

处理方法

减少前端Batch写相关的操作。

10.55 Alarm-02.000.0002.00024-盘古Master 选举队列过长告警

维护工具检测到集群中盘古Master 选举队列过长的时候，产生该告警，检测周期为两分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	盘古	盘古Master

可能原因

网络状态可能有问题。

影响范围

影响盘古的可用性。

处理方法

修复网络相关的错误。

10.56 Alarm-02.000.0002.00025-盘古Master 紧急操作队列过长告警

维护工具检测到集群中盘古Master 紧急操作队列过长时，产生该告警，检测周期为两分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	盘古	盘古Master

可能原因

紧急操作比较多。

影响范围

影响盘古的可用性。

处理方法

降低前端读写，减小盘古自身的压力。

10.57 Alarm-02.000.0002.00026-盘古Master 心跳队列告警

维护工具检测到集群中盘古Master 心跳队列过长时，产生该告警，检测周期为两分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P1	盘古	盘古Master

可能原因

网络状态可能有问题。

影响范围

影响盘古的可用性。

处理方法

修复网络相关的错误。

10.58 Alarm-02.000.0002.00027-盘古Master高优先级队列过长告警

维护工具检测到集群中盘古Master 高优先级队列过长时，产生该告警，检测周期为两分钟。

告警信息

告警类型	告警级别	告警对象	告警模块
阈值告警	P2	盘古	盘古Master

可能原因

高优先级的操作比较多。

影响范围

影响盘古的可用性。

处理方法

降低前端读写，减小盘古自身的压力。