

---

## 文章

[Michael Lei](#) · 五月 12, 2021 阅读大约需 17 分钟

# InterSystems 最佳实践系列之--数据平台和性能 – 虚拟机备份和 Caché 冻结/解冻脚本

在本帖中，我将展示使用外部备份来备份 Caché 的策略，以及与基于快照的解决方案集成的示例。如今，大多数解决方案部署在基于 VMware 的 Linux 上，因此许多帖子都以展示解决方案如何集成 VMware 快照技术为例。

## Caché 备份 - 包括电池？

Caché 安装后即包含 Caché 在线备份，可提供不间断的 Caché 数据库备份。但随着系统规模的扩大，您应该考虑更高效的备份解决方案。集成了快照技术的外部备份是推荐的系统（包括 Caché 数据库）备份解决方案。

## 外部备份有特殊注意事项吗？

[外部备份](#)的在线文档包含了全部详细信息。一个关键考虑事项是：

“为确保快照的完整性，Caché 提供了在创建快照时冻结数据库写操作的方法。在创建快照期间，只冻结对数据库文件的物理写入，从而允许用户进程继续在内存中不间断地执行更新。”

还需要注意的是，虚拟化系统上的部分快照过程会导致正在备份的虚拟机短暂暂停，这段时间通常称为关闭时间。该时间通常不到一秒，因此不会被用户注意到，也不会影响系统运行，但在某些情况下，关闭时间可能较长。如果关闭时间长于 Caché 数据库镜像的 QoS 超时时间，那么备份节点将认为主节点出现故障，并将进行故障转移。在本帖的后面部分，我将说明在需要对镜像 QoS 超时时间进行更改时如何查看关闭时间。

---

[这里是 InterSystems 数据平台和性能系列的其他帖子的列表。](#)

对于本帖，您还应该查看 Caché 在线文档中的[备份和还原指南](#)。

---

## 备份选择

### 最低限度的备份解决方案 - Caché 在线备份

如果您没有其他备份方案，可使用 InterSystems 数据平台附带的这一功能，以实现零停机时间的备份。请记住，Caché 在线备份只备份 Caché 数据库文件，它将捕获数据库中为数据分配的所有块以及写入到顺序文件的输出。Caché 在线备份支持累积备份和增量备份。

在 VMware 的上下文中，Caché 在线备份是一种客户机内备份解决方案。与其他客户机内解决方案一样，无论应用程序已虚拟化还是直接在主机上运行，Caché 在线备份的操作都基本相同。Caché 在线备份必须与系统备份配合使用，才能将 Caché 在线备份输出文件与应用程序使用的所有其他文件系统一起复制到备份介质中。最低限度的系统备份必须包括安装目录、日志和备用日志目录、应用程序文件以及包含应用程序使用的外部文件的任何目录。

Caché 在线备份应被视为小型站点的入门级方法，这些站点希望实施低成本解决方案，只备份 Caché 数据库或临时备份，例如，该方案在设置镜像时非常有用。然而，随着数据库规模的扩大，加上 Caché 通常只是客户数据环境的一部分，建议将外部备份与快照技术和第三方实用工具相结合来作为最佳做法。这样做的优点包括可备份非数据库文件、更快的还原时间、企业范围的数据视图以及更好的目录和管理工具。

---

## 推荐的备份解决方案 - 外部备份

以 VMware 为例；VMware 上的虚拟化为保护整个虚拟机增加了额外的功能和选择。将解决方案虚拟化后，您就有效地将系统（包括操作系统、应用程序和数据）全部封装在 .vmdk（和其他一些）文件中。需要时，可以非常方便地管理这些文件，并使用它们恢复整个系统，这与物理系统上的操作截然不同。在物理系统中，必须分别恢复和配置各个组件 - 操作系统、驱动程序、第三方应用程序、数据库和数据库文件等。

---

## VMware 快照

VMware 的 vSphere Data Protection (VDP) 和其他用于虚拟机备份的第三方备份解决方案（如 Veeam 或 Commvault）利用 VMware 虚拟机快照的功能来创建备份。以下是 VMware 快照的高阶说明，有关更多详细信息，请参见 VMware 文档。

快照应用于整个虚拟机，操作系统和所有应用程序或数据库引擎都不知道快照的进行，记住这一点很重要。此外，还要记住：

VMware 快照本身不是备份！

快照使备份软件进行备份，但它们本身并不是备份。

VDP 和第三方备份解决方案将 VMware 快照过程与备份应用程序相结合来管理快照的创建及非常重要的删除。从高级别看，使用 VMware 快照进行外部备份的过程和事件顺序如下：

- 第三方备份软件请求 ESXi 主机触发 VMware 快照。
- 虚拟机的 .vmdk 文件被置于只读状态，并且为虚拟机的每个 .vmdk 文件都创建一个子 vmdk 增量文件。
- 使用写入时复制机制，对虚拟机的所有更改都将写入增量文件。任何读取都先从增量文件进行。
- 备份软件管理只读父 .vmdk 文件到备份目标的复制。
- 备份完成后，将提交快照（虚拟机磁盘恢复写入，并且增量文件中更新的块将写入父文件）。
- VMware 快照此时已删除。

备份解决方案还使用其他功能，如变更块跟踪 (CBT) 允许增量或累积备份，以提高速度和效率（对节省空间尤为重要），并且通常还增加其他重要功能，如重复数据删除和压缩、调度、挂载已更改 IP 地址的虚拟机以进行完整性检查等、完整的虚拟机和文件级恢复，以及目录管理。

管理不当或长时间不运行的 VMware

快照可能会占用过多存储（随着越来越多的数据被更改，增量文件会持续增长），还会拖慢虚拟机的速度。

在生产实例上运行手动快照之前，您应该非常仔细地考虑。为什么要这么做？

如果时光倒流到创建快照的时候，会发生什么？从创建到回滚之间的所有应用程序事务会发生什么？

如果您的备份软件创建和删除快照，没有问题。快照应该只存在很短的时间。

备份策略的关键部分是选择系统使用率较低的时间，以最大程度地减少对用户和性能的影响。

---

## 快照的 Caché 数据库注意事项

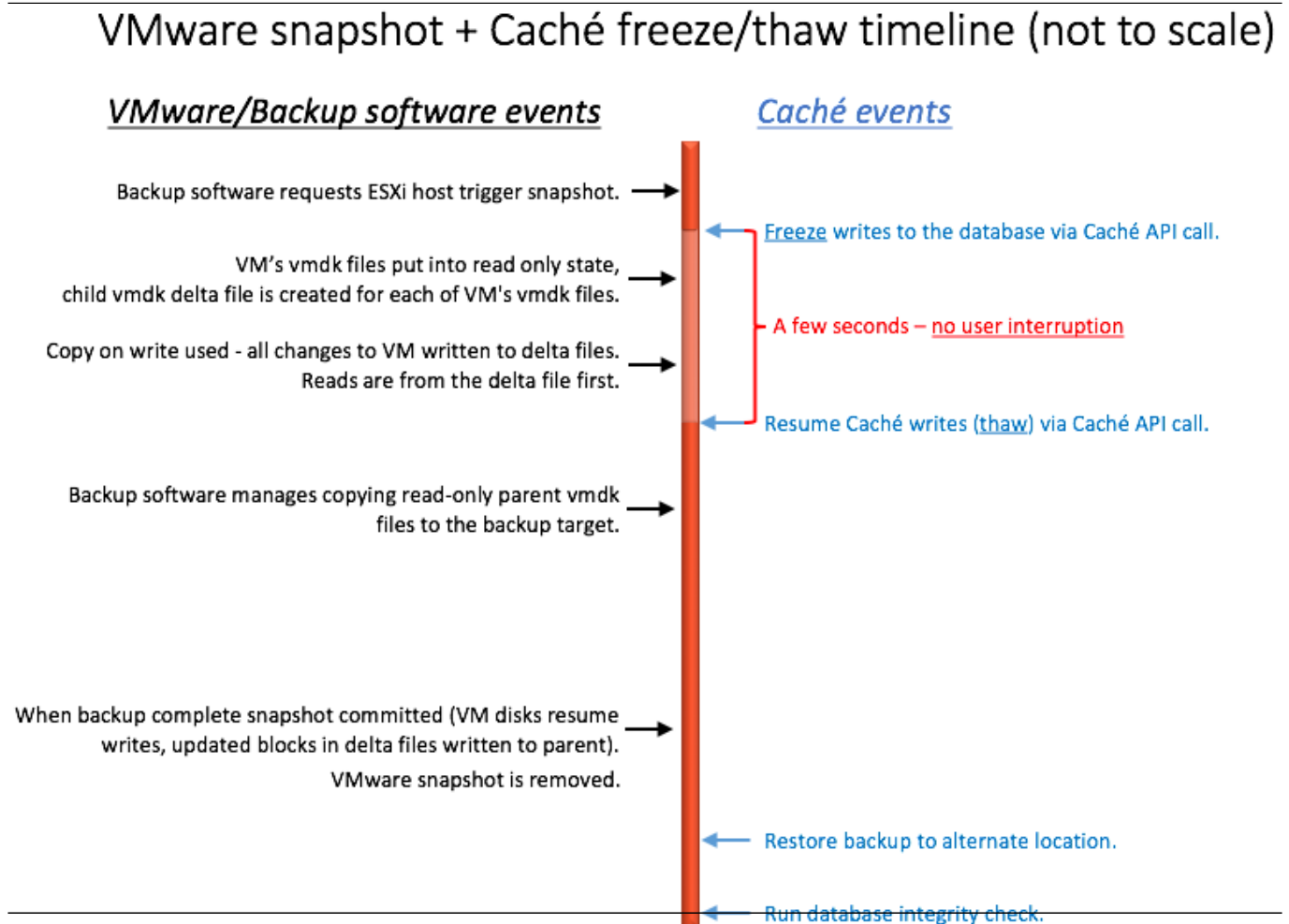
拍摄快照之前，数据库必须处于静默状态，以便提交所有待处理的写入，并且数据库处于一致状态。Caché 提供了方法和 API，以在创建快照时先提交对数据库的写入，再短时间冻结（停止）写入。这样，在创建快照期间，只冻结对数据库文件的物理写入，从而允许用户进程继续在内存中不间断地执行更新。触发快照后，将解冻数据库写入，备份继续将数据复制到备份介质。冻结和解冻的间隔应该非常短（几秒钟）。

除了暂停写入，Caché 冻结还会处理日志文件的切换并向日志写入备份标记。

在物理数据库写入被冻结时，日志文件继续正常写入。

如果在物理数据库写入被冻结时系统崩溃，则在启动过程中，会照常从日志恢复数据。

下图显示了为创建具有一致数据库映像的备份，所执行的冻结和解冻以及 VMware 快照步骤。



请注意，冻结和解冻之间的短暂时间只是创建快照的时间，而不是将只读父文件复制到备份目标的时间。

## 整合 Caché 冻结和解冻

vSphere 允许在创建快照的任一侧自动调用脚本，此时会调用 Caché 冻结和解冻。

注意：为了使此功能正常工作，ESXi 主机请求客户机操作系统通过 VMware Tools 静默磁盘。

客户机操作系统中必须安装 VMware Tools。

脚本必须遵守严格的名称和位置规则。还必须设置文件权限。对于 Linux 上的 VMware，脚本名称为：

```
# /usr/sbin/pre-freeze-script
# /usr/sbin/post-thaw-script
```

以下是我们团队对内部测试实验室实例进行 Veeam 备份时使用的冻结和解冻脚本示例，这些脚本应该也适用于其他解决方案。这些示例已在 vSphere 6 和 Red Hat 7 上测试和使用。

虽然这些脚本可用作示例并说明了方法，但您必须针对您自己的环境进行验证！

### pre-freeze-script 示例：

```
#!/bin/sh
#
# Script called by VMWare immediately prior to snapshot for backup.
# Tested on Red Hat 7.2
#

LOGDIR=/var/log
SNAPLOG=$LOGDIR/snapshot.log

echo >> $SNAPLOG
echo "`date`: Pre freeze script started" >> $SNAPLOG
exit_code=0

# Only for running instances
for INST in `ccontrol qall 2>/dev/null | tail -n +3 | grep '^up' | cut -c5- | awk '{
print $1}`'; do

    echo "`date`: Attempting to freeze $INST" >> $SNAPLOG

    # Detailed instances specific log
    LOGFILE=$LOGDIR/$INST-pre_post.log

    # Freeze
    csession $INST -U '%SYS' "##Class(Backup.General).ExternalFreeze(\"$LOGFILE\",,,,
,,1800)" >> $SNAPLOG $
    status=$?

    case $status in
        5) echo "`date`: $INST IS FROZEN" >> $SNAPLOG
            ;;
        3) echo "`date`: $INST FREEZE FAILED" >> $SNAPLOG
            logger -p user.err "freeze of $INST failed"
            exit_code=1
            ;;
        *) echo "`date`: ERROR: Unknown status code: $status" >> $SNAPLOG
            logger -p user.err "ERROR when freezing $INST"
            exit_code=1
            ;;
    esac
    echo "`date`: Completed freeze of $INST" >> $SNAPLOG
done

echo "`date`: Pre freeze script finished" >> $SNAPLOG
exit $exit_code
```

## 解冻脚本示例：

```
#!/bin/sh
#
# Script called by VMWare immediately after backup snapshot has been created
# Tested on Red Hat 7.2
#

LOGDIR=/var/log
SNAPLOG=$LOGDIR/snapshot.log

echo >> $SNAPLOG
echo "`date`: Post thaw script started" >> $SNAPLOG
exit_code=0

if [ -d "$LOGDIR" ]; then

    # Only for running instances
    for INST in `ccontrol qall 2>/dev/null | tail -n +3 | grep '^up' | cut -c5- | awk '{print $1}'`; do

        echo "`date`: Attempting to thaw $INST" >> $SNAPLOG

        # Detailed instances specific log
        LOGFILE=$LOGDIR/$INST-pre_post.log

        # Thaw
        csession $INST -U%SYS "##Class(Backup.General).ExternalThaw(\"$LOGFILE\")" >>
$SNAPLOG 2>&1
        status=$?

        case $status in
            5) echo "`date`: $INST IS THAWED" >> $SNAPLOG
                csession $INST -U%SYS "##Class(Backup.General).ExternalSetHistory(\"$LOGFILE\")" >> $SNAPLOG$
                ;;
            3) echo "`date`: $INST THAW FAILED" >> $SNAPLOG
                logger -p user.err "thaw of $INST failed"
                exit_code=1
                ;;
            *) echo "`date`: ERROR: Unknown status code: $status" >> $SNAPLOG
                logger -p user.err "ERROR when thawing $INST"
                exit_code=1
                ;;
        esac
        echo "`date`: Completed thaw of $INST" >> $SNAPLOG
    done
fi

echo "`date`: Post thaw script finished" >> $SNAPLOG
exit $exit_code
```

## 记得设置权限：

```
# sudo chown root.root /usr/sbin/pre-freeze-script /usr/sbin/post-thaw-script
# sudo chmod 0700 /usr/sbin/pre-freeze-script /usr/sbin/post-thaw-script
```

## 测试冻结和解冻

要测试脚本是否正确运行，可以在虚拟机上手动运行快照并检查脚本输出。 以下截图显示了 “ Take VM Snapshot ” （ 拍摄虚拟机快照 ） 对话框和选项。

**取消选中** - “ Snapshot the virtual machine's memory ” （ 拍摄虚拟机内存快照 ）。

**选中** - “ Quiesce guest file system (Needs VMware Tools installed) ” （ 静默客户机文件系统(需要安装 VMware Tools) ） 复选框以暂停客户机操作系统上运行的进程，以便在拍摄快照时文件系统内容处于已知的一致状态。

**重要！** 完成测试后，记得删除快照！！！

如果静默标志为真，并且在拍摄快照时虚拟机已开机，将使用 VMware Tools 静默虚拟机中的文件系统。静默文件系统是使磁盘上的数据进入适合备份的状态的过程。此过程可能包括将脏缓冲区从操作系统的内存中缓存刷新到磁盘等诸如此类的操作。

以下输出显示了在运行将快照包括为操作一部分的备份后，上述冻结/解冻脚本示例中设置的 \$SNAPSHOT 日志文件的内容。

```
Wed Jan 4 16:30:35 EST 2017: Pre freeze script started
Wed Jan 4 16:30:35 EST 2017: Attempting to freeze H20152
Wed Jan 4 16:30:36 EST 2017: H20152 IS FROZEN
Wed Jan 4 16:30:36 EST 2017: Completed freeze of H20152
Wed Jan 4 16:30:36 EST 2017: Pre freeze script finished

Wed Jan 4 16:30:41 EST 2017: Post thaw script started
Wed Jan 4 16:30:41 EST 2017: Attempting to thaw H20152
Wed Jan 4 16:30:42 EST 2017: H20152 IS THAWED
Wed Jan 4 16:30:42 EST 2017: Completed thaw of H20152
Wed Jan 4 16:30:42 EST 2017: Post thaw script finished
```

此示例显示，冻结和解冻之间经过了 6 秒 (16:30:36-16:30:42)。在此期间，用户操作不会中断。

*您将必须从您自己的系统中收集指标*

，但为说明起见，此示例来自在虚拟机上运行应用程序基准测试的系统，该虚拟机没有瓶颈，且平均指标为每秒超过 2 百万 Gloref、每秒 17 万 Gloupd、每秒 1100 次物理读取以及每个写守护进程周期 3000 次写入。

请记住，内存不是快照的一部分，因此在重启时，虚拟机将重启并恢复。数据库文件将保持一致。您不希望 “ 恢复 “ 备份，而是需要在某个时间点的文件。恢复文件后，您可以前滚日志，并执行保持应用程序和事务一致性所需的任何其他恢复步骤。

为实现额外的数据保护，系统还可以自行完成[日志切换](#)，并将日志备份或复制到其他位置，例如每小时备份一次。

以下是上述冻结/解冻脚本示例中的 \$LOGFILE 的输出，其中显示了快照的日志详细信息。

```
01/04/2017 16:30:35: Backup.General.ExternalFreeze: Suspending system

Journal file switched to:
/trak/jnl/jrnpri/h20152/H20152_20170104.011
01/04/2017 16:30:35: Backup.General.ExternalFreeze: Start a journal restore for this
backup with journal file: /trak/jnl/jrnpri/h20152/H20152_20170104.011

Journal marker set at
```

```
offset 197192 of /trak/jnl/jrnpri/h20152/H20152_20170104.011
01/04/2017 16:30:36: Backup.General.ExternalFreeze: System suspended
01/04/2017 16:30:41: Backup.General.ExternalThaw: Resuming system
01/04/2017 16:30:42: Backup.General.ExternalThaw: System resumed
```

---

## 虚拟机关闭时间

在创建虚拟机快照时以及完成备份和提交快照后，虚拟机需要短时间冻结。这个短时间冻结通常指关闭虚拟机。[这里](#)是一篇不错的关于关闭时间的博客文章。我在下面总结了细节，并将其放在 Caché 数据库注意事项的上下文中。

根据那篇关于关闭时间的文章：“要创建虚拟机快照，需要“关闭”虚拟机，以便 (i) 将设备状态序列化到磁盘，以及 (ii) 关闭当前正在运行的磁盘并创建快照点。进行整合时，“关闭”虚拟机是为了关闭磁盘并将其置于适合整合的状态。”

关闭时间通常是几百毫秒；但是，如果在提交阶段有非常高的磁盘写入活动，关闭时间可能为几秒。

如果虚拟机是参与 Caché 数据库镜像的主成员或备份成员，并且关闭时间长于镜像服务质量 (QoS) 超时时间，则镜像将报告主虚拟机故障并启动镜像接管。

**2018 年 3 月更新：** 我的同事 Peter Greskoff

向我指出，在虚拟机关闭期间或主镜像成员不可用的任何其他时间，备份镜像成员可以在稍微超过半个 QoS 超时的时间内启动故障转移。

有关 QoS

注意事项和故障转移方案的详细说明，请参见以下精华帖：[镜像服务质量超时指南](#)，不过关于虚拟机关闭时间和 QoS 的简述是：

如果备份镜像在半个 QoS

超时的时间内未收到来自主镜像的任何消息，它将发送一条消息以确保主镜像仍处于活动状态。

然后，备份镜像将再等待半个 QoS 超时的时间，以接收来自主镜像的响应。

如果主镜像没有任何响应，则认为主镜像已关闭，备份镜像将接管。

在繁忙的系统上，日志不断从主镜像发送到备份镜像，并且备份镜像无需检查主镜像是否仍在运行。

但是，在安静的时间内（更可能发生备份），如果应用程序处于空闲状态，则主镜像和备份镜像之间可能在超过半个 QoS 超时的时间内没有消息。

以下是 Peter 的示例；考虑以下空闲系统时间框架，其中 QoS 超时为 :08 秒，虚拟机关闭时间为 :07 秒：

- :00 主成员使用保持连接信号 ping 仲裁器，仲裁器立即响应
- :01 备份成员向主成员发送保持连接信号，主成员立即响应
- :02
- :03 虚拟机关闭开始
- :04 主成员尝试向仲裁器发送保持连接信号，但没有发送成功，直到关闭完成
- :05 备份成员 ping 主成员，因为 QoS 已过半
- :06
- :07
- :08 仲裁器在整个 QoS 超时时间内未收到主成员的响应，因此关闭连接
- :09 备份成员没有收到主成员的响应，并与仲裁器确认连接丢失，因此开始接管
- :10 虚拟机关闭结束，太迟了！！

另请阅读上面链接的帖子中的“[配置服务质量超时的陷阱和注意事项](#)”部分，以理解让 QoS 超时只有必要长度的这一平衡。QoS 太长（尤其是超过 30 秒）也会导致问题。



2018 年 3 月更新结束：

有关镜像 QoS 的更多信息，另请参见[文档](#)。

将关闭时间保持在最低限度的策略包括：在数据库活动较少并且存储设置良好时运行备份。

如上文所述，在创建快照时，有几个选项可以指定，其中一个选项是在快照中包含内存状态 - 请记住，Caché 数据库备份不需要内存状态。如果设置了内存标志，则快照中会包含虚拟机内部状态的转储。创建内存快照需要较长时间。内存快照可用于恢复到拍摄快照时运行中的虚拟机状态。数据库文件备份并不需要。

拍摄内存快照时，虚拟机的整个状态将关闭，**关闭时间是可变的**。

如前所述，对于备份，必须针对手动快照或通过备份软件将静默标志设置为真，以保证备份的一致性和可用性。

## 查看 VMware 日志以了解关闭时间

从 ESXi 5.0 开始，快照关闭时间会记录在每个虚拟机的日志文件 (vmware.log) 中，消息类似于：

```
2017-01-04T22:15:58.846Z| vcpu-0| I125: CheckpointUnstun: vm stopped for 38123 us
```

关闭时间以微秒为单位，所以在上面的示例中，38123 us 为 38123/1000000 秒或 0.038 秒。

要确保关闭时间在可接受的限制范围内，或者要在怀疑关闭时间长而导致问题时进行故障排除，可以从您感兴趣的虚拟机文件夹中下载并查看 vmware.log 文件。下载后，您可以提取日志并进行排序，例如使用下面的 Linux 命令示例。

## 下载 vmware.log 文件的示例

下载支持日志有几种方法，包括通过 vSphere 管理控制台或从 ESXi 主机命令行创建 VMware 支持包。有关所有详细信息，请查阅 VMware 文档，但下面给出一个简单方法来创建和收集小得多的支持包，其中包括 vmware.log 文件，以便查看关闭时间。

您将需要虚拟机文件所在目录的长名称。使用 ssh 登录到运行数据库虚拟机的 ESXi 主机，并使用命令：vim-cmd vmsvc/getallvms 列出 vmx 文件以及与它们唯一关联的长名称。

例如，本帖中使用的数据库虚拟机示例的长名称输出为：26 vsan-tc2016-db1 [vsanDatastore]  
e2fe4e58-dbd1-5e79-e3e2-246e9613a6f0/vsan-tc2016-db1.vmx rhel764Guest vmx-11

接下来运行命令以只收集和捆绑日志文件：  
vm-support -a VirtualMachines:logs。

该命令将回显支持包的位置，例如：To see the files collected, check '/vmfs/volumes/datastore1 (3)/esx-esxvsan4.iscinternal.com-2016-12-30--07.19-9235879.tgz'。

您现在可以使用 sftp 将文件传输到主机，以进行进一步处理和查看。

在此示例中，解压缩支持包后，导航到与数据库虚拟机长名称相对应的路径。例如，在此例中：<bundle name>/vmfs/volumes/<host long name>/e2fe4e58-dbd1-5e79-e3e2-246e9613a6f0。

您将看到几个已编号的日志文件，最新的日志文件没有编号，即 vmware.log。该日志可能只有几百 KB，但是有很多信息，不过我们只关心关闭/取消关闭时间，使用 grep 可以轻松找到。例如：



```
$ grep Unstun vmware.log
2017-01-04T21:30:19.662Z| vcpu-0| I125: Checkpoint_Unstun: vm stopped for 1091706 us
---
2017-01-04T22:15:58.846Z| vcpu-0| I125: Checkpoint_Unstun: vm stopped for 38123 us
2017-01-04T22:15:59.573Z| vcpu-0| I125: Checkpoint_Unstun: vm stopped for 298346 us
2017-01-04T22:16:03.672Z| vcpu-0| I125: Checkpoint_Unstun: vm stopped for 301099 us
2017-01-04T22:16:06.471Z| vcpu-0| I125: Checkpoint_Unstun: vm stopped for 341616 us
2017-01-04T22:16:24.813Z| vcpu-0| I125: Checkpoint_Unstun: vm stopped for 264392 us
2017-01-04T22:16:30.921Z| vcpu-0| I125: Checkpoint_Unstun: vm stopped for 221633 us
```

我们可以看到，示例中有两组关闭时间，一组来自快照创建，另一组是删除/整合快照（例如，备份软件复制只读 vmx 文件完毕后）45 分钟后每个磁盘的关闭时间。

在上面的示例中，我们可以看到大多数关闭时间为亚秒级，尽管初始关闭时间刚刚超过一秒。

短暂的关闭时间对于最终用户来说并不明显。但是，Caché 数据库镜像等系统进程会持续监视实例是否“活动”。如果关闭时间超过镜像 QoS 超时时间，那么节点可能被视为不可联系并已“死亡”，并且将触发故障转移。

提示：要查看所有日志或进行故障排除，一个方便的命令是 grep 所有 vmware\*.log 文件，并查找关闭时间接近 QoS 超时时间的任何异常值或实例。以下命令将输出传输到 awk 进行格式化：

```
grep Unstun vmware* | awk '{ printf ("%""d", $8)} {print " ---" $0}' | sort -nr
```

---

## 总结

您应该在系统正常运行期间定期监视系统，以了解关闭时间以及它们可能对 HA（如镜像）的 QoS 超时有何影响。如前所述，将关闭/取消关闭时间保持在最低限度的策略包括：在数据库和存储活动较少并且存储设置良好时运行备份。要持续监视，可以使用 VMware Log insight 或其他工具处理日志。

我将在将来的帖子中重新讨论 InterSystems 数据平台的备份和还原操作。但现在，如果您基于您的系统工作流程有任何意见或建议，请通过下面的评论部分分享。

[#InterSystems 业务解决方案和架构](#) [#备份](#) [#提示和技巧](#) [#系统管理](#) [#部署](#) [#镜像](#) [#Caché](#) [#InterSystems IRIS](#)  
[#InterSystems IRIS for Health](#) [#文档](#)

---

### 源

URL:

<https://cn.community.intersystems.com/post/intersystems-%E6%9C%80%E4%BD%B3%E5%AE%9E%E8%B7%B5%E7%B3%BB%E5%88%97%E4%B9%8B-%E6%95%B0%E6%8D%AE%E5%B9%B3%E5%8F%B0%E5%92%8C%E6%80%A7%E8%83%BD-%E2%80%93%E8%99%9A%E6%8B%9F%E6%9C%BA%E5%A4%87%E4%BB%BD%E5%92%8C-cach%C3%A9-%E5%86%BB%E7%BB%93%E8%A7%A3%E5%86%BB%E8%84%9A%E6%9C%AC>