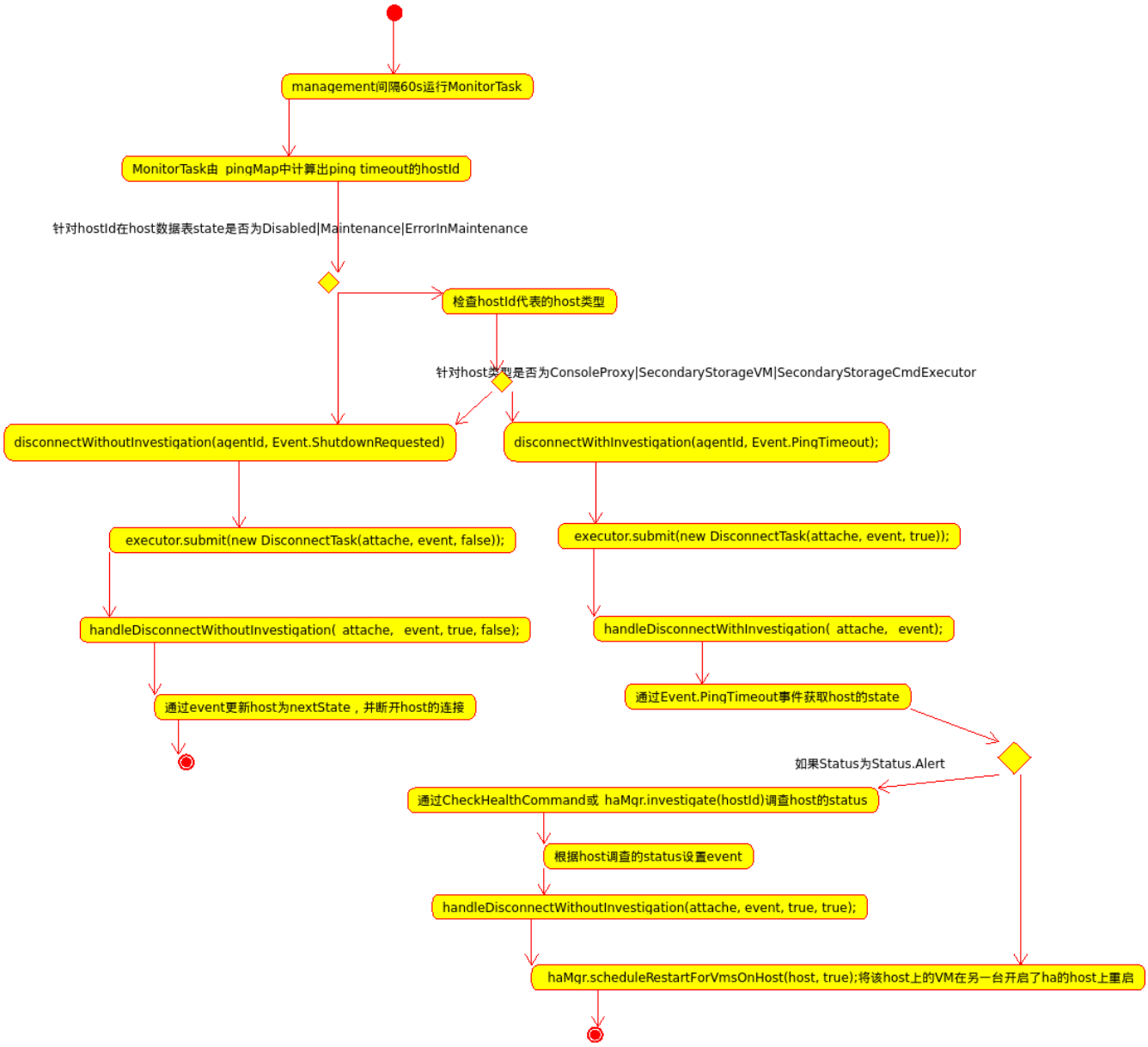
cloudstack 对 VM 的高可用有两种方式

（1）VM ha：由 cloudstack management 间隔发送 pingTask 收集 host 机器上 VM 的 state 信息，更新数据库表 vm_instance 的 power_state 字段，之后通过消息通知机制 MessageBusBase 发行该主题，订阅者比较数据库表 vm_instance 的 power_state 和 state 字段，若果 power_state 为 PowerOff 或 PowerReportMissing，而 state 为 Starting|Stopping|Running|Stopped|Migrating 中一种，则调用 HighAvailabilityManagerImpl 的 scheduleRestart(VMInstanceVO vm, boolean investigate)方法来重启 VM。

图 1、VM ha 执行流程图

（2）host ha：由 cloudstack management 间隔发送 MonitorTask 对 ping 超时的 host，查询其状态，如果状态不为 Disabled，Maintenance，ErrorInMaintenance（异常断开连接），调用 HighAvailabilityManagerImpl 的 scheduleRestartForVmsOnHost(final HostVO host, boolean investigate)来重启该 host 上的所有 VM。



AgentManagerImpl 类中有一个
`private final ConcurrentHashMap<Long, Long> _pingMap = new ConcurrentHashMap<Long, Long>(10007);`

在处理 host 的主动连接时，对于 host 类型不为 TrafficMonitor 和 SecondaryStorage 的 host _pingMap 中保存着 hostid 和连接时的当前时间

`_pingMap.put(host.getId(), InaccurateClock.getTimeInSeconds());`

在处理 host 的主动断开连接时

`_pingMap.remove(agentId);`

////

在 AgentManagerImpl 组件启动时(start 方法)

```
_monitorExecutor.scheduleWithFixedDelay(new MonitorTask(), PingInterval.value(),
PingInterval.value(), TimeUnit.SECONDS);
```

_monitorExecutor 间隔指定时间调度执行 MonitorTask 任务

MonitorTask 任务内容为

（1）根据当前时间计算出需要进行 ping 检测的 hostid

```
List<Long> agentsBehind = new ArrayList<Long>();
long cutoffTime = InaccurateClock.getTimeInSeconds() - getTimeout();
for (Map.Entry<Long, Long> entry : _pingMap.entrySet()) {
  if (entry.getValue() < cutoffTime) {
      agentsBehind.add(entry.getKey());
      }
  }

if (agentsBehind.size() > 0) {
    s_logger.info("Found the following agents behind on ping: " + agentsBehind);
    }

return agentsBehind;
```

（2）对于每一个需要进行 ping 检测的 hostid，从 host 数据库表查询其 state

2.1 若状态为 Disabled，Maintenance，ErrorInMaintenance ，则断开 host 的连接

```
if (resourceState == ResourceState.Disabled || resourceState ==
ResourceState.Maintenance || resourceState == ResourceState.ErrorInMaintenance) {
                            /*
                             * Host is in non-operation state, so no
                             * investigation and direct put agent to
                             * Disconnected
                             */
 status_logger.debug("Ping timeout but host " + agentId + " is in resource state of
" + resourceState + ", so no investigation");
   disconnectWithoutInvestigation(agentId, Event.ShutdownRequested);
}
```

2.2 如果 state 不为上述三种（异常断开连接），但 host 类型为 ConsoleProxy，SecondaryStorageVM，SecondaryStorageCmdExecutor，不调查检测，关闭连接

2.3 对 host 调查检测，并发出告警事件。并对 host 上 VM 生成调用 HighAvailabilityManagerImpl 的 scheduleRestartForVmsOnHost 来重启 host 上的 VM，scheduleRestartForVmsOnHost 对每一个 VM 调用 scheduleRestart。

#######

HighAvailabilityManagerImpl 的 scheduleRestart 会在数据库表 op_ha_work 中插入一条 VM 的 ha 类型的记录，然后唤醒 WorkerThread 来处理。

HighAvailabilityManagerImpl 组件在 configure 根据数据库表 configuration 的记录 ha.workers 的值（5）来构建 5 个 WorkerThread。WorkerThread 取 op_ha_work 的记录，对 WorkType.HA 类型的记录执行 restart(work)该函数会重启 VM。