УДК 621.311.1

Авторы:
Кашеманов А.В.,
Прокопьев В.В.,
Чернов В.К.,
ООО НПП «ЭКРА»,
г. Чебоксары, Россия.

Kashemanov A.V.,
Prokopiev V.V.,
Chernov V.K.,
EKRA Research and Production Enterprise LTD,
Cheboksary, Russia.

ПОСТРОЕНИЕ ОТКАЗОУСТОЙЧИВОГО КЛАСТЕРА ПТК ЦПС НА ПЛАТФОРМЕ ВИРТУАЛИЗАЦИИ

BUILDING A FAULT TOLERANCE CLUSTER ON THE VIRTUALIZATION PLATFORM FOR DIGITAL SUBSTATION

Аннотация: рассмотрена реализация кластерной системы виртуализации для сегмента серверов уровня подстанции. Описаны особенности работы кластера, ее архитектура и организация взаимодействия между устройствами в системе.

Ключевые слова: виртуализация, кластер, виртуальная машина, контейнер, АСУ ТП, гипервизор, отказоустойчивость, SCADA.

Abstract: the implementation of a cluster virtualization system for the substation-level server segment is considered. The article describes the features of the cluster operation, its architecture and the organization of interaction between devices in the system.

Keywords: virtualization, cluster, virtual machine, container, APCS, hypervisor, fault tolerance, SCADA.

Введение

Сервера АСУ ТП/ССПИ, сервера и автоматизированное рабочее место (АРМ) SCADA, входящие в состав программно-технических средств (ПТК) цифровых подстанций (ЦПС) [1], в общем случае представлены в виде отдельных физических машин (рис. 1). Каждая такая машина имеет набор программного обеспечения (ПО) для реализации определенных функций (сбор, передача, преобразование, отображение, сохранение информации и т.п.) и полностью использует выделенные ей для этого аппаратные ресурсы. Отказоустойчивость такого решения обеспечивается за счет использования аппаратного резервирования машин, что соответственно удваивает количество элементов и связей в системе, приводит к ее усложнению.

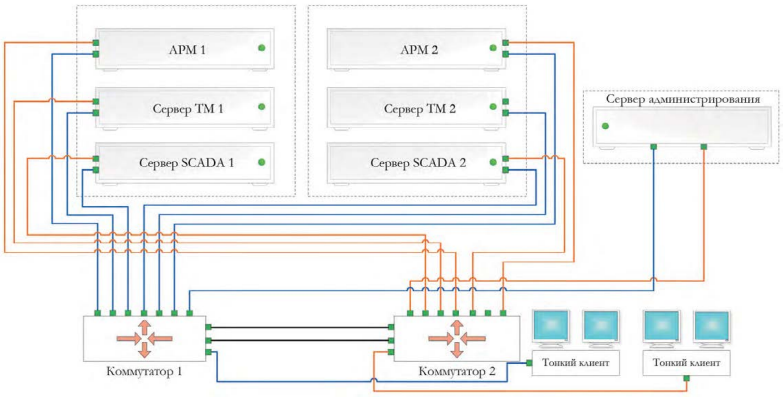


Рис. 1. Общая структурная схема ПТК ЦПС, состоящая из отдельных серверов

Иной подход состоит в применении виртуализации в кластере. Под кластером виртуализации подразумевается целостная система из физических серверов, которая объединена между собой сетью (рис. 1) и управляется через единую платформу виртуализации. Данная платформа, обычно выступающая в качестве ПО, необходима для централизованного управления кластером. Согласно [2], уровень подстанции должен быть образован серверами, объединенными в отказоустойчивый кластер, на платформе виртуализации которого работают сервера и АРМ. Виртуализация – группа технологий, основанных на преобразовании формата или параметров программных или сетевых запросов к компьютерным ресурсам с целью обеспечения независимости процессов обработки информации от программной или аппаратной платформы информационной системы [3]. Другими словами, в пределах одной физической машины можно развернуть несколько виртуальных систем, которые логически изолированы друг от друга средствами процессора (технологии Intel VT, AMD-V). При этом виртуальная система имеет свой логический набор ресурсов (процессорных, оперативной памяти, устройств хранения и др.), что делает возможным свободный перенос ее на другую машину. Предоставление ресурсов из общего пула аппаратных ресурсов, свободного на оборудовании, и изоляцию опе-

Рис.2. Структура основных типов виртуализации

рационной системы (ОС) осуществляет гипервизор (hypervisor). Гипервизор в среде виртуализации бывает двух типов:

I тип (рис. 2 а). Устанавливается на сервер без установленной предварительно ОС и предоставляет виртуальным системам или гостевым ОС, запущенным под его управлением, службу виртуальной машины (ВМ).

II тип (рис. 2 б). Программный гипервизор, устанавливается поверх основной ОС (чаще всего это Linux).

Следовательно, используя гипервизор, можно развернуть несколько ВМ с гостевой ОС Linux или Windows, к примеру. Гостевыми обычно называют ОС, работающие именно внутри ВМ.

Кроме приведенных двух типов имеется «метод виртуализации», реализуемый на уровне ядра ОС и использующий для изоляции не аппаратные ресурсы, а ресурсы операционной системы, так называемое пространство имен. Такой подход называется «контейнеризацией».

Каждый контейнер представляет собой исполняемый пакет ПО, работающий с основной ОС и требующий минимальных ресурсов для запуска. При этом контейнеры, как и ВМ, логически изолированы друг от друга (рис. 2 в). Использование однородной среды не требует дополнительных расходов на эмуляцию виртуального оборудования и развертывание полноценной гостевой ОС. Примером систем контейнеризации может служить LXC (Linux Containers) и Docker.

С целью повышения отказоустойчивости, надежности и степени использования оборудования целесообразно функции сервера телемеханики (ТМ), SCADA-системы и АРМ реализовывать в виде кластера виртуализации. Отказоустойчивость кластера виртуализации основана на возможности минимизации времени перемещения и последующего запуска экземпляра ВМ/контейнера на исправном хосте в случае аварийного отказа другого. При этом стоит учитывать, что при перемещении ВМ на другой хост необходимо обеспечить аппаратный резерв для корректной работы всех ВМ, что требует использования более высокопроизводительных серверов.

Архитектура кластерной системы

Указанные выше типы виртуализации требуют минимум три физических узла для функционирования системы, взаимосвязь между которыми обеспечивается через коммутаторы (рис. 3).

Два сервера, на которых располагаются ВМ/контейнеры, резервируют друг друга в случае неисправности. Оценка аварийного отказа возложена на третий сервер, в концепции виртуализации называемый «свидетель» (witness, quorum), основная задача которого состоит в определении вышедшего из строя узла и перераспределении ресурсов на оставшийся узел. Стоит отметить, что SCADA-сервера можно разверт-

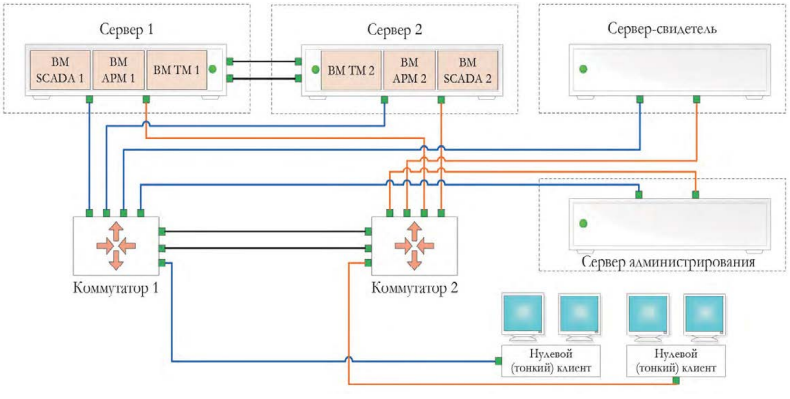
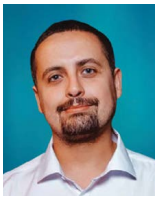


Рис. 3. Обобщенная структурная схема кластера



Кашеманов Антон Владимирович
В 2011 г. окончил ЧГУ им. И.Н. Ульянова по специальности «Релейная защита и автоматизация электроэнергетических систем».
Руководитель продуктового направления АСУ ТП центра инжиниринга АСУ ООО НПП «ЭКРА».

Прокопьев Вадим Валентинович
В 2004 г. окончил ЧГУ им. И.Н. Ульянова, магистр техники и технологий.
Руководитель направления АСУ ТП центра инжиниринга АСУ ООО НПП «ЭКРА».

Чернов
Владимир Константинович
В 2021 г. окончил ЧГУ им. И.Н. Ульянова, магистр техники и технологий.
Инженер 3 кат. продуктового направления АСУ ТП центра инжиниринга АСУ ООО НПП «ЭКРА».

нуть внутри LXC-контейнеров вместо ВМ, используя механизм контейнеризации [4]. LXC-контейнер, он же Linux контейнер, не нуждается в полноценной ОС, а использует ядро системы. Внутри контейнера устанавливаются все необходимые «службы» для корректной работы SCADA-системы и задаются его сетевые параметры.

Подключение к ВМ осуществляется через нулевого или тонкого клиента. Связь нулевого клиента и ВМ реализуется через проприетарный протокол PCoIP, а связь с тонким клиентом через протокол RDP. Главное отличие нулевого клиента от тонкого клиента состоит в том, что нулевой клиент декодирует изображение аппаратно, а тонкий клиент с помощью программных средств. Это связано с тем, что у тонкого клиента имеется своя ОС, оперативная память и процессор общего назначения, у нулевого клиента же нет своей ОС. Наличие ПО делает тонкого клиента более универсальным, но менее производительным.

Управление всей системой виртуализации, а также дальнейшая ее модификация осуществляется через веб-интерфейс браузера с помощью централизованной платформы управления. Доступ через веб-интерфейс браузера является проблемой с точки зрения информационной безопасности. Проблему неограниченного доступа к кластеру можно решить добавлением сервера администрирования системы виртуализации и дополнительно разграничить сеть на VLAN.

Организация взаимодействия между устройствами в кластере осуществляется посредством коммутаторов, назначение которых – передать трафик управления между узлами и диагностическую информацию о состояния двух серверов узлу-свидетелю. Коммутаторы аппаратно резервируют друг друга. Прямое соединение между хостами, цель которого передать файлы ВМ и контейнеры, требует высокой пропускной способности канала связи. Для этого применяется протокол агрегирования (LACP), где скорость каждого соединения составляет 10 Гб/с.

Все ВМ и контейнеры в ПТК ЦПС используют одну систему хранения данных (СХД), состоящую из дисковых накопителей самих узлов. Цель организации СХД в кластере состоит в обеспечении доступности серверов к ВМ (контейнерам) в случае аварийных неисправностей. СХД платформы виртуализации в данном случае строится на технологии vSAN (программная СХД) или с применением распределенного блочного устройства DRBD.

СХД vSAN

Технология vSAN [5] позволяет объединить локальные диски нескольких физических устройств в общее виртуальное хранилище, решая проблему выделения отдельного сервера для этих целей. Для организации общего хранилища на основных серверах может использоваться одна или несколько дисковых групп с несколькими HDD накопителями и, как минимум, одним SSD диском для возможности кэширования часто запрашиваемых данных. ПО гипервизора устанавливается на отдельный накопитель, а объем остальных полностью выделяется под общее хранилище. Дисковое пространство в составе узла-свидетеля, состоящее из одного SSD и одного HDD накопителей, также должно быть добавлено в структуру vSAN, но не для организации общего СХД, а для хранения метаданных (структура, состояние кластера и т.п.).

Схема vSAN для двух узлов использует механизм отказоустойчивости по технологии RAID-1, т.е. полное зеркалирование объекта с размещением копии на другом сервере. У каждой ВМ имеется свой параметр, который определяет количество доступных отказов (failures to tolerate - FTT). При FTT равном 1, создается одна реплика (копия) ВМ, и пространство, занимаемое ею, становится в два раза больше в сравнении с полезной ёмкостью. Соответственно, чем больше FFT, тем больше реплик хранятся в хранилище. В кластере с двумя узлами достаточно одной реплики ВМ. К примеру, при использовании трех локальных HDD объемом по 300 Гб и одного SSD объемом 480 Гб в виртуальное хранилище будет добавлено 600 Гб памяти с одного сервера. Наличие идентичных дисковых устройств второго сервера добавит к СХД еще 600 Гб, и общий объем хранилища составит 1.2 Тб, но при использовании RAID-1 полезный объем будет уменьшен в два раза.

СХД на основе DRBD

DRBD – это распределённые реплицируемые блочные устройства, предназначенные для построения отказоустойчивых кла-

стерных систем на ОС Linux. DRBD занимается полным отражением по сети всех операций. Можно считать, что это сетевой RAID-1 [6]. Как и в случае со vSAN, DRBD использует несколько накопителей для организации СХД и один для установки ПО платформы. Свободное дисковое пространство каждого из узлов делится на одинаковое количество блочных устройств, равное числу ВМ (контейнеров) в пределах одного сервера, т.е. сколько ВМ – столько и блочных устройств. К примеру, при наличии 600 Гб объема памяти можно создать 3 ресурса по 200 Гб для ВМ АРМ, SCADA и ТМ, при этом объем ресурсов в случае необходимости можно уменьшить или увеличить.

DRBD обычно находится в двух состояниях: первичный и вторичный (запасной). Сервер, на котором устройство находится в первичном состоянии, будет производить операции записи и чтения с файловой системой ВМ (контейнера) и отсылать все произведенные изменения на вторичное устройство. В случае выхода из строя первичного узла запасной перейдет в активное состояние и с помощью технологии отказоустойчивого резервирования запустит ВМ. После восстановления работы отказавшего узла DRBD-устройства начнут синхронизироваться между собой. Задача управления и наблюдения за работоспособным состоянием ресурсов возложена на linstor-controller, который устанавливается в качестве LXC-контейнера в отдельное DRBD-устройство. Для того чтобы linstor-controller мог наблюдать за ресурсами хостов, необходимо установить linstor-satellite на узлах, тогда система будет выявлять сбой в работе ресурса DRBD.

Непосредственно резервирование ВМ и контейнеров реализуется с помощью технологий отказоустойчивости (high availability - HA). Основное назначение данной технологии состоит в обнаружении неисправностей кластера, дальнейшей миграции и перезапуске ВМ с целью повышения степени использования оборудования и сокращения простоев работы приложений внутри ВМ. Время восстановления составляет от нескольких секунд до нескольких минут в зависимости от используемой гостевой ОС, СХД и платформы виртуализации.

Для виртуализации с СХД vSAN можно использовать также технологию непрерывного резервирования ВМ (Fault Tolerance), при которой копия ВМ находится в режиме ожидания, постоянно обмениваясь данными с основной машиной. После отказа одного сервера копия моментально запускается на другом. Обычно время «включения» составляет несколько секунд. Но за наличие быстрого отклика на неисправность (авария, отключение питания сервера и т.п.) приходится расплачиваться отсутствием возможности создания снимков (snapshots) ВМ, использованием всего одного виртуального процессора и некоторыми другими ограничениями [7]. Такую технологию стоит использовать для приложений с требованием постоянной доступности.

Анализ времени восстановления и переключения серверов

Произведем оценку времени восстановления серверов для всех типов виртуализации в сравнении с отдельными физическими машинами. Для оценки времени восстановления работы ВМ был выполнен внезапный отказ одного из физических серверов.

Для оценки времени восстановления при отказоустойчивом резервировании с сервера свидетеля посылаются ICMP-запросы до ВМ АРМ и контейнера SCADA-сервера. Так как контейнер более «легковесен», то он запускается быстрее, чем ВМ на гостевой ОС Windows. Перед изоляцией неисправного сервера и переноса работы ВМ платформа виртуализации ожидает 120 с, после этого перемещает машину и восстанавливает ее работу. На рис. 4 показан график работы при отказоустойчивом резервировании, где красной линией показано время восстановления сервера для I и II типа резервирования, синей линией отображается время восстановления сервера при использовании контейнера в кластере (по оси ординат показано количество передаваемой информации в байтах за промежуток времени в 10 с). Суммарное время перезапуска ВМ составляет около 180 с, а контейнера – 165 с. Как уже было упомянуто, контейнер использует ОС гипервизора в качестве основной и не требует установки гостевой ОС, которая и требует дополнительное время для запуска.

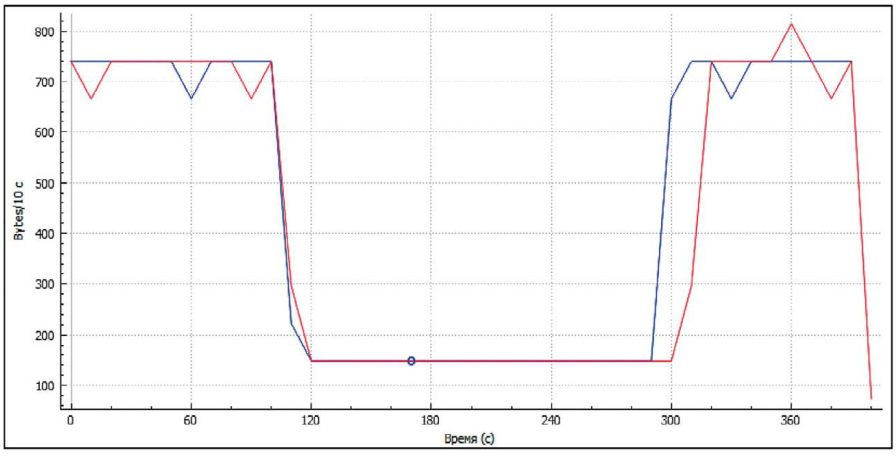


Рис. 4. Отказоустойчивое резервирование ВМ и контейнера в кластере

Рис. 5. Непрерывное резервирование ВМ

Тестирование работы непрерывного резервирования ВМ осуществлялось путем отключения по питанию одного из серверов. Для измерения времени задержки при переключении работы ВМ в локальную сеть был добавлен испытательный комплекс Omicron CMC356 и контроллер присоединения (КП) ЭКРА 243. С испытательного комплекса подавался ступенчатый пилообразный сигнал до КП с шагом 0,1 с, который затем посредством MMS-сообщений передавал телеинформацию на ВМ с EKRASCADA [8]. Время переключения фиксировалось посредством просмотра значений кривой на графике EKRASCADA ARM (рис. 5). Аварийное отключение узла приводит к моментальному перезапуску ВМ, время отключения и включения по графику составляет соответственно 09 ч 15 мин 59 с, 09 ч 16 мин 00 с, задержка переключения равна 1,4 с.

В дополнение на платформе виртуализации имеется возможность настроить отправку SNMP Trap'а до одного из серверов SCADA в случае изменения состояния работы ВМ и места ее размещения. При отключении ВМ, работающей по технологии непрерывного резервирования, платформа виртуализации сгенерировала и отправила SNMP Trap, показывающий изменения статуса ВМ (рис. 6). Такие события позволяют диспетчеру оценивать состояние работоспособности ВМ.

Тестирование работы непрерывного резервирования контейнера в кластере с помощью SCADA-системы выполнить не представляется возможным, так как в контейнере отсутствует ОС, необходимая для разворачивания SCADA-системы.

В случае аппаратного резервирования при отключении патч-кордов EKRASCADA не заметила потерь сигналов, как показано на рис. 7. Связано это с тем, что в EKRASCADA сбор данных выполняется обоими серверами одновременно и оба сервера синхронизируют между собой базы данных. При потере сервера, к которому был подключен АРМ, АРМ переключается на соседний сервер и подкачивает данные с работающего сервера.

На рис. 8 показан журнал событий, где в столбце статус события имеется сообщение о недоступности основного сервера АСУ ТП.

Стоит отметить, что испытания для проверки времени переключения в кластере виртуализации проводились с одной ВМ, на которой был запущен сервер SCADA. При наличии двух ВМ время переключения будет нулевым, как и при аппаратном резервировании, но с одним весомым отличием: две ВМ SCADA все равно будут доступны на одном физическом сервере. Следовательно, имеется еще одна точка отказа при работе серверов SCADA.

Таким образом, по сравнению с решением, в котором все функции системы возложены на отдельные сервера, виртуализация позволяет увеличить непрерывность работы ВМ за счет переноса выполняемых процессов автоматически на другой физический узел. То есть все ВМ/контейнеры в лице основных и резерв-

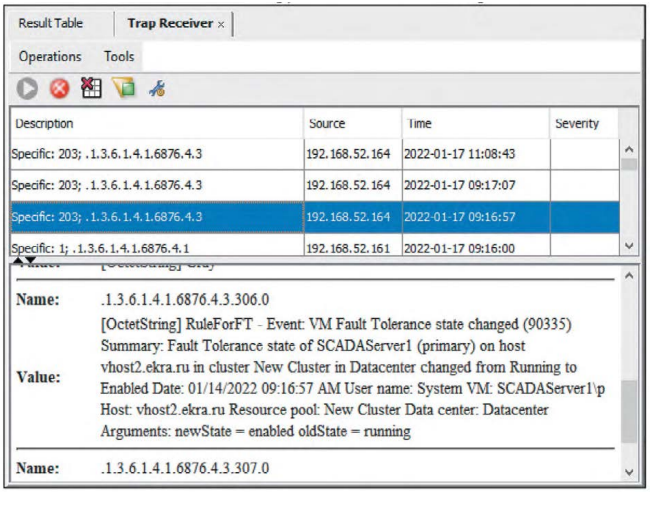


Рис. 6. SNMP Trap, показывающий изменение статуса работы ВМ

ных АРМ, SCADA-серверов и серверов ТМ будут работать на одном узле в случае аварии другого. Стоит заметить, что если выйдет из строя основной узел до восстановления другого узла, то все ВМ приостановят свою работу, так как узел для миграции отсутствует. Другими словами, имеется всего лишь одна аппаратная точка для отказа. После возобновления работы узлов все ВМ/контейнеры восстановятся, но данные на время будут потеряны. Поэтому необходимо постоянно производить мониторинг работы кластера и его компонентов (ВМ, СХД, платформа виртуализации и т.п). Избежать такой проблемы можно путем добавления еще одного сервера виртуализации, что позволит увеличить отказоустойчивость системы с одной стороны, но с другой стороны увеличит стоимость всей системы.

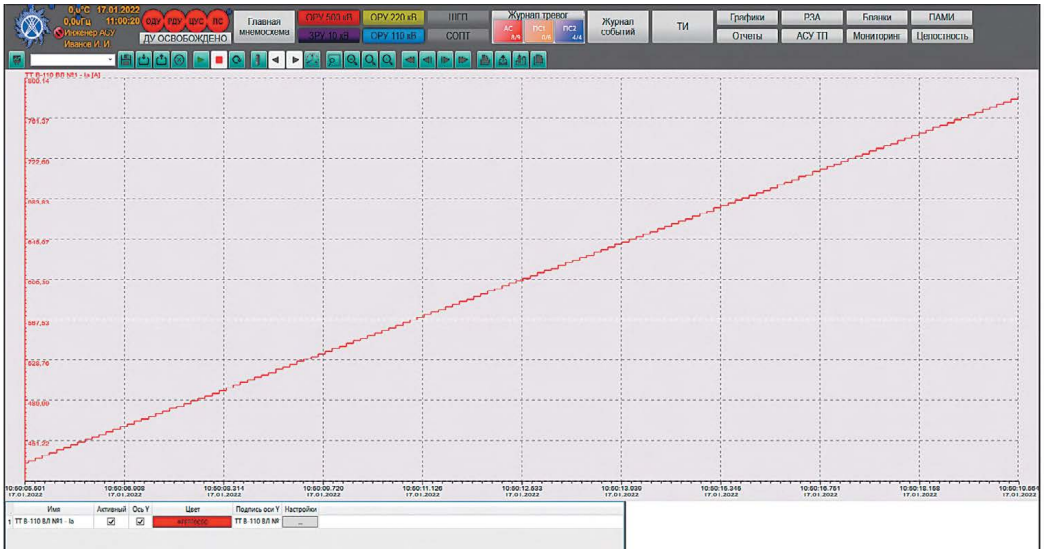


Рис. 7. Непрерывное аппаратное резервирование

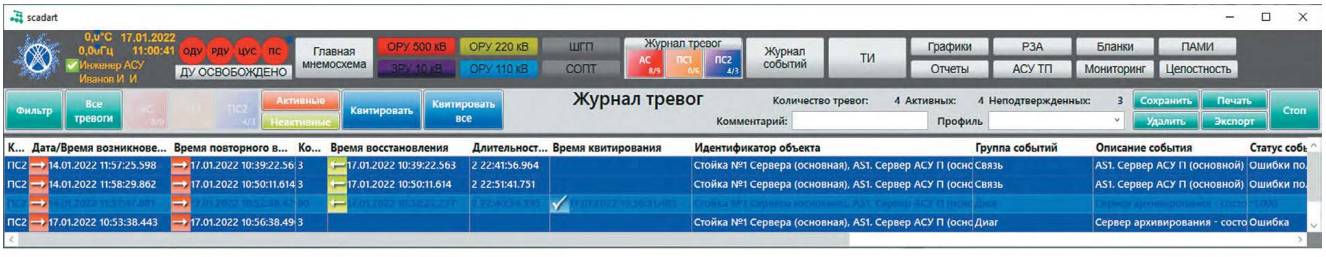


Рис. 8. Непрерывное аппаратное резервирование. Журнал событий

Выводы

ПТК ЦПС на основе технологий виртуализации позволяет уменьшить количество оборудования за счет переноса выполняемых функций в ВМ и контейнеры, что повышает отказоустойчивость и надежность системы. Виртуализация также позволяет облегчить масштабируемость системы, которая достигается за счет установки под определенные задачи дополнительной ВМ без необходимости приобретения нового полноценного физического сервера. Эффективность использования ресурсов имеющихся серверов увеличивается ввиду использования нескольких ВМ различных гостевых ОС на единой аппаратной платформе с возможностью распределения ресурсов по загруженности CPU, RAM, сетевого трафика и дискового пространства. Использование кластера виртуализации исключает единую точка отказа, обеспечивает непрерывную работу системы при отказе аппаратной или программной части.

Литература:

1. СТО 56947007-29.240.10.302-2020 Типовые технические требования к организации и производительности технологических ЛВС в АСУ ТП ПС ЕНЭС. – ПАО «ФСК ЕЭС»: 26.02.2020. – 94 с.

2. СТО 34.01-21-004-2019. Цифровой питающий центр. Требования к технологическому проектированию цифровых подстанций напряжением 110-220 кВ и узловых цифровых подстанций напряжением 35 кВ. – ПАО «Россети»: 29.03.2019. – 114 с.

3. ГОСТ Р 56938-2016 Защита информации. Защита информации при использовании технологий виртуализации. Общие положения. – Введ. с 01.06.2017. - М.: Стандартинформ, 2018. – 28 с.

4. Proxmov VE Administration Guide. Release 6.2. – Proxmox Server Solutions Gmbh: 2020 – 416 p.

5. vSAN 2 Node Guide. – VMware: 2020. – 192 p.

6. Игорь Чубин. DRBD / Чубин Игорь // xgu.ru/wiki. – 2016. – URL: http://xgu.ru/wiki/DRBD (дата обращения: 12.01.2022).

7. vSphere Features Not Supported with Fault Tolerance // VMware Docs. – 2020. – URL: https://docs.vmware.com/en/VMware-vSphere/7.0/com.vmware.vsphere.avail.doc/GUID-F5264795-11DA-4242-B774-8C3450997033.html (дата обращения: 12.01.2022).

8. Техническая информация на EKRASCADA // soft.ekra.ru. – 2020. – URL: https://soft.ekra.ru/ekrascada/ru/downloads/documents (дата обращения 12.01.2022).