Игорь Афонин

# Решение для обеспечения постоянной доступности информационных систем

**В статье даётся обзор уровней доступности информационных систем и способов их достижения. Показана общая методика определения необходимого уровня отказоустойчивости. Представлено решение AdvantiX для обеспечения постоянной готовности, описана его архитектура и ключевые особенности.**

## Введение

Высокая доступность информационной системы (ИС) управления является ключевым требованием обеспечения непрерывности функционирования современного производства. Необходимый коэффициент готовности системы зависит от специфики производственных процессов и определяется прежде всего максимально допустимым временем простоя, а также связанным с ним нанесённым ущербом и возможными последствиями. Химическое производство, система безопасности аэропорта, управление транспортом и финансовые биржи требуют максимального уровня доступности системы, а для самых критичных приложений обеспечения производственного процесса розничного магазина или склада достаточно значительно более низкого уровня, хотя в этом случае всё зависит от масштаба последствий, вызванных простоем.

Попробуем разобраться в том, что такое уровень доступности и какими средствами он обеспечивается.

## Доступность: цели и метрики

Информационные технологии обеспечивают необходимый уровень доступности автоматизированных систем управления производством. Высокая доступность является стратегической необходимостью для многих задач автоматизации. Но не все задачи имеют одинаковый приоритет, поэтому важно отметить различие преимуществ и стандартов высокой доступности.

*Высокая доступность* (*High Availability, HA*) означает, что приложения остаются доступными в течение очень значительной части общего времени работы и способны быстро восстанавливаться после локализованных сбоев. Пользователи видят короткое прерывание работы или в некоторых случаях вообще не замечают этого.

*Постоянная доступность* (*Fault Tolerance*) – это наивысший уровень доступности и означает непрерывную доступность сервисов, невзирая на ошибки приложения и сбои оборудования. Постоянная доступность (устойчивость к отказам) не допускает простоев и связанных с ними потерь данных и транзакций.

*Аварийное восстановление* (*Disaster Recovery*) – восстановление ИТ-системы после региональных катастроф, таких как перебои электропитания, наводнения, землетрясения и другие катаклизмы.

Новейшие технологии, с одной стороны, значительно снижают расходы на ИТ, но с другой стороны, повышение уровня доступности значительно увеличивает количество требуемых ресурсов, таких как аппаратное и программное обеспечение, время и средства на развёртывание и поддержание системы в работоспособном состоянии и, в конечном итоге, денежные средства, необходимые для этого. Обеспечить постоянную доступность для всех задач во всевозможных ситуациях нереально и слишком дорого. Наилучшим подходом для обеспечения доступности реальных задач будет расчёт трёх основных метрик:

- время восстановления RTO (Recovery Time Objective) – время, в течение которого приложение должно быть восстановлено, или максимальное время недоступности сервиса;
- точка восстановления RPO (Recovery Point Objective) определяет максимально допустимые потери данных и транзакций;
- совокупная стоимость владения TCO (Total Cost of Ownership) – прямые и косвенные затраты, связанные с достижением параметров RTO и RPO в течение срока жизни приложения.

## Уровни доступности

После того как установлены допустимая цена простоя и связанные с ним метрики, можно выбрать соответствующий уровень доступности – Availability Level, или AL (рис. 1, уровни доступности указаны по аналитическим отчётам IDC). Стоит отметить, что существуют два вида показателей доступности: вероятность незапланированных простоев или доступность от общего време-

ни (выраженное в девятках рабочее время). Обе метрики широко используются, и возможно применение разных метрик для различных задач.

Итак, что собой представляют уровни доступности?

*Высоконадёжный уровень* (*Reliable*) – начальный уровень. Многие приложения не требуют защиты или просто используется оборудование с аппаратным резервированием и возможностью «горячей» замены. Требуется восстановление, но процесс управления не зависит от этих приложений.

*Восстанавливаемый уровень* (*Recoverable*) – резервируются некоторые компоненты инфраструктуры, такие как Web-сервисы, серверы DNS и Active Directory с автоматическим восстановлением после сбоев. Но в ряде случаев и для этих сервисов простой недопустим.

*Высокая доступность* (*Highly Available*) – уровень доступности для ERP-систем, баз данных, почтовых и других сервисов, которые обеспечивают производственные процессы. Когда сервисы становятся недоступными, возможна потеря данных, что может существенно сказаться на цене простоя. Для этого уровня необходим расчёт метрик RTO и RPO.

*Постоянная доступность* (*Continuously Available*) требуется для небольшого количества задач, когда недопустимо малейшее время простоя, таких как критически важные SCADA и MES-системы, приложения для транспорта, безопасности, трейдинговых площадок и банков.

## Обеспечение высокой доступности

Начальный уровень Reliable обеспечивается использованием серверов с резервированием компонентов.

Уровень Recoverable может быть достигнут путём использования резервных серверов с асинхронной репликацией данных. При отказе сервера необходимый сервис запускается на резервном и он становиться активным. Всё происходит на уровне приложения, и дополнительное ПО не требуется.

Для более высоких уровней необходимо дополнительное ПО, так называемое программное обеспечение промежуточного слоя (ПО ПС), и некоторая инфраструктура. ПО ПС играет важную роль в обеспечении высокой доступности приложений с полностью прозрачным для пользователей переключением на резервные ресурсы.


**Рис. 1. Уровни доступности**

Основные свойства ПО ПС, существенные для обеспечения высокой доступности:

- берёт на себя часть функций, которые локально выполняются операционной системой;
- осуществляет маршрутизацию запросов, позволяя тем самым обеспечить живучесть прозрачным для пользователей образом;
- осуществляет балансировку загрузки вычислительных мощностей, что также способствует повышению доступности данных;
- осуществляет тиражирование любой информации, а не только содержимого дисков, делая приложение устойчивым к отказам серверов;
- отслеживает состояние приложений и при необходимости тиражирует и перезапускает программы, что гарантирует живучесть программных систем;
- даёт возможность прозрачным для пользователей образом выполнять переконфигурирование (и, в частности, наращивание) серверных компонентов, что позволяет масштабировать систему, сохраняя инвестиции в прикладные системы.

Стабильность прикладных систем – важный фактор повышения доступности данных.

## Отказоустойчивый кластер

*Отказоустойчивый кластер* (*Failover Cluster*) широко используется для уровня Highly Available и представляет собой решение, которое сочетает в себе два или более серверов, общее дисковое хранилище и специальное программное обеспечение для автоматического реагирования на отказы оборудования и восстановления после сбоя (рис. 2). Серверы в кластере общаются друг с другом постоянно, проверяя так называемое сердцебиение (Heartbeat), которое подтверждает, что другие серверы в кластере работают. Если один из узлов кластера выходит из строя, другой автоматически принимает на себя его задачи и обеспечивает доступность ресурсов. Кластеризация часто требует специализированных экспертиз для развёртывания решения и поддержания его в работоспособном состоянии. Кроме того, кластеры для доступа к данным используют общее дисковое пространство (Shared Storage), которое может стать единой точкой отказа и потенциальным источником простоя.

Следует отметить, что кластеры ориентированы прежде всего на высокую производительность, балансировку нагрузки и имеют большие возможности по масштабированию. Как было отмечено ранее, они сложны в установке и эксплуатации. Для исключения единой точки отказа необходимо создавать специальную инфраструктуру и прежде всего – сеть хранения данных SAN (Storage Area Network).

По данным аналитических агентств (Aberdeen Group, Analyst Insight, June 2013), отказоустойчивые кластеры с использованием кластеризации имеют время незапланированного простоя 4,38 часа в год и обеспечивают доступность на уровне 99,95%.

## Решение AdvantiX Intellect FT

В статье [1] было рассмотрено решение AdvantiX Intellect уровня HA на основе программного обеспечения Stratus Avance, которое использует синхронизацию данных между двумя узлами и проактивный мониторинг параметров системы на уровне BMC (Baseboard

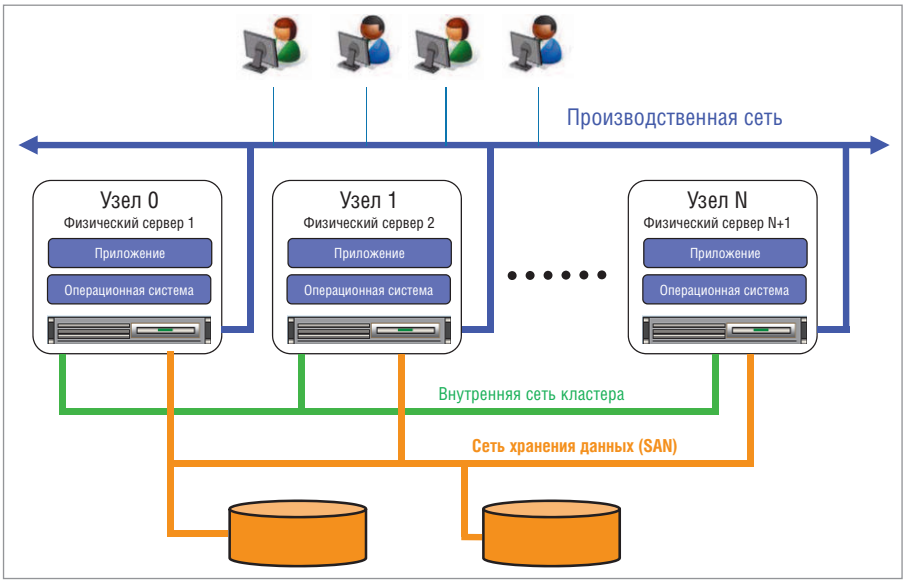


**Рис. 2. Отказоустойчивый кластер**

Management Controller) контроллера управления системной платы, что позволило уменьшить время простоя до менее чем 50 минут в год и довести степень доступности системы до 99,99%+, что означает немного больше, чем 99,99%.

Как было указанно ранее, для некоторых критически важных задач недопустимо даже малейшее время простоя.

В качестве решения задач уровня Fault Tolerance предлагается решение AdvantiX Intellect FT, которое поддерживает постоянную доступность и обеспечивает непрерывность бизнес-процессов и целостность данных. Решение обладает уровнем доступности 99,999% и позволяет уменьшить расчётное время простоя до 5,25 минут в год, а на практике избежать его вовсе.

Решение содержит следующие компоненты (рис. 3):

- два сервера Advantix архитектуры x86 с поддержкой аппаратной виртуализации;
- внутренние (межузловые) и внешние (производственные) сетевые соединения;
- виртуализация, использующая CentOS и гипервизор KVM (KVM Virtualization);
- программное обеспечение промежуточного слоя Stratus everRun Enterprise для обеспечения отказоустойчивости (Availability Engine).

В отличие от других решений Stratus everRun Enterprise предотвращает простои, а не только выполняет восстановление после сбоев. Это преимущество решения оказывает большое влияние на безопасность и финансовые показатели.

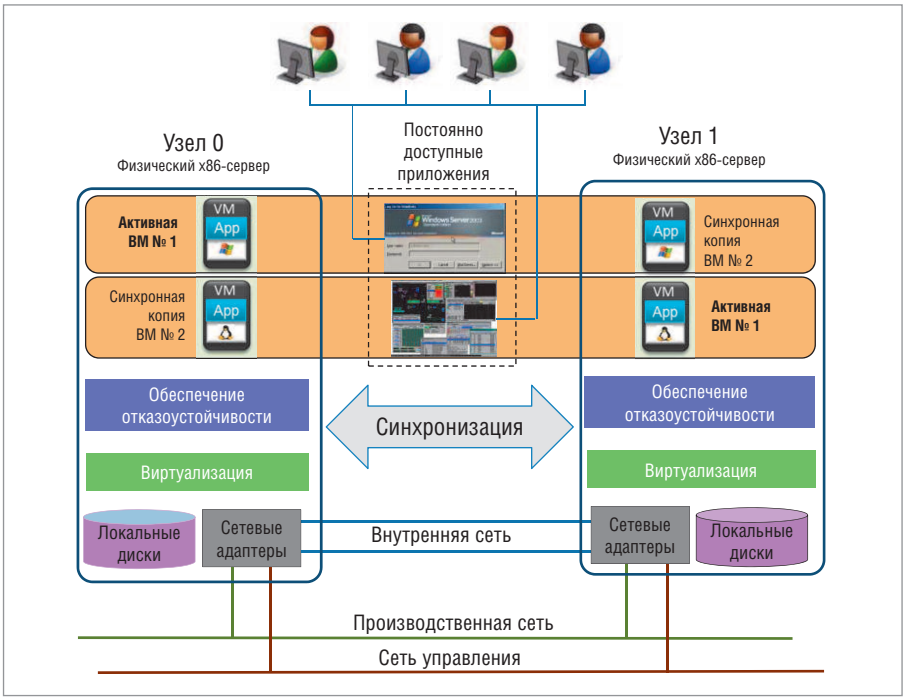


**Рис. 3. Архитектура решения AdvantiX Intellect FT**

## Как это работает

Одно приложение выполняется на двух виртуальных машинах (ВМ). В случае отказа одной машины приложение продолжит выполняться на другой машине без прерывания или потерь данных. В случае отказа компонента в одной системе вместо него используется исправный компонент второй системы.

Ввод-вывод автоматически зеркалируется на резервном сервере. Функция обеспечения отказоустойчивости сохраняет в памяти все выполняющиеся транзакции, а также данные и содержимое кэша. Перезапуска машины не требуется.

Использование двух физических серверов обеспечивает 100% резервирование на уровне аппаратных компонентов. Программные модули отвечают за зеркалирование и постоянную синхронизацию всех элементов каждой виртуальной машины на физическом уровне. Это приводит к отсутствию единой точки отказов для каждого приложения, находящегося в режиме защиты (Protected VM). Специальные модули системы управляют и поддерживают режим синхронизации операций между серверами. Они отвечают за выявление сбойных компонентов и логически удаляют их из конфигурации. После устранения неисправности эти компоненты возвращаются (реинтегрируются) в конфигурацию. Процессы логического удаления и реинтеграции выполняются бесшовно и незаметны для операционных систем и приложений. Устранение сбоев происходит без потери данных, транзакций и состояния приложения, и обеспечивается нулевое время простоя.

## Обзор архитектуры решения

**Процесс обеспечения защиты приложений**

- Загружается каждая защищённая виртуальная машина.
- Availability Engine находит парную ВМ на другом узле и объединяет их.
- Процесс объединения представляет собой следующее:
  - определяется состояние присоединяемых компонентов второго узла;
  - анализируется необходимость зеркального отображения данных, памяти и их синхронизации.
- Инициируется процесс синхронизации, и менеджер пользовательского интерфейса обновляет информацию о процессе восстановления и состоянии защищённой ВМ.

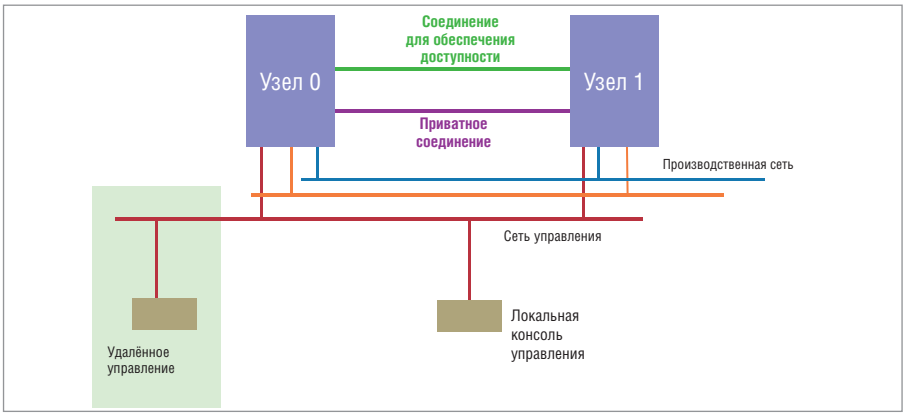


**Рис. 4. Сетевая архитектура решения**

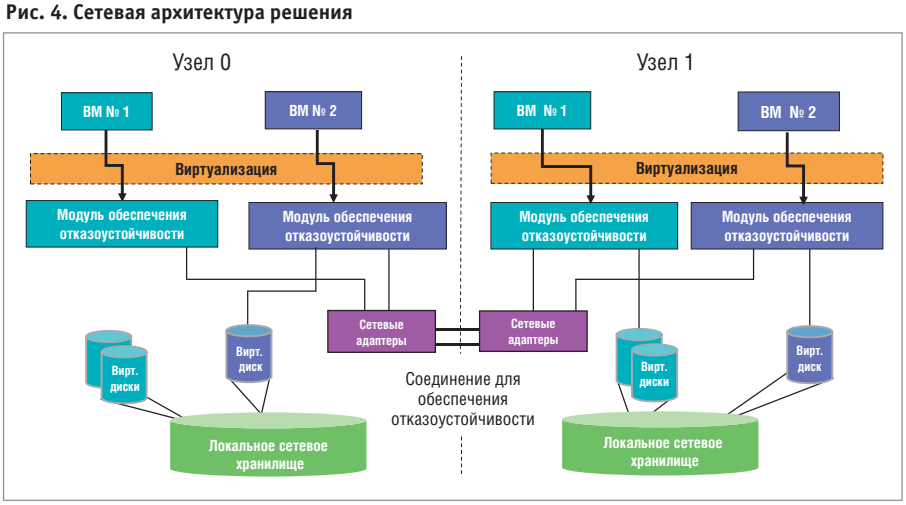


**Рис. 5. Перенаправление ввода-вывода**

**Сетевая архитектура**

Для работы необходимы четыре сетевых соединения (рис. 4).

1. *Сеть управления* (*Management Link*)
   - Общая сеть между консолью управления и хостами.
   - Каждый хост имеет свой частный IP-адрес в этой сети.
   - Оба хоста получают общий IP-адрес (Cluster IP).
2. *Закрытое соединение* (*Private Availability Link*) *Private A-Link*
   - Соединение точка-точка между узлами.
   - Резервное соединение для Private A-Link.
   - Агрегация A-Link.
   - Обеспечивает управление отказоустойчивостью и синхронизацию узлов.
3. *Сеть обеспечения доступности* (*Availability Links*)
   - Соединение точка-точка между узлами.
   - Рекомендуется использовать соединение 10 Гбайт для увеличения производительности.
   - Возможно использование до восьми портов.
4. *Производственная сеть* (*Business Networks*)
   - Резервируемая сеть для приложений.
   - Может быть совмещена с сетью управления.
   - Возможно использование до 20 портов.

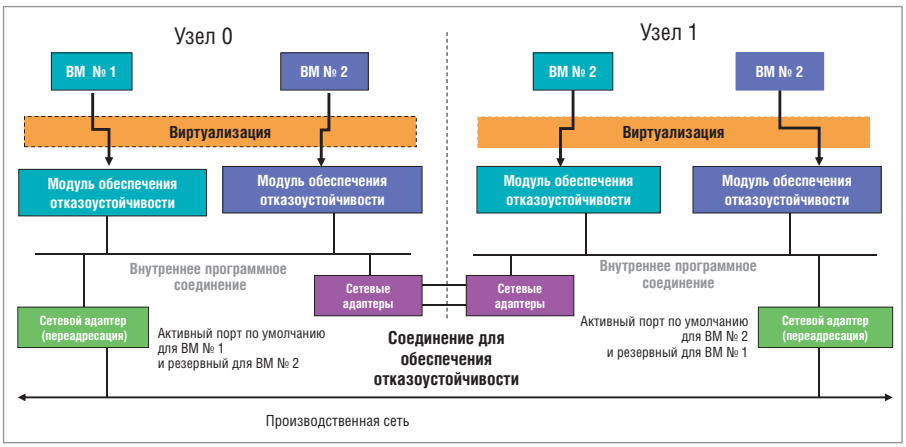


**Рис. 6. Перенаправление дисковых операций**

## Основные компоненты системы

**Модуль перенаправления ввода/вывода**

В виртуализации без резервирования ввод-вывод направляется с виртуальной машины на физический уровень и в обратном направлении, с физического уровня на виртуальную машину. В рассматриваемом решении ввод-вывод, кроме того, направляется на другой узел с получением подтверждения о записи, чем обеспечивается резервирование данных в случае сбоя одного из устройств. Если устройство неисправно, то оно удаляется из процесса работы и предпринимаются соответствующие восстановительные действия, прозрачные для приложения. В случае выхода из строя сетевого адаптера сетевой трафик будет маршрутизироваться через другой узел (рис. 5). В случае сбоя диска виртуальная машина будет работать с исправным диском на другом сервере (рис. 6). Согласованное состояние приложений и данных между узлами системы будет обеспечено и в случае, когда один из серверов выйдет из строя. Данные и транзакции не потеряются.

**Модуль работы с контрольными точками**

Этот модуль отвечает за синхронизацию состояния работающих виртуальных машин. Он определяет, когда и как эффективнее передавать изменения оперативной памяти в резервную виртуальную машину. Во многом скорость работы приложений зависит от этого модуля и производительности межузлового соединения. Синхронизация осуществляется в реальном масштабе времени и происходит по мере изменения состояния оперативной памяти виртуальной машины (рис. 7). Для исключения потери согласованного состояния активной и резервной виртуальных машин при высокой интенсивности операций с памятью и недостаточной пропускной способности приватного соединения этот модуль замедляет работу виртуальной машины до получения полной синхронности набора данных на активной и резервной виртуальной машине.

**Модуль работы с дисковой подсистемой**

Каждая виртуальная машина имеет своё виртуальное дисковое пространство, выделенное из общего пула дисковой подсистемы физических серверов. Модуль работы с дисковой подси-

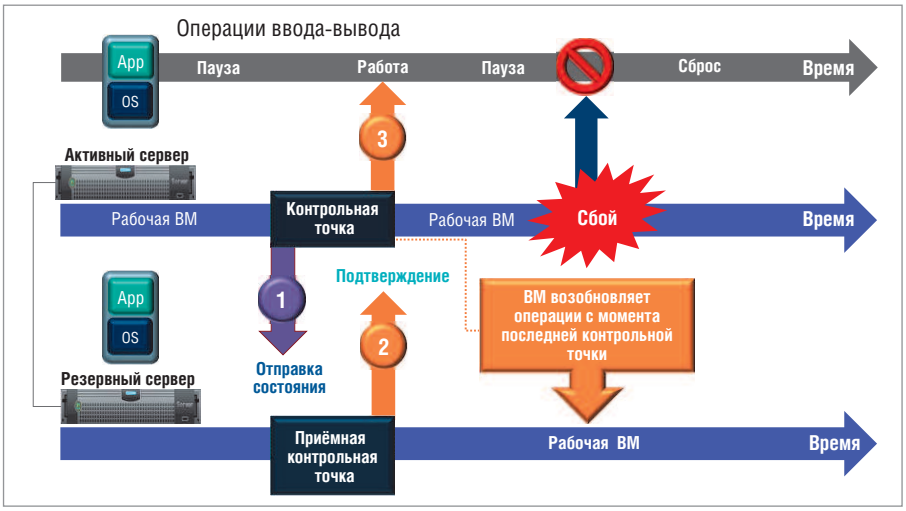


**Рис. 7. Работа с контрольными точками**

стемой отвечает за согласованное состояние наборов данных на дисках, включая образы операционных систем и приложений.

Как говорилось ранее, модуль перенаправления ввода-вывода для дисковых операций обеспечивает зеркалирование данных между дисками, размещёнными на разных физических серверах. Если обе части (расположенные на двух серверах) зеркала виртуального диска в рабочем состоянии, то данные на них записываются синхронно. Если один из дисков неисправен, то данные будут записываться на рабочую сторону, также будет записываться служебная информация о том, что не удалось записать на другую сторону. После того как неисправность будет устранена, недостающая информация перезаписывается и зеркалирование восстанавливается.

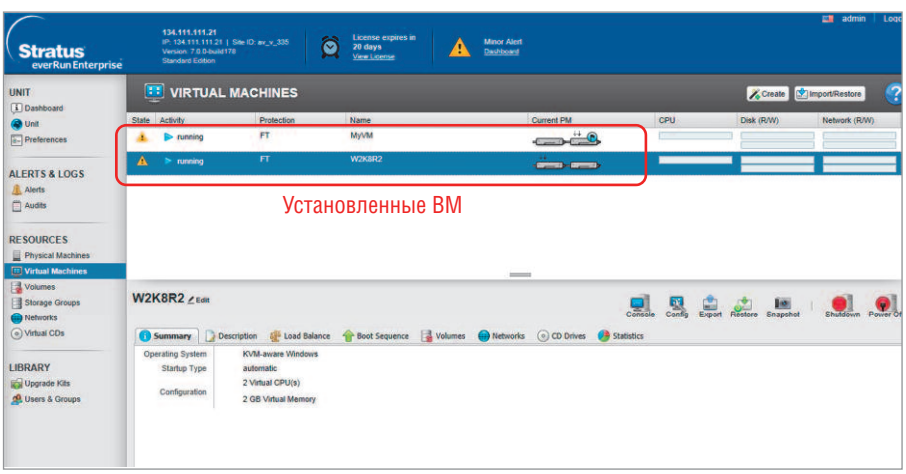


**Рис. 8. Консоль управления**

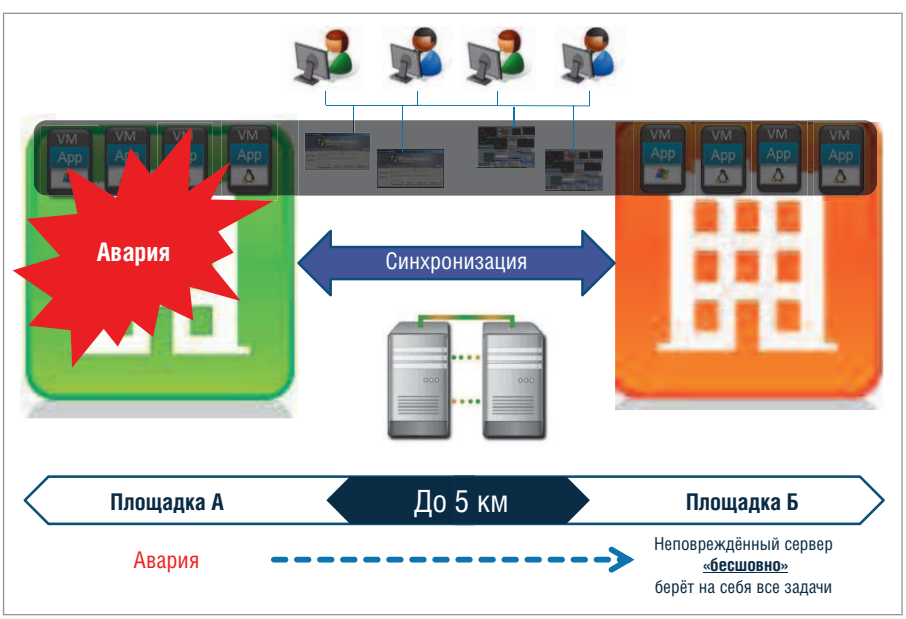


**Рис. 9. Раздельное размещение узлов (Split/Site)**

**Модуль обработки ошибок**

Этот модуль отвечает за логику работы системы при возникновении ошибок. Он фиксирует ошибку, логически удаляет сбойный компонент и отправляет сообщение оператору через коммуникационный интерфейс. Когда компонент снова обнаруживается в системе, он проверяется на работоспособность, и если всё в порядке, возвращается в конфигурацию и синхронизируется с аналогичным активным компонентом (своим партнёром). Обработчик ошибок использует набор компонентов от каждого сервера: диски, сетевые контроллеры, процессоры, память. Если хотя бы один компонент из набора исправен, приложение продолжает работу. Например, это может означать, что есть исправный сетевой адаптер на одном узле и исправная дисковая подсистема на другом узле.

**Консоль управления**

Пользовательский интерфейс управления (UI management) на основе браузера обеспечивает богатую среду для установки, мониторинга и управления конфигурацией (рис. 8). Через этот интерфейс виртуальные машины могут быть созданы, защищены, экспортированы и импортированы. Компоненты системы, такие как сетевые подключения виртуальных ЦП, памяти и размеры хранения, могут быть изменены. Аналогичным образом физическое оборудование можно контролировать и управлять им с использованием таких операций, как запуск/выключение, перезагрузка и реконфигурации компонентов.

Некоторые работы по обслуживанию выполняются автоматически. Другие сценарии обслуживания требуют замены устройств, что легко выполнить через консоль управления.

## Обеспечение катастрофоустойчивости

Если в число рассматриваемых рисков входят серьёзные аварии поддерживающей инфраструктуры, приводящие к выходу из строя производственной площадки организации, следует предусмотреть распределённые меры обеспечения живучести, такие как создание или аренда резервного вы-

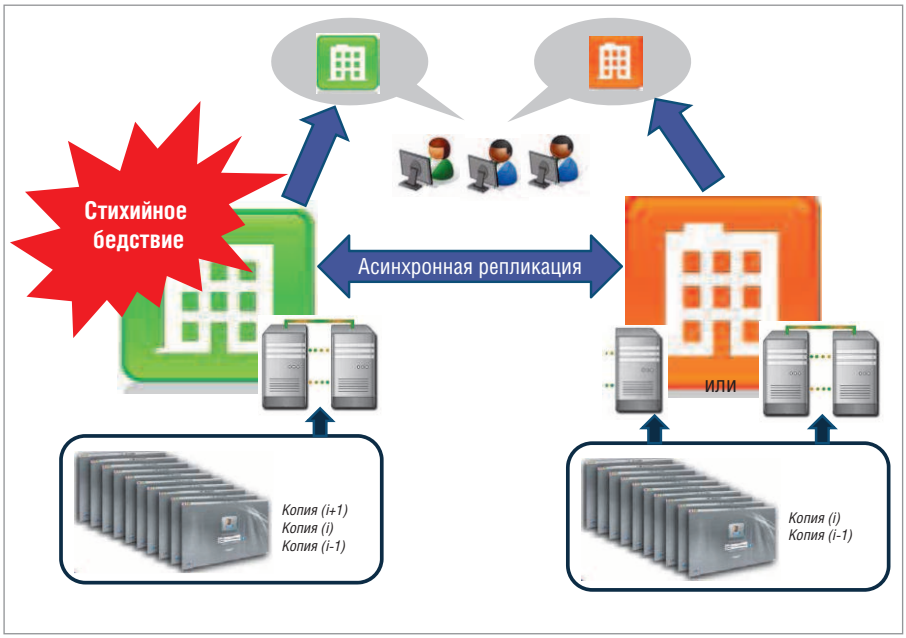


**Рис. 10. Катастрофоустойчивое размещение**

числительного центра. При этом, помимо дублирования и/или тиражирования ресурсов, необходимо иметь средства автоматического или быстрого ручного переконфигурирования компонентов ИС, чтобы обеспечить переключение с основной площадки на резервную.

Решение позволяет обеспечить отказоустойчивость как путём разнесения установки узлов – это решение Split/Site, так и значительного географического разнесения, используя коммутируемые соединения – Disaster Recovery. Решение Split/Site использует прямое сетевое соединение и синхронную репликацию данных и оперативной памяти и позволяет обеспечить непрерывность процессов без потери данных и транзакций. Узлы можно разнести на расстояние до 5 км (рис. 9). Решение даёт возможность обеспечить непрерывность бизнеса при выходе из строя одной из площадок установки, например в результате пожара, разрушения площадки или иного сбоя в эксплуатации.

Решение Disaster Recovery позволяет смягчить последствия стихийных бедствий, поддерживая асинхронные репликации между географически разделёнными площадками по глобальной сети связи (рис. 10).

*Таблица 1*

**Конфигурация решений AdvantiX Intellect FT**

| Модель | Виртуальные процессоры* | Память | Дисковые накопители | ЛВС |
|---|---|---|---|---|
| **FT-ER** | 2 | 4 Гбайт | 1×SSD | 3×GbE |
| **FT-E3** | 4 | 16 Гбайт | 4×HDD | 4×GbE |
| **FT-E5** | 24 | 32 Гбайт | 8×HDD | 2×10 GbE 2×GbE |

**Число виртуальных процессоров, доступных приложениям.*

*Таблица 2*

**Сравнение решений**

| Простой и влияние | Отдельный сервер | Отказоустойчивый кластер | AdvantiX Intellect |
|---|---|---|---|
| Незапланированный простой* | Более 43 часов в год | Более 4 часов в год | 4 минуты в год |
| Простой при техобслуживании | Значительный | Требуется перезапуск | Нет |
| Время восстановления после отказов | Часы | Часы | Нет |
| Потери данных | Да | Да | Нет |
| Работы по восстановлению | Вручную | Требуется переключение на другой ресурс при отказе. Требуются сценарии и проверки | Полностью автоматически, не требуется перезапуск |
| Обнаружение отказов | Нет | Нет | Да, автоматически |
| **Развёртывание** | | | |
| Время настройки | Часы | Дни | Часы |
| Администрирование | Нет данных | Сложно, вручную | Просто |
| Квалификация | Минимальная | Высокая | Минимальная |
| **Расходы** | | | |
| Цена первоначального приобретения | Низкая | Высокая | Низкая |
| Расходы на обслуживание | Низкие | Очень высокие | Низкие |
| SAN/внешнее хранилище данных | Не требуется | Требуется | Не требуется |
| Встроенные средства управления | Нет | Нет | Да |

** Приблизительное значение за год. Источник: Analyst Insight компании Aberdeen Group, июнь 2013 г.*

Резервная площадка всегда имеет последнюю копию данных и позволяет быстро перезапустить приложения, обеспечивая низкие значения точки восстановления (RPO) и времени восстановления (RTO).

## Модель использования

Использование архитектуры x86 и стандартных методов виртуализации, прозрачных для широкого круга программного обеспечения, открывает широкие возможности по применению данного решения для критически важных задач, требующих высокой доступности:

- SCADA (АСУ ТП, АСКУЭ, АСТУЭ);
- MES (Data Collection, Enterprise portal, Terminal Services и другие модули);
- безопасность, контроль доступа и видеонаблюдение;
- автоматизация транспорта и зданий.

Решение предлагается в различных исполнениях для разных условий эксплуатации и соответствующих конфигураций и покрывает практически все возможные варианты использования для управления производственными процессами.

**Исполнение ER**

Решение для жёстких условий эксплуатации на основе платформы AdvantiX Intellect FT. Основной особенностью платформы является отсутствие компонентов с вращающимися элементами. Используются пассивное охлаждение и SSD-диски, что позволяет применять решение в широком диапазоне температур в необслуживаемых помещениях.

**Исполнение E3**

Решение в промышленном исполнении на базе Server Grade IPC-шасси. Платформа имеет пылезащитные фильтры и виброустойчивые крепления для жёстких дисков. Может эксплуатироваться в производственных помещениях.

**Исполнение E5**

Высокопроизводительное решение с широкими возможностями по расширению для эксплуатации в стандартных серверных помещениях.

Основные характеристики приведены в таблице 1.

## Совокупная стоимость владения

Выбор решения для обеспечения высокой доступности не отличается от поиска решений по управлению рисками, когда необходимо балансировать между предельными затратами и издержками, связанными с рисками от потерь. Необходимо рассчитать затраты для увеличения доступности по нескольким пунктам (или даже долям пунктов), затем определить, оправдываются ли они ожидаемыми потерями от простоев.

Ключевым моментом является баланс между ожидаемыми издержками и выгодами. Совокупная стоимость владения (TCO) для любого уровня включает первоначальные прямые затраты по созданию инфраструктуры, закупке аппаратного и программного обеспечения, прямые и косвенные затраты по внедрению, операционные накладные расходы, зарплату персонала по штатному расписанию, расходы на консультации и многое другое. Только тщательный анализ TCO и метрик высокой доступности RTO и RPO позволит создать наиболее эффективное решение для конкретной задачи.

Сравнение решения AdvantiX Intellect FT с другими решениями для обеспечения готовности приведено в таблице 2.

## Заключение

Современные технологии виртуализации резко изменили структуру затрат по обеспечению высокой доступности и во многих случаях заменили дорогостоящие серверные кластеры и экзотические стратегии удалённого резервирования, инновационно сочетая стандартное оборудование и высокопроизводительное ПО для ряда задач, предъявляющих самые высокие требования по надёжности к информационным системам, обеспечивающим их реализацию. Цена сбоя и ликвидации его возможных последствий может быть очень высока. Решение AvantiX Intellect FT на базе ПО Stratus everRun Enterprise обеспечивает высокую доступность с нулевым временем простоя и нулевое администрирование, то есть постоянную доступность приложений без потери данных и транзакций при минимальных затратах на обеспечение его работоспособности.

Решение может применяться для различных критически важных задач, с различными требованиями по производительности и условиям эксплуатации оборудования. ●

## Литература

1. Игорь Афонин. Решение AdvantiX Intellect для обеспечения высокой доступности информационных систем // Современные технологии автоматизации. – 2013. – № 4.

**Автор – сотрудник фирмы ПРОСОФТ**
**Телефон: (495) 234-0636**
**E-mail: info@prosoft.ru**