



Технологии серверов IBM System x

Июль 2007

Содержание

Стратегия построения	3
IBM X-Architecture	3
Вертикальная и горизонтальная масштабируемость	3
IBM Enterprise X-Architecture	4
IBM Xtended Design Architecture	5
Процессоры	8
IBM XceL4v	8
IBM XrandOnDemand	8
Системные разделы	10
Процессоры	11
Многоядерность	11
Виртуализация	11
64-разрядность	12
Защита от вирусов	12
Технология энергосбережения	12
Hyper Transport	12
Память	13
Память DDR2	13
Буферизированная память	13
Fully Buffered DIMM	14
Технология памяти ChipKill	14
Active Memory: зеркалирование, замена/добавление без отключения системы	15
Memory ProteXion	16
Диагностика памяти	16
Диски	17
Интерфейсы для прямого подключения дисков	17
Внешние системы хранения	18
Уровни RAID для дисковых подсистем	19
Подсистема ввода/вывода	22
PCI	22
PCI-X	22
PCI-Express	23
Сеть	24
TCP Offload Engine	24
Wake-on-LAN	24
Jumbo Frame	24
Adapter Teaming	24
Системное управление	25
Predictive Failure Analysis (PFA)	25
Light Path Diagnostics	26
Baseboard Management Controller (BMC)	26
Remote Supervisor Adapter (RSA)	27
IBM Director	28
Remote Deployment Manager	29
Capacity Manager	30
Software Distribution Premium Edition	30
Virtualization Manager	30
Power Executive	31

Стратегия построения

IBM X-Architecture

IBM X-Architecture – это стратегия построения серверов стандартной архитектуры IBM System x, основывающаяся на применении инновационных технологий и обширного опыта компании IBM в области создания серверных комплексов. Реализация этой стратегии позволила создать лучшую стандартную платформу с функциональными возможностями уровня предприятия – производительностью, масштабируемостью, готовностью, управляемостью и обслуживаемостью – по привлекательной цене для удовлетворения всех потребностей конечных пользователей. Реализация концепции X-Architecture – это интеграция проверенных временем технологий RISC-систем и мэйнфреймов компании IBM в стандартные системы IBM System x на базе процессоров с архитектурами x86 и x86-64.

В рамках стратегии X-Architecture на стандартных серверах были реализованы такие инновационные технологии, как Active PCI-X, оперативная память с восстановлением ChipKill™, средства предсказания сбоев аппаратных компонентов Predictive Failure Analysis®, поиск и устранение неисправностей сервера с помощью Light Path Diagnostics™, а также комплекс аппаратно-программных интеллектуальных средств системного управления – интегрированный сервисный процессор, Remote Supervisor Adapter, IBM Director и дополнения к нему.

Вертикальная и горизонтальная масштабируемость

Для обеспечения защиты инвестиций и распределения затрат по времени чрезвычайно важна возможность постепенного наращивания вычислительных ресурсов, то есть масштабируемость, которая может быть вертикальной (Enterprise X-Architecture) и горизонтальной (Xtended Design Architecture). Под вертикальной масштабируемостью подразумевается усиление вычислительных возможностей системы, а под горизонтальной – объединение систем в единый виртуальный вычислительный ресурс. Каждый из этих подходов рассчитан на применение в различных областях. Так, горизонтальное масштабирование лучше всего подходит для балансировки нагрузки Web-приложений, терминальных ферм и серверов приложений, а вертикальное масштабирование – для больших баз данных систем ERP/CRM, управлять которыми на одной системе проще и эффективнее, а также для консолидации. Вертикальная

масштабируемость – это улучшение характеристик используемых серверов путем увеличения количества процессоров, емкости памяти и наращивания подсистемы ввода/вывода с целью повышения общей производительности сервера, а горизонтальная масштабируемость – увеличение количества серверов для распределения нагрузки между ними.

IBM Enterprise X-Architecture

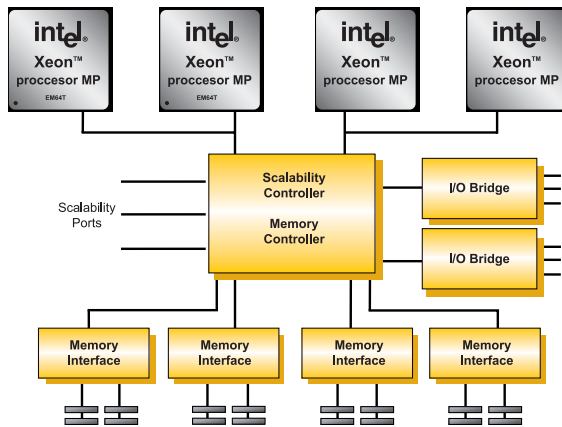
Стратегия Enterprise X-Architecture – это эволюционное развитие стратегии X-Architecture для вертикально масштабируемых стандартных серверов.

Революционные усовершенствования процессоров, подсистем памяти и ввода/вывода привели к существенному увеличению производительности, масштабируемости и готовности серверов System x. Принципиально новые подходы к проектированию серверов позволили внедрить гибкий подход к построению высокопроизводительных систем на базе 32- и 64-разрядных серверов System x, обеспечивающий возможность «оплаты по мере роста».

Технология EXA третьего поколения обеспечивает следующие возможности:

- масштабируемость XpandOnDemand™ до 32 сокетов;
- разбиение вычислительной системы на разделы (partitioning);
- подсистемы ввода/вывода Active PCI-X 2.0 и Active PCI-Express x8;
- память Active Memory™:
 - поддержка до 512 ГБ оперативной памяти;
 - высокоскоростная память DDR2;
 - трехуровневая защита памяти – ChipKill, Mirroring, ProteXion;
 - память с возможностью добавления и замены модулей без выключения системы;
- работа с шиной обмена с памятью на частотах до 667 МГц;
- xceL4v™ Dynamic Server Cache;
- технологии Enterprise X-Architecture используются в серверах IBM System x3800, x3850, x3950.

Архитектура EXA-3G



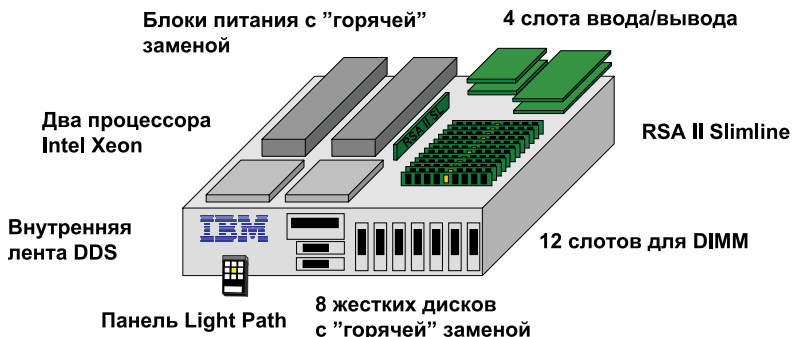
IBM Xtended Design Architecture

Технологическое лидерство, глубокое понимание потребностей конечных пользователей, а также большой опыт в области расширения функциональных возможностей стандартных компонентов являются ключевыми элементами, на основе которых компания IBM разработала новую концепцию построения серверных систем IBM Xtended Design Architecture, обеспечивающую дополнительные конкурентные преимущества вашему бизнесу. Серверная архитектура третьего поколения IBM Xtended Design Architecture, поддерживающая новые функциональные возможности серверов уровня предприятия, является дальнейшим развитием признанных и проверенных временем технологий X-Architecture.

Технологии IBM Xtended Design Architecture используются в новых серверных платформах IBM System x и IBM BladeCenter, которые соответствуют передовым отраслевым стандартам, обеспечивают новые уровни производительности, готовности, управляемости и гибкости, а также способны адаптироваться к меняющимся требованиям бизнеса.

Xtended Design Architecture в работе

Сервер приложений x3650 высокой готовности и производительности с Calibrated Vektored Cooling™



Технологии IBM Xtended Design Architecture обеспечивают:

- Производительность и масштабируемость.
 - Поддержка технологии 64-разрядных расширений, увеличение максимального объема устанавливаемой оперативной памяти, увеличение объемов внутреннего дискового пространства, увеличение пропускной способности подсистемы ввода/вывода, более эффективное наращивание подсистемы ввода/вывода, благодаря интеграции расширенных функций (поддержка RAID, аппаратные средства системного управления) обеспечивают исключительную производительность и масштабируемость серверов IBM System x.
- Готовность.
 - Технологии Memory Mirroring и Hot Spare Memory (зеркалирование и «горячий» резерв памяти), а также резервирование блоков питания и вентиляторов с возможностью «горячей» замены обеспечивают новый уровень готовности системы для поддержания непрерывности бизнес-процессов.
 - Технология Calibrated Vektored Cooling является высокоэффективной архитектурой охлаждения, которая обеспечивает расширенную функциональность и высокую степень интеграции компонентов новых систем IBM System x.
- Управление.
 - Новая версия ПО для системного управления IBM Director 5.2 поддерживает аппаратные платформы различных архитектур (System x, System p и System i), расширенные функции управления операционной системой Linux, а также

новые отраслевые стандарты системного управления (ASF 2.0, IPMI 2.0). Система диагностики IBM Light Path Diagnostics четвертого поколения обеспечивает простое и быстрое сервисное обслуживание системы за счет возможности идентификации отказавших компонентов без открытия корпуса сервера. Диагностика Light Path – наиболее полная и простая в использовании автономная система диагностики в отрасли, включающая поддержку предупреждений отказов процессоров, памяти, жестких дисков, блоков питания, вентиляторов и модулей управления напряжением (VRM).

- Гибкость.

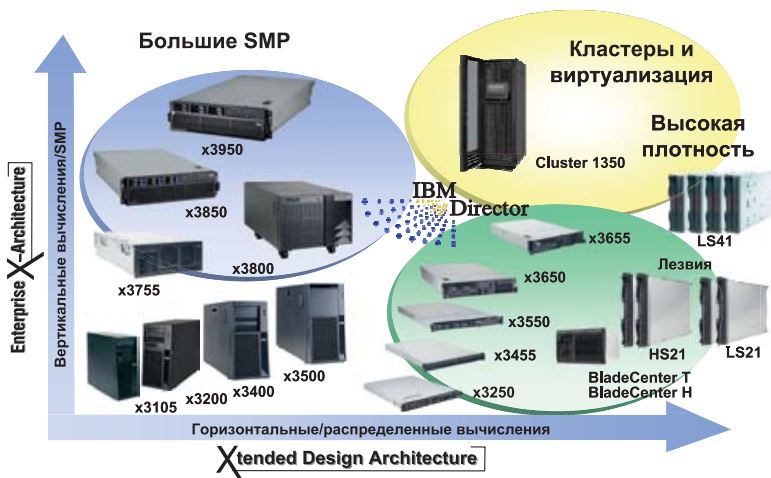
- Отличная внутренняя масштабируемость, работа с 32- и 64-разрядными приложениями, модульные серверные решения и системы хранения, а также широкий выбор стандартных операционных систем обеспечивают исключительную гибкость решений, построенных на основе технологий IBM Xtended Design Architecture.
- Оптимизация рабочей нагрузки, обеспечение непрерывности бизнес-процессов, снижение стоимости вычислений, увеличение производительности ИТ-систем, упрощение инфраструктуры, возможность «оплаты по мере роста» и защита инвестиций являются основными достоинствами нового подхода к построению серверных систем.

Технологии IBM Xtended Design Architecture используются в новых моделях:

IBM System x3105, x3200, x3400, x3500, x3250, x3550, x3650, x3655, а также в IBM BladeCenter.

Модульные вычисления

Платформа IBM System x: широчайший выбор!



Процессоры

IBM Xcel4v

Применение этой технологии является ключевым фактором повышения производительности серверов IBM System x на базе Enterprise X-Architecture. Кэш Xcel4v позволяет уменьшить загрузку системной шины и увеличить производительность процессора за счет применения встроенного фильтра snoop filter. В результате 4-процессорным серверам IBM System x не требуются процессоры с кэшем третьего уровня для достижения максимальной производительности. Кэш Xcel4v обеспечивает выборочное кэширование с настолько низкой задержкой при обработке памяти, что отпадает потребность в физическом кэш-чипе L4. В качестве хранилища, поддерживающего данное кэширование, используется область основной высокопроизводительной памяти сервера (до 256 МБ на 4 процессора).

IBM XpandOnDemand

Одним из направлений оптимизации производительности серверов является построение подсистем памяти и ввода/вывода, максимально эффективно использующих возможности архитектуры процессоров. Серверы, созданные на базе стандартной архитектуры, начинают работать менее эффективно при использовании более 4 процессоров именно из-за недостаточного использования ресурсов памяти и ограничения шин ввода/вывода. Технология EXA поддерживает расширенные системы ввода/вывода и современные архитектуры подсистем управления памятью, а также

х3950 и х3950 Е: мощность по требованию

Масштабируемость XpandOnDemand™

Модульная масштабируемость обеспечивает простой путь наращивания мощности высокопроизводительных SMP

Оптimalен для
баз данных и виртуализации

х3950 + (3) х3950 Е
Четыре шасси 16-way
до 256 Гб памяти

х3950 + (1) х3950 Е
Два шасси 8-way
до 128 Гб памяти

х3950 2w-4w одно шасси
до 64 Гб памяти

х3950 + (7) х3950 Е
Восемь шасси 32-way
до 512 Гб памяти



работу с высокоскоростным разделяемым кэшем. Новая архитектура предлагает пользователям эффективное решение для построения симметричных мультипроцессорных систем (SMP) из высокопроизводительных 4-процессорных модулей расширения SMP. Из таких блоков можно строить 4-, 8-, 16- и 32-процессорные системы.

Масштабируемый узел (scalable enterprise node) можно представить как автономный компьютер, включающий процессоры, подсистему ввода/вывода, память и другие компоненты. На каждом блоке может работать своя операционная система, либо одна ОС может исполняться на нескольких блоках за счет создания системных разделов (system partitioning). Блоки объединяются между собой с помощью выделенных высокоскоростных шин, получивших название портов расширения SMP (SMP Expansion Port) и обеспечивающих высокоэффективное совместное использование ресурсов системы. Описанный подход позволяет либо конфигурировать систему как один большой вычислительный комплекс, либо разбить ее на несколько вычислительных ресурсов с возможностью впоследствии менять конфигурацию системы. Такая технология получила название XrandOnDemand – масштабируемость по требованию.

Масштабирование системы может быть проведено по нескольким сценариям:

- увеличение количества процессоров путем добавления новых блоков;
- увеличение объема оперативной памяти путем добавления новых модулей памяти;
- расширение состава периферийных устройств и дискового пространства путем добавления внешних модулей и карт;
- наращивание вычислительной мощности путем добавления блока с новыми высокоскоростными процессорами;
- оптимизация производительности комплекса благодаря правильному выбору ОС для критичных приложений и конфигурированию системных разделов;
- и т.д.

Причем добавление в систему новых модулей расширения обеспечивает более значительный прирост производительности по сравнению с линейным увеличением количества процессоров благодаря средствам организации эффективного взаимодействия между блоками по выделенным шинам.

Системные разделы

Системные разделы – еще одна из возможностей построения высокоэффективных серверных систем на платформе Intel, заложенная в Enterprise X-Architecture. К выгодам разделения системы на разделы, можно отнести следующее:

- консолидация оборудования;
- миграция программного обеспечения и организация его совместного функционирования;
- поддержка разработки, тестирования и сопровождения решений;
- изоляция ресурсоемких приложений;
- автономное копирование и восстановление данных для раздела;
- и т.д.

Разбиение системы на разделы в рамках IBM EXA может быть выполнено двумя способами: физическое разбиение и логическое разбиение.

При физическом разбиении сервер может одновременно выполнять множество экземпляров одной операционной системы в отдельных разделах (так же как и различные версии операционной системы или даже различные типы операционных систем). Сервер может иметь до восьми связанных блоков, каждый из которых содержит независимо работающие процессоры, память и систему ввода/вывода и способен исполнять собственную операционную систему и свой набор приложений. Раздел может включать несколько блоков, например все восемь. Каждый блок может управляться независимо от другого с применением специального программного обеспечения. Логическое разбиение, в свою очередь, более гибкое, поскольку границы не определены физически. Благодаря этой гибкости приложения могут максимально эффективно использовать все системные ресурсы по мере необходимости.

Логическое разбиение на разделы предусматривает установку на сервер специальной операционной системы (например, VMware ESX Server), которая обеспечивает виртуализацию аппаратных ресурсов сервера и их распределение между виртуальными машинами. Внутри виртуальной машины работают стандартная операционная система и приложения, при этом виртуальные машины полностью изолированы друг от друга. Такой подход

обеспечивает динамическое перераспределение ресурсов между виртуальными машинами, а также эффективность использования существующих ресурсов сервера.

Процессоры

Использование широко распространенных в отрасли процессорных технологий позволяет предлагать клиентам качественные решения по доступной цене, разрабатывая современные вычислительные комплексы с наилучшими соотношениями «цена/качество» и «цена/производительность». В серверах IBM System x используются процессоры Intel и AMD — новейшие двухъядерные процессоры обоих производителей и четырехъядерные процессоры Intel Xeon серии 5300.

Многоядерность

Двухъядерный процессор – это фактически два процессора, выполненных конструктивно на одном кристалле. Четырехъядерная архитектура объединяет четыре процессора в единый процессорный чип, что приводит к повышению эффективности работы по сравнению с двухъядерными процессорами.

Суммарная производительность системы, построенной на базе многоядерных процессоров, возрастает за счет более эффективного согласования работы процессоров с данными и устройствами.

Виртуализация

Данная технология позволяет абстрагировать аппаратную среду сервера путем разделения его на виртуальные машины, причем на аппаратном уровне. Благодаря этому производителям программного обеспечения можно не беспокоиться о программной эмуляции виртуализации на процессоре. Таким образом, обеспечивается гибкая и защищенная консолидация множества операционных систем и приложений на единой платформе, повышается эффективность использования ресурсов, упрощается ИТ-инфраструктура и снижаются расходы на управление. Технология виртуализации у рассматриваемых производителей процессоров называется Intel-VT и AMD-V.

64-разрядность

Новая микроархитектура оперирует 64-разрядными регистрами, благодаря чему адресуемое пространство оперативной памяти расширяется за пределы ограничения в 4 ГБ, что немаловажно при построении систем обработки информации. 64-разрядная технология позволяет обеспечить новый уровень быстродействия как 64-разрядных, так и 32-разрядных систем. У AMD данная технология называется AMD64, а у Intel – EM64T.

Защита от вирусов

Предотвращение атак, направленных на переполнение буфера. Рассматриваемая функция на процессорах AMD носит название AMD64 Execution Protection NX-bit, а на процессорах Intel – Execute Disable Bit.

Технология энергосбережения

Оптимизация потребляемой процессором энергии в зависимости от его загрузки. Процессор имеет несколько комбинаций рабочих частот и напряжений питания, между которыми он может переключаться в зависимости от выполняемой задачи. Технология энергосбережения у Intel имеет обозначение Demand-based switching, а у AMD – PowerNow!

Hyper Transport

Данная технология является составляющей архитектуры AMD64 и представляет собой высокопроизводительную шину, обеспечивающую пиковую пропускную способность до 22,4 ГБ/с. Применение Hyper Transport в вычислительных системах способствует увеличению общей производительности за счет устранения «узких» мест при передаче данных, увеличения пропускной способности и уменьшения задержек доступа.

Память

В настоящее время в серверах используется и непрерывно совершенствуется технология микросхем памяти Synchronous Dynamic Random Access Memory (SDRAM), поддерживающая технические решения современных процессоров. Память SDRAM обеспечивает высокоскоростной обмен данными при последовательном доступе к блокам памяти.

Память DDR2

Технология Double Data Rate (DDR) позволяет увеличить скорость передачи данных за счет того, что обмен данными выполняется как на восходящем уровне синхронизирующего сигнала, так и на нисходящем. Благодаря мультиплексированию обеспечивается одновременная пересылка 64 бит информации, причем за один такт выполняются две пересылки. Технология DDR2 является новым поколением технологии DDR. Главное ее преимущество состоит в том, что она обеспечивает более высокую пропускную способность. В настоящее время память DDR2 работает на частотах от 400 МГц (это верхний предел частоты для DDR) до 667 МГц. Кроме того, DDR2 требует меньших энергозатрат, поскольку работает с пониженным до 1,8 В напряжением электропитания, по сравнению с 2,5 В – 2,8 В для DDR. Кроме того, стандарты несовместимы и по количеству контактов на модуле: в DDR – 184, в DDR2 – 240.

Существует 3 вида памяти: небуферизированная, буферизированная (регистрационная) и полностью буферизированная.

В небуферизированной памяти контроллер памяти обращается непосредственно к чипам памяти, что дает небольшой выигрыш в скорости работы. Однако небуферизированная память более энергоемка, что ограничивает количество устанавливаемых в систему модулей. Для настольных систем такое ограничение не является проблемой, а для серверов оно становится критичным, поэтому в них чаще используется буферизированная память.

Буферизированная память

Технология ее построения такова: в каждом модуле памяти имеется буфер, работающий напрямую с чипами памяти, а контроллер, в свою очередь, связан лишь с буферами. Таким образом, количество устанавливаемых в систему модулей памяти может быть значительно увеличено. Небуферизированная и буферизированная память несовместимы и не могут одновременно устанавливаться в одну систему.

Fully Buffered DIMM

Этот интерфейс памяти обеспечивает масштабируемость и максимальную полосу пропускания для процессоров и подсистем обмена данными в мощных серверных платформах. Использование стандарта FBDIMM позволяет разместить больше модулей на одном канале. В настоящее время стандарт FBDIMM поддерживается серверными платформами на базе процессоров Intel. Полностью буферизированная память позволяет в три раза увеличить пропускную способность канала «процессор – память» – до 21,3 ГБ/с при четырех каналах полностью буферизированных DIMM-модулей PC2-5300. (Два канала небуферизированной памяти PC2-3200 обеспечивают пропускную способность 6,4 ГБ/с.) Возможность одновременного выполнения операций чтения и записи устраняет задержки, вызываемые блокировкой пересекающихся операций, которая имела место в памяти предшествующего поколения. Кроме того, новые серверы оснащены инновационными средствами обеспечения надежности и защиты данных, включая усовершенствованную технологию CRC, повторение операции чтения или записи при обнаружении ошибки в данных, дополнительные буферные регистры для локализации ошибок.

Технология памяти ChipKill

В большинстве современных серверов используется стандартная память с коррекцией ошибок ECC (Error Checking and Correcting). Память ECC обнаруживает и самостоятельно исправляет любую однобитовую ошибку, а также обнаруживает, но не исправляет двухбитовую ошибку. Ошибки в большем количестве бит не обнаруживаются. Увеличение объемов используемой в серверах памяти требует применения новых решений, обеспечивающих повышение ее надежности. IBM реализовала новую технологию, получившую название ChipKill Protect ECC DIMM, которая обеспечивает надежное функционирование модуля памяти даже в том случае, когда из строя выходит целиком один чип памяти. Данная технология использует стандартные модули памяти ECC. В основе технологии лежит принцип организации массива RAID, используемый для дисковой подсистемы. Память ChipKill используется в средних и старших моделях серверов System x (x3400, x3500, x3800, x3455, x3550, x3650, x3655, x3750, x3850, x3950) и IBM BladeCenter (HS20, HS21, LS20, LS41, JS21).

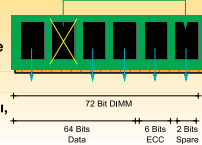
Active Memory: зеркалирование, замена/добавление без отключения системы

Потребность в наращивании памяти серверов определила появление новых технологий ее организации. Зеркалирование оперативной памяти используется для повышения устойчивости памяти к сбоям. При этом контроллер памяти одновременно работает с двумя портами доступа к памяти, действуя абсолютно прозрачно и незаметно для операционной системы и приложений, работающих на сервере. Любой сбой одного из модулей памяти не приводит к нарушению работоспособности системы, поскольку контроллер переключается на доступ к данным из другого модуля. Возможность «горячей» замены и добавления модулей позволяет изменять конфигурацию памяти сервера и устранять неисправности модулей памяти без остановки системы. Модуль может быть заменен только на эквивалентный по объему, типу и скорости. Подобная замена незаметна для операционной системы и может быть выполнена, только если включено зеркалирование. Добавление модулей памяти без выключения сервера позволяет увеличить объем оперативной памяти для операционной системы и приложений. В настоящее время такая возможность реализована только для Windows Server 2003 Enterprise Edition и Datacenter Edition.

Active Memory™ – многоуровневая защита

Memory ProteXion™ - Redundant Bit Steering

- › Redundant Bit Steering подобна диску "hot-spare" в RAID-массиве
- › Использует резервные биты в модуле памяти (hot spare bits)
- › Удваивает устойчивость Chipkill
- › Стандартная функция, независимость от операционной системы, не требуется дополнительного оборудования



Memory Mirroring

- › Непрерывная работа серверов на основе процессоров Intel
- › Значительное увеличение времени работы
- › Возможности и надежность мэйнфреймов
- › Независимость от операционной системы
- › Впервые реализована в x440 и x445



Память Chipkill™

- › Интегрирована в набор микросхем EXA
- › Увеличение надежности памяти
- › Память Chipkill увеличивает готовность за счет обнаружения и исправления многобитовых ошибок
- › Третье поколение Chipkill (1: 7000 M10, 2: 7600/6000)



Memory ProteXion

Технология Memory ProteXion позволяет использовать резервные биты микросхем памяти для хранения данных при выходе из строя одной микросхемы модуля памяти, удваивая тем самым надежность ChipKill. Дело в том, что для реализации алгоритма восстановления битовых ошибок в памяти ECC реально используются лишь 6 из 8 резервных битов на чип. Двух свободных битов чипа, заимствуемых контроллером из разных чипов модуля, достаточно для размещения данных в случае выхода из строя целого чипа. В случае возникновения такой неисправности контроллер сам примет решение об использовании резервных битов памяти без выдачи диагностического сообщения администратору системы. Если выйдет из строя еще один чип модуля, то контроллер сформирует сообщение PFA и включит диагностический световой индикатор.

Диагностика памяти

Диагностика памяти проводится периодически в автоматическом режиме и обеспечивает выявление ошибок памяти до того, как они приведут к нарушениям в работе сервера. Диагностика памяти работает совместно с технологией Memory ProteXion и не требует включения зеркалирования памяти. В ходе диагностики выявляются ошибки памяти, а также определяется, исправимы эти ошибки или нет. Если ошибка исправима, то включается механизм Memory ProteXion, данные из поврежденных областей переносятся в новые участки и генерируется сообщение об ошибке. До тех пор пока есть резервное пространство, никаких дополнительных действий, кроме записи сообщений в журнал, не производится. Если ошибку исправить не удастся, то включается световая диагностика неисправности, четко указывающая на неисправный модуль. Если включено зеркалирование памяти, то контроллер перестает использовать поврежденный модуль и работает с резервным до тех пор, пока система не будет выключена и не будет произведена замена модуля. Если BIOS поддерживает механизм «горячей» замены модулей памяти, то перезагрузки системы не требуется и новый модуль может быть задействован «на лету».

Диски

Интерфейсы для прямого подключения дисков

EIDE (Enhanced Integrated Drive Electronics)

В интерфейсе EIDE используется параллельная технология ATA. Стандарт EIDE применяется в серверах для управления периферийными устройствами, такими как CD-ROM и DVD.

SATA (Serial Advanced Technology Attachment)

Интерфейс Serial ATA пришел на смену старому параллельному интерфейсу ATA и все шире используется в серверах начального уровня. Стандарты интерфейса описывают скорость передачи данных в 150, 300 и 600 МБ в секунду.

SCSI (Small Computer System Interface)

Данный интерфейс обеспечивает параллельную передачу данных для группы дисков. В настоящее время практически исчерпан резерв интерфейса SCSI для увеличения производительности дисковой подсистемы. Наиболее эффективным для серверов на основе SCSI является интерфейс Ultra320 (его еще называют Ultra4). Сигналы управления для Ultra320 передаются по шине с частотой 80 МГц, что обеспечивает передачу данных на скорости до 320 МБ/сек. Поддерживается до 16 устройств при длине кабелей до 12 метров и 2 устройства при длине кабелей до 25 метров. Используется метод передачи управляющих сигналов LVD.

SAS (Serial Attached SCSI)

Интерфейс SAS является развитием обычного SCSI. Как это следует из названия, интерфейс использует последовательную передачу данных, в отличие от параллельной в SCSI, что обеспечивает увеличение скорости передачи данных и возможность удлинения интерфейсных кабелей. За счет каскадного соединения устройств SAS контроллер может обслуживать до 72 устройств по 4 портам. Стандарт SAS 1.0 задает скорость обмена до 300 МБ в секунду на 1 порт. Таким образом, современный адаптер ServeRAID с поддержкой SAS с 8 портами теоретически поддерживает скорость обмена до 2,4 ГБ в секунду. Однако в реальности скорость будет ограничена пропускной способностью шины PCI-X или PCI Express. В отличие от SCSI технология SAS обладает большей производительностью, масштабируемостью и надежностью.

Внешние системы хранения

iSCSI (Internet SCSI)

Реализуется как инкапсуляция протокола SCSI в виде пакетов TCP/IP, что позволяет организовывать связь между серверами и удаленными системами хранения на основе существующих сетей Ethernet. iSCSI – это возможность перейти к внешнему консолидированному хранилищу данных с использованием существующей инфраструктуры Ethernet или создав новую. В настоящее время существуют реализации протокола iSCSI для сетей 1 Гб Ethernet.

Для подключения сервера к iSCSI может использоваться как специализированный адаптер iSCSI, так и стандартный сетевой адаптер сервера вместе с установленным в операционной системе драйвером iSCSI.

SAN

SAN (Storage Area Network) – это специализированная выделенная высокоскоростная сеть, предназначенная для пересылки данных между системами хранения и серверами. Иногда такую сеть называют сетью за серверами. Как и в обычной локальной сети LAN, в SAN допускаются соединения «каждый с каждым», для чего используются такие элементы построения сети, как маршрутизаторы, шлюзы, концентраторы и коммутаторы. Чаще всего для построения SAN применяются оптоволоконные соединения (Fiber Channel), однако допускаются и другие технологии, например, существуют IP-сети SAN на базе технологии iSCSI. Поскольку для организации SAN требуется специализированное оборудование, обрабатывающее передаваемые пакеты данных, скорость передачи данных будет ниже, чем при прямом подключении дисков, – до 800 МБ в полнодуплексном режиме.

NAS

NAS (Network Attached Storage) представляет собой совокупность устройств, оптимизированных для совместного использования файлов в сети. Решения NAS базируются на классическом протоколе TCP/IP в локальной сети Ethernet. Данные передаются и принимаются устройствами по локальной сети на основе протокола TCP/IP. Создание специализированных устройств хранения, адресуемых по локальной сети, позволяет не «привязывать» систему хранения к конкретному серверу, и соединение «каждый с каждым» обеспечивается только средствами локальной сети.

Устройство NAS сочетает в себе сервер, операционную систему и набор устройств хранения. Для операций обмена, в отличие от SAN или iSCSI, используются протоколы уровня работы с файлами. Детали размещения данных в системе хранения маскируются ОС устройства. Поскольку устройства NAS интегрируются в уже готовую сетевую инфраструктуру, их легко устанавливать, настраивать и сопровождать.

Уровни RAID для дисковых подсистем

Аббревиатура RAID (Redundant Array of Independent Disks, надежный массив независимых дисков) давно уже перестала ассоциироваться только дисковыми массивами и используется для обозначения множества технологий повышения надежности функционирования группы устройств путем обеспечения избыточности.

Рассмотрим подробно основные уровни RAID для внутренних дисковых подсистем на базе контроллеров ServRAID.

RAID-0

Данные разбиваются на блоки, которые, в свою очередь, распределяются между дисками. Таким образом, можно одновременно работать с несколькими блоками данных. В данном массиве избыточность полностью отсутствует, вследствие чего повреждение одного диска приводит к повреждению всего массива, и информация полностью теряется. С другой стороны, RAID-0 обладает самой высокой скоростью передачи информации по сравнению с другими уровнями RAID. Стоимость реализации данной архитектуры минимальна, так как избыточные диски отсутствуют.

Минимальное количество дисков в массиве – 2.

RAID-1

В этом случае обеспечивается полное дублирование информации, которая записывается одновременно на два носителя, что гарантирует безопасность ее хранения. Если на одном диске произойдет потеря данных, то на другом останется их копия, таким образом, информация будет сохранена. Операция чтения в этом случае происходит быстрее, чем с одного носителя, благодаря тому, что на парные носители можно одновременно отправлять разные запросы. А запись осуществляется немного

медленнее, так как ОС нужно сформировать две команды для одного блока информации.

RAID-1 медленнее чем RAID-0, но быстрее чем RAID-5.

Минимальное количество дисков в массиве – 2.

RAID-10

RAID-10 состоит из массива RAID-0, который, в свою очередь, включает несколько массивов RAID-1. Таким образом, RAID-10 обеспечивает производительность на уровне RAID-0 и отказоустойчивость на уровне RAID-1. Недостатками этой архитектуры являются высокая стоимость исполнения и требование четного количества дисков в массиве.

Минимальное количество дисков в массиве – 4.

RAID-1E

Данная архитектура, разработанная IBM, призвана устранить ограничения по количеству дисков RAID-1 при сохранении характеристик производительности и отказоустойчивости RAID-1.

Минимальное количество дисков в массиве – 3.

RAID-1E0

RAID-1E0 представляет собой массив RAID-0, сформированный из массивов RAID-1E. В результате можно объединить в один массив до 60 дисков. В этом состоит наиболее важное преимущество данной архитектуры перед другими.

RAID-5

Архитектура RAID-5 основана на вычислении контрольной суммы данных и записи их на разные носители массива. Таким образом, для хранения контрольных сумм необходим объем памяти, равный одному физическому носителю, независимо от общего числа дисков. Такое решение позволяет максимально эффективно использовать дисковое пространство, обеспечивая высокие уровни производительности и отказоустойчивости.

Минимальное количество дисков в массиве – 3.

RAID-50

Данная архитектура обеспечивает отказоустойчивость на уровне RAID-5 при вдвое большей производительности. RAID-50 представляет собой массив RAID-0, построенный из блоков RAID-5.

Минимальное количество дисков в массиве – 6.

RAID-5E

Это массив RAID-5 с интегрированным диском hot spare. Преимущество данной технологии перед RAID-5 заключается в наличии на каждом диске области hot spare, которая резервируется под «горячий» резерв. Таким образом, при отказе одного из дисков информация по-прежнему может считываться параллельно с нескольких носителей, что обеспечивает высокую производительность.

Минимальное количество дисков в массиве – 4.

RAID-5EE

Это массив RAID-5E с более эффективным распределением резервных областей на дисках, что позволяет значительно ускорить процесс восстановления массива.

Минимальное количество дисков в массиве – 4.

Уровень RAID	Обеспечивается ли надежность	Описание
RAID-0	нет	Данные распределяются по всем устройствам (по страйпам)
RAID-1	да	Зеркальная копия данных на двух дисках
RAID-1E	да	Данные зеркалируются, но для количества дисков больше двух
RAID-5	да	Данные записываются на несколько дисков по страйпам. Вычисляется контрольная сумма страйпов одного уровня, и эта сумма распределенно записывается на диски
RAID-5E	да	Аналогично 5, но имеется дополнительное пространство для замены вышедшего из строя диска, также распределенное по массиву
RAID-5EE	да	Аналогично 5E, но с большей скоростью восстановления после сбоя
RAID-10	да	Запись данных по страйпам (RAID-0) по нескольким массивам RAID-1
RAID-50	да	Запись данных по страйпам (RAID-0) по нескольким массивам RAID-5

Подсистема ввода/вывода

Подсистема ввода/вывода компьютера обеспечивает обмен данными с устройствами, являющимися внешними по отношению к процессору и оперативной памяти. Эти устройства могут находиться как в одном корпусе с процессором, так и снаружи, на различных расстояниях от него. В состав компьютера должны входить различные контроллеры, поддерживающие требуемые протоколы обмена и обеспечивающие соединение и необходимую пропускную способность взаимодействий с шинами внешних устройств, а также локальными и глобальными сетями передачи данных. Зачастую производительность компьютеров определяется не столько быстродействием процессоров, сколько эффективностью подсистемы ввода/вывода. Ниже представлены основные технологии организации ввода/вывода, используемые в серверах System x.

PCI-X

В связи с повышением тактовой частоты процессоров, ростом пропускной способности локальных сетей и появлением высокоскоростных периферийных устройств, пропускной способности PCI оказалось недостаточно, и в 1998 г. была разработана новая спецификация расширения шины PCI под названием PCI-X.

Эта спецификация опирается на существующую технологию PCI, но за счет ряда усовершенствований протокола позволяет значительно увеличить производительность шины. Технология PCI-X обеспечивает стабильную работу устройств на частотах 66, 100 и 133 МГц. Пропускная способность при частоте 66 МГц достигает 533 МБ/с. А если вся периферия работает на частоте 100 МГц, то возможное число устройств, подключаемых к PCI-X, сокращается до двух, однако пропускная способность 64-разрядной шины достигает 800 МБ/с. Ширина полосы пропускания при подключении единственного устройства, работающего на частоте 133 МГц, достигает 1066 МБ/с. Благодаря обеспечению обратной совместимости PCI-X использует те же порты, что и классическая шина PCI (32- или 64-разрядная). Спецификация PCI-X требует, чтобы адаптеры при установке в PCI-систему поддерживали любые ее режимы. И наоборот, если обычный PCI-адаптер устанавливается на шину PCI-X, то он и все остальные адаптеры данного шинного сегмента работают по протоколу PCI. При этом также возможно наличие в одной системе как шины PCI, так и PCI-X.

Существует 2 спецификации PCI-X: PCI-X 1.0 и PCI-X 2.0. Спецификация 2.0 поддерживает пропускную способность до 2133 МБ/с и обладает такими дополнительными возможностями, как, например, разделение транзакций

и коррекция ошибок (Error Checking and Correction, ECC). Также реализован стандарт Active PCI-X, обеспечивающий возможность «горячей» замены.

PCI-Express

Основой данной технологии является последовательный интерфейс и применяемые дифференциальные сигнальные пары контактов, которые совершают обмен данными по схеме «точка – точка». Две такие пары составляют один канал. Благодаря такой топологии достигается ряд преимуществ: удешевление конструкции, уменьшение габаритов, более простая разводка печатных дорожек с упрощенными требованиями к борьбе с паразитными излучениями, что в конечном итоге выражается в возможности работы на гораздо более высоких частотах, с поддержкой «горячей» замены периферийных устройств. Также снимается необходимость такого важного для параллельного интерфейса параметра, как синхронизация сигнальных линий всей шины. Пропускная способность одного канала достигает 2,5 ГБ/с, при этом скорость передачи данных (без служебных битов) достигает 200 МБ/с в каждом направлении. PCI-Express может включать от одной до нескольких линий. Каждая такая конфигурация имеет одно из следующих обозначений: x1, x2, x4, x8, x16 или x32, где цифрой обозначено количество каналов. Если реализация x1 содержит 4 контакта, то x16 – 64. Таким образом, объясняется различие в физических размерах слотов. Можно устанавливать карту в больший по размеру слот, но не наоборот. Пропускная способность PCI-Express в полнодуплексном режиме составляет 400 МБ/с, 800 МБ/с, 1,6 ГБ/с, 3,2 ГБ/с или 6,4 ГБ/с для соответствующих вариантов исполнения – x1, x2, x4, x8, x16.

В новых моделях высокопроизводительных серверов IBM System x реализован стандарт Active PCI-Express, в котором функции Active PCI-X применены для высокопроизводительной шины PCI-Express.

Преимущества перед PCI-X

Высокая производительность: повышение пропускной способности благодаря линейному наращиванию производительности путем линейного расширения шины. Помимо этого, PCI-Express является дуплексной шиной.

Упрощение разводки периферии: стандартизация там, где ранее использовались всевозможные варианты PCI – AGP, PCI-X и другие.
Снижение комплексных затрат на разработку и внедрение систем.

Простота использования: производить модернизацию и доработку систем с устройствами PCI-Express значительно легче. Существует возможность использовать карты PCI-Express с «горячим» подключением.

Сеть

Сеть обеспечивает передачу данных между серверами, системами хранения и пользователями. Большая часть информации проходит через сеть, таким образом, быстродействие информационной системы в целом зависит от пропускной способности сети и быстродействия аппаратуры, обрабатывающей сетевой поток данных. Существует множество технологий, направленных на увеличение быстродействия передачи и обработки информации, далее перечисляются наиболее важные из них на сегодняшний день, применяемые в серверах System x.

TCP Offload Engine

Для обработки трафика TCP/IP необходимы значительные объемы ресурсов процессора, так как ему требуется выполнять множественные вычисления: обработку пакетов, перемещение данных, обработку прерываний. Данная технология позволяет с помощью аппаратных средств разгружать процессор, освобождая тем самым ресурсы системы для выполнения поставленных задач.

Wake-on-LAN

Данная технология упрощает дистанционное обслуживание серверов. WoL позволяет включить удаленную станцию с использованием сетевого интерфейса, поддерживающего эту технологию. Таким образом, обеспечивается удаленная централизованная установка и настройка ПО в запланированное время, что значительно сокращает трудозатраты и обеспечивает высокую эффективность работы ИТ-персонала.

Jumbo Frame

Поддержка больших пакетов (Jumbo Frame) уменьшает количество пакетов в сети, упрощая тем самым обработку потока данных, особенно в гигабитных сетях. Таким образом, повышается производительность системы. Для использования данной технологии необходимо, чтобы ее поддерживали все устройства сети, участвующие в обмене информацией.

Adapter Teaming

Технология, позволяющая объединять несколько сетевых адаптеров в один логический для повышения производительности и/или отказоустойчивости.

Системное управление

Средства системного управления предоставляют пользователям серверов System x ряд ключевых преимуществ, наиболее важными из которых являются сокращение затрат на управление и обслуживание ИТ-инфраструктуры и защита инвестиций.

Сокращение затрат на управление ИТ-инфраструктурой обеспечивается благодаря:

- интеллектуальным средствам управления;
- интегрированному централизованному управлению;
- средствам расширенной обработки событий;
- реализации удаленной сервисной поддержки.

Сведение к минимуму продолжительности, а значит – и стоимости простоев обеспечивает сокращение затрат на обслуживание ИТ-инфраструктуры.

Специальные механизмы позволяют осуществлять:

- предсказание сбоев оборудования (дисков, процессоров, памяти, модулей управления напряжением, вентиляторов, блоков питания) и программного обеспечения;
- диагностику оборудования без прерывания работы пользователей;
- планирование и управление производительностью системы;
- быстрый поиск и устранение неисправностей (диагностика Light Path).

Кроме того, благодаря поддержке как оборудования IBM, так и оборудования других производителей, обеспечивается защита инвестиций.

Ниже приведено краткое описание предлагаемых технологий системного управления.

Predictive Failure Analysis (PFA)

Позволяет заблаговременно (до выхода оборудования из строя) осуществлять анализ неисправностей:

- жестких дисков;
- процессоров;
- модулей памяти;
- модулей управления напряжением;
- вентиляторов;
- блоков питания.

Для каждой из неисправностей формируется отчет, который является основанием для замены сервисным центром потенциально сбойного компонента до его реального выхода из строя в течение гарантийного срока.

Light Path Diagnostics



Позволяет легко найти и заменить вышедший из строя модуль благодаря световой индикации неисправности. Система световой индикации имеет трехуровневую иерархическую систему светодиодов:

- на передней панели сервера отображается информация о сбое;
- на специальной выдвижной панели – информация о подсистеме сервера, в которой произошел сбой;
- внутри сервера горящим индикатором подсвечивается соответствующий аппаратный компонент: диск, процессор, модуль памяти, модуль управления напряжением, вентилятор, блок питания, адаптер ввода/вывода.

Важной особенностью Light Path Diagnostics является автономное питание. Система работает даже тогда, когда сервер физически отключен от источников питания.

Baseboard Management Controller (BMC)

BMC является своего рода компьютером внутри компьютера. Данное устройство выполняет задачи системного управления, помогая поддерживать работоспособность сервера. Будучи интегрированным в некоторые из серверов System x (x3400, x3500, x3800, x3455, x3555, x3650, x3650, x3655, x3755, x3850, x3950) и IBM BladeCenter (HS20, HS21, LS20, LS21, LS41), BMC непрерывно осуществляет диагностику системы и сообщает о потенциальных сбоях.

Через ПО IBM Director BMC сигнализирует об изменениях температуры системы, напряжения питания, скорости вентиляторов, производительности памяти и жестких дисков. Кроме того, он предоставляет преимущества в области управления конфигурациями, такие как удаленное обновление микрокода, удаленный контроль энергопотребления, а также функции автоматического перезапуска сервера – Automatic Server Restart (ASR).

Remote Supervisor Adapter (RSA)

RSA – это специальный адаптер, выполняющий ряд важных функций по обслуживанию сервера. Это устройство дополняет возможности BMC, обеспечивая выполнение расширенных функций системного управления в удаленном режиме.

RSA поддерживает:

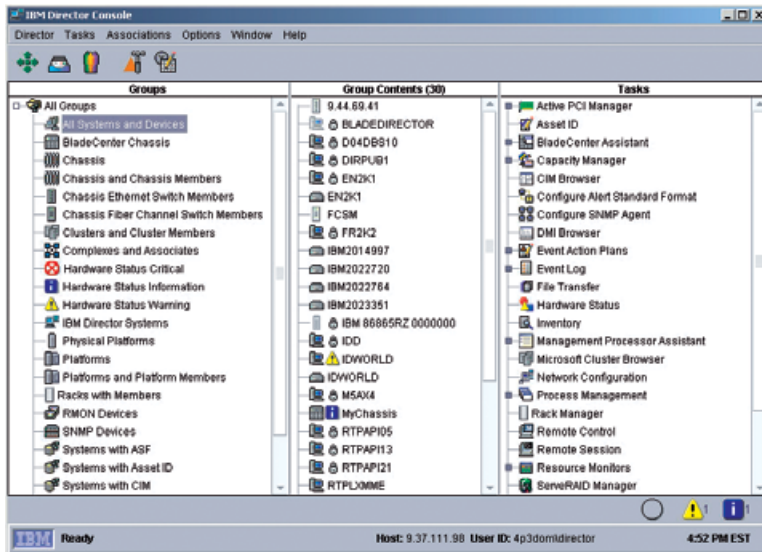
- удаленное управление оборудованием и ОС;
- управление через Web-интерфейс;
- удаленные флоппи-дискет и CD;
- формирование SNMP-событий от аппаратных датчиков PFA.

RSA интегрируется с IBM Director с использованием программных средств:

- Management Processor Assistant (MPA) plug-in в IBM Director;
- Management Processor Command-Line Interface (MPCLI).

В случае необходимости удаленного управления, встроенный видео-адаптер обеспечивает превосходную производительность с минимальным сетевым трафиком. RSA функционирует независимо от центрального процессора, обеспечивая возможность управления, даже когда сервер выключен.

RSA позволяет сокращать TCO благодаря повышению эффективности службы поддержки ИТ-инфраструктуры.



IBM Director

ПО IBM Director является полнофункциональным решением для системного управления. Основываясь на отраслевых стандартах, это ПО поддерживает серверные платформы на базе процессоров стандартной архитектуры, а также некоторые модели IBM System i™ и IBM System p®.

IBM Director включает мощные инструментальные средства и утилиты, которые позволяют автоматизировать процессы, необходимые для эффективного управления серверными системами с прогнозированием неисправностей, в том числе:

- планирование производительности;
- инвентаризация оборудования и программного обеспечения;
- диагностика и сервисная поддержка;
- поиск и устранение неисправностей.

Графический пользовательский интерфейс позволяет выполнять управление системами локально или удаленно. IBM Director обеспечивает сокращение трудозатрат на управление системами, повышая эффективность работы ИТ-персонала.

Средства упреждающего анализа и прогнозирования позволяют свести к минимуму время простоев систем, поддерживающих критически важные приложения. Обеспечение безопасности является первостепенной задачей архитектуры IBM Director. Система аутентификации пользователей интегрирована в систему безопасности операционной системы с возможностью управления доступом к задачам и системам IBM Director.

Средства аутентификации IBM Director обеспечивают защиту от несанкционированного доступа к системам и получения полного контроля над ними со стороны других серверов управления.

Дополнительный уровень безопасности в управляемой среде обеспечивается шифрованием информационных потоков по протоколу SSL с возможностью выбора алгоритма шифрования DES или 3DES.

IBM Director поддерживает стандартные операционные системы и оборудование различных производителей (гетерогенные среды), а также интегрируется в системы управления предприятиями, такие как IBM Tivoli® и другие.

Модульная архитектура IBM Director обеспечивает возможность расширения функций управления путем добавления новых утилит или модулей. К таким модулям относятся Server Plus Pack, Software Distribution Premium Edition, Remote Deployment Manager, Real Time Diagnostics и другие.

Remote Deployment Manager

Программный инструмент, позволяющий удаленно развертывать ОС и другое ПО, обновлять микрокоды системы, а также выполнять резервное копирование и восстановление данных жесткого диска. Работает на предустановленном IBM Director Server и DHCP-сервере.

RDM позволяет настраивать одновременно множество систем без непосредственного контакта с ними. Важной особенностью RDM является поддержка оборудования других поставщиков. RDM включает утилиту, чрезвычайно важную для повышения уровня безопасности бизнеса, которая обеспечивает полное удаление данных в случае необходимости. Жизненно важные данные могут быть предварительно сохранены в безопасном месте и впоследствии перенесены обратно на сервер.

Capacity Manager

Инструмент управления производительностью системы, включающий экспертную систему анализа производительности системы, средства поиска «узких» мест с выдачей рекомендаций по их устранению, а также режим прогнозирования рабочих нагрузок.

Software Distribution Premium Edition

Инструмент Software Distribution позволяет создавать и устанавливать на системы программное обеспечение в удаленном режиме, обеспечивая экономию времени и сокращение затрат. Software Distribution предлагается в двух версиях:

- Версия Standard включена в IBM Director. Она обеспечивает распространение ПО, полученного от IBM. Версия Premium Edition – это дополнительная утилита для IBM Director.
- Версия Premium Edition позволяет создавать и распространять собственное ПО (для сред Windows и Linux). Использование мастера упрощает формирование пакетов ПО, «упакованных» с применением Windows Installer Package, InstallShield Package, Red Hat Package Manager, IBM Update Assistant.

Virtualization Manager

Модуль Virtualization Manager (VM) является дополнением к IBM Director, обеспечивающим интеграцию ПО для виртуализации в инфраструктуру IBM Director, чтобы в полной мере использовать преимущества инструментов IBM Director при управлении виртуальными машинами.

Power Executive

Инструмент IBM Power Executive позволяет управлять электропитанием из среды IBM Director. Этот инструмент обеспечивает контроль над энергопотреблением и более эффективное использование имеющихся энергоресурсов, помогая:

- лучше планировать создание новых центров обработки данных или их модернизацию;
- правильно выбирать мощность питания для имеющихся физических систем;
- согласовывать приобретение дополнительных аппаратных средств с имеющимися ресурсами электропитания;
- лучше использовать существующие ресурсы.



Дополнительная информация:

Информационный портал IBM System x
www.ibm.com/systems/ru/x/

Информация о совместимости

IBM ServerProven

[www.ibm.com/servers/eserver/
serverproven/compat/us/](http://www.ibm.com/servers/eserver/serverproven/compat/us/)

Поддержка IBM System x

www.ibm.com/systems/support/t

Конфигураторы для IBM System x:

Standalone Solution Configuration Tool

Configuration and Options Guide

www.ibm.com/systems/x/configtools.html

IBM BladeCenter – Innovative Leadership!

[www.ibm.com/servers/eserver/
bladecenter/advantage/competitive.html](http://www.ibm.com/servers/eserver/bladecenter/advantage/competitive.html)

IBM Cool Blue energy management portfolio

[www.ibm.com/systems/x/about/ power/
bladecenter.html](http://www.ibm.com/systems/x/about/power/bladecenter.html)

© IBM Восточная Европа/Азия

123317, Москва,
Краснопресненская наб., 18.
Факс: +7 (095) 940-2070,
Тел.: +7 (095) 775-8800,
+7 (095) 940-2000.
ibm.com/ru.

Отпечатано в России.
Все права защищены.

Логотип IBM, eServer, Lotus, Notes, Tivoli, X-Architecture, BladeCentre, ServeRAID, Xtended Design Architecture, ServerProven, TotalStorage, ChipKill, IntelliStation, pSeries, iSeries и xSeries являются торговыми марками или зарегистрированными торговыми марками International Business Machines Corporation в США и других странах. Intel, Pentium, Xeon и Itanium являются торговыми марками или зарегистрированными торговыми марками Intel.

Linux является торговой маркой Линуса Торвальдса.

Microsoft, Windows, Windows Server, Ахapta являются зарегистрированными торговыми марками Microsoft Corporation.

Наименования других компаний, продуктов и услуг могут быть торговыми или сервисными марками третьих лиц.

Все заявления относительно намерений и перспективных планов IBM могут быть изменены без уведомления.