

# Технологии архивирования данных ERP-систем SAP

Обзор современных решений по архивированию данных ERP-систем SAP, представленных на региональном рынке.

## Введение

Технологии архивирования онлайн-данных стали активно развиваться с 2003 г. Основная причина этого — стремление максимально снизить эксплуатационные и капитальные издержки, связанные с сопровождением данных на высокодоступных онлайн-СХД.

Примерно с 2007 г. к архивированию данных возрос интерес со стороны российских компаний. Во-первых, это было обусловлено тем, что ряд компаний в силу специфики своего бизнеса должны выполнять регулирующие требования европейского союза или/и американского законодательства. Во-вторых, также мощным стимулом явилось и то, что за последние несколько лет законодательно оформлены требования к российским компаниям по хранению и представлению данных со стороны федеральных органов. В качестве примеров регулирующих требований, связанных с необходимостью обеспечения архивирования данных, можно привести следующие.

В течение 2009 г. Центробанком РФ планируется внедрение в России Basel II (инструкции, разработанные и предложенные Базельским комитетом по банковскому надзору), в соответствии с которыми банки будут обязаны собирать и хранить всю детальную информацию о своих процессах и операциях. При этом данные не только должны быть собраны со всех подразделений, но и проверены, и консолидированы. Обычно это выливается в создание новых или расширение существующих центров хранения данных (хотя многие организации уже имеют центры хранения данных, большинство все же не удовлетворяет требованиям Basel II в этом плане). Еще одним очень важным элементом Basel II является ИТ-безопасность, в соответствии с которой банк должен внедрить средства контроля над центром хранения данных.

Российское законодательство устанавливает различные сроки хранения на документы бухгалтерского, налогового учета, персональных данных сотрудников предприятия. При создании любой системы архивирования необходимо учитывать эти особенности. Сроки хранения документов достаточно велики, например:

- бухгалтерская отчетность — 5 лет;
- налоговый учет — 4 года;
- проектная документация по капитальному строительству — 20 лет;
- технологическая и конструкторская документация — 20 лет;
- научная документация — 15 лет;
- кино- и фотодокументы и др. — 5 лет;
- лицевые счета работников — 75 лет.

Необходимо учитывать, что в системах SAP в зависимости от задач проекта, может возникнуть потребность хранить документы бухгалтерской отчетности и связанные с ними отсканированные копии договоров, которые в свою очередь могут содержать требуемые документы с технологической и конструкторской документацией, что будет существенно влиять на срок хранения документа в целом.

Наиболее продолжительны сроки хранения в системах, где настроен расчет заработной плат сотрудников. Однажды рассчитанная заработная плата сотрудника должна храниться 75 лет для возможности получения лицевого счета на любой отрезок времени. Данные документы после выполнения расчета неизменны и являются наиболее подходящими для перевода в архивное хранилище.

Телекоммуникационные российские компании, работающие в странах Евросоюза, с сентября 2007 г. должны обеспечивать выполнение директивы 2002/58/ЕС, которая была принята в рамках мер по борьбе с международной организованной преступностью. При каждом звонке по мобильному телефону создается отдельная запись call data record (CDR), которая затем может использоваться для биллинга, сегментации рынка, выявления случаев мошенничества и др. Данные, которые нужно сохранять в соответствии с директивой, — это вся информация CDR за исключением голосовых данных. Также нужно сохранять данные обо всех неудачных звонках. В соответствии с директивой, данные CDR должны храниться от 6 месяцев до 2 лет со времени сеанса связи и, если государственные или местные органы правоохранения запросили эти дан-

ные, то оператор обязан предоставить их немедленно. За один рабочий день может накопиться в среднем 10 млн CDR (до миллиарда CDR). При ориентировочном размере одной CDR — 500 байт — за день, в среднем, накапливается 5 Тбайт, которые нужно сохранить.

Кроме того, операторы из стран, не входящих в ЕС, должны выполнять требования по сбору информации, действующие в ЕС. Это означает, что данные о звонке с мобильного телефона в России и на Украине абоненту, зарегистрированному в ЕС, должны быть сохранены оператором, обслуживающим этого абонента.

Российские компании, которые хотели бы стать участниками фондового рынка США, должны будут выполнить требования закона Sarbanes-Oxley (Sarbanes-Oxley Act — SOX). Согласно этому закону, все компании, которые либо представлены на фондовом рынке США, либо планируют выйти на эти рынки, должны в обязательном порядке соответствовать требованиям SOX в сфере создания эффективной системы внутреннего контроля при составлении финансовой отчетности, а также отчитываться в надежности и достоверности данной финансовой отчетности. А данные по финансовой деятельности в электронной форме должны предоставляться по запросу в течение 48 часов, даже если с момента события прошло 5 лет. Личную ответственность за соблюдение этих требований несут генеральный и финансовый директора компании.

## Архивирование и резервирование данных

Часто термин “архивирование” ставят в соответствие с резервированием. И под архивом подразумевается хранение вторичных резервных копий на более дешевых носителях, чем первичные (рис. 1).



Рис. 1. Архив в “классическом” понимании (в “традиционной” схеме обеспечения надежности данных) обеспечивал хранение резервных копий на более дешевых носителях.



- Архивирование статической информации:**
  - улучшает производительность основных приложений;
  - уменьшает ТСО, используя многоуровневые системы хранения;
  - освобождает емкость на системах хранения 1-го уровня;
  - прозрачно для пользователей и приложений.
- Резервное копирование активной информации:**
  - не нужно копировать и восстанавливать архивные данные;
  - сокращение окна резервного копирования;
  - повышение надежности, высокий процент полных копий.
- Извлечение из архива или восстановление из резервной копии:**
  - запросы на восстановление обрабатываются быстрее;
  - ранее недоступная информация теперь стала доступна.

Рис. 2. При данной архитектуре архивирования удалось в значительной степени оптимизировать процедуры резервного копирования/восстановления.

При этом архив использовался только в случае, если первичная резервная копия повреждена, а по объему соответствовал сумме размеров всех ранее сохраненных резервных копий.

Существенным недостатком подобных решений являлось то, что с ростом активной или основной базы данных резервные копии начинали “раздуваться”, вследствие чего начинали расти затраты на их хранение и, соответственно, требовалось все больше времени на процедуры резервного копирования/восстановления. В ряде случаев, когда основная база данных использовалась еще, например, для тестирования и разработки, размер резервных копий в сравнении с размером основной БД мог расти в геометрической прогрессии. А при высоких требованиях по доступности – еще больше.

Ответом на эти издержки бизнеса явился шаг со стороны разработчиков ПО, позволяющий из производственных данных (а, соответственно, и из первичных резервных копий) удалять редко используемую, но необходимую информацию с переносом ее в архивное хранилище (рис. 2). Вместо последовательной двухэтапной схемы, стали популярны решения, где процедуры резервного копирования и архивирования работают совместно, дополняя друг друга. Все процедуры были значительно оптимизированы и, соответственно, были снижены затраты на их поддержание. Дополнительно появилась возможность использовать архивные данные в режиме “nearonline”. Архив стал активной компонентой в составе ILM-решений.

## Обзор рынка решений архивирования данных для ERP-систем SAP

Стандартное решение, поставляемое компанией SAP AG, позволяет выполнять архивирование документов на SAP ContentServer с хранением их либо в базе SAP MaxDB, либо на файловой системе сервера. Решения по архивированию документов SAP позволяют в значительной степени расширить возможности архивирования, позволяя архивировать документы на внешних СХД, существенно снижая нагрузку на продуктивную БД и IT-бюджет на ее поддержание, одновременно повышая ее доступность

Под процессом архивирования в системах SAP ERP, SAP BI понимается фи-

зическое перемещение документов из базы данных системы SAP в некое другое хранилище с возможностью on-line поиска и просмотра документов через стандартные меню, предусмотренное в интерфейсе SAP. Основными кандидатами на архивирование являются документы с большим сроком хранения, документы, содержащие графические объекты или прикрепленные файлы, не используемые в режиме изменения документов.

В системе SAP ERP определено большое количество различных документов, почти все они могут быть архивированы. Наиболее целесообразно использовать процесс архивирования для следующих типов документов (по модулям системы):

- **модуль SAP HR (человеческие ресурсы):**
  - результаты оценки времени;
  - результаты расчета заработной платы;
  - долгосрочные документы менеджмента персонала;
- **модуль SAP FI (финансы):**
  - архивирование спецрегистров;
  - архивирование гибкой главной книги;
  - архивирование открытых позиций;
- **модуль SAP CO (контролинг):**
  - документы отдельных позиций регистрации затрат и выручки;
  - документ перерасчета затрат CO;
  - документы регистра материалов;
  - документы отдельных позиций калькуляционного учета результатов;
- **дополнительные документы, кандидаты для архивирования:**
  - сканированные документы, фотографии;
  - документы в форматах MS Word, MS Excel, Adobe и др.

В общем случае перечень документов, которые целесообразно архивировать, зависит от организации бизнес-процессов в компании.

В настоящее время многие компании-разработчики пошли по пути создания собственных решений для архивирования документов при работе с системами SAP. Большинство решений основано на замене стандартного SAP Content Server на собственную разработку. Этому способствует простота реализации методов работы сервера содержимого на языках программирования, доступность документации, понятная программа сертификации собственной разработки в компании SAP. Фактически сервер содержимого должен обеспечивать выполнение нескольких основных функций, а именно:

- сортировка, обновление или добавление документов в выбранный репозиторий документов;
- получение документов или их частей из репозитория документов;
- получение информации о документах или их содержимом;
- поиск входящего строки в документе;

- поиск документов по их атрибутам;
- удаление документов.

При прохождении сертификации встроенными в SAP систему функциями производится тестирование переданного на сертификацию продукта на соответствие спецификации SAP HTTP API 4.6, например, по сценарию SAP BC-HCS. Несмотря на возраст данной спецификации, она актуальна и для систем SAP с версиями ядра 7.0. Скорее всего, поддержка данного механизма архивирования будет сохранена и в последующих версиях SAP, несмотря на появление более современных механизмов для обеспечения стратегии ILM в SAP.

Рассмотрим наиболее часто встречаемые продукты, предназначенные для архивирования данных из систем SAP.

### IBM DB2 CommonStore for SAP

Это решение от компании IBM для архивирования данных из систем SAP. Несмотря на появления у SAP, начиная с версии 7.0, новых возможностей по управлению жизненным циклом информации основным методом работы IBM CommonStore является интерфейс, совмещенный с интерфейсом SAP ContentServer.

Одной из особенностей данного продукта является наличие прослойки – клиента, который позволяет через Интернет-браузер управлять архивированием объектов SAP, а также просматривать ранее сохраненные объекты. Другая черта продукта – интеграция с программным обеспечением резервного копирования – Tivoli Storage Manager (TSM). Именно эта связь продуктов и позволяет через дополнительный модуль для TSM выполнять архивирование на EMC Centera. Архитектура решения приведена на рис. 3.

При рассмотрении данного решения для выполнения архивирования из систем SAP на дисковый массив EMC Centera проявляются следующие недостатки:

- громоздкость и сложность решения (используются продукты IBM

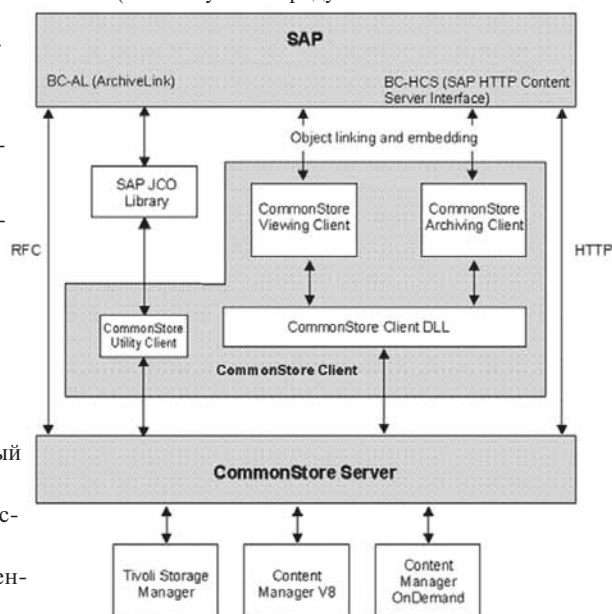


Рис. 3. Архитектура решения IBM DB2 CommonStore for SAP.



CommonStore, Tivoli Storage Manager, EMC Centera and backup recovery module);

- стоимость решения – все продукты лицензируются отдельно;
- ориентировано на инфраструктуру хранения одного производителя (IBM).

#### Archive Services for SAP (EMC)

Архитектура решения для архивирования представлена на рис. 4. Это комплексное решение одного из ведущих производителей ПО с большим количеством дополнительных функций. В качестве базового уровня архивного хранения предлагается EMC Centera.

Кроме выполнения архивирования на устройства EMC Centera, в решении присутствует множество интересных особенностей:

- просмотр архивированных документов через интернет-браузер осуществляется с использованием продукта ExternalViewing Services for SAP. Иными словами, архивные отчеты и документы могут просматриваться непользователями систем SAP;
- разработаны транзакции для работы с архивными данными и предоставляющие их в том же виде, что и стандартные транзакции SAP (Viewpoint for SAP);
- полная интеграция с продуктами линейки Documentum.

Также возможно использовать дополнительные опциональные модули, расширяющие функционал решения. Это, например, модуль управления жизненным циклом документов (Content Storage Services), если возможностей модуля окажется недостаточно, то модуль Retention Policy Services обеспечит расширенный набор функций ILM, вплоть до автоматического перемещения архивированных документов по различным носителям в зависимости от этапа жизни документа. Для повышения информационной безопасности разработан модуль, обеспечивающий возможности шифрования, цифровой подписи, цифрового шрединга (Trusted Content Services).

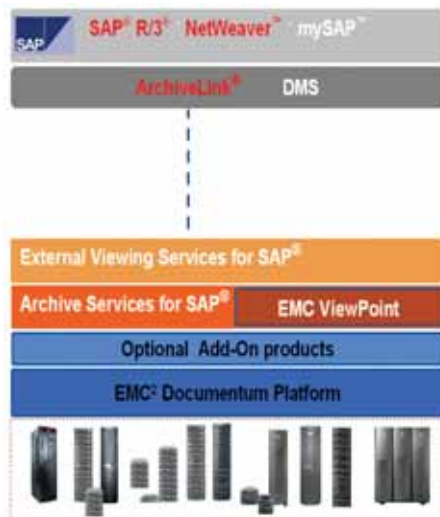


Рис. 4. Общая архитектура решения по архивированию данных модулей SAP с помощью решения Archive Services for SAP (EMC).

К недостаткам решения можно отнести:

- необходимость приобретения для обеспечения полноценности решения ряда дополнительных сервисов EMC и закупка лицензий;
- высокую стоимость решения.

#### Technoserv Content Server

Данное решение – результат разработки российской компании Техносерв АС – представляет собой универсальный инструмент архивирования для SAP. Решение Technoserv Content Server 1.2 предназначено для выполнения архивирования документов, находящихся в системе SAP, на специализированную систему архивного хранения, например, EMC Centera, доступ к которой осуществляется при использовании API архивного устройства (EMC Centera SDK) по локальной сети предприятия. Также данный продукт позволяет выполнять архивирование документов SAP на файловую систему сервера с установленным Technoserv Content Server (рис. 5).



Рис. 5. Общая архитектура решения по архивированию данных модулей SAP с помощью решения Technoserv Content Server.

Technoserv Content Server может работать на всех основных платформах, поддерживаемых SAP J2EE, а именно: Sun SPARC Solaris, IBM AIX, Windows.

TechnoServ Content Server, аналогично SAP Content Server, поддерживает работу с файловой системой сервера, на котором он установлен. Данная возможность позволяет организовать процесс разработки и тестирования задачи в стандартном, трехсистемном ландшафте SAP, без задействования реальной архивной системы хранения для систем разработки и тестирования.

Отличительные особенности решения Technoserv Content Server 1.2 заключаются в следующем:

- использование в качестве WEB сервера SAP J2EE, входящую почти в любой системный ландшафт SAP;
- полная совместимость с SAP Content Server. Настройка и работа с Technoserv Content Server осуществляются с использованием стандартных транзакций SAP;
- возможность модификации решения под конкретные требования заказчика;
- возможность использования в качестве места хранения архивной информации (помимо EMC Centera) широкого спектра архивных хранилищ при нали-

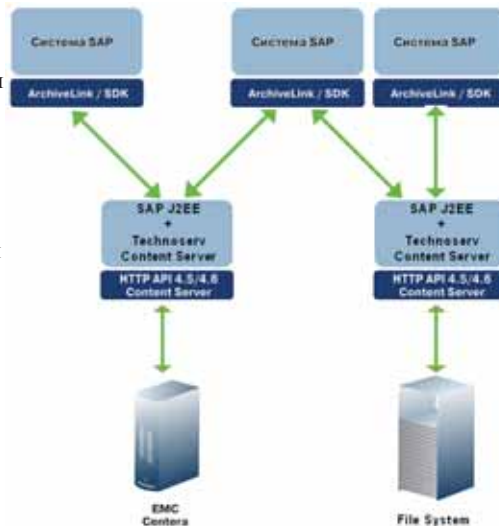


Рис. 6. Возможная архитектура использования решения Technoserv Content Server с несколькими SAP-системами и одновременной поддержкой двух способов архивации – на СХД и на файловую систему сервера.

чии у них возможности доступа к хранимой информации по API;

- наличие поддержки решения на территории РФ.

Возможная архитектура использования решения Technoserv Content Server с несколькими SAP-системами и одновременной поддержкой двух способов архивации – на СХД и на файловую систему сервера представлена на рис. 6. В настоящее время решение Technoserv Content Server сертифицировано компанией SAP AG для интеграции с SAP NetWeaver 7.00 (2004s) по сценарию интеграции BC-HCS.

В данном решении была реализована поддержка работы с архивным хранилищем EMC Centera. Его выбор был обусловлен хорошими эксплуатационными и функциональными характеристиками устройства, в частности, SDK EMC Centera поставляется под широкий перечень операционных систем и имеет поддержку режима 64 бита, что немаловажно при интеграции библиотек SDK EMC Centera и SAP J2EE.

Важным аспектом является обеспечение информационной безопасности архивируемых документов. Сама EMC Centera не поддерживает какого-либо шифрования объектов. Это можно сделать на уровне приложений – программно или на базе специализированного устройства. Вследствие невысокой интенсивности обращения к архивному хранилищу, возможно использование достаточно большого множества решений по шифрованию. Они могут быть реализованы как на базе отечественных разработок с поддержкой гостированных алгоритмов шифрования, так и на базе соответствующих разработок от западных вендоров. В основном, это решения на основе специализированных серверов (appliance), обеспечивающих шифрование/дешифрование данных “на лету”. При необходимости возможно встраивание механизма шифрования в разработанный Technoserv Content Server.

## Заключение

Процедуры архивирования документов SAP, сохраняя полную их прозрачность при доступе со стороны модулей/клиентов SAP, позволяют существенно уменьшать объемы продуктивных баз данных SAP, а общие объемы хранимых данных за счет многократных процедур бэкапирования сокращать в отдельных случаях многократно, что снижает не только эксплуатационные, но и капитальные затраты. Архивирование документов, сохраняемых только для аудита, также существенно повышает доступность онлайн-баз данных, одновременно снижая загрузку ИТ-персонала, а, соответственно, уменьшая и риски, связанные с сервисной поддержкой из-за снижения производительности приложений. Снижение загрузки процессоров серверов гарантирует их предсказуемую работу даже в период пиковых нагрузок и дает возможность совмещать во времени их дорогостоящее обновление.

## Комплексное решение для оценки рисков и DR-восстановления

**Апрель 2009 г.** — Корпорация Symantec анонсировала Veritas CommandCentral Disaster Recovery Advisor (VCC DRA) — программный инструмент управления, который заблаговременно выявляет риски и снижает расходы. Этот комплекс выявляет уязвимости конфигурации в любом центре обработки данных и гарантирует возможность восстановления приложений в случае внеплановых простоев.

VCC DRA быстро начинает оправдывать инвестиции, производя полную оценку надежности в течение двух дней после установки. Это достигается благодаря безагентной архитектуре и автоматической настройке его базы знаний, содержащей свыше 3000 характеристик рисков. Disaster Recovery Advisor проводит комплексную оценку рисков с помощью глубокого анализа использующихся приложений, серверов и систем хранения данных от всех ведущих поставщиков программного обеспечения и оборудования.

### Заблаговременное выявление риска гарантирует успешное послеаварийное восстановление

Исследование, проведенное Symantec в августе 2008 г., показало, что более чем в 30% случаев испытания систем послеаварийного восстановления терпят неудачу ввиду того, что организации остаются уязвимыми к недокументированным изменениям конфигурации или “отклонению конфигурации”. Для успешного послеаварийного восстановления требуется полный контроль и регистрация всех изменений конфигурации приложений, серверов и систем хранения данных, а также гибкий способ тестирования HA/DR. VCC DRA представляет собой комплексное решение для заблаговременного автоматического выявления скрытых рисков, связанных

с “отклонением конфигураций”, которые ставят под сомнение возможность успешного послеаварийного восстановления деятельности предприятия. При помощи VCC DRA ИТ-службы смогут автоматически выявлять тысячи таких проблем, оценивать риск для своих приложений, серверов и систем хранения данных и принимать действенные меры, эффективно используя имеющиеся у них ресурсы.

### Два дня на оценку рисков благодаря обширной базе знаний и возможности работы без программных агентов

VCC DRA обладает одной из самых обширных в отрасли баз знаний с тысячами описаний уязвимостей и критических проблем — почти всего, что может помешать ЦОД восстановить работу после аварии. Эта богатая база знаний позволяет заказчикам выявить риски в своей среде в течение двух дней после установки. Установка VCC DRA осуществляется быстро и гладко, так как решение работает без программных агентов и не требует никакой настройки.

База знаний непрерывно и автоматически пополняется, что гарантирует успешную реализацию планов HA/DR и защиту ЦОД от тех проблем HA/DR, которые только что обнаружены. Для любого из выявленных рисков VCC DRA автоматизирует процесс устранения, предоставляя рекомендации, которые помогают решить проблему до выполнения необходимых тестов DR, а это экономит время и другие ресурсы.

### Широкая поддержка обеспечивает восстановление и экономию

В рамках успешного плана тестирования HA/DR Disaster Recovery Advisor поддерживает ЦОД с неоднородным оборудованием и ПО — серверами, системами хранения и базами данных. Он поддерживает также ведущие решения кластеризации и управления хранением данных. VCC DRA работает со всеми основными операционными системами, а также дисковыми массивами и базами данных.

Disaster Recovery Advisor — это новейшее предложение Symantec в категории продуктов для хранения данных и обеспечения высокой готовности, которое преследует цель преобразования ЦОД путем повышения производительности, заблаговременного выявления рисков и сокращения затрат. С целью эффективной защиты и администрирования информации, Symantec недавно анонсировала Veritas Operations Services, платформу онлайн-услуг для выявления скрытых рисков при одновременном повышении производительности и эксплуатационной готовности.

Согласно исследованиям, проведенным Symantec в 2008 году:

- почти трети (30%) организаций приходится приводить в действие свой план послеаварийного восстановления каждый год, между тем, только 47% способны полностью восстановить работу в течение одной недели;
- 30% испытаний системы послеаварийного восстановления заканчива-

ются неудачей и почти треть (32%) организаций утверждает, что эти испытания повлияют на их заказчиков, а примерно пятая часть (21%) признает, что такие испытания могут повлиять на их продажи и доходы;

- 43% организаций испытывают проблемы, связанные с управлением изменениями конфигурации.

## Новые решения Cisco для центров обработки данных

**Январь 2009 г.** — Cisco расширила семейство коммутаторов Cisco Nexus™ и модернизировала коммутаторы серии Cisco Catalyst. В результате Cisco вывела на рынок новые устройства Cisco Nexus 7018, Cisco Nexus 5010 и Cisco Nexus 2000, специально адаптированные для жестких требований ЦОД нового поколения. Платформы Nexus предоставляют ИТ-отделам преимущества унифицированной коммутации. Они оптимизированы для поддержки архитектурных изменений и внедрения виртуализации, приложений Web 2.0 и сетевых вычислений (cloud computing). Технология Cisco облегчает виртуализацию ЦОД за счет повышения “интеллектуальности” сетей и их прозрачности для виртуализации. Новые технологии Cisco реализуют стратегию Cisco Data Center 3.0, направленную на внедрение унифицированной коммутации и унифицированных вычислений в центрах обработки данных.

### Новые продукты Nexus и усовершенствованные коммутаторы Catalyst 6500:

- коммутатор Cisco Nexus 7018 из серии продуктов Nexus 7000 имеет слоты с 18 слотами. 16 из этих слотов можно использовать для установки модулей ввода-вывода, что позволяет поддерживать до 512 портов 10 Gigabit Ethernet и выполнять требования самых крупных центров обработки данных;
- коммутатор Cisco Nexus 5010 с 28 портами устанавливается на одной полке (высота — 1 RU) и поддерживает технологии 10 Gigabit Ethernet, Cisco Data Center Ethernet (DCE), Fibre Channel over Ethernet (FCoE) и Fibre Channel. Это позволяет консолидировать трафик локальных сетей (LAN), сетей хранения (SAN) и серверных кластеров в единой унифицированной матрице коммутации;
- устройство Cisco Nexus 2000 Fabric Extender помогает ИТ-менеджерам ЦОД решать две важные задачи: повышать количество серверов и удовлетворять растущий спрос на полосу пропускания для каждого сервера. Устройства Cisco Nexus 2148T Fabric Extender, подключенные к сдвоенным коммутаторам Cisco Nexus 5020, могут поддерживать до 2.496 серверов Gigabit Ethernet, что резко повышает масштабируемость на уровне доступа без увеличения количества точек управления в сети.