EMC VMAX 3 – корпоративная платформа управления данными

В начале июля 2014 г. корпорация EMC объявила о выпуске систем семейства VMAX 3, которые трансформируют VMAX® из корпоративной системы хранения в корпоративную платформу управления данными. Такое решение позволяет достичь в центре обработки данных уровней гибкости, эффективности и контроля, сравнимых с облачными решениями, и тем самым принципиально меняет набор возможностей, который до сегодняшнего дня был недоступен для корпоративных систем хранения.

Евгений Пухов – технический эксперт направления унифицированных систем хранения, EMC Россия и СНГ.

Введение

7 июля корпорация EMC объявила о выпуске систем семейства VMAX 3, которые архитектурно являются продолжением модельного ряда VMAX второго поколения (10K, 20K, 40K), а идеологически предлагают целый ряд новых подходов, призванных упростить управление жизненным циклом информации.

Почему новая VMAX называется корпоративной платформой управления, а не хранения данных? Безусловно, VMAX это, в первую, очередь система хранения класса High-end для бизнес критичных нагрузок первого уровня. Она является платформой управления потому, что позволяет заказчикам иметь полный контроль над сервисами данных в терминах SLA и инфраструктурой там, где размещены приложения — в центре обработки данных или в публичном облаке. Если раньше такое было возможно только при интеграции сторонних сервисов управления данными и внешнего ПО, то теперь все управление жизненным циклом информации, а также реализация модели «хранение данных как услуга» в гибридном облаке с предсказуемым уровнем обслуживания возможна напрямую базовыми сервисами системы хранения/управления данными VMAX.

Новый модельный ряд состоит из трех моделей – VMAX 100K, 200K, 400K. Принципиально новая архитектура VMAX 3 основана на операционной системе HYPERMAX OS и архитектуре Dynamic Virtual Matrix. HYPERMAX OS — это первое в отрасли открытое конвергентное решение, объединяющее гипервизор систем хранения и операционную систему. Благодаря этому в VMAX3 можно встраивать сервисы инфраструктуры хранения данных (такие как облачный доступ, поддержка мобильности данных и защита данных) напрямую в массив. Это принципиально повышает уровень эффективности и консолидации центра обработки данных за счет уменьшения занимаемых площадей и снижения энергопотребления. Архитектура Dynamic Virtual Matrix позволяет динамически выделять вычислительные ресурсы для повышения производительности и обеспечения предсказуемых уровней обслуживания в масштабах крупного предприятия.

Архитектура

Новая версия VMAX, как и все предыдущие, построена на многоконтроллерной аппаратной архитектуре INTEL. Только на этот раз контроллеры СХД объединены отказоустойчивой шиной данных – Infiniband 56 Гбит/с. Суммарный когерентный кэш достигает 16 Тбайт в старших моделях, количество вычислительных ядер доходит до 384 (рис. 1).

В VMAX 3 произошло множество конструктивных изменений. Изменились контроллеры, дисковые полки, изменились и сами стойки: Теперь VMAX 3 поставляется в стандартных 19” шкафах и допускает поддержку стоек сторонних производителей. Таким образом, не нужно как-то особенно планировать размещение VMAX, что особенно актуально для больших датацентров. В VMAX больше нет разделения на system bay и storage bay. Все стойки полностью равноправны. В каждой стойке может размещаться от 1 до 2 Engine (1 Engine = 2 контроллера), интерконнект и некоторое количество дисковых полок. Существует два варианта дисковых полок – и оба с плотным размещением: 60 в 3.5 дюймовом формате и 120 в 2.5 дюймовом формате (рис. 2).

Рис. 2. Полка с плотным размещением дисков.

За счет этого увеличилась общая плотность массива (рис. 3): в одной стойке можно поставить 720 дисков и два контроллера (1 Engine) или 480 дисков и четыре контроллера (2 Engine). Максимальная конфигурация массива помещается в 8 шкафов: 5760 дисков и 8 Engine.

Стойки VMAX можно разносить на расстояние до 25 метров от первой, в которой установлена внутренняя коммутация (рис. 4).

VMAX 3 поддерживает следующие типы интерфейсов ввода-вывода:

– FC 8 и 16 Гбит/с;

Рис. 1. Базовые аппаратные характеристики VMAX 3.

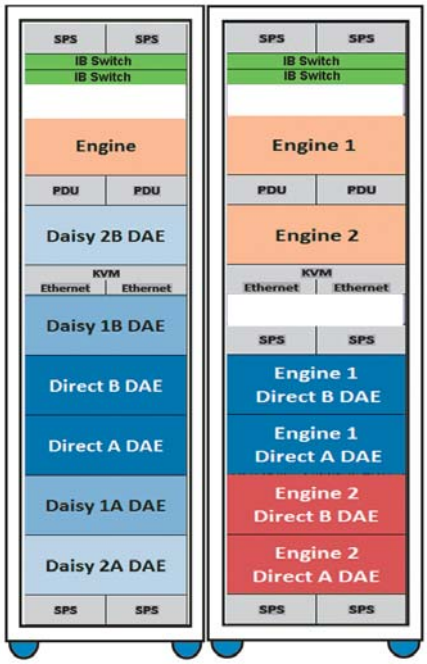

Рис. 3. Варианты размещения контроллеров и полок VMAX внутри стоек. 1-Engine и 2-Engine конфигурации.

Рис. 4. Возможность разнесения стоек VMAX на расстояние до 25 метров от system bay 1.

– FCoE/ iSCSI 10 Гбит/с.

Ассортимент дисков постоянно расширяется, и по состоянию на август 2014 г. поддерживаются:

– SSD 200, 400, 800 Гбайт;

– SAS 15K 300 Гбайт;

– SAS 10K 300, 600, 1200 Гбайт;

– NL-SAS 7K 2, 4 Тбайт.

Впервые для защиты кэш-памяти применена технология Vault to flash. Раньше для этой цели был предусмотрен специальный раздел на жестких дисках vault, теперь же это отдельная Flash-память, установленная в сами контроллеры. Это позволило снизить емкость батарей, используемых для защиты кэша: поддержание работоспособности жестких дисков vault теперь не нужно.

Основная операционная система массива поменяла свое название на HYPERMAX (ранее Enginuity), и название это выбрано неслучайно. Все дело в том, что, помимо операционной системы, на контроллерах работает специализированный гипервизор, который позволяет запускать большое количество дополнительных сервисов, таких как: мониторинг, управление, файловый доступ. Особенно следует отметить возможность интеграции виртуальной версии VPLEX для создания катастрофоустойчивых инфраструктур без необходимости дополнительного оборудования.

Рис. 5. Динамическая балансировка процессорных ядер между задачами.

Появилась динамическая внутренняя балансировка нагрузки. Если раньше в Enginuity была жесткая привязка процессорных ядер к портам ввода-вывода, то в HYPERMAX она исчезла, и вычислительная мощность балансируются между Front-end, back-end и встроенным гипервизором (рис. 5).

Учитывая огромное количество ресурсов (VMAX 400K в максимальной конфигурации поддерживает 384 Intel ядра и 16 Тбайт кэш-памяти) в сочетании с невероятной гибкостью динамического перераспределения ресурсов, можно говорить о новом VMAX как одной из самой мощной СХД класса High-End на сегодняшний день.

Планирование ресурсов VMAX

В новой версии принципиально изменился подход к сайзингу системы и динамическому выделению ресурсов. Если раньше внедрение и настройку системы VMAX можно было сравнить с целым небольшим проектом, то теперь VMAX поставляется в полностью предконфигурированном состоянии с разбивкой на RAID, с преднастроенными виртуальными пулами ресурсов и упрощенным интерфейсом управления. Предварительная конфигурация с соблюдением уровней SLA собирается на заводе в полном соответствии с профилем нагрузки, который был согласован с заказчиком на этапе сайзинга и планирования. На этапе финального внедрения в эксплуатацию на площадке заказчика необходимо просто подключить преднастроенные логические тома к приложениям, и можно начинать работу с гарантированным уровнем SLA!

Стандартно предлагаются несколько уровней SLA: DIAMOND, PLATINUM, GOLD, SILVER, BRONZE (точнее, в терминологии VMAX это называется SLO – Service Level Objective), разделенным допустимым для приложения временем отклика, а также профилем нагрузки (рис. 6). Принимаются во внимание следующие параметры:

– среднее время отклика;

– соотношение чтение/запись;

– профиль нагрузки sequential/random;

– средний блок чтения/записи;

– объем данных.

Если на системе планируется запускать несколько разных задач, необходимо предопределить профиль нагрузки каждой из них. Отсутствие интерференции полностью гарантируется базовыми сервисами HYPERMAX.

Чем больше информации о нагрузке было известно на этапе проектирования, тем точнее VMAX сможет планировать свои ресурсы предсказывая узкие места. Говоря иначе, настройка VMAX 3 начинается не с момента подключения системы в датацентре заказчика, а с момента сбора информации перед заказом системы.

Федеративное хранение

Одна из ключевых функциональностей, появившихся в VMAX некоторое время назад, – Federated Tiered Storage (FTS), которая в новой версии была существенно расширена и дополнена.

FTS, во-первых, позволяет консолидировать на базе VMAX массивы различных производителей, а, во-вторых, распространить на подключенный пул массивов такие технологии, как FAST, SRDF, TimeFinder. Ключевое отличие FTS от похожих реализаций других производителей заключается в том, что VMAX при записи данных на сторонний массив обязательно производит проверочное чтение и контроль CRC.

Применимость FTS весьма разнообразна: это использование сторонних массивов в составе многоуровневого хранения FAST; для хранения снимков Timefinder совместно с SRDF для удаленной репликации. Интересна реализация совместно с VPLEX для создания Active/Active конфигураций с обеспечением непрерывного доступа к данным и полной устойчивости относительно отказа одного из датацентров целиком. Совместно с VPLEX увеличивается мобильность данных, а также появляется возможность совместной работы с единым географически распределенным кросс-платформенным пулом ресурсов (рис 7).

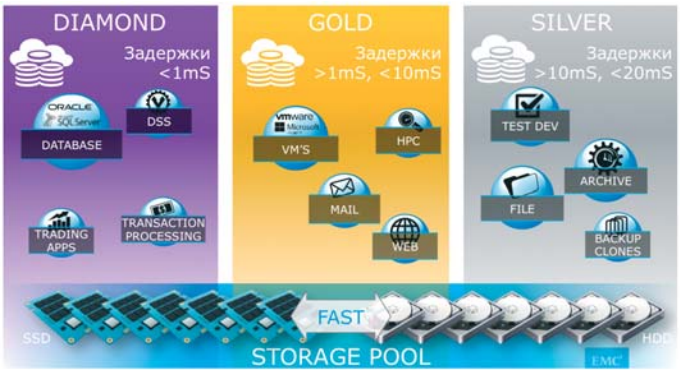

Рис. 6. Уровни SLA с классификацией по времени отклика.

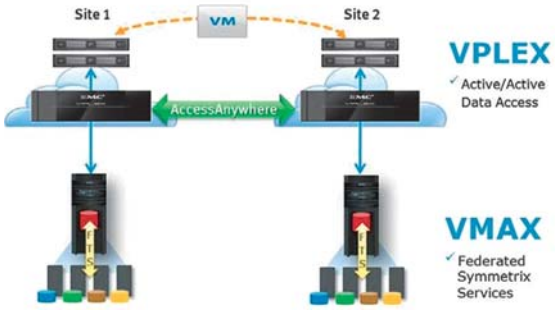

Рис. 7. FTS совместно с VPLEX.

В VMAX 3 появилась совершенно новая технология, основанная на алгоритмах FTS: EMC ProtectPoint, позволяющая подключать систему резервного копирования EMC DataDomain (DD) как Federated Tiered Storage и осуществлять резервное копирование и восстановление данных напрямую с DD.

Работает все это следующим образом: выполняется приостановка БД, агент ProtectPoint дает команду на снятие мгновенного снимка TimeFinder, работа БД восстанавливается. После этого в фоновом режиме VMAX осуществляет копирование содержимого снимка на DD.

С точки зрения базы данных и приложений, при внедрении ProtectPoint процедура РК не меняется, и все выглядит очень похоже на обычное резервное копирование с использованием мгновенных снимков, за тем исключением, что трафик РК не прогоняется по всей сети между сервером приложений и сервером РК, а локализован между VMAX и DD.

Восстановление данных также выполняется по команде агента ProtectPoint, установленного на сервере приложения/БД. Администратор вручную или посредством любого поддерживаемого ПО резервного копирования, выбирает нужную контрольную точку: начинается процесс восстановления данных с резервной копии DD. Ключевая особенность ProtectPoint – в том, что пользоваться восстановленным томом можно сразу после инициализации процедуры восстановления, не дожидаясь завершения копирования данных. Все запрашиваемые блоки будут приоритетно прочитаны с DD и предоставлены приложению. Конечно, это вызовет дополнительные задержки, но промежуток времени до полного физического восстановления тома можно потратить на служебные функции проверки содержимого, прав доступа, совместимости и т.п. В совокупности с высокой скоростью канала между DD и VMAX это сильно снижает время восстановления после сбоя (рис. 8).

Многоуровневое хранение

Многоуровневое хранение остается самой сильной стороной массивов EMC уже на протяжении 6 лет, когда в 2008 г. корпорация впервые на рынке предложила Flash-диски для использования в составе одного из типов носителей в VMAX.

Алгоритмы, по которым перемещаются данные между уровнями хранения, нацелены на поддержание заранее заданного SLA (SLO). Если раньше решение о перемещении между уровнями было продиктовано частотой запросов того или иного блока, то теперь решение принимается исходя из поддержания заранее заданного времени отклика и других параметров (при условии, что мы их предопределили).

Анализ данных и перемещение между уровнями хранения теперь работает непрерывно 24x7. На производительность это не влияет, т.к. VMAX наделен огромными вычислительными ресурсами даже в минимальной конфигурации (рис. 9).

Другое важное изменение в FAST: если раньше центр принятия решения о перемещениях между уровнями находился непосредственно в ядре VMAX, то теперь в этом помогают агенты, устанавливаемые на серверы приложений и баз данных. От агентов поступает информация по наиболее востребованным областям данных и возможным изменениям профиля нагрузки. FAST на VMAX стал проактивным.

Анализ производительности СУБД

Теперь с помощью Unisphere можно не только управлять массивом, но и анализировать производительность баз данных: в интерфейс добавлена утилита DBclassify, которая умеет собирать информацию непосредственно с сервера СУБД и передавать ее в Unisphere. Это позволяет рассматривать Unisphere как единую точку мониторинга производительности, как со стороны массива, так и со стороны сервера приложений. DBclassify может сравнить время отклика IO на стороне системы хранения и время отклика на стороне приложения/ базы данных и сделать вывод о том находится ли узкое место внутри массива или его нужно искать где-то вовне: в сетевом стеке, файловой системе и т.п. DBclassify дает табличное представление, таким образом можно увидеть “узкое” место в какой-то конкретной области БД, и с помощью технологий FAST дать команду на перемещение таблицы в другой пул ресурсов. Применение DBclassify нацелено преимущественно на Oracle, но ожидается поддержка для других БД.

Unisphere, DBclassify и все встроенные функции по мониторингу, управлению системы находятся в ядре ОС HYPERMAX и не требуют сторонних серверов управления.

Заключение

Новое поколение систем VMAX – не только быстрее но и функциональнее предыдущих. Целый ряд новых подходов в сочетании с высокой производительностью позволяет говорить о VMAX не как о системе хранения, а как о системе управления данными. VMAX – это уже не просто СХД для облаков, а полноценная облачная инфраструктура хранения для решения задач “гиперконсолидации” бизнес-критичных сервисов.

Позднее в новые системы будут интегрированы технологии только что приобретенной компании TwinStrata, благодаря которым можно будет отправлять редко используемые данные в публичные облачные хранилища. Дисковые массивы VMAX 3, как ожидается, начнут поставляться в третьем квартале 2014 г. Они будут также снабжены дополнительными функциями, обеспечивающими поддержку виртуализованных вычислительных ресурсов.

Евгений Пухов,
EMC Россия и СНГ.

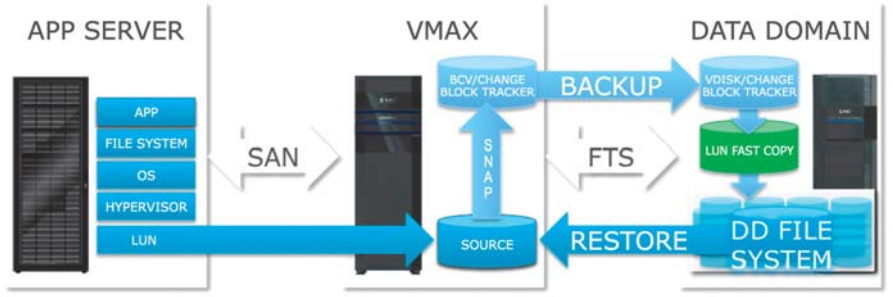

Рис. 8. EMC ProtectPoint.

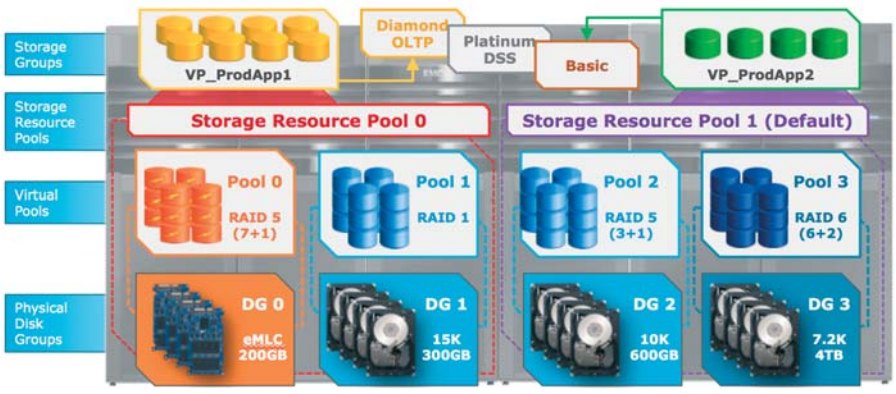

Рис. 9. Service Level Objective (SLO) и виртуальные пулы.