

# SAP HANA для петабайтных хранилищ



**Вадим Табаков** — к.э.н., менеджер по развитию направления "Базы данных и технологии", САП СНГ.

Общее ускорение процессов практически всех сфер жизнедеятельности вызывает лавинообразный рост объемов информации, повышение уровня сложности данных, их многообразия и формализации. Стремление к удовлетворению полноты и объективности информации требует применения технологий и методов, способных не только хранить и обрабатывать данные, избавляя их от противоречивости и множественности, но также позволяющих своевременно обеспечить ту или иную форму производственной деятельности необходимой целостной и полезной информацией (репрезентативным набором данных), избежать фрагментации при обработке данных и в целом информационного кризиса не только в рамках некой замкнутой, изолированной системы, такой как бизнес-процесс, предприятие, или даже социальная группа, но и на уровне целых отраслей и экономик.

Существующая практика фрагментации обработки данных в рамках различных процессов, систем, приложений, видов деятельности и групп пользователей приводит к эффекту комбинационного взрыва, в котором каждые вновь полученные данные увеличивают число взаимосвязей в геометрической прогрессии. Повышение многообразия состава вычислительной инфраструктуры вызывает чрезмерную сложность ее операционной эксплуатации и, как следствие, приводит к полной парализации управления и обработки информации.

Отдельные технологии, способные в той или иной степени решать проблему фрагментации данных, позволяют хранить и обрабатывать огромные объемы информации, преодолевая информационный барьер. Однако основной вопрос применения технологии, как и ее ценность, заключается не только в скорости обработки и передачи потоков данных в единицу времени, но и в возможности формирования репрезентативной выборки на основе генеральной совокупности данных, и в последующем применении индуктивных методов для установления закономерностей и построения достоверных аналитических прогнозов.

Таким образом, современная технология для работы с данными стремится к совмещению целого ряда решений (суммы технологий) на базе единой интеграционной платформы, обладающей целостным единством (что, тем не менее, не подразумевает ее абсолютную универсальность!), и призванной решать сложнейшие задачи управления данными в конвергентных информационных средах в режиме реального времени. В SAP такая концепция реализована в интеграционном технологическом комплексе SAP Real Time Data Platform, центральным компонентом которого является SAP HANA<sup>1)</sup>.

Снижение стоимости вычислительных мощностей является предпосылкой к применению SAP HANA не только в качестве акселератора отдельных задач, а к полному переходу от использования дисковых подсистем к реализации задач хранения и управления данными в оперативной памяти. С другой стороны, необходимо отметить, что ежедневный анализ терабайтов данных в оперативной памяти снижает общую производительность вычислительной архитектуры. В этом случае полезный объем оперативной памяти целесообразно использовать для хранения и обработки данных в рамках горизонта анализа (как правило, такой горизонт составляют данные полученные не менее чем за последние три года деятельности организации), а дисковые подсистемы и традиционные СУБД — для хранения исторических и архивных сегментов, требующих значительных объемов дискового пространства.

Одним из примеров подобной архитектуры служит использование СУБД SAP Sybase IQ в качестве NearLineStore (NLS) для SAP NetWeaverBusinessWarehouse. Такая реализация позволяет эффективно и гибко управлять жизненным циклом и обменом информации, учитывая все преимущества указанных технологий.

Использование SAP HANA совместно с такой масштабируемой распределенной инфраструктурой как Hadoop расширяет пределы информационной емкости, позволяя организовывать хранилища данных, поддерживающие петабайты информации<sup>2)</sup>. Преимущества Hadoop предоставляют возможность реализовывать алгоритмы пакетной обработки данных (в задачах, для которых время отклика не является критичным), архивации и ретроспективного анализа данных (доступ к которым не требует оперативности реального времени). Hadoop в этом случае используется для агрегации результатов из различных источников, предметных областей с последующей репликацией репрезентативной выборки в SAP HANA для реализации сложных аналитических расчетов и формирования отчетности.

1) SAP HANA (HighPerformanceAnalyticAppliance) – технология, предоставляющая возможности хранения и обработки данных оперативной и обеспечивающая высокую производительность при обработке огромных массивов данных.

2) The SAP HANA One Petabyte Test: <http://www.saphana.com/community/blogs/blog/2012/11/12/the-sap-hana-one-petabyte-test>

К примеру, в области телекоммуникаций SAP HANA решает вопрос анализа количества обработанных вызовов, в то время как на Hadoop ложится задача измерения нагрузки сети.

Объединение сегментов данных из Hadoop с данными оперативного анализа SAP HANA позволяет расширить область применения аналитики. Так, к примеру, в области розничной торговли данные с кассовых терминалов (PointofSale, POS), хранящиеся в SAP HANA, через определенные временные интервалы могут быть агрегированы с данными о состоянии складских запасов из файловой системы Hadoop, что позволяет реализовывать ABC/XYZ анализ, прогнозировать характер потребительского спроса и определять оптимальные объемы складских запасов.

В настоящее время SAP поддерживает партнерские отношения с тремя ключевыми поставщиками Hadoop, такими как Cloudera, Hortonworks и Intel.

Выполнение прикладных вычислений и реализации статистических алгоритмов осуществляется динамически непосредственно в SAP HANA с помощью подключаемых библиотек, таких как: IMSL (International Math&Statistics Lib), R, BFL (Business Function Lib), PAL (Predictive Analysis Lib), L. Это позволяет удовлетворить такое требование парадигмы BigData, как Velocity (скорость обработки), снижая время отклика систем до 90% по сравнению с дисковой архитектурой.

Одним из наиболее красноречивых примеров прикладного использования технологий SAP HANA и Hadoop-инфраструктуры является архитектура, реализованная компанией Mitsui Knowledge Industry (MKI), которая применяется для сборки и анализа генома человека при диагностировании и лечении онкологических заболеваний в процессе сравнения ДНК заболевших с ДНК здоровых пациентов. В реализованной архитектуре сотни терабайт неструктурированных данных, содержащих информацию о фрагментах структур ДНК, собираются для предварительной обработки в Hadoop-кластере, после чего с помощью аналитического инструментария R в SAP HANA выполняется анализ полученных фрагментов и восстановление последовательностей ДНК-хромосом. Результаты тестирования показывают, что производительность SAP HANA для решения этой задачи в 408 тысяч раз выше, чем у традиционных дисковых СУБД. Такой подход позволяет снизить время клинического исследования (в процессе которого пациент находится в стационаре) с 2–3 дней до 20 минут, т.е. в 216 раз быстрее, чем это происходит сейчас. Это позволяет врачу буквально на месте в режиме реального времени собрать геном пациента, провести его сравнительный анализ и выбрать метод лечения. Также повышение скорости анализа в общем снижает время и затраты на прикладные исследования, поиск и синтезирование новых лекарственных препаратов.

Таким образом, комплементарное использование SAP HANA, Hadoop и традиционных технологий управления данными позволяет снизить сложность ИТ-инфраструктуры, эффективно распределить вычислительную нагрузку и логически сегментировать большие массивы данных. При этом Hadoop может выступать в роли "горячего архива данных", поскольку позволяет не только хранить данные, но также обеспечивает пользователям постоянный и оперативный доступ к необходимой информации. Hadoop и in-метогу хранилище на SAP HANA также технологически дополняют друг друга в рамках процесса информационного обмена. При помощи Hadoop может быть также организован "холодный архив данных", выступающий альтернативой традиционному ленточному резервному хранилищу, а масштабируемые вычислительные мощности Hadoop позволяют проводить параллельную обработку сложных запросов и структур данных, а также реализовывать механизмы поиска скрытых закономерностей и полезных знаний в массивах данных.

## RSA: адаптивное управление идентификацией и доступом

**Май 2013 г.** — Компания RSA, подразделение безопасности корпорации EMC® (NYSE: EMC), опубликовала новый обзор RSA "Adaptive IAM: Defending the Borderless Enterprise" (Адаптивные системы IAM: защита для предприятия без границ), в котором содержится призыв к организациям модернизировать концепцию и подходы в отношении управления идентификацией и доступом (Identity and Access Management — IAM). В обзоре рассматриваются причины, по которым системы IAM занимают передовые позиции в борьбе с кибератаками, и способы трансформации традиционных решений для обеспечения их соответствия требованиям предприятия и реальности современного ландшафта угроз.

Задача, которая традиционно выполняется решениями IAM, — это обеспечение доступа соответствующих пользователей к соответствующим корпоративным ресурсам. На сегодняшний день наблюдается новый всплеск интереса к решениям IAM как к возможному способу дальнейшего усиления безопасности современных расширенных и не имеющих границ корпоративных сред. Однако традиционным решениям IAM еще только предстоит пройти определенный путь развития, чтобы обеспечить поддержку растущего числа пользователей, партнеров и облачных сервисов, которые используют для доступа к корпоративным ресурсам конечные устройства и приложения, часто выпадающие из сферы непосредственного контроля корпоративных ИТ-отделов.

В обзоре вводится концепция адаптивной системы IAM, которая в ближайшие несколько лет трансформирует традиционные системы IAM в более динамичные,

гибкие, интеллектуальные системы, способные учитывать и оценивать риски. В условиях, когда корпоративные системы идентификации постоянно подвергаются повышенной опасности, одна успешная операция входа в систему больше не является достаточным основанием для доверия. Уровень контроля безопасности должен корректироваться на основании каждой транзакции, участвующих сторон и информационных ресурсов, подвергаемых риску. Адаптивная система IAM будет контролировать ситуационный периметр, который будет принудительно активировать необходимые меры безопасности при каждом доступе пользователей к корпоративным данным и ресурсам.

Адаптивная система IAM определяется четырьмя новыми характеристиками:

- *расширенные профили пользователей* используются для сравнения действий и поведения пользователей в реальном времени с историческим базовым показателем, в результате чего выявляются существенные отклонения от "нормального" поведения, свидетельствующие о наличии проблем с безопасностью;
  - *аналитика больших данных* применяется к массивным наборам данных для оценки рисков и различения "нормального" и "отклоняющегося" поведения;
  - *функции мониторинга и вмешательства на основе оценки рисков* отслеживают действия пользователей после прохождения первоначальной аутентификации и корректируют средства контроля доступа в соответствии с рассчитанным уровнем риска. Работа пользователей прерывается запросами дополнительной аутентификации только в случае обнаружения небезопасных действий;
  - *интерфейс потребительского уровня* предполагает, что контроль идентификации и оценка рисков должны выполняться незаметно для пользователя, не мешая его работе, и визуализироваться в интерфейсе только при необходимости.
- Чтобы ускорить переход заказчиков на более адаптивные системы IAM, компания RSA также анонсировала выпуск новых средств интеграции:
- *расширенный профиль пользователя*. Разработанный RSA лидирующий в отрасли механизм оценки рисков реализован в недавно выпущенных ПО RSA® Authentication Manager 8 и RSA® Adaptive Authentication и предназначен для прозрачного сбора информации от различных устройств, пользователей и сред для определения "нормального" поведения пользователей. Для повышения надежности решений об аутентификации и авторизации последняя версия ПО RSA® Adaptive Directory 6.1 позволяет организациям агрегировать и централизованно администрировать информацию об идентификации в локальных хранилищах данных и в облачных приложениях, чтобы затем на ее основе создавать расширенные профили пользователей;
  - *аналитика в реальном времени для оценки рисков и интеграция со средствами контроля на базе оценки рисков*. Более тесная интеграция между ПО RSA® Access

Manager 6.2, ПО RSA Adaptive Authentication и ПО RSA Authentication Manager 8 позволяет заказчикам связать анализ рисков с более надежной аутентификацией и средствами контроля доступа;

- *удобство работы пользователей*. Новые выпуски ПО RSA® Adaptive Federation 1.5 (для модели "ПО как услуга") и RSA® Federated Identity Manager (для локальной установки) обеспечивают прозрачный единый вход для облачных приложений.

Системы IAM следующего поколения позволяют создать унифицированное представление о пользователях, способное учитывать возрастание числа пользователей за счет использования облачных и мобильных платформ, и лучше идентифицировать попытки мошеннического и злонамеренного доступа к корпоративным ресурсам, причем все это достигается с минимальным влиянием на работу обычных пользователей. По прогнозам RSA, адаптивные системы IAM в будущем станут ключевым компонентом программ обеспечения безопасности на базе интеллектуальных алгоритмов и помогут организациям обеспечить защиту ценных корпоративных ресурсов и идентификационных данных в гетерогенных средах, объединяющих доверенные и не доверенные ИТ-инфраструктуры.

## Seagate: новые SSD

**Май 2013 г.** — Компания Seagate Technology анонсировала свой новый портфель решений на основе технологии флэш-памяти. Компания представила свой первый твердотельный накопитель (SSD) для конечных потребителей и следующее поколение SSD корпоративного уровня. Новая линейка накопителей, включающая Seagate® 600 SSD, Seagate® 600 Pro SSD и Seagate® 1200 SSD, была создана для обеспечения высочайшей скорости передачи данных и их целостности. В качестве дополнения к новому портфелю компания Seagate также представила решение Seagate® X8 Accelerator для СХД, являющееся картой памяти PCIe. Карта ускорения основана на технологии компании Virident и предлагает возможность работы со всеми ключевыми интерфейсами, включая SATA и SAS.

*Seagate 600 SSD* — один из самых передовых на текущий момент накопителей для апгрейда ноутбука. Пользователи ноутбуков смогут загружать свой компьютер в четыре раза быстрее и запускать приложения в два раза быстрее по сравнению со стандартными жесткими дисками, значительно сокращая время ожидания доступа к контенту.

Seagate 600 SSD обеспечивает высокую степень быстродействия и прост в установке в качестве апгрейда традиционного жесткого диска. Новый SSD от Seagate доступен в различных вариантах высоты корпуса: модель SSD высотой 5 мм идеальна для ультратонких устройств, а модель традиционной высоты подойдет для стандартных ноутбуков. Диск обладает емкостью до 480 Гбайт и поставляется в форм-факторе 2,5 дюйма. Накопитель оснащен интерфейсом SATA 6 Гбит/с, обеспечивающим быструю передачу данных и простоту подключения.