



<https://doi.org/10.21122/1683-6065-2023-2-64-69>
УДК 669.004

Поступила 29.03.2023
Received 29.03.2023

ТЕХНИЧЕСКИЙ ПОТЕНЦИАЛ ПЛАТФОРМЫ SAP HANA

М. А. КОВАЛЕВ, ОАО «БМЗ – управляющая компания холдинга «БМК»,
г. Жлобин, Гомельская обл., Беларусь, ул. Промышленная, 37. Тел.: 80233455740.

В статье описаны технические возможности платформы и базы данных SAP HANA, являющейся основой для развертывания инновационной ERP-системы SAP S/4 HANA и производных от нее модулей. Описаны принципы работы, основные инструменты и возможности платформы. Определены требования к программному и аппаратному обеспечению для развертывания и сопровождения данной платформы на крупных производственных предприятиях. Проведены практические исследования скорости работы на данной платформе и обзор некоторых инструментов. Платформа разработана с использованием передовых технологий, обеспечивающих высокую скорость обработки информации и безопасное ее хранение, удовлетворяя требованиям современного бизнеса и пользователей.

Ключевые слова. SAP HANA, In-Memory Computing, ACID, NewSQL, Unified Tables, Persistent Storage, SAP HANA Studio, Machine Learning, SAP Cloud Platform, SAP HANA PlanVisualizer.

Для цитирования. Ковалев, М. А. Технический потенциал платформы SAP HANA / М. А. Ковалев // Литье и металлургия. 2023. № 2. С. 64–69. <https://doi.org/10.21122/1683-6065-2023-2-64-69>.

TECHNICAL POTENTIAL OF THE SAP HANA PLATFORM

М. А. KOVALYOV, OJSC “BSW – Management Company of Holding “BMC”,
Zhlobin city, Gomel region, Belarus, 37, Promyshlennaya str. Tel.: 80233455740.

The article describes the technical capabilities of the SAP HANA platform and database, which is the basis for the deployment of the SAP S/4 HANA innovative ERP system and modules derived from it. The principles of operation, the main tools and capabilities of the platform are described. The requirements for software and hardware for the deployment and maintenance of this platform at large manufacturing enterprises are defined. Practical studies of the speed of work on this platform and an overview of some tools have been carried out. The platform is developed using advanced technologies that provide high-speed information processing and secure storage, meeting the requirements of modern business and users.

Keywords. SAP HANA, In-Memory Computing, ACID, NewSQL, Unified Tables, Persistent Storage, SAP HANA Studio, Machine Learning, SAP Cloud Platform, SAP HANA PlanVisualizer.

For citation. Kovalyov M. A. Technical potential of the SAP HANA platform. Foundry production and metallurgy, 2023, no. 2, pp. 64–69. <https://doi.org/10.21122/1683-6065-2023-2-64-69>.

Современные бизнес-пользователи должны намного быстрее реагировать на изменение ситуации у клиентов и на рынке. Им необходим динамический доступ к необработанным данным в реальном времени. SAP HANA предоставляет пользователям функции гибкого и оперативного моделирования данных и сводит к нулю для пользователя время ожидания внесения изменений в модели данных и выполнения задач по администрированию баз данных.

SAP HANA – это in-memory база данных и платформа для разработки приложений. Она хранит данные в оперативной памяти, используя для этого поколоночный способ хранения. Степень сжатия данных при этом может достигать 10:1. Архитектура SAP HANA обеспечивает как высокоскоростную обработку транзакций, так и работу со сложными аналитическими запросами, совмещая решение этих задач в рамках единой платформы.

In-Memory Computing – технология размещения данных непосредственно в ОЗУ (RAM) выделенных серверов, а не в сложных реляционных базах данных, работающих на сравнительно медленных дисках. Технология In-Memory Computing получила широкое применение в обработке больших массивов данных (Big Data), помогая сократить временные издержки на обработку запросов, для быстрого обнаружения паттернов, анализа больших объемов данных на лету для различных бизнес-клиентов, в том числе розничной торговли, банков, ресторанов, казино, коммунальных услуг и т. д.

SAP HANA использует технологию NewSQL, сочетающую в себе преимущества NoSQL и транзакционные требования классических баз данных. Основные преимущества:

- приложения взаимодействуют в основном посредством SQL;
- транзакции полностью поддерживают требования ACID;
- при управлении не применяется механизм блокировок – таким образом, исключается вероятность конфликта между записываемыми и считываемыми в реальном времени данными;
- архитектура shared-nothing, при которой каждый узел системы отвечает за свой набор данных, являясь независимым и самодостаточным, подразумевает легкую и быструю масштабируемость;
- производительность узла системы управления базами данных (СУБД) на NewSQL намного выше, чем производительность традиционных реляционных СУБД;
- скорость отклика системы в 50 раз выше скорости отклика традиционных реляционных СУБД.

SAP HANA использует подход Unified Tables (рис. 1), который обеспечивает высокую скорость чтения и записи данных в таблицу поколоночного хранения. Такой механизм позволяет быстро осуществлять транзакции (т.е. запись новых строк), анализировать данные с высокой скоростью за счет поколоночного хранения в сжатом виде, параллельной обработки данных, а также хранить все данные в оперативной памяти (in-memory). При проведении записи изменения не сразу вносятся в основное место хранения таблиц. Вместо этого все правки заносятся в отдельную структуру данных – дельта-хранилище (L1-delta). Здесь данные хранятся в оптимизированном для записи формате. Когда необходимо перенести изменения из дельта-хранилища, то запускается специальный процесс Delta merge – слияние дельты. Сначала данные из L1-delta преобразуются в поколоночный формат в L2-delta, а затем объединяются с основным хранением данных (main store). А для механизма чтения данных все три области хранения информации (L1-delta, L2-delta и main store) предоставляют данные в целостном виде. Благодаря этому процессу получается обеспечить высокую скорость записи и анализа данных (рис. 2–4).

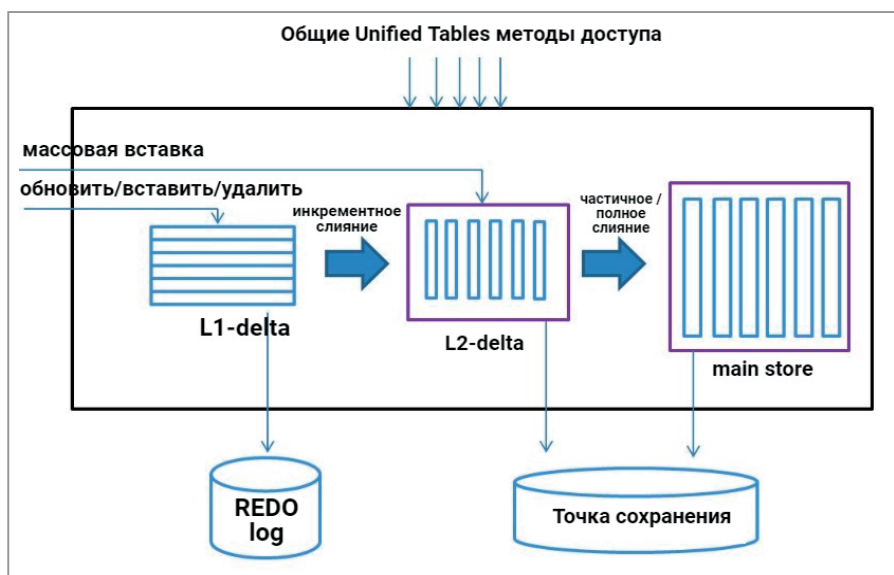


Рис. 1. Схема работы с данными (подход Unified Tables)

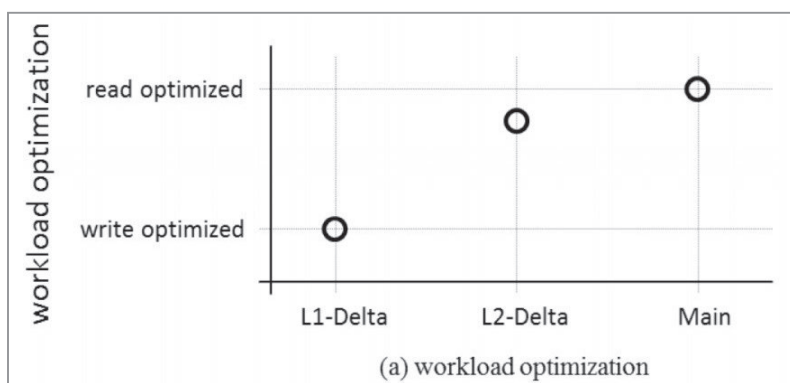


Рис. 2. Оптимизация рабочей нагрузки (чтение и запись данных) на каждом дельта-хранилище

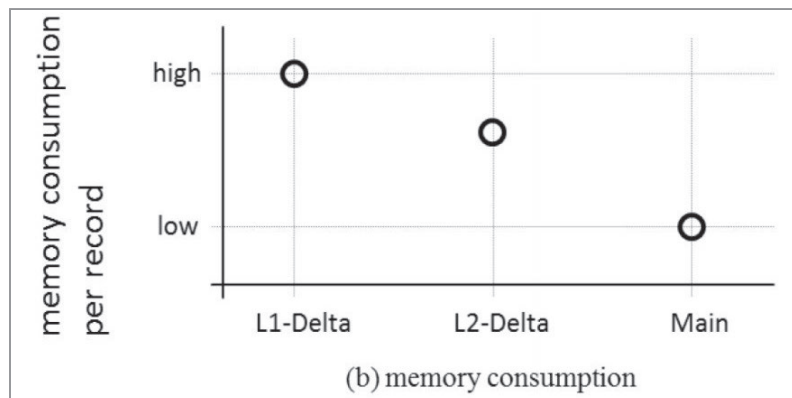


Рис. 3. Потребление памяти на одну запись на каждом дельта-хранилище

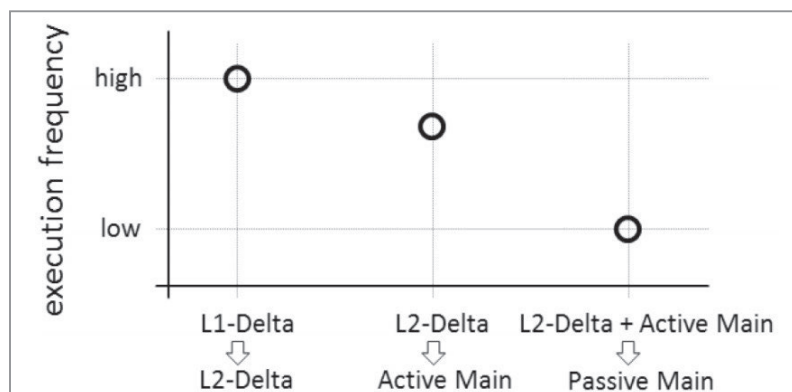


Рис. 4. Частота выполнения операций на каждом дельта-хранилище

Для таблиц с поколонным хранением данных SAP HANA использует сжатие со словарем. Словарь сортируется по отдельным значениям столбца. Каждый столбец использует массив целочисленных значений, которые представляют позиции фактических значений в словаре. Таблицы с постолбцовым хранением данных рекомендуется использовать тогда, когда в таблице содержится огромное количество данных, которые часто просматриваются или агрегируются, и когда таблица содержит много столбцов, а типичные запросы имеют доступ только к нескольким из них. На этапе L2-Delta проходит кодирование словаря, на этапе main store – его сортировка и сжатие. Пример сжатия со словарем приведен на рис. 5.

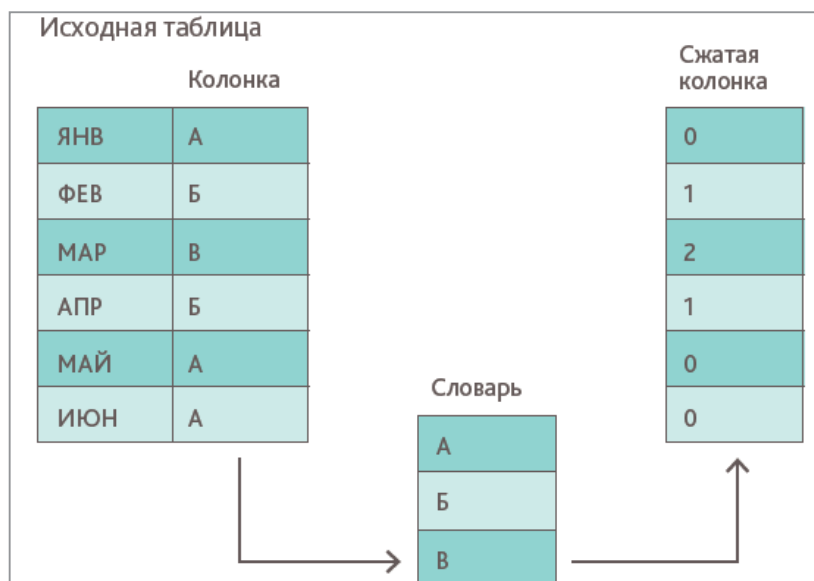


Рис. 5. Сжатие со словарем (пример)

Чтобы исключить потерю данных при отказе в работе сервера или ОЗУ, в SAP HANA используется концепция Persistent Storage (рис. 6).

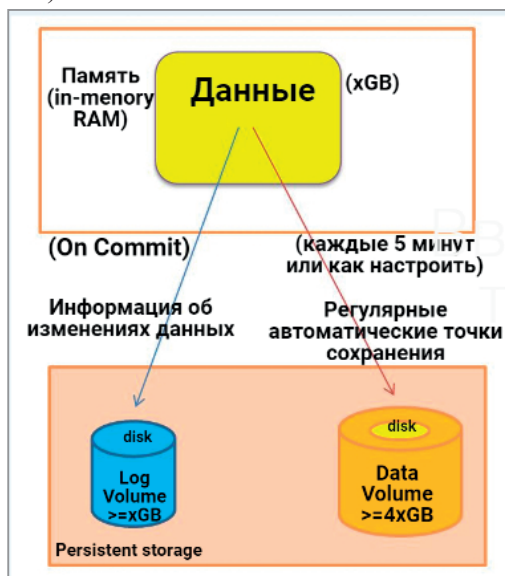


Рис. 6. Концепция Persistent Storage

Транзакционные данные обновляются в лог-журнале при каждом коммите. Данные сохраняются в объеме данных каждые 300 с или в соответствии с настройками. Они создают точки сохранения. В случае перезагрузки или запуска питания система может быть возвращена в последнюю согласованную точку сохранения, а затем воспроизвести данные, зафиксированные в протоколе.

Объем дискового пространства, необходимого для хранения данных, рассчитывается по формулам:

$$RAM_d = RAM_{st},$$

$$RAM_{total} = RAM_d + RAM_{st} = 2 \times RAM_{st},$$

$$DISK_{pers} = 4 \times RAM_{total},$$

$$DISK_{log} = RAM_{total},$$

$$DISK_{total} = DISK_{pers} + DISK_{log} = 4 \times RAM_{total} + RAM_{total} = 5 \times RAM_{total}.$$

В SAP HANA есть не только база данных, но и целый набор сервисов (рис. 7) для разработки приложений, средства интеграции и очистки данных, библиотеки для аналитической обработки данных, включая Machine Learning, а также возможности для хранения и обработки специальных данных. SAP HANA позволяет без дополнительных инструментов загружать данные из различных источников, разрабатывать различные формы для ввода, редактирования и анализа данных. Также доступны инструменты для сложной интеллектуальной обработки данных: преобразование трансформация, поиск закономерностей, исследования. И, конечно, платформа открыта для визуального анализа данных через различные инструменты.



Рис. 7. Сервисы платформы SAP HANA

Требования по развертыванию SAP HANA в режимах Appliance и TDI на примере сертифицированных серверов BullSequana S и Bullion S приведены в таблице.

Требования по развертыванию SAP HANA в режимах Appliance и TDI

Характеристика	BullSequana S	Bullion S
Процессор	Intel Xeon Scalable	Intel Xeon E7 v4 Family
Максимальное количество процессоров	32	16
ОЗУ	64 ГБ – 48 ТБ	128 ГБ – 24 ТБ

Appliance – преднастроенное решение, включающее сервер, СХД и пакет ПО для внедрения «под ключ», с централизованной службой поддержки и оговоренным уровнем производительности. Здесь SAP HANA поставляется в виде предварительно настроенного аппаратного и программного обеспечения, полностью интегрированного и сертифицированного.

TDI (Tailored Data center Integration) позволяет выбирать конкретных производителей и компоненты инфраструктуры в зависимости от пожеланий заказчика – с учетом выполняемых задач и рабочей нагрузки.

Развертывание SAP HANA в совместном режиме Appliance + TDI на примере ПАО «ГМК «Норильский никель» представлено на рис. 8.

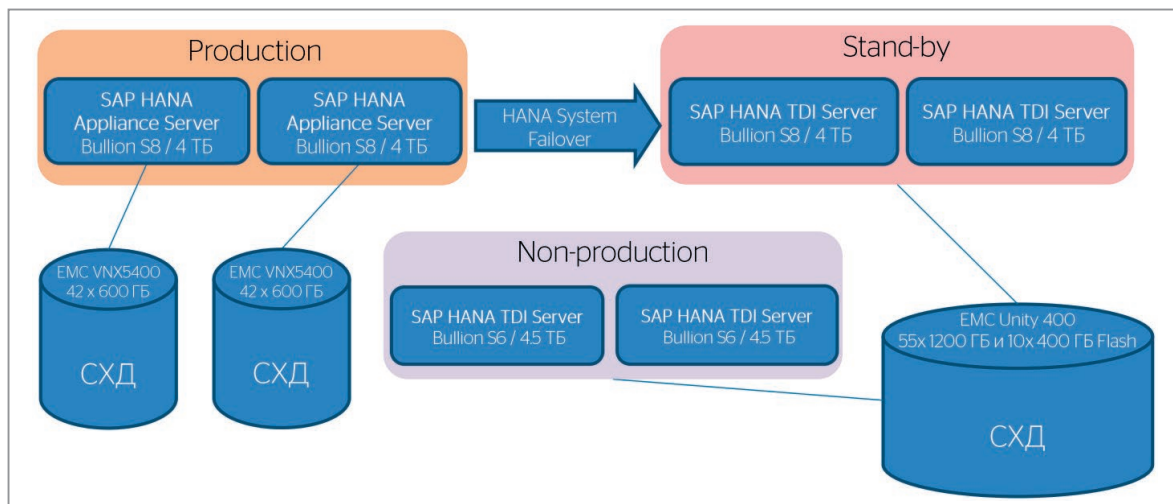


Рис. 8. Развертывание SAP HANA на примере ПАО «ГМК «Норильский никель»

SAP HANA Studio – это интегрированная среда разработки (IDE) на основе Eclipse для разработки и администрирования базы данных SAP HANA в форме инструмента с графическим интерфейсом. SAP HANA Studio работает на компьютере клиента / разработчика и подключается к серверу SAP HANA. SAP HANA Studio позволяет пользователю управлять локальной или удаленной базой данных SAP HANA, управлять авторизацией пользователей, создать новые или изменить существующие модели данных, разрабатывать приложения.

SAP S/4 HANA изначально построена на базе SAP HANA с упрощенной моделью данных, без индексов, агрегирующих выражений и без повторов. Она использует ее функциональность и может быть развернута на месте, в облаке или используя оба способа. S/4 HANA разработана специально для SAP Fiori, предлагая интегрированный пользовательский интерфейс.

SAP Fiori – это единый (доступный на любых устройствах, включая планшеты и смартфоны), ролевой (предусматривающий реализацию так называемых факт-листов для разных должностей и разных участков) интерфейс.

SAP Cloud Platform – это открытая облачная платформа SAP, которая предоставляет возможности для хранения и обработки данных в SAP HANA и ASE, а также в open source-системах PostgreSQL, MongoDB и Hadoop. SAP Cloud Platform предназначена для разработки мобильных и HTML-приложений, которые могут быть легко интегрированы с любыми облачными или локальными системами благодаря встроенному в SAP Cloud Platform сервису интеграции. В настоящее время SAP Cloud Platform предоставляет порядка 70 различных сервисов, включая интернет вещей, алгоритмы прогнозирования, машинный

перевод и другие методы машинного обучения, имеет собственную веб-среду разработки WebIDE и инструмент для быстрого прототипирования Build, поддерживает разработку приложений на Java, XSJS, C++, Python, Ruby, Node.js, Go, PHP, .Net.

SAP HANA PlanVisualizer (PlanViz) – инструмент для вычисления производительности выполнения запросов к базе данных SAP HANA. На рис. 9 показаны этапы и общая производительность выборки 500 тыс. записей из таблицы колоночного типа ZBKPF (копия структуры таблицы BKPF в системе SAP ERP), хранящейся в базе данных на SAP Cloud Platform.

SQL-код запроса на выборку: `select * from "SYSTEM"."ZBKPF"`.

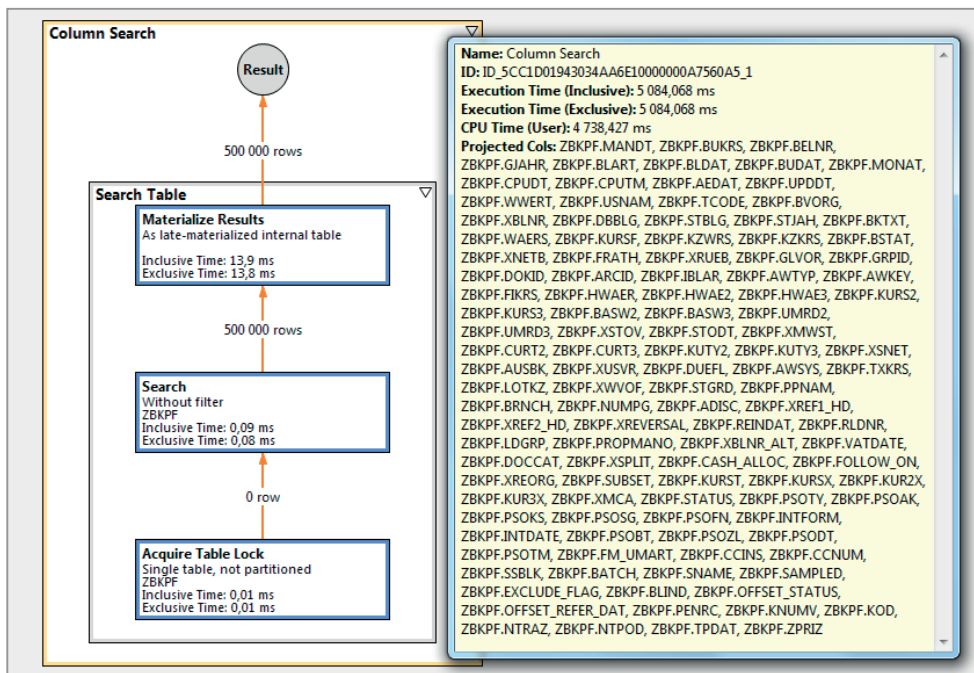


Рис. 9. Практическое измерение скорости выборки данных

Общее время на выполнение запроса составляет примерно 5 с. Аналогичный тест был проведен в системе SAP ERP. Общее время на выполнение запроса на основе данных транзакции ST05 составляет примерно 15 с. Исходя из результатов, выборка данных в SAP HANA выполняется в 3 раза быстрее.

Таким образом, платформа SAP HANA действительно превосходит предыдущие поколения систем SAP ERP, имея большой набор инструментов (в том числе и облачных) для ее администрирования и анализа данных, обеспечивая высокую скорость доступа к информации и ее безопасное хранение, предоставляя различные варианты развертывания и способы хранения данных, удовлетворяя требованиям бизнес-пользователей. Реализованные на базе данной платформы ERP-системы будут популярны среди потребителей и смогут успешно конкурировать на мировом рынке долгое время.