

# ПРИМЕНЕНИЕ ПЛАТФОРМЫ SAP HANA ДЛЯ АНАЛИЗА БОЛЬШИХ ОБЪЁМОВ ДАННЫХ

Минковский Д. В.

Кафедра информационных технологий автоматизированных систем, Белорусский государственный университет информатики и радиоэлектроники

Минск, Республика Беларусь

E-mail: dzmitry.minkouski@gmail.com

*Промышленные предприятия постоянно сталкиваются с необходимостью обработки данных, которые в большом объеме поступают с заводов или других объектов. Компании собирают триллионы байт информации о своих клиентах, поставщиках и производственных операциях. Согласно исследованию консалтинговой компании IDC, к 2023 г. объем ценных данных увеличился в два раза, и 60% всех собранных данных можно будет использовать в процессе принятия решений. В статье рассматривается применение платформы SAP HANA для обработки и анализа больших объёмов данных*

## ВВЕДЕНИЕ

Современные системы автоматизации позволяют хранить данные с высоким уровнем детализации, что позволяет аналитикам создавать отчёты практически любой степени сложности. Однако, часто такие отчёты могут оказаться бесполезны из-за проблем с производительностью. Формирование отчета может занимать несколько часов, а иногда даже несколько десятков часов. Поэтому такие отчеты становятся неудобны, и часто от них отказываются предприятия, т.к. они не могут предоставить информацию своевременно. В таких случаях правильным решением будет использование специально разработанной для обработки больших объёмов данных платформы SAP HANA.

## I. ОБЗОР SAP HANA

SAP HANA – это многоцелевое решение для хранения и обработки информации. Одна из особенностей SAP HANA – это встроенный механизм вычислений, который позволяет переносить выполнение операций по планированию с уровня приложений на уровень базы данных SAP HANA. С помощью современной архитектуры аппаратной платформы вычисления проходят эффективнее – вся «лавина» обрабатываемых данных разбивается на строго определённое количество потоков, число которых равно общему количеству ядер платформы. Такой подход позволяет максимально эффективно использовать вычислительную мощность каждого ядра каждого процессора. SAP HANA также предоставляет технологии для хранения и обработки данных in-memory. SAP HANA как база данных позволяет хранить данные в построчном и в поколонном виде. Технология хранения и обработки данных in-memory обеспечивает быструю обработку транзакций, а вместе технологией анализа данных Calculation View гарантирует высокое быстродействие при выполнении аналитических запросов [1]. Основным ядром в SAP HANA является компонент СУБД. В осно-

ве СУБД SAP HANA используется реляционная модель данных, но также существует возможность обращения к данным с помощью «графового» языка запросов WIPE. Гибкость в выборе языка запросов обусловлена архитектурными возможностями SAP HANA и заключается в использовании единого представления данных в In-Memory хранилище. Таким образом, у пользователя есть возможность обращения к данным с помощью различных семантических конструкций, используя при этом единую копию данных в памяти СУБД. Классический подход, принятый в ряде других OpenSource СУБД, отличается от вышеуказанного, потому что подразумевает использование как минимум двух хранилищ данных и разделение способа хранения графовых структур и реляционных таблиц.

## II. КОНЦЕПЦИЯ УПРАВЛЕНИЯ ДАННЫМИ

На рисунке 1 отражена общая схема управления данными в SAP HANA и суть концепции управления с помощью различных языков – в частности, SQL и WIPE. Используя движок Data Processing, можно сформировать на уровне Data Manipulation новый семантический уровень для работы с данными, но при этом будет применена единая копия исходных данных, что существенно повышает возможности платформы SAP HANA для решения задач, где требуется представление информации в виде графовых структур.

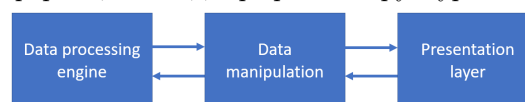


Рис. 1 – Концепция управления данными

Технология In-Memory в СУБД SAP HANA позволяет хранить и обрабатывать данные в памяти, используя уникальные алгоритмы, разработанные в компании SAP и на базе платформы Intel x86. Недавно SAP также анонсировала поддержку платформы IBM Power для SAP Hana. Уникальность и высокая скорость обработки запросов к данным заключается в возможности их

хранить и выполнять. Они находятся в сжатом виде в памяти RAM. Благодаря разработанному алгоритму обработки данных в SAP HANA удалось реализовать подход Unified Tables, который обеспечивает высокую скорость чтения и записи данных в таблицу поколочного хранения [2]. Поэтому одним из главных преимуществ SAP HANA является возможность выполнять аналитические запросы сразу на транзакционных данных, которые добавляются в реальном времени. При этом система автоматически берёт на себя обеспечение прозрачного доступа к данным. Таким образом, новые данные в таблице сразу доступны для анализа без предварительной обработки.

### III. АНАЛИТИЧЕСКИЕ СЕРВИСЫ SAP HANA

В SAP HANA есть не только база данных, но и целый набор сервисов для разработки приложений, средства интеграции и очистки данных, библиотеки для аналитической обработки данных, включая Machine Learning, а также возможности для хранения и обработки специальных типов 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. На рисунке 2 можно увидеть все сервисы, доступные на платформе SAP HANA [1].



Рис. 2 – Аналитические сервисы SAP HANA

Давайте рассмотрим наиболее популярные. SAP HANA включает в себя движок для хранения и обработки геоданных – данных, которые описывают положение, форму и ориентацию объектов в пространстве. SAP HANA поддерживает пространственные типы данных и методы их обработки. Существует специальный метод для обработки такой структуры – граф. SAP

HANA в этом случае предоставляет возможности для обработки гиперсвязанных данных и их отношений. Движок для обработки данных имеет встроенные алгоритмы поиска окрестностей, кратчайших путей, сильно связанных компонент, сопоставления образцов и многое другое.

В SAP HANA также есть сотня предварительно упакованных алгоритмов машинного обучения и прогнозирования с такими возможностями, как объединение, кластеризация, классификация, регрессия, распределение вероятности, временные ряды и многое другое [3]. Кроме этого, вы можете использовать библиотеку TensorFlow и язык R.

SAP HANA имеет встроенные возможности для обработки и анализа текстовых файлов, включая различные функции по интеллектуальному анализу текстов – например, нечеткая логика, поиск синонимов, семантический разбор и т.д.

SAP HANA Streaming Analytics может фиксировать, фильтровать, анализировать и воздействовать на миллионы событий в секунду в режиме реального времени, сохраняя данные или результаты в базу данных SAP HANA и направляя менее критические данные в более дешевые решения для хранения – такие, как Hadoop. SAP HANA Streaming Analytics также интегрирована с системой сообщений Apache Kafka[3].

### ЗАКЛЮЧЕНИЕ

Крупномасштабная инфраструктура хранения, обработки и анализа данных SAP HANA делает платформу наиболее подходящим местом для загрузки, обработки и последующего хранения терабайтов менее структурированных внешних данных – потоков кликов, журналов, IoT, текста, изображений и видео. Анализ обработанных данных позволяет пользователям использовать проверенную платформу для вычисления в памяти и использовать ее высокую производительность и параллелизм для аналитики с малыми задержками.

### СПИСОК ЛИТЕРАТУРЫ

1. Practical Guide to SAP HANA and Big Data Analytics / Dominique Alfermann, Stefan Hartmann // N. Engl. SAP Press. – 2018. – Vol. 44, № 8. – P. 238–243.
2. SAP HANA analytics cloud [Electronic resource] / – Mode of access: <https://help.sap.com/viewer>. - Date of access: 15.10.2020.
3. Advanced Analytics | SAP HANA 2.0 [Electronic resource] / – Mode of access: <https://blogs.sap.com/2019/11/02/advanced-analytics-sap-hana-2.0/>. – Date of access: 14.10.2020.