

# Data Lake — универсальное хранилище для аналитики больших данных

В апреле 2014 г. компания Pivotal начала поставки Pivotal Big Data Suite — программного пакета с годовой подпиской, включающего программное обеспечение, обслуживание и поддержку. В состав пакета вошли продукты Pivotal Greenplum Database, Pivotal GemFire, Pivotal SQLFire, Pivotal GemFire XD, Pivotal HAWQ и Pivotal HD, предоставляющие заказчикам гибкие возможности обработки больших данных. Этот анонс позволил перейти от концепции Data Lake в плоскость продвижения конкретных решений для аналитики больших данных.



Сергей Золотарёв — директор по продажам, Pivotal Россия и СНГ.



Александр Ермаков — старший системный инженер, Pivotal Россия и СНГ.

## Введение — предыстория

Аналитика больших данных (АБД) в настоящее время является ключевым направлением, связанным с развитием бизнеса. Одним из принципиальных его отличий от традиционных приложений для ВА/ВИ с использованием хранилищ данных — data warehouse (DW) — состоит в том, что для анализа используются не только структурированные данные, получаемые в результате функционирования транзакционных OLTP-систем, но и весь объем неструктурированных или плохо структурированных данных, который, например, может быть получен из социальных сетей и других источников. Другая важная особенность АБД в том, что за счет развития, прежде всего, тех-

нологий аппаратных платформ, результаты ВА/ВИ могут быть получены в режиме, приближенном к реальному времени.

Учитывая активно развивающийся сегмент рынка информационных хранилищ нового поколения, компания EMC объявила аналитику больших данных своим приоритетом в 2012 г. Впервые EMC анонсировала свою программу развития аналитики больших данных в мае 2011 г. В частности, EMC представила комплексную стратегию дистрибуции, интеграции и поддержки ПО с открытым исходным кодом Apache Hadoop.

В результате интеграции Hadoop с базой данных Greenplum появляется возможность в рамках одного хранилища обрабатывать как структурированные, так и неструктурированные данные.

Все решения, объявленные в мае 2011 г., появились на рынке в конце сентября 2011 г. В частности, стала доступна EMC Greenplum Modular Data Computing Appliance (DCA) — первая в индустрии платформа для аналитики корпоративного класса, которая на основе модульной архитектуры позволяет комбинировать MPP-обработку (Massively Parallel Processing) реляционных баз данных и неструктурированных данных на базе Apache Hadoop-кластеров.

При этом все BI-приложения и ELT/ETL-инструменты могут развертываться в одном и том же DCA с помощью новых Greenplum Data Integration Accelerator модулей. Как результат — унифицированная Big Data платформа для структурированных и неструктурированных данных и приложений в составе одной инфраструктуры с общим мониторингом, управлением и поддержкой от одного вендора — EMC.

В декабре 2011 г. EMC анонсировала дальнейшее развитие платформы DCA — EMC Greenplum™ Unified Analytics Platform (UAP). Новая платформа соединила в себе три решения для поддержки аналитики больших данных: 1) EMC Greenplum database — для структурированных данных; 2) EMC Greenplum HD — для аналитики и обработки неструктурированных данных; 3) EMC Greenplum Chorus™, представляющее собой производительную машину и социальную сеть для коллективов аналитиков<sup>1)</sup>. Основ-

ная цель создания UAP — максимальная интеграция результатов аналитики больших данных с бизнес-процессами.

В конце января 2012 г. EMC анонсировала интеграцию файловой горизонтально-масштабируемой СХД EMC Isilon с распределенной файловой системой Hadoop (Hadoop Distributed File System — HDFS), что в значительной степени расширило возможности Hadoop-кластеров по хранению данных с точки зрения надежности, доступности/скорости доступа, а также простоты управления и масштабирования.

По сути, интеграция EMC Isilon с протоколом HDFS положила начало переходу использования для хранения данных в Hadoop-кластерах — от кустарных ненадежных СХД к промышленным.

Другой важный этап развития DW-платформ EMC для АБД был связан с началом поставок Pivotal<sup>1)</sup> Big Data Suite (апрель 2014 г.) — программного пакета с годовой подпиской, включающего программное

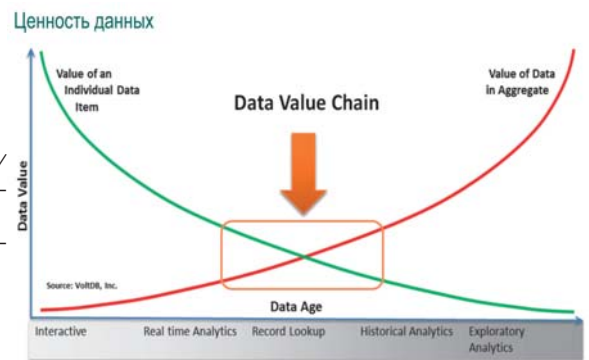


Рис. 1. Ценность данных возрастает по мере увеличения их агрегирования.

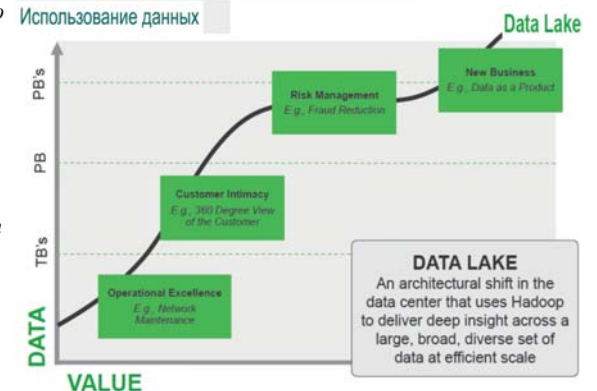


Рис. 2. С ростом ценности данных возрастают их объемы, которые необходимо анализировать.

1) 1 апреля 2013 г. EMC и VMware объявили о создании новой компании — Pivotal, которая объединила в себе стратегические технологии, персонал и программы EMC и VMware, включая Greenplum, Cloud Foundry, Spring, Cetas, Pivotal Labs, GemFire и другие продукты, входящие в пакет VMware vFabric Suite.

2) В настоящее время Chorus — это проект open source и Pivotal не поддерживает его.

обеспечение, обслуживание и поддержку. В состав пакета вошли продукты Pivotal Greenplum Database, Pivotal GemFire, Pivotal SQLFire, Pivotal GemFire XD, Pivotal HAWQ и Pivotal HD, предоставляющие заказчикам гибкие возможности обработки больших данных.

Пакет Pivotal Big Data Suite связал воедино все ведущие технологии компании и дал возможность эффективно использовать любую из них по мере необходимости — от анализа данных в памяти в реальном времени с помощью Pivotal GemFire, Pivotal SQLFire и Pivotal GemFire XD до быстрых запросов посредством Pivotal HAWQ в Pivotal HD. Также он позволил в полной мере реализовать одну из актуальных задач — слияние традиционных сред хранения данных и инфраструктуры следующего поколения, позволяющей создавать так называемое «Business Data Lake» (озеро бизнес-данных).

### Зачем нужна аналитика больших данных?

Достаточно широко распространено мнение, что бизнесу вполне достаточно, той информации, которую собирают компании при работе транзакционных приложений. При этом обогащение данные вовсе не обязательно. Однако это не так (рис. 1, 2).

Значение агрегированных данных возрастает по мере возрастания роли исторической и прогнозной аналитики (см. рис. 1) и риск-менеджмента. А именно эта аналитика, во-первых, дает возможность “заглянуть за горизонт” и по-новому переосмыслить существующий бизнес, а во-вторых, правильно и своевременно оценивать существующие и возможные риски.

И, если отсутствие прогнозной аналитики в компании, может лишь замедлять бизнес, не позволяя ему “идти в ногу со временем”, то во втором случае это может привести к его краху бизнеса/компании. К этому следует добавить то, что многие противоправные действия (например, мошеннические операции в банковской сфере, утечки данных из-за инсайда, выявление злонамеренных приложений и др.) стали намного изощреннее и их выявление становится уже невозможным без аналитики больших данных, в ряде случаев онлайн-аналитики больших потоков событий данных (измеряемых сотнями гигабайт и терабайт в секунду).

Однако не все так просто. Решение подобных задач требует специализированных технологических платформ, прежде всего, вследствие того, что оперировать приходится очень большими объемами (измеряемых петабайтами, см. рис. 2) и потоками данных. На это накладывается еще условие: результаты аналитики должны быть получены в режиме, приближенном к реальному времени или за ограниченное время. В противном случае, их ценность может снижаться или становится ничтожной. Т.е. основная цель аналитики больших данных — работать на предотвращение возможных нежелательных событий, а не бороться с их последствиями.

### Архитектура Data Lake

Data Lake дает возможность интеграции всех источников данных и всех потребителей аналитики на базе единой платформы.

Архитектура Data Lake не имеет in-box решения и может меняться в зависимости от задач, существующего оборудования, экономических ограничений и т.п. В полной комплектации это интегрированная система трех типов кластеров (например, на базе стандартных серверов): 1) кластер для развертывания GemFire; 2) кластер для развертывания GreenPlum; 3) hadoop-кластер. Все кластеры интегрированы между собой и позиционируются по возрастанию времени реакции, соответственно.

Архитектура Data Lake в горизонтальном представлении — в интеграции с источниками данных, платформой обработки и потребителями запросов, а также вертикальном представлении — по типам данных и по соответствующим им технологиям обработки дана на рис. 3, 4.

Data Lake имеет три уровня (см. рис. 4) обработки данных — GemFire (Pivotal GemFire, Pivotal SQLFire и Pivotal GemFire XD; самый быстрый уровень — аналитика в реальном времени), MPP GreenPlum (near online аналитика) и Pivotal HAWQ (обработка данных в Pivotal HD — быстрая аналитика за приемлемое время).

Data Lake предлагается как в качестве ПО, которое может устанавливаться на оборудование на базе стандартных серверов (включая виртуальные серверы) и на программно-аппаратный комплекс DCA.

#### Аппаратные компоненты Data Lake

В качестве аппаратной платформы в составе Data Lake могут использоваться различные компоненты в зависимости от того, какие объемы данных необходимо анализировать и за какое время (или при каких допустимых задержках).

Для наиболее высокопроизводительной аналитики (в реальном времени) предлагается использовать аналитику in-memory на базе распределенных кластеров GemFire с общим полем памяти.

GemFire — это in-memory data grid или распределенный кэш, где каждый узел — это ячейка общей памяти. При этом все они синхронизированы между собой, т.е.,

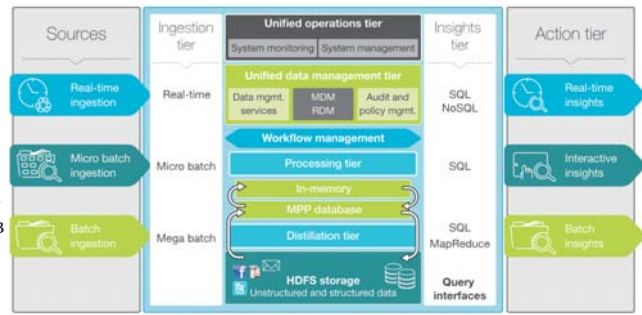


Рис. 3. Архитектура платформы Data Lake в интеграции с источниками данных, платформой обработки и потребителями запросов.

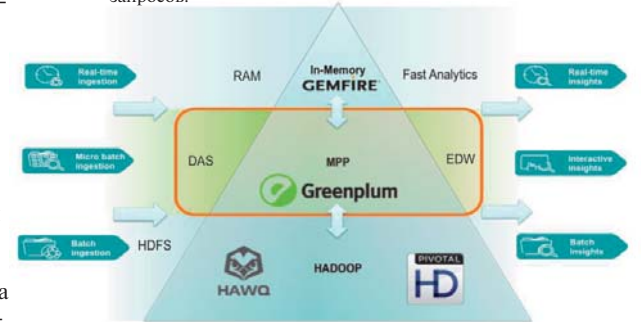


Рис. 4. Архитектура платформы Data Lake по типам данных и по соответствующим им технологиям обработки.

если что-то записывается в один из узлов, эта информация сразу становится доступна другим узлам. Это, во-первых, позволяет создавать отказоустойчивые решения, а, во-вторых, создавать платформы для высоконагруженных систем, которые аккумулируют OLTP-нагрузку и которым требуется низкое время отклика.

GemFire может хранить данные в ОП, на локальных дисках этих узлов, в hadoop-кластере, используя соответствующие протоколы, а также перемещать данные на другой уровень в зависимости от того, что необходимо выполнить (рис. 5).

Там, где задержки могут составлять минуты и десятки минут, предлагается использовать GreenPlum. Здесь ускорение достигается за счет погружения модулей обработки аналитических запросов на уровень хранилища при максимальном распараллеливании самой обработки.

СУБД Pivotal Greenplum DB (Pivotal GPDB, рис. 6) построена на основе архитектуры симметричных вычислений с массовым параллелизмом (Symmetric MPP) без разделения ресурсов, которая была разработана для бизнес-аналитики и аналитической обработки. Именно такая архитектура идеально подходит для типичных операций аналитических БД, например, сортировки, агрегирования, объединения огромных таблиц, за счет

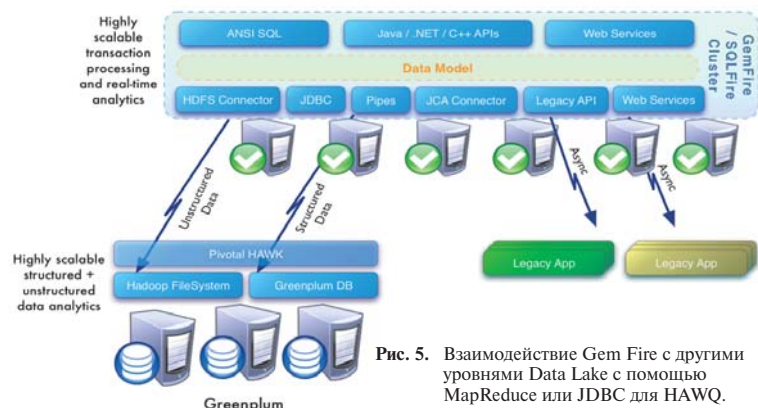


Рис. 5. Взаимодействие Gem Fire с другими уровнями Data Lake с помощью MapReduce или JDBC для HAWQ.

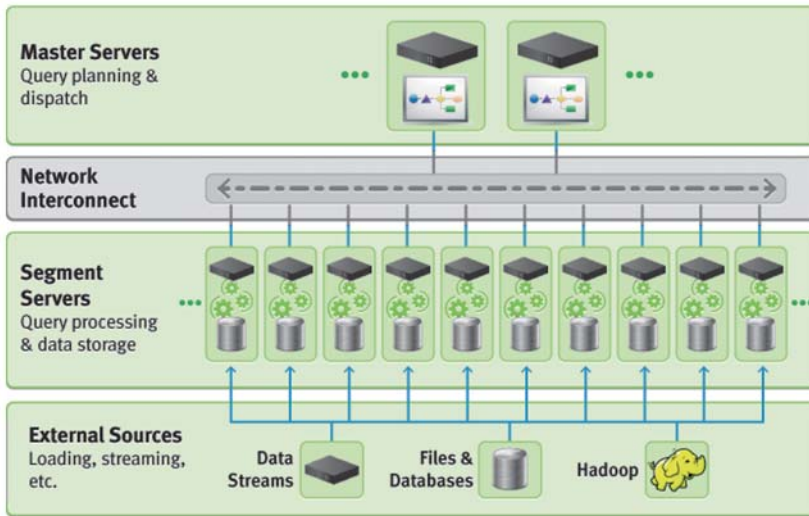


Рис. 6. МРР-архитектура.

возможности параллельной обработки данных на всех узлах кластера базы данных и отсутствия «узких мест», например, недостаточная пропускная способность сети передачи данных или кластерного интерконнекта, недостаточная производительность системы хранения и т.д. Массово параллельная архитектура без разделения ресурсов, подразумевает физическое распределение данных таблиц на небольшие подмножества на отдельных серверах сегментов, каждый из которых имеет выделенный, независимый широкополосный канал подключения к локальным дискам.

Серверы сегментов способны эффективно обрабатывать каждый запрос полностью параллельно, одновременно использовать все подключения к дискам и эффективно распределять потоки данных между сегментами согласно плану запросов. При этом каждый из узлов использует собственные ресурсы (память, дисковое пространство и т.д.) для выполнения задачи.

Поскольку базы данных без разделения ресурсов автоматически распределяют данные и обеспечивают параллельность рабочих нагрузок при запросах на всем доступном оборудовании, они намного более производительны и масштабируемы, чем СУБД общего назначения для задач бизнес-аналитики.

При работе с большими объемами данных чрезвычайно важно выполнить максимально возможное количество ресур-

соемких операций «максимально близко» к данным. В Pivotal GPDB рекомендуется организовать хранение данных на встроенных дисках серверов кластера (direct attached storage, DAS). В этом случае данные будут передаваться только по высокоскоростной внутренней шине узла, обрабатываться его процессорами в локальной оперативной памяти, без передачи по сетевой шине. Промежуточные результаты (по объемам меньше исходных данных на порядки) – передаются по внутренней шине (программная технология называется Pivotal gNet™).

Pivotal GPDB обеспечивает возможность организации хранения данных таблицы, как по записям, так и по столбцам. Это чрезвычайно важно для задач аналитики. Способ хранения (по записям или по столбцам) можно чередовать даже на уровне разделов (партиций) одной таблицы. Технология называется Polymorphic Data Storage™.

Наконец, там, где требуется анализировать петабайты данных (а допустимое время обработки – часы), предлагается использовать PivotalHD (рис. 7) – коммерческий дистрибутив Hadoop с поддержкой Enterprise-уровня, построенный на базе open source фреймворка Apache Hadoop. Помимо стандартных средств для управления кластером и обработки данных, в данный дистрибутив включены средства централизованного мониторинга, балансировки и масштабирования

кластера – Pivotal Command Center, средства эффективной виртуализации Hadoop Virtualization Extensions (HVE), а также транслятор SQL-on-Hadoop – HAWQ с полноценной поддержкой ANSI SQL синтаксиса. При этом на этот уровень могут быть погружены и обрабатываться как структурированные (БД GreenPlum), так и неструктурированные данные.

Следует отметить, что при наличии уже подходящей у заказчика платформы, могут предлагаться чисто программные решения.

#### Особенности Data Lake

Принципиальной особенностью новой платформы для аналитики больших данных, в сравнении с более ранними решениями EMC для этих задач, явилось появление новых интерфейсов, которые значительно упростили и ускорили работу с неструктурированными данными на уровне стандартных SQL-запросов. Это сделало доступной платформу Data Lake широкому спектру внешних приложений без каких-либо или минимальных изменений.

Это было достигнуто, во-первых, за счет разработки оптимизатора запросов, а во-вторых, за счет разработки Pivotal собственного транслятора SQL-on-Hadoop – HAWQ.

HAWQ стал доступен в феврале 2013 г. и явился весьма удачной разработкой в серии аналогичных – Impala (10.2012 – Cloudera), Stinger (02.2013 – Hortonworks), Vertica (HP – 2014), BigSQL (2014 – IBM) – значительно упрощающей работу с hadoop-системами, заменяя собой HiveQL<sup>3</sup>.

Особенности транслятора SQL-on-Hadoop – HAWQ:

- СУБД Greenplum, портированная на HDFS;
- написан на C;
- оптимизатор запросов с возможностью выбора оптимального плана выполнения на базе структуры данных и статистики;
- полноценная поддержка ANSI SQL-92, аналитических расширений SQL-2003;
- поддержка сложных запросов с коррелированными подзапросами, сложными оконными функциями и различными join'ами;
- поддержка ANSI типов данных;

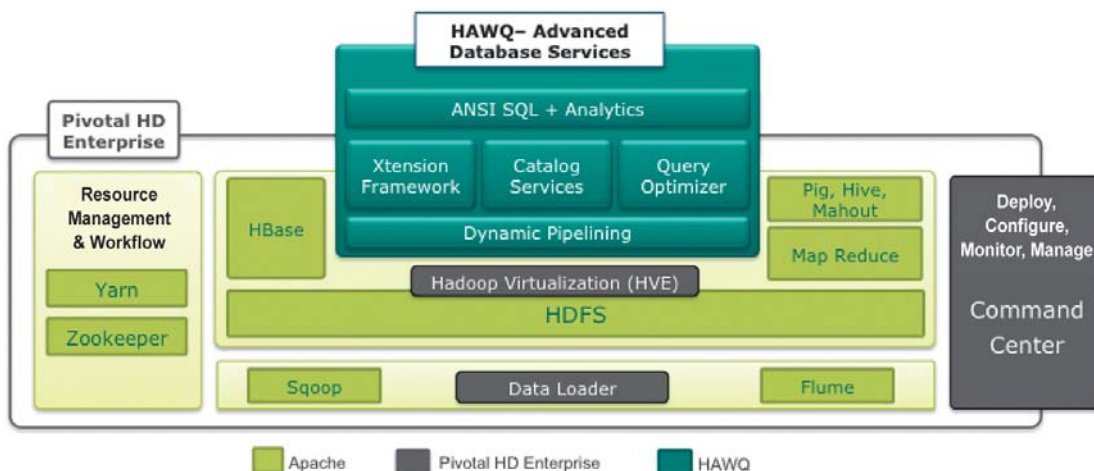


Рис. 7. Архитектура PivotalHD.

# Будущее аналитики за “сырыми” данными

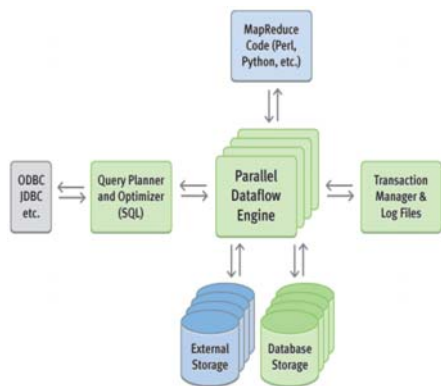


Рис. 8. Компоненты, участвующие в обработке ODBC/JDBC-подключения.

— не использует MapReduce при загрузке и чтении данных.

HAWQ работает с HDFS через фреймворк PXF, что позволяет:

- читать различные типы данных с HDFS (текстовые файлы, в т.ч. сжатые; Avro; JSON; Parquet-файлы; в дальнейшем планируется поддержка новых форматов хранения данных и новых типов сжатия);
- читать данные из внешних источников (GemFireXD, HBase, Hive, Cassandra, Redis, Accumulo);
- иметь расширяемый API.

## Интеграция Data Lake в внешними приложениями

Все уровни архитектуры используют не проприетарные интерфейсы подключения, в частности, для интеграции с GemFire/Greenplum/HAWQ. Возможно использование ODBC/JDBC-подключения для осуществления запросов к данным. На рис. 8 схематично отображено, какие компоненты участвуют в обработке данного подключения.

После получения запроса в случае Greenplum и HAWQ вступает в работу оптимизатор для распараллеливания и транслирования частей запроса к рабочим узлам. Такой подход позволяет использовать любые внешние приложения и платформы, поддерживающие синтаксис SQL без каких-либо модификаций.

Также для Greenplum и HAWQ возможно интеграция на уровне использования MapReduce Jobs, то есть приложения могут транслировать в Greenplum или HAWQ задачи с использованием данного синтаксиса. В частности, например, возможно использовать Sqoop или Flume для записи данных из HDFS непосредственно в Greenplum.

Для GemFire возможно также использование JDBC Connector для подключения к IMDG-данным, а также использование различных программных API (для подключения из внешних приложений, написанных, например, на Java или .NET). Также возможно подключение к Hive с помощью MapReduce или JDBC для HAWQ (см. рис. 5).

## Интеграция данных между уровнями Data Lake

Интеграция данных между уровнями Data Lake организована внутренними протоко-

3) Hive — разработка Facebook, используется для анализа данных в их хранилище. Синтаксис HiveQL не соответствует ANSI SQL-92, имеет множество ограничений на подзапросы.



Андрей Кондратов — главный архитектор, компания GlowByte Consulting.

Суть концепции «Data Lake» довольно проста — данных много, они разнообразны. В моменте мы знаем (анализируем, используем, собираемся использовать в ближайшее время) только часть того, что храним, соответственно, не можем предложить оптимальную модель данных для их хранения с учетом всех потенциально возможных вариантов анализа, в том числе и неизвестных на текущий момент. Поэтому необходимо хранить данные в «сыром» виде и обеспечить максимально гибкую возможность их интерпретации, работы с ними. Это является ключевым отличием таких систем от классических систем класса Аналитические Хранилища Данных.

Несмотря на то, что термин «Data Lake» прозвучал впервые в 2011 г., практическая реализация концепции в отрасли аналитики больших данных в реальных проектах началась совсем недавно — с середины 2013 г. Данный факт в первую очередь связан с тем, что в течение некоторого времени не было enterprise-ready решений в виде законченного стека продуктов с бесшовной интеграцией. Изначально MPP СУБД и решения, основанные на Hadoop, развивались параллельно и независимо, без должного уровня ин-

лами, например, Greenplum общается с HDFS посредством протокола GPHDFS, а GemFire — с HDFS с помощью HDFS Connector/MR (MapReduce).

## Интеграция данных из источников

Не менее важное требование для аналитической системы — высокоскоростная, линейно-масштабируемая загрузка данных. Доступное на рынке сетевое оборудование обеспечивает ширину полосы пропускания от 1 до более 2,5 Гбайт/с для каждого узла кластера Pivotal GPDB. Для каждой строки вставленных данных система вычисляет хэш значений этих столбцов, чтобы равномерно распределить строки между всеми сегментами

теграции и с существенным опережением MPP СУБД в части поддержки требований enterprise-уровня. Одним из таких полнофункциональных решений стала линейка продуктов Pivotal Big Data Suite компании Pivotal/EMC. Концепция, таким образом, получила поддержку в виде платформ, программных или программно-аппаратных, которые позволили решать стоявшие задачи обработки сверх больших объемов данных за ограниченное время, снизив при этом стоимость владения данными.

Отдельно хотел бы остановиться на некоторых ключевых особенностях организации Data Lake в масштабах компании:

- особенностью реализации Data Lake в компаниях с существующей инфраструктурой является интеграция существующих Хранилищ Данных и создаваемого Data Lake. Он, в свою очередь, начинает выступать источником данных для Хранилища, а также и приемником данных Хранилища (хранение истории, дублирование данных для различных задач в рамках Data Lake). Другим интересным направлением, с точки зрения развития отрасли, является создание Хранилищ Данных и аналитических систем поверх Data Lake, но на базе его инфраструктуры и программной платформы;
- довольно важным нюансом является реализация задачи управления данными (Data Governance) в проектах создания и развития Data Lake. Более того, качественное решение данной задачи имеет высокую критичность, так как без ее решения невозможна качественная реализация концепций Data as a Service, Analytics as a Service, Self Service BI, Data Science, по причине непрозрачности и непонимания данных, хранящихся в системе.



системы. Такой подход наряду с секционированием позволяет отказаться от индексов в большинстве случаев. Это положительно сказывается на скорости загрузки данных (тем не менее, Pivotal GPDB поддерживает и растровые индексы, и бинарные деревья для тех случаев, где их применение необходимо).

Технология Scatter/Gather Streaming™ (SG Streaming) позволяет линейно масштабировать скорость загрузки данных в Pivotal GPDB. За счет того, что загрузка может выполняться всеми узлами кластера одновременно, наращивая количество узлов в кластере БД и источнике, можно добиться линейного роста скорости и соответствующей

шего сокращения времени загрузки. На фазе «рассредоточения» (scatter) каждый узел загружает данные в кластер. На фазе сбора данные перераспределяются между узлами в зависимости от значения хэш-функции ключа распределения.

## Применения Data Lake

Пример 1. Банки/финансовые организации: проведение таргетированных маркетинговых кампаний

В настоящее время выбор целевых групп для подобных акций проводится по очень узкому кругу параметров. Например, выборка для рассылки определяется какой-то условной возрастной группой и полом. И после этого несколько десятков тысяч людей получают одинаковые SMS, письма и пр. В результате получатель подобных рассылки воспринимает их как навязчивый спам и должный коммерческий эффект не достигается. Однако в Европе уже есть примеры другого подхода, когда, например, банку удалось увеличить прибыль в несколько раз, поставив у себя систему аналитики, которая позволяла выбирать более узкую фокус-группу по гораздо большему числу параметров (которые у них уже были собраны).

Пример 2. Телеком-компания: своевременное устранение причин плохого сервиса

Достаточно часто, когда клиент переходил к другому оператору (например, из-за частого прерывания связи), не удавалось установить причину этого вследствие того, что эта информация хранилась только на определенном оборудовании. После того, как была установлена система аналитики, способная анализировать большое количе-

ство плохо структурированной информации, появилась возможность выявлять первопричину плохой связи за короткое время. Это, во-первых, позволило быстро устранять неисправности или модернизировать оборудование. А во-вторых, предотвращать уход клиента к другому оператору. Например, когда клиенту не удавалось позвонить на какой-то номер после пяти попыток (вследствие того, что его телефон не поддерживает какую-то сеть или в его телефоне устаревшая прошивка), ему поступало, например, сообщение "у Вас не работал такой-то сервис такого-то числа из-за того, что Вам необходимо обновить такое-то ПО". В результате, клиент видел контроль его связи со стороны провайдера услуг и оставался лояльным к нему.

Пример 3. Телеком-компания: геомаркетинг

Это можно сделать, когда о человеке есть некая информация (его профиль, например, что он любит, что часто посещает и т.п.). И когда человек находится в определенном месте, в определенный момент времени ему поступает определенная информация. Например, "уважаемый .... рядом с Вами находится наш партнер, который предлагает такие-то услуги". Этим уже пользуется один из операторов связи в Европе на базе решения с использованием GemFire.

Пример 4. Продажа билетов (внедрения – "Китайские ж/д", сайт Expedia)

Одно из применений GemFire – продажа билетов (ж/д, авиа). У многих компаний есть центры обработки данных, которые могут быть размещены в разных частях страны. И когда происходит покупка билета, то, как правило, продажа соверша-

ется через ЦОД, который расположен в данном регионе, но т.к. покупку билета можно приобрести в любой точке страны, то информация о конкретной продаже должна сразу попасть в другие ЦОДы. Сделать это стандартными средствами без in-memory технологий достаточно сложно. В результате часто происходили ситуации, когда билет на одно место покупали сразу несколько человек вследствие того, что информация о покупке поступала в другой ЦОД с недопустимой задержкой. GemFire позволяет эту проблему решить, т.к. это распределенный кэш. Вследствие этого, его узлы могут находиться в разных местах, но они синхронизируются между собой достаточно быстро (при этом для связи ЦОД необходимо обеспечить канал порядка 10Gb) и это полностью снимает проблему консистентности данных.

Пример 5. Поточковая обработка событий на базе решения GemFire и Data Lake

Например, требуется онлайн-обработка/фильтрация каких-то событий, например, когда человек что-то делает (разговаривает по телефону, посещает сайты и др.). GemFire начинает собирать эту информацию из логов и записывать ее на свои узлы. Дополнительно он может дополнить ее данными с уровня GreenPlum или hadoop для того, чтобы сразу проанализировать ее. Далее, используя заданные правила, GemFire, например, через web-сервис может послать эти результаты для дальнейшего анализа или принятия решения.

*Сергей Золотарёв,  
Александр Ермаков,  
Pivotal Россия и СНГ.*

## Smart Logger II для

## платформы Microsoft Lync

**Октябрь 2014 г.** – Система речевой аналитики и автоматической оценки качества работы операторов контакт-центра (КЦ) Smart Logger II (<http://www.speechpro.ru/>) была интегрирована с платформой унифицированных коммуникаций Microsoft Lync. Теперь у компаний, использующих MS Lync, появилась возможность автоматизировать решение таких задач, как: выявление непрофильной нагрузки, анализ упоминаний маркетинговых акций, выявление причин повторных вызовов, обеспечение информационной безопасности. Smart Logger II регистрирует голосовой трафик внешних и внутренних соединений серверов MS Lync и сохраняет их в базе данных. Автоматически распределять зафиксированные вызовы по темам обращений клиентов позволяют встроенные алгоритмы полнотекстового распознавания слитной речи и семантического анализа содержимого переговоров.

Также аудиофайлы проходят обработку инструментами оценки качества взаимодействия «клиент-оператор», которые автоматически предоставляют результаты оценки и существенно облегчают работу супервизоров. Текстовые расшифровки записей переговоров позволяют пользователю быстро и эффективно искать ключевые слова в больших массивах фоно-

грамм при решении задач, связанных с оценкой качества, или задач информационной безопасности.

Несколько ранее Smart Logger II был расширен алгоритмом определения эмоционального фона разговоров, который разделяет все записи переговоров на положительные, нейтральные и негативные, фиксируя все случаи, когда клиент доволен и благодарит оператора контактного центра, или наоборот, выражает свое недовольство. За рубежом данный метод анализа, известный как метод определения тональности высказываний (Sentiment Analysis), успел зарекомендовать себя как эффективное средство автоматизации КЦ.

Модуль определения удовлетворенности клиентов анализирует не только параметры голоса и отдельные ключевые слова-индикаторы эмоций, но и проводит полный семантический анализ диалога с клиентом, что значительно повышает точность поиска претензионных обращений клиентов.

## Комплексное решение для мобильной безопасности

**Октябрь 2014 г.** – Компания Check Point® Software Technologies Ltd. представила Check Point Capsule – комплексное мобильное решение, которое обеспечивает многоуровневую систему защиты данных и мобильных устройств, и включает:

– **безопасный доступ к работе:** защита корпоративных данных на мобильном

устройстве без необходимости управления всем девайсом. Check Point Capsule создает безопасную бизнес-среду и отделяет корпоративные данные и приложения от персональных. Это позволяет сотрудникам безопасно работать с бизнес-приложениями через простой пользовательский интерфейс, который обеспечивает доступ к корпоративной почте, файлам, каталогам, контактам и календарю "в одно касание" и не влияет на личные данные пользователя;

- **защищенные корпоративные документы:** предотвращение внутренних и внешних утечек данных за счет обеспечения безопасности самого документа на всем протяжении его пути. Check Point Capsule защищает корпоративные документы, где бы они ни находились; авторизованные пользователи могут получить доступ к защищенному документу беспрепятственно и просто на любом устройстве;
- **полная защита от угроз:** расширение сферы действия корпоративной политики безопасности позволяет защищать устройства, даже когда они находятся за пределами корпоративной сети. Check Point Capsule сканирует весь трафик с мобильных устройств в облако и предотвращает доступ к вредоносным файлам и веб-сайтам, заражение ботами и другие киберугрозы. При этом поддерживаются различные платформы мобильных устройств и операционные системы, включая iOS, Android, Windows и MacOS.