SYBASE
An SAP Company

АНАЛИТИКА БОЛЬШИХ ДАННЫХ ОТ SYBASE

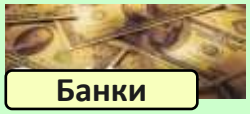
АНДРЕЙ ХРОМОВ
ВЕДУЩИЙ ТЕХНИЧЕСКИЙ КОНСУЛЬТАНТ SYBASE CIS

ФОРУМ «BIG DATA 2012», МОСКВА, 22 МАРТА 2012 Г.

ЧТО ТАКОЕ «БОЛЬШИЕ ДАННЫЕ»

ГДЕ ЖИВУТ БОЛЬШИЕ ДАННЫЕ

Индустрии



Банки

telecom

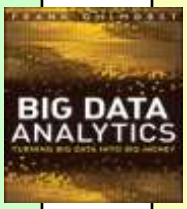
GLOBAL CAPITAL MARKETS

Трейдинг, инвестиции

Торговля

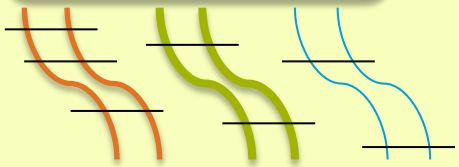
Гос. органы

Интернет-компании



Задачи

Маркетинг
Анализ эфф. каналов



Отслеживание посещений сайтов, анализ эффективности каналов: email, sms, соц.сети, пр.

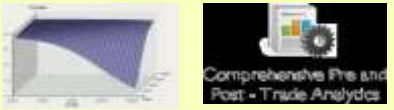
Продажи
Поиск корреляций

Оценка рисков на основе комплексного анализа всей детальной информации по клиенту

Анализ машинных данных

Анализ показателей RFID, логов вебсайтов, SMS, показаний датчиков

Финансы
Моделирование



Оценка ликвидности, инвестиционных. портфелей, кредитных рисков, стресс тесты

SYBASE | An SAP Company

SYBASE IQ: АНАЛИТИКА БОЛЬШИХ ДАННЫХ

Широкое использование во всем мире в сферах, где требуется работа с большими и очень большими объемами данных - с скоростью секунд

Управление и анализ всех статистических данных по всему населению Канады

Хранение и анализ всех налоговых данных по всем физ. и юр. лицам США

Сложная финансовая аналитика в более чем 200 финансовых институтах мира

Крупнейшие банки, страховые компании, телеком операторы во всем мире используют Sybase IQ как платформу для своих Хранилищ Данных

Хранение и анализ огромных массивов отраслевых данных для топ 30 крупнейших в мире информационных провайдеров, включая Transunion, Nielsen , Thomson Reuters, TNS Media

SYBASE | An SAP Company

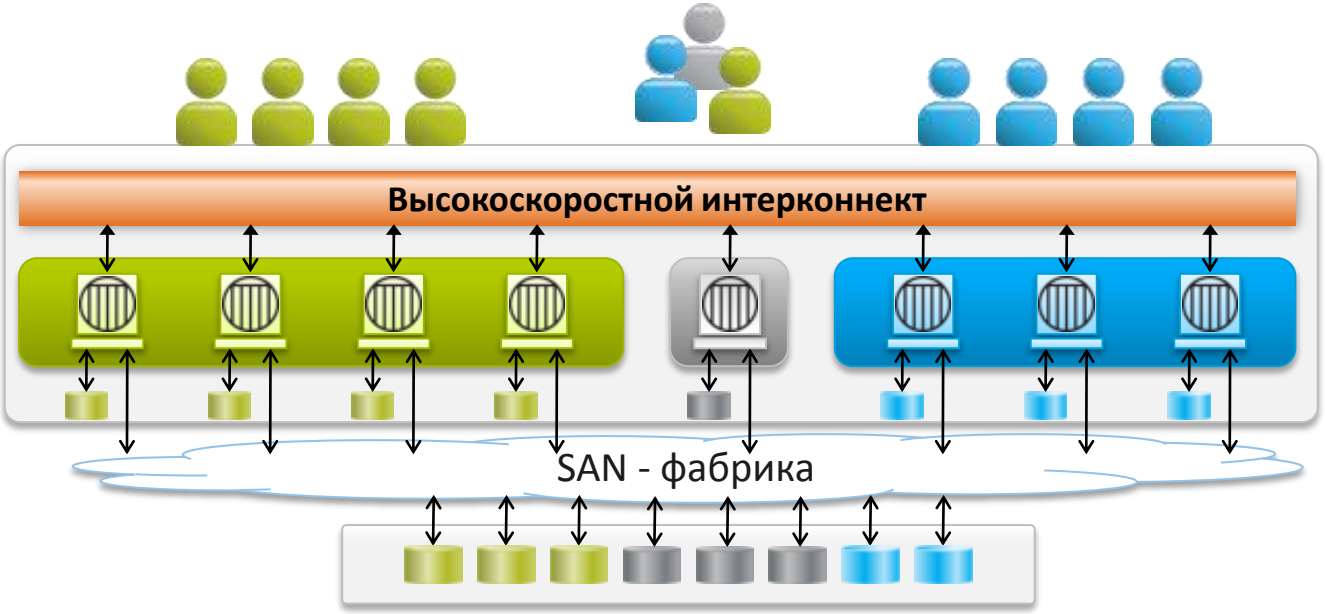
SYBASE IQ 15

Вехи развития VLDB-платформы

SYBASE | An SAP Company

ПЛАТФОРМА SYBASE IQ 15

Высокопроизводительная работа с данными для Аналитики Больших Данных

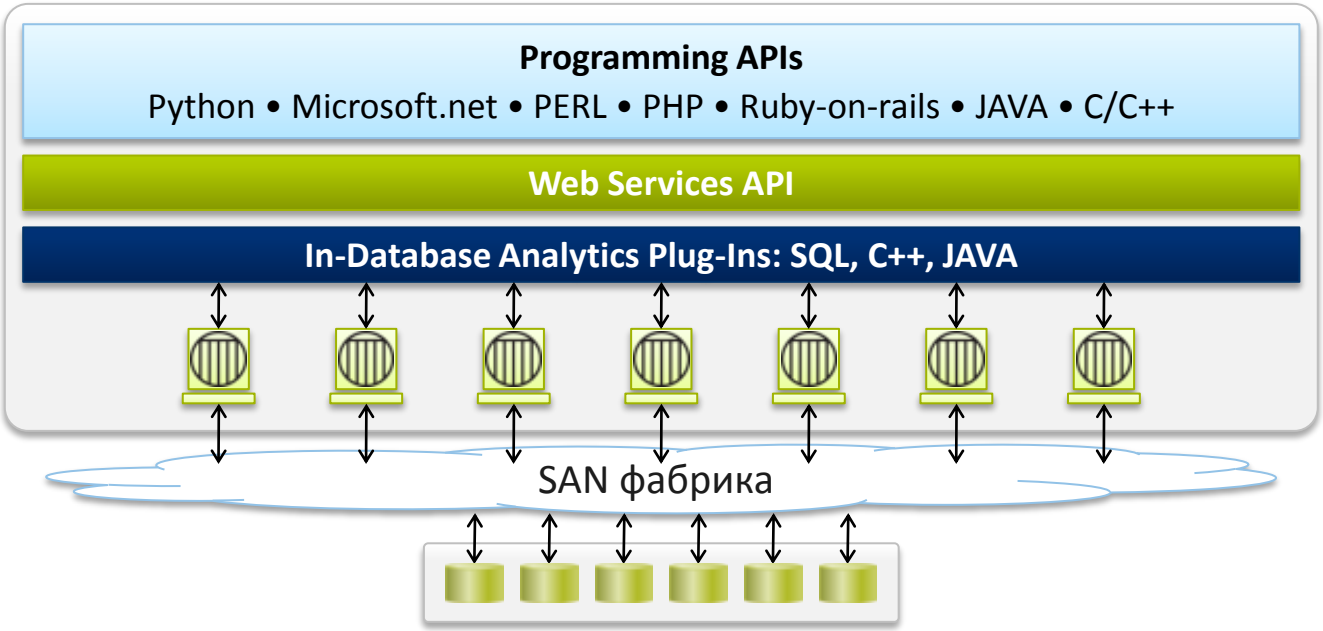


- **Поколоночная организация данных – скорость, компрессия, ad-hoc аналитика**
- **Кластерная MPP-архитектура PlexQ™**
  - Использование массивного параллелизма для обработки сложных запросов
  - Виртуальные витрины данных, основа для организации Облачных IQ
- **Логические и физические партиции - для управления жизненным циклом данных**
- **Удобные и мощные средства администрирование и мониторинг – низкий TCO**

SYBASE | An SAP Company

ПЛАТФОРМА SYBASE IQ 15

Прикладные сервисы для Аналитики Больших Данных

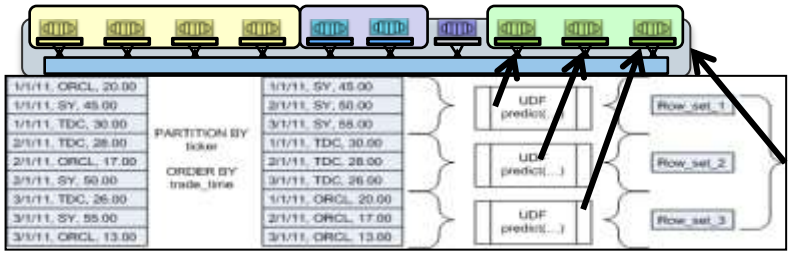


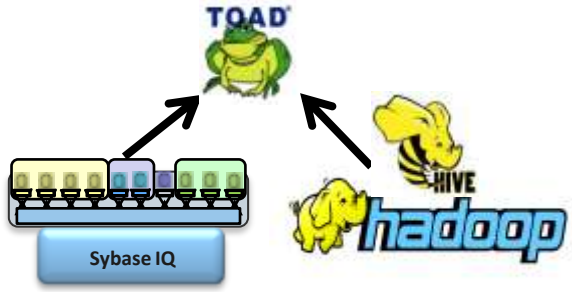
- **Среда PlexQ™**
  - Развитый ANSI SQL , соответствующий спецификации ANSI 2008
  - Встраиваемые в базу данных внешние модули аналитики: SQL, C++, JAVA для статистического анализа, для датамайнинга и т.п.
  - In-database web services with SOAP API
  - Query and data federation via SQL queries
- **API для множества языков программирования:**
  - C, JAVA, PHP, PERL, Python, Ruby-on-rails, ADO.NET

SYBASE | An SAP Company

SYBASE IQ 15.4
СПЕЦИАЛЬНО ДЛЯ БОЛЬШИХ ДАННЫХ

- **Поддержка MapReduce**
  - с помощью TPF API (функции с параметрами типа таблица)
- **Интеграция с Hadoop**
  - Федерация на клиенте
    - с помощью Toad от Quest
  - ETL из Hadoop в IQ
    - с помощью SQOOP от Cloudera
  - Федерация данных и запросов
    - Использование данных Hadoop в тексте запросов IQ
- **UDF с параметрами типа таблица**
  - Новый API для Java UDF как основа для внешних модулей аналитики





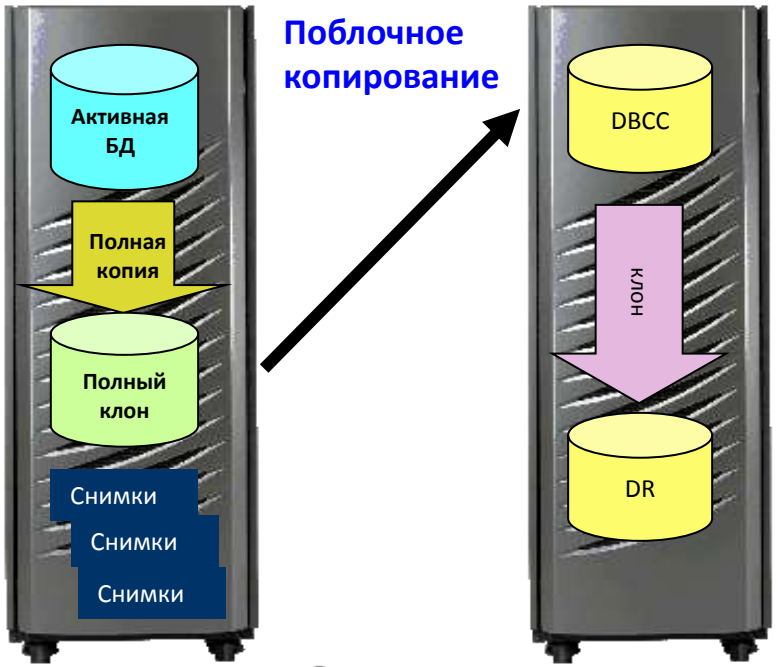
БОЛЬШИЕ ДАННЫЕ – БЭКАПЫ И ЗАЩИТА > NONSTOPIQ

Проблемы защиты Больших Данных

- Защита данных КХД
  - Традиционный бэкап – медленно, ненадежно, дорого
- Обеспечение доступности КХД в режиме 24x7
  - Обычный DR-сайт – сложно, ненадежно, дорого

Решение

- Технология NonStopIQ
  - Sybase IQ + поблочная репликация массива
- Несколько копий данных IQ в онлайне, снимки данных и клоны
- Моментальное переключение IQ с копии на копию
- Выполнение «бэкапа» многотерабайтной базы за минуты
- Копия данных для проверки базы на целостность, для тестеров, для разработчиков

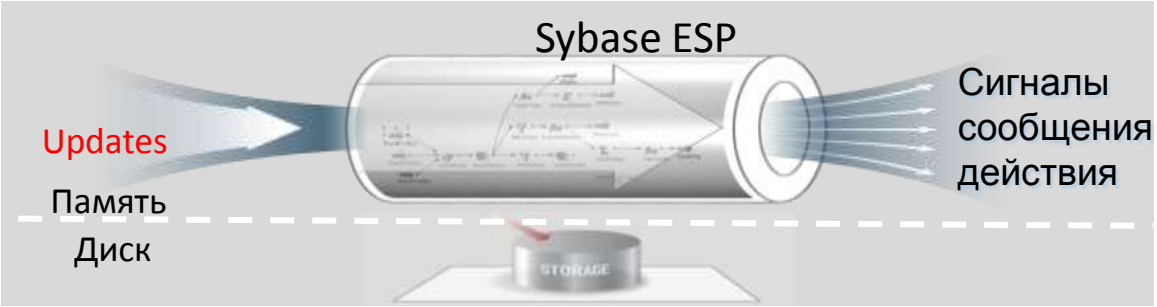


SYBASE | An SAP Company

ПРОБЛЕМА БОЛЬШИХ ДАННЫХ №2 - БЫСТРОТА РЕАКЦИИ

Непрерывная аналитика потоков данных

- Sybase Event Stream Processer
- Sybase RAP



Аналитика в реальном масштабе времени

- Sybase Replication Server Real-Time Load

SYBASE | An SAP Company

ПРИМЕРЫ КЛИЕНТОВ

TELEFONICA

Компания

Telefonica является одним из крупнейших в мире телеком-операторов, 300 миллионов клиентов, работает в 25 странах в Европе, Латинской Америке и Китае. 285,000 сотрудников. Оборот за 2011 год 46 миллиардов евро.

Ситуация

11 независимых и разрозненных систем, объем данных 70 ТБ (более 2 млрд. записей),15000 запросов в день, более 1000 активных пользователей.

Требовалось повысить производительность работы и снизить сложность обслуживания

TELEFONICA

Что дало решение на Sybase IQ:

- **Объемы –** единое хранилище объемом **15 TB**, объединившее все 11 систем (2 млрд. строк - **70 TB исходных данных**)
- **Быстрота –** вся информация обновляется каждые 5 минут, в день в систему загружается порядка 1 ТБ новых данных
- **Быстрота –** скорость выполнения запросов улучшилась в 200 раз (**15,000 запросов в день, 1,000 активных пользователей)**

“Новая система позволила уменьшить объем требуемого дискового пространства в 4-6 раз по сравнению в другими решениями на рынке, одновременно с заметным сокращением затрат на оборудование, поддержку, администрирование. Кроме этого новая система позволила нам получить время отклика более чем в 200 раз лучше , чем на наших предыдущих системах.”

- Pedro Romera, Systems Engineering Manager, Telefonica, Spain

SYBASE | An SAP Company

COMSCORE

Company

ComScore является глобальным лидером в области мониторинга и сбора статистической информации о работе интернет систем, является наиболее известным источником для маркетинговой аналитики по цифровым коммуникация. Компания осуществляет треккинг более 3 млн. сайтов по всему миру, помогает своим клиентам лучше понять кто, как и зачем использует их веб-ресурсы. Маркетинговые агентства, издательства, рекламные агентства, финансовые аналитики обращаются к comScore за аналитической информацией, помогающей им ориентироваться в цифровом бизнесе.

Ситуация

Требовалось платформа для аналитического хранилище данных comScore’s - системы Customer Knowledge Platform, которая должна успешно справляться с быстрорастущими объемами данных, быть масштабируемой и достаточно экономичной. Объем – более 20TB данных (плановый рост до 140 TB). При этом хранилище должно также обеспечивать высокую скорость обработки запросов.

COMSCORE

Что дало решение на Sybase IQ:

- **Объемы –** единое хранилище данных ~**140 TB (1.7 триллионов строк),** которое продолжает быстро расти
- **Быстрота –** Hadoop хранилище ComScore ежедневно принимает по 2-to-3 TB данных, которые сперва анализируются и потом из них отбирается 200-300 GB данных для загрузки в базу данных Sybase IQ
- **Быстрота -** хранилище Sybase IQ обслуживает **1000x (тысячи) одновременных пользователей** по всему миру, обрабатывая их регламентные и ad-hoc запросы. В режиме 24x7x365

“Мы знали, что нам будет необходимо отслеживать данные по миллионам веб пользователей по всему миру. Поэтому нам была нужна база данных, способная очень хорошо масштабироваться. Теперь мы уверены, что Sybase IQ способна справиться с нашими растущими объемами. Скорость Sybase IQ позволила нам теперь значительно быстрее анализировать данных и выдавать результаты нашим клиентам. Это помогает нам успешнее вести наш бизнес”

- Ric Elert, Vice President of Engineering, comScore

IRS, INTERNAL REVENUE SERVICE

“Sybase IQ это наше секретное оружие”

Department of the Treasury
Internal Revenue Service

Компания

IRS – бюро департамента казначейства США, один из крупнейших в мире и наиболее эффективных сборщиков налогов. В 2004 г. IRS суммарно собрал налогов на сумму более 2 триллионов USD, обработав более 224 млн. налоговых деклараций.

Ситуация

Требовалось построить новое современное хранилище данных (CDW, Compliance Data Warehouse) для хранения не менее 10 лет налоговой информации. Каждая отдельная налоговая декларация должна быть доступна для немедленного анализа.

Решение на Sybase IQ

- Единое хранилище данных **объемом около 360 TB**
- Около 1000 пользователей

ЗАКЛЮЧЕНИЕ
SYBASE IQ ГОТОВ К БОЛЬШИМ ДАННЫМ

1. ВЫСОКАЯ СКОРОСТЬ

- Вертикальное хранение
- Битмап-индесы
- MPP кластер PlexQ
- Аналитика, встраиваемая в базу

2. БОЛЬШИЕ ОБЪЕМЫ

- Компрессия данных
- Партиции логические и физические
- MPP кластер PlexQ

SYBASE IQ

3. НЕРЕЛЯЦИОННЫЕ ДАННЫЕ

- Поддержка Map Reduce
- Интеграция с Hadoop
- Текстовый поиск

4. ЭКОНОМИЧНОСТЬ

- Простое обслуживание
- Низкая стоимость владения
- Эффективное использование оборудования

SYBASE | An SAP Company


SYBASE®
An SAP Company