

СИСТЕМЫ BIG DATA, ОСНОВНЫЕ ПОНЯТИЯ

Новогрудская Р.Л., Заика А.В.

Институт телекоммуникационных систем НТУУ «КПИ», Украина

E-mail: zaika.nastya7@gmail.com

Big Data systems, basic concepts

The basic characteristics of Big Data systems for storing and manipulating large amounts of data are described. The properties of the software platform for analytics large amounts of data as a set of tools of systems Big Data are specified.

На сегодняшний день широко распространенными являются системы Big Data. Понятие Big Data представляет собой серию подходов, инструментов и методов обработки структурированных и неструктурированных данных больших размеров для получения результатов, которые легко воспринимаются человеком. Эти системы стали очень популярными в последние годы из-за ряда преимуществ которыми они наделены. В работе описаны базовые понятия, связанные с построением систем Big Data.

К основным преимуществам использования систем Big Data можно отнести:

- Получение качественно новых знаний за счет комплексного анализа всей информации в едином аналитическом хранилище.

- Расширение функциональности существующих информационных систем поддержки бизнеса.

- Увеличение эффективности использования аппаратных ресурсов серверов.

- Обеспечение минимальной стоимости использования всех видов информации за счет возможности использования ПО с открытым кодом и облачных технологий хранения информации.

Системы BigData (SBD) это системы для хранения и манипулирования информацией, а также генерирования новых знаний на основании подхода BigData. Такие системы должны учитывать все особенности и характеристики BigData в рамках обеспечения доступа к информации сохраняя специфику информационных потоков в зависимости от предметной области и направления исследований пользователя.

Основной характеристикой SBD является территориальная распределенность больших объемов информации в разных областях жизнедеятельности человека.

Большие объемы данных определяется 5-ю основными характеристиками:

- объем – X1,
- скорость – X2,
- разнообразие – X3,
- стоимость – X4,
- сложность X5.

Учет таких характеристик при проектировании SBD предполагает минимизацию характеристик X4 и X5 с учетом постоянного возрастания X1, X2 и X3.

Основными особенностями больших объемов информации являются:

- территориальная распределенность информации,
- агрегация информации разных предметных областей и сфер жизнедеятельности пользователей.

Исходя из этого SBD должна включать:

- программные модули, построенные на основе интеллектуальных методов унификации информационных структурных единиц различных предметных областей,
- программные сервисы, реализующие интеграцию информационных и вычислительных ресурсов.

SBD должны предоставлять пользователю возможности работы с информацией, хранящейся в базах данных SBD, а также информацией, находящейся в открытом доступе глобальной сети (Рис. 1.1).

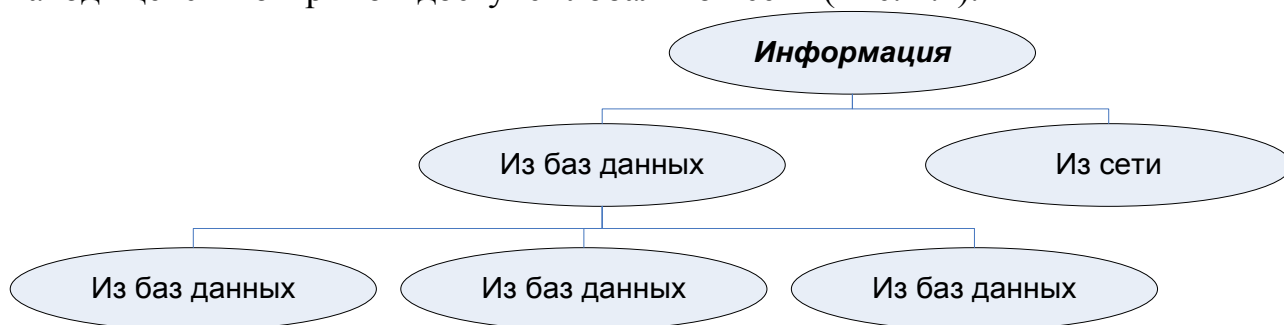


Рис. 1.1. Особенности хранения информации, представленной в SBD.

SBD предполагает манипулирование информацией в рамках выполнения следующих пользовательских функций: хранение информации, поиск, моделирование, аналитика и прогнозирование, а также выполнение расчетных заданий пользователей (Рис 1.2).

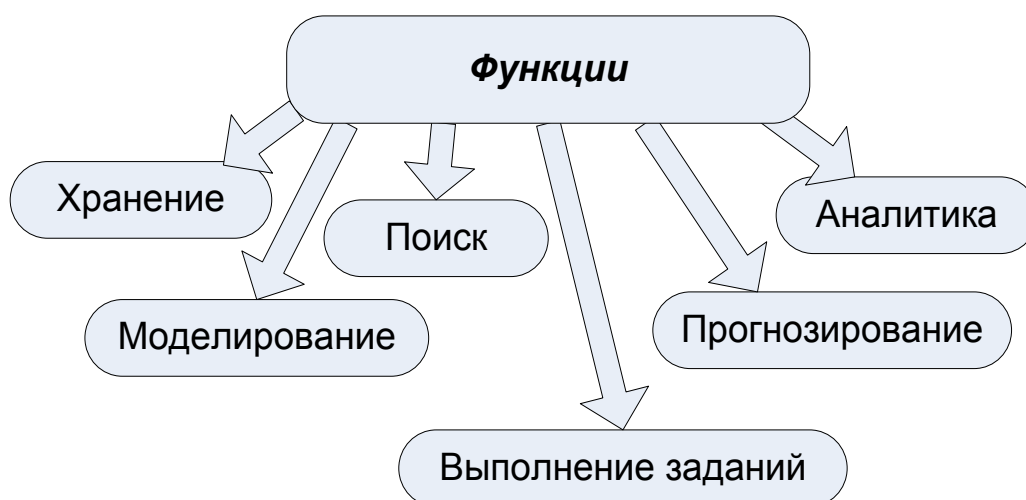


Рис. 1.2. Функциональные возможности SBD

Для осуществления таких функций в рамках комплекса инструментальных средств SBD необходимо разработать программную платформу аналитики больших объемов данных (BigData analytics platform) которая выполняет следующие функции:

- автоматический поиск и захват информации в Интернете в режиме реального времени, анализ информации (новости, BBS, блоги или другие средства массовой информации), например, поиск и анализ информации о наиболее горячих темах, которые обратили на себя внимание общественности, основываясь на ключевых словах поиска;
- прогнозирование тенденций развития новостных событий или прогнозирование возможности возникновения чрезвычайных ситуаций в социуме;
- интеллектуальный анализ имеющихся корпоративных информационных ресурсов, например, оперативный анализ, прогноз и технологическое предвидение тенденций развития бизнеса в будущем, прогноз и технологическое предвидение социально-экономического развития больших организационных систем.
- выполнение задач пользователей ориентированных определенную сферу деятельности учитывая специфические математические и расчетные методы предметной области.

Комплекс инструментальных средств SBD базируется на новых поколениях программных технологий и архитектур, а также средств программного обеспечения, предназначенных для извлечения экономической выгоды из очень больших объемов информации.

В работе описаны основные особенности, характеристики и функции систем Big Data, основываясь на которых становится возможным разработать специализированную программную платформу аналитики больших объемов данных выступающую как комплекс инструментальных средств SBD.

Литература

1. A. Halevy, P. Norvig, and F. Pereira, "The Unreasonable Effectiveness of Data," IEEE Intelligent Systems, vol. 24, no. 2, 2009, pp. 8–12.
2. V. Gudivada, D. Rao, and V. Raghavan, "Renaissance in Data Management Systems: SQL, NoSQL, and NewSQL," Computer (in press).
3. R. Baeza-Yates. "Big Data or Right Data?" Proc. 7th Alberto Mendelzon Int'l Workshop on Foundations of Data Management (AMW 13), 2013, vol. 1087, paper 14; <http://ceur-ws.org/Vol-1087/paper14.pdf>
4. C. Bizer et al., "The Meaningful Use of Big Data: Four Perspectives—Four Challenges," SIGMOD Record, vol. 40, no. 4, 2012, pp. 56–60.
5. Дописувачі Вікіпедії. Big Data [Інтернет]. Вікіпедія, ; 2015 жов 7, [дата звернення 2016 бер 15]. Доступно з: https://uk.wikipedia.org/w/index.php?title=Big_Data&oldid=16856338.