

ТЕХНОЛОГИЯ БОЛЬШИХ ДАННЫХ КАК СТРАТЕГИЧЕСКОЕ НАПРАВЛЕНИЕ



Л.Ю. Шилин

Декан факультета
информационных тех-
нологий и управления
БГУИР, доктор
технических наук,
профессор



А.А. Навроцкий

Заведующий кафедрой
информационных
технологий автома-
тизированных систем
БГУИР, кандидат физико-
математи-
ческих наук, доцент



О.В. Герман

Доцент кафедры
информационных
технологий авто-
матизированных
систем БГУИР,
кандидат техни-
ческих наук, доцент



Л.С. Стригалеv

старший препода-
ватель кафедры
информационных
технологий автома-
тизированных сис-
тем БГУИР,

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

E-mail: kafitas@bsuir.by

Abstract: We discuss the methodological issues of technology Big Data.

Растущий вклад информационной индустрии в мировой ВВП, обусловленный прямой зависимостью ВВП от информационных технологий, требует постоянного совершенствования их методов и средств. Сложность, высокая динамичность развития и стоимость современных информационных технологий порождают многочисленные проблемы, в числе которых и проблемы технологии больших данных (Big Data), слабоструктурированная проблематика которых имеет тесную связь с другими проблемами информационных технологий; так что Big Data носит и глобальный (мировой), и локальный характер (как ценность для бизнеса). Именно ценность для бизнеса и составляет «движительную силу» Big Data.

Технология Big Data в бизнесе позволяет снизить расходы, повысить эффективность ведения бизнеса и сократить риски бизнеса; при этом риск-менеджмент оказывается связанным с политикой информационной безопасности предприятия. Эффект применения Big Data, обеспечивающий конкурентные преимущества, зависит от предметной области и проблемной среды предприятия. Наиболее востребована технология Big Data в финансовой и телекоммуникационной сферах, в которых применение систем NoSQL (Not Only SQL) является технологической необходимостью, так как данные сферы имеют дело с неструктурированной информацией. Системы NoSQL обеспечивают горизонтальное масштабирование, параллельную обработку и возможность

быстрого получения результатов. По существу биржевые и банковские систем послужили катализатором развития Big Data.

Технологию Big Data принято характеризовать «четырьмя V»: объем (Volume), разнообразие (Variety), изменчивость (Variability) и скорость (Velocity); или, применительно к бизнесу, несколько иначе: объем, разнообразие, скорость и ценность (Value). Такого описания понятия технологии Big Data, которая в компьютерной индустрии занимает второе место после облачных технологий, явно недостаточно. Необходима системная проработка данного понятия.

Представляется, что наиболее конструктивным в этом деле является подход, связанный с вовлечением в рассмотрение четверки: система, структура, цель, технология [1, 2]. В такой четверке вещественный, энергетический и информационный метаболизм системы интерпретируется как ее «технология», как способ ее существования; при этом «движущей силой» является технология. Она же и приводит к существенным структурным изменениям системы в процессе технологического прорыва (необходимость таких структурных изменений обусловлена целевой установкой обеспечения живучести и конкурентоспособности системы). При этом система, как владелец технологии, должна рационально управлять технологическим пространством свободы.

В глобальном плане возникновение технологии Big Data обусловлено экстенсивным ростом мировой экономики в условиях усложнения технологических процессов, вызвавших энергетический дефицит. Истощение запасённой Землей солнечной энергии в виде угля, нефти, газа и др. требует коренных изменений в энергопотреблении. В результате все больше энергии уходит на информационную составляющую технологических процессов, необходимую как для управления, так и оптимизации энергетических затрат. При этом доля таких затрат будет постоянно возрастать.

При огромных объемах вычислений решающим фактором является величина потребляемой энергии [3], так что узким местом становится все же не масса и быстродействие процессора, а потребляемая им энергия. Так, например, масса нашей Вселенной позволяет записать при минимальных энергетических затратах (постоянная Планка на один бит) в некоторый «бесплатно предоставленный» регистр не более чем полное множество 67 буквенных комбинаций русского языка (около $1,4 \times 10^{101} \text{ бит}$). Предположим, что мы хотим задействовать мировой компьютерный парк для поиска в данном полном множестве всех шедевров стихотворных миниатюр; так вот по энергетическим соображениям сделать это мы не сможем. Очевидно, что никакое распараллеливание здесь не поможет. Заметим, что скорость света не позволит это сделать и по временным соображениям. Ханс Бреммерман показал, что процессор массой равной массе Земли за 10^{10} лет может обработать не более. Число 10^{93} бит называется пределом Бремерманна [4], а задача, объем вычислений которой превышает это число – трансвычислительной. И таких задач много. Замеченный дуализм можно использовать, положив энергетические

затраты в основу классификации сложности задач Big Data.

Технология Big Data не является чем то новым. Особенность приведенной выше запредельной гипотетической задачи состоит в том, что она офлайн во времени, а при ее распараллеливании имеет место онлайн в пространстве (по горизонтали). Данная задача позволяет уяснить основные достоинства Big Data. Нетрудно заметить, что значительная часть 67 буквенных комбинаций уже содержится в мировом компьютерном пространстве. При этом с точки зрения обработки она избыточна, но уже не так как в книжной индустрии, где имеются многочисленные миллионные издания. В компьютерном плане можно пойти дальше, положив в «облака» по одному экземпляру существующих файлов, с доступом к ним за небольшую плату. Далее, если речь идет о статистической обработке данных, то мы имеем возможность получить оценки по множеству данных, в то время как для получения таких же оценок по временной реализации необходима эргодичность процесса, что не всегда имеет место. Наконец, в «информационной руде» мировом компьютерном пространства содержатся неизвестные закономерности, которые может выявить специализированный робот-анализатор.

Огромный и всевозрастающий объем информации, который становится и энергетически накладным, требует оптимизации. Необходимы интеллектуальные средства для устранения «информационного мусора», роботы-анализаторы, средства для устранения информационной избыточности; а также методы и средства, обеспечивающие корпоративной системе надлежащую информационную безопасность [5].

Литература

- [1]. Strigalev L.S, German O.V. Methodological aspects of the IT-specialists training // Информационные технологии и системы 2011: Материалы Международной конференции, Минск, БГУИР, 2011 - С.199,200.
- [2]. Стригалева Л.С. Слабоструктурированные аспекты технологии дистанционного обучения. //Дистанционное обучение - образовательная среда XXI века: Материалы VI Междунар.научн.-метод.конференции,22-23 нояб.2007 г. - Минск:БГУИР,2007 - С.230-232.
- [3]. Стригалева Л.С. Экономико-энергетический аспект информационных технологий. // Экономическое развитие общества: инновации, информатизация, системный подход: Материалы Междунар.научно-экономической конференции, 22-23 апреля.2008 г. - Минск: «ПАРАДОКС»,2008 - С.257-260.
- [4]. Клир, Дж. Системология. Автоматизация решения системных задач: Пер. с англ. /Дж. Клир .– М. Радио и связь; 1990. – 544 с.
- [5]. Навроцкий А. А., Герман О.В., Стригалева Л.С. Методы оценки качества средств защиты информации // Технические средства защиты информации: Материалы XII Белорусско-российской научно-технической конференции, 28-29 мая.2014 г. - Минск:БГУИР,2014.