

## РОЛЬ БОЛЬШИХ ДАННЫХ В ЖИЗНЕННОМ ЦИКЛЕ ИЗДЕЛИЯ. СОВРЕМЕННЫЕ ПОДХОДЫ К УПРАВЛЕНИЮ БОЛЬШИМИ ДАННЫМИ

Е.Ю. Киселева

eanufrieva94@gmail.com

SPIN-код: 7751-2065

А.В. Федотова

afedotova.bmstu@gmail.com

SPIN-код: 4099-0349

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация

---

### Аннотация

Представлены цели, для которых определены методы управления большими данными на конкретных этапах жизненного цикла изделия (ЖЦИ). Проанализированы три основных периода ЖЦИ и в соответствии с ними представлены примеры различных видов входных и выходных данных. Также представлены современные подходы к обработке больших данных, включающие в себя обход использования реляционной системы базы данных (СУБД). В работе рассмотрены следующие подходы к управлению большими данными: MapReduce, NoSQL, Hadoop, Disco. Рассмотрены подходы, используемые СУБД, такие как СУБД с вертикальным хранением и In-memory СУБД. Особое внимание уделено Data Mining.

### Ключевые слова

Жизненный цикл изделия, большие объемы данных, MapReduce, NoSQL, Hadoop, Disco, In-Memory, реляционная СУБД

Поступила в редакцию 22.06.2018

© МГТУ им. Н.Э. Баумана, 2018

---

**Введение.** Большие данные в информационных технологиях — серия подходов, инструментов и методов обработки структурированных и неструктурированных данных больших объемов и значительного многообразия для получения воспринимаемых человеком результатов, эффективных в условиях непрерывного прироста, распределения по многочисленным узлам вычислительной сети, сформировавшихся в конце 2000-х годов, альтернативных традиционным системам управления базами данных (СУБД) и решениям класса Business Intelligence. В данную серию включают средства массово-параллельной обработки неопределенно структурированных данных, прежде всего решениями категории NoSQL, алгоритмами MapReduce, программными каркасами и библиотеками проекта Hadoop.

В качестве определяющих характеристик для больших данных выбирают: объем, скорость (в смысле как скорости прироста, так и необходимости высокоскоростной обработки и получения результатов), многообразие (в смысле возможности одновременной обработки различных типов, структурированных и полуструктурированных данных). Сегодня также выделяют изменчивость и ценность данных как новые характеристики (<https://plmandbigdata.wordpress.com/>) [1].

Некоторые компании, разрабатывающие сложные технические системы, определяют термин «большие данные» (большие объемы данных) как информацию, которую невозможно обрабатывать и анализировать традиционными методами по причине ее большого объема [2]. Такая информация хранится и циркулирует в одной из самых необходимых систем, которая может улучшить создание новых наукоемких изделий и снизить производственные издержки путем управления данными о нем на протяжении всего жизненного цикла — системе управления жизненным циклом изделия (ЖЦИ).

Актуальность темы обосновывается постоянным совершенствованием технологий больших данных. На сегодняшний день создаются все новые методы и подходы к организации и обработке большого объема информации, проводится тотальный анализ уже существующих методов и возможные их изменения.

Большинство производителей имеют трудности в хранении и управлении таким объемом данных, поскольку традиционные методы не справляются с управлением «большими данными».

Целью данной работы является ознакомление с существующими методами обработки больших данных и их влиянием на жизненный цикл изделия. Для достижения данных целей были поставлены следующие задачи:

- определить на каком этапе жизненного цикла изделия используется управление большими данными;
- определить существующие методы обработки больших данных, используемые СУБД.

В настоящей работе рассмотрена роль больших данных на различных этапах жизненного цикла изделия и дальнейшие методы обработки и управления большими данными.

**Использование больших данных в жизненном цикле изделия.** Существует множество научных работ, связанных с управлением большим объемом данных в ЖЦИ (<https://plmandbigdata.wordpress.com/>) [1–6]. Так, в работе Ли [1] описывается роль методов по управлению «большими данными» на определенных этапах ЖЦИ.

Указанные в работах методы используются для следующих целей [1]:

- управление данными в составлении расписаний;
- управление цепочками поставок;
- массовое производство по индивидуальным заказам.

Рассмотрим каждую из этих целей подробнее. Для управления в составлении расписаний необходимо создать хранилища большого объема данных с высокой скоростью и для большого количества пользователей, где задачи по составлению расписаний упрощаются посредством использования динамических характеристик реального времени. Что касается управления цепочками поставок, то необходимы специальные технологии, которые будут улучшать сотрудничество компаний друг с другом, для создания аналитических прогнозов в управлении цепочками поставок целых сетей компаний, такие, например, как SCM-системы (Supply Chain Management System). Для развития производства, требуется собирать дан-

ные каждого потребителя и распределять необходимые ресурсы, что и обеспечит массовое производство под каждого заказчика.

Каждый жизненный цикл изделия можно разделить на три основных периода: начало жизни, середина жизни и конец жизни изделия, т. е. проектирование, производство и эксплуатация [7]. На каждом периоде жизни изделия большие данные играют важную роль.

Примеры различных видов входных и выходных данных на стадиях жизненного цикла (проектирование, производство, эксплуатация) представлены в таблице.

**Виды входных и выходных данных на стадиях жизненного цикла**

Входные данные		Выходные данные	
Категория	Главные данные	Категория	Главные данные
Потребительский спрос	Функция изделия, конфигурация, инструкция и т. д.	Проектировочная спецификация	Список материалов, список поставщиков, программный код и т. д.
Техническая поддержка	Главные причина поломки, важные компоненты и т. д.	Информация о изделии	Инструкция по сборке, история производства и т. д.
Информация компании	Альтернативный поставщик, аутсорсинговая фирма и т. д.	—	—
Инструкция пользователя	Функции изделия, инструкция по установке	Статус производства	Степень качества каждого компонента.
Заводская информация	Инструкция по сборке, спецификация и т. д.	Среда использования	Средняя влажность, температура и т. д.
Информация по поддержке изделия	Инструкция по обслуживанию	План поддержки	Инструменты, места, стоимость
История поддержки	Идентификационный номер компонентов, дата установки, список замененных частей	Информация о переработке	Производственная информация
Информация о статусе изделия	Условия использования	Информация об утилизации	Срок жизни изделия и частей, стоимость утилизации, информация о перепроизводстве

Рассмотрим роль данных на этапах жизненного цикла изделия.

В начале жизни изделия генерируют его концепцию, разрабатывают проект, который потом реализуют [1]. На этом этапе выявляют и анализируют определенные типы данных: исторические данные, показывающие, покупал ли заказ-

чик аналогичный продукт ранее; данные исследования рынка и сведения о посещенных страницах по предпочтительным веб-сайтам. Сочетание этих данных нуждается в технологии обработки больших данных для принятия решений и ориентации клиентов. После составления маркетингового анализа осуществляют закупку соответствующих материалов для производства необходимых изделий. При производстве продукта создается огромный объем документации (спецификации, чертежи и т. д.). Контроль качества продукции и моделирование также являются видами деятельности, имеющими тесные связи с обработкой больших данных. Что же касается оборудования, используемого при производстве изделия, то здесь тоже можно заметить применения «больших данных» — управление оборудованием, энергосбережение и т. д.

В середине жизни изделия происходит распределение, дальнейшая эксплуатация и обслуживание изделия. Одной из важных частей этапа середины жизни является управление складом. Управление складом — это сложная структура, которая рассматривает отношения между факторами принятия решения на глобальном уровне [1]. По сравнению с традиционными методами использование технологий больших данных поможет решить множество задач по управлению складом.

Методы больших данных вводятся в процесс обслуживания клиентов, а именно изделия, для удовлетворения требований клиента при использовании изделия.

На этапе конца жизни изделия происходит утилизация или, если необходимо, восстановление и повторное восстановление [1]. Сбор информации об утилизированных материалах, изделиях, также осуществляется с помощью методов больших данных.

В следующей главе представлены конкретные подходы (методы) по управлению и обработке «больших данных».

**Современные подходы к обработке больших данных.** Если рассматривать в качестве основного хранилища данных реляционную систему управления базами данных (СУБД), то можно столкнуться с рядом задач, с которыми реляционная архитектура не справилась бы. Поэтому, инженеры разработали новые решения, среди которых [2]:

1) покупка более мощного оборудования. Данный способ является наименее затратным, но не эффективным, поскольку объемы данных увеличиваются с геометрической прогрессией, а бесконечно мощного оборудования не существует;

2) оптимизация запросов и создание новых индексов. Данный способ является наиболее затратным. Такой метод может облегчить работу системы ненадолго, поскольку создание новых индексов порождает дополнительные операции. В свою очередь, дополнительные операции приводят к росту объема обрабатываемых данных, что приводит к внедрению избыточности.

Существуют и другие методы, которые способствуют усложнению клиентского программного обеспечения, а также поддержания работоспособности и отказоустойчивости разработанного решения.

Если анализировать выводы о реляционной СУБД, то прогнозы на следующие 5 лет неоптимистичны. Поэтому инженеры предлагают новые архитектуры, способные адаптироваться к возрастающим объемам данных.

Рассмотрим современные подходы к управлению большими данными более подробно.

**MapReduce** — данная модель предназначена для реализации параллельных вычислений над очень большими наборами данных (несколько петабайт). Модель MapReduce была разработана компанией Google в 2004 г. Главным преимуществом MapReduce заключается в том, что она позволяет распределенно производить операции предварительной обработки и свертки [2].

**NoSQL** — архитектура, предназначенная для описания реляционной базы данных без использования SQL [2]. Иными словами, если при работе с традиционными реляционными СУБД мы имеем один сервер баз данных для нескольких приложений, то в парадигме NoSQL одному приложению или модулю приложения предлагается отдельное решение по работе с данными.

*Основные черты NoSQL [2]:*

- высокая пропускная способность;
- способность к горизонтальному масштабированию по требованию для некоторого набора операций на многих серверах;
- отказ от строгой согласованности данных друг с другом и внутренняя непротиворечивость (консистентность данных). Жертвуя консистентностью данных, многие NoSQL базы данных реализуют доступность данных и устойчивость, следуя теореме CAP (или теореме Эрика Брюера).

**Hadoop** — основа архитектуры данного метода заключается в отказе от произвольного обновления данных и последовательной обработке всего набора для достижения серьезного прироста производительности.

Метод Hadoop используется для реализации поисковых и контекстных механизмов многих высоконагруженных веб-сайтов. Ценность Hadoop заключается в том, чтобы создать репрезентативную выборку из гигантских объемов первичных данных, поддерживая принцип Джона Фон Неймана — построение надежных систем из ненадежных элементов, и затем передать эту выборку для оперативной обработки в прикладные системы [2].

**Disco** — метод, который подходит для выполнения задач по распределению пакетной обработки больших объемов данных. Особенностью Disco является то, что ядро системы разработано на функциональном языке Erlang [2], отлично подходящем для программирования распределенных вычислений на основе модели акторов.

**СУБД с вертикальным хранением** данных позволяет значительно повысить скорость считывания больших объемов данных благодаря отсутствию необходимости при запросе нескольких ячеек в строке считывать всю строку целиком, как это реализовано в СУБД со строковой организацией. Данный метод предпочтительнее использовать при обработке в несколько миллионов строк в таблице [3].

*In-Memory СУБД* позволяют хранить и обрабатывать данные непосредственно в оперативной памяти, что значительно повышает скорость обработки [3].

Особый интерес представляет *Data Mining* и прогнозная аналитика, поскольку они имеют прямую связь с интеллектуальными системами.

Отметим, что отдельные технологии, способные в той или иной степени решать проблему фрагментации данных позволяют хранить огромные объемы информации, оперативно их обрабатывать, преодолевая информационный барьер [7]. Однако основной вопрос применения технологии, как и ее ценность, заключается не в скорости обработки и передачи потоков данных в единицу времени, а в возможности формирования действительно репрезентативной выборки на основе генеральной совокупности данных, и последующем применении индуктивных методов для установления закономерностей и построении достоверных аналитических прогнозов. И в этом отношении интеллектуальные технологии управления Большими Данными решающим образом приходят на помощь науке, делая возможным, кроме алгоритмических вычислений — моделирование процессов с помощью таких инструментов как генетические алгоритмы, нейронные сети и т. д. [8]

Инструменты управления потоками событий предназначены для обработки больших объемов разнородных данных, неупорядоченных, поступающих в реальном времени из самых разных источников, таких как смартфоны, сети, датчики, электронные письма, сделки с ценными бумагами и т. д. Выделим следующие характерные сценарии для решений подобного класса.

1. *Выявление ситуации.* Мониторинг поступающих событий на предмет обнаружения шаблонов, свидетельствующих о появлении возможности либо проблемы, т. е. такой ситуации, которая требует реакции либо должна быть зафиксирована [9]. Мониторинг может осуществляться с помощью как простого фильтра, так и сложного набора правил, идентифицирующих корреляцию между входящими событиями и просматривающих их в поиске наборов условий (в том числе и таких, как отсутствие событий). В результате генерируются высокоуровневые события, сигнализирующие о наступлении соответствующей ситуации.

2. *Агрегация и анализ данных — непрерывные вычисления* [10]. В данных ведется поиск корреляций, они группируются и агрегируются, затем выполняются расчеты, результатом которых является новая информация, такая как сводные данные, высокоуровневая статистика, либо значения для коррекции основных производственных параметров.

3. *Сбор данных.* Работа технологий по управлению потоками событий сопровождается формированием наборов исходных данных о событиях или сводных данных более высокого уровня. Собранные данные могут быть использованы в качестве контекста для вновь поступающих событий, а также сохранены в исторической базе данных для последующего анализа, отчетности или в целях аудита.

**Заключение.** При работе с большими объемами данных возникают две задачи: как управлять большим объемом информации и как его эффективно применять. В данной работе были рассмотрены основные цели использования

больших данных на различных стадиях жизненного цикла изделия, а также, методы по обработке больших данных. Использование представленных методов повысит эффективность проектирования производства, процесса обслуживания, проектирования изделия и его качества. Заметим, однако, что ни один из этих подходов не предлагает решение всех поставленных задач, которые возникают при управлении большими объемами данных. Поэтому необходимо осуществлять синтез решений в зависимости от поставленных задач.

### Литература

- [1] Li J., Tao F., Cheng Y., Zhao L. Big Data in product lifecycle management. *Int. J. Adv. Manuf. Technol.*, 2017, vol. 81, no. 1–4, pp. 667–684.
- [2] Марфа Э. Управление большими данными, работа с поставщиками и защита информации. *Машиностроение и смежные отрасли*, 2014, № 4, с. 38–40.
- [3] Табаков В.В., Федотова А.В. Интеллектуализация технологий управления большими данными. *Интеллектуальные системы и технологии: современное состояние и перспективы. Сб. науч. тр. 3-й Межд. школы-семинара*. Тверь, ТвГТУ, 2017, с. 57–62.
- [4] Woo T., Krensky P. Optimizing product lifecycle management using BigData analytics. URL: <http://aberdeen.com/research/11657/11657-rr-big-data-analytics/content.aspx> (дата обращения 18.04.2018).
- [5] Hayes J. Using PLM to enable IoT and big data in an old industry. URL: <http://www.engineering.com/PLMERP/ArticleID/10908/Using-PLM-to-Enable-IoT-and-Big-Data-in-an-Old-Industry.aspx> (дата обращения 18.04.2018).
- [6] Федотова А.В., Ветров А.Н., Тарасов В.Б. Грануляция информации при моделировании жизненного цикла сложных технических систем. *Науковедение*, 2013, № 5(18). URL: <http://naukovedenie.ru/PDF/53tvn513.pdf>.
- [7] Глушков В.М. *Основы безбумажной информатики*. Москва, Наука, 1987, с. 9–11.
- [8] Табаков В.В. SAP HANA: сумма технологии. URL: <https://sapland.ru/kb/blogs/sap-hana-summa-tehnologii.html> (дата обращения 05.02.2017).
- [9] Анализ и использование динамических данных: введение в обработку сложных событий. URL: <http://docplayer.ru/51335498-Analiz-i-ispolzovanie-dinamicheskikh-dannyh-vvedenie-v-obrabotku-slozhnyh-sobytiy.html> (дата обращения 14.07.2017).
- [10] Агрегирование и группировка данных. URL: [https://www.politerm.com/zuludoc/sql\\_aggreg.html?sphrase\\_id=23692](https://www.politerm.com/zuludoc/sql_aggreg.html?sphrase_id=23692) (дата обращения 20.01.2017).

**Киселёва Екатерина Юрьевна** — магистр кафедры «Компьютерные системы автоматизации производства», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

**Федотова Алена Валерьевна** — кандидат технических наук, доцент кафедры «Компьютерные системы автоматизации производства», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

**Научный руководитель** — Тарасов Валерий Борисович, кандидат технических наук, доцент кафедры «Компьютерные системы автоматизации производства», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

---

## THE ROLE OF GREAT DATA IN THE LIFE CYCLE OF THE PRODUCT. MODERN APPROACHES TO MANAGEMENT OF GREAT DATA

E.Yu. Kiseleva

eanufrieva94@gmail.com

SPIN-code: 7751-2065

A.V. Fedotova

afedotova.bmstu@gmail.com

SPIN-code: 4099-0349

**Bauman Moscow State Technical University, Moscow, Russian Federation**

---

### Abstract

The article presents the purposes for which methods for managing large data at specific stages of the product life cycle (PLC) are defined. Three main periods of the PLC are analyzed and, according to them, examples of different types of input and output data are presented. Modern approaches to big data processing, including bypassing the use of the relational database management system (DBMS) are also presented. The paper considers the following approaches in managing big data: MapReduce, NoSQL, Hadoop, Disco. Approaches used by the DBMS, such as DBMS with vertical storage and In-memory DBMS, are considered. Particular attention is paid to Data Mining.

### Keywords

Product life cycle, big data processing, MapReduce, NoSQL, Hadoop, Disco, In-Memory, relational DBMS

Received 22.06.2018

© Bauman Moscow State Technical University, 2018

---

### References

- [1] Li J., Tao F., Cheng Y., Zhao L. Big Data in product lifecycle management. *Int. J. Adv. Manuf. Technol.*, 2017, vol. 81, no. 1–4, pp. 667–684.
- [2] Marfa E. Big data management, supplier management and information protection. *Mashinostroenie i smezhnye otrasli*, 2014, no. 4, pp. 38–40.
- [3] Tabakov V.V., Fedotova A.V. Intellektualizatsiya tekhnologiy upravleniya bol'shimi dannymi [Intellectualization of big data management technology]. *Intellektual'nye sistemy i tekhnologii: sovremennoe sostoyanie i perspektivy. Sb. nauch. tr. 3-y Mezhd. shkoly-seminara* [Intellectual systems and technologies: contemporary state and prospects. Proc. 3<sup>rd</sup> Int. School-Workshop]. Tver', TSTU publ., 2017, pp. 57–62.
- [4] Woo T., Krensky P. Optimizing product lifecycle management using BigData analytics. Available at: <http://aberdeen.com/research/11657/11657-rr-big-data-analytics/content.aspx> (data obrashcheniya 18 April 2018).
- [5] Hayes J. Using PLM to enable IoT and big data in an old industry. Available at: <http://www.engineering.com/PLMERP/ArticleID/10908/Using-PLM-to-Enable-IoT-and-Big-Data-in-an-Old-Industry.aspx> (accessed 18 April 2018).
- [6] Fedotova A.V., Vetrov A.N., Tarasov V.B. Information granulation in life cycle modeling for complex technical systems. *Naukovedenie*, 2013, no. 5(18). Available at: <http://naukovedenie.ru/PDF/53tvn513.pdf>.
- [7] Glushkov V.M. Osnovy bezbumazhnoy informatiki [Basics of paper-free informatics]. Moscow, Nauka publ., 1987, pp. 9–11.
- [8] Tabakov V.V. SAP HANA: summa tekhnologii [SAP HANA: technology sum]. Available at: <https://sapland.ru/kb/blogs/sap-hana-summa-tehnologii.html> (accessed 05.02.2017).

- [9] Analiz i ispol'zovanie dinamicheskikh dannykh: vvedenie v obrabotku slozhnykh sobytiy [Analysis and application of dynamic data: introduction into complex event processing]. Available at: <http://docplayer.ru/51335498-Analiz-i-ispolzovanie-dinamicheskikh-dannyh-vvedenie-v-obrabotku-slozhnyh-sobytiy.html> (accessed 14 July 2017).
- [10] Agregirovanie i gruppirovka dannykh [Data aggregation and grouping]. Available at: [https://www.politerm.com/zuludoc/sql\\_aggreg.html?sphrase\\_id=23692](https://www.politerm.com/zuludoc/sql_aggreg.html?sphrase_id=23692) (accessed 20 January 2017).

**Kiseleva E.Yu.** — Master's Degree student, Department of Computer Systems of Manufacturing Automation, Bauman Moscow State Technical University, Moscow, Russian Federation.

**Fedotova A.V.** — Cand. Sc. (Tech.), Assoc. Professor, Department of Computer Systems of Manufacturing Automation, Bauman Moscow State Technical University, Moscow, Russian Federation.

**Scientific advisor** — V.B. Tarasov, Cand. Sc. (Tech.), Assoc. Professor, Department of Computer Systems of Manufacturing Automation, Bauman Moscow State Technical University, Moscow, Russian Federation.