

# ГОЛЕМИТЕ ДАННИ – ФАКТОР ЗА КОНКУРЕНТНО ПРЕДИМСТВО В ГЛОБАЛИЗИРАЩИЯ СЕ СВЯТ

хон. ас. Станислава Клисарова-Белчева

ПУ „П. Хилендарски“, ФИСН

## BIG DATA – A SOURCE OF COMPETITIVE ADVANTAGE IN A GLOBALIZING WORLD

Stanislava Klisarova-Belcheva

Plovdiv university „P. Hilendarski“, FESS

**Abstract:** *In the last few years, IT professionals and business analysts focused their attention on cloud computing and virtualization, but also on big data. Big data are a set of structured and unstructured data of large volume, which are generated at a high pace from many and varied sources, have different formats and need real-time analysis. Big data help managers foresee changes in business environment faster and react more effectively with optimal solutions to specific cases. The aim of this paper is making an overview of a possible implementation of this innovative technology as a tool for developing a competitive advantage. The proposed application is based on a combination of relational and non-relational databases for storing and redistribution of the growing volume of business information.*

**Key words:** Big Data, Business Analysis, Databases

**Резюме:** *През последните няколко години на вниманието на ИТ специалистите и бизнес анализаторите наред с облачните услуги и виртуализацията се наредиха и големите данни. Те представляват съвкупност от структурирани и неструктурирани данни с голям обем, които се генерират с огромна скорост от различни и многобройни източници. Притежават различен формат и се нуждаят от анализ в реално време. Големите данни помагат на бизнеса да може да предвиди промените в бизнес средата по-бързо и да реагира по-ефективно и адекватно с оптимално предложение за конкретния случай. Целта на настоящата работа е да направи обзор на възможните имплементации на иновативната технология като инструмент за постигане на конкурентно предимство. Предложеният вариант се базира на комбинация от релационни и нерелационни бази данни за съхранение и преразпределение на нарастващия обем информация.*

**Ключови думи:** големи данни, бизнес анализ, бази данни

През последните няколко години във фокуса на вниманието на ИТ специалистите и бизнес анализаторите наред с облачните услуги и виртуализацията се наредиха и големите данни. Настоящият доклад анализира ролята на големите данни като източник на конкурентно предимство при вземане на бизнес решения. Целта е да се докаже значимостта на големите данни за бизнеса и да сравнят алтернативните възможности за реализация на тази иновативната технология.

### 1. Същност на понятието Big Data

Big data или големи данни е всеобхватно понятие за съвкупност от данни, които са толкова комплексни и с огромен обем, че стават трудно обработваеми от традиционния софтуер за бази данни. Големите данни представляват колекция от данни с размери, надвишаващи възможностите на често използваните софтуерни инструменти за събиране, обработка и управление. Източниците на големи данни са традиционни или дигитални, в или извън компанията. Въпреки че терминът големи данни често се свързва с обема данни, той има и друго значение. С големи данни се обозначава и технологията, включваща инструменти и процеси, чрез които организациите и компаниите се справят с големи обеми от данни.

Основните характеристики на големите данни са следните:

1. Обем – той е толкова голям и неструктуриран, че е невъзможен за обработка от стандартите системи за управление на бази данни;
2. Разнообразие – големите данни включват както структурирани, така и неструктурирани и много разнообразни данни, което ги прави трудни за изследване и анализ;
3. Вариации – данните могат да имат много различни представяния или да се срещат множество пъти. Това е сериозна пречка пред анализаторите на данните;
4. Сложност – поради факта че данните се събират от много източници управлението на данните става доста сложен процес;
5. Скорост – не на последно място по значение е скоростта на обработка, съхранение и извличане на информация, която те предлагат.

Големите данни са мощни, хаотични и всеобхващащи, но те могат да бъдат контролирани по начин, който да донесе икономическо предимство. Приложението им се обуславя от факта, че все повече компании генерират нарастващо количество данни и обменят информация помежду си. Чрез анализ на събраната информация освен на традиционния въпрос „Какво се е случило?“ вече може да се отговори и на въпроса „Какво ще се случи?“. Това е много важно за бизнеса поради възможността да се предвидят с по-голяма точност бъдещи събития като се елиминират странични и несъществени фактори.

Големите данни включват както структурирани данни, събрани от бизнес приложенията за продажби, парични потоци, взаимодействие между звена, център за обслужване на клиенти, така и неструктурирани данни от социални медии и интернет, а също и от различни датчици.

„Големите данни да не просто огромно количество данни. Големите данни в бизнес приложенията са технология и концепция“, посочва анализаторът Марио Майер-Хубер, IDC [1].

Според проучване на KPMG, 7 от 10 висши ръководители на фирми са на мнение, че анализът на данните е „най-важната“ или „много важна“ движеща сила на бизнеса. Интересът към тази тема е още по-голям поради факта, че обемите данни продължават да растат с бързи темпове. Проучване на IDC и EMC сочи, че на всеки две години цифровата вселена удвоява обема си и към 2020 г. тя ще достига 40 трилиона гигабайта. Все повече компании имат достъп до информация с по-голям обем, отколкото са в състояние да управляват. Тези организации се нуждаят от подходящи ИТ инструменти и квалифицирани сътрудници, за да извлекат от „големите данни“ максимална полза.

## **2. Big Data като фактор за конкурентно предимство на фирмите в условията на глобализация**

Според концепцията за конкурентното предимство конкурентоспособни са организациите, които предлагат на потребителите по-голяма стойност, или чрез пониски цени или чрез предоставяне на по-големи ползи и услуги, което оправдава по-високите цени. Големите данни предоставят на организациите точна и навременна информация за всеки аспект от дейността им въз основа на прецизен анализ.

Ефективността на процеса на бизнес планиране и увеличаването на производителността на компаниите зависят до голяма степен от внимателното изследване на данните и съставянето на нови печеливши бизнес модели. Динамичната ико-

номическа среда налага да се разработват нови стратегии и да се минимизират рисковете при вземане на дългосрочни решения, което е невъзможно без внимателно изследване и прилагане на резултатите от анализа на големите данни.

Защо големите данни са съществен фактор за конкурентоспособността на бизнеса може да се открие в някои актуални проучвания на големи консултантски компании като Gartner, ИТС и др.

Големите данни осигуряват по-висока информираност – колкото повече информация има, толкова по-подробен анализ може да се направи и толкова по-полезни резултати са извлечените резултати. Така може да се работи в диалогов режим с потребителя като му се отправят подходящи предложения. Например, ако даден клиент посети виртуален банков клон, той може веднага да получи подходящите предложения според потребителския му профил, извлечен от големите данни. С помощта на големите данни търговците на дребно генерират най-подходящата оферта за конкретния клиент.

В други случаи резултатите от анализа на големите данни са основа за редизайн на продуктите с цел тяхното усъвършенстване или подпомагат тяхната промоция. Анализът на неструктурираните данни, събрани в социалните медии, може да разкрие нов потенциал и допълнителни клиенти за бизнеса. Полезна е и възможността за анализ на риска, за да се предскаже вероятност за просрочие на клиенти и доставчици. Чрез прогнозен анализ на големите данни може да се следят съвременните тенденции в дадена индустрия.

Конкурентоспособността се повишава и чрез създаването на нови и иновативни продукти, които точно съответстват на изискванията на клиентите. Конкурентоспособността на компанията нараства, защото предлага уникален и ценен продукт, създаден за конкретния клиент.

Резултатите от анализите могат да генерират и косвени приходи. Например, получената обобщена статистическа информация може да се разпространява срещу заплащане. Данните могат да поддържат организацията в безопасност като разкриват потенциални вътрешни заплахи. Това би позволило на мениджърите да откриват чувствителна информация, която не е защитена и съхранена в съответствие с нормативните изисквания.

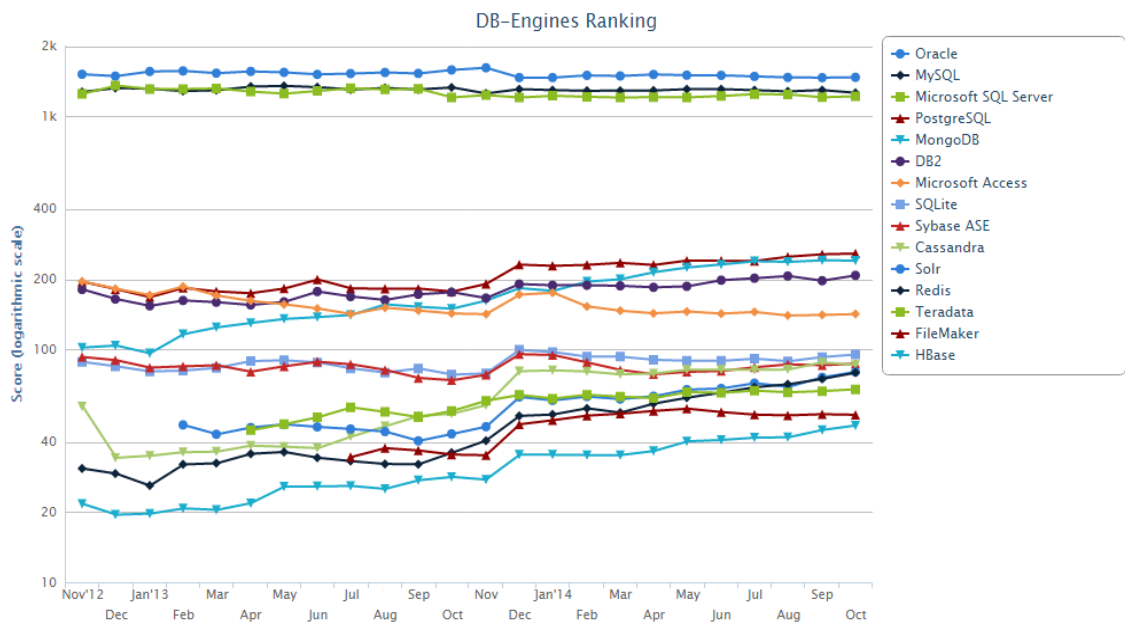
Примери за използване на големи данни може да се намерят в много части на света. Човешкият поток в Осло потокът управлява интелигентното улично осветление, което намалява консумацията на енергия с 62%. В Орегон с големи данни се създава „зелена вълна“, намаляваща вредните емисии.

Според изследователите прилагането на софтуер за бизнес анализи на големи данни ще нарасне значително през следващите години. Големите данни са надежден инструмент за постигане и запазване на конкурентно предимство в нестабилната икономическа обстановка, тъй като те:

- Генерират по-точна информация;
- Улесняват процеса на вземане на бизнес решения;
- Подобрят маркетинговата стратегия и посоката на развитие на компанията;
- Повишават приходите поради увеличение на клиентите и намаляват разходите.

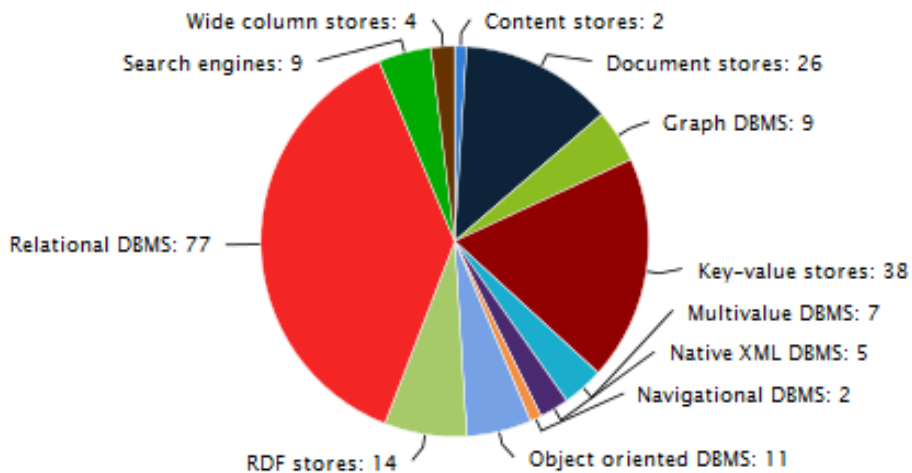
### 3. Сравнителен анализ на алтернативите за реализация на Big Data

Необходимостта да се усъвършенства работата с големи обеми информация в организациите е трудоемка задача за производителите на бизнес софтуер. Според проучване сред разработчиците най-често използваният модел за съхранение и обработка на информацията е реляционният. Подобна е ситуацията при Dynamics на Microsoft, Business suite на SAP, както и при много други използващи съответно: MS SQL, Oracle, DB2, MySQL, PostgreSQL, Firebird и т.н. съгласно данните от проучване на DB-Engines от м. октомври 2014 [2].



Фиг. 1. Проучване на използваните СУБД сред разработчиците

Тъй като големите данни включват огромен обем от структурирана и неструктурирана информация, реляционните бази данни не са в състояние да се справят с управлението им. За тази цел през последните години активно се разработват и прилагат иновативни специализирани решения.



Фиг. 2. Разпределение на използваните СУБД

Според проучване, проведено сред разработчиците на софтуерни решения за големи данни [3] все по-често в практиката се прилагат NoSQL бази данни като Cassandra, Redis, HBase, MongoDB.

NoSQL (Not Only SQL) са разширение на познатия реляционен модел, в който се използват SQL запитвания. Основните видове NoSQL бази данни са:

- документно ориентирани;
- граф;
- ключ-стойност;
- обектно-ориентирани;
- RDF.

Исторически първите NoSQL бази данни са BigTable на Google и Dynamo на Amazon, които въвеждат част от термините и показват, че управлението на големите данни не е невъзможно или непосилна задача. Тези системи могат да се използват само на сървърите на създателите им, но поради отворената им документация силно влияят на останалите: Cassandra от Dynamo и BigTable. На базата на Google GFS и BigTable стартира Hadoop. Съхранението във вид ключ-стойност или документния модел са реализирани в MongoDB и CouchDB [4].

Hadoop е една от най-разпространените в практиката към момента технологии за големи данни. Това е инфраструктурен софтуер, който извлича, съхранява и обработва огромни структури от данни. Тази платформа е представена от Apache и има творен код. Основните Hadoop характеристики са следните:

1. Как се съхраняват данните (файловете)?
2. Как се обработват данните?

Hadoop е система, която използва клъстери за съхранение на информацията. Идеята за съхраняването на данните е заимствана от Google GFS (Google File System), като се използва HDFS (Hadoop Distributed File system). Системата може да съхранява множество файлове с практически неограничен размер, като те се разпределят между множество сравнително евтини компютри.

Втората специфична характеристика е възможността за обработка на данните. За целта е създадена специална платформа, наречена MapReduce. При нея методът за извличане на данни е разделен на две: **Map** и **Reduce**. **Map** функцията връща 0, 1 или няколко записа, а **Reduce** обобщава резултатите в една стойност. Базата данни, включена в този пакет, се нарича HBase.

Друга технология, с чието развитие се свързва бъдещето на големите данни е In-Memory Computing. Тя служи за обработка на големи данни в реално време. Името на технологията показва, че тя е ориентирана към изчисления в паметта, т.е. технологията дава достъп до данните директно от паметта. По този начин компаниите анализират данните в реално време, подобряват планирането, усъвършенстват моделирането, реализират мониторинг, оптимизират производствените процеси и намаляват производствените разходи. Пример за реализация на In-Memory Computing е SAP HANA. Платформата интегрира софтуер и хардуер, като си сътрудничи с водещи компании за хардуер като IBM, HP, Fujitsu, Cisco, Dell, Hitachi и други.

## Заклучение

Появата на големите данни като следствие от глобализацията се свят и развитието на информационните технологии за тяхната обработка, насърчава използването им от бизнес анализаторите и мениджърите поради многобройните

предимства, които те предлагат за бизнеса. Нарастващият обем на големите данни може да се контролира от NoSQL бази данни и технологии, базирани на NoSQL решения, които подпомагат процеса на вземане на решения в постоянно променящата се бизнес среда. Прилагането на големите данни е предпоставка за адекватна, бърза и ефективна реакция на бизнеса, ще улесни намирането на оптималното решение за проблемните ситуации и в крайна сметка, ще увеличи конкурентоспособността на бизнеса.

#### **ЛИТЕРАТУРА**

1. IDC Big Data and Business Analytics Forum 2014 <http://idc-cema.com/>
2. Forbs – What is big data – <http://www.forbes.com/sites/lisaarthur/2013/08/15/what-is-big-data/>
3. DBMS popularity by database model – [http://db-engines.com/en/ranking\\_categories](http://db-engines.com/en/ranking_categories)
4. Lith, Adam at all. Investigating storage solutions for large data: A comparison of well performing and scalable data storage solutions for real time extraction and batch insertion of data. Department of Computer Science and Engineering, Chalmers University of Technology, 2010.