Zeszyty Naukowe PWSZ w Płocku
Nauki Ekonomiczne, t. XXIII, 2016.

Katarzyna Racka
Państwowa Wyższa Szkoła Zawodowa w Płocku

BIG DATA – ZNACZENIE, ZASTOSOWANIA I ROZWIĄZANIA TECHNOLOGICZNE

Streszczenie:

Technologie Big Data i ich zastosowanie do procesów biznesowych rozwijają się w tempie dynamicznym. Przedsiębiorstwa analityczno-doradcze specjalizujące się w zagadnieniach strategicznego wykorzystania technologii IT informują, że z roku na rok zwiększa się liczba przedsiębiorstw wdrażających lub planujących wdrożenie rozwiązań technologicznych związanych z Big Data. Dużo przedsiębiorstw uważa, że analizy danych niestrukturalnych będą kluczem do głębszego zrozumienia zachowań klienta. Uważają one, że analityka jest absolutnie niezbędna lub bardzo ważna dla prowadzenia ogólnej strategii biznesowej przedsiębiorstwa oraz do poprawy wyników operacyjnych.
Celem tego artykułu jest wyjaśnienie co dokładnie oznacza pojęcie Big Data, co to są dane niestrukturalne oraz jakie mogą mieć zastosowania. Ponadto, w artykule prezentuję wyniki raportów dotyczących wdrażanie technologii Big Data i omawiam przykładowe technologie związane z Big Data.

Słowa kluczowe: Big Data, NoSQL, MapReduce, Hadoop.

Wprowadzenie

Szybko rozwijający się postęp informatyczny, ciągłe gromadzenie bardzo dużych ilości danych, zwanych Big Data, spotyka się z problemami wynikającymi z ograniczeń sprzętowych i czasu potrzebnego do przeprowadzania analiz tych danych. Zmusza to do tworzenia coraz bardziej skomplikowanych algorytmów potrzebnych do analizy danych.

Z drugiej jednak strony czym większe zbiory danych są analizowane tym większe prawdopodobieństwo odkrycia bardziej znaczącej wiedzy w nich ukrytej, którą można wykorzystać w procesach biznesowych i produkcyjnych. Poprawa strategii biznesowych, na podstawie analizy zgromadzonych danych, to także szansa na osiągnięcie przez przedsiębiorstwo przewagi konkurencyjnej.

Na stonie internetowej EMC[2] w 2014 r. przedstawiono wyniki badań EMC Digital Universe with Research & Analysis by IDC *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*

[www 2]. Z badań tych wynika, że do roku 2020 wskaźnik danych, które tworzymy i kopiujemy rocznie sięgnie 44 zettabajty, lub 44 biliony gigabajtów.

Wiele przedsiębiorstw posiada dane zgromadzone za pomocą kamer, czujników, anten, nagrań. Te zróżnicowane dane, są często danymi niestrukturalnymi i nie mogą być przechowywane w tradycyjnych bazach relacyjnych. Jednak wiedza w nich zgromadzona może być bardzo znacząca, pozwalając na analizę trendów, anomalii rozwoju firmy, oceny klienta a także przewidywania istotnych zagrożeń finansowych.

Sposobem na powyższe problemy są nowe rozwiązania technologiczne Big Data, oferujące nierelacyjne bazy danych umożliwiające przechowywanie danych niestrukturalnych oraz przetwarzanie rozproszone dużych zestawów danych. Rozwiązania te pozwalają na obniżenie czasu i kosztów związanych z analizą bardzo dużych danych.

Analiza Big Data i jej zastosowanie do procesów biznesowych rozwija się w tempie dynamicznym – z badań przeprowadzonych przez firmę Gartner wynika, że do 2020 roku, ponad 80 procent procesów biznesowych w firmach będzie oparte o Big Data [www 5].

1. Definicja Big Data

Pojęcie Big Data oznacza dane lub zbiór danych, które są tak duże i złożone, że tradycyjne aplikacje przetwarzania danych są niewystarczające do analizy tych danych.

W praktyce pojęcie Big Data jako dużych danych jest pojęciem względnym. Biorąc pod uwagę rozwój technologii informacyjnych i komunikacyjnych, inne rozmiary i problem złożoności występował w dużych danych 19 lat temu, inny 15 a jeszcze inny teraz.

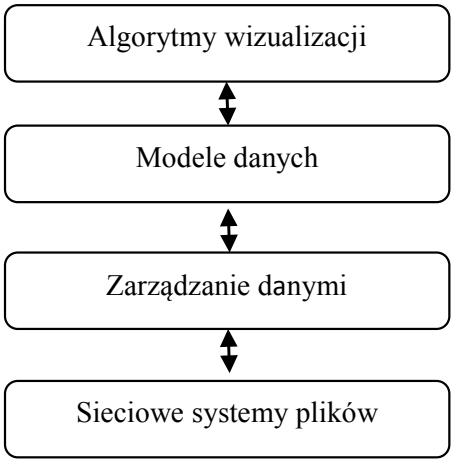
Pojęcie Big Data było omawiane już w roku 1997 przez M. Cox, D. Ellsworth w *Managing Big Data for Scientific Visualization* [Cox i Ell, 1997]. Zauważyli oni, że istnieją dwa różne znaczenia pojęcia Big Data, takie jak:

- duże zbiory danych (ang. big data collections)
- dane będące dużymi obiektami (ang. big data objects),

które są zbyt duże, aby być przetwarzane za pomocą standardowych algorytmów i oprogramowania na pojedynczym sprzęcie.

Celem M. Cox, D. Ellsworth było: wydobywanie informacji z danych, maksymalizacja danych i zbiorów danych, tak aby nadal mogły być analizowane, rozważając również problem ich interaktywności i wizualizacji w czasie rzeczywistym. Zaprezentowali oni schemat architektury warstwowej wizualizacji i zarządzania danymi.

Rysunek 1. Schemat architektury warstwowej wizualizacji i zarządzania danymi

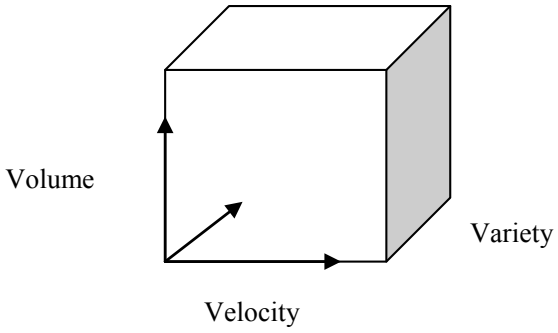


Źródło: Opracowanie własne na podstawie: Cox i Ell [1997]

W roku 2001 Doug Laney z firmy META Group (obecnie Gartner – firma analityczno-doradcza) w publikacji *3D Data Management: Controlling Data Volume, Velocity, and Variety* [Laney, 2001], pojęcie Big Data definiuje w modelu „3V":

- Objętość (ang. volume) – ilość danych.
- Prędkość (ang. velocity) – szybkość, z jaką dane są generowane i przetwarzane.
- Różnorodność (ang. variety) – rodzaj i charakter danych.

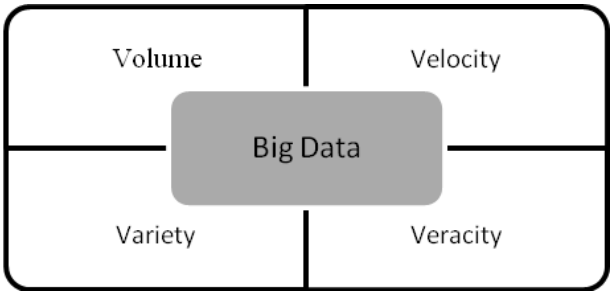
Rysunek 2. Trójwymiarowy model Big Data 3V



Źródło: Opracowanie własne na podstawie: Laney [2001].

W kolejnych latach model 3V został rozbudowany o dodatkowy wymiar wiarygodność (ang. veracity), tworząc model 4V.

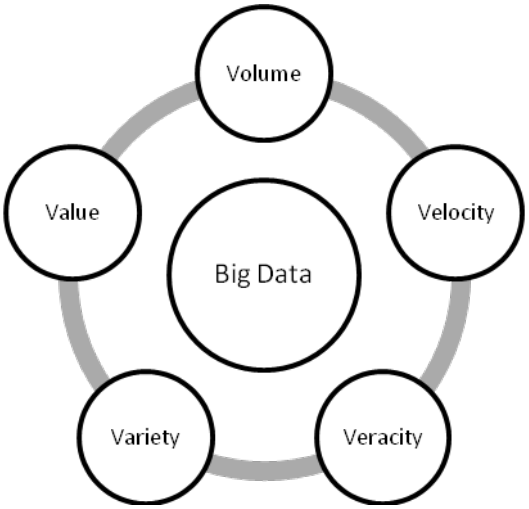
Rysunek 3. Big data 4V



Źródło: Opracowanie własne.

Dodanie do modelu 4V cechy wartość (ang. value) pozwoliło przedstawić Big Data w modelu 5V.

Rysunek 4. Big data 5V



Źródło: Opracowanie własne na podstawie B. Marr, *Big Data: The 5 Vs Everyone Must Know,* 2014 r. [www 4]

1.1. Charakterystyka Big Data

Pojęcie Big Data jest więc charakteryzowane za pomocą objętości – wielkości (ang. volume), prędkości (ang. velocity), różnorodności (ang. variety), wartości (ang. value) i wiarygodności (ang. veracity).

Volume – objętość

Obecnie gromadzone dane osiągają ogromne rozmiary, liczone w terabajtach, petabajtach. Wszystkie badania potwierdzają, że zbieranych danych

będzie coraz więcej. Tak duża ilość danych jest obecnie wyzwaniem dla tradycyjnych systemów IT. Większość firm albo nie ma odpowiednich struktur nadających się do przechowywania dużej ilości danych, albo brakuje im zdolności do przetwarzania takich danych przy użyciu efektywnej techniki przetwarzania zapytań. Dzięki technologii Big Data istnieją rozwiązania umożliwiające przechowanie i wykorzystywanie tych danych za pomocą systemów rozproszonych, w których części danych są przechowywane w różnych miejscach i zbierane przez odpowiednie oprogramowanie.

Velocity – prędkość

Prędkość danych odnosi się do szybkości, z jaką dane napływają oraz czasu w jakim są analizowane. By wydobyć ważne informacje, z danych szybko napływających, a co za tym idzie szybko się zmieniających, dane te muszą być analizowane w czasie rzeczywistym. Dane szybko przekazywane mogą być przesyłane strumieniowo, jednak wymagają one skutecznych technik składowania. Technologia Big Data pozwala na analizowanie danych podczas ich generowania, bez wprowadzania ich do bazy danych.

Variety – różnorodność

Dane pochodzą z wielu źródeł. Różnorodność tych danych, odnosi się do zróżnicowanego charakteru danych źródłowych, typu danych (strukturalnych i niestrukturalnych), formatu w jakim zostały zapisane. Dane mogą zawierać tekst z portali społecznościowych, dane z czujników, zdjęcia, filmy, nagrania dźwiękowe. Takie zróżnicowane dane, często są danymi niestrukturalnymi i nie mogą być przechowywane w tradycyjnych relacyjnych systemach baz danych ze względu na statyczny charakter schematów. Aby zapisać takie dane, korzysta się z baz NoSQL, które nie wymagają tradycyjnego schematu danych.

Veracity – wiarygodność

Wiarygodność danych dotyczy prawdziwości danych, jak również ich prawidłowości integralności. Dane, które będą poddawane analizie i eksploracji (data mining) muszą zawierać prawdziwe informacje. Nie powinny zawierać błędów, którymi mogą być np. błędy wynikające z pomyłek operatora (ortograficzne, literówki), różne formy danych oznaczające te same informacje, niezgodności wartości pola i jego opisu, brakujące wartości, błędy wynikające z ograniczeń systemu (brak pól na niektóre ważne dane), wielokrotne wprowadzenie tej samej danej.

Value – wartość

Przed przystąpieniem do analizy Big Data należy ustalić, które dane będą poddane analizie, czyli które dane będą istotne, wartościowe, a które bezużyteczne. To przyspieszy proces analizy danych.

2. Zastosowania Big Data

Jak podają autorzy V. Mayer-Schönberger, K. Cukier książki pt. BIG DATA *Rewolucja, która zmieni nasze myślenie, pracę i życie* [Mayer – Schönberger i Cukier, 2014, s. 13], w 2009 roku odkryto nowy szybko rozprzestrzeniający się wirus grypy, nazwany H1N1. Łączył w sobie fragmenty wirusów, które powodowały świńską i ptasią grypę. Instytucje zajmujące się zdrowiem na całym świecie obawiały się wybuchu pandemii. Aby ograniczyć prędkość rozprzestrzeniania się wirusa, służby te musiały znać miejsca, gdzie pojawiają się nowe przypadki zachorowań. W Stanach Zjednoczonych lekarze informowali o nowych przypadkach grypy Centrum Zwalczania i Zapobiegania Chorobom (CDC). Mapa pandemii była nieaktualna, ponieważ niektórzy chorzy chorowali kilka dni, zanim poszli do lekarza. Przekazywanie informacji do instytucji centralnych również wymagało czasu, ponadto CDC zestawiało dane tylko raz w tygodniu. Pomocne rozwiązanie przyniosła firma Google, która była w stanie sprawdzić jakie słowa ludzie wpisują w wyszukiwarkę internetową. Ponieważ każdego dnia w okno wyszukiwarki Google wpisywanych jest ponad trzy miliardy zapytań i wszystkie są archiwizowane, firma dysponowała ogromną liczbą informacji. Firma Google porównała 50 milionów najczęstszych fraz wyszukiwanych przez Amerykanów w internecie z danymi CDC dotyczącymi rozprzestrzeniania się sezonowej grypy w latach 2003–2008. Pomysł polegał na tym, żeby zidentyfikować obszary, na których pojawiła się grypa, na podstawie tego, czego ludzie szukali w internecie. Inżynierowie z tej firmy stworzyli system, z wykorzystaniem technologii Big Data, który został tak zaprojektowany, by szukać zależności między częstością pojawiania się pewnych zapytań a rozprzestrzenianiem się grypy. W celu sprawdzenia wyszukiwanych terminów przetworzyli 450 milionów różnych modeli matematycznych, porównując własne przewidywania z rzeczywistymi danymi dotyczącymi grypy z lat 2007–2008 dostarczonymi przez CDC. W efekcie końcowym, ich program znalazł kombinację 45 wyszukiwanych fraz, które wykorzystane w modelu matematycznym, dawały dużą korelację między przewidywaniami, a oficjalnymi liczbami z całego kraju. Inżynierowie firmy Google, podobnie jak CDC, mogli wywnioskować, gdzie pojawiła się grypa, jednak w przeciwieństwie do tej instytucji mogli to stwierdzić w czasie zbliżonym do rzeczywistego, a nie z jedno- lub dwutygodniowym opóźnieniem.

Technologie Big Data, oprócz przytoczonego powyżej przykładu medycznego mają również zastosowanie w wielu innych sektorach gospodarki, takich jak np. w :

- bankowości,
- finansach,
- ubezpieczeniach,
- doradztwie,
- telekomunikacji,
- turystyce,
- energetyce,

- przemyśle,
- budownictwie,
- logistyce,
- reklamie,
- administracji publicznej,
- informatyce – IT,
- handlu,
- nauce.

Na świecie inwestycje w technologie Big Data cały czas się rozwijają.

Firma Gartner z Stamford w USA, w roku 2012, 2013 [www 6] oraz 2014 [www 7] przeprowadziła badania, zorganizowane w celu określenia planów inwestycyjnych organizacji w technologie Big Data. Na podstawie badań z roku 2014 stwierdzono, że 73 procent respondentów zainwestowało lub zamierza inwestować w Big Data w ciągu najbliższych 24 miesięcy. W badaniach z 2013 roku wartość ta wynosiła jedynie 64 procent, a w roku 2012 tylko 58 procent. Z badań wynika również, że zmniejszyła się liczba respondentów deklarujących, że nie ma żadnych planów inwestycyjnych w Big Data, z 31 procent (w 2013 roku) do 24 procent (w 2014 roku). Na pytanie czy inwestycje w technologie Big Data będą nadal prowadzone, odpowiedziało pozytywnie 47 procent organizacji w roku 2014 r., a w roku 2013 zaledwie 37,8 procent. Jednak ten wzrost inwestycji w technologie Big Data nie doprowadził do wzrostu liczby organizacji zgłaszających wdrożenie projektów Big Data. Podobnie jak w 2013, również w 2014 roku wiele firm było na etapie tworzenia strategii rozwoju i projektów pilotażowych oraz eksperymentalnych.

W raporcie *2015 State of Analytics* opracowanym przez Salesforce [www 1] pokazano rosnące znaczenie analityki jako filaru działalności gospodarczej. W publikacji zebrano trendy opracowane na podstawie danych i opinii 2091 liderów biznesu z USA, Kanady, Brazylii, Wielkiej Brytanii, Francji, Niemiec, Japonii i Australii. Firmy biorące udział w badaniu podzielono na trzy grupy: przedsiębiorstwa, które wykazują najwyższe wyniki w swojej działalności (High Performers) w porównaniu z pozostałymi przedsiębiorstwami, przedsiębiorstwami z przeciętnymi wynikami (Moderate Performers), oraz przedsiębiorstwa mające najniższe wyniki (Underperformers). Najważniejsze cztery wnioski z badań *2015 State of Analytics* to:

- Analiza staje się czołowym elementem strategii biznesowych. 90 procent respondentów z grupy przedsiębiorstw o najlepszych wynikach mówi, ze analityka jest absolutnie niezbędna lub bardzo ważna dla prowadzenia ogólnej strategii biznesowej przedsiębiorstwa oraz do poprawy wyników operacyjnych.
- Sukces przedsiębiorstwa wymaga poszerzania analizy w wielu obszarach działalności. Przedsiębiorstwa o najlepszych wynikach analizują średnio ponad 17 różnych rodzajów danych – jest to prawie dwukrotnie większy wynik od pozostałych badanych przedsiębiorstw. Prowadzenie efektywności operacyjnej i ułatwie-

nie rozwoju są priorytetami dla współczesnych przedsiębiorstw. W celu poprawienia wydajności, przedsiębiorstwa koncentrują się na bardziej zaawansowanych przypadkach, takich jak: automatyzacja operacji biznesowych, tworzenie nowych modeli biznesowych i przewidywanie zachowań klientów.

- Rozpoczęła się era analityki w czasie rzeczywistym. Analitycy biznesowi coraz częściej odczuwają potrzebę podejmowania decyzji na podstawie analizy wykonywanej w czasie rzeczywistym. Przedsiębiorstwa osiągające najlepsze wyniki 5,1-krotnie częściej od innych potwierdzają, że ze swoich narzędzi analitycznych uzyskują informacje niezbędne do podejmowania decyzji na czas.
- Przedsiębiorstwa o najlepszych wynikach budują kulturę analityki, tzn. udostępniają informacje analityczne dla wszystkich działów przedsiębiorstwa. Przedsiębiorstwa z najlepszymi wynikami, dwa razy częściej niż pozostałe firmy potwierdzają, że ich połowa pracowników używa narzędzi analitycznych do codziennego użytkowania.

W badaniach *2015 State of Analytics* zwrócono również uwagę na:

- Analizowanie danych niestrukturalnych. Rynek analityki danych rozwija się bardzo dynamicznie. Zastosowanie analizy Big Data do procesów biznesowych stale rośnie. Z badań *2015 State of Analytics* wynika, że do 2020 roku, ilość analizowanych zasobów danych wzrośnie do 83 procent. Przedsiębiorstwa o najlepszych wynikach i największym wykorzystaniu analizy danych w strategiach biznesowych, w największym również stopniu zwiększają jej stosowanie w celu poprawienia korzyści.76 procent przedsiębiorstw o najlepszych wynikach potwierdza, że ich firma wykorzystuje narzędzia analityczne, aby uzyskać strategiczną wiedzę z Big Data. Ponadto te same przedsiębiorstwa 5,3 razy częściej potwierdzają, że analiza danych niestrukturalnych będzie kluczem do głębszego zrozumienia zachowań klienta. Dodatkowo 5,4-krotnie częściej, od przedsiębiorstw z najniższymi wynikami, wykorzystują narzędzia do analizy Big Data.

Rynek technologii i usług Big Data reprezentuje szybko rosnące, a zarazem rozwijające się, wielomiliardowe światowe możliwości. Amerykańska firma International Data Corporation (IDC), która jest dostawcą informacji rynkowych, usług doradczych i rozwiązań dla aplikacji informacyjnych, podaje że technologia i usługi na rynku Big Data w zeszłym roku rozwijały się w tempie sześciokrotnie szybszym niż cały rynek IT. Ponadto prognozy IDC pokazują że rynek technologii i usług Big Data będzie rosnąć o 6,4 procent potęgując roczny wzrost do 41,5 mld dolarów do 2018 r. [www 8]

3. Rozwiązania technologiczne Big Data

Tak jak już wcześniej zostało opisane, Big Data są bardzo dużymi zbiorami danych osiągającymi ogromne rozmiary. Ich dane są często niestrukturalne i nie mogą być przechowywane w tradycyjnych relacyjnych systemach baz danych. Prędkość z jaką dane napływają oraz czas w jakim się zmieniają wymusza aby dane te były analizowane w czasie rzeczywistym.

Zamiast kupować coraz lepsze maszyny do analizy Big Data (proces ten nazywany jest skalowaniem pionowym), w nowych rozwiązaniach Big Data, systemy rozbudowuje się poprzez dodawanie kolejnych maszyn, co jest określane jako skalowanie poziome.

Technologia Big Data pozwala na analizowanie danych podczas ich generowania, bez wprowadzania ich do bazy danych. Tworzone są również rozwiązania baz danych, które nie wymagają tradycyjnego schematu danych.

3.1. NoSQL

Nazwa NoSQL wywodzi się od słów „non SQL". Bazy NoSQL nazywane są również jako Not OnlySQL. NoSQL są to bazy danych zapewniające mechanizm przechowywania i udostępniania danych, które są modelowane w sposób inny niż tabelaryczny stosowany w relacyjnych bazach danych. Dzięki temu pozwalają one na gromadzenie i korzystanie z danych niestrukturalnych. Bazy NoSQL znane były już od 1960 roku, ale ich rozwój i największe zastosowanie zaczęło się wraz z pojawieniem Big Data. NoSQL to prostota konstrukcji, możliwość skalowania pojedynczych baz danych, dokładniejszy dostęp do danych i ich kontrola. Duża dostępność danych jest uzyskiwana w bazach NoSQL kosztem spójności danych, informacje zgromadzone w danych mogą różnić się w zależności od klastra. Dlatego bazy te nie mogą być stosowane w miejscach gdzie istotna jest dokładność danych. Kolejnym problemem, który wiąże się z bazami NoSQL jest brak jasnego sformalizowanego a zarazem prostego języka zapytań.

Typy baz danych NoSQL:

- Bazy dokumentowe (ang. documentstore) – dane przechowywane są tu jako dokumenty.
- Bazy klucz-wartość (ang. key-value stores) – w tym modelu, dane są reprezentowane jako zbiór par klucz-wartość.
- Bazy kolumnowe (ang. columnoriented stores) – w tym modelu dane zapisywane są w kolumnach.
- Bazy grafowe (ang. graphdatabase) – bazy danych wykorzystujące struktury grafów z węzłami, krawędziami – relacjami (teoria grafów).
- Bazy obiektowe – bazy o modelu obiektowym. Dane udostępniane są w postaci obiektowej.

3.2. MapReduce

Wiele z systemów Big Data zostało zapoczątkowanych dzięki platformie MapReduce, która jest produktem firmy Google.

MapReduce jest to model programowania służący do przetwarzania i generowania dużych zestawów danych. Pozwala on na przetwarzanie równoległe. Podstawowym założeniem tego modelu jest podział problemu na dwa główne etapy nazywane mapowaniem i redukcją. Rozproszony system plików z MapReduce pozwala na przetwarzanie danych w miejscu ich przechowywania. Dzięki temu rozwiązaniu nie ma potrzeby przesyłania informacji z komputerów magazynujących dane do serwerów. Zamiast przesyłać duże ilości danych wysyłany jest program MapReduce o rozmiarach kilku kilobajtów. Dzięki takiemu rozwiązaniu można zyskać na czasie, który traci się jeśli przesyła się dane. Ponadto MapReduce został tak opracowany by potrafił sobie radzić z awariami maszyn.

Godnym uwagi rozwiązaniem, w dziedzinie Big Data, było stworzenie przez firmę Amazon rozproszonego magazynu danych typu klucz-wartość o nazwie Dynamo.

W kolejnych latach projektami do pracy z Big Data były: HBase, MongoDB, Cassandra czy RabbitMQ. Oczywiście nie są to jedyne rozwiązania technologiczne, ich lista cały czas się zwiększa. Jednym z popularniejszych rozwiązań Big Data, według portalu Kdnuggets, który zajmuje się eksploracją danych i odkrywaniem wiedzy a w tym również Big Data, jest Hadoop.

3.3. Apache Hadoop

W ramach projektu Apache™ Hadoop® rozwijane jest oprogramowanie typu open-source, które umożliwia przetwarzanie rozproszone dużych zestawów danych w klastrach komputerów za pomocą prostych modeli programowania. Zostało ono zaprojektowane aby skalować z jednego serwera do tysiąca komputerów, oferując możliwość obliczeń i przechowywania danych. Biblioteka oprogramowania Apache Hadoop jest odporna na uszkodzenia, przeznaczona do wykrywania i obsługi uszkodzeń w warstwie aplikacji.

Projekt Apache Hadoop obejmuje następujące moduły:

- Hadoop Common – wspólne narzędzia, które wspierają inne moduły Hadoop.
- Hadoop Distributed File System (HDFS™) – rozproszony system plików, który zapewnia dostęp do wysokiej przepustowości danych aplikacji.
- Hadoop YARN – platforma programistyczna (framework) do planowania pracy i zarządzania zasobami klastra.
- Hadoop Map Reduce – system oparty na YARN przeznaczony do równoległego przetwarzania dużych zbiorów danych.

Podsumowanie

Coraz więcej ludzi i przedsiębiorstw korzysta z Internetu, smart – urządzeń, czujników, różnego rodzaju elektronicznych rejestratorów, produkując i gromadząc ogromne ilości danych zwane Big Data. Analiza bardzo dużych zbiorów danych (Big Data) może prowadzić do dokładniejszej i trafniejszej wiedzy. Reguły i zależności odkryte na dużych zbiorach danych będą zatem wiarygod-

niejszą wiedzą, niż gdyby były pozyskiwane z tradycyjnych hurtowni danych. W konsekwencji decyzje podejmowane na podstawie wiedzy uzyskanej z dużych zbiorów danych, mogą prowadzić do większej efektywności operacyjnej, redukcji kosztów i zmniejszenia ryzyka.

Jednak analiza Big Data jest trudna, czasem niemożliwa w tradycyjnych relacyjnych bazach danych i wymaga wdrożenia i stosowania nowych rozwiązań technologicznych takich jak, np. bazy NoSQL. Ponadto, należy zwrócić uwagę na fakt, iż bazy NoSQL w niektórych aspektach są bardziej skomplikowane niż tradycyjne relacyjne bazy danych, a w innych są od nich prostsze. Systemy te można co prawda skalować do coraz większych zbiorów danych, ale użycie ich wymaga zastosowania nowych technik, które wcale nie są uniwersalne. Dodatkowo wymagają one, na etapie początkowym, poniesienia kosztów związanych na przykład z przeszkoleniem pracowników lub wdrożeniem nowego oprogramowania.

Ustalenie zakresu wdrożenia, czy wybór narzędzi, powinny być zatem poprzedzone szczegółową analizą.

Przedsiębiorstwa chcą lepiej korzystać ze zgromadzonych przez siebie danych zwiększając listę celów i obszarów, w których mają być zastosowane narzędzia analityczne. Analiza Big Data obecnie jest uważana jako panaceum na poprawę zysków przedsiębiorstwa oraz osiągnięcie przez nią przewagi konkurencyjnej. Z badań przeprowadzonych przez firmy analityczno – doradcze, opisanych w tym artykule, wynika, że analiza Big Data i ich zastosowanie do procesów biznesowych rozwija się w tempie dynamicznym. Z roku na rok coraz więcej przedsiębiorstw deklaruje, że zainwestowało lub zamierza zainwestować w Big Data w ciągu najbliższych dwóch lat. Uważają one, że analizy danych niestrukturalnych będą kluczem do na przykład głębszego zrozumienia zachowań klienta. Jednak z badań wynika również, że wzrost inwestycji w technologie Big Data nie doprowadził do znaczącego wzrostu liczby organizacji zgłaszających wdrożenie projektów Big Data. Nadal wiele przedsiębiorstw jest na etapie tworzenia strategii rozwoju i projektów pilotażowych oraz eksperymentalnych.

Aby wdrożenie rozwiązań Big Data nie zakończyło się rozczarowaniem, ważne jest dobre zrozumienie, co analityka Big Data naprawdę oznacza oraz ustalenie jakie korzyści dla ich biznesu może przynieść i czy będą one opłacalne.

Literatura

- Busłowska Eugenia, Juźwiuk Łukasz. „Wprowadzenie do optymalnego wykorzystania MapReduce". *Logistyka 4/2014*. [dostęp online http://www.czasopismologistyka.pl] (dostęp: 14.03.2016).
- Cox Michael, Ellsworth David. 1997. *Managing Big Data for Scientific Visualization*. Siggraph [dostęp online
- www.dcs.ed.ac.uk/teaching/cs4/www/visualisation/SIGGRAPH/gigabyte_datasets2.pdf] (dostęp: 14.03.2016).
- Ledwoń Paweł. 2009. *Bezbolesne wprowadzenie do MapReduce*. Wrocławski Portal Informatyczny [dostęp online http://informatyka.wroc.pl] (dostęp: 14.03.2016).

- Laney Doug. 2001. *3D Data Management: Controlling Data Volume, Velocity, and Variety. META Group* (obecnie Gartner) [dostęp online http://blogs.gartner.com] (dostęp: 14.03.2016).
- Mayer-Schönberger Viktor. Cukier Kenneth. 2014. *BIG DATA Rewolucja, która zmieni nasze myślenie, pracę i życie*. Warszawa: MT Biznes.
- Marz Nathan, Warren James. 2016. *Big Data. Najlepsze praktyki budowy skalowalnych systemów obsługi danych w czasie rzeczywistym*. Gliwice: Helion.
- Racka Katarzyna. 2015. „Metody eksploracji danych i ich zastosowanie". *Zeszyty Naukowe PWSZ w Płocku. Nauki Ekonomiczne* tom XXI: 143.
- Sadalage Pramod J., Martin Fowler. 2014. *NoSQL. Kompendium wiedzy*, Gliwice: Helion.
- Tabakow Marta, Korczak Jerzy, Franczyk Bogdan. 2014. „Big Data – definicje, wyzwania i technologie informatyczne". *Informatyka Ekonomiczna Business Informatics* 1 (31). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu: 138.
- Venner Jason. 2009. *Pro Hadoop*. New York: Apress.
- White Tom.2016. *Hadoop. Kompletny przewodnik. Analiza i przechowywanie danych*. Gliwice: Helion.

Źródła internetowe

- 2015 State of Analytics. Salesforce. 2015
- [www 1] www.salesforce.com (dostęp: 14.03.2016)
- EMC Digital Universe with Research & Analysis by IDC *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*
- [www 2] https://www.emc.com (dostęp: 14.03.2016).
- Christof Strauch. NoSQL Databases.
- [www 3] http://www.christof-strauch.de/nosqldbs.pdf (dostęp: 14.03.2016).
- Marr Bernard. 2015. Big Data: The 5 Vs Everyone Must Know
- [www 4] https://www.linkedin.com (dostęp: 14.03.2016).
- Gartner Says It's Not Just About Big Data; It's What You Do With It: Welcome to the Algorithmic Economy
- [www 5] http://www.gartner.com (dostęp: 14.03.2016).
- Gartner Survey Reveals That 64 Percent of Organizations Have Invested or Plan to Invest in Big Data in 2013. STAMFORD. Connecticut. 2013
- [www 6] http://www.gartner.com (dostęp: 14.03.2016).
- Gartner Survey Reveals That 73 Percent of Organizations Have Invested or Plan to Invest in Big Data in the Next Two Years. STAMFORD. Connecticut . 2014
- [www 7] http://www.gartner.com (dostęp: 14.03.2016).
- IDC. Big Data Research
- [www 8] https://www.idc.com/prodserv/4Pillars/bigdata (dostęp: 14.03.2016).
- Jak dostęp do informacji warunkuje wyniki i rozwój firm – raport
- [www 9] http://it-manager.pl (dostęp: 14.03.2016).
- Neal Leavitt. 2010.Will NoSQL Databases Live Up to Their Promise?, IEEE Computer Society
- [www 10] http://www.leavcom.com (dostęp: 14.03.2016).
- Top Big Data Processing Frameworks
- [www 11] http://www.kdnuggets.com (dostęp: 14.03.2016).

- What Is Apache Hadoop
- [www 12] hadoop.apache.org/index.html (dostęp: 14.03.2016).

BIG DATA – MEANING, APPLICATIONS AND TECHNOLOGY SOLUTIONS

Summary:

Big Data technologies and their application to business processes is growing rapidly. Analytical and consulting enterprises specializing in issues of strategic use of IT technology indicate that the number of companies implementing or planning to implement technological solutions related to Big Data is increasing annually. A lot of companies believe that the analysis of unstructured data will be the key to a deeper understanding of customer behavior. They believe that the analyst is absolutely essential or very important to conduct the overall business strategy and improve operational results.

The purpose of the article is to define Big Data, explain what the unstructured data are and how to apply them. Furthermore, in the article I present the results of reports on the Big Data technologies implementation and discuss the associated technologies.

Keywords: Big Data, NoSQL, MapReduce, Hadoop.