

W świecie Wielkiego Brata

Od pewnego czasu hasło Big Data zyskuje coraz większą popularność nie tylko w kręgach związanych z IT, ale również w środowisku biznesowym. Z wykorzystywaniem wielkich zbiorów danych przedsiębiorstwa łączą bowiem wiele nadziei oraz planów na zyskanie przewagi konkurencyjnej na rynku.

Próby zdefiniowania, czym jest Big Data, przybierają różne kształty. Przyjmuje się, że jednym z wyróżników Big Data jest ilość zbieranych danych, których magazynowanie za pomocą tradycyjnych hurtowni danych jest praktycznie niemożliwe. Oczywiście postęp techniczny sprawia, że pojawiają się nowe możliwości zbierania, sortowania i przechowywania dużych wolumenów danych, których magazynowanie jeszcze kilka lat temu nie byłoby możliwe. Dlatego pojęcie Big Data w odniesieniu do rozmiaru baz danych jest płynne i wraz z rozwojem techniki jego zakres będzie ewoluował.

Czy Big Data to tylko rozmiar baz danych?

Według Gartnera Big Data można opisać za pomocą trzech wymiarów. Oprócz wielkości (volume) są to prędkość (velocity) oraz różnorodność (variety). Dane powinny być dostępne na bieżąco, bez opóźnień. Dodatkowo, powinny być stale aktualizowane w czasie rzeczywistym, tak aby jak najdokładniej odzwierciedlały rzeczywistość. Tradycyjne rozwiązania Business Intelligence pracujące na dużych zbiorach danych potrzebują czasu na przetworzenie informacji i prezentują wyniki z opóźnieniem. Pociąga to za sobą ryzyko przedawnienia, przeterminowania analizowanych danych jeszcze przed ukończeniem ich analizy. W odniesieniu do Big Data taka sytuacja nie powinna mieć miejsca. W Big Data analizowane informacje pochodzą z różnych źródeł, nie są ujednolicone, przybierają różne kształty. Mogą to być zdjęcia, nagrania, informacje pochodzące z telefonów komórkowych, dane udostępniane poprzez media społecznościowe oraz wszystkie inne nieuporządkowane informacje.

Do wymienionych powyżej wyróżników IBM dodał jeszcze jeden wymiar – prawdziwość (veracity). Dane znajdujące się w Big Data powinny być zgodne z rzeczywistością, dzięki czemu ich analiza będzie prowadziła do wiarygodnych wniosków.

Aby przechowywane dane miały jakąkolwiek wartość, powinny dawać możliwość skutecznego i szybkiego wyciągania konkretnych informacji pozwalających na podejmowanie decyzji. W przypadku dużych, rozproszonych danych taką możliwość daje Hadoop, narzędzie stworzone przez fundację Apache. To platforma open source służąca do magazynowania, przetwarzania i zarządzania dużymi zbiorami informacji. Koncepcja platformy zakłada podział takich zbiorów na mniejsze części, których przetwarzanie może następować równolegle na kilku serwerach. Pierwsza wersja Apache Hadoop została wydana w grudniu 2011 roku. Druga miała premierę w sierpniu ubiegłego roku. Użytkownikami systemu są m. in. Facebook, Adobe oraz Ebay.

Jaka czeka nas przyszłość?

Próba przewidzenia, w jaki sposób dążenie do gromadzenia danych, ich przetwarzania i analizy wpłynie na nasze życie, jest niezwykle trudna. Z jednej strony istnieje możliwość znacznego ułatwienia życia, z drugiej firmy wykorzystujące dane powinny pamiętać o nieprzekraczaniu delikatnej granicy, którą wyznacza prywatność ich klientów. Po jej przekroczeniu zaufanie społeczne do firmy zdecydowanie spadnie, co pociągnie za sobą odpływ klientów. Gartner w swoich raportach przewiduje, że do 2016 roku wykorzystywanie danych zdobytych legalnie, jednak powszechnie uważanych za prywatne spowoduje odpływ klientów u 25 proc. firm, które zdecydują się na użycie takich informacji. W tym momencie pojawia się pytanie: gdzie leży granica prywatności? W rzeczywistości po części jest wyznaczana przez obowiązujące prawo, a po części sami ją wyznaczamy. Dane dostępne firmom nie mogą być pozyskiwane nielegalnie, bez naszej wiedzy. Zakładając dowolne konto w sieci, np. na portalu społecznościowym, akceptujemy regulamin, który nie zawsze jest skrupulatnie czytany, udostępniamy i publikujemy informacje na swój temat. Pobierając aplikację na smartfona, zgadzamy się na dostęp do naszych danych zapisanych w telefonie.

Przedsiębiorstwa często zamiast tworzyć własne bazy danych, korzystają z dostępnych na rynku zbiorów danych osobowych. Oczywiście sprzedawane dane powinny być pozyskiwane z legalnych źródeł, z zapewnieniem wszystkich praw osób, których dotyczą, m.in. zgody na przekazanie informacji na ich temat innemu podmiotowi. Należy pamiętać, że o zgodną z prawem wymianę danych powinien zadbać zarówno sprzedający, jak i nabywca baz, ponieważ obie strony ponoszą odpowiedzialność za

prawidłowość wymiany. Prognozy przewidują, że w ciągu kilku najbliższych lat rynek wymiany danych będzie się zwiększał. Według Gartnera w 2016 roku 30 proc. wszystkich przedsiębiorstw będzie czerpało korzyści ze sprzedaży posiadanych przez siebie informacji o klientach bądź zdecyduje się na wymianę takich danych z innymi przedsiębiorcami.

Gartner przeprowadził badanie, według którego największym wyzwaniem związanym z Big Data stojącym przed przedsiębiorstwami jest określenie, w jaki sposób osiągnąć zysk z dostępu do Big Data. Analiza Big Data niesie ze sobą szansę usprawnienia pracy przedsiębiorstw i zminimalizowania ryzyka związanego z podejmowaniem decyzji. Z pewnością dostęp do dużych zbiorów danych może interesować firmy ubezpieczeniowe. Dzięki analizie danych medycznych system wykorzystywany przez takie przedsiębiorstwo miałby możliwość sprawdzenia stanu zdrowia klienta. Informacje o stanie zdrowia, trybie życia, wykonywanej pracy oraz zainteresowaniach, np. uprawianiu sportów ekstremalnych, częstych podróżach etc., pozwoliłyby precyzyjnie dobrać pakiet ubezpieczeń dla konkretnego klienta. Z podobnych założeń mogłyby wyjść banki. Dzięki uzyskaniu informacji udostępnianych przez klientów banki mogłyby szacować ich zdolność kredytową, proponować atrakcyjne formy lokowania wolnych środków bądź oferować usługi dodatkowe. Przykładem wykorzystania Big Data mogą też być działania rządowe podjęte w momencie uzyskania informacji o zagrożeniu zamachem terrorystycznym. W takim przypadku system analizujący Big Data mógłby w szybkim czasie zbadać portale społecznościowe, wysłane SMS-y, e-maile, określić miejsca odwiedzone w ostatnim czasie, wykonane podróże i w ten sposób wytypować określoną grupę podejrzanych osób.

Czy Big Data może poprawić standard życia?

Wyobraźmy sobie samochód, który analizując informacje z kamer drogowych, sam zaplanuje najszybszą trasę podróży, będzie ją zmieniał na bieżąco w zależności od natężenia uchu, omijając zakorkowane ulice. Dodatkowo, nie będzie wymagał kontroli kierowcy i sam dowiezie pasażerów do celu. Gdy poziom paliwa spadnie poniżej wyznaczonej przez nas granicy, sam pojedzie je uzupełnić, natomiast, gdy czujniki wykryją zużycie się poszczególnych podzespołów, informacja o konieczności wymiany pojawi się na naszym smartfonie. Oczywiście dodatkowo przedstawione zostaną orientacyjne koszty części zamiennych i ceny naprawy w różnych serwisach samochodowych. Nie jest to wizja oderwana od rzeczywistości. Już w tej chwili w ofercie producentów nawigacji samochodowych pojawia się opcja pracy online, pozwalająca na wybranie trasy przejazdu z uwzględnieniem aktualizowanej informacji o utrudnieniach na drodze. Postęp techniczny pozwala przypuszczać, że to, co obecnie można zaliczyć do science fiction, za kilka lat będzie czymś naturalnym.

Próbując odpowiedzieć na pytanie, czy Big Data jest przyszłością przedsiębiorstw, czy tylko chwilowym trendem, który zostanie zastąpiony przez inny modny temat, należy zastanowić się nad przeniesieniem dostępu do informacji na wyniki ekonomiczne firmy. Wykorzystanie pochodzących z różnych źródeł danych o klientach jest kuszące, jednak obecnie większość przedsiębiorstw nie jest w stanie przeprowadzić analizy posiadanych informacji, która przekładałaby się na poprawę wyników finansowych. Według Gartnera w 2015 roku ponad 90 proc. przedsiębiorców będzie postrzegać dostęp do informacji za atut pozwalający na osiągnięcie przewagi konkurencyjnej, jednak mniej niż 10 proc. będzie w stanie przenieść tę wartość na dane ekonomiczne. Zatem, czy dostęp do teoretycznie nieograniczonych informacji spowoduje wzrost wydajności? Bez odpowiedniego systemu do analiz wydaje się to niewykonalne.

Autor: Piotr Kurowski, specjalista ds. wsparcia sprzedaży, Comarch

Artykuł pochodzi z magazynu Nowoczesne Zarządzanie nr 1/2014: www.NZ.Comarch.pl