

Wielkie dane w Polsce i na świecie

ALEKSANDRA WOŹNIAK

Działające w Polsce przedsiębiorstwa najczęściej analizują dane pochodzące ze swoich systemów wewnętrznych. Z kolei firmy zagraniczne – dane pochodzące z aplikacji mobilnych i położenie geograficzne. W najbliższym czasie obie grupy zamierzają wdrażać rozwiązania analityczne, głównie w celu lepszego zrozumienia potrzeb konsumentów.

Z badania wynika, że najczęściej analizowanymi danymi w przedsiębiorstwach działających w Polsce są te pochodzące z systemów wewnętrznych firm, w dalszej kolejności – zachowania użytkowników na stronach internetowych firmy, a także dane pochodzące z aplikacji mobilnych. Najbardziej brane pod uwagę dane to z kolei wpisy na forach, dane generowane przez rozwiązania Internet Rzeczy (IoT) oraz korespondencja emailowa. Za granicą nacisk jest położony na analizę: danych pochodzących z aplikacji mobilnych, położenie geograficzne, dane pochodzące z systemów wewnętrznych firm, zachowania użytkowników na stronach internetowych firmy i Internet rzeczy/przedmiotów.

Rodzaje danych poddawane analizie w polskich przedsiębiorstwach różnią się oczywiście w zależności od sektora. Na przykład w branży telekomunikacyjnej najczęściej analizuje się położenie geograficzne, potem zaś dane pochodzące z systemów wewnętrznych i dane o zachowaniu użytkowników na stronach internetowych firm. W branży IT/E-commerce najczęściej analizowane są zachowania użytkowników na stronach internetowych firm, następnie dane pochodzące z systemów wewnętrznych i aplikacji mobilnych. W sektorze mediów i reklamy, a także w branży FMCG, najchętniej podejmowane są analizy danych pochodzących z aplikacji mobilnych oraz geolokalizacji. Sektory energetyki oraz bankowości i ubezpieczeń najczęściej biorą pod uwagę dane pochodzące z ich własnych systemów wewnętrznych.

Dla porównania, biorące udział w badaniu firmy telekomunikacyjne spoza Polski nacisk kładą przede wszystkim na analizę danych generowanych przez Internet rzeczy/przedmiotów, a dopiero w dalszej kolejności – dane geolokalizacyjne oraz dane pochodzące z aplikacji mobilnych. Z kolei przedsiębiorstwa wywodzące się z branży IT/E-commerce badają dane pochodzące z aplikacji mobilnych, dane dotyczące położenia geograficznego, a także zachowania użytkowników na stronach internetowych firmy.

Zastosowania prowadzonych analiz Big Data – aktualne...

Firmy działające w Polsce, podobnie jak i ich zagraniczne odpowiedniki, wykorzystują obecnie analitykę Big Data przede wszystkim w celu lepszego zrozumienia potrzeb klientów i budowania ich wartości (tj. zdyskontowanej różnicy pomiędzy przyniesionymi przez klientów przychodami, a sumą kosztów sprzedaży, obsługi i ich pozyskania).

Udzielane odpowiedzi przez polskich respondentów dotyczyły kolejno: lepszego zrozumienia potrzeb klientów, budowania wartości klienta, wykrywania nadużyć, optymalizacji sieci usług i punktów dostępu, a także zapobiegania odejściom i optymalizacji wyglądu stron internetowych oraz aplikacji.

Rozkład wskazań udzielonych przez zagraniczne przedsiębiorstwa był zbliżony do odpowiedzi grupy polskich uczestników badania. Obecnie analitykę Big Data zagraniczne firmy wykorzystują najczęściej do: budowania wartości klienta i lepszego zrozumienia jego potrzeb, zwiększania możliwości w zakresie wykrywania nadużyć, a tak-



że w celu optymalizowania aplikacji i zapobiegania odejściom klientów.

Udział poszczególnych analizowanych zagadnień różni się w zależności od sektorów gospodarki, w których działają badane przedsiębiorstwa. Przykładowo, obecnie w polskiej branży telekomunikacyjnej najczęściej wskazywanym celem analiz Big Data jest optymalizacja sieci usług i punktów dostępu, a także budowanie wartości klienta. Respondenci spoza Polski działający w tej branży deklarowali zaś przede wszystkim nastawienie na lepsze zrozumienie potrzeb klientów oraz budowanie ich wartości.

Te dwa rodzaje analiz były najczęściej wymieniane również przez polskich ankietowanych wywodzących się z sektorów bankowości, FMCG i energetyki, a także IT/E-commerce. W przypadku badanych z zagranicy, przedstawiciele firm IT/ E-commerce chętniej wybierali odpowiedzi dotyczące wykrywania nadużyć i budowania wartości klienta.

Polski sektor mediów i reklamy aktualnie koncentruje się na analizach związanych z budowaniem wartości klienta, zaś branża ubezpieczeniowa i sektor publiczny – na analizie wykrywania nadużyć.

... i planowane

Zasadnicze cele i rodzaje badań prowadzonych obecnie i w najbliższej przyszłości nie zmieniają się. Aktualnie przedsiębiorstwa działające w Polsce wykorzystują analitykę wielkich zbiorów danych, przede wszystkim w celu lepszego zrozumienia potrzeb klientów i budowania ich wartości. W najbliższej przyszłości zmieni się rozłożenie akcentów między tymi dwiema kwestiami. Najczęściej realizowane będą zatem analizy związane z budowaniem wartości klienta, potem zaś te mające na celu lepsze zrozumienie jego potrzeb. Analogiczna zmiana priorytetów nastąpi także wśród zagranicznych uczestników badania.

Choć rodzaje analiz, które polskie firmy będą przeprowadzać w najbliższym czasie, różnią się w zależności od sektorów gospodarki, to większość będzie się koncentrowała na kwestiach związanych z budowaniem wartości klienta. Inne priorytety będą dominowały w sektorach FMCG (zapobieganie odejściom) oraz administracji publicznej (optymalizacja aplikacji i wykrywanie nadużyć).

Technologie dobierane pod potrzeby

Rozwiązania Big Data cieszą się dużą popularnością w obszarze marketingu i sprzedaży (blisko 40% odpowiedzi ankietowanych firm w Polsce i za granicą). Rzadziej są natomiast wykorzystywane w działach finansów i controllingu (20% - 30% wskazań) oraz bezpieczeństwa (20%).

Wyniki badania wskazują, że do analizy rozproszonych danych o konsumencie przedsiębiorstwa (tak polskie, jak i zagraniczne) używają zarówno tradycyjnych rozwiązań Business Intelligence (najczęściej firm: Oracle, Microsoft, SAS), jak i różnych technologii Big Data (Hadoop, Cassandra, MongoDB). Wykorzystywane są przy tym zarówno rozwiązania komercyjne, jak i open source (np. R, Python).

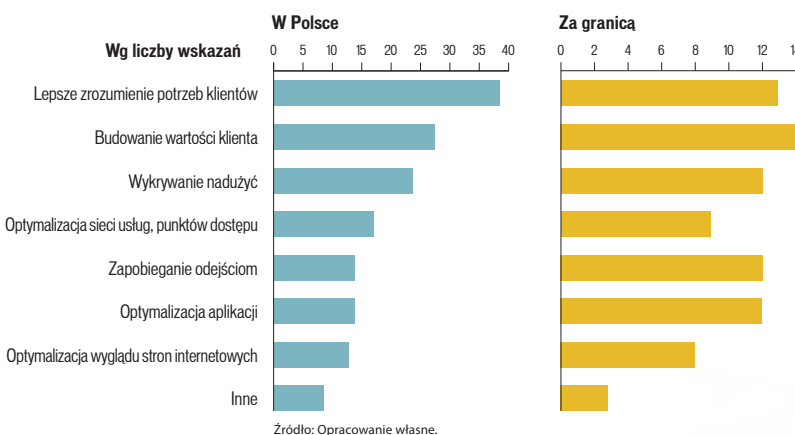
Tradycyjne komercyjne rozwiązania są z reguły stosowane do procesów pozyskiwania danych ETL (Extract, Transform and Load), do platform Business Intelligence, dużych hurtowni danych oraz do zarządzania danymi.

Narzędzi Big Data - bazujących przeważnie na otwartym oprogramowaniu - używa się na ogół do groma-

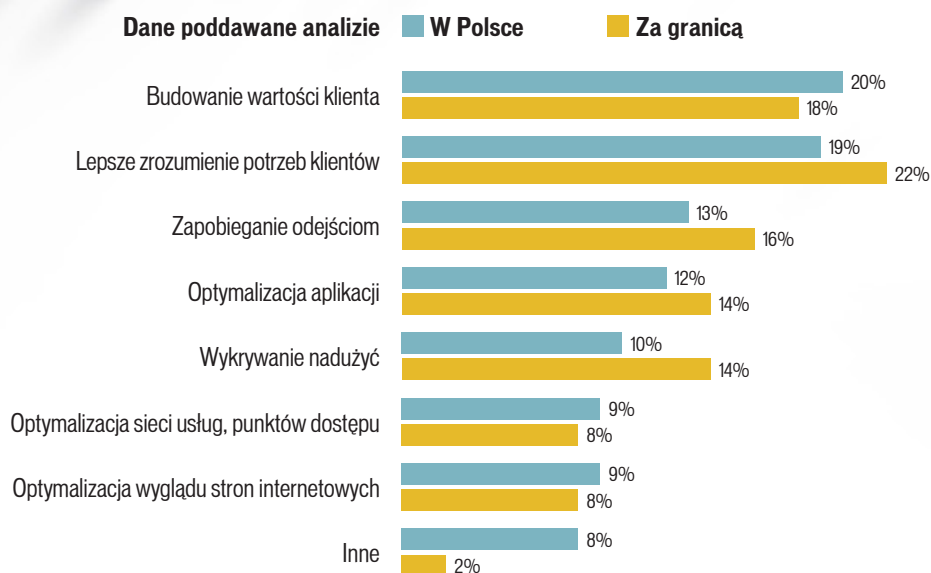
▶▶ Dane (ustrukturyzowane i nieustrukturyzowane) poddawane analizie w firmach w Polsce i za granicą (procentowy udział poszczególnych odpowiedzi w ogóle udzielonych przez daną grupę respondentów)



▶▶ Rodzaje analiz prowadzone obecnie w polskich i zagranicznych przedsiębiorstwach.



▶ Rodzaje analiz w obszarze Big Data, jakie firmy w Polsce i za granicą zamierzają wdrożyć w najbliższym czasie (procentowy udział poszczególnych odpowiedzi w ogóle udzielonych przez daną grupę respondentów)



Źródło: Opracowanie własne.

Próba wysokiej jakości

Badanie towarzyszące pracy dyplomowej Aleksandry Woźniak pt. „Przegląd narzędzi Big Data do analizy rozproszonych danych o konsumencie” przeprowadzone zostało w II kwartale 2015 roku na próbie 55 polskich firm oraz 19 respondentów zagranicznych. Badanie ankietowe miało na celu zgromadzenie informacji na temat używanych obecnie narzędzi Big Data do analizy danych rozproszonych o konsumencie w polskich i zagranicznych przedsiębiorstwach i instytucjach, sprawdzeniu, jakie dane są obecnie poddawane analizie oraz jakie analizy firmy zamierzają przeprowadzić w tym obszarze w najbliższym czasie. W badaniu udział wzięły m. in. takie przedsiębiorstwa i instytucje, jak: PKN Orlen, PZU, T-Mobile, Polkomtel, Play, Wirtualna Polska, Onet, Interia, Allegro, Alior Bank, Procter & Gamble, Ministerstwo Administracji i Cyfryzacji, Facebook, Twitter, LinkedIn, eBay, Amazon.com, Alibaba, Apple, Motorola, Vodafone oraz Tesco.

dzenia wielkich wolumenów danych i baz typu NoSQL/NewSQL, ich analizy, a także na rzecz rozwiązań Business Intelligence i wizualizacji danych.

Największe przedsiębiorstwa (np. Facebook czy Alibaba) często rozwijają własne rozwiązania klasy Big Data, bazując przy tym przeważnie na dostępnych narzędziach typu open source. Tego typu trend jest coraz wyraźniejszy – wspiera go dostępność interfejsów API oraz wsparcie zaangażowanych społeczności użytkowników i deweloperów.

Przeważająca większość ankietowanych osób (76% polskich i 79% zagranicznych respondentów) jest zadowolona z obecnie stosowanych narzędzi i oprogramowania do analiz Big Data. Jednocześnie większość respondentów biorących udział w badaniu – zarówno z Polski (75%), jak i z zagranicy (95%) – zadeklarowała, że stosowane w ich organizacjach rozwiązania Big Data zmieniały się w ciągu ostatnich kilku lat. W tym czasie pojawiały się bowiem inne, lepsze (57% odpowiedzi polskich respondentów i 56% zagranicznych) lub tańsze (odpowiednio 37% i 28%) narzędzia tego typu. Aktualnie ponad połowa badanych (59% firm działających w Polsce i 47% za granicą) nie planuje zmiany narzędzi do analityki wielkich wolumenów danych stosowanych w swoich organizacjach.

Trendy Big Data zmieniające biznes

Większość ankietowanych osób dostrzega trendy w obszarze Big Data. Kilka ważniejszych, które zostały wymienione w badaniu to np.: analityka Big Data w chmurze, coraz większa popularność baz NoSQL, SQL na Hadoop, analityka w czasie rzeczywistym i analityka in-memory (w pamięci operacyjnej serwera), „jezioro danych”, bazy grafowe, uczenie maszynowe, „głębokie uczenie” (deep learning) czy Internet rzeczy/przedmiotów.

Chociaż Hadoop został stworzony z myślą o pracy na fizycznych maszynach, obecnie dostępnych jest coraz więcej technologii umożliwiających przetwarzanie danych w chmurze. Wydaje się, że w przyszłości rozwiązania Big Data będą hostowane hybrydowo – lokalnie oraz w chmurze, na co wskazywali ankietowani w badaniu. Apache Spark jest jednym z narzędzi open source umożliwiającym błyskawiczną analizę danych w czasie rzeczywistym, w technologii in-memory. Coraz częściej firmy w celu zwiększenia elastyczności środowisk analitycznych rozwijają Hadoopa o kolejne rozwiązania, np. uruchamiając platformę Apache Spark i bazy NoSQL.

Technologia „głębokie uczenie” (deep learning – oparty na sieciach neuronowych zestaw technik uczenia się maszyn) i uczenie maszynowe (machine learning), będące wciąż na etapie rozwoju, już teraz wykazują ogromny potencjał biznesowy, zarówno w polskich, jak i zagranicznych przedsiębiorstwach.

Respondenci obu grup zauważają, że coraz częściej na znaczeniu zyskuje też wykorzystanie reprezentacji grafowej relacji między danymi w eksploracji danych. Ważną rolę w rozwoju Big Data zaczął odgrywać również Internet Rzeczy. Zagraniczne firmy (np. telekomunikacyjne) już teraz analizują dane generowane przez IoT.

Więcej na temat wyników badania dowiedzą się Państwo podczas konferencji Computerworld „Data+ Big Data & Business Analytics”, która odbędzie się w dniach 26-27 listopada 2015 w Warszawie.

Podczas konferencji swoimi pomysłami i rekomendacjami w obszarze analizy danych podzielą się m. in. menedżerowie z General Motors, BP, mBanku, Allegro, ITAKI i wielu innych firm. Szczegóły na dataplus.computerworld.pl.