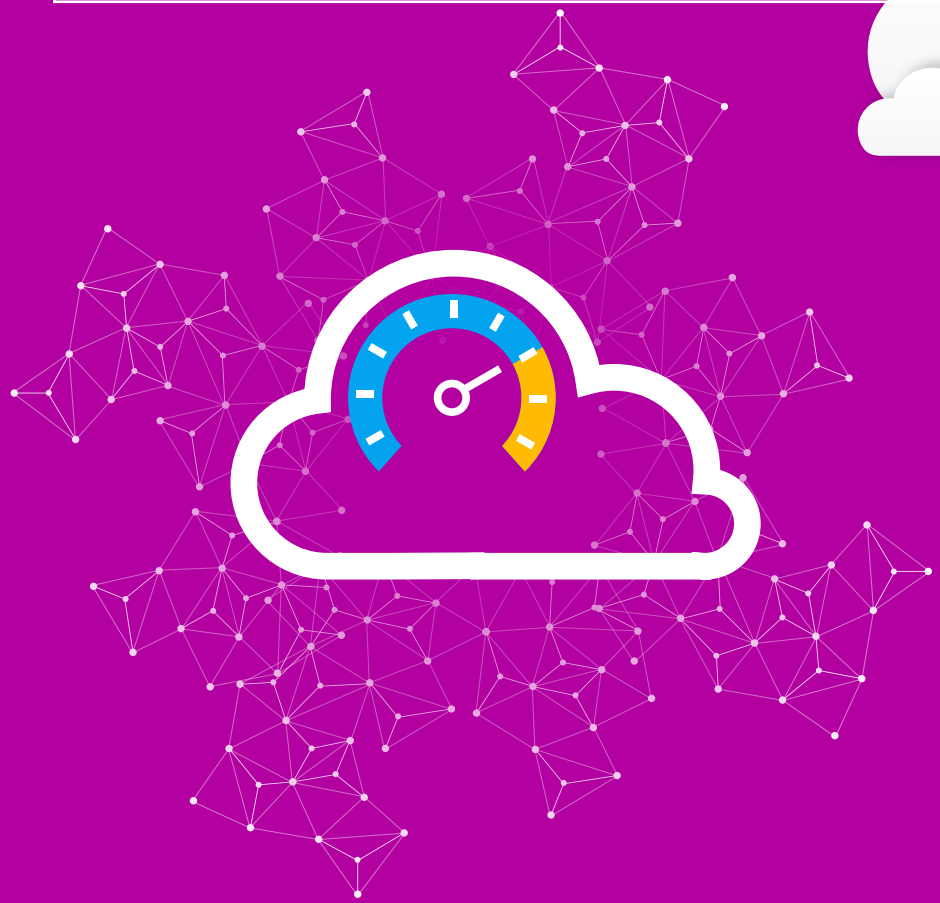
RAPORT



Narzędzia analityczne

w chmurze publicznej

ITWIZ
TO CO ISTOTNE W IT

Microsoft

ADAM JADCZAK

JAK OBNIŻYĆ KOSZTY

analityki Big Data dzięki chmurze publicznej

Firma Rockwell Automation potrzebowała lepszych, szybszych i tańszych sposobów zarządzania i analizowania coraz większej ilości danych. Budowa własnej infrastruktury wiązała się z dużymi kosztami. Firma podjęła więc strategiczną decyzję, aby skorzystać z rozwiązań oferowanych w chmurze Microsoft Azure. Takie podejście sprzyja szybszemu wprowadzaniu rozwiązań analitycznych, a ponadto znacząco obniża koszty inwestycyjne w porównaniu z tymi ponoszonymi na własną infrastrukturę.

Rockwell Automation to jedna z największych na świecie firm wg przychodów w branży automatyki przemysłowej. W latach 2010–2013, czyli w najgorszych warunkach gospodarczych od pokoleń, firma pozyskiwała nowych klientów, a jej zysk netto wzrósł o 240%. Wszystko dzięki ofercie, bazującej m.in. na analizie dużych zbiorów danych.

KONIECZNOŚĆ WYKORZYSTANIA DODATKOWYCH ZASOBÓW W CHMURZE

Więcej klientów to oczywiście coś pozytywnego. Ale to również wyzwanie, ponieważ związane jest z coraz większą liczbą systemów sterowania, które trzeba monitorować, analizować i zapewniać im wsparcie. Tymczasem pojedynczy klient może mieć tysiące czujników i programowalnych sterowników logicznych, a każdy dostarcza dziesiątek tysięcy punktów danych na sekundę. Są to pomiary temperatury, ciśnienia, wibracji itp.

Trzy lata temu firma dotarła do kresu możliwości własnych centrów danych. Zdecydowano wówczas o wykorzystaniu dodatkowych zasobów w chmurze Microsoft Azure. To pomogło, ale firma potrzebowała rozwiązania, które wspierałoby analizy dużych zbiorów danych – Big Data. *„Generujemy zbyt dużo danych dla naszej hurtowni danych. Potrzebowaliśmy rozwiązania, aby obsłużyć rosnącą ilość danych"* – mówi Juan Asenjo, Lead Architect for Remote Monitoring w Rockwell Automation.

Dzięki usłudze Azure Data Factory programiści Rockwell Automation mogą tworzyć – oparte na danych – przepływy zadań, które łączą, agregują i przekształcają dane pochodzące ze źródeł lokalnych, chmurowych, platformy Hadoop i usług internetowych. Mogą też tworzyć złożone scenariusze przetwarzania danych za pomocą niewielkiej ilości kodu. Usługa Azure Data Factory zawiera narzędzia umożliwiające łatwy, całościowy monitoring i zarządzanie tymi źródłami, co zmniejsza czas i koszty tradycyjnie związane z ich obsługą.

Budowanie tego rozwiązania samodzielnie wymagałoby integracji z istniejącymi systemami firmy, a także integracji systemów lokalnych z chmurą i przetwarzania wsadowego w trybie ciągłego streamingu. Cała ta integracja wymusi-

POWRÓT

łaby z kolei bardzo niepraktyczne i kosztowne projekty programistyczne. Przedstawicielom Rockwell Automation brakowało jednocześnie jednolitego, holistycznego widoku, którego firma ta potrzebowała do usprawniania zarządzania urządzeniami klientów oraz ich konserwacją.

ZARZĄDZAĆ DANYMI SZYBCIEJ I W SPOSÓB BARDZIEJ INNOWACYJNY

Konieczność zarządzania większą ilością danych stanowiła jedynie część wyzwania stojącego przed Rockwell Automation. Istniała jeszcze potrzeba zarządzania danymi szybciej i w sposób bardziej innowacyjny. Przykładowo, dla klientów z branży naftowej firma potrzebowała oprogramowania nie tylko do monitorowania złożonego łańcucha dostaw obejmującego platformy morskie, odległe rafinerie i stacje paliw. Firma ta potrzebowała także rozwiązania, które umożliwiłoby klientom *„poprawę wydajności, wprowadzanie innowacji i utrzymywanie strumienia paliwa w całym łańcuchu dostaw"*.

„To, o czym mówimy, jest stopniem współpracy, o którym jeszcze nie słyszano w przemyśle naftowym i gazowym. Aby to zrobić, potrzebujemy jednak rozwiązania umożliwiającego nieprzerwany przepływ dużych ilości danych" – twierdzi Doug Weber, Business Manager, Remote Application Monitoring w Rockwell Automation.

Aby przekazywać użyteczne analizy pracownikom w terenie – zarówno własnym technikom wsparcia, jak i klientom – Rockwell Automation opracowuje bardzo intuicyjne, interaktywne pulpity nawigacyjne z wykorzystaniem Microsoft Power BI. Natomiast do zarządzania obciążeniami dużą ilością danych, firma korzysta z kilku baz danych Azure SQL i stosuje różne poziomy usług. Dzięki temu możliwe jest spełnienie różnych wymagań dotyczących przetwarzania danych klientów.

PAKIET USŁUG CHMUROWYCH WSPIERAJĄCY DATA GOVERNANCE

Aby umożliwić przepływ dużych ilości danych, Rockwell Automation opracowała – i wciąż udoskonala – strategię Data Governance opartą na technologiach Microsoftu. By zarządzać i analizować ogromne ilości danych, firma korzysta z usług **Azure HDInsight**, zarządzanej usługi Hadoop działającej w chmurze. Aby przenieść wiedzę czerpaną z danych do dziedziny analizy predykcyjnej, Rockwell Automation rozwija możliwości w dziedzinie planowania wydajności i konserwacji predykcyjnej za pomocą **Azure Machine Learning**. Zarządzając zaś złożonymi interakcjami danych w wielu systemach, firma korzysta z usługi **Azure Data Factory**.

W celu przekazywania użytecznej analizy pracownikom w terenie – zarówno własnym technikom wsparcia, jak i klientom — Rockwell Automation opracowuje bardzo intuicyjne, interaktywne pulpity nawigacyjne z wykorzystaniem **Microsoft Power BI**. Natomiast do zarządzania obciążeniami dużą ilością danych, firma korzysta z kilku baz danych **Azure SQL** i stosuje różne poziomy usług. Dzięki temu możliwe jest spełnienie różnych wymagań dotyczących przetwarzania danych klientów. Firma korzysta też z funkcji georedundancji i przywracania danych dla określonego punktu w czasie *(Point-in-Time Restore)*. Ma to zapewnić wymagany poziom niezawodności i łatwości zarządzania.

UCZENIE SIĘ LEPSZEJ WYDAJNOŚCI DZIAŁANIA URZĄDZEŃ IOT

Strategia Data Governance firmy – oparta na usługach Microsoft Azure – zaprezentowana została na bazie projektu pilotażowego (Proof-of-Concept). Wykorzystano w nim analizę danych, w celu zwiększenia efektywności energetycznej. Projekt koncentrował się na optymalizacji wykorzystania wielu pomp zasilających kolektor silnika. Taka konfiguracja jest typowa dla miejskich zakładów wodociągowych, ale także innych zakładów z wieloma kotłami energetycznymi. Tradycyjnie kierownicy zakładów decydują o tym, które pompy są uruchamiane i kiedy je uruchamiać na podstawie własnych algorytmów i orientacyjnych reguł, które mogą być dalekie od optymalnych.

W projekcie Proof-of-Concept, firma Rockwell Automation wykorzystała narzędzia Azure Machine Learning, aby zrozumieć, które pompy i z jaką prędkością działają. Dzięki temu jej przedstawiciele chcieli zmaksymalizować zaopatrzenie w wodę, minimalizując jednocześnie zużycie energii. Wynik wykazał 1-proc. wzrost efektywności energetycznej. Niby niewiele, ale przekładało się to na znaczne oszczędności, szczególnie w przypadku dużych systemów pomp. *„Takie rozwiązanie, oparte na Azure Machine Learning, może zaoszczędzić dużym systemom miejskim kilka milionów dolarów rocznie”* – skomentował Doug Weber.

FABRYKA DANYCH W CHMURZE

W innym rozwiązaniu Rockwell Automation wykorzystała kilka technologii Azure do rozwiązywania problemu integracji złożonych systemów danych. Wykorzystuje też hostowaną w chmurze usługę Azure Data Factory do komponowania, koordynowania i zarządzania źródłami danych o wysokiej dostępności i niezawodności. Składają się one ze zbiorów danych i usług orkiestracji, w celu lepszego ich przetwarzania.

Dzięki usłudze Azure Data Factory programiści Rockwell Automation mogą tworzyć – oparte na danych – przepływy zadań, które łączą, agregują i przekształcają dane pochodzące ze źródeł lokalnych, chmurowych, platformy Hadoop i usług internetowych. Mogą też tworzyć złożone scenariusze przetwarzania danych za pomocą niewielkiej ilości kodu. Usługa Azure Data Factory zawiera narzędzia umożliwiające łatwy, całościowy monitoring i zarządzanie tymi źródłami, co zmniejsza czas i koszty tradycyjnie związane z ich obsługą.

Właśnie takiego rozwiązania poszukiwała firma Rockwell Automation, aby wspierać użycie platformy Hadoop do monitorowania i analizy wydajności systemów produkcyjnych i przetwarzania swoich klientów. Testowała też usługi Azure Data Factory jeszcze przed ich oficjalną premierą. Juan Asenjo postanowił skorzystać z usługi Azure Data Factory w projekcie pilotażowym, by zaspo-

POWRÓT

koić pilną potrzebę szybkiego, niezawodnego sposobu interpolowania czasoszeregu danych z czujników IoT przed przeprowadzeniem ich analizy. Firma prowadziła już interpolację danych, jednak na stosunkowo ograniczoną skalę i przy użyciu relacyjnych baz danych oraz czasochłonnego procesu tworzenia własnego kodu.

Juan Asenjo miał nadzieję, że dzięki zastosowaniu usługi Azure Data Factory będą mogli łatwiej interpolować więcej danych, zmniejszając ilość surowych danych, które musieli zbierać z czujników, przekazując je do chmury. Celem było przyspieszenie procesu i obniżenie jego kosztu. Programiści użyli usługi Azure Data Factory, aby utworzyć nowe źródło danych zaledwie za pomocą kilku kliknięć. Następnie połączyli je z bazą Azure SQL Database, pamięcią Blob Storage i usługą HDInsight. Mieli również możliwość zautomatyzowania koordynacji wykonania skryptów Pig i Hive w usługach Hadoop/HDInsight, zamiast uruchamiać te skrypty ręcznie. Nie musieli też martwić się o zarządzanie klastrem.

W ramach testu surowe dane z czujników były przechowywane w usłudze Blob Storage i przetwarzane w HDInsight. Kompletny zestaw interpolowanych danych został zwrócony do usługi Blob Storage, a świeże – lub najbardziej aktualne dane – do bazy danych Azure SQL, w której warstwa analityczna przedstawia użytkownikom

dane. Rezultatem było przekształcenie surowych danych w aktywa informacyjne, które mogą być wykorzystane przez Azure Machine Learning i Power BI, w celu uzyskania wiarygodnych wniosków.

Rockwell Automation określiła wykorzystanie oferowanych przez Microsoft technologii przetwarzania danych „dynamicznym sukcesem", który może mieć daleko idący wpływ na jej działalność. Dzięki tym technologiom czas wprowadzania produktu jest krótszy, źródło danych ma pełną widoczność, koszty są niższe, a jednocześnie udostępniane są innowacyjne sposoby obsługi klientów.

REDUKCJA CZASU TWORZENIA NOWYCH ROZWIĄZAŃ O 80%

Juan Asenjo ze współpracownikami opracował i wdrożył rozwiązanie integracji danych z platformy Hadoop w ciągu zaledwie 2 tygodni. Trwałoby to 5 razy dłużej, gdyby nie wykorzystano usługi Azure Data Factory. Uzyskali oni 80-proc. oszczędność czasu i kosztów. Nie trzeba było bowiem instalować i konfigurować systemu harmonogramów, budować szablonów do koordynacji konwersji danych ani ręcznie tworzyć skryptów Pig & Hive. *„Oszczędziliśmy kilka tygodni czasu programowania, dzięki automatyzacji w usłudze Azure Data Factory, i mogliśmy skoncentrować się na konkretnych skryptach ETL, które zoptymalizowały rozwiązanie"* – wspomina Juan Asenjo.

Juan Asenjo i Doug Weber oceniają krótszy czas wprowadzenia nowego rozwiązania dla klientów jako jedną z zalet usługi Azure Machine Learning. *„Dzięki Azure Machine Learning mo-*

POWRÓT

W projekcie Proof-of-Concept, firma Rockwell Automation wykorzystała narzędzia Azure Machine Learning, aby zrozumieć, które pompy i z jaką prędkością działają. Dzięki temu jej przedstawiciele chcieli zmaksymalizować zaopatrzenie w wodę, minimalizując jednocześnie zużycie energii. Wynik wykazał 1-proc. wzrost efektywności energetycznej. Niby niewiele, ale przekładało się to na znaczne oszczędności, szczególnie w przypadku dużych systemów pomp.

żemy przenieść analitykę do zupełnie nowego poziomu, poziomu predykcyjnej, przydatnej analizy" – mówi Doug Weber. *„Ale możemy zrobić jeszcze więcej. Czas na opracowanie rozwiązania przy użyciu Azure Machine Learning jest znacznie krótszy w porównaniu z pisaniem własnego kodu. To umożliwia wprowadzenie tego rozwiązania u większej liczby klientów i budowanie bardziej rozbudowanego rozwiązania, niż byłoby możliwe w innych przypadkach"* – dodaje.

PEŁNY DOSTĘP DO ŹRÓDEŁ DANYCH TO REDUKCJA KOSZTÓW NAWET O 90%

W środowisku produkcyjnym rozwiązanie będzie generować stałe oszczędności dzięki zmniejszeniu kosztów utrzymania i wsparcia. Według Johna Dycka, Director of Software Business Development w Rockwell Automation, oszczędności mogą wynieść nawet 90%.

„Korzystając z Azure Data Factory, otrzymujemy dostęp do wszystkich źródeł naszych danych. Oznacza to, że możemy szybciej i łatwiej je obsługiwać" – mówi.

„Planujemy wdrożenie usługi Microsoft Azure we wszystkich aplikacjach do monitorowania, ponieważ umożliwiają one bardziej efektywne organizowanie danych w coraz bardziej złożonych środowiskach" – stwierdza John Dyck. *„Możemy zdalnie obsługiwać klientów – niezależnie od miejsca, w jakim znajdują się na świecie – na poziomie efektywności, który nie był dotychczas możliwy. Przykładowo, wcześniej koordynacja ruchu danych między krajami była bardzo trudna. Nowe technologie łagodzą tę złożoność"* – dodaje.

SZYBSZE I SZERSZE WYKORZYSTANIE ROZWIĄZAŃ BIZNESOWYCH DZIĘKI BIG DATA

Możliwość skalowania danych na platformie Hadoop sprawia, że jest to idealne rozwiązanie do łatwiejszej i szybszej pracy z większymi ilościami danych. Dlatego, według Johna Dycka, Rockwell Automation prawdopodobnie przyspieszy wdrożenie Hadoop. Zamiast przetwarzać zbiory danych tylko na żądanie, firma będzie mogła dzięki temu wykonać interpolację danych, które są gotowe do wykorzystania przez użytkowników. To sprawia, że czas przetwarzania raportów skraca się z kilku minut do kilku sekund.

Rockwell Automation będzie także korzystać z usług Azure HDInsight i Azure Data Factory, aby pomóc klientom spełniać wymogi dotyczące sprawozdawczości amerykańskiej agencji ochrony środowiska, które nakładają na producentów obowiązek przechowywania danych przez 5 lat. Rozwiązania te mogą posłużyć do przechowania danych dotyczących zarządzania jakością tak długo, nawet przez 10 lat, jeśli zażyczą sobie tego klienci.

„Dzięki zastosowaniu Big Data różnica w ilości wykorzystywanych danych i sposobie ich użycia będzie ogromna. To kolejny aspekt ogromnego wpływu technologii przetwarzania danych oferowanych przez Microsoft na naszą działalność" – podsumowuje John Dyck.

PAWEŁ **OKOPIEŃ**

ANALITYKA W CHMURZE:

o się opłaca

Chmura Microsoft Azure daje wiele możliwości, a wśród nich te związane z zaawansowaną analityką i Business Intelligence. Dzięki dedykowanym narzędziom i odpowiednim planom taryfowym, firmy otrzymują rozwiązanie szyte na miarę i łatwo skalowalne.

Narzędzia Business Intelligence to wiele aplikacji, które pozwalają zebrać i przetworzyć olbrzymią liczbę danych, także tych nieustrukturyzowanych z wewnętrznych i zewnętrznych systemów IT. To dokumenty firmowe, dane zdrowotne, pliki, e-maile, wyniki finansowe, zdjęcia, filmy wideo, nagrania z call center i dużo więcej. Narzędzia te ułatwiają analitykę danych firmowych potrzebnych do przygotowania raportów, dashboardów czy wizualizacji danych. W rezultacie pracownicy i menedżerowie mają ułatwioną pracę i znacznie szybciej otrzymują wyniki potrzebnych im analiz. BI w chmurze ułatwia podejmowanie decyzji biznesowych dzięki dostarczeniu kompleksowych informacji, poprawienie efektywności operacyjnej, łatwiejszej identyfikacji trendów rynkowych i nowych szans.

Przykładowo, Microsoft oferuje narzędzia do analizy i wizualizacji danych Power BI jako dodatkową usługę dla Office 365, ale przede wszystkim jest to usługa dostępna w chmurze Azure. W tej wersji znacznie więcej jest dostępnych funkcjonalności. Power BI to cały zbiór narzędzi Business Intelligence. Podstawowa wersja dla Office 365 jest bezpłatna dla biznesowych użytkowników tej usługi. Wersja pro, przy umowie na rok, kosztuje 8,40 euro. Podobnie jest z usługą Power BI embedded dla chmury Azure. 100 pierwszych sesji jest bezpłatnych. Natomiast później płacimy wyłącznie za te sesje, które wykonujemy. Kolejny próg to 4,22 euro miesięcznie.

Ministerstwo Finansów, zamiast inwestować ponad 20 mln zł we własną infrastrukturę na potrzeby zbierania Jednolitego Pliku Kontrolnego, wykupiło usługi w chmurze Azure. Szacunkowy roczny koszt ich wykorzystania – dotyczy to nie tylko projektu JPK, lecz także kilku innych, w tym platformy e-learningowej – **to ok. 1 mln zł rocznie.** Przedstawiciele resortu finansów wspominali także o znacznie szybszej budowie potrzebnych rozwiązań dzięki zastosowaniu gotowych narzędzi na platformie Azure.

Na dedykowanej stronie Microsoft zamieszcza kalkulator, który pozwala obliczyć koszty utrzymania chmury w naszej firmie (goo.gl/1pI9aN). Co ważne, w przypadku środowiska Azure płacimy wyłącznie za infrastrukturę i narzędzia, które rzeczywiście w danym momencie wykorzystujemy. Kiedy nie korzystamy aktywnie z systemu, płacimy mniej albo wcale. Natomiast w każdej chwili moc może zostać zwiększona do naszych potrzeb, gdy intensywnie będziemy stosować te narzędzia. Przykładowo, za godzinę wykorzystania platformy Hadoop (usługa) HDInsight na podstawowej maszynie z 4 rdzeniami, 7 GB RAM, 256 GB dysku zapłacimy 0,27 euro.

Większość z narzędzi Microsoftu można bezpłatnie przetestować, firma oferuje też bogate dokumentacje, bezpłatne szkolenia, również w formie wideo, a także odpowiednie wsparcie.

Azure oferuje rozwiązania z obszaru sztucznej inteligencji AI oraz internetu rzeczy (IoT Hub, Event Hub). Platforma Azure pozwala rozłożyć koszty wykorzystywanych narzędzi analitycznych. Nie musimy inwestować jednorazowo sporych sum w licencje na oprogramowanie i sprzęt.

Przykładowo, Ministerstwo Finansów, zamiast inwestować ponad 20 mln zł we własną infrastrukturę na potrzeby zbierania Jednolitego Pliku Kontrolnego, wykupiło usługi w chmurze Azure. Szacunkowy roczny koszt ich wykorzystania – dotyczy to nie tylko projektu JPK, lecz także kilku innych, w tym platformy e-learningowej – to ok. 1 mln zł rocznie. Przedstawiciele resortu finansów wspominali także o znacznie szybszej budowie potrzebnych rozwiązań dzięki wykorzystaniu gotowych narzędzi na platformie Azure.

Rozwiązania cloud computing zapewniają też ciągłe wsparcie techniczne. Jednocześnie Microsoft stale udoskonala platformę Azure, dodaje nowe elementy. Firma stara się cały czas zwiększać świadomość użytkowników, stąd liczne programy edukacyjne, ułatwiające korzystanie z dostępnych narzędzi. Atutem jest również wspomniana skalowalność.

Dzięki narzędziom analitycznym w chmurze organizacje otrzymują inteligentne, interaktywne wizualizacje danych. Dostarczane są one w ciągu kilku minut, a nie godzin czy dni, w zależności od wydajności naszej platformy. Oprócz szerokiego zakresu gotowych form wizualizacji danych można wszystko spersonalizować pod potrzeby własnej firmy. Z narzędzi analitycznych Microsoft Azure korzystają takie firmy, jak Rolls-Royce, Coca-Cola czy Infosys.

Power BI to ostatnia z łańcucha usług zaawansowanej analityki na platformie Microsoft Azure, który tworzą:

- Azure Analysis Services – modelowanie danych w chmurze klasy korporacyjnej,
- Azure Bot – usługa botów,
- Azure Data Factory – pozwalająca na organizowanie i przenoszenie danych,
- Azure Data Lake Store – repozytorium do przechowywania obciążeń wynikających z danych Big Data,
- Azure Data Lake Analytics – rozproszona usługa analizy Big Data,
- Azure Data Catalog,
- Azure Stream Analytics,
- Azure Machine Learning,
- Azure HDInsight – aprowizacja klastrów Hadoop, Spark, R Server, Hbase i Storm w chmurze,
- Azure SQL DB,
- Azure SQL DW
- Azure Stream Analytics,
- Azure Cognitive Services.

PAWEŁ OKOPIEŃ

W JAKI SPOSÓB

przedsiębiorstwa wykorzystują analitykę w chmurze

Według najnowszych badań, już 43% firm korzysta z rozwiązań Business Intelligence as a Service. Firmy te uznają, że analityka i Business Intelligence w chmurze zaczynają być jednym ze współczesnych źródeł przewagi konkurencyjnej. W jaki sposób korzystają z nich organizacje z całego świata?

Na początku roku BARC (Business Applications Research Center) przedstawił raport o wykorzystaniu Business Intelligence oraz Data Management w chmurze. Jak wynika z raportu, na przestrzeni ostatnich lat nastąpił znaczący wzrost zainteresowania takimi narzędziami. W roku 2016 43% badanych przedsiębiorstw wykorzystywało analityczne rozwiązania oparte na chmurze. W roku 2013 było to zaledwie 29%. Ponadto aż 78% ankietowanych planuje rozszerzenie projektów związanych z użyciem narzędzi analitycznych o rozwiązania cloud computing. W jaki sposób wykorzystują analitykę w chmurze organizacje ankietowane przez BARC?

RAPORTOWANIE TO GŁÓWNA FUNKCJONALNOŚĆ BI W CHMURZE

Narzędzia w chmurze w największym stopniu stosowane są przez tzw. Power Userów – 54% w wysokim stopniu z nich korzysta, a 32% w średnim. Wśród tradycyjnych użytkowników intensywnie z narzędzi BI w chmurze korzysta 20% ankietowanych, a 46% w średnim stopniu. Aż 76% ankietowanych na pytanie, jaki jest pierwszorzędny sposób wykorzystania cloud Business Intelligence, odpowiedziało: przetwarzanie raportów oraz tzw. Dashboard Consumption. Pod tym drugim kryje się przede wszystkim odczyt wygenerowanych raportów z paneli samoobsługowych.

Dla 57% przedsiębiorstw ważnym sposobem wykorzystania BI w chmurze jest analiza danych ad hoc. Dla 55% firm BI w chmurze to przydatne rozwiązanie podczas tworzenia raportów, 39% ankietowanych wskazało również użyteczność przy przygotowaniu danych do analizy. Mniej przedsiębiorstw wykorzystuje narzędzie Business Intelligence w modelu cloud computing do prototypowania (27%), analizy predykcyjnej (23%), planowania i prognozowania operacji (23%), procesowania danych w czasie pików (14%), strategicznego planowania i przeprowadzania symulacji (13%).

Raport BARC prezentuje też zmiany, jakie zachodzą w sposobie użycia narzędzi analitycznych w chmurze. W 2013 roku aż 60% wskazywało jako priorytet analizę ad hoc. W roku 2016 było

to 57%. Jednocześnie, aż 84% przedsiębiorstw zamierza w ten sposób wykorzystywać narzędzia dostępne w chmurze. Znacząco rośnie zainteresowanie przygotowywaniem danych do analizy. Dużo firm myśli również o stosowaniu chmury przy analizie predykcyjnej, strategicznym planowaniu czy planowaniu i prognozowaniu operacyjnym.

FIRMY KORZYSTAJĄ Z CORAZ WIĘKSZEJ LICZBY KOMPONENTÓW W CHMURZE

Na pytanie, jakie komponenty BI są obsługiwane przez chmurę, ankietowani stwierdzili, że są to: narzędzia BI (62%), serwery niezbędne do funkcjonowania BI (51%), narzędzia do eksploracji danych w chmurze Visual Discovery (49%), narzędzia do predykcyjnych i zaawansowanych analiz (25%), aplikacje analityczne (22%), narzędzia do zarządzania wydajnością (18%). Co ciekawe, w badaniu z 2013 roku zabrakło aplikacji analitycznych i narzędzi do zarządzania wydajnością. Jak widać, coraz więcej obszarów analitycznych przejmowanych jest przez narzędzia oferowane w modelu cloud computing.

Jeśli chodzi o zarządzanie danymi, to najczęściej w chmurze magazynowane są dane oraz prowadzona jest integracja danych ETL/Batch. W 33% ankietowanych przedsiębiorstw w chmurze przygotowuje się dane dla biznesowych użytkowników systemu, a w 31% wykorzystuje się dane z chmury w czasie rzeczywistym.

UDZIAŁ BI W CHMURZE STALE ROŚNIE

Zgodnie z wynikami raportu BARC, znaczenie rozwiązań analitycznych w chmurze jest coraz ważniejsze dla coraz szerszej grupy przedsiębiorstw. Widać wyraźnie, że firmy sięgają po dedykowane narzędzia w chmurze.

Coraz więcej rozwiązań cloud computing w firmach to rozwiązania hybrydowe i publiczne. Przedsiębiorstwa coraz chętniej sięgają po chmurę publiczną i hybrydową. W roku 2016 wśród badanych przez BARC organizacji tylko 24% korzystało wyłącznie z chmury prywatnej, 30% użytkowało chmurę hybrydową, a 46% – publiczną. Dla porównania, pod koniec 2013 roku było to odpowiednio: 29%, 27% i 44%.

Dostawcy usług w chmurze nie tylko świadczą usługi hostingowe, ale przede wszystkim dostarczają odpowiednie narzędzia (62% ankietowanych). Na podstawie tych rozwiązań zarządza się i uruchamia konkretne środowisko pracy (25%).

W 2013 roku aż 60% wskazywało jako priorytet analizę ad hoc w chmurze. W roku 2016 było to 57%. Jednocześnie, aż 84% przedsiębiorstw zamierza w ten sposób wykorzystywać narzędzia dostępne w chmurze.

Znacząco rośnie zainteresowanie przygotowywaniem danych do analizy. Dużo firm myśli również o stosowaniu chmury przy analizie predykcyjnej, strategicznym planowaniu czy planowaniu i prognozowaniu operacyjnym.

ADAM JADCZAK

W JAKI SPOSÓB WYKORZYSTAĆ

potencjał Machine Learning w analityce

O analizie dużych zbiorów danych, wykorzystaniu ich do poznania preferencji konsumentów, rozwiązaniach opartych na Reinforcement Learning, Generative Models i Deep Learning oraz sztucznej inteligencji rozmawiamy z Jackiem Dąbrowskim, dyrektorem ds. technologii w So1.

W jaki sposób w So1 wykorzystujecie dane?
Tworzymy algorytmy, które – w połączeniu z kartami lojalnościowymi i aplikacjami mobilnymi – pozwalają dużym, niemieckim sieciom handlowym tworzyć dla klientów dedykowane zestawy promocji. W tradycyjnych, papierowych ulotkach mamy kilkanaście, kilkadziesiąt stron z setkami różnych promocji. Dzięki naszym rozwiązaniom klient supermarketów otrzymuje kilka trafnie dobranych promocji, takich, z których prawdopodobieństwo skorzystania jest największe.

Przewidujemy m.in. czy dany produkt ktoś kupiłby bez promocji i jaki jest wpływ zniżek różnej wysokości. Dzięki temu budujemy tzw. nadwyżkę zakupową, maksymalizując sprzedaż. Korzyść zyskują jednak nie tylko sieci handlowe. Dla klienta będzie to otrzymanie sprofilowanej pod niego promocji na produkty, które go najbardziej zainteresują z indywidualnymi, atrakcyjnymi zniżkami. Dla supermarketów – zwiększanie obrotu. Dodatkowo zaś zachęta do kupowania produktów, po które dany klient nigdy wcześniej nie sięgał. Dla marek będzie to natomiast np. zachęcenie konsumenta do rezygnacji z produktów konkurenta. Supermarkety i marki rozliczają się z nami w ramach *success fee*.

W jaki sposób analizowane są dane kupujących?

Opracowany w So1 system jest w pełni anonimowy. Klient nie musi podawać żadnych informacji na swój temat. Nie musi się rejestrować. Nie wiemy, kim jest z imienia i nazwiska. Wiemy jednak, co kupuje i prognozujemy, co będzie kupował. Bazujemy na wszystkich, dokonywanych w danej sieci transakcjach – produktach, porach dnia, dniach tygodnia, lokalizacji sklepu i wydawanych kwotach. Choć tych parametrów jest znacznie więcej. Do indywidualnej analizy preferencji zakupowych wystarczą nam dane z 3 paragonów klienta. Już ta ich minimalna liczba pozwala nam dowiedzieć się prawie wszystkiego o danym konsumencie. Zysk dla algorytmów z kolejnych 5 koszyków zakupowych jest niewielki, choć pozwala poznać szczegółowe preferencje klienta dotyczące niszowych produktów.

Celem stojącym za naszym działaniem jest jak najdokładniejsze modelowanie profili oraz uzyskanie jak najtrafniejszej informacji o tym, kiedy i gdzie dany konsument pójdzie do sklepu, a następnie co i w jakiej ilości kupi. Stosowane przez nas modele probabilistyczne *(Generative Models – przyp. red.)* generują scenariusze zachowania klienta w przyszłości o największym prawdopodobieństwie. Dzięki temu jesteśmy od 5 do 8 razy bardziej skuteczni niż losowo wysyłane promocje. Zerknięcie na przygotowany przez nas zestaw 8 promocji kończy się wykorzystaniem co najmniej jednej z nich aż w 60%, a obrót sklepów wzrasta o 3%.

Współpracujemy z firmami, które prowadzą badania panelowe o konsumentach sieci handlowych. My zaś – na bazie paragonów – możemy z dużym prawdopodobieństwem odcyfrować dane socjodemograficzne, takie jak płeć, stan cywilny, liczba osób w gospodarstwie domowym. Łącząc te informacje, możemy targetować kampanie np. na rodziny z dwójką dzieci.

Wykorzystujemy algorytmy Deep Learning. Każdy produkt, paragon i konsument jest reprezentowany wektorem. Po każdym kolejnym zakupie dodawane są informacje, które aktualizują te wektory.

Nasze algorytmy rozpoznają nawet nienazwane właściwości produktów. Łącznie jest to ok. 800 cech. Informacji o estetyce, walorach smakowych czy modach i trendach nie znajdziemy na paragonie, w tabelach cech produktów ani na etykiecie. Dostarczają nam ich zachowania konsumentów.

Z jakich narzędzi korzystacie w pracy?

Podstawowym są algorytmy Deep Learning. Dysponując zestawem danych, otrzymujemy problem do rozwiązania. Jednak za każdym razem znamy prawidłową odpowiedź. Chcąc przewidywać, czy ktoś kupi w supermarkecie bułkę – bazując na danych historycznych – wiemy, czy to zrobił i w jakiej sytuacji. Deep Learning bazuje na tzw. ukrytej reprezentacji. Modele Deep Learning dla każdego problemu tworzą własną reprezentację. W przypadku Google Translate było to kodowanie treści językowych na potrzeby mechanizmu tłumaczenia. Podobnie jest np. z rozpoznawaniem mowy.

Tworząc rozwiązania sztucznej inteligencji, będziemy jednak musieli wykorzystać inne narzędzia – Reinforcement Learning i Generative Models. Pierwsze z nich oparte jest na sposobie, w jakim się uczymy. W procesie tym staramy się doprowadzać do tego, aby jak najczęściej otrzymać nagrodę. Wiemy, jak to zrobić na podstawie informacji zwrotnych. Nie wiemy jednak, kiedy otrzymamy nagrodę. Jesteśmy także zdani na siebie, jeśli chodzi o rozpoznanie sytuacji, w której się znajdujemy. Maszyny „uczą" się maksymalizacji sytuacji, w których otrzymają nagrodę, grając, korzystając np. z platformy OpenAI Gym. Reinforcement Learning to w uproszczeniu systemy „rzucone" w nieznane środowisko i niemające nauczyciela. Same muszą nauczyć się tak działać, aby jak najczęściej otrzymywać nagrodę.

Z kolei wspomniane już przeze mnie wcześniej Generative Models oparte są na modelach probabilistycznych. Korzystając z nich, otrzymujemy rozkład prawdopodobieństwa wszystkich możliwych zachowań czy obserwacji. Ten matematyczny model jest więc bliski temu, jak zachowuje się człowiek. Uczymy się nie tylko tego, jak się

zachować w danej sytuacji, lecz także przewidywać świat wokół nas. Dzięki temu wiemy, które zjawiska są dla nas zaskakujące i muszą zwrócić naszą uwagę. Pozostałe sytuacje dzieją się niejako w tle. Przykładowo, początkowo dużo uwagi poświęcamy prowadzeniu samochodu, później nasze reakcje są nieświadome.

W jaki sposób wykorzystujecie narzędzia Deep Learning?

Każdy produkt, paragon i konsument jest reprezentowany wektorem. Po każdym kolejnym zakupie dodawane są informacje, które aktualizują te wektory. Nasze algorytmy rozpoznają nawet nienazwane właściwości produktów. W uproszczeniu, dokonywane przez nie operacje można opisać takim równaniem:

(Banany bio) – banany + cebula = (cebula bio)

To bardzo prosty przykład właściwości, którą rozumiemy pod nazwą „eko produktu". Analizujemy jednak dużo trudniejsze do „wyłapania" zależności. Wpływ na preferencje zakupowe może mieć np. ekskluzywne opakowanie produktu. To zaś można powiązać z tym, że dany produkt nie będzie zawierał produktów GMO, będzie dedykowany osobom z nietolerancją laktozy itp. Informacji o estetyce, walorach smakowych czy modach i trendach nie znajdziemy na paragonie, w tabelach cech produktów ani na etykiecie. Dostarczają nam ich zachowania konsumentów. Opierając się na tych informacjach, możemy proponować produkty podobne do lubianych przez klienta. Nasze algorytmy wykrywają i analizują ok. 800 cech. Kolejne 500 wymiarów opisuje konsumentów.

Co ciekawe, nasze algorytmy ignorują kategorie i podkategorie stosowane przez sieci handlowe. Tworzą własne, lepiej dopasowane do preferencji klientów hierarchie produktów wybieranych w pewnym kontekście, np. śniadaniowych. Inne rzeczy zaproponujemy więc w środę rano, a inne w piątek wieczorem. Rozkład tych 800 cech obrazuje wykorzystywana przez nas mapa wymiarów (rys. 1).

Obecnie większość działań koncentruje się na analizie dużych zbiorów danych o zachowaniach użytkowników...

Widzimy przyszłość w tym, że dane zbierane z wielu źródeł i w różnorodnym kontekście – nawet anonimowe – będą następnie w sposób algorytmiczny syntetyzowane. Będą poszerzać naszą perspektywę. Firmy telekomunikacyjne np. gromadzą dane o połączeniach, SMS-ach, naszej lokalizacji, zapytaniach do serwerów DNS itp. Być może przyszłością dla nich jest – przy spadających dochodach z połączeń i dostępu do internetu – przestawienie się na model wykorzystujący synergię tych danych.

Innym źródłem informacji są oczywiście pliki cookies i cała historia przeglądania przez nas stron internetowych. Poza tym każda ze stosowanych przez nas przeglądarek dysponuje własną sygnaturą, tzw. fingerprintem. Jej wersja, rozdzielczość, zainstalowane dodatki i wtyczki sprawiają, że nabiera ona unikalnych cech. Na tej podstawie strony internetowe są w stanie łatwo odróżnić nasze kliknięcia i dokładniej śledzić to, co robimy.

Na bazie danych operatora i informacji z sieci supermarketów jesteśmy w stanie zidentyfikować, że w obu przypadkach była to ta sama osoba. Dokonuje się tego na podstawie czasu transakcji i geolokalizacji. **Już nawet 1% zidentyfikowanych osób pozwala na ekstrapolację danych na całą społeczność. Na tej bazie otrzymujemy bardzo skuteczny model probabilistyczny. Do skutecznego targetowania w zupełności wystarczą anonimowe dane.**

Połączenie tych wszystkich danych daje praktycznie kompletny obraz możliwych zachowań konsumentów. Przykładowo, na bazie danych operatora i informacji z sieci supermarketów jesteśmy w stanie zidentyfikować, że w obu przypadkach była to ta sama osoba. Dokonuje się tego na podstawie czasu transakcji i geolokalizacji.

Już nawet 1% zidentyfikowanych osób pozwala na ekstrapolację danych na całą społeczność. Na tej bazie otrzymujemy bardzo skuteczny model probabilistyczny. Do skutecznego targetowania w zupełności wystarczą anonimowe dane. Nie ma bowiem znaczenia to, jak dana osoba się nazywa, lecz to, co robi, kiedy to robi, co lubi...

Czy to nie przerażające, że „maszyny" tak dużo o nas wiedzą?

Z jednej strony, może wydawać się to przerażające, z drugiej – jest nieuniknione. Koncepcja sklepów Amazon Go bazuje na tym, że są one najeżone kamerami i czujnikami analizującymi nasze zachowania. Docelowo tego typu rozwiązania znajdą się także w naszych domach. Nie ma odwrotu do czasów, gdy społeczeństwa funkcjonowały na podstawie blokad informacyjnych, gdzie tylko część osób miała pełny ogląd tego, co dzieje się na świecie.

Zmierzamy do społeczeństwa, w którym technologia będzie gromadziła ogromne ilości danych na nasz temat. Moim zdaniem, nie będzie sensu panować nad ich wyciekiem. Tym bardziej że istnieje ryzyko, iż nierówny dostęp do informacji wszystkich podmiotów na rynku zostanie wykorzystany przeciwko nam. Musimy raczej dążyć do pełnej transparentności, aby wszyscy mieli dostęp do zbieranych danych. Zwłaszcza że – jak wspomniałem wcześniej – dla przetwarzających je algorytmów nie ma znaczenia nasze imię i nazwisko.

W jaki sposób można wykorzystać różne źródła danych Big Data?

Na podstawie danych operatora wiemy np., że ktoś ma dobry dzień. Jego wzmożona aktywność na serwisie społecznościowym, dzielenie się dobrą informacją, może oznaczać, że otrzymał w pracy podwyżkę. Utwierdzić nas w tym może fakt zakupu dobrego wina i polędwicy wołowej dla 2 osób, co system zidentyfikuje jako przygotowania do uroczystej kolacji. Oczywiście wiemy, że tego dnia nie ma on urodzin. W tym momencie nasz algorytm mógłby mu zaproponować np. promocję na wycieczkę z ograniczonym czasem trwania tej oferty. Sam zainteresowany o tym nie pomyślał, ale być może będzie chciał uczcić podwyżkę wyjazdem na weekend z towarzyszką lub towarzyszem życia.

Jeśli nałożymy dodatkowo na to preferencje żywnościowe, które łączą się z naszym charakterem – a dokładnie z poziomem neurotransmiterów – będziemy wiedzieli, jakie jest prawdopodobieństwo podjęcia takiej spontanicznej decyzji. Gdy ludzie odżywiają się zdrowo, ich zakupy mają mniej impulsywny charakter. Z kolei osoby jedzące tłusto, słono i słodko skłonne są do bardziej ryzykownych działań.

Na razie nasze działania analizują zaawansowane algorytmy. Jednak w przyszłości zastąpi je zapewne sztuczna inteligencja. Nad rozwiązaniami tego typu pracują naukowcy z kilku ośrodków. Najważniejsze z nich to: OpenAI założony przez Elona Muska, Google DeepMind, Facebook AI Research oraz dział R&D Microsoft. Czym się zajmują?

Instytut OpenAI chce publicznie udostępniać wyniki prac nad sztuczną inteligencją. Chcą rozpowszechniać wszystkie osiągnięcia, wraz z kodami źródłowymi. Firmy komercyjne wyniki swoich prac raczej utajniają albo opóźniają ich udostępnianie. OpenAI dąży do pełnej transparentności.

Elon Musk – i stojący za nim OpenAI – wspiera także społeczność osób zajmujących się tymi tematami. Udostępniają im np. narzędzia, do których dotychczas dostęp miały tylko największe korporacje. Przykładem jest platforma OpenAI Gym. Jest to środowisko, w którym znalazło się kilkaset gier. Gdy tworzysz algorytm oparty na Reinforcement Learning, potrzebujesz dostępu do platformy, na której – grając w gry – będzie się on mógł uczyć. Ręczne podłączenie kilkuset gier jest praktycznie niemożliwe. W OpenAI Gym wszystko jest zautomatyzowane. Co ciekawe, społeczność skupiona wokół OpenAI – po uruchomieniu tej platformy – była w stanie odtworzyć wszystkie kluczowe, nieopublikowane dotąd kody źródłowe algorytmów Google Deep Mind.

Nawiasem mówiąc, w OpenAI pracuje aż 6 Polaków. Tymczasem w Polsce nikt tematami związanymi ze sztuczną inteligencją się nie zajmuje, i to mimo bardzo silnych fundamentów matematycznych w naszym kraju. Nie istnieją praktycznie zespoły zajmujące się pracami badawczo-rozwojowymi nad sztuczną inteligencją i jej zastosowaniami. Jeśli szukamy kogoś do naszego zespołu, musimy uczyć te osoby zarówno teorii, jak i praktyki.