



# GENERATION DATA

USING DATA FOR PROFIT

1.0





## WPROWADZENIE

### 1. KRÓTKA HISTORIA DANYCH

Dane nie są czymś nowym. Stają się jednak tak duże i tak wszechobecne, że mają wpływ na wszystkich, niezależnie od tego, czy tego chcemy, czy nie. Całkowicie zmieniają sposób, w jaki prowadzimy działalność i wpływają na większość innych aspektów naszego życia.

### 2. CO TO JEST BIG DATA?

Teraz, gdy wiemy już, co się zmieniło, czym tak naprawdę jest big data?

- Oficjalna definicja to “Dane o bardzo dużych rozmiarach, operowanie i zarządzanie którymi stanowią poważne wyzwania logistyczne”. Oxford English Dictionary, 2013. Ta definicja ma jednak ograniczenia, ponieważ koncentruje się wyłącznie na aspektach technicznych i przedstawia dane jako coś problematycznego.
- Alternatywna definicja: „Big data to zbiory danych o dużej objętości, dużej prędkości i różnorodności, które wymagają opłacalnych, innowacyjnych form przetwarzania informacji w celu uzyskania lepszego wglądu i podejmowania decyzji.” Gartner, 2014<sup>1</sup>. Ta definicja daje nam szerszy obraz. Podstawową ideą leżącą za zwrotem „Big Data” jest zatem to, że wszystko, co robimy, pozostawia coraz większy ślad cyfrowy, który możemy wykorzystać i przeanalizować. Big Data odnosi się zatem do samych danych, ale także do naszej zdolności do korzystania z coraz większej ilości danych.
- Niektórzy ludzie idą dalej: „Big data to nowe podejście firm, organizacji non-profit, agencji rządowych i osób fizycznych, które łącząc dane z wielu źródeł może prowadzić do podejmowania lepszych decyzji.” Gil Press, współpracownik Forbes<sup>2</sup>.
- 

#### 2. 1 Zasada 5V

To, co zmieniło się w przypadku danych w ciągu ostatnich 10 lat, czyniąc ten temat tak bardzo interesującym, to zasada 5V:

VOLUME (WIELKOŚĆ) - wielkość generowanych danych.

VELOCITY (PRĘDKOŚĆ) - Szybkość, z jaką dane są generowane i agregowane.

VARIETY (RÓŻNORODNOŚĆ) - Rodzaje dostępnych danych.

VERACITY (WIARYGODNOŚĆ) - Dokładność lub wiarygodność danych.

VALUE (WARTOŚĆ) - zakres, w jakim dane generują cenne ekonomicznie spostrzeżenia lub korzyści.

#### Wielkość

Tam, gdzie zaledwie kilka lat temu Gigabajty (GB) były limitem, Terabajty (TB) i Petabajty (PB) stały się teraz wspólną walutą życia menedżerów danych.

- W ciągu każdej sekundy przez Internet przesyła się więcej danych, niż 20 lat temu znajdowało się w całym Internecie
- Według IBM, 90% danych, które istnieją dzisiaj, zostało wygenerowanych w ciągu poprzednich dwóch lat.

Większość danych stało się tak dużymi, że nie możemy już przechowywać i analizować ich przy użyciu tradycyjnej technologii baz danych. Obecnie korzystamy z systemów rozproszonych, w których części danych są przechowywane w różnych lokalizacjach i łączą się przez oprogramowanie.

#### Prędkość

- Z każdą sekundą każdego dnia wzrasta liczba danych. Nie tylko trzeba te dane przeanalizować, ale także należy zapewnić odpowiednio dużą szybkość transmitowania tych danych, aby umożliwić dostęp do nich w czasie rzeczywistym (np. dostęp do strony internetowej, weryfikacja karty

<sup>1</sup> <http://www.gartner.com/itglossary/big-data/> (dostępne online 18 lipca 2019 r.)

<sup>2</sup> <https://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours/#21c1558d13ae>



kredytowej, wysyłanie szybkich wiadomości).

- W przypadku wielu aplikacji szybkość tworzenia danych jest nawet ważniejsza niż ich wielkość, ponieważ dzięki analizie informacji w czasie rzeczywistym lub prawie w czasie rzeczywistym firma może być znacznie bardziej elastyczna niż jej konkurenci.

### Różnorodność

Big data ma postać wiadomości, aktualizacji i obrazów umieszczanych w sieciach społecznościowych; odczytów z czujników; sygnałów GPS z telefonów komórkowych i nie tylko. W miarę cyfryzacji coraz większej części działalności biznesowej, nowe źródła informacji i coraz tańszy sprzęt łączą się, aby dostarczyć nam ogromne ilości niezwykle różnorodnych rodzajów danych. Dostępne dane są często nieuporządkowane - nie są zorganizowane w bazie danych - i nieporęczne, ale ogromna ilość po prostu czeka na ich opublikowanie.

### Wiarygodność

W dzisiejszych czasach obowiązkiem jest weryfikacja wiarygodności danych. Dane, w większości prawdziwe, mogą okazać się jednak niedokładne lub niekompletne. Przykładem może być system GPS, który każe nam lawirować ulicami miasta, co oznacza, że nie jest w stu procentach dokładny. Dzieje się tak dlatego, ponieważ sygnał "gubi się" odbijając się od wysokich budynków. Wówczas dane dotyczące lokalizacji muszą być powiązane z innymi danymi, takimi jak dane drogowe, dane z akcelerometra, żeby dostarczyć dokładnych informacji.

### Wartość

Posiadanie dostępu do dużych ilości danych nie daje korzyści, o ile nie potrafi się ich zamienić w realną wartość. Przykładami wartości wynikającej z posiadania danych mogą być:

- Dane, które pomagają firmom lepiej zrozumieć i wspierać swoich klientów. Dane zebrane przez Netflix dotyczące naszych nawyków pomagają im w lepszym doborze rekomendacji.
- Dane, które pomagają zoptymalizować procesy. Przykładem jest Uber, który jest w stanie przewidzieć zapotrzebowanie, szacować opłaty i wysłać najbliższego kierowcę do klienta na podstawie zebranych danych.
- Dane, które umożliwiają bycie innowacyjnym. Firmy sportowe umieszczają sensory w piłkach i GPS w ubraniach aby zwiększyć ich przydatność dla sportowców, którzy mogą dzięki temu analizować własne postępy.

## 2.2 ZROZUMIEĆ DANE

Dane mogą być rozumiane i klasyfikowane zgodnie z tym jak (i przez kogo) są generowane i/lub jak są organizowane lub strukturyzowane.

Dostępne są nieskończone rodzaje danych, ale samo posiadanie WIĘCEJ danych nie zawsze jest lepsze. Jakość danych oraz ich przydatność lub wartość dla Twoich celów są tak samo ważne, jak ich ilość.

## 3. OD DUŻYCH DANYCH DO INTELIGENTNYCH DANYCH

Musimy być realistyczni, pragmatyczni, a czasem nawet sceptyczni co do tego, co można osiągnąć. Musimy wiedzieć też jaką wartość możemy uzyskać i jak ją zmaksymalizować.

- Dane odnoszą się do wszelkich nieprzetworzonych i nieuporządkowanych informacji (tekst, liczba, media).
- Duże dane odnoszą się do pompowania ilości danych, niekiedy również obniżając przez to ich jakość.
- Inteligentne dane polegają na tworzeniu oszczędnych i innowacyjnych sposobów przetwarzania dużych ilości danych w celu uzyskania lepszego wglądu, poprawienia procesu podejmowania decyzji i umożliwienia automatyzacji realizowanych procesów.<sup>3</sup>
- Definicja może także odnosić się do danych, które wymagają niewielkiej interwencji/koordynacji ze strony człowieka, aby były przydatne.

(W przeszłości większość analiz odbywała się przy przetwarzaniu wsadowym. Dane były gromadzone zgodnie z harmonogramem, konwertowane do pożądanego stanu, wprowadzane do bazy danych i

<sup>3</sup> <https://securities.bnpparibas.com/insights/what-is-smart-data-2.html>



przetwarzane co godzinę, co noc lub co tydzień. Wadą tego podejścia jest to, że w czasie, w którym dane są analizowane, mogą być już one nieaktualne. W przeciwieństwie do tego podejścia, inteligentne dane monitorują dane u źródła, oceniają je, podejmują decyzję i dzielą się wynikami - wszystko w określonym przedziale czasu składającym się z sekund lub ułamków sekundy.)

### **Jak zmierzyć, czy dane będą inteligentne?**

Zadawaj pytania, aby upewnić się, że szybko i skutecznie doprowadzi to do powstania lepszej, bardziej opłacalnej usługi lub produktu.

### **Jak stać się mądrym? Nie polegaj na HiPPO!**

Jednym z najbardziej krytycznych aspektów dużych zbiorów danych jest ich wpływ na sposób podejmowania decyzji i tego, kto je podejmuje. Gdy danych jest mało, są drogie do uzyskania lub nie są dostępne w formie cyfrowej, rozsądne jest, aby właściwi ludzie podejmowali decyzje, ponieważ robią to na podstawie zgromadzonego wcześniej doświadczenia oraz wzorców i relacji, które uprzednio zaobserwowali.

W społeczności osób zajmujących się dużymi zbiorami danych wielu jej członków utrzymuje, że firmy często podejmują większość ważnych decyzji, opierając się na „HiPPO” – „Highest-Paid Person’s Opinion”<sup>4</sup> (opinii najlepiej opłacanej osoby w danej firmie). Oczywiście wielu kierowników wyższego szczebla rzeczywiście opiera się na danych i są gotowi pominąć własną intuicję, gdy dane się z nią nie zgadzają, ale w dzisiejszym świecie biznesu ludzie zbyt często jednak polegają na doświadczeniu i intuicji, a za mało na posiadanych danych.

### **Czy Twoja firma jest „inteligentna”?**

Udowodniono, że firmy oparte na danych są bardziej produktywne i rentowne. Badanie NESTA wykazało, że Datavores (firmy polegające na danych przy podejmowaniu decyzji biznesowych) i Data Builders (firmy pracujące na dużych zbiorach danych) są ponad 10% bardziej wydajni niż Dataphobes (firmy korzystające z niewielkich ilości danych i nie korzystające z nich przy podejmowaniu decyzji).<sup>5</sup>

## **4. KORZYŚCI Z DANYCH DLA BIZNESU**

Znaczenie dużych zbiorów danych nie zależy od tego, ile danych ma firma, ale w jaki sposób firma je wykorzystuje. Każda firma wykorzystuje dane na swój własny sposób; im efektywniej to robi, tym większy jest jej potencjał.

Przejawia się to w innowacyjnych produktach i usługach opartych na danych, ulepszeniach procesów i bardziej świadomym podejmowaniu decyzji w całej gospodarce i społeczeństwie.

Korzyści można podzielić na dwie główne kategorie:

- Proces. Firma może pobierać dane z dowolnego źródła i analizować je w celu znalezienia odpowiedzi, które umożliwią ulepszenie wewnętrznych procesów.
- Produkt/innovacja. Dane nie tylko tworzą wartość, pozwalając nam robić te same rzeczy lepiej, ale także pozwalają nam robić zupełnie nowe rzeczy.

<sup>4</sup> <https://hbr.org/2012/10/big-data-the-management-revolution>

<sup>5</sup> <https://www.nesta.org.uk/report/skills-of-the-datavores-talent-and-the-data-revolution/>