
BigData & Cloud

Wprowadzenie

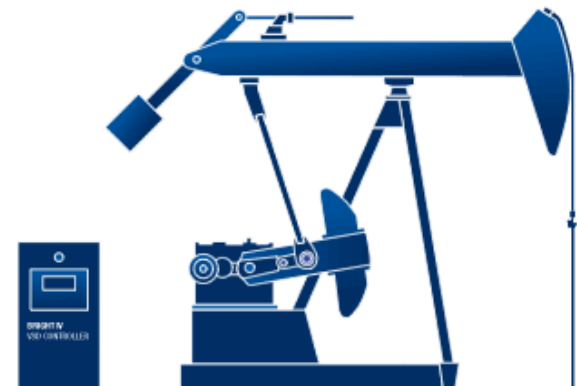
Poznań 29-30 wrzesień

Agenda

- Dane, dużo danych
- Przykłady
- Wyzwania i ograniczenia technologiczne
- Wbudowana ekspertyza
- Podsumowanie

Dane jako na nowo odkrywany zasób naturalny

- Dane istnieją wokół nas od zawsze
- Ich przetwarzanie ma coraz większe znaczenie
- Umiejętność ich przetwarzania przekłada się na konkretne zyski
- Z roku na rok generujemy ich coraz więcej
 - Systemy transakcyjne
 - Media społecznościowe
 - Systemy monitorowania
 - ...



Boom danych



2.5 million items
per minute



5 TB per flight



300,000 tweets
per minute



> 1 PB per day
gas turbines



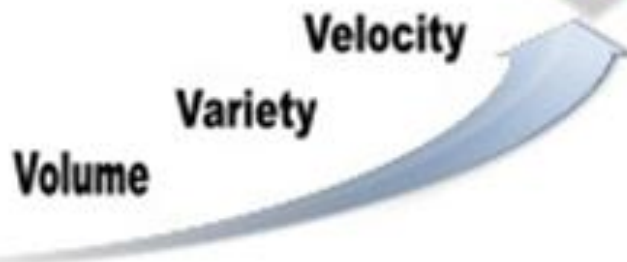
200 million emails
per minute



220,000 photos
per minute



2020
44 zettabytes



80%

Of world's data
is unstructured

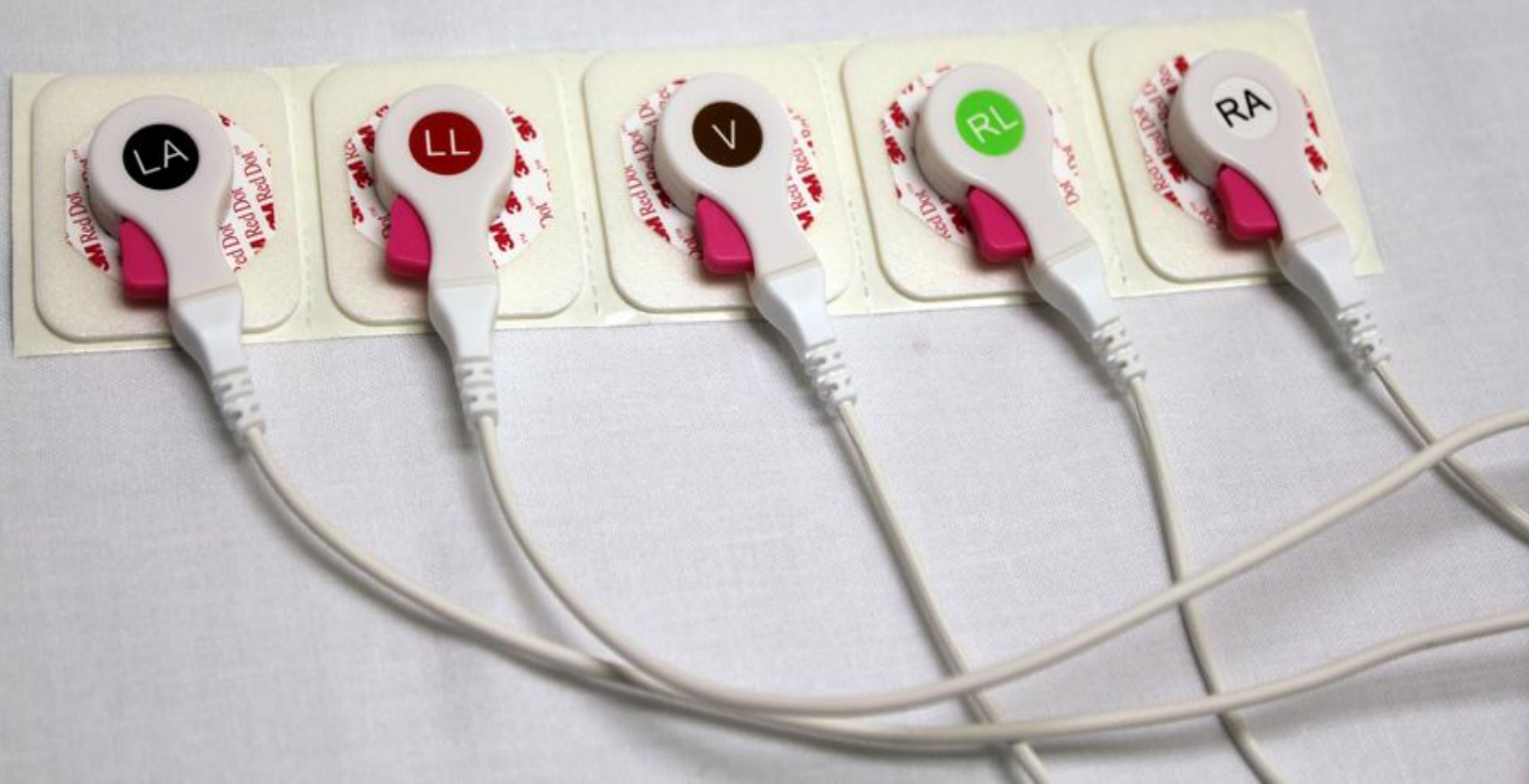
Ludzka matematyczna intuicja

- Dlaczego często nie wystarczy?

Przykład 1: Insurethebox – telemetria stosowana

- Składka w zależności od tego jak bezpiecznie się jeździ
- 2 lata działalności, 100 000 sprzedanych polis
- Gromadzenie i analizowanie następujących danych:
 - pora (dzień lub noc) przejazdu,
 - trasa przejazdu i prędkość,
 - płynność jazdy – przeciążenia,
 - przerwy przy dłuższych podróżach,
 - liczba przejechanych kilometrów (w tym po autostradach),
 - liczba podróży,
 - szczegóły wypadków





WordPress OneNote

Social Media



Facebook



Twitter



LinkedIn



Pinterest



Google+



Tumblr



Flickr



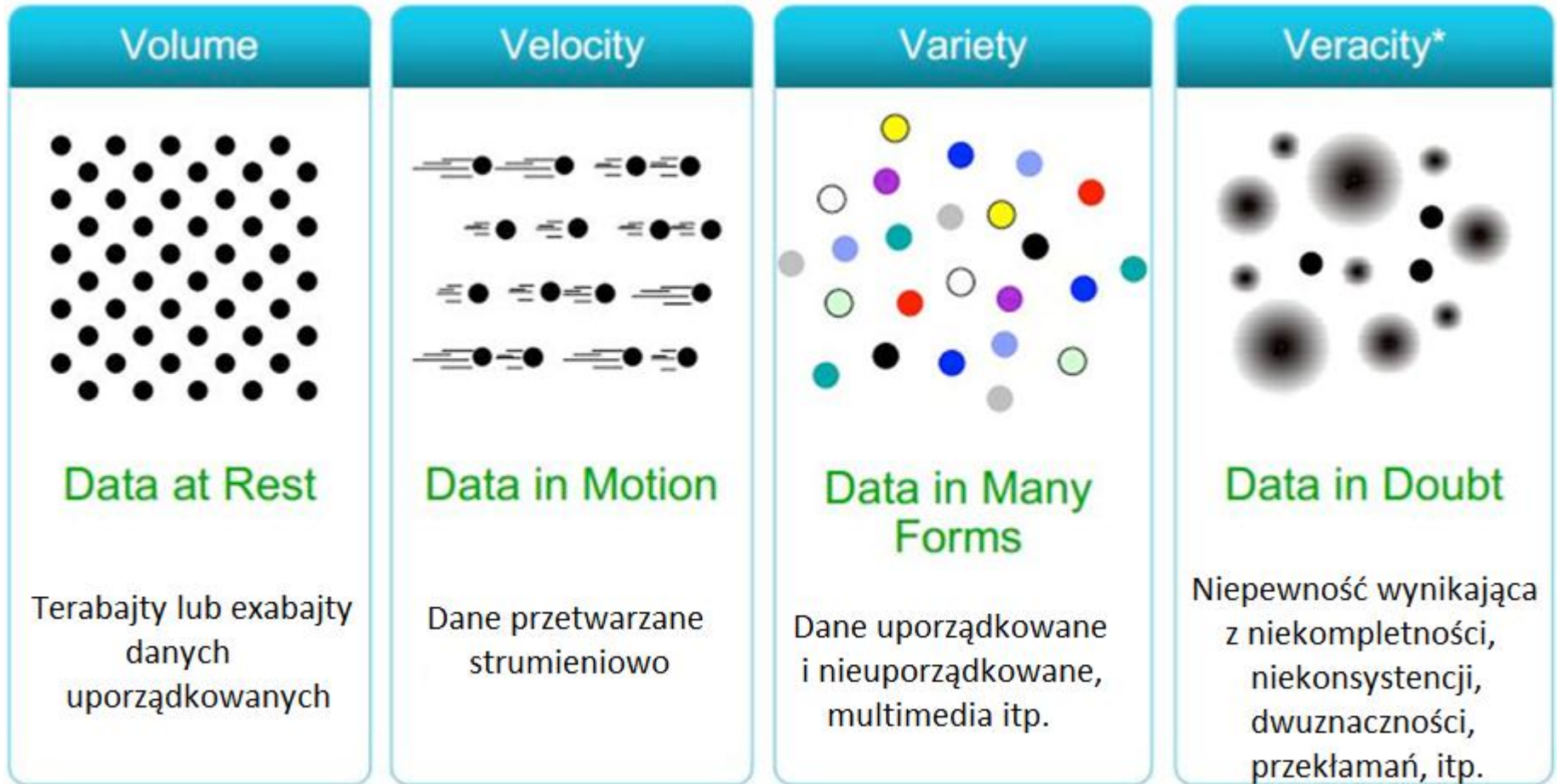
Big Data – ile to danych?

- Kilobajt – 1000 Bajtów
- Megabajt – 1000 Kilobajtów
- Gitabajt – 1000 Megabajtów
- Terabajt – 1000 Gigabajtów
- Petabajt – 1000 Terabajtów
- Exabajt – 1000 Petabajtów
- Zettabajt – 1000 Exabajtów
- ...



**Boeing 787
generuje ~5TB
danych na lot (wg.
Virgin Atlantic)**

BigData – co oznacza?



Zmiana podejścia

Tradycyjne podejście



Biznes

ustala pytania

IT

organizuje dane tak aby
aby udzielić odpowiedzi

Miesięczne raporty
Analizy rentowności
Ankiety konsumenckie

BigData



IT

dostarcza platformę
do kreatywnego
eksplorowania danych

Biznes

określa jakie jakie pytania
można zadać

Brand sentiments
Strategie produktu
Max. utylizacja zasobów

Główne obszary zastosowań

■ Finanse

- Ryzyka kredytowe,
- Analizy ryzyka
- Rozpoznanie rynku
- High Frequency Trading



■ Służba zdrowia

- Monitorowanie w czasie rzeczywistym czynności życiowych, analiza i informowanie personelu medycznego
- Przewidywanie ryzyka na podstawie historii pacjenta i jego rodziny



■ Telekomunikacja

- Optymalizacja wydajności sieci
- Poprawa jakości promocji



■ Logistyka

- Monitorowanie przesyłek
- Optymalizacja sieci przesyłek i centrów dystrybucyjnych



■ Media społecznościowe

- Analiza sentymentów
- Optymalizacja reklam

Wymagania dla infrastruktury pod hurtownię

- **Pojemność**
 - Tera, Peta, Exa...
- **Szybkość**
 - Zapytania analityczne wykonywane w kilka minut
 - Zapytania operacyjne z natychmiastową odpowiedzią
 - Ładowanie wielu Terabajtów na godzinę
- **Niezawodność**
 - Systemy używane przez kluczowe osoby w przedsiębiorstwach
 - Używane do podejmowania strategicznych decyzji
- **Skalowalność**
- **Funkcjonalność**
 - Funkcje analityczne z wielu obszarów
- **Koszt (początkowy i operacyjny)**

Dostępne technologie – CPU

■ CPU

–Możliwości

- < 10GHz
- 64bitowe rejestry
- Przepustowość GT/s

–Oczekiwania

- Przetworzenie terabajtów danych na minutę

■ Przestrzeń dyskowa

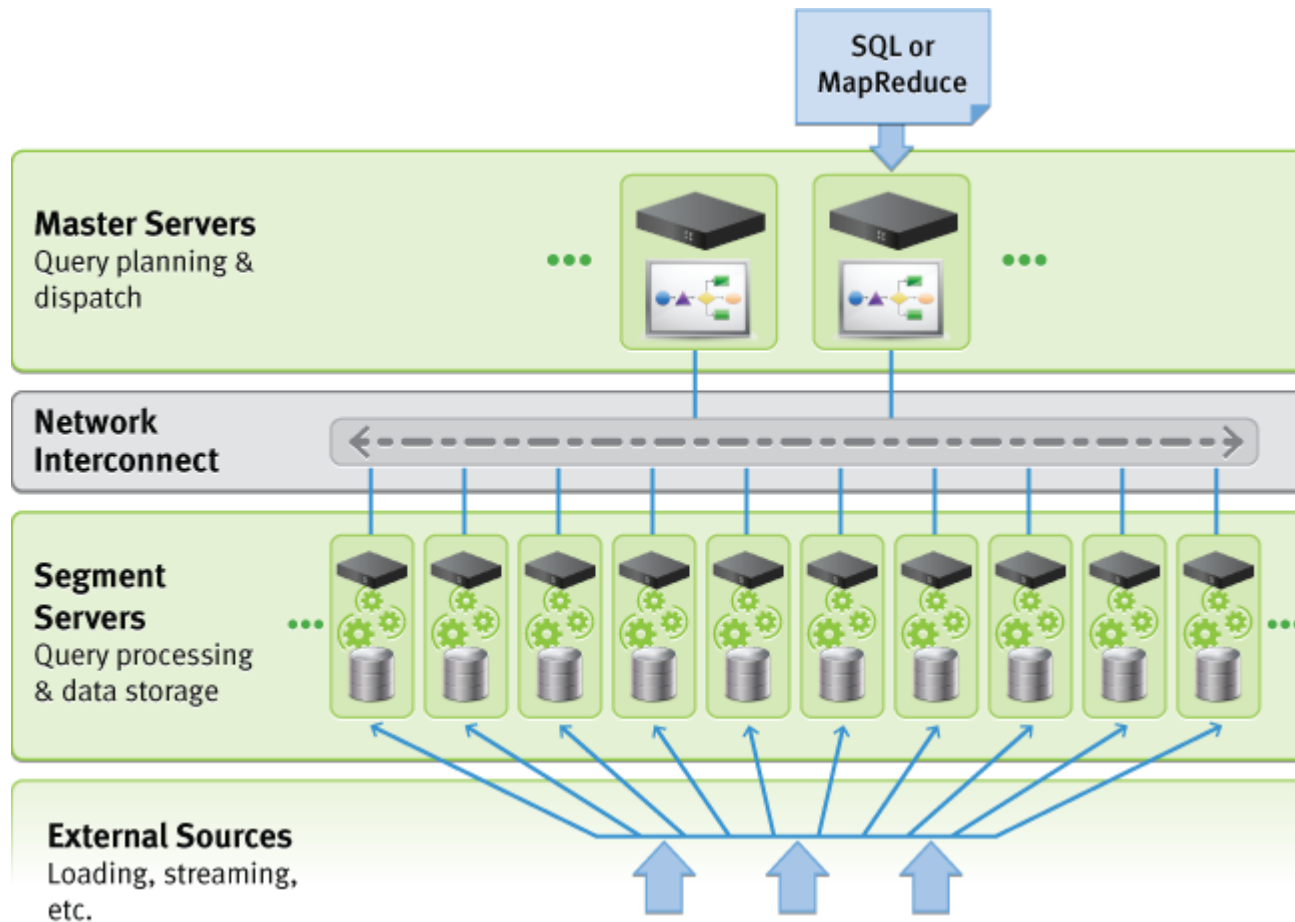
–Możliwości

- Maksymalna szybkość odczytu: setki megabajtów
- Maksymalna szybkość zapisu: setki megabajtów

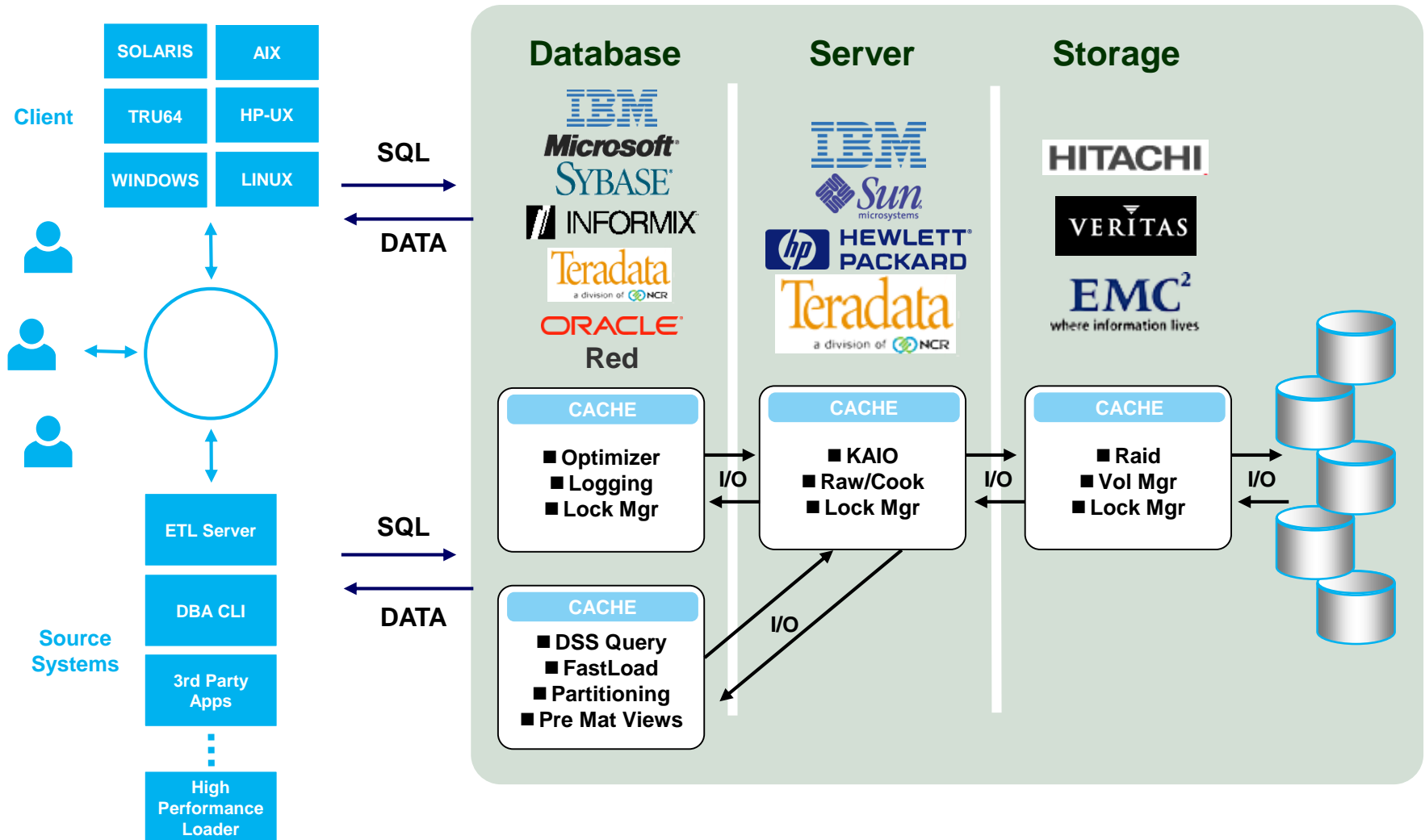
–Oczekiwania

- Możliwość skanowania terabajtów danych na minutę

Komputery masowo równoległe (MPP i AMPP)



Jak uruchomić hurtownię danych ?

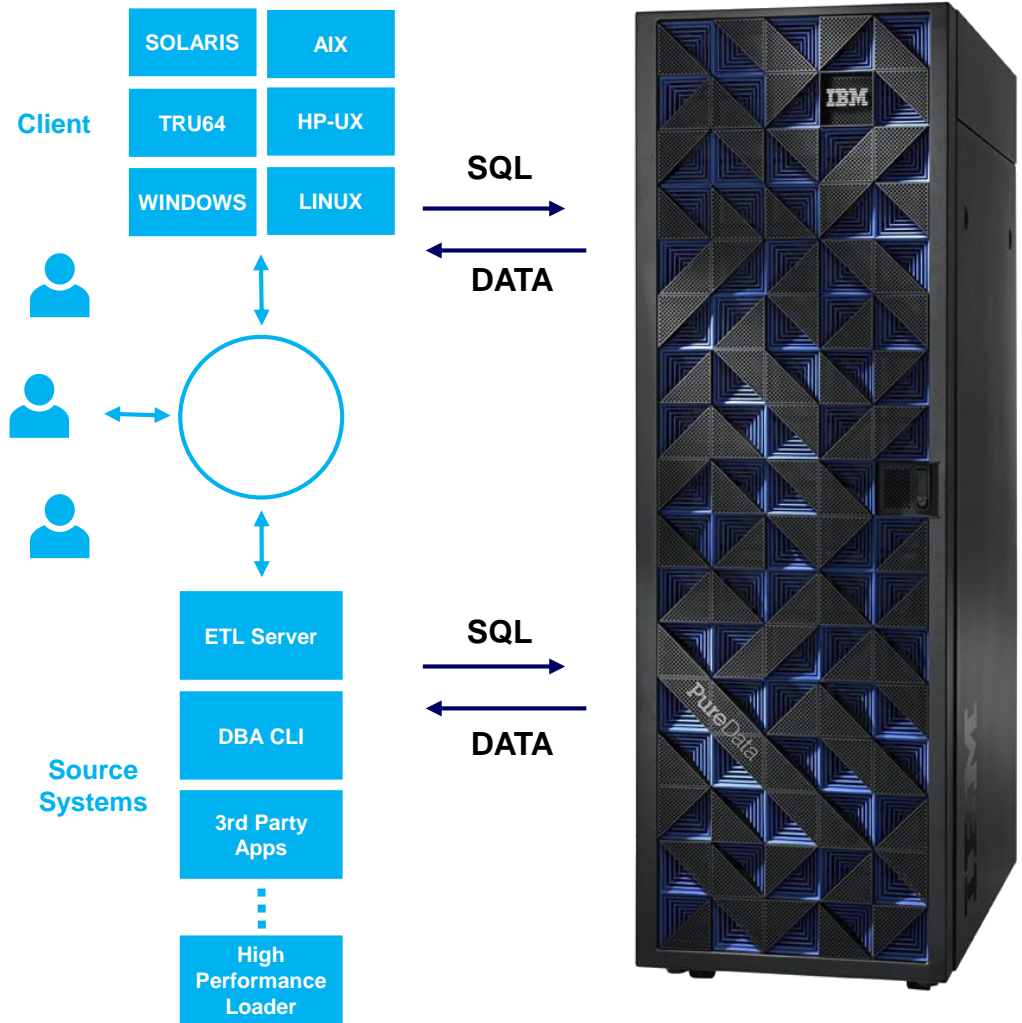


Appliance

- Czym jest?
 - Dedykowane urządzenie
 - Kompletne rozwiązanie
 - Szybka instalacja
 - Łatwe zarządzanie
- Typy
 - Storage
 - Network
 - Anti-spam
 - Search engines
- Plusy
 - Niski koszt całkowity (TCO)
 - Wbudowana ekspertyza (bezpieczeństwo, optymalizacja wydajności, itp.)
 - Możliwość szybkiej wymiany
 - Skalowalność
 - Skrócony czas przygotowania rozwiązania



Jak **szybko i bezboleśnie** uruchomić hurtownię danych?

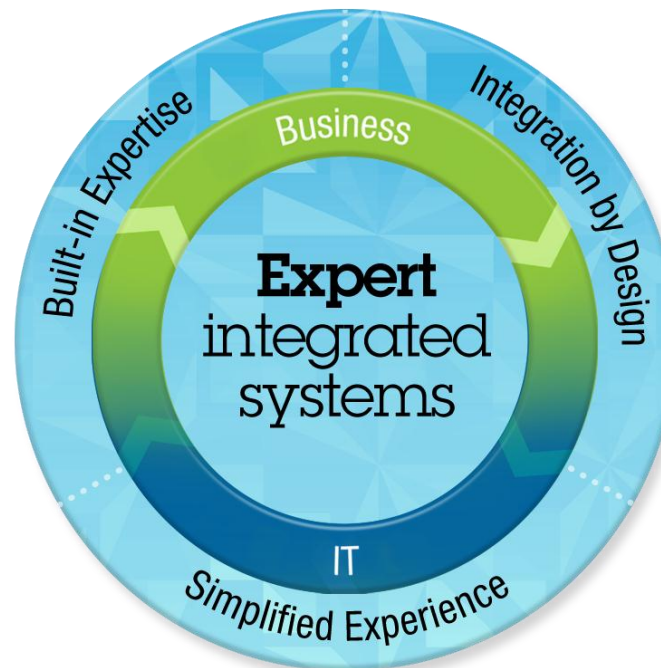


Systemy ze integrowaną wiedzą ekspercką

PureSystems

Wbudowana wiedza specjalistyczna

Rejestrowanie i automatyzowanie działań ekspertów – od tworzenia infrastruktury po obsługę aplikacji



Pełna integracja

Integracja i optymalizacja zasobów sprzętowych i programowych w ramach gotowego systemu dopasowanego do specyfiki konkretnych obciążeń

Uproszczona obsługa

Zmiana cyklu życia rozwiązań informatycznych – redukcja kosztów, oszczędność czasu i zasobów, a także szybsze wdrażania innowacji w przedsiębiorstwie

Rodzina IBM PureData Systems

Szybkie i łatwe ujarzmienie BigData



PureData

System for Transactions

Na przykład systemy typu E-commerce...

Platforma bazodanowa zaprojektowana i dostrojona specjalnie pod kątem obsługi danych transakcyjnych

PureData

System for Analytics

Na przykład analizowanie informacji o klientach

Prosta w obsłudze hurtownia danych zoptymalizowana na super szybką analitykę w skali PetaBajtów

**Powered by Netezza
technology**

PureData

System for Operational Analytics

Na przykład wykrywanie defraudacji w czasie rzeczywistym

Operacyjna hurtownia danych zoptymalizowana pod kątem złożonych analiz i obsługi nawet wielu tysięcy interaktywnych zapytań w ciągu sekundy

IBM PureData System for Analytics



Gotowy system do analiz danych

- ✓ Urządzenie specjalnego przeznaczenia – hurtownia danych
- ✓ Zintegrowane: silnik bazy danych, serwery oraz macierze dyskowe
- ✓ Standardowe interfejsy
- ✓ Niski koszt TCO (Total cost of Ownership)

Speed: 10-100x szybsza niż tradycyjne systemy

Simplicity: Minimalny nakład na administrację i strojenie

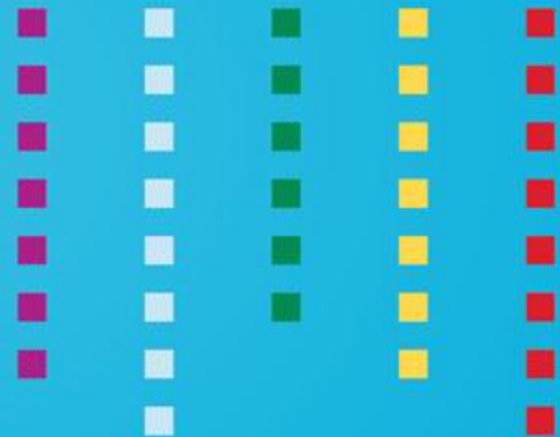
Scalability: Liniowa skalowalności sięgająca petabajtów

Smart: Superwydajna zaawansowana analityka

Made in Poland
Rozwój i utrzymanie
także w IBM Poland Lab

85%

nowego oprogramowania
jest **dostarczana**
w chmurze



Source: IDC

Learn more at ibm.com/cloud

