

# Systemy Business Intelligence, Big Data W13

*dr inż. Janusz Granat*

# Plan wykładu

---

- BI – wprowadzenie
- Przegląd rozwiązań BI ( w tym bazujących na Bigdata)
- Przykład analizy danych z wykorzystaniem BI
- Trendy rozwoju

# BI - Wprowadzenie



# Business intelligence

---

Business intelligence (BI) to zbiór technik i narzędzi przekształcających dane operacyjne w użyteczną informację biznesową. Informacja ta może być wykorzystana w procesach podejmowania decyzji oraz do wspierania działalności operacyjnej.

BI pozwala na analizę danych historycznych i aktualnych oraz dostarcza informacje o charakterze prognostycznym.

# Techniki BI

---

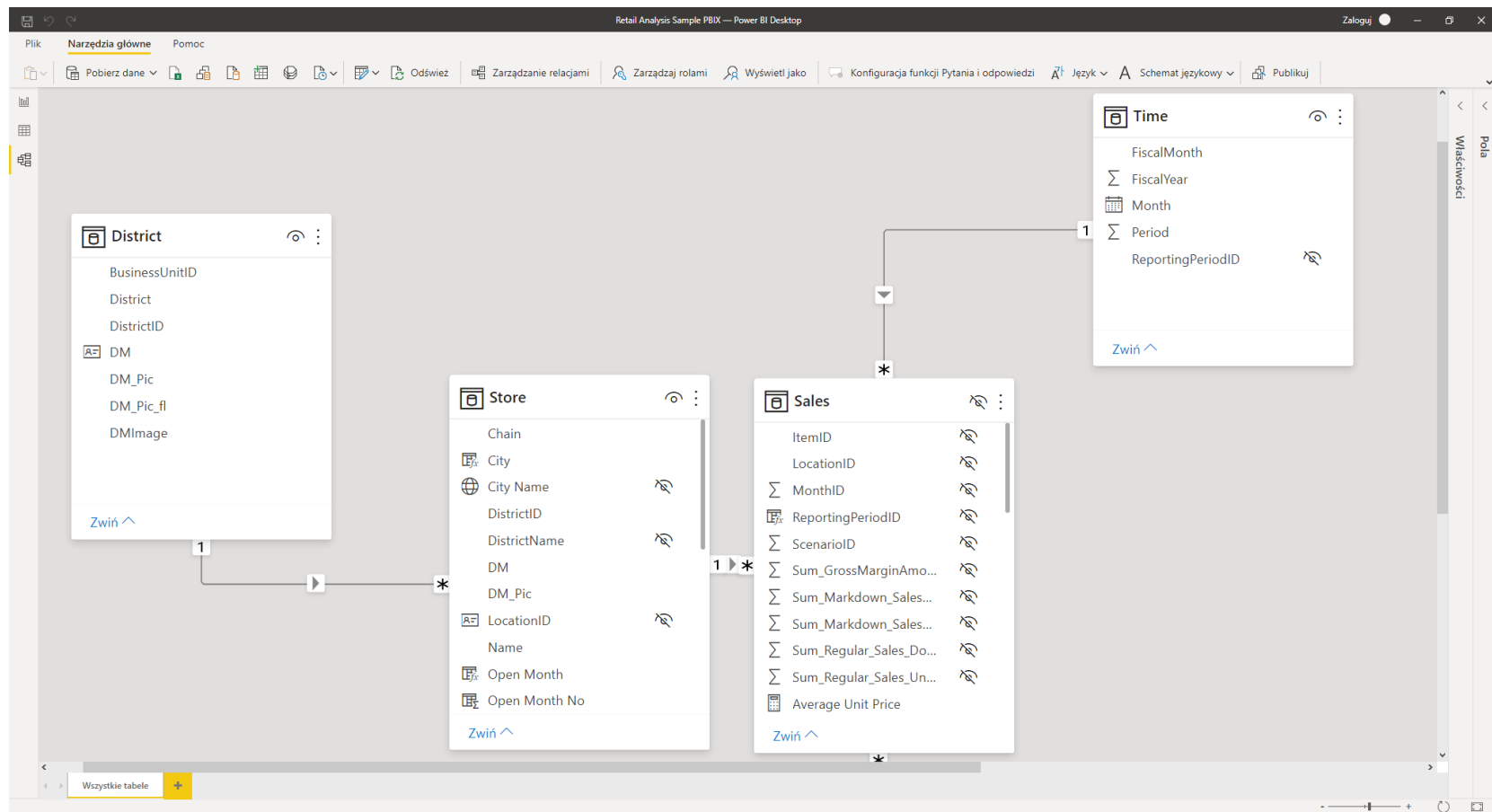
- Raportowanie
- OLAP (OnLine Analytical Processing)
- Analytics
- Data and text mining
- CEP (Complex Event Processing)

# Przechowywanie danych

---

- Bazy danych
- Hurtownie danych
- Bazy wielowymiarowe
- Specjalizowane rozwiązania przechowywania danych typu Big Data (3V Volume, Velocity Variety; duża ilość danych, duża zmienność danych, duża różnorodność danych)

# Bazy danych i hurtownie danych



# Relacyjna baza danych — fundament systemów transakcyjnych

## Czym jest?

Zorganizowany zbiór danych w tabelach powiązanych relacjami. Silnik: PostgreSQL, MySQL, Oracle, MS SQL Server.

## Cel i zakres

Obsługa bieżących transakcji (zamówienia, płatności, rejestracja). Optymalizacja pod kątem zapisu i odczytu pojedynczych rekordów.

## Model danych

Schemat relacyjny (ERD), normalizacja do 3NF. Silna

*Rola w BI: źródło surowych danych → zasilanie ETL → hurtownia lub data lake.*

## Struktura

Ustrukturyzowana  
(tabele, kolumny)

## Skalowalność

Pionowa (scale-up)  
lub sharding

## Latencja

Bardzo niska  
(ms)

## BI-gotowość

Ograniczona —  
zapytania OLAP wolne

Przykład: system ERP (SAP), sklep internetowy, aplikacja bankowa — dane operacyjne w czasie rzeczywistym.

# Centralne repozytorium analityczne — serce klasycznego BI

---



## Dwa podejścia architektoniczne:

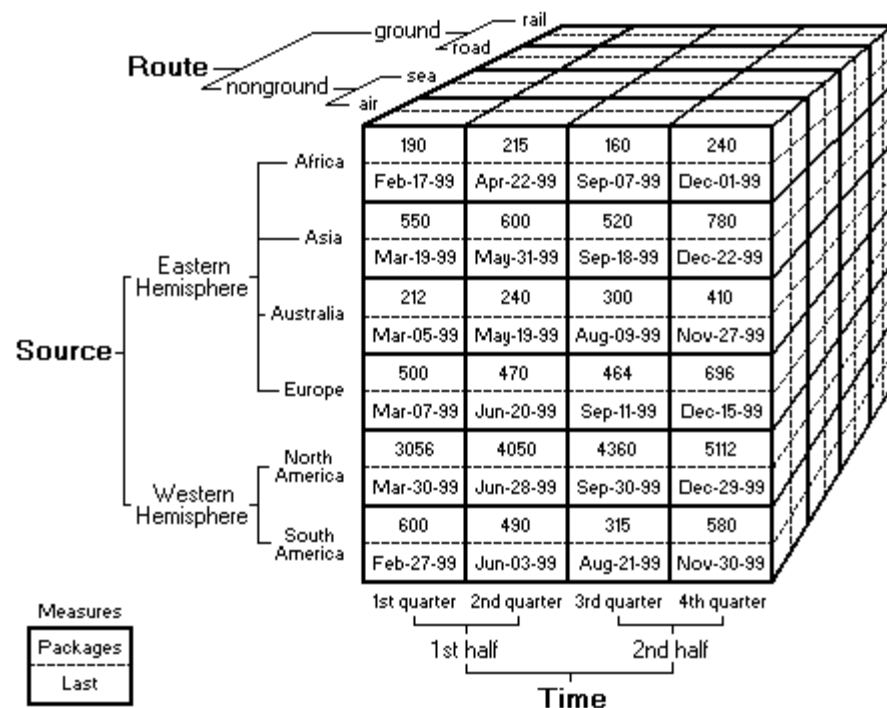
### Kimball (Bottom-Up)

- Zaczynij od Data Martow (departamentowych)
- Schemat gwiazdy (Star Schema)
- Szybsze wdrożenie, łatwa rozbudowa
- Stosowany w: Retail, e-commerce, finanse

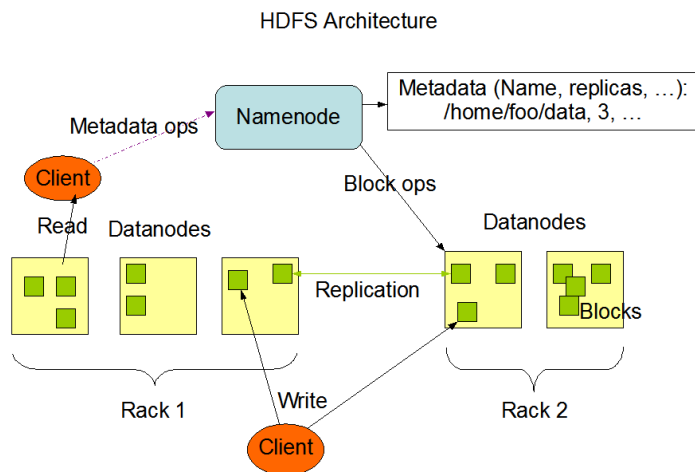
### Inmon (Top-Down)

- Najpierw centralny DWH (3NF), potem Marty
- Dane znormalizowane, jedna wersja prawdy
- Dłuższe wdrożenie, lepsza spójność
- Stosowany w: bankowość, ubezpieczenia

# Tabele w MOLAP



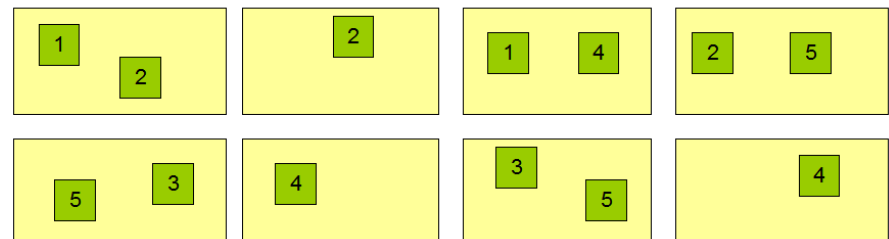
# Hadoop Distributed File System (HDFS)



## Block Replication

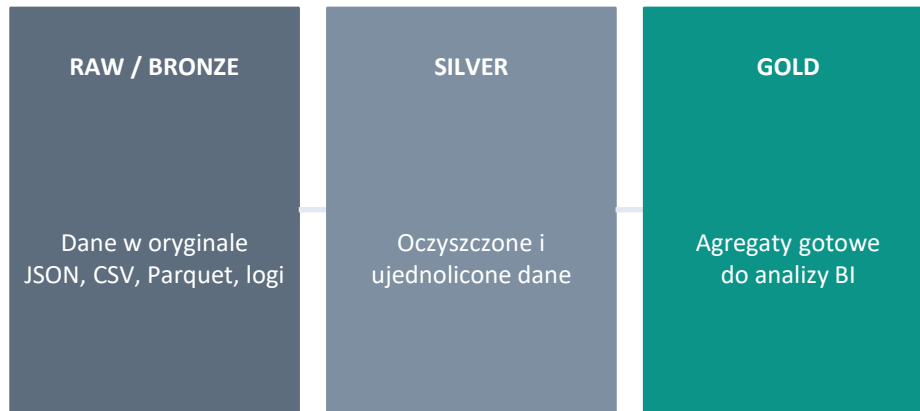
NameNode (Filename, numReplicas, block-ids, ...)  
/users/sameerp/data/part-0, r:2, {1,3}, ...  
/users/sameerp/data/part-1, r:3, {2,4,5}, ...

## Datanodes



# Surowe dane w natywnym formacie — schema-on-read

## Strefy (warstwy) Data Lake:



### Dane nieustrukturyzowane

Zdjęcia, wideo, teksty, logi, dane IoT, JSON — wszystko jest akceptowane bez wcześniejszego schematu.

### Schema-on-read

Schemat definiuje się przy odczycie, nie przy zapisie. Elastyczność vs ryzyko Data Swamp.

### Technologie

AWS S3 + Glue, Azure Data Lake Storage Gen2, GCS + Dataproc, Apache Hadoop / Spark, Delta Lake.

### Data Lakehouse

Ewolucja: Delta Lake / Apache Iceberg łączy elastyczność jezior z ACID transakcjami DWH.

# Zdecentralizowana architektura danych jako produktu

---

## 4 zasady Data Mesh (Zhamak Dehghani, 2019):

1

### Domain Ownership

Każda domena biznesowa (marketing, finanse, logistyka) jest właścicielem swoich danych i odpowiada za ich jakość.

2

### Data as a Product

Dane traktowane jak produkt: mają właściciela, dokumentację, SLA, testy jakości i API dla konsumentów.

3

### Self-serve Platform

Infrastruktura chmurowa umożliwia każdej domenie samodzielne publikowanie i konsumowanie danych bez IT.

4

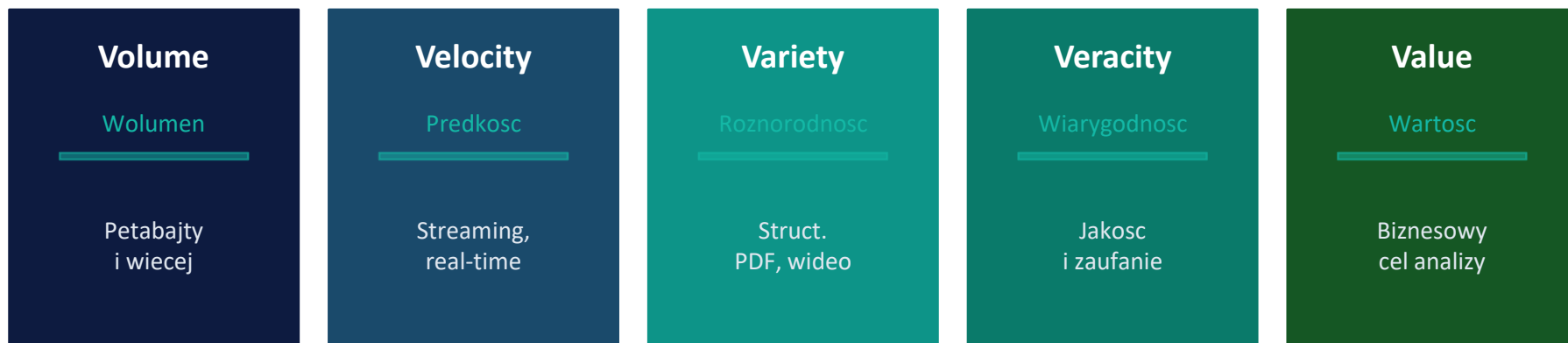
### Federated Governance

Federacyjne zarządzanie: globalne standardy (bezpieczeństwo, prywatność) + lokalna autonomia domen.

*Data Mesh to NIE jest technologia — to organizacyjny paradygmat zarządzania danymi.*

# Kiedy dane są za duże, za szybkie lub za różnorodne dla klasycznych narzędzi

## Model 5V Big Data:



## Stos technologiczny Big Data:

| Przetwarzanie batch                                                         | Streaming / real-time                                                                     | Skladowanie                                                                   |
|-----------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------|
| Apache Hadoop / MapReduce<br>Apache Spark (PySpark)<br>Hive — SQL na Hadoop | Apache Kafka (kolejki)<br>Apache Flink / Spark Streaming<br>AWS Kinesis, Azure Event Hubs | HDFS, S3, ADLS Gen2<br>Delta Lake, Apache Iceberg<br>Cassandra, HBase (NoSQL) |

# Porównanie — kiedy używać jakiego podejścia?

| Kryterium  | Baza danych   | Hurtownia    | Data Lake   | Data Mesh     | Big Data    |
|------------|---------------|--------------|-------------|---------------|-------------|
| Dane       | Ustrukturyz.  | Ustrukturyz. | Dowolne     | Dowolne       | Duże / mix  |
| Skala      | GB–TB         | TB–PB        | PB+         | Dowolna       | PB+         |
| Latencja   | ms            | sekundy      | minuty      | Różna         | sekundy     |
| Zapytania  | OLTP / SQL    | OLAP / MDX   | Spark / SQL | API domeny    | Spark/Flink |
| Główny cel | Transakcje    | Raportowanie | Eksploracja | Samodzielność | Analityka   |
| Dojrzałość | Bardzo wysoka | Wysoka       | Średnia     | Niska/rośnie  | Wysoka      |

Nowoczesne architektury (Lakehouse, Data Mesh) łączą zalety wielu podejść w jednym ekosystemie.

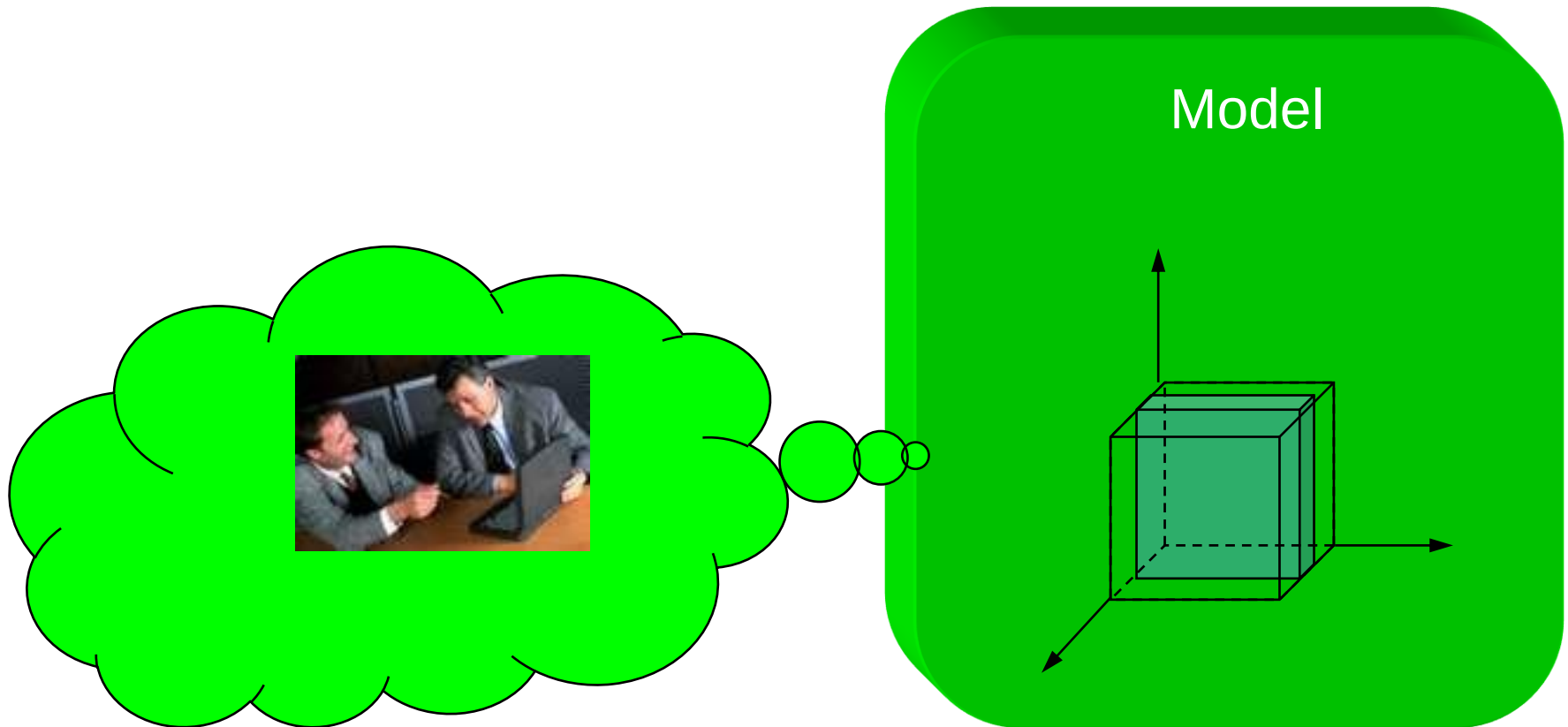
# OLAP - Definicja

---

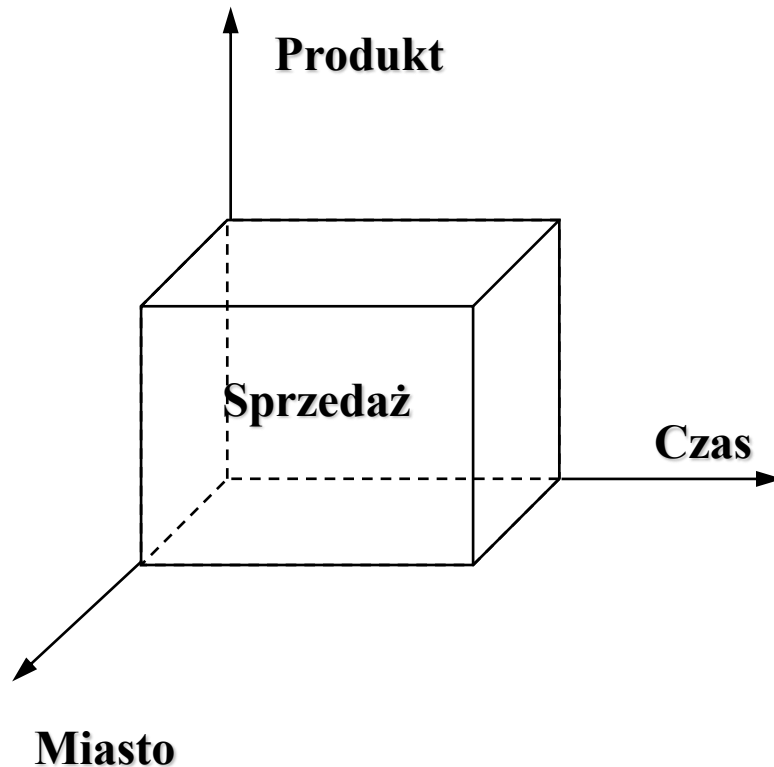
OLAP – On-Line Analytical Processing

Narzędzie informatyczne przeznaczone  
do wielowymiarowych i hierarchicznych  
analiz danych

# OLAP jako model myślowy



# Model wielowymiarowy (widok biznesowy)



**Wymiary**

Czas

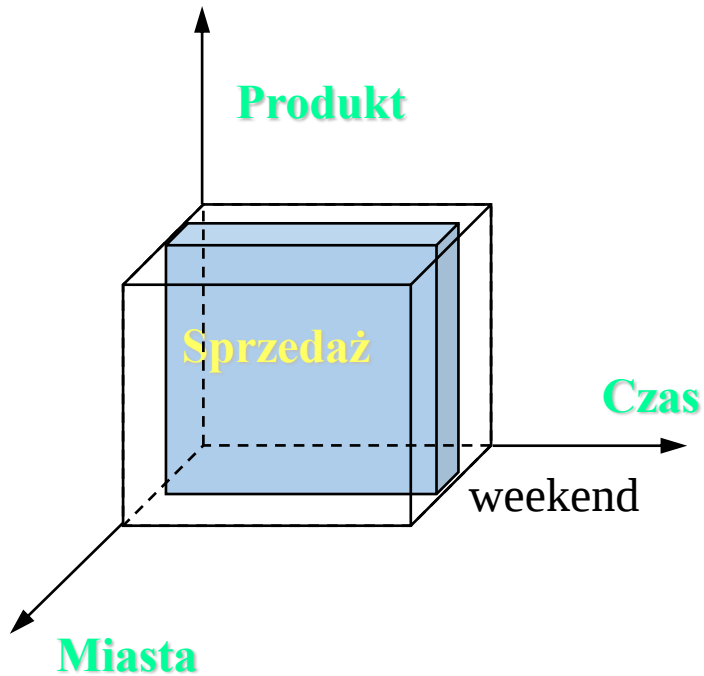
Produkt

Miasto

**Miary**

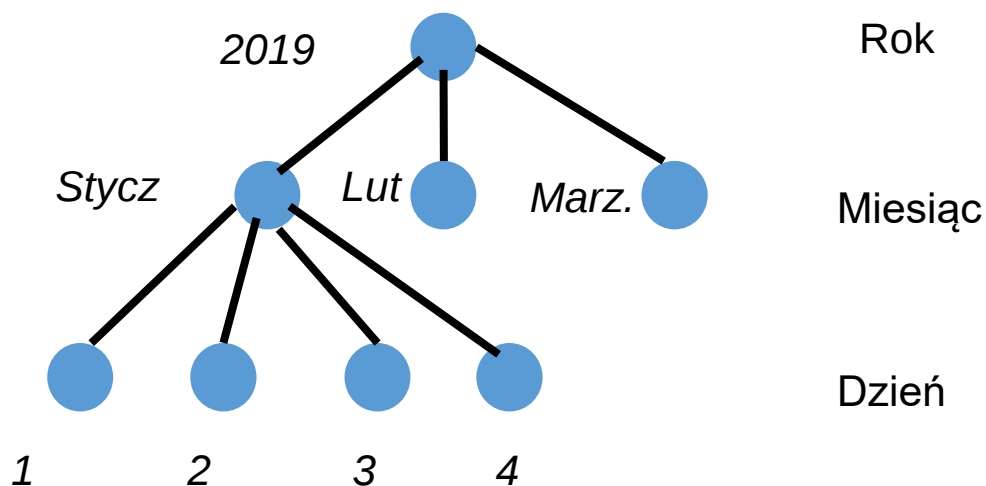
Wielkość sprzedaży

# Arkusz analiz



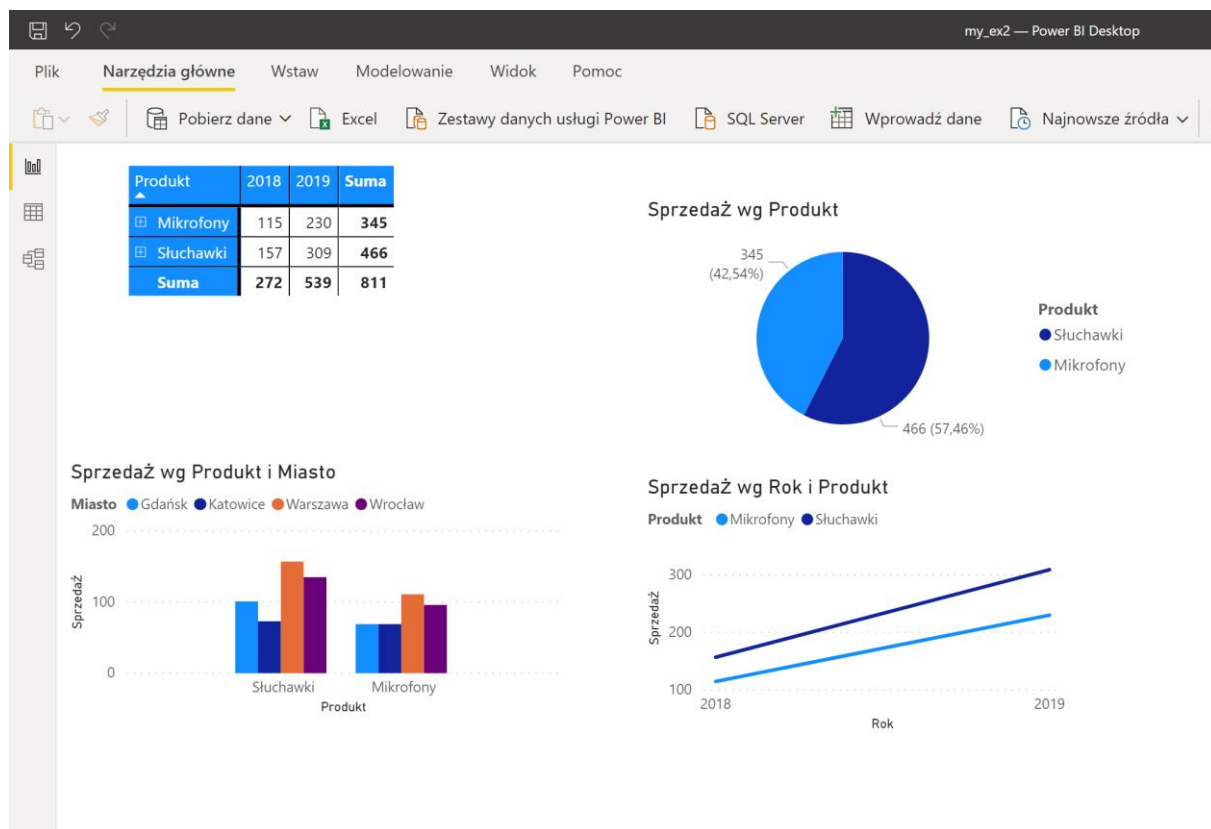
| Produkt<br>▼ | 2018 | 2019 | Suma |
|--------------|------|------|------|
| Słuchawki    | 157  | 309  | 466  |
| Mikrofony    | 115  | 230  | 345  |
| Suma         | 272  | 539  | 811  |

# Hierarchia

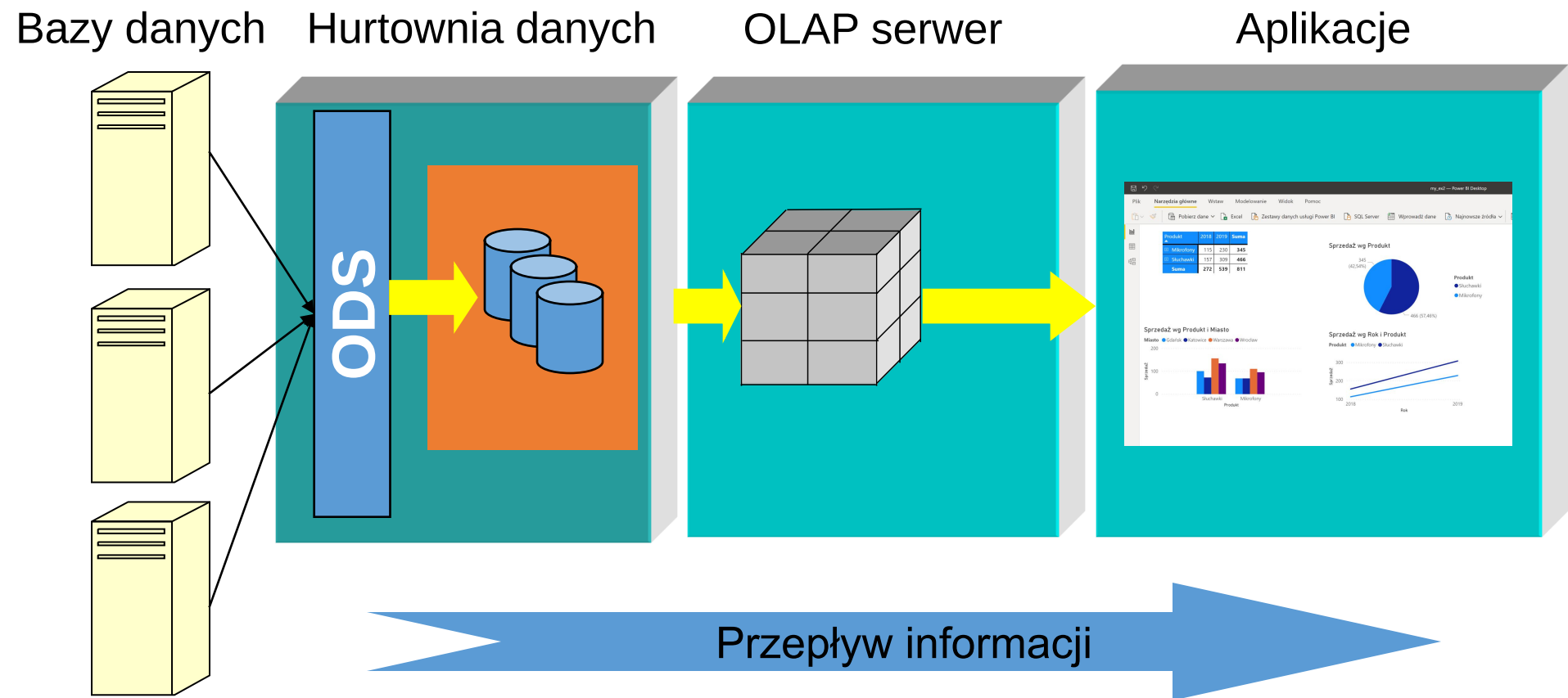


| Rok          | 2018 |         |      | 2019 |         |      | Suma |
|--------------|------|---------|------|------|---------|------|------|
| Produkt<br>▼ | Luty | Styczeń | Suma | Luty | Styczeń | Suma |      |
| Słuchawki    | 66   | 91      | 157  | 134  | 175     | 309  | 466  |
| Mikrofony    | 65   | 50      | 115  | 130  | 100     | 230  | 345  |
| Suma         | 131  | 141     | 272  | 264  | 275     | 539  | 811  |

# Kokpit zarządzania



# Umieszczenie w architekturze systemu analitycznego



# Przegląd rozwiązań BI

# Wybrane rozwiązania

---

- Rozwiązania dla małej ilości danych
  - *MS Excel*
- Rozwiązania z łatwo konfigurowalnymi analizami
  - *Power BI, QlikView, Tableau*
- Analityka z wykorzystaniem chmury
  - *Qubole*
- Analityka z wykorzystaniem BigData
  - *Apache Kylin*
- Rozwiązania z zaawansowaną analityką
  - *SAS, IBM, SAP*

# Tabela przestawna - EXCEL

| Dzień | Miesiąc  | Rok  | Miasto   | Produkt   | Sprzedaż |
|-------|----------|------|----------|-----------|----------|
| 1     | Grudzień | 2009 | Warszawa | Słuchawki | 10       |
| 1     | Grudzień | 2009 | Katowice | Słuchawki | 1        |
| 1     | Grudzień | 2009 | Wrocław  | Słuchawki | 5        |
| 1     | Grudzień | 2009 | Gdańsk   | Słuchawki | 8        |
| 2     | Grudzień | 2009 | Warszawa | Słuchawki | 10       |
| 2     | Grudzień | 2009 | Katowice | Słuchawki | 1        |
| 2     | Grudzień | 2009 | Wrocław  | Słuchawki | 5        |
| 2     | Grudzień | 2009 | Gdańsk   | Słuchawki | 8        |
| 3     | Grudzień | 2009 | Warszawa | Słuchawki | 7        |
| 3     | Grudzień | 2009 | Katowice | Słuchawki | 2        |
| 3     | Grudzień | 2009 | Wrocław  | Słuchawki | 6        |
| 3     | Grudzień | 2009 | Gdańsk   | Słuchawki | 6        |
| 4     | Grudzień | 2009 | Warszawa | Słuchawki | 11       |
| 4     | Grudzień | 2009 | Katowice | Słuchawki | 4        |
| 4     | Grudzień | 2009 | Wrocław  | Słuchawki | 6        |
| 4     | Grudzień | 2009 | Gdańsk   | Słuchawki | 1        |
| 1     | Styczeń  | 2010 | Warszawa | Słuchawki | 10       |
| 1     | Styczeń  | 2010 | Katowice | Słuchawki | 1        |
| 1     | Styczeń  | 2010 | Wrocław  | Słuchawki | 5        |
| 1     | Styczeń  | 2010 | Gdańsk   | Słuchawki | 8        |
| 2     | Styczeń  | 2010 | Warszawa | Słuchawki | 10       |
| 2     | Styczeń  | 2010 | Katowice | Słuchawki | 1        |
| 2     | Styczeń  | 2010 | Wrocław  | Słuchawki | 5        |
| 2     | Styczeń  | 2010 | Gdańsk   | Słuchawki | 8        |
| 3     | Styczeń  | 2010 | Warszawa | Słuchawki | 7        |
| 3     | Styczeń  | 2010 | Katowice | Słuchawki | 2        |
| 3     | Styczeń  | 2010 | Wrocław  | Słuchawki | 6        |
| 3     | Styczeń  | 2010 | Gdańsk   | Słuchawki | 6        |
| 4     | Styczeń  | 2010 | Warszawa | Słuchawki | 11       |



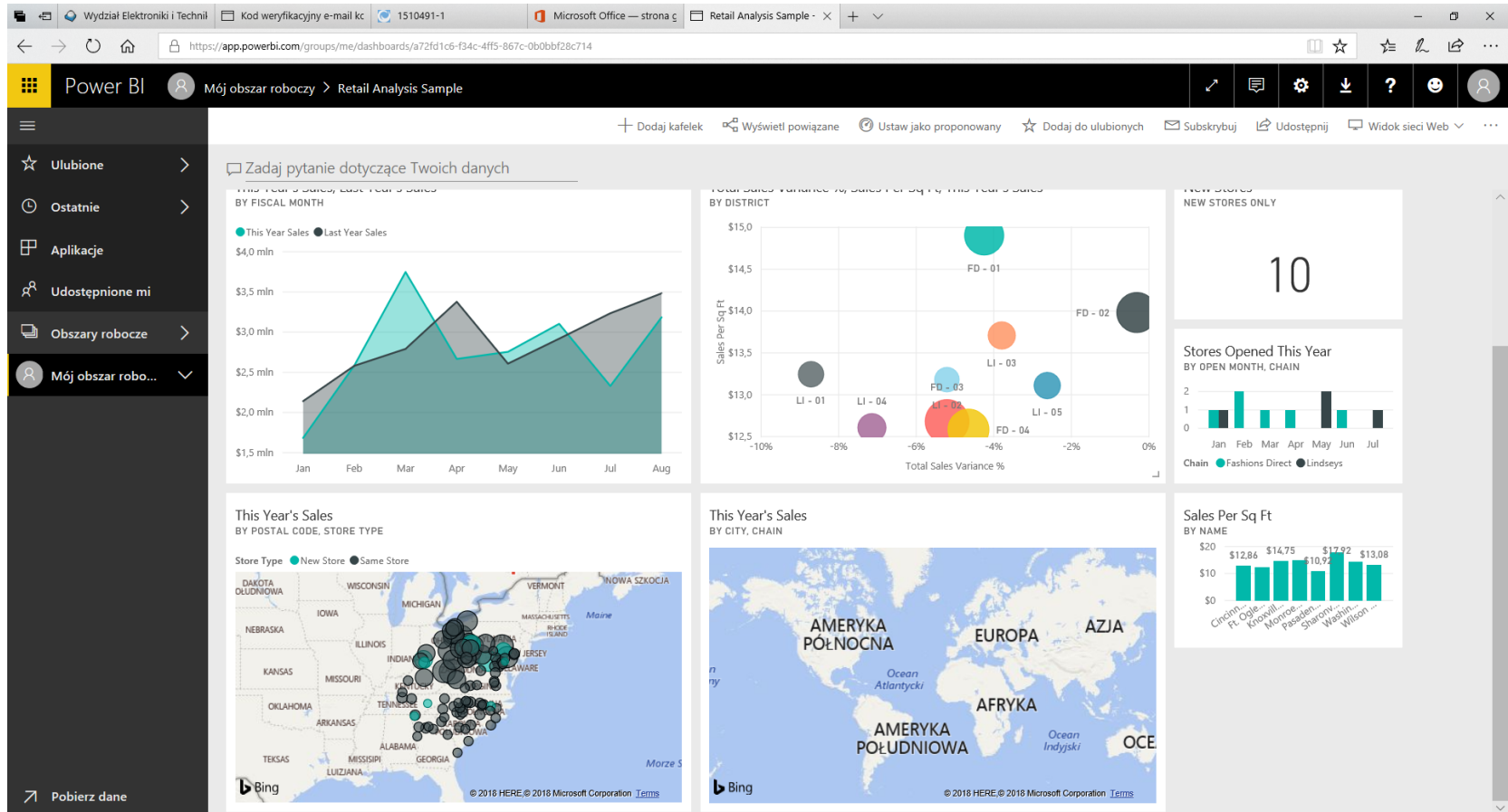
|                 |         |      |              |
|-----------------|---------|------|--------------|
| Rok             | 2010    |      |              |
| Suma z Sprzedaż | Miesiąc |      |              |
| Produkt         | Styczeń | Luty | Suma końcowa |
| Mikrofony       | 50      | 65   | 115          |
| Słuchawki       | 91      | 66   | 157          |
| Suma końcowa    | 141     | 131  | 272          |

# Tabela przestawna - Excel

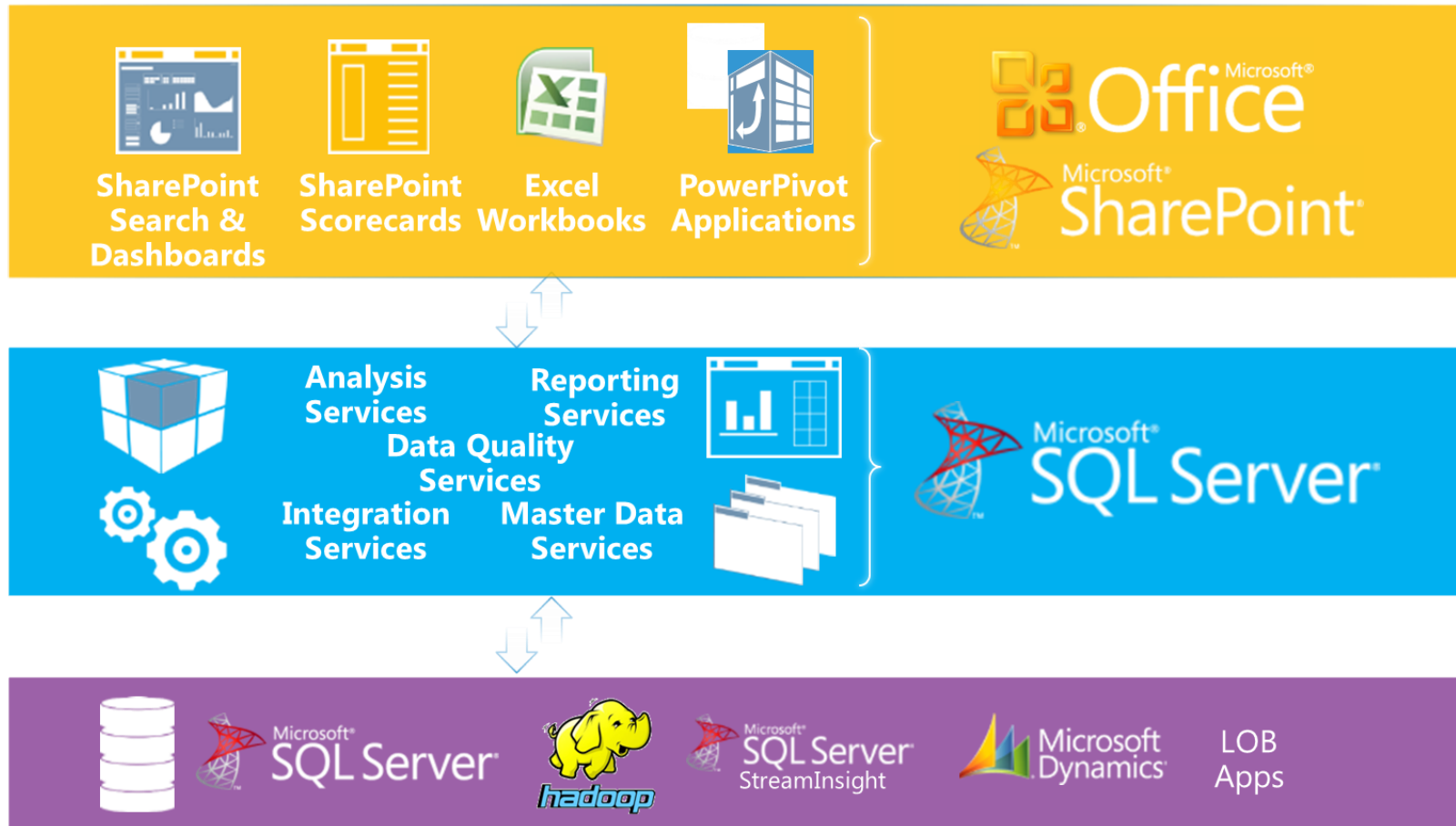
---

- Zalety
  - szybkie analizy
  - znane narzędzie
- Wady
  - Zarządzanie danymi
  - Mała elastyczność

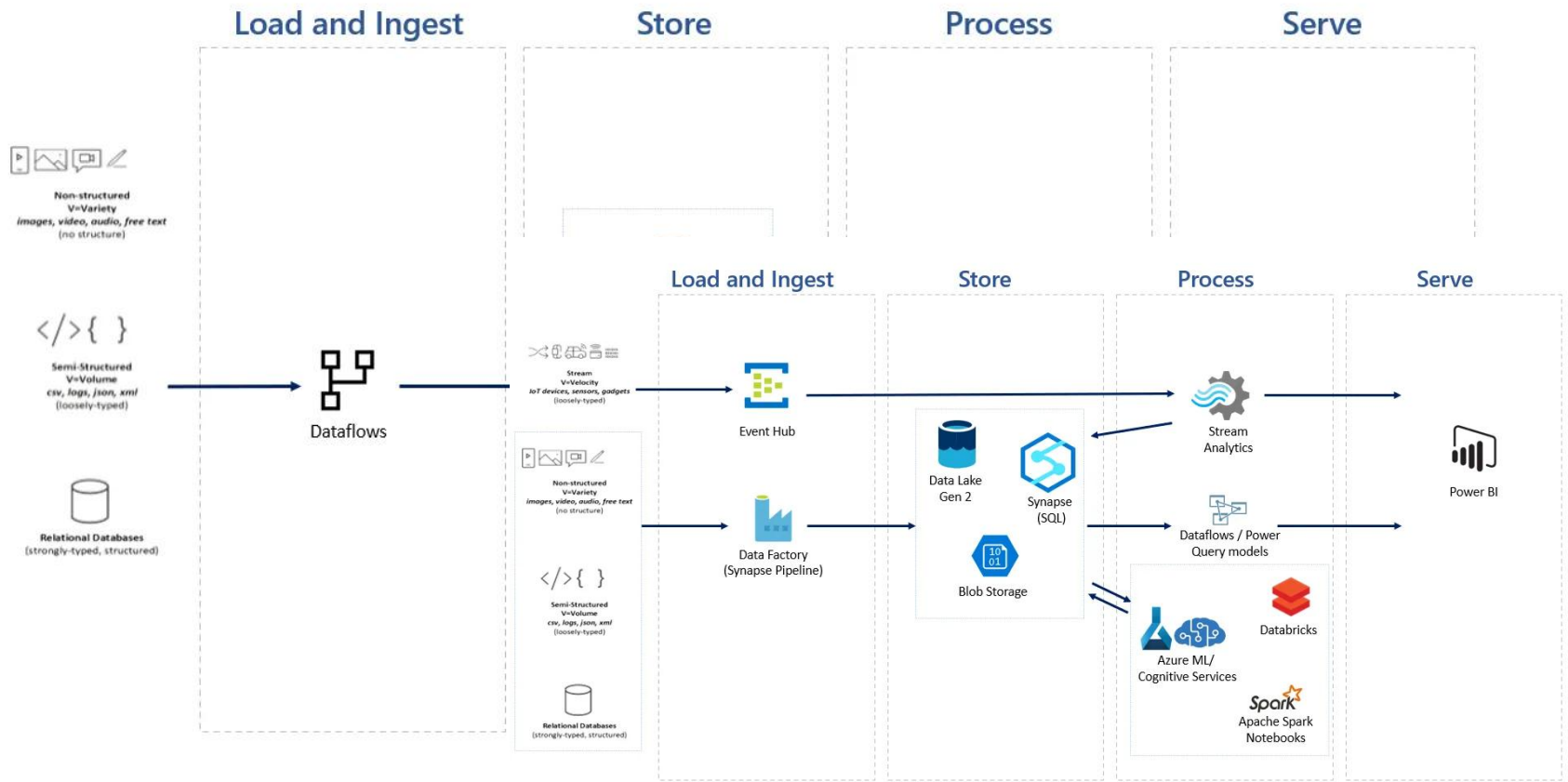
# Power BI (MICROSOFT)



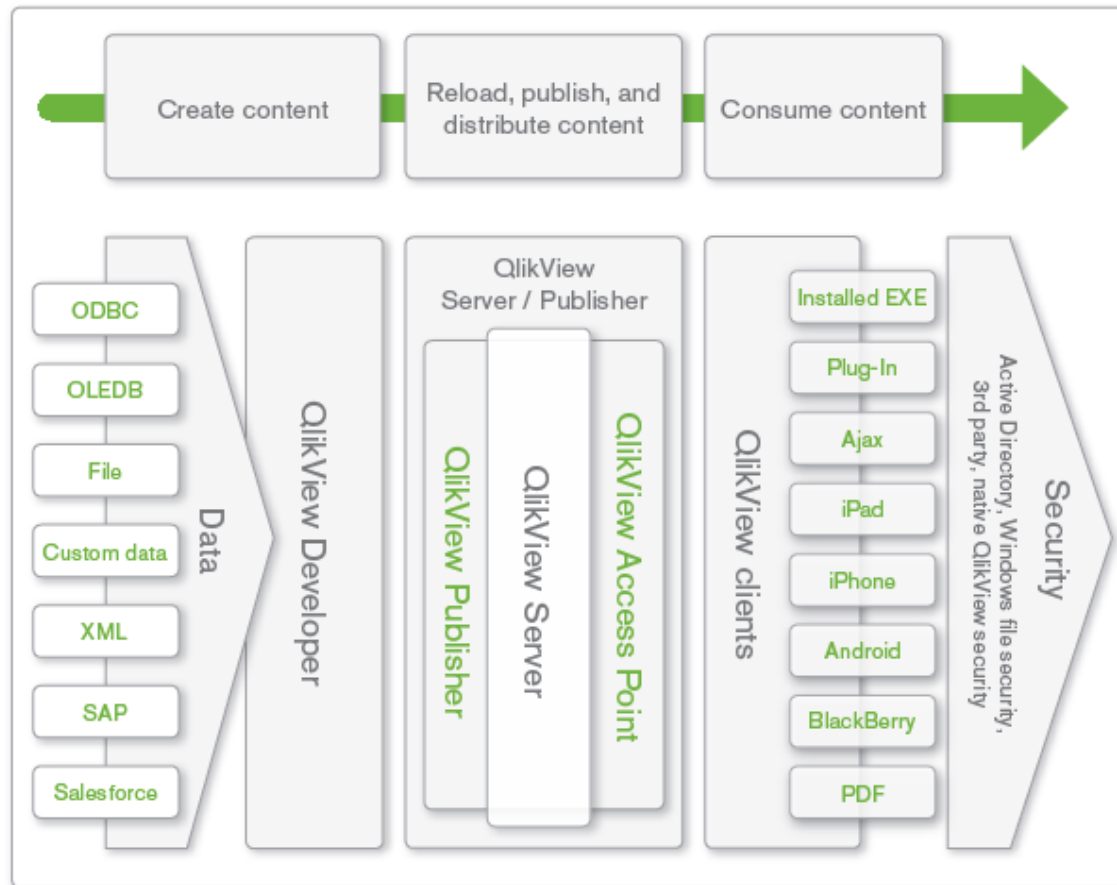
# MICROSOFT



# Power BI - architektura

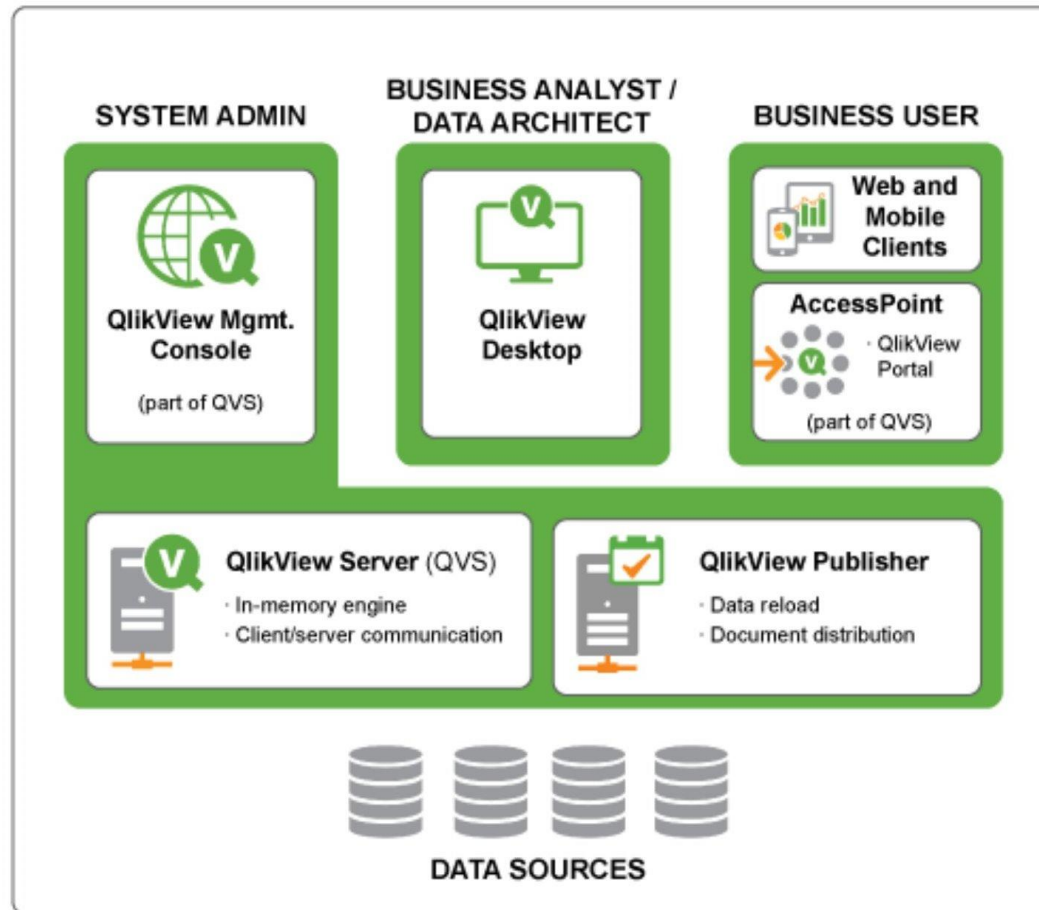


# QlikView

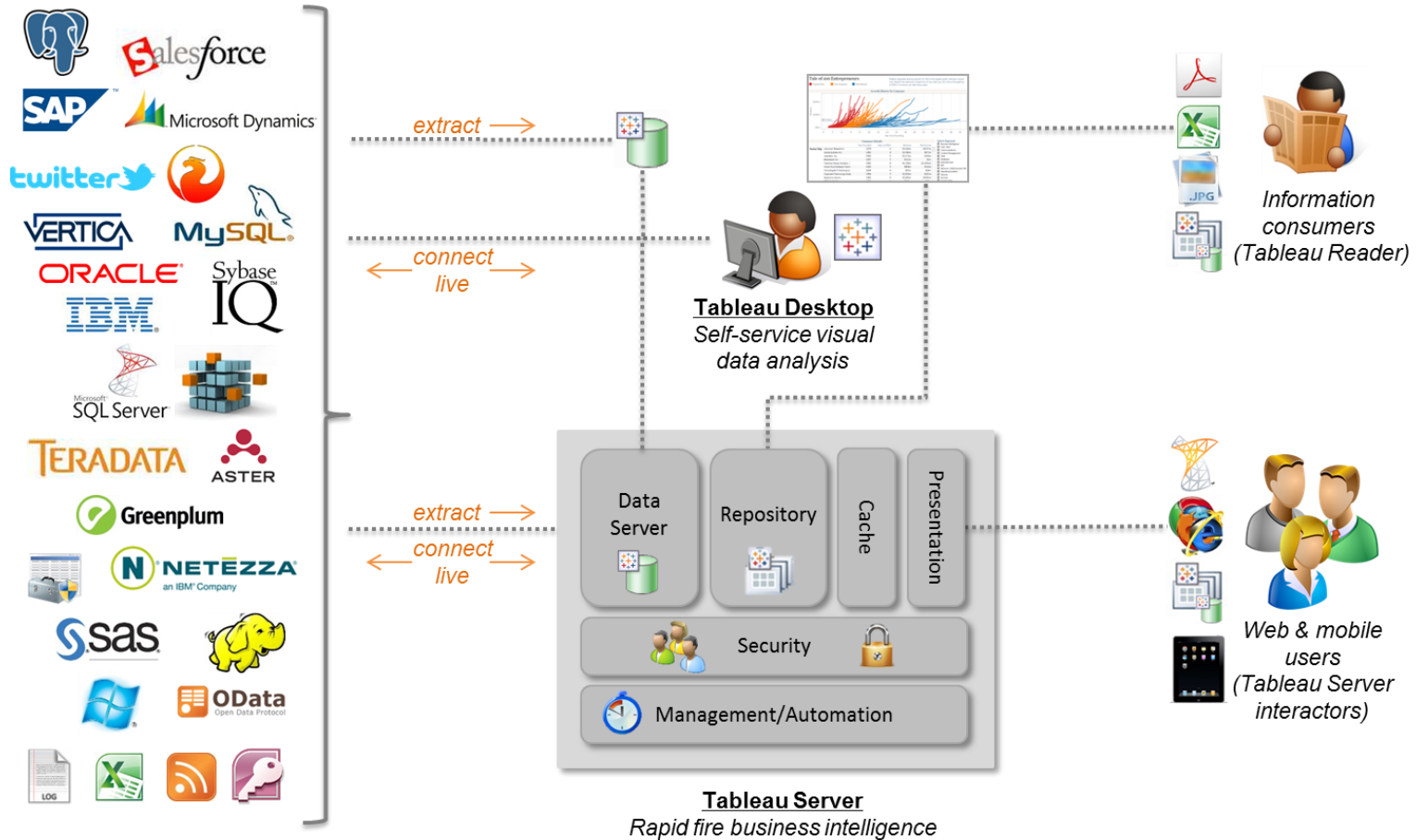


In-memory

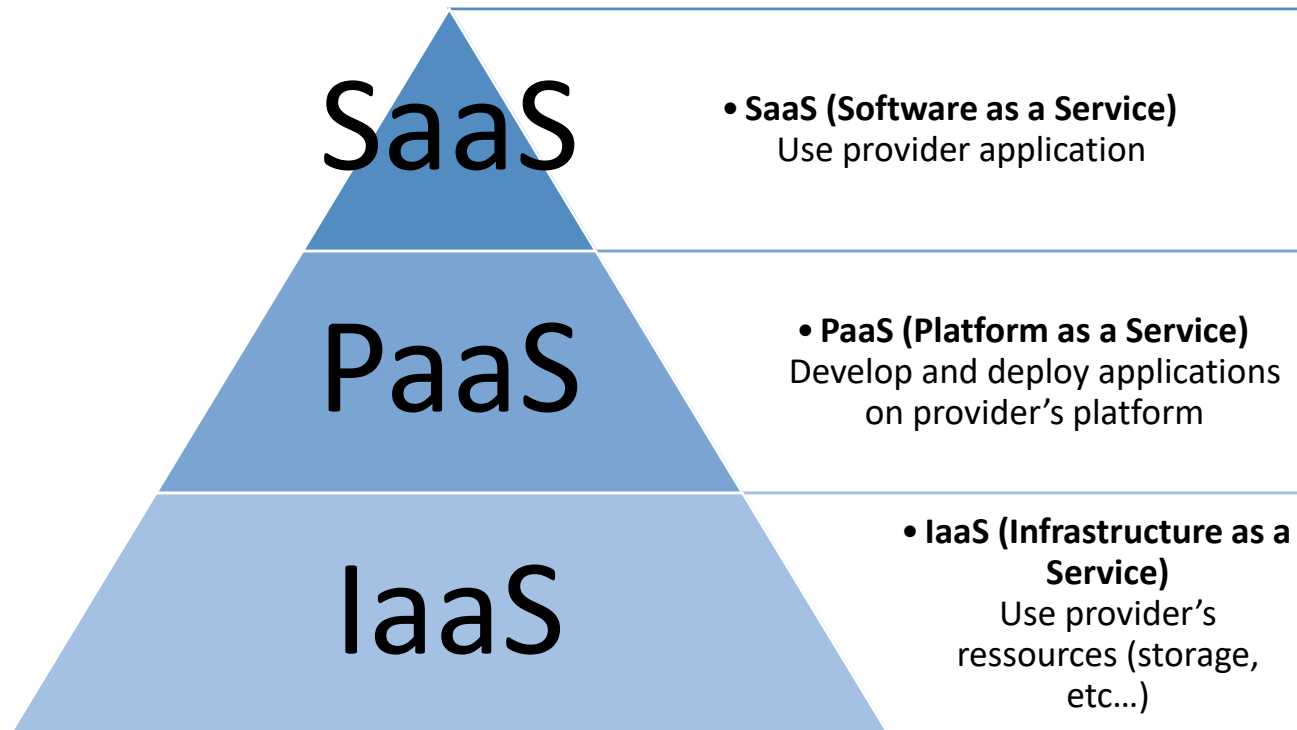
# QlikView



# Tableau



# Rozwiązania chmurowe

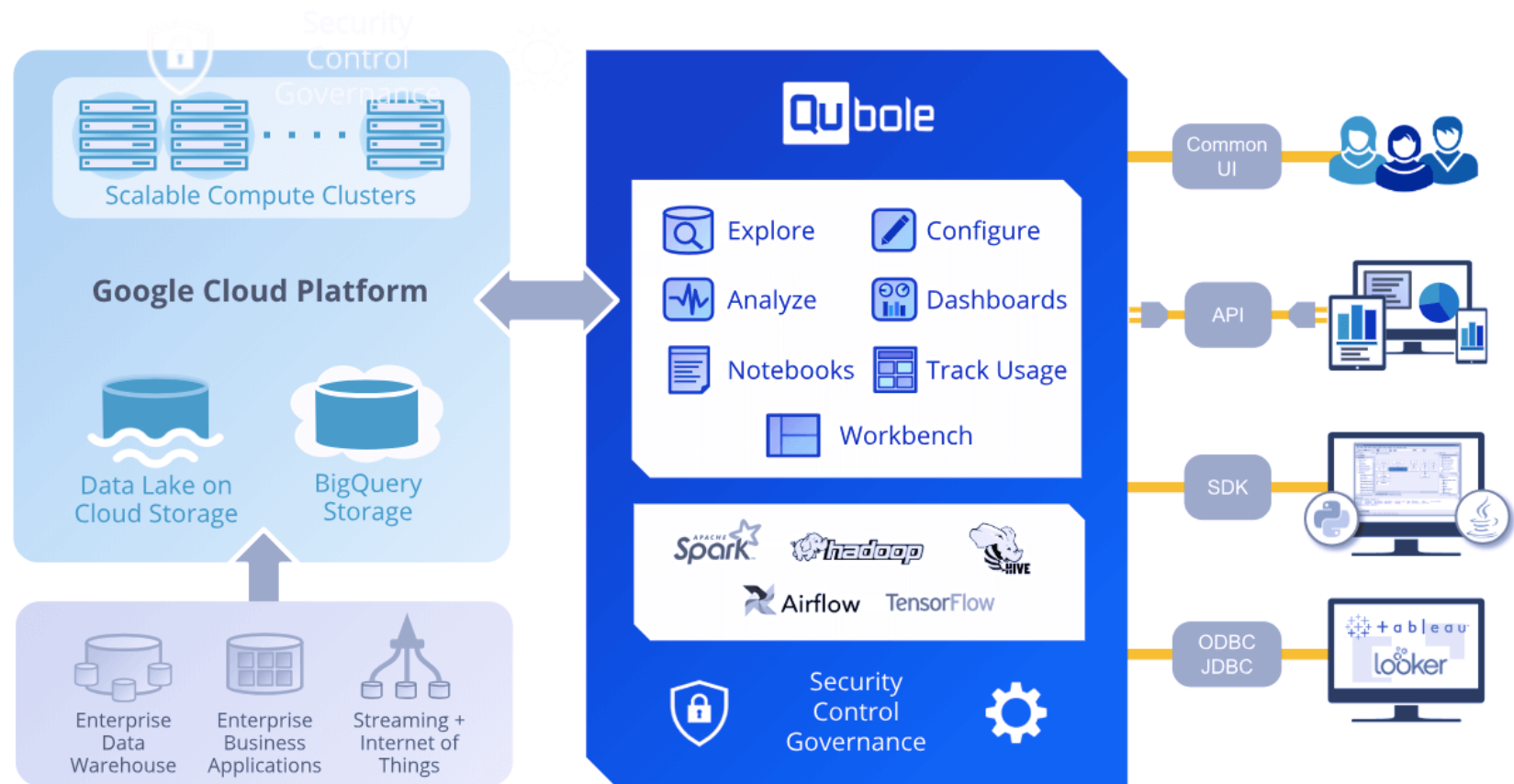


# Rozwiązania analityczne w chmurze

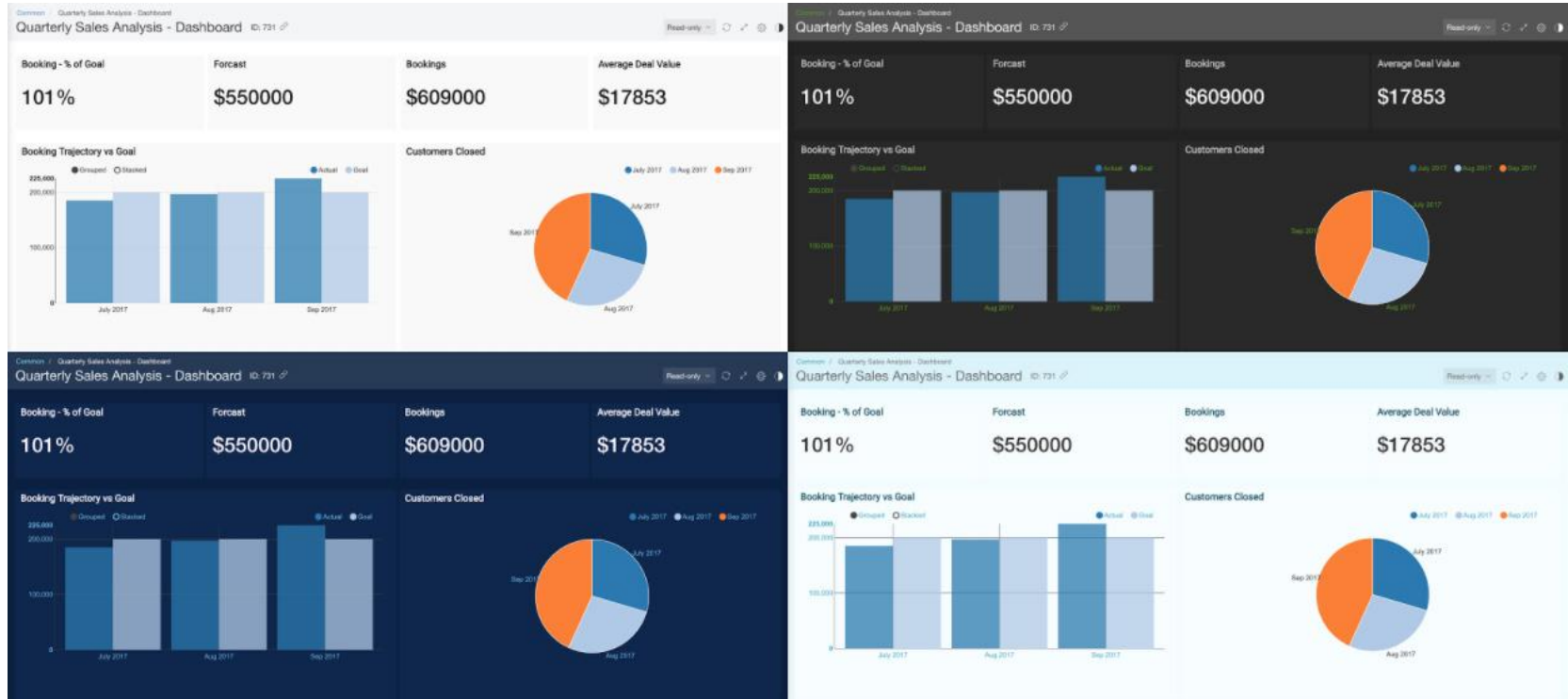
---

- Sisense
- Periscope data
- Microsoft Power BI Pro
- IBM Cognos Analytics
- Amazon QuickSight
- Tableau online
- SAP Business Objects BI
- Looker
- Qubole
- ...

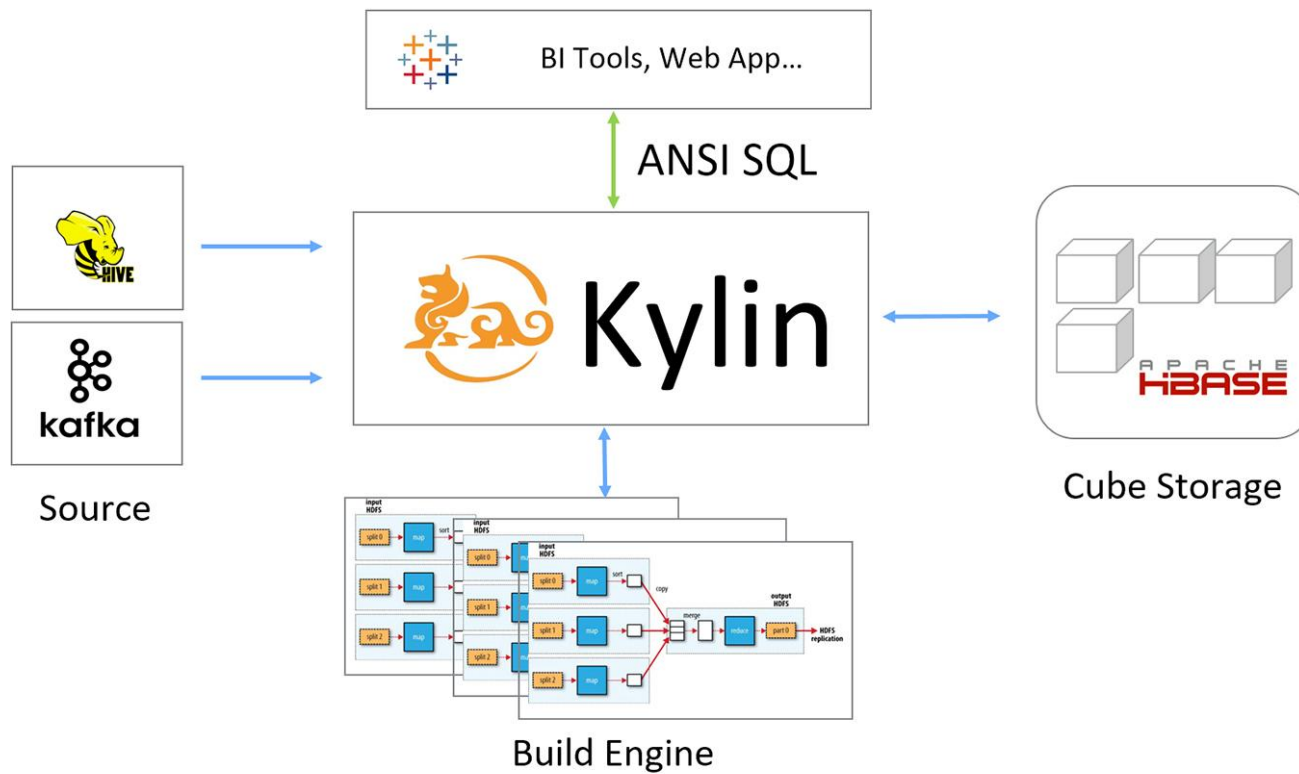
# Qubole



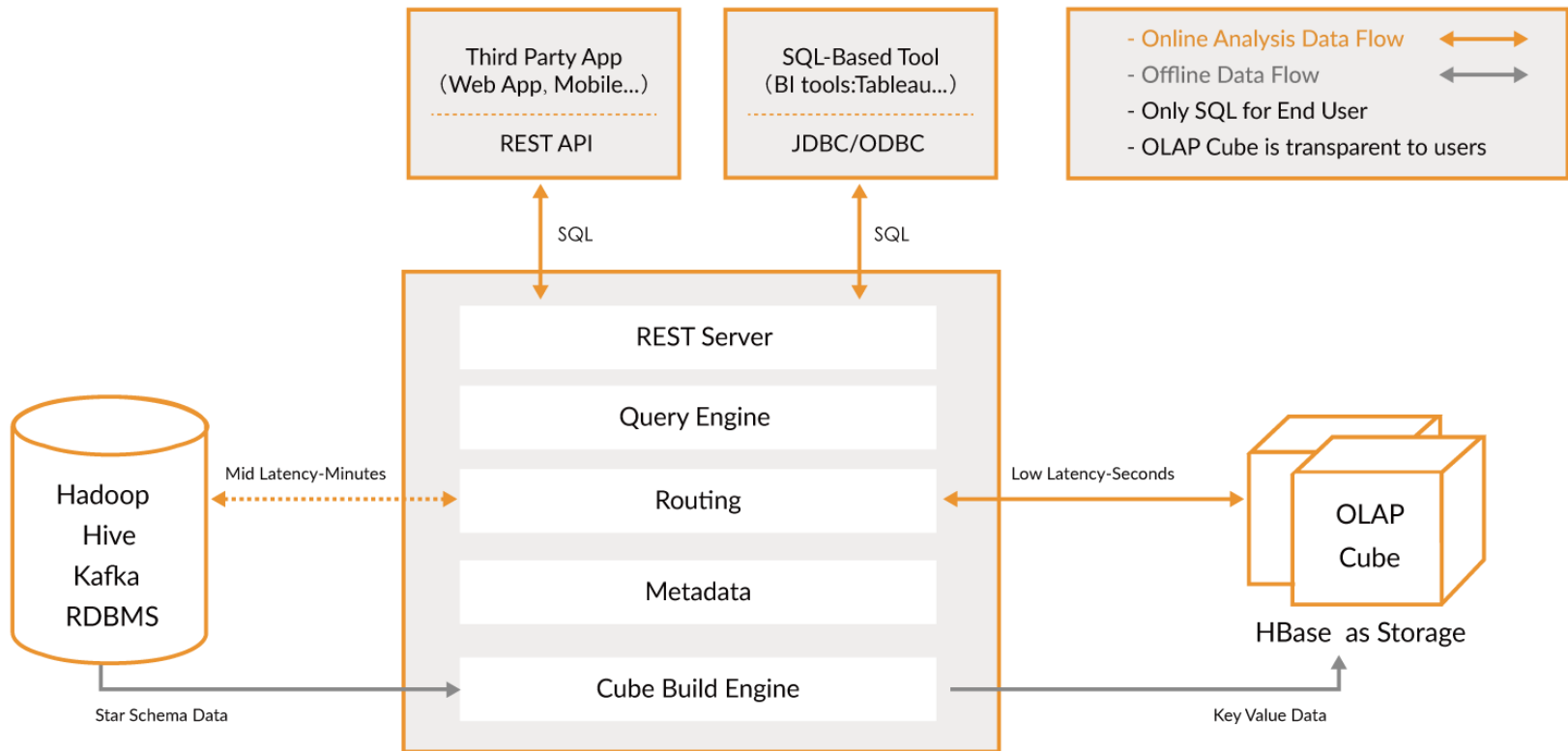
# Qubole analytics



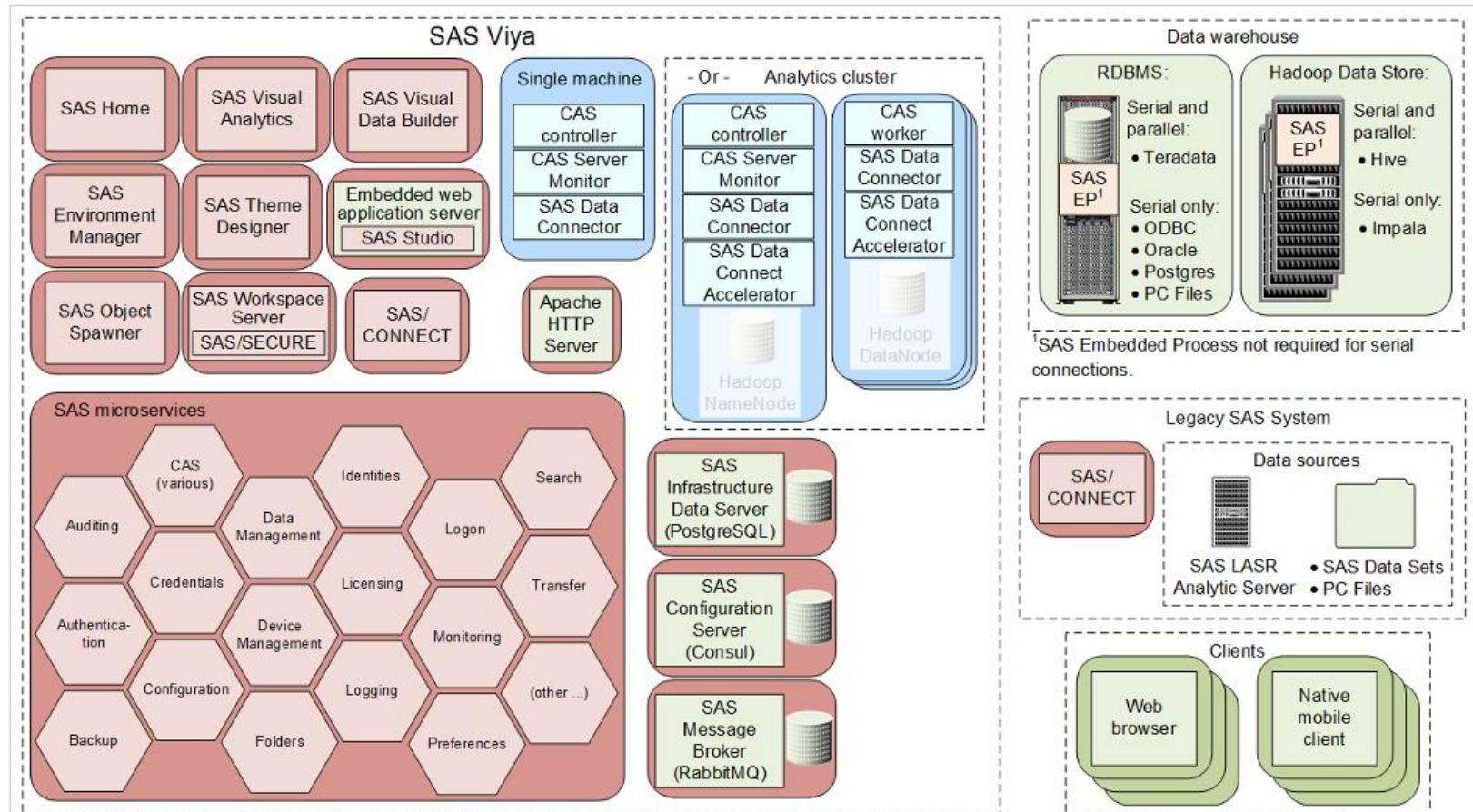
# Architektura- Apache Kylin™



# Apache Kylin™



# Architektura SAS Viya



## IBM Business Analytics

Industry  
Solutions



**Financial  
Services**



**Public**



**Distribution**



**Industrial**



**Communications**

Functional  
Solutions

Customer  
Acquisition

Social Media  
Analytics

Budgeting &  
Forecasting

Financial  
Consolidation

Resource  
Optimization

Asset  
Management

Risk  
Identification

Risk & Control  
Assessment

Customer  
Lifetime  
Value

Customer  
Loyalty  
& Retention

Disclosure  
Management

Profitability  
Modeling &  
Optimization

Production  
Planning

Decision  
Management

Risk Mitigation  
Planning

Risk Aware  
Decisioning

Core  
Capabilities

**Report**

**Analyze**

**Model**

**Plan**

**Collaborate**

**Predict**

Forecast

Visualize

Decide

Govern

Discover

Score

Mine

Simulate

Survey

Software  
Categories



**Business  
Intelligence**



**Predictive  
Analytics**

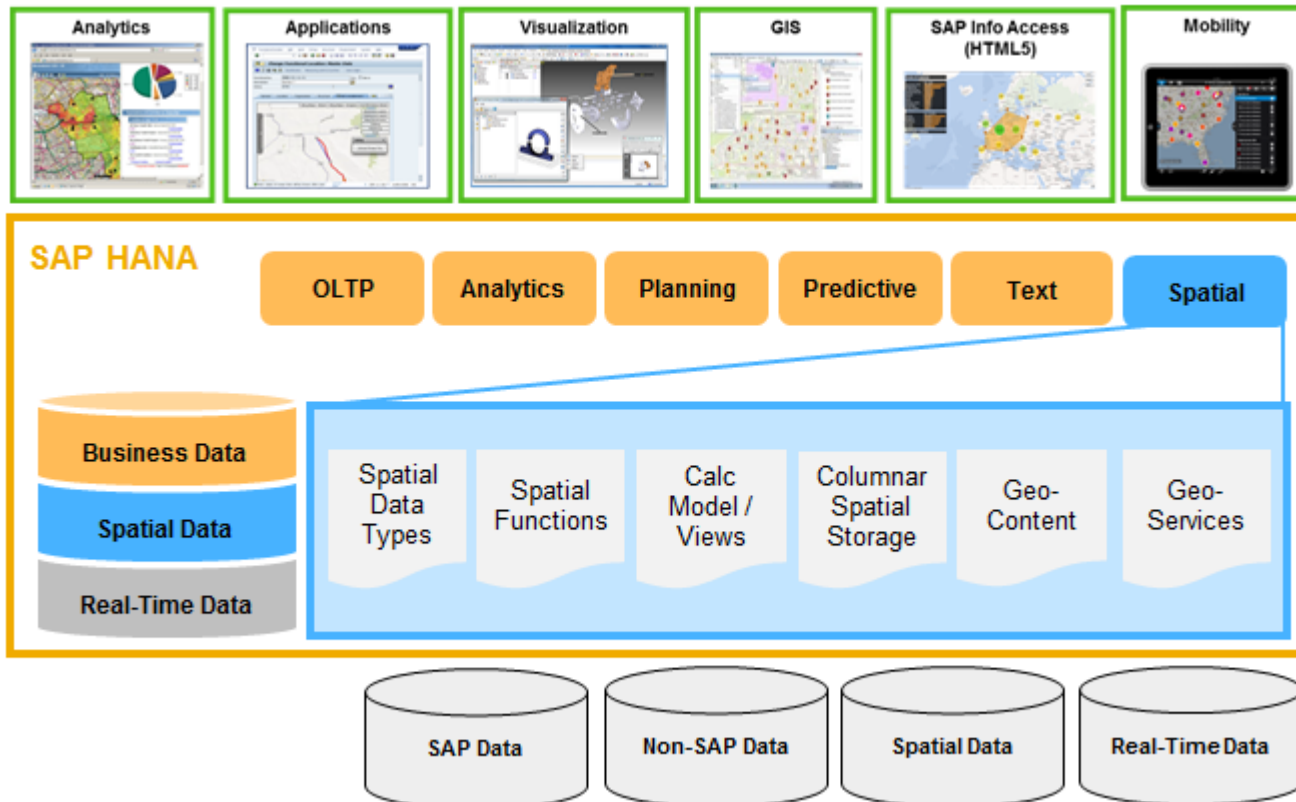


**Performance  
Management**



**Risk  
Analytics**

# Sap HANA



# Open Source OLAP

---

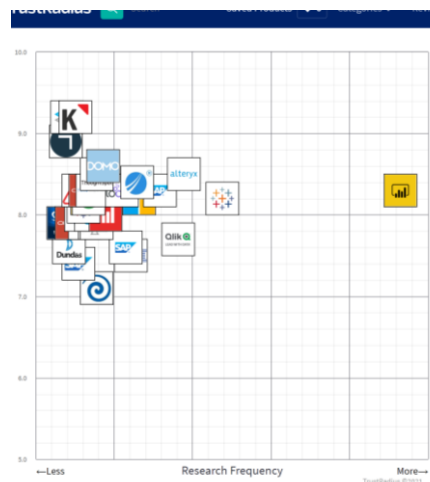
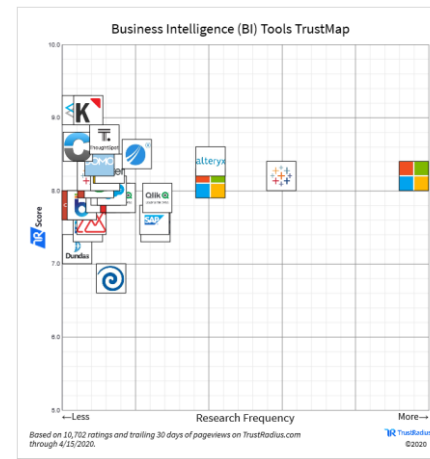
- Pentaho from Hitachi Vantara
- JasperReports Library
- Mondrian
- Cubes
- Apache Kylin™

# Analiza dostępnych narzędzi

Figure 1: Magic Quadrant for Analytics and Business Intelligence Platforms



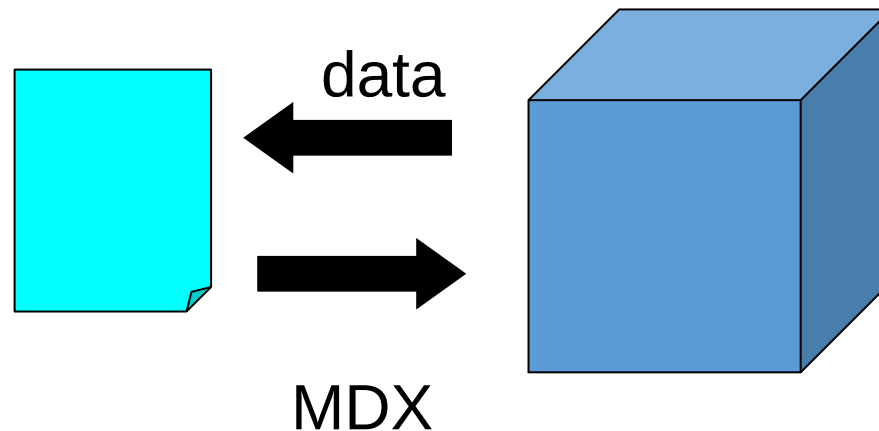
Source: Gartner (February 2021)



Klipfolio,  
InsightSquared Sales  
Analytics,  
ThoughtSpot,  
Cyfe,  
TIBCO Spotfire,  
Alteryx Platform,  
Domo,  
Looker,  
Sisense,  
Microsoft Power BI,  
Tableau Desktop,  
TIBCO JasperSoft,  
Tableau Online,  
Microsoft BI,  
Google Charts,  
Oracle Analytics Cloud.

# MDX (MultiDimensional eXpressions )

---



# MultiDimensional eXpressions (MDX)

---

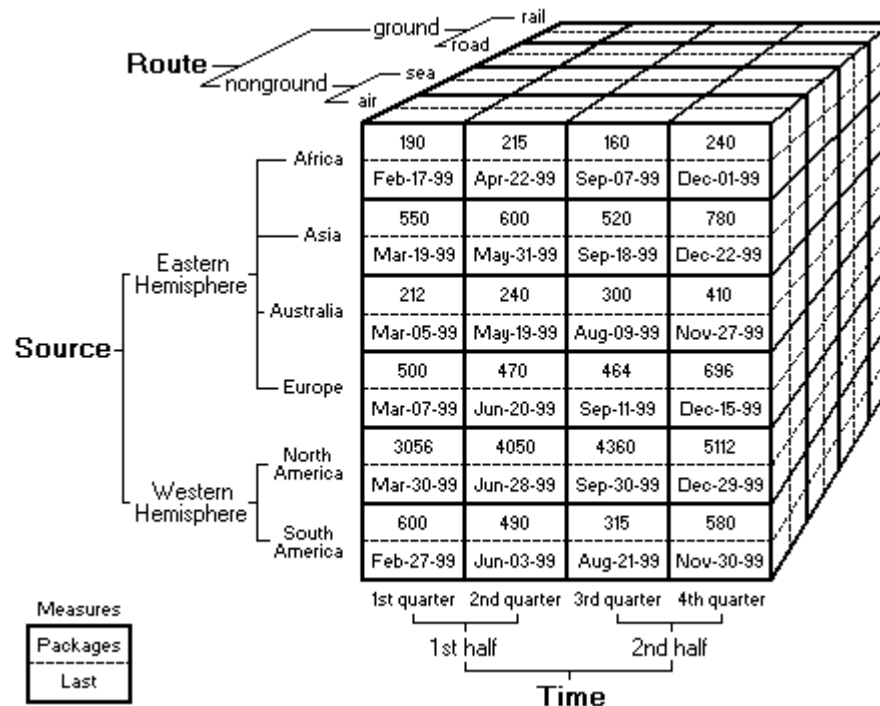
[WITH <specyfikacja formuły> [,<specyfikacja formuły> ...]]

SELECT [<specyfikacja osi> [,<specyfikacja osi> ...]]

FROM [<specyfikacja kostki>]

[WHERE [<specyfikacja warunków>]]

# Dostęp do wymiarów



(Source.[Eastern Hemisphere].Africa,  
Time.[2nd half].[4th quarter], Route.Air, Measures.Packages)

# Przykład analizy I

# Przykład – Analiza wydatków IT

---

IT Spend Analysis (pulpit nawigacyjny, raport i zestaw danych) służy do analizy planowanych i rzeczywistych kosztów działu IT.

Analiza pozwala zrozumieć, jak dobrze firma zaplanowała rok oraz zbadać obszary wykazujące duże odchylenia od planu.

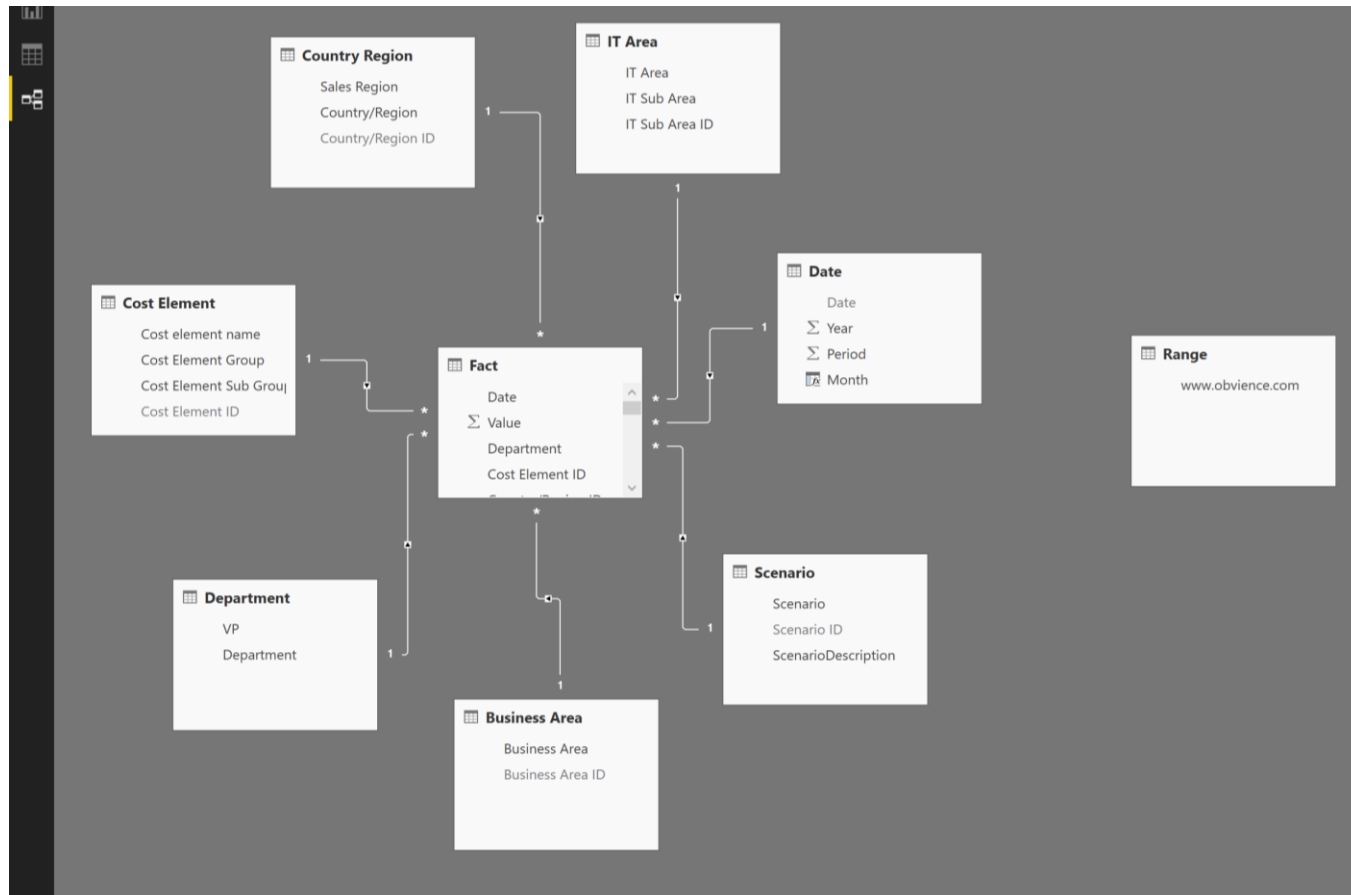
Firma w tym przykładzie ma roczny cykl planowania, a następnie co kwartał wydaje nowe oszacowanie kosztów, aby pomóc w analizowaniu zmian w wydatkach na zasoby informatyczne w roku obrachunkowym.

Prawdziwe dane pochodzące z firmy obviEnce (<http://obvience.com/>), przedstawione w sposób anonimowy

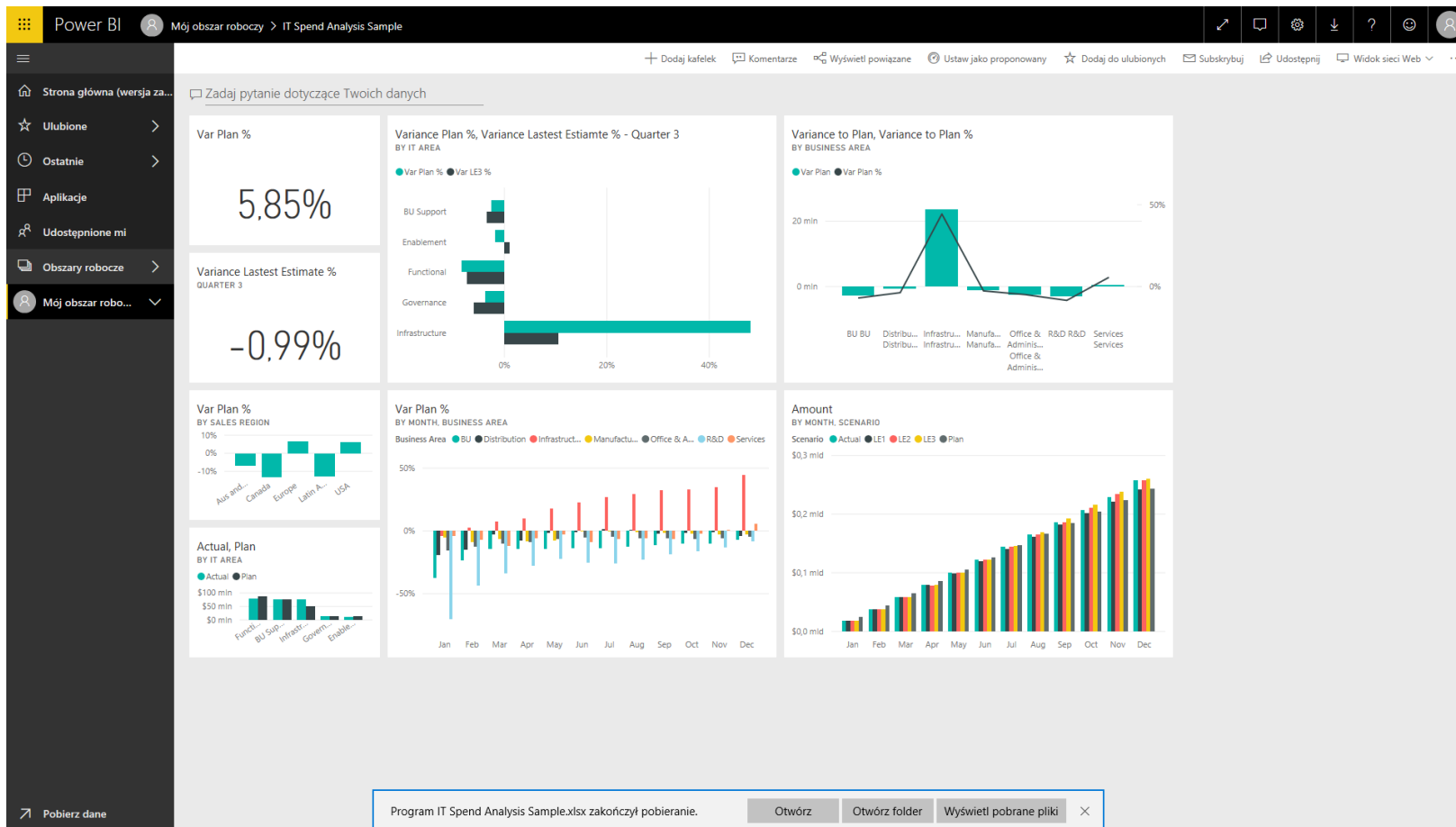
# Dane

| Month | Var Plan % | Var Plan            | Actual      | LE1         | LE2         | LE3         | Plan        |
|-------|------------|---------------------|-------------|-------------|-------------|-------------|-------------|
| Jan   | -27,29%    | -6600558.762000006  | \$17585804  | \$17613109  | \$17585804  | \$17585804  | \$24186363  |
| Feb   | -15,75%    | -6968695.622999974  | \$37275296  | \$37353501  | \$37275296  | \$37275296  | \$44243992  |
| Mar   | -9,82%     | -6419827.122000001  | \$58936954  | \$58952338  | \$58875173  | \$58936954  | \$65356781  |
| Apr   | -8,25%     | -7085575.172999933  | \$78846303  | \$79473261  | \$78784522  | \$78846303  | \$85931878  |
| May   | -4,37%     | -4608780.914999932  | \$100779875 | \$98903492  | \$100718094 | \$100779875 | \$105388656 |
| Jun   | -2,79%     | -3520328.366999954  | \$122816815 | \$119716778 | \$122755034 | \$122816815 | \$126337143 |
| Jul   | -1,70%     | -2504727.4019999504 | \$144618783 | \$140797761 | \$144557002 | \$145900670 | \$147123510 |
| Aug   | -0,61%     | -1008144.1319999397 | \$165203070 | \$161125751 | \$165141289 | \$169042978 | \$166211214 |
| Sep   | 0,63%      | 1167816.6450001001  | \$186669058 | \$181860379 | \$186607277 | \$192456801 | \$185501241 |
| Oct   | 1,27%      | 2594264.139000118   | \$207538488 | \$202491225 | \$210520162 | \$216654625 | \$204944224 |
| Nov   | 2,43%      | 5450280.5070001185  | \$229353238 | \$221652272 | \$234355603 | \$238633490 | \$223902957 |
| Dec   | 5,85%      | 14240290.12800017   | \$257525697 | \$242827941 | \$257369776 | \$260093394 | \$243285407 |

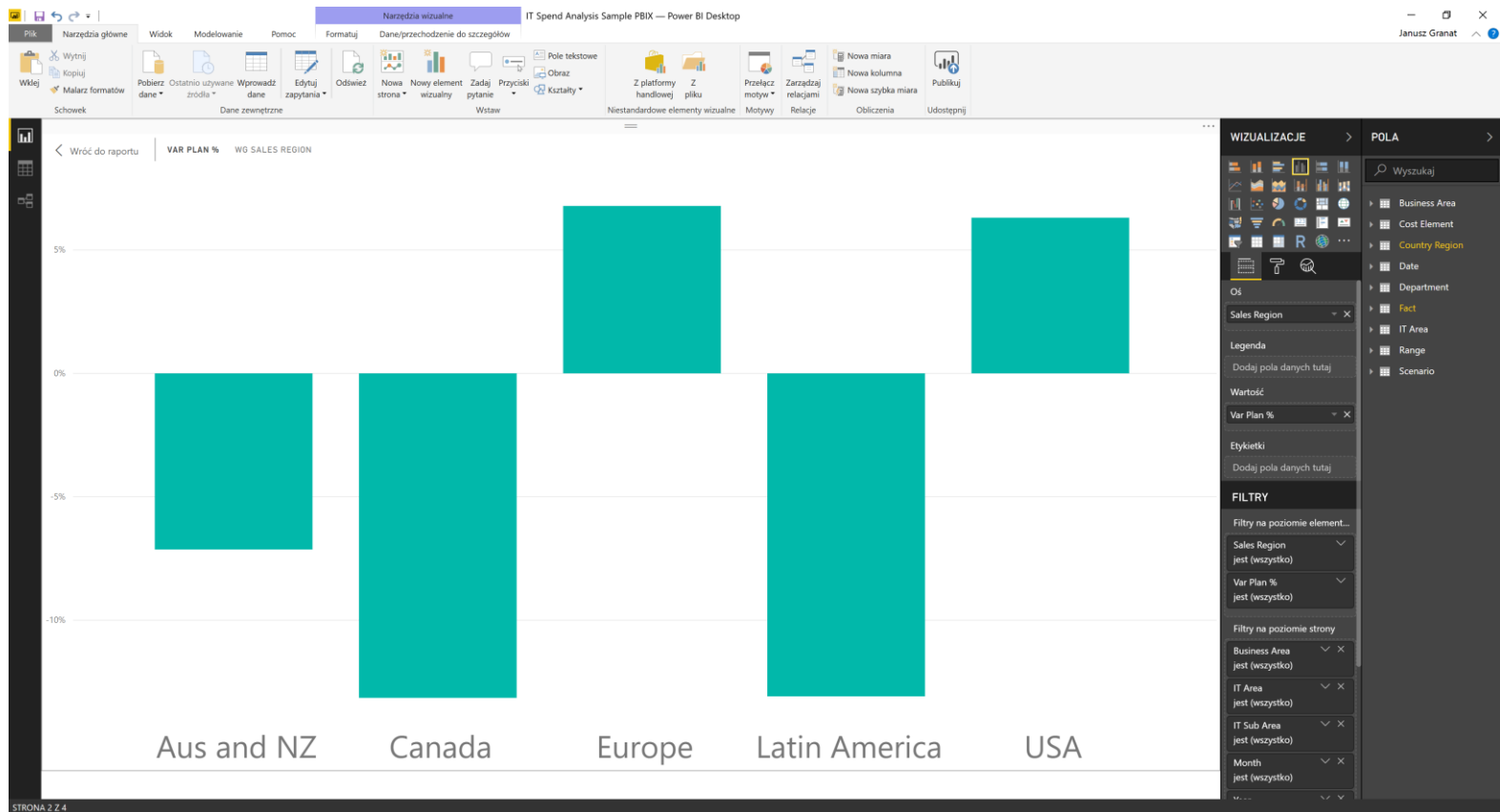
# Model danych



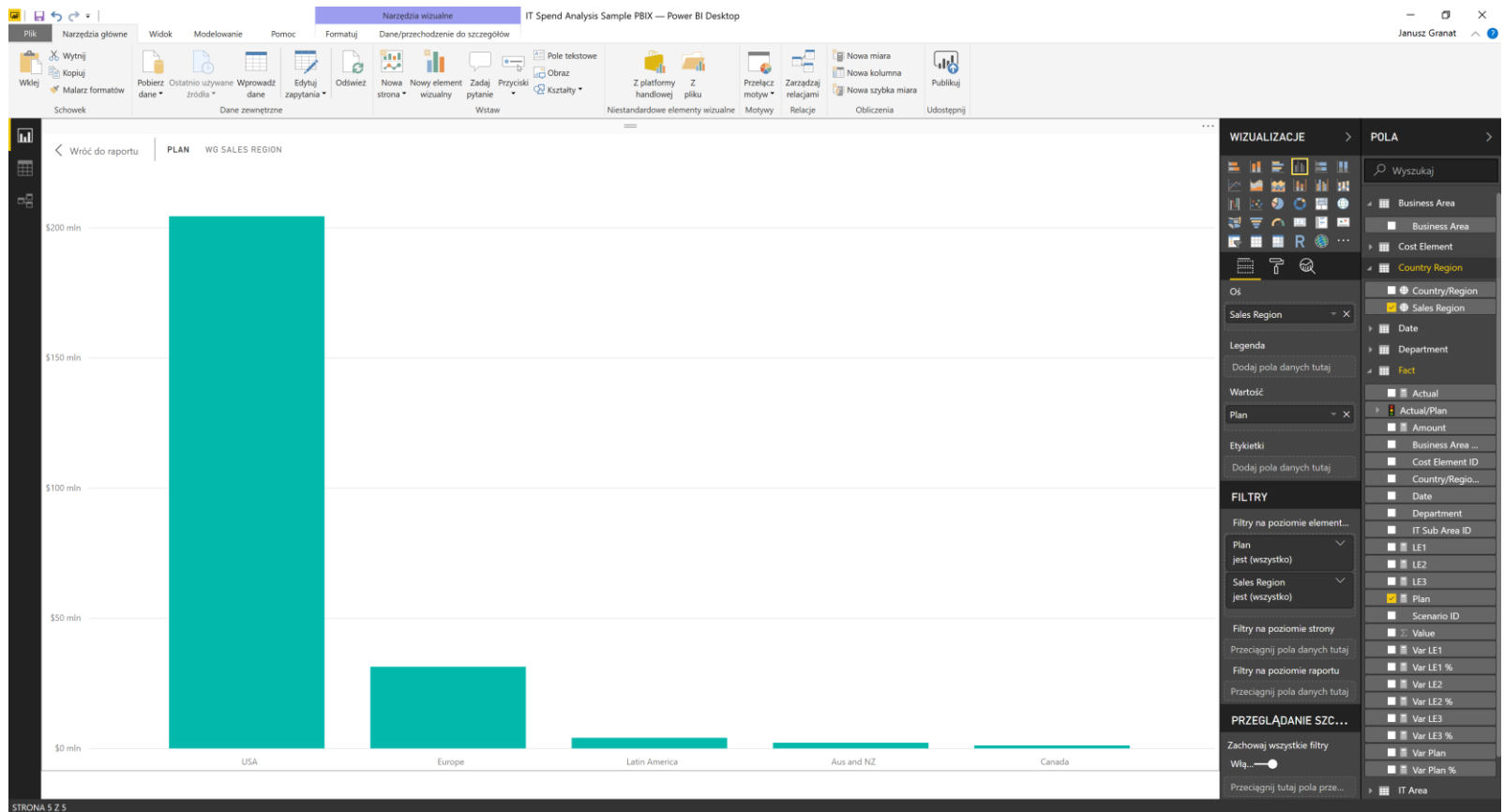
# Przykład analizy



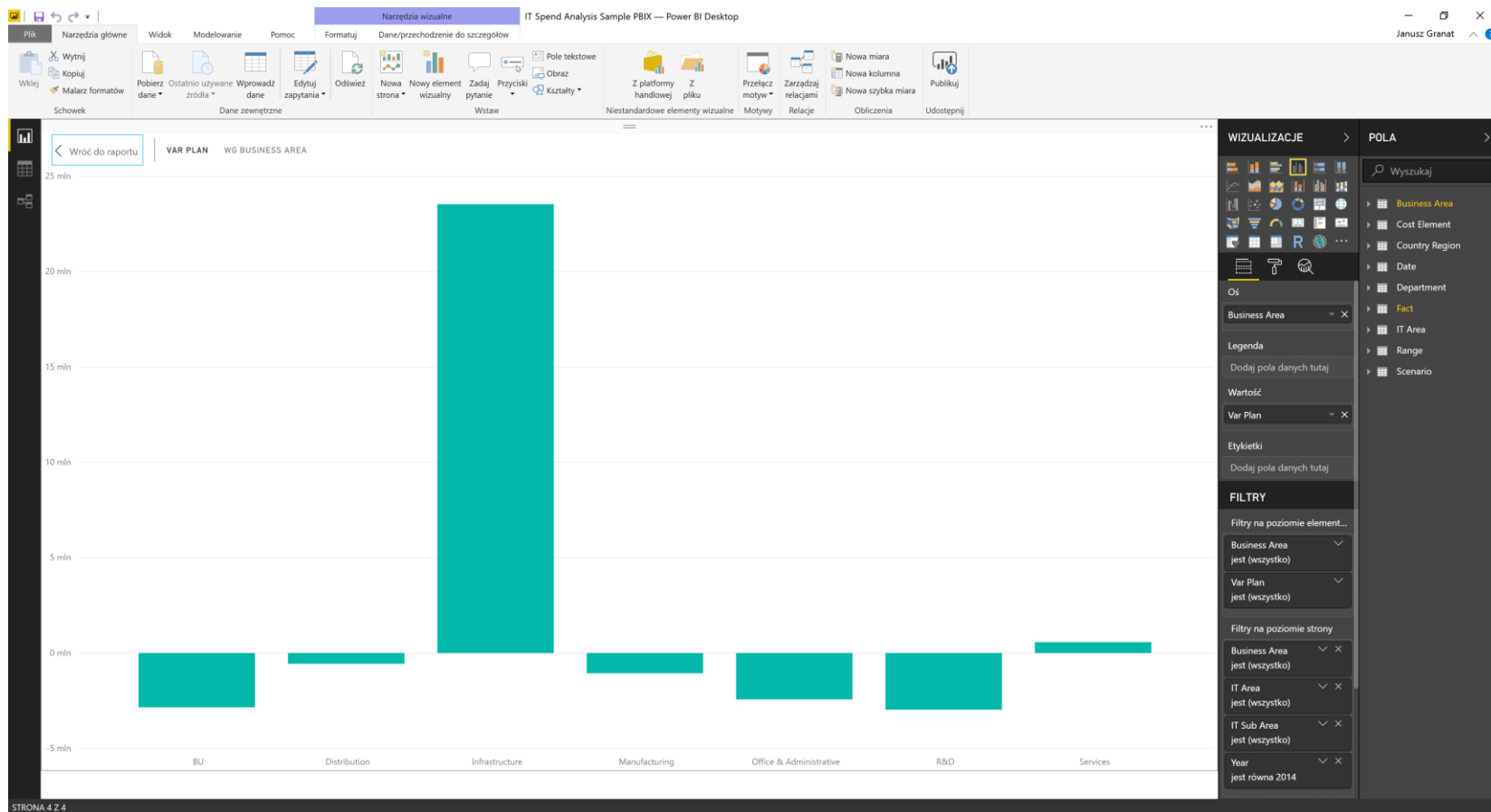
# Regiony



# Wydatki /Regiony



# WYDDKI/OBSZARY



## Przykład analizy II

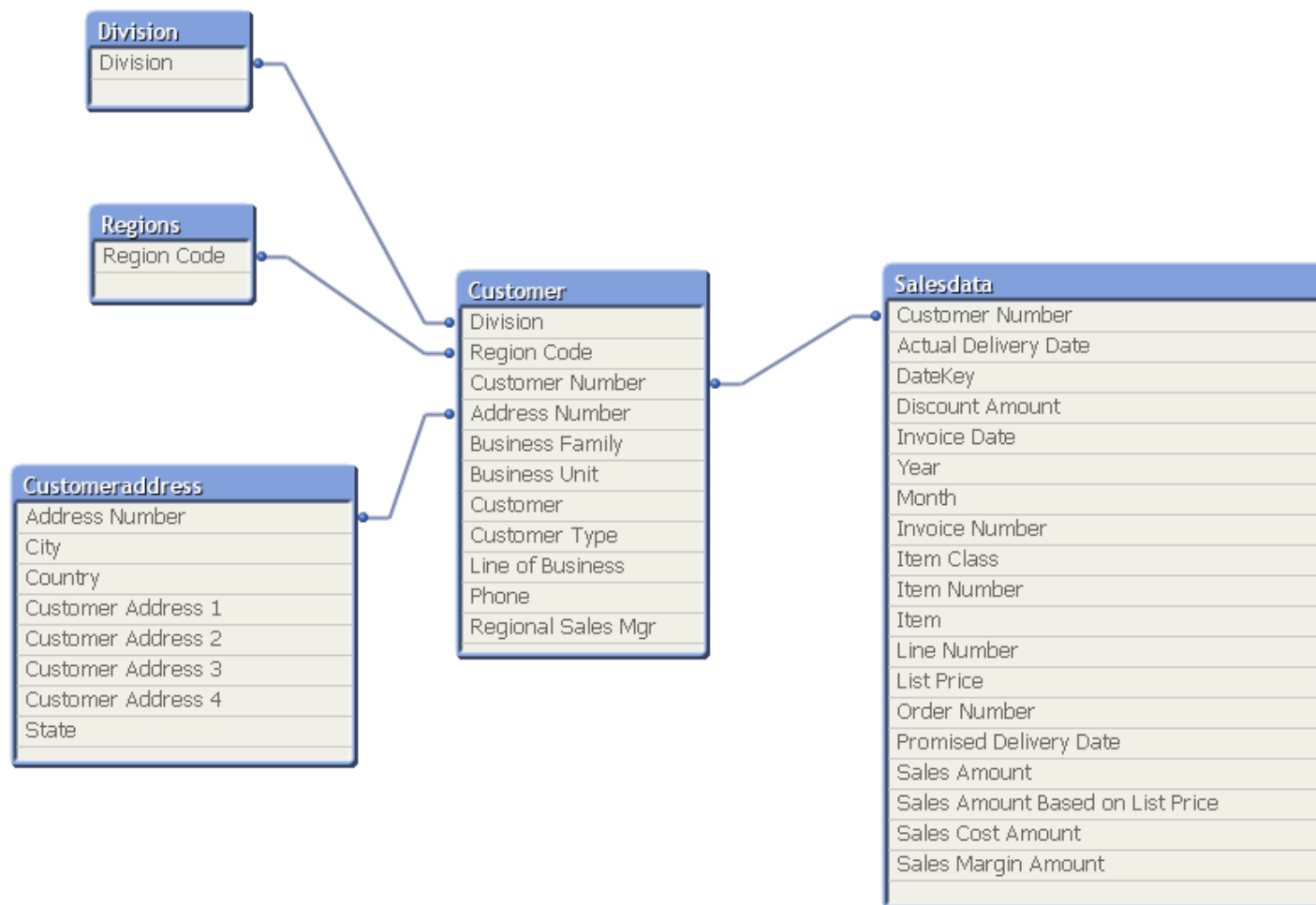
# Przykład – Analiza sprzedaży

---

Analiza pozwala ocenić sprzedaż produktów  
Dane pochodzące z przykładu QlickView

# Analiza danych sprzedażowych

## Dane Źródłowe



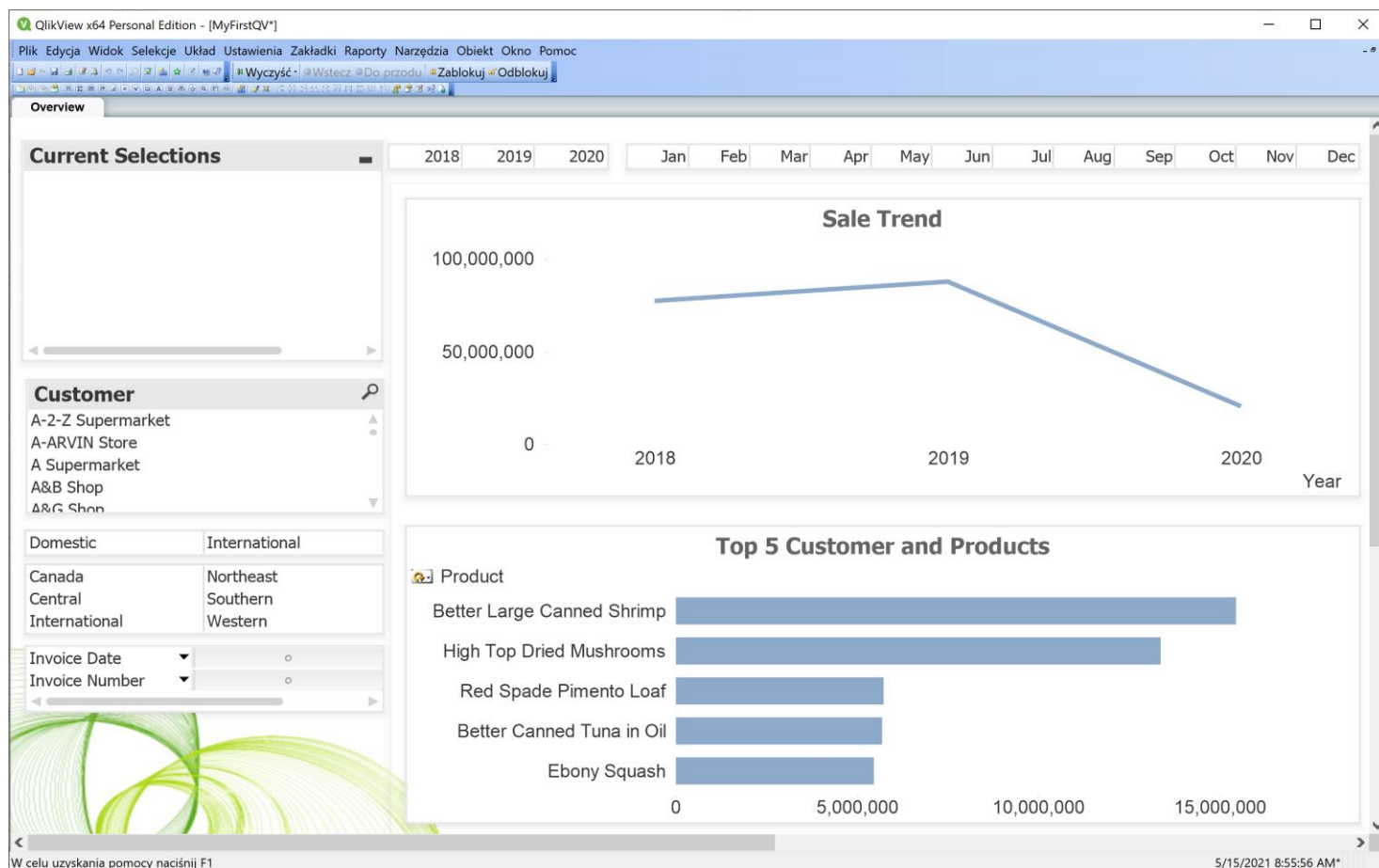
# Tabela sprzedaży

Okno dialogowe

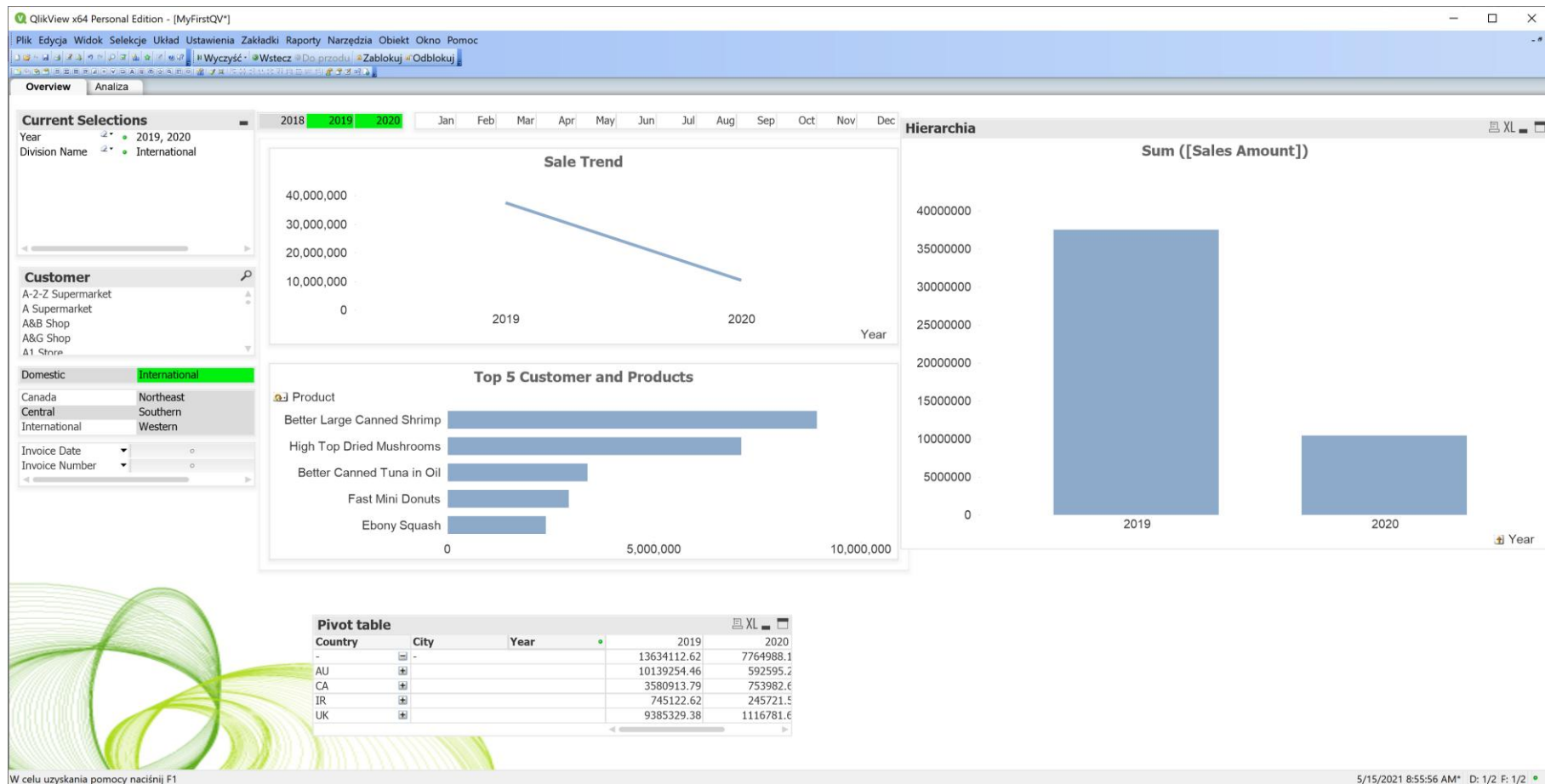
| Invoice... | Item Class | Item Num... | Item                            | Line Number | List Price | Order Number | F ^ |
|------------|------------|-------------|---------------------------------|-------------|------------|--------------|-----|
| 119351     | P01        | 34901       | Better Noodle Soup              | 21601.88    | 51.88      | 216007       |     |
| 119351     | P01        | 62725       | Tell Tale Walnuts               | 21601.88    | 51.88      | 216007       |     |
| 119351     | P01        | 28929       | Nationeel Potato Chips          | 21607.36    | 37.36      | 216007       |     |
| 119352     | P01        | 15236       | Great Muffins                   | 21607.05    | 37.05      | 216014       |     |
| 119351     | P01        | 36015       | Big Time Apple Cinnamon Waffles | 21607.05    | 37.05      | 216007       |     |
| 119352     | P01        | 36015       | Big Time Apple Cinnamon Waffles | 21607.05    | 37.05      | 216014       |     |
| 119352     | P01        | 36001       | Big Time Frozen Cheese Pizza    | 21602.03    | 412.03     | 216014       |     |
| 119352     | P01        | 28929       | Nationeel Potato Chips          | 21607.36    | 37.36      | 216014       |     |
| 119352     | P01        | 37673       | Landslide French Roast Coffee   | 21607.5     | 207.5      | 216014       |     |
| 119352     | P01        | 38076       | Cutting Edge Foot-Long Hot Dogs | 21608.66    | 548.66     | 216014       |     |
| 119352     | P01        | 60794       | High Top Walnuts                | 21601.88    | 51.88      | 216014       |     |
| 119352     | P01        | 62018       | Big Time Grape Popsicles        | 21601.82    | 51.82      | 216014       |     |
| 119352     | P01        | 62871       | Fast Dried Apples               | 21601.98    | 21.98      | 216014       |     |
| 119352     | P01        | 61484       | Super Creamy Peanut Butter      | 21601.88    | 51.88      | 216014       |     |
| 119351     | P01        | 36001       | Big Time Frozen Cheese Pizza    | 21602.03    | 412.03     | 216007       |     |
| 119351     | P01        | 38076       | Cutting Edge Foot-Long Hot Dogs | 21608.66    | 548.66     | 216007       |     |
| 119351     | P01        | 61484       | Super Creamy Peanut Butter      | 21601.88    | 51.88      | 216007       |     |

Zamknij Pomoc

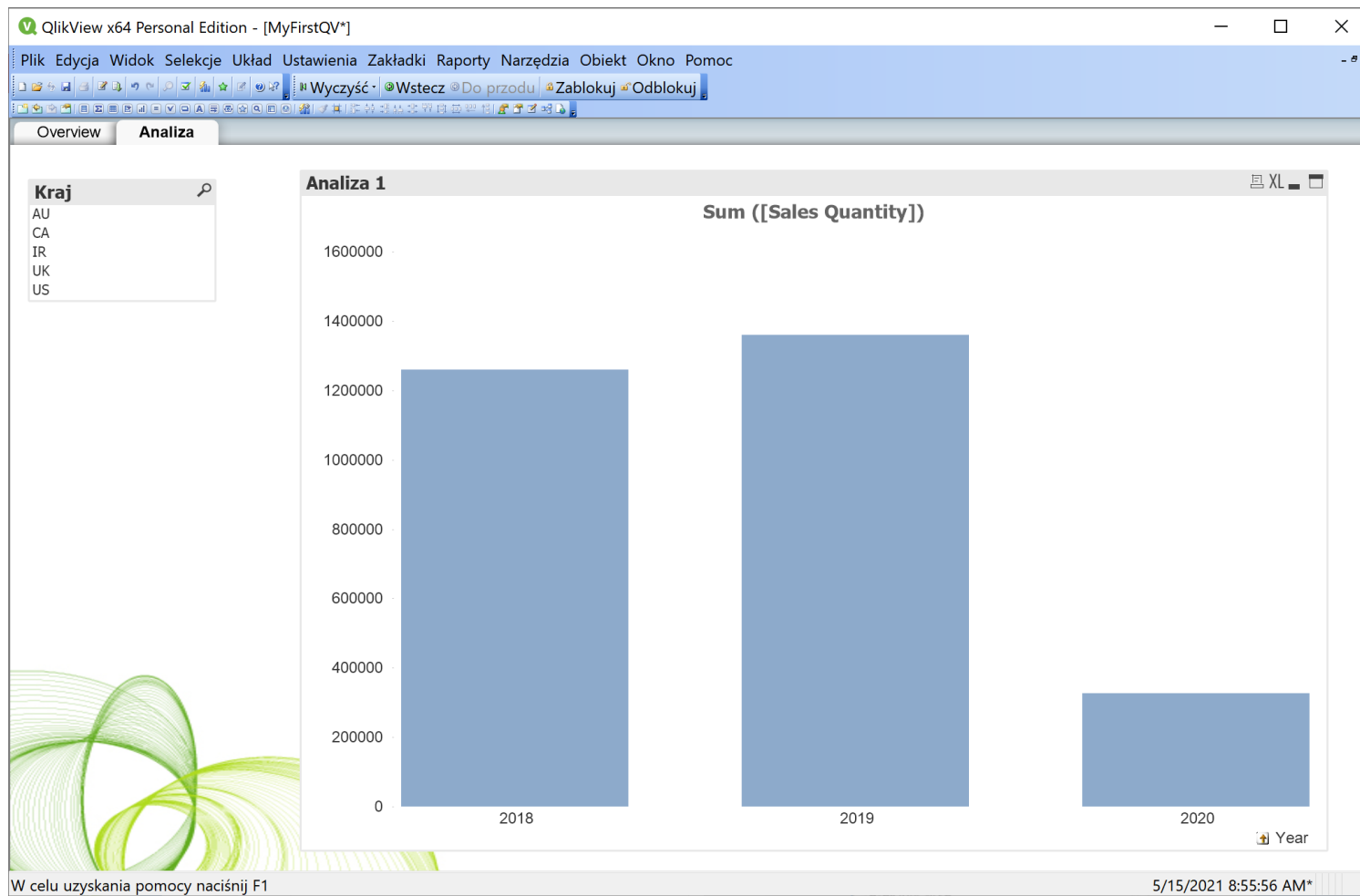
# Ekran analizy



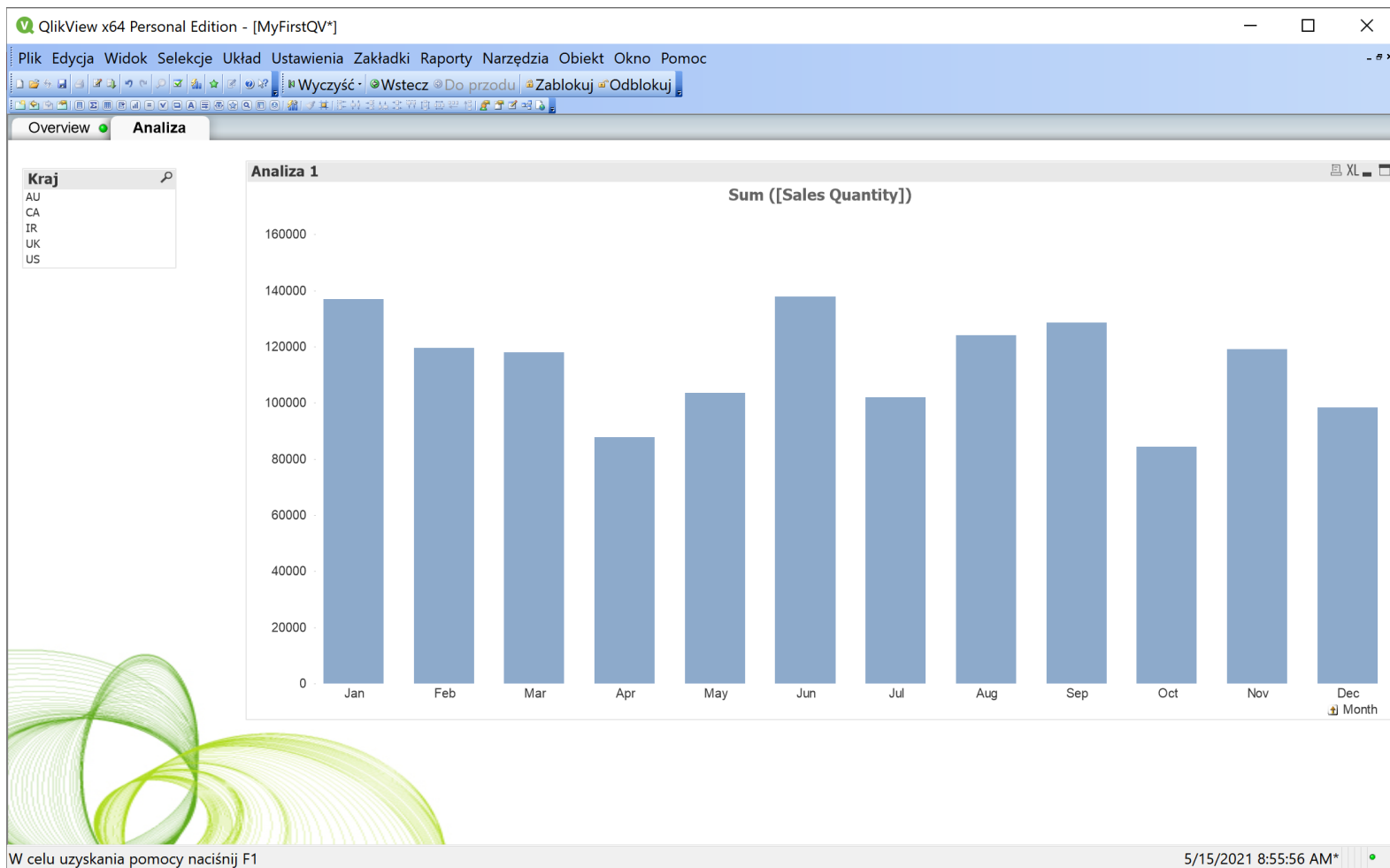
# Ekran analizy (selekcja)



# Hierarchia (poziom 1)



# Hierarchia (poziom 2)



# Tabela typu PIVOT

QlikView x64 Personal Edition - [MyFirstQV\*]

Plik Edycja Widok Selekcje Układ Ustawienia Zakładki Raporty Narzędzia Obiekt Okno Pomoc

Wyczyść Wstecz Do przodu Zablokuj Odblokuj

Overview Analiza **Pivot table**

**Pivot table** XL

| Country | City                | Year | 2019        |
|---------|---------------------|------|-------------|
| -       | -                   |      | 13634112.62 |
| AU      |                     |      | 10139254.46 |
| CA      |                     |      | 3580913.79  |
| IR      |                     |      | 745122.62   |
|         |                     |      | 1554339.91  |
|         | 00-394 Warsaw       |      | 413724.2    |
|         | 1980 Santiago       |      | 150000      |
|         | 4100 Porto          |      | 131719.75   |
|         | A 5102 Anthering    |      | 69593.93    |
|         | Beijing 100088      |      | 56910.6     |
|         | Bogota              |      | 48471.61    |
| UK      | Bremen              |      | 96022.27    |
|         | Cali                |      | 8183.22     |
|         | Carcas              |      | 52391.48    |
|         | D 95326 Kulmbach    |      | 1162716.12  |
|         | Dubai               |      | 150358.38   |
|         | EE 0026 Tallinn     |      | 8422.58     |
|         | FIN 00710           |      | 186667.73   |
|         | Industriestrasse... |      | 53829.64    |

W celu uzyskania pomocy naciśnij F1

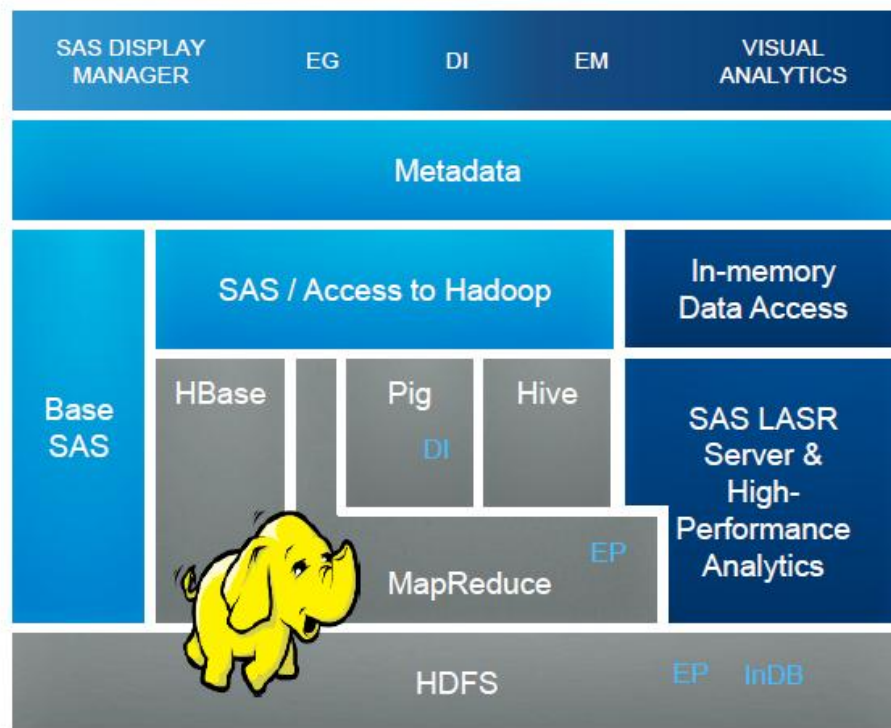
5/15/2021 8:55:56 AM\* 32 X 4

# Big Data Analytics

---

- **Volume** - duża ilość danych
- **Variety** - duża różnorodność danych (structured, unstructured, text, sensor data, audio, video, click streams, log files)
- **Velocity** - duża szybkość pojawiania się nowych danych i ich analizy w czasie rzeczywistym
- **Value** - znacząca wartość danych dla biznesu

# Big Data – Apache Hadoop and SAS



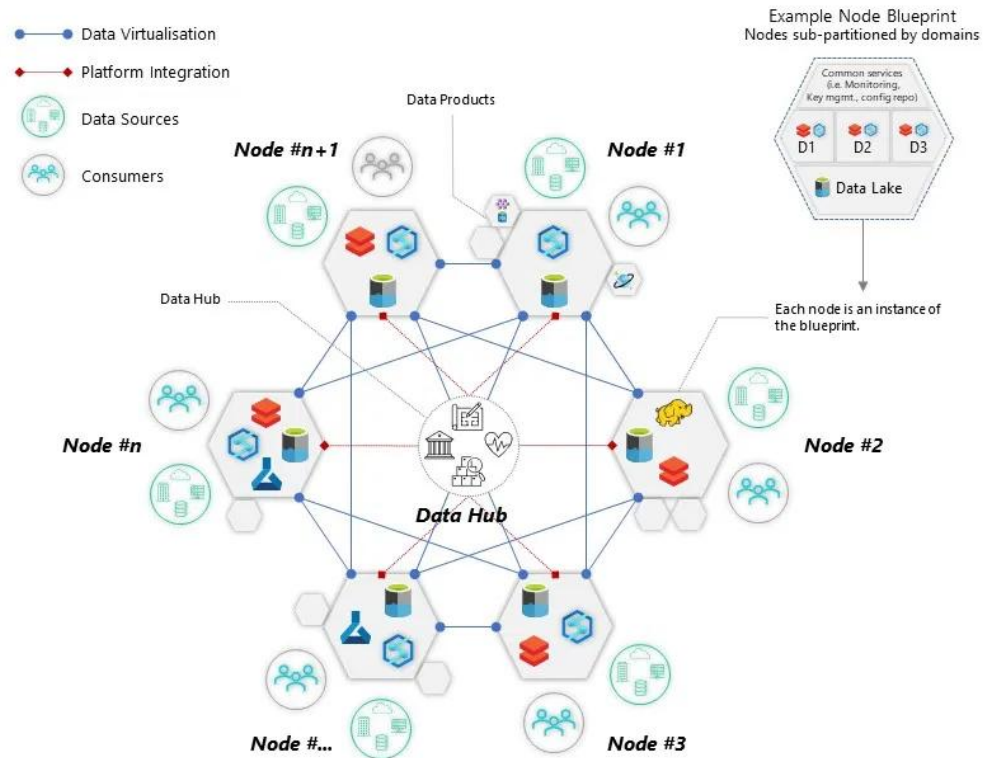
Paul Kent – SAS Institute

Regresja logistyczna  
SAS - 20h  
SAS and Hadoop – 2 min

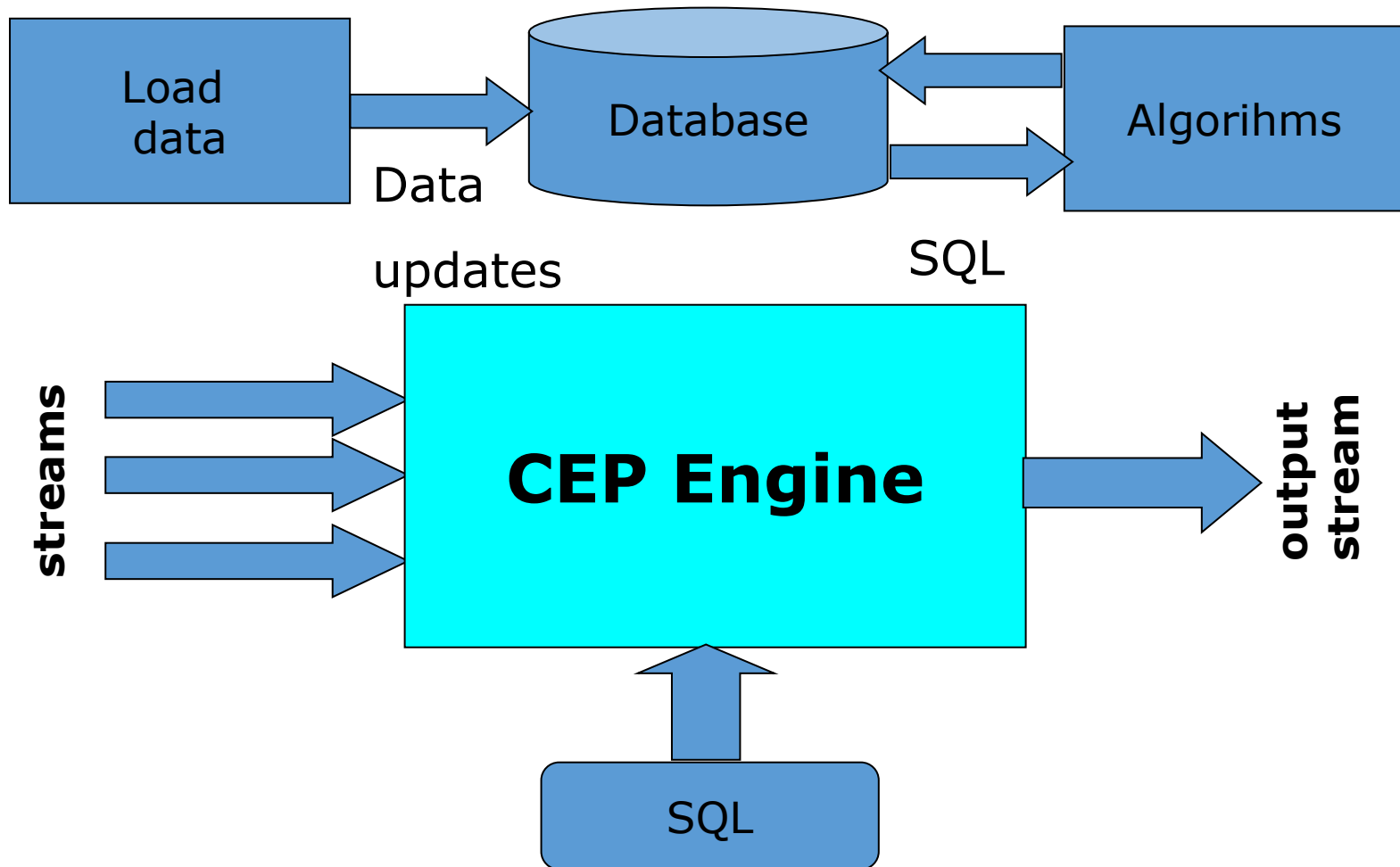
# Data mesh

## Harmonised Mesh

- Azure Harmonised Mesh allows multiple groups within an organisation to operate their own analytics platform whilst adhering to common policies and standards.
- The central datahub hosts data catalogue, mesh wide audit capabilities, monitoring, and auxiliary services for automation.
- The central data platform group defines blueprints that encompass baseline security, policies, capabilities and standards.
- New nodes are instantiated based on these blueprints, which encompass key capabilities to enable enterprise analytics (i.e.. Storage, monitoring, key management, ELT, analytical engines, and automation)
- Node instances can be augmented to serve respective business requirements, i.e. deploying additional domains, customising domains and data products within the node.
- Nodes are typically split by either org-division, function, or region.



# ESP (Event Stream Processing)



# Analiza zdarzeń na stronach www

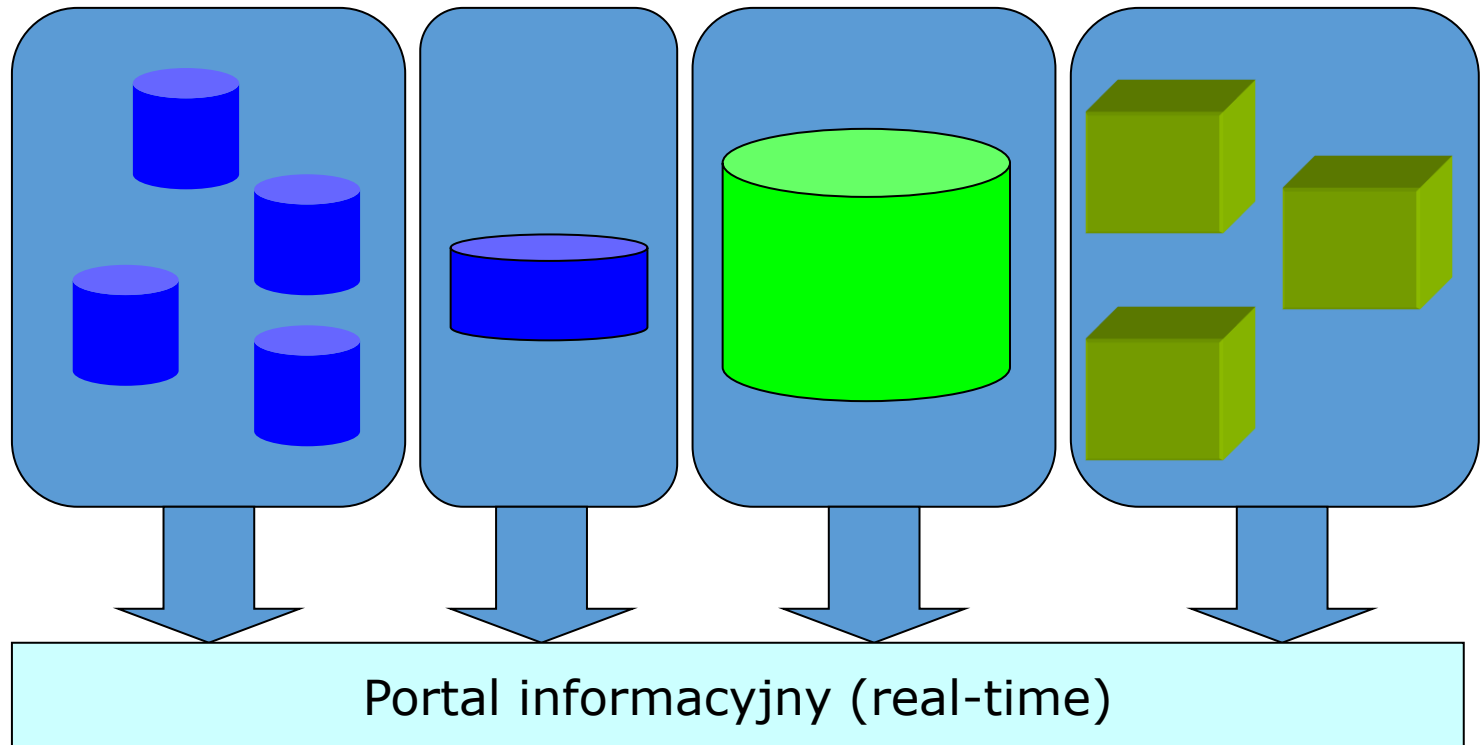
The screenshot shows the website of Politechnika Warszawska (Warsaw University of Technology) with several annotations for event analysis:

- Zmiana rekomendacji** (Change of recommendation) - points to the navigation bar.
- Przetwarzania i analityka danych strumieniowych** (Streaming data processing and analytics) - points to a large image of a person working on a red surface.
- Akcje dla sesji o zadanej długości** (Actions for sessions of a given length) - points to the same large image.
- Nieprawidłowy CTR (clickthrough rate)** (Incorrect CTR) - points to the 'Badania i nauka' (Research and Science) section.

The website content includes:

- Navigation bar: Aktualności, Kalendarium, Badania i nauka, Rekrutacja, Studenci, Pracownicy, Uczelnia, Kontakt, Szukaj.
- Logo: Politechnika Warszawska.
- Section: **Badania i nauka** (Research and Science).
- Section: **Aktualności** (News).
- Section: **Nowy rozdział w rozwoju drogownictwa w Polsce** (A new chapter in the development of road transport in Poland).
- Text: Opublikowano: 08.03.2021 13:56.
- Text: Zakończyły się prace badawcze w ramach projektu „Optymalizacja konstrukcji asfaltowej nawierzchni drogi dzięki zastosowaniu asfaltów modyfikowanych”. Otrzymane wyniki dowiodły, że opracowane koncepcje konstrukcji nawierzchni z wykorzystaniem asfaltów wysokomodyfikowanych HiMA pozwalają budować trwalsze i tańsze drogi. W projekcie uczestniczyli naukowcy z Politechniki Warszawskiej.
- Section: **Naukowcy zajmujący się geoinformacją łączą siły** (Scientists working on geoinformation join forces).
- Text: Opublikowano: 08.03.2021 13:14.
- Text: Za nami pierwsze po powołaniu pełnego składu posiedzenie Rady Naukowej Sieci Naukowej Analiz Geoprzestrzennych. Liderem sieci jest Politechnika Warszawska.
- Section: **1 nowa wiadomość na koncie J.Granat@ia.pw.edu.pl** (1 new message on the account J.Granat@ia.pw.edu.pl).
- Text: Re: sprawozdania. Zbyszek, Przypomniała mi się jeszcze jedna zgłoszona publikacja; przy okazji poprawilem drobniutki. Pozdrawiam, Marek.

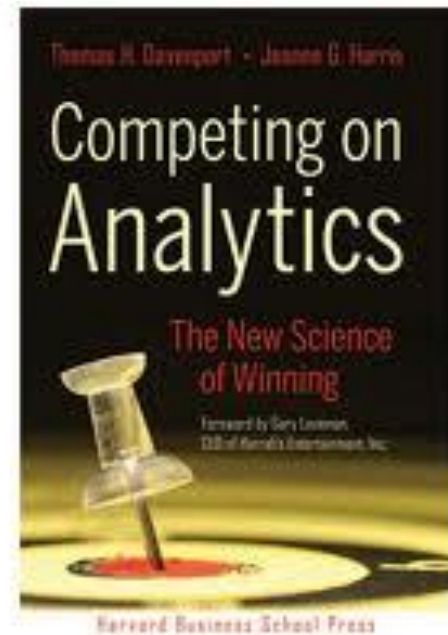
# BI – następna generacja



# Analityka a przewaga konkurencyjna (1)

---

Davenport, T.H., Harris, J.G.:  
Competing on Analytics: the New  
Science of Winning.  
Harvard Business School,  
Boston (2007)

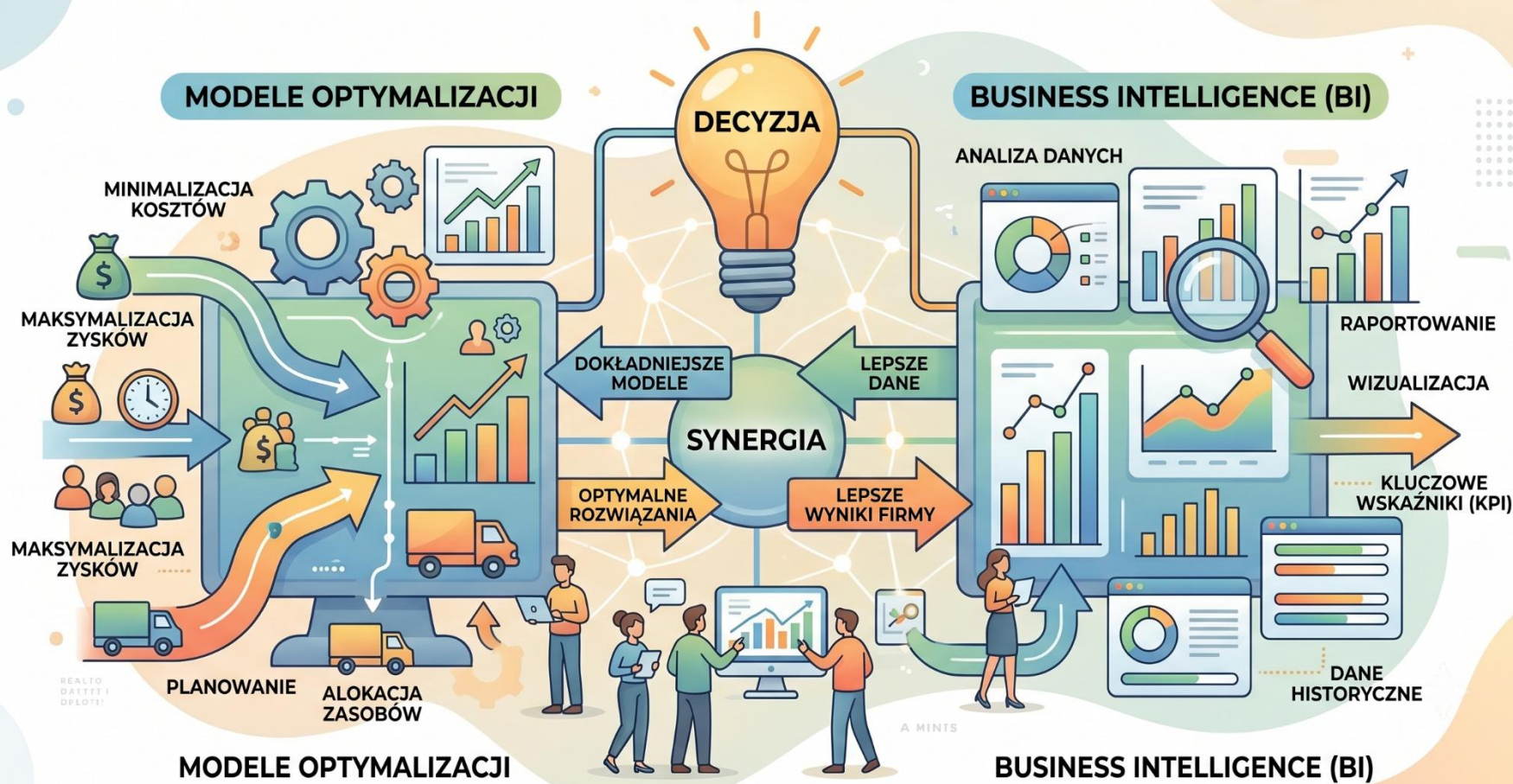


# Analityka a przewaga konkurencyjna (2)

|                                 |                                      |              |
|---------------------------------|--------------------------------------|--------------|
| Optymalizacja                   | Poszukiwanie najlepszego rozwiązania | Analityka    |
| Modelowanie predykcyjne         | Przewidywanie                        |              |
| Przewidywanie/<br>ekstrapolacja | Czy trend będzie kontynuowany?       |              |
| Analiza statystyczna            | Dlaczego coś się wydarzyło?          |              |
| Alerty                          | Niezbędne działania                  | Raportowanie |
| Zapytania/ drążenie danych      | Poszukiwanie przyczyn                |              |
| Raporty ad-hoc                  | Raporty na żądanie                   |              |
| Raporty                         | Prezentacja danych                   |              |

# Optymalizacja i Business Intelligence

## MODELE OPTYMALIZACJI I BUSINESS INTELLIGENCE



# Rola modeli optymalizacyjnych w BI

## Optymalizacja w BI odpowiada na pytanie:

*Jak nie tylko zrozumieć przeszłość,  
Ale i kształtować przyszłość?*

Klasyczna analityka opisuje CO się dzieje.  
Modele optymalizacyjne odpowiadają JAK działać.

## Przykłady zastosowań:

- Minimalizacja kosztów operacyjnych
- Maksymalizacja zysku / przychodów
- Optymalna alokacja zasobów
- Zarządzanie ryzykiem finansowym

↑ 15–35%

poprawa efektywności  
operacyjnej

↓ 20%

redukcja kosztów  
w łańcuchu dostaw

3×

szybsze decyzje  
przedsiębiorstwa

# OLAP vs Modele optymalizacyjne

| OLAP                                                                                     | vs        | OPTYMALIZACJA                           |
|------------------------------------------------------------------------------------------|-----------|-----------------------------------------|
| Co się stało? Ile? Kiedy?                                                                | Pytanie   | Co robić? Jak działać optymalnie?       |
| Eksploracja i raportowanie danych                                                        | Cel       | Znalezienie najlepszej decyzji          |
| Historyczne — kostki OLAP, agregaty                                                      | Dane      | Historyczne + model matematyczny        |
| Dashboard, pivot table, wizualizacja                                                     | Wynik     | Wartość optymalna + plan działania      |
| Sprzedaz w Q3 wg regionu i kategorii                                                     | Przykład  | Ile zamówić, żeby zminimalizować koszt? |
| Power BI, Tableau, SAP BW, SSAS                                                          | Narzędzia | Python PuLP, Gurobi, OR-Tools, CPLEX    |
| OLAP mówi CO jest — optymalizacja mówi CO ROBIC. Razem tworzą pełny cykl decyzyjny w BI. |           |                                         |

# AI i Business Intelligence



# Architektura: Tradycyjne BI vs. AI-driven BI

## Tradycyjne BI

### ETL

Ekstrakcja → Transform → Load (ręczne reguły)

### OLAP

Analizy wielowymiarowe, kostki danych

### SQL

Zapytania pisane przez analityka

### Dashboardy

Statyczne raporty (Tableau, Power BI)

### Ograniczenie

Opisuje CO było — nie przewiduje CO BĘDZIE

## AI-driven BI

### Auto-ETL

Pipeline danych z detekcją anomalii (Great Expectations)

### ML Models

Predykcja, klastrowanie, rekomendacje (scikit-learn, XGBoost)

### NLP Query

Pytania do danych w języku naturalnym (LangChain + SQL)

### Auto-Insights

Narracje generowane przez LLM (GPT-4, Claude)

### Możliwość

Przewiduje, rekomenduje i WYJAŚNIA decyzje

# Predykcja Sprzedaży — Random Forest w Pythonie

*Zadanie: Przewidzieć sprzedaż tygodniową sklepu na podstawie: dnia tygodnia, promocji, temperatury, święta.*

Python · scikit-learn

```
import pandas as pd
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_absolute_error, r2_score
import numpy as np

# 1. Wczytanie danych
df = pd.read_csv("sprzedaz.csv")

# 2. Feature engineering
df["dzien_tyg"] = pd.to_datetime(df["data"]).dt.dayofweek
df["miesiac"] = pd.to_datetime(df["data"]).dt.month

# 3. Cechy i target
features = ["dzien_tyg", "miesiac", "temperatura",
            "promocja", "swieto", "zapas_stock"]
X = df[features]
y = df["sprzedaz_szt"]

# 4. Podział train/test (80/20)
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)

# 5. Model
model = RandomForestRegressor(
    n_estimators=200,
    max_depth=8,
    random_state=42
)
model.fit(X_train, y_train)
```

## Wyniki modelu:

**MAE: 43 szt.**

Średni błąd bezwzględny

**RMSE: 61 szt.**

Pierwiastek błędu kwadratowego

**R<sup>2</sup>: 0.91**

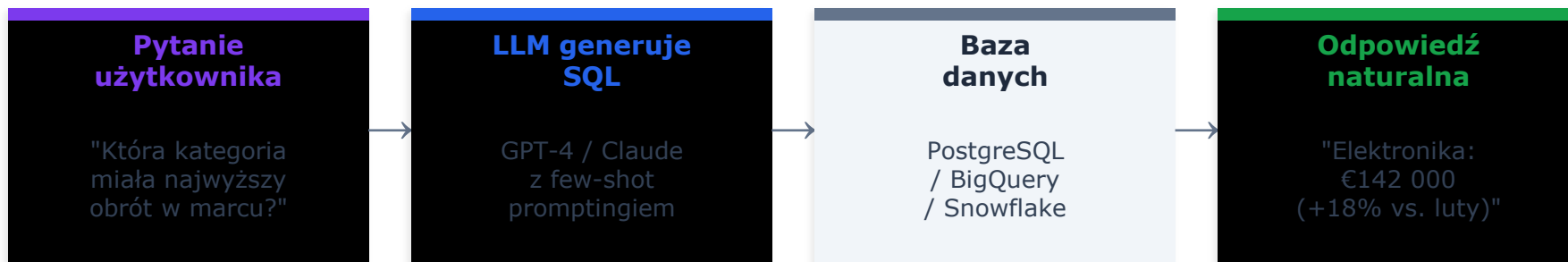
91% wariancji wyjaśnionej modelem

## Ważność cech (feature importance):

|             |                 |
|-------------|-----------------|
| zapas_stock | <div></div> 34% |
| promocja    | <div></div> 27% |
| dzien_tyg   | <div></div> 18% |
| temperatura | <div></div> 12% |
| swieto      | <div></div> 6%  |
| miesiac     | <div></div> 3%  |

# NLP w BI — Pytania do danych w języku naturalnym

*Cel: Użytkownik pisze po polsku, system generuje SQL i zwraca odpowiedź.*



Python ·  
LangChain · GPT-4

```
from langchain.agents import create_sql_agent
from langchain.sql_database import SQLDatabase
from langchain.agents.agent_types import AgentType
from langchain_openai import ChatOpenAI

# Połącz z bazą danych
db = SQLDatabase.from_uri(
    "postgresql://user:pass@localhost/bi_db",
    include_tables=["sprzedaz", "produkty", "sklepy"]
)
llm = ChatOpenAI(model="gpt-4o", temperature=0)
# Utwórz agenta SQL
agent = create_sql_agent(
    llm=llm,
    db=db,
    agent_type=AgentType.OPENAI_FUNCTIONS,
    verbose=True # pokaż wygenerowany SQL
)
# Pytanie po polsku
odpowiedz = agent.invoke(
    "Która kategoria miała najwyższy obrót w marcu?"
)
print(odpowiedz["output"])
# → 'Elektronika: €142 000 (+18% vs. luty)'
```

## Wygenerowany SQL:

```
SELECT
    p.kategoria,
    SUM(s.wartosc) AS obrot,
    LAG(SUM(s.wartosc)) OVER
        (ORDER BY MONTH(s.data))
        AS prev_month
FROM sprzedaz s
JOIN produkty p
    ON s.produkt_id = p.id
WHERE MONTH(s.data) = 3
GROUP BY p.kategoria
ORDER BY obrot DESC
LIMIT 1;
```

# Generatywna AI — Automatyczne Narracje Raportów

*Problem: analityk spędza 3–5h tygodniowo pisząc komentarze do wykresów. Rozwiązanie: LLM generuje narrację automatycznie.*

Python ·  
Anthropic API

```
import anthropic
import json

client = anthropic.Anthropic() # ANTHROPIC_API_KEY z env

# Dane KPI z bazy (np. wynik zapytania SQL)
kpi_data = {
    "okres": "Kwiecień 2025",
    "sprzedaz_total": 2_340_000,
    "zmiana_mom": +0.08, # +8% vs. marzec
    "zmiana_yoy": -0.03, # -3% vs. kwiecień 2024
    "top_kategoria": "Elektronika",
    "top_sklep": "Warszawa Mokotów",
    "zwroty_pct": 0.042,
    "nps": 67
}

prompt = f"""Jesteś analitykiem BI w firmie retail.
Na podstawie poniższych KPI napisz profesjonalne
podsumowanie wyników miesiąca (max 120 słów),
przeznaczone dla zarządu. Wskaż trendy,
niepokojące sygnały i rekomendacje.

Dane: {json.dumps(kpi_data, ensure_ascii=False)}
"""

message = client.messages.create(
    model="claude-opus-4-5",
    max_tokens=300,
    messages=[{"role": "user", "content": prompt}]
)

narracja = message.content[0].text
print(narracja)
```

## Wygenerowana narracja:

Kwiecień 2025 przyniósł wzrost sprzedaży o 8% MoM do €2,34M, jednak wynik jest o 3% poniżej poziomu sprzed roku — trend wymagający uwagi.

Elektronika dominuje sprzedaż (31% udziału), a sklep Warszawa Mokotów osiąga najlepsze wyniki w sieci. Wskaźnik zwrotów 4,2% jest powyżej normy (cel: <3%) — rekomendowane jest zbadanie przyczyn w kategoriach AGD.

## 💡 Dobre praktyki promptingu:

- Definiuj rolę modelu ("Jesteś analitykiem...")
- Podaj limit słów — unikasz rozwlekłości
- Wyraź format wyjścia (JSON, Markdown, proza)