POZNAN UNIVERSITY OF TECHNOLOGY

Data Warehouses and Business Intelligence: Introductory Technology Overview

Robert Wrembel
Poznan University of Technolog
Institute of Computing Science
Robert.Wrembel@cs.put.poznan.pl
www.cs.put.poznan.pl/rwrembel

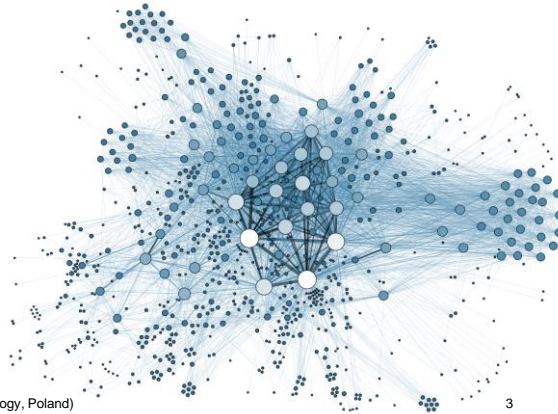# Business Intelligence

⊃ **OLAP - On-Line Analytical Processing**
- **classical analysis**
  - **analysis of trends in the past**
  - **predicting trends**
  - **what-if analysis**
- **typically implemented by means of SQL select**

Business Intelligence

○ **BI = OLAP+**
- **data mining**
- **text analytics (Facebook, Tweeter, ...)**
  - **opinions, sentiments**
- **graph analytics**
- **log analytics**

Typical BI applications

○ **Ad-hoc queries (about 10% corporate applications)**
- **ad-hoc computing**
- **drill-down, drill-accross**

○ **Reports (about 90% corporate applications)**

○ **Dashboards**
- **indicates the status at a specific point in time**

○ **(Performance) Scorecards**
- **monitors progress of a goal over time**
- **answers question: "how well are we doing business?"**
- **includes: KPIs, goals, alerts**

Example reports

- ➲ **Transactions made by payment cards grouped by age categories <20,30),<30;40), ...**
- ➲ **Sum of payments by cards by month in a given year by sex**
- ➲ **$ of granted loans by month in a given year**
- ➲ **Total assets on active debit accounts by month in a given year converted to PLN**

KPI

- ➲ **A Key Performance Indicator (KPI) is a metric (measure) for evaluating factors that are crucial to the success of an organization wrt the business the organization is doing**
- ➲ **Assesses the most important factors for making progress towards declared goals, e.g.**
  - ▪ **yearly profit from sales > 1M EUR**
  - ▪ **# of faulty products < 1%**
  - ▪ **# of satisfied customers > 90%**
  - ▪ **# of cars sold per month > 20**
- ➲ **Represented as visual objects (varius types of gauges)**

KPI

⮩ **Relates the current value of a metric with a target value**

Leading Indicator

⮩ **LI measures key drivers of business value**
⮩ **Allow to predict what will happen soon**
⮩ **Difficult to figure out which LI would be the most appropriate for a given business**
  ▪ **e.g.: LI(customer satisfaction) → KPI(customer retention)**

Balanced Score Card

⊃ **The balanced scorecard is a management system aimed at**

- **translating an organization's strategic goals into a set of performance objectives**
- **the objectives are measured, monitored, and changed if necessary to ensure that the organization's strategic goals are met**

Balanced Score Card

⊃ **Performance measurement with the following 4 perspectives:**

- **financial - includes measures such as: operating income, return on capital, and economic value added**
- **customer - includes measures such as: customer satisfaction, customer retention, and market share in target segments**
- **business process (procurement, production, and order fulfillment) - includes measures such as: cost, throughput, and quality**
- **learning & growth - includes measures such as: employee satisfaction, employee retention, skill sets**

Balanced Score Card

⮑ **Each perspective includes:**

- major **objectives** to be achieved
- **metrics** to measure progress toward reaching the objective
- specific **target** values for the measures
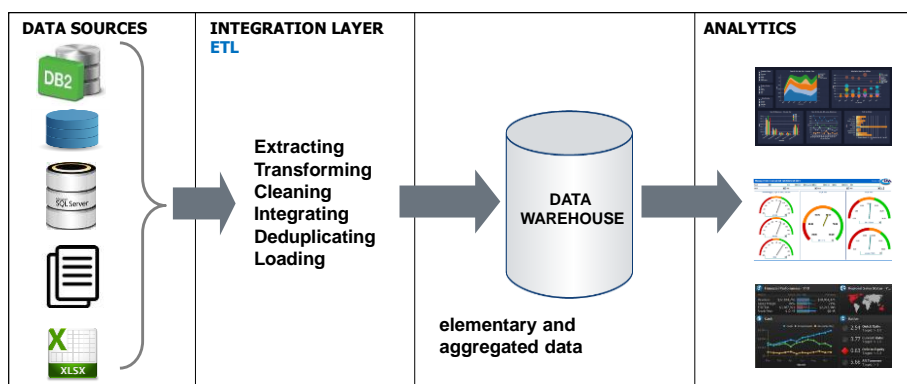- **action**s to be initiated in order to meet the objectives

Charts

cycle plot

highlight table
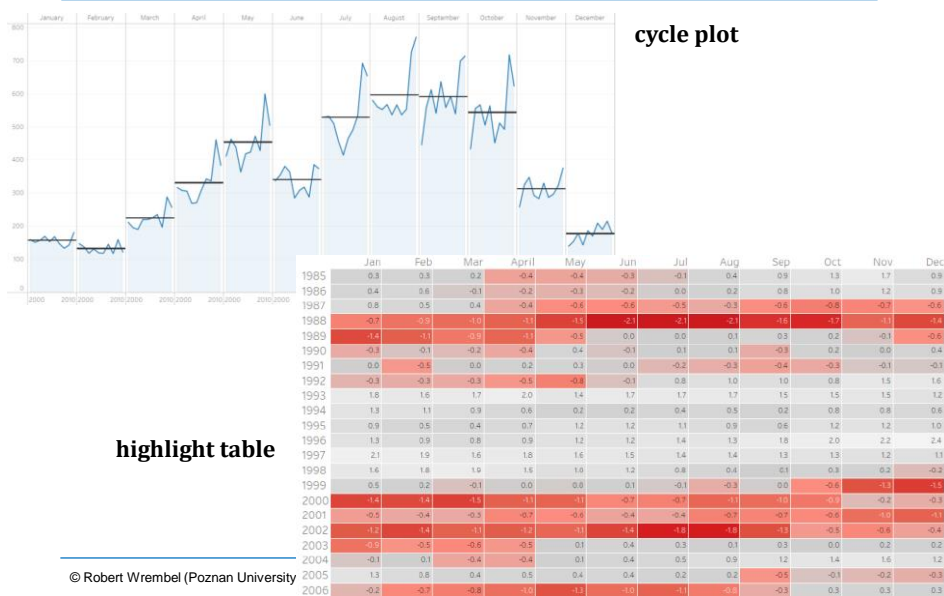
BI users

- ⮥ **Active - 10% of all BI users**
- ⮥ **Working at the same time: 1% of all BI users**

Technology

- ⮥ **Data wareshouse system (see Topic 4)**

Commercial DW systems

⮑ **Traditional**
- **Oracle11g, Hypersion Essbase - Oracle Corporation**
- **DB2 UDB - IBM**
- **Sybase IQ - Sybase**
- **MS SQL Server - Microsoft**
- **SAP Business Warehouse - SAP**
- **Teradata - Teradata**

⮑ **Main memory (in-memory)**
- **Netezza - IBM**
- **Exadata - Oracle**
- **SAP Hana - SAP**
- **Teradata DW Appliance - Teradata**

- **Big Data**
  - **Cloudera Impala**
  - **Apache Kyline**
  - **Apache Drill**

> **DB size > 100 TB: 14% of enterprises**
> **DB size > 60TB: 59% of enterprises**
> **(Forrester Consulting, Jul 2017)**
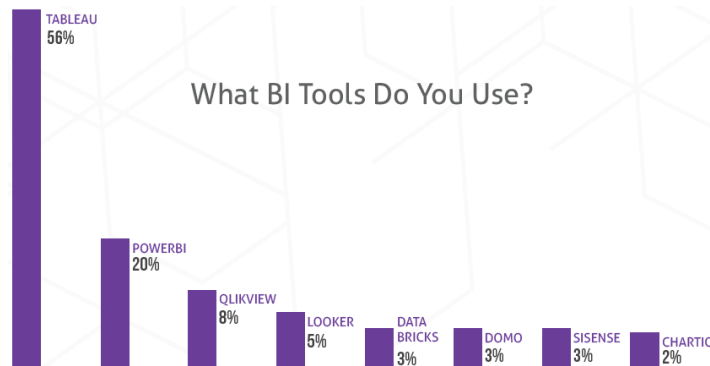
Data governance

⮑ **The management of:**
- **data availability**
- **relevancy**
- **usability**
- **integrity**
- **lineage (provenance)**
- **backups**
- **security in an enterprise**

⮑ **Regulated by a company's policies and rules**

BI tools

➲ **Panoply Data Warehouse Trends Report 2018**

What BI Tools Do You Use?

TABLEAU 56%
POWERBI 20%
QLIKVIEW 8%
LOOKER 5%
DATA BRICKS 3%
DOMO 3%
SISENSE 3%
CHARTIO 2%

Gartner Report

➲ **http://www.gartner.com/technology/reprints.do?id =1-1DZLPEP&ct=130207&st=sb**

➲ **Assessment criteria**

➲ **Integration**
- ▪ **BI infrastructure**
- ▪ **Metadata management**
- ▪ **Development tools**
- ▪ **Collaboration**

➲ **Information Delivery**
- ▪ **Reporting**
- ▪ **Dashboards**
- ▪ **Ad hoc query**
- ▪ **Microsoft Office integration**
- ▪ **Mobile BI**

➲ **Analysis**
- ▪ **Online analytical processing (OLAP) - multidimensional analysis, what-if**
- ▪ **Interactive visualization**
- ▪ **Predictive modeling and data mining**
- ▪ **Scorecards- aligining KPIs with a strategic objective**
- ▪ **Prescriptive modeling, simulation and optimization**
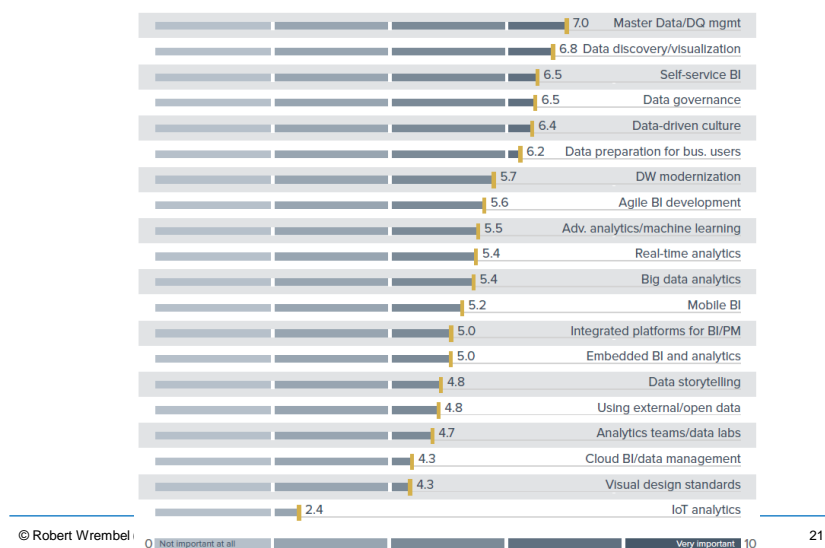
Gartner Report

Magic Quadrant for Analytics and Business Intelligence Platforms

Gartner Report

Magic Quadrant for Data Science and Machine Learning Platforms

BI/DW Trends

➲ **BI Trend Monitor 2019. BARC Research Study**

Value	Category
7.0	Master Data/DQ mgmt
6.8	Data discovery/visualization
6.5	Self-service BI
6.5	Data governance
6.4	Data-driven culture
6.2	Data preparation for bus. users
5.7	DW modernization
5.6	Agile BI development
5.5	Adv. analytics/machine learning
5.4	Real-time analytics
5.4	Big data analytics
5.2	Mobile BI
5.0	Integrated platforms for BI/PM
5.0	Embedded BI and analytics
4.8	Data storytelling
4.8	Using external/open data
4.7	Analytics teams/data labs
4.3	Cloud BI/data management
4.3	Visual design standards
2.4	IoT analytics

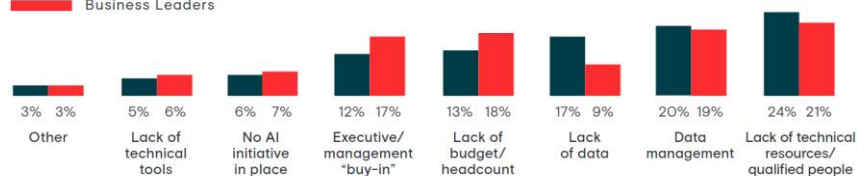0 Not important at all   Very important 10

Data management for Data Science

What do you consider the biggest
bottleneck to any of your AI initiatives or project?

- ■ Technologist
- ■ Business Leaders

3% 3%	5% 6%	6% 7%	12% 17%	13% 18%	17% 9%	20% 19%	24% 21%
Other	Lack of technical tools	No AI initiative in place	Executive/ management "buy-in"	Lack of budget/ headcount	Lack of data	Data management	Lack of technical resources/ qualified people

Participants surveyed: 374

❖ appen

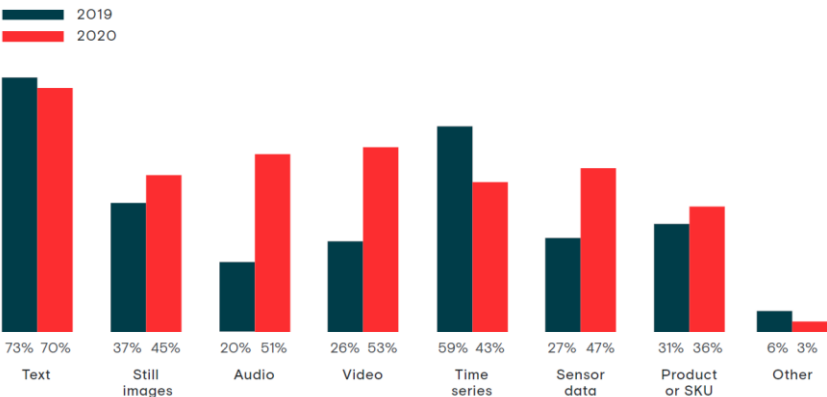The State of AI and Machine Learning. An Appen Whitepaper, 2020

Data management for Data Science

What kinds of data do you work with?

■ 2019
■ 2020

	73% 70%	37% 45%	20% 51%	26% 53%	59% 43%	27% 47%	31% 36%	6% 3%
	Text	Still images	Audio	Video	Time series	Sensor data	Product or SKU	Other

Participants surveyed: 374 in 2020 & 251 in 2019

❖ appen

The State of AI and Machine Learning. An Appen Whitepaper, 2020