Ukrainian-Turkish
Scientific and Technological Research
Coordination Center

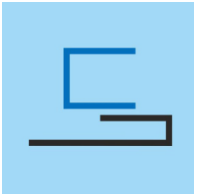
NTU
"KHPI"

İTÜ

Recent Research in Data Analysis and Visualization Issues within Business Intelligence Domain

Department of Software Engineering and Management Information Technology
Computer Science and Software Engineering Faculty
National Technical University “Kharkiv Polytechnic Institute”
Kharkiv, Ukraine

SEMiT



Computer Science
& Software
Engineering Faculty

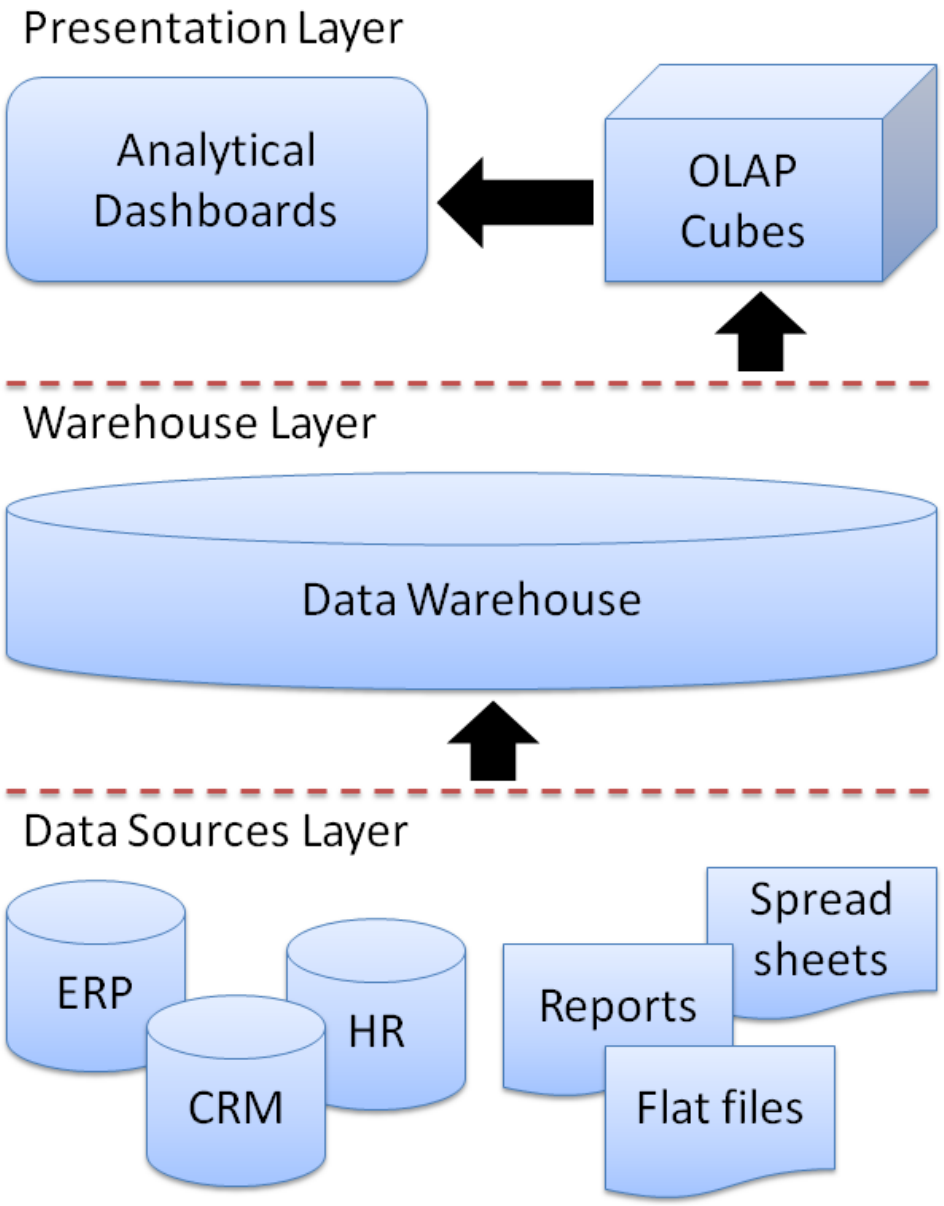
Authors

Andrii Kopp
andrii.kopp@khpi.edu.ua
kopp93@gmail.com

SPEAKER
MSc in Information Technology, Assistant Lecturer, Postgraduate Student, Department of Software Engineering & Management Information Technology, NTU “KhPI”

Dmytro Orlovskyi
dmytro.orlovskyi@khpi.edu.ua
orlovskyi.dm@gmail.com

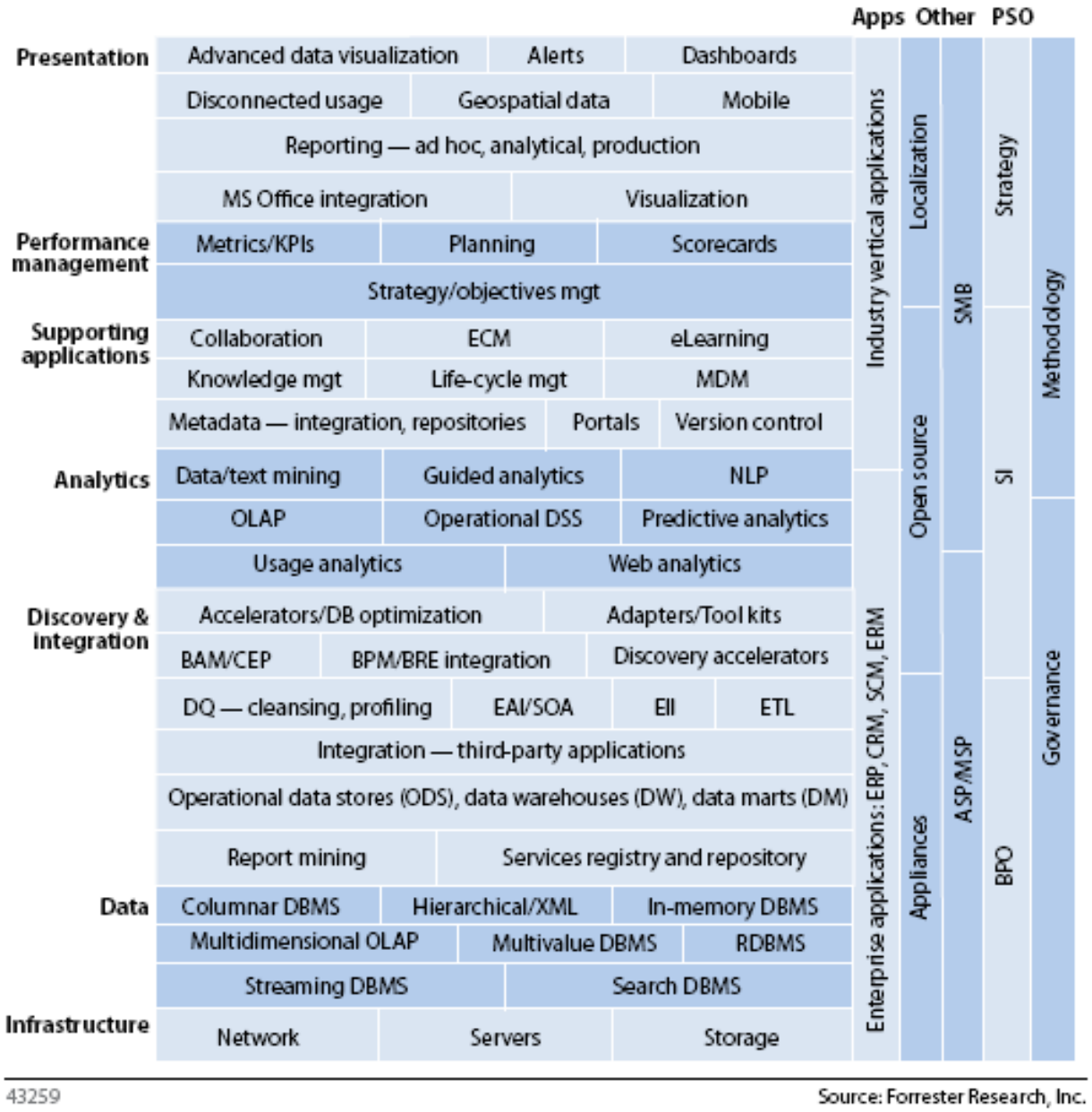
PhD in Information Technology, Associate Professor, Deputy Dean of Computer Science & Software Engineering Faculty, Department of Software Engineering & Management Information Technology, NTU “KhPI”



What is Business Intelligence?

… is an umbrella term that includes the applications, infrastructure and tools, and best practices that enable access to and analysis of information to improve and optimize decisions and performance

Gartner IT Glossary



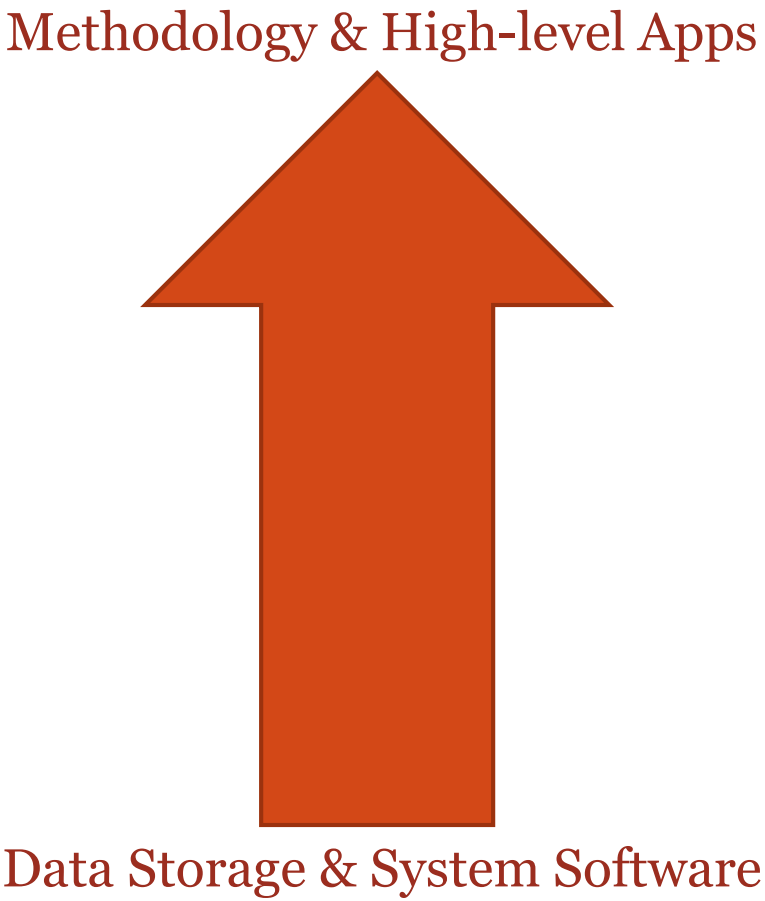
Business Intelligence Stack by Forrester Research, Inc.

- Dashboards
- Metrics/KPIs
- Scorecards
- Knowledge Mgt.
- Repositories
- Version Control
- OLAP
- Predictive Analytics
- BAM
- CEP
- BPM
- ETL
- Data Warehouses
- RDBMS

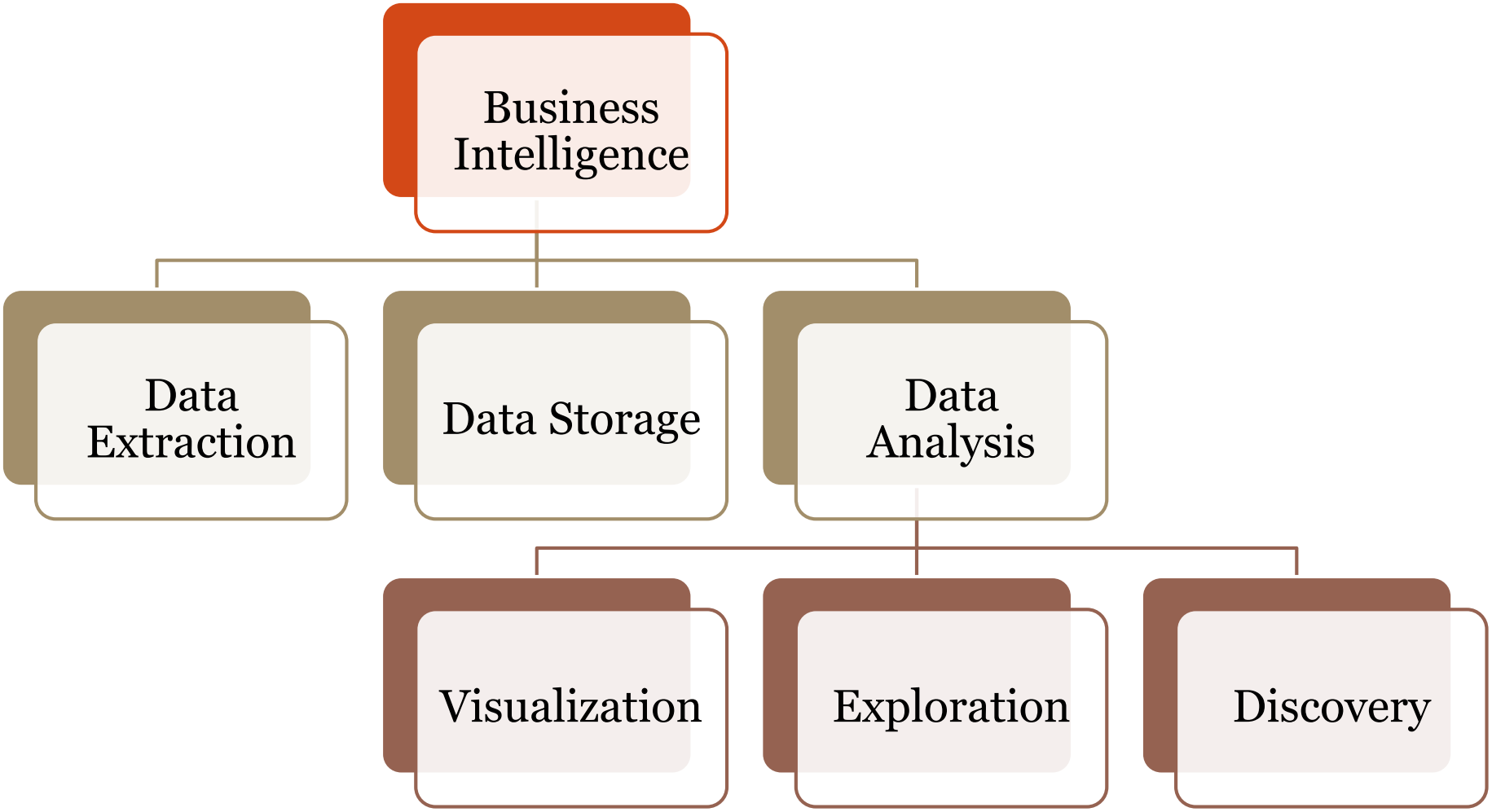
http://www.dbms2.com/2008/01/14/forrester-collects-business-intelligence-buzzwords/

Presentation Structure

- Business Intelligence Stack Overview & Areas of Expertise
- Presentation: Dashboards
- Performance Management:
  - Metrics/KPIs
  - Scorecards
- Supporting Applications:
  - Knowledge Management
  - Repositories
  - Version Control
- Analytics:
  - OLAP
  - Predictive Analytics
- Discovery & Integration:
  - BAM, CEP, BPM, ETL
  - Data Warehouses
- Data: RDBMS



Core BI Capabilities



BI Systems Capabilities

Reporting

It means taking raw data and turning it into information that can be used to make intelligent business decisions.

Dashboards

Pages where you can insert reports, graphs, and charts in order to create a central location for critical business information.

Ad-hoc Querying

It is essentially reporting in real time. It is a tool for exploring questions while you are looking through your data.

OLAP

Aims at efficient multidimensional processing of large data volumes (fast, interactive answers to large aggregate queries).

Data Mining

Capability to dig through huge amounts of data, find the relevant stuff, and come up with predictions. Also known as Machine Learning.

Scorecards

Designed to measure progress toward a goal. Telling you how well you are meeting a specific target goal.

Business Intelligence Systems

Figure 1. Magic Quadrant for Analytics and Business Intelligence Platforms

Source: Gartner (February 2018)

- Power BI

Microsoft

- QlikView

Qlik

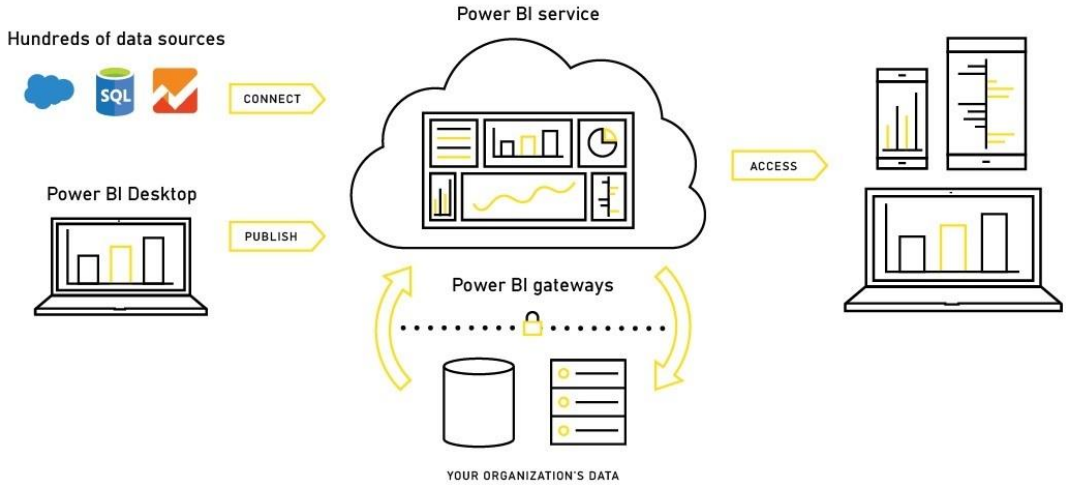
- Tableau

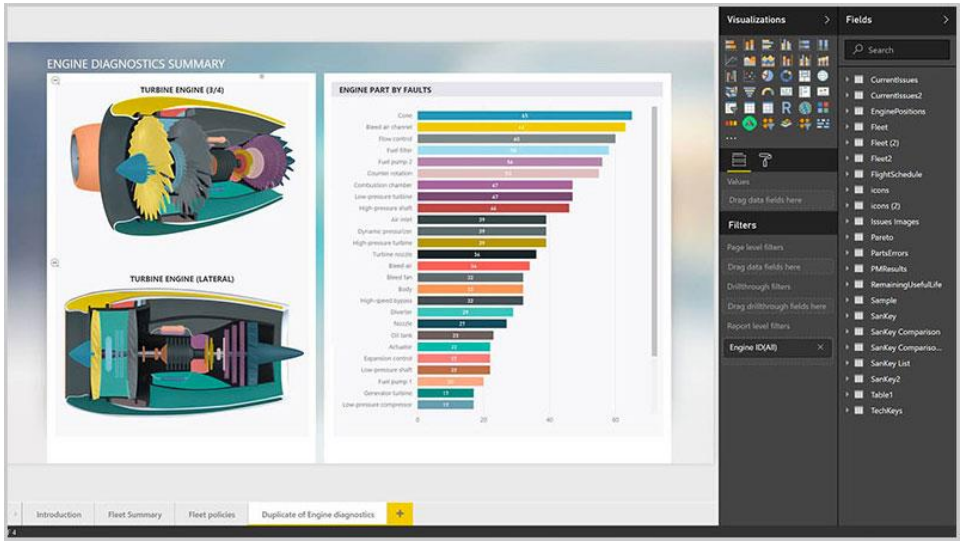
+ a b l e a u

https://www.bmc.com/blogs/gartner-magic-quadrant-analytics-business-intelligence/

Microsoft Power BI

- Various data sources can be connected to the system: third-party applications, cloud services, streaming data, Excel workbooks
- Through the API, you can connect your own applications to the service
- Interactive dashboards are available on any device and display real-time data
- Users can share information in several ways
- The service works on all platforms: cloud, desktop and mobile

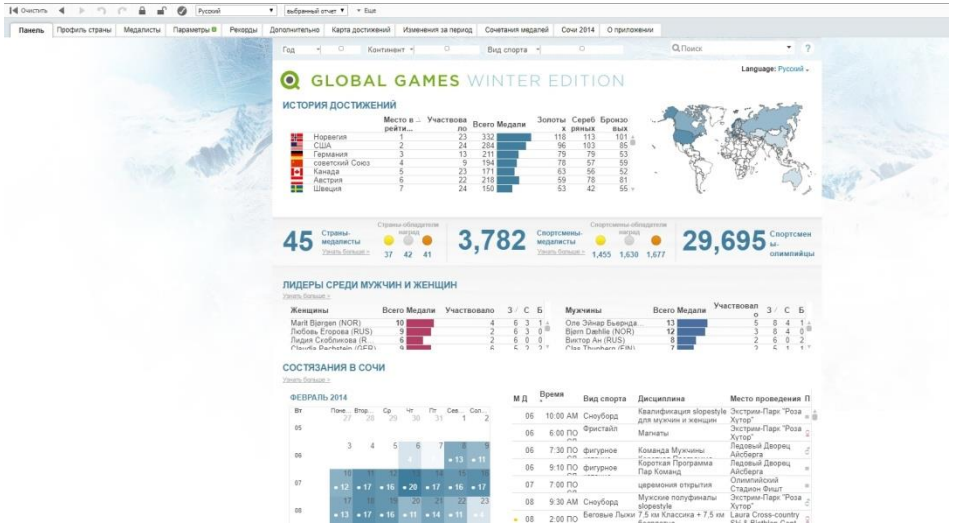


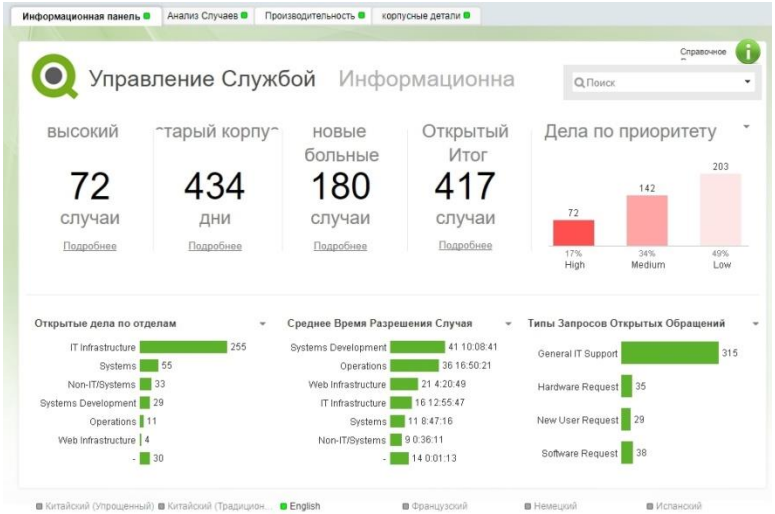


https://powerbi.microsoft.com/en-us/what-is-power-bi/

QlikView

- The uniform analytics algorithm works in all versions
- Products provide fast, interactive data visualization
- The service can be used together with other members of the team, you can share any applications created
- Finished visualizations can be used for samples and studies

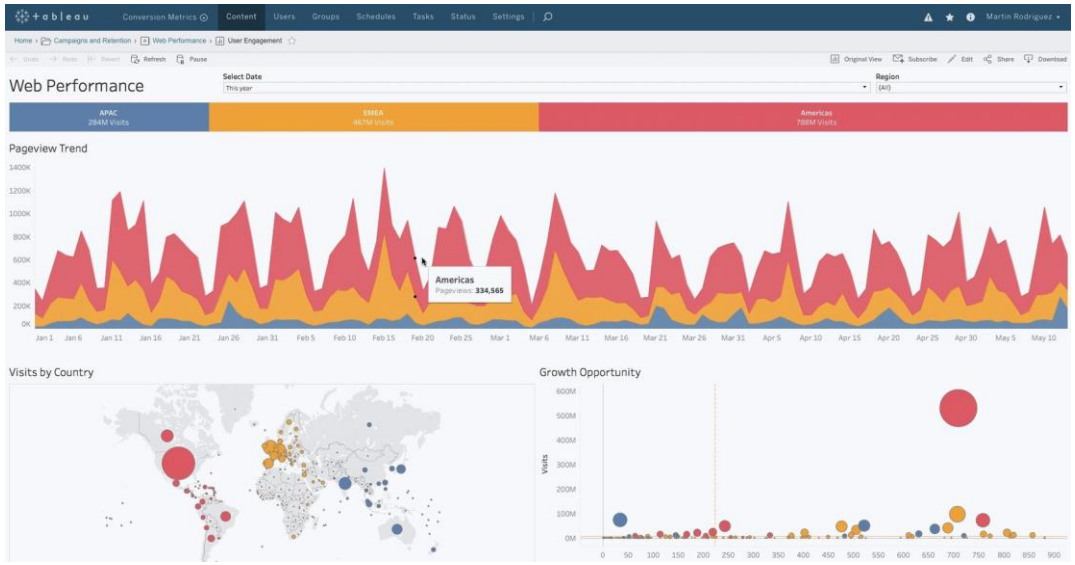


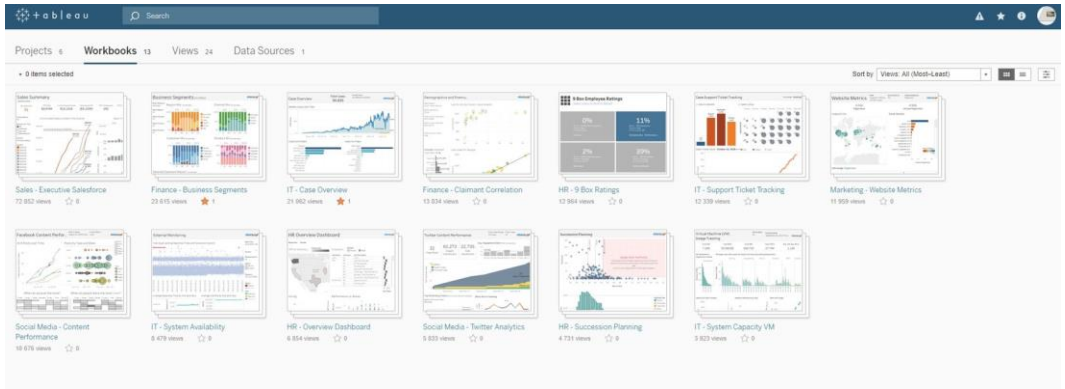


https://www.qlik.com/us/

Tableau

- Users can create tools for dashboards and analytics
- The system works with all devices where there are data streams – no need to worry about hardware requirements or software requirements
- Information panels have access to data warehouses
- Applications for creating dashboards can be created by business users themselves
- Several users can work on the report at once

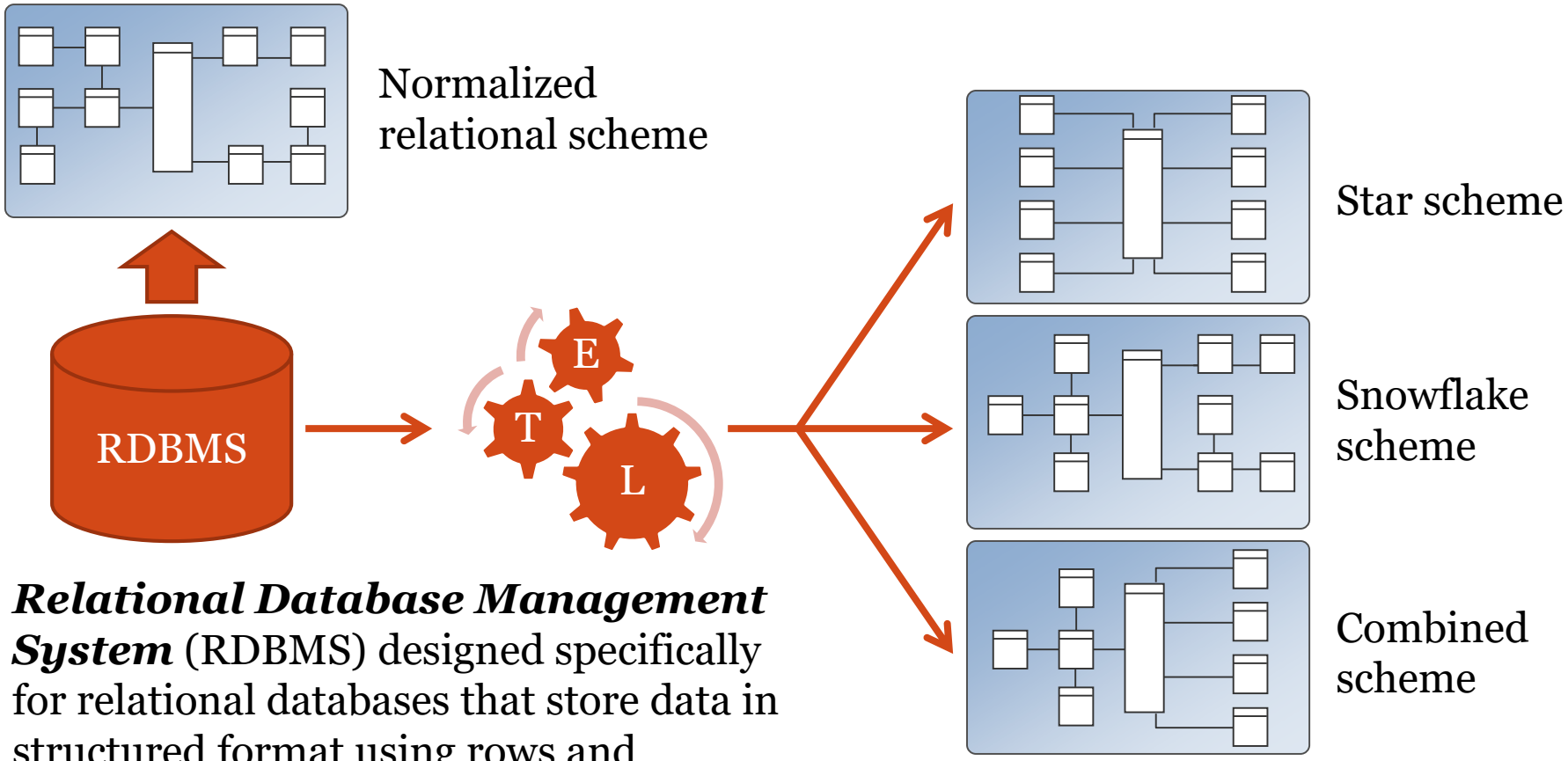




https://www.tableau.com/

Data Warehouses and RDBMS

Data Warehouse (DW) is a central database created for the purpose of making data from all heterogeneous sources useful and accessible



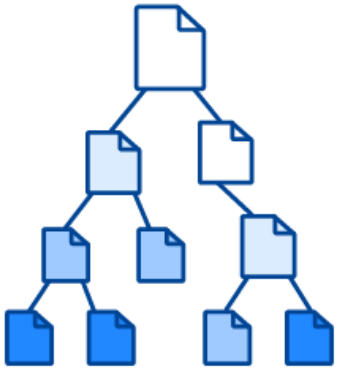
Relational Database Management System (RDBMS) designed specifically for relational databases that store data in structured format using rows and columns of related tables

Other Data Storage Models and DBMS

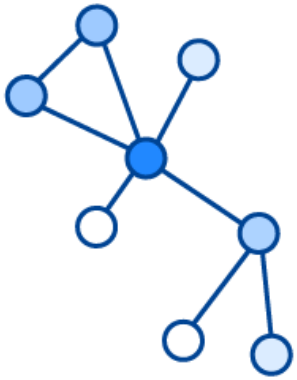
- "NoSQL" does not mean that these systems do not use SQL – almost all of them support limited subset of SQL statements

- "NoSQL" stands for "Not only SQL", since such systems replace some features of RDBMS like consistency and integrity with flexibility and scalability, which are vital for modern cloud applications

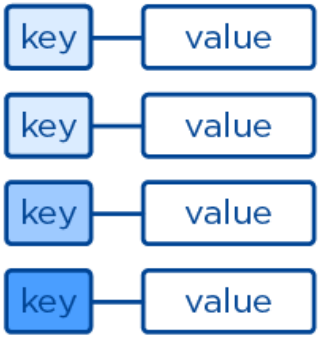
Document



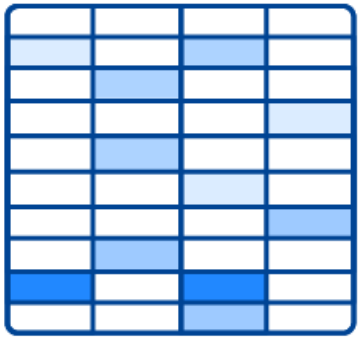
Graph



Key-Value



Wide-column

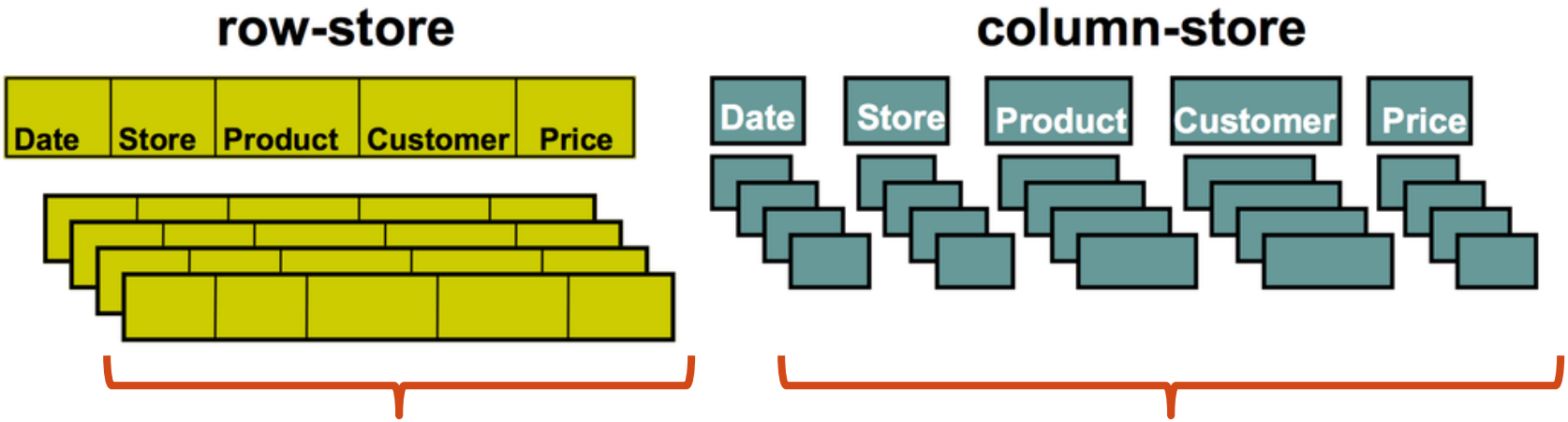


https://www.dbbest.com/technologies/nosql-databases/

Other Data Storage Models and DBMS

Columnar Databases

- Store content by columns rather than rows (unlike RDBMS)
- "Traditional" row-by-row approach of RDBMS keeps all information about entities together
- "Columnar" approach (column-by-column) keeps all information about attributes together



- Easy to add or modify a record
- Slower since need to read unnecessary data

- Reading only relevant data
- Read or modify operations require multiple accesses

https://slideplayer.com/slide/5134369/

Traditional Row Based Storage

ID	Continent	City	Dept	Value
1	Europe	Paris	Sales	£500
2	USA	New York	Sales	£300
3	Europe	Paris	Sales	£700
4	Europe	London	Sales	£500
5	USA	New York	Sales	£200
6	Europe	London	Web	£100

Faster data "slicing"

Column Based Storage

ID	1	2	3	4	5	6
Continent	Europe	USA	Europe	Europe	USA	Europe
City	Paris	New York	Paris	London	New York	London
Team	Sales	Sales	Sales	Sales	Sales	Web
Value	£500	£300	£700	£500	£200	£100

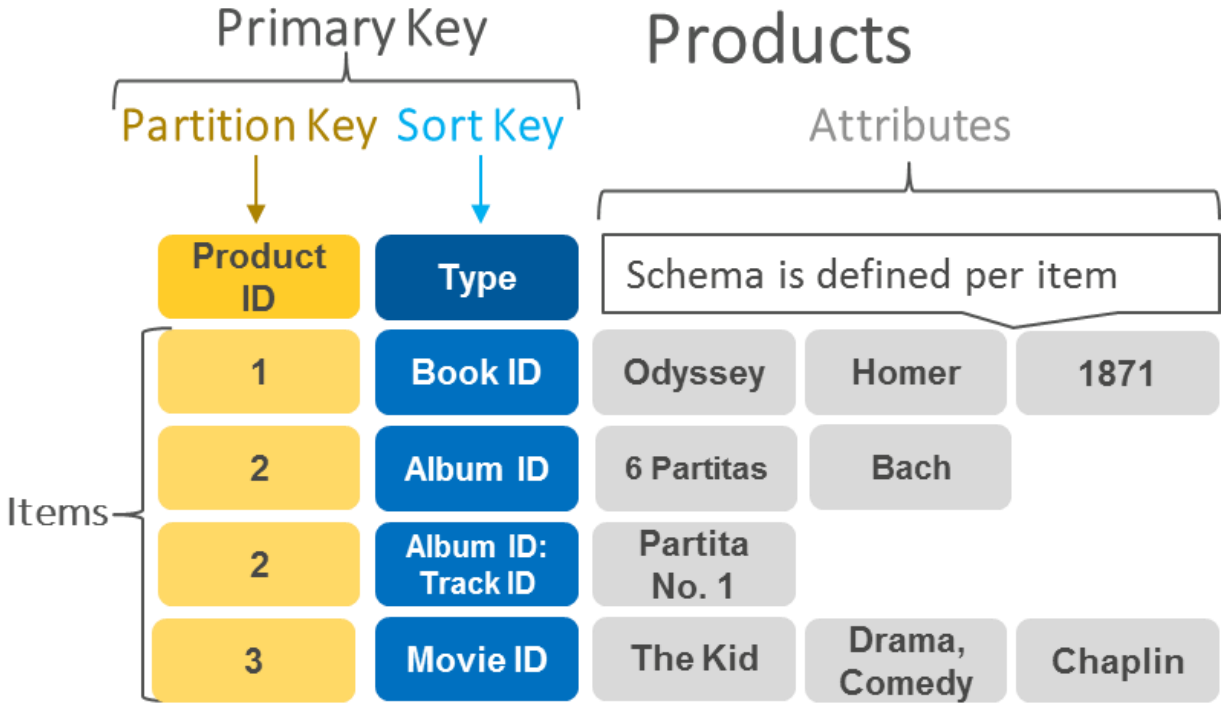
Faster data aggregation

https://dzone.com/articles/oracle-vs-hadoop

Other Data Storage Models and DBMS

Key-Value Stores

- Use simple key-value method to store data
- Data is stored as a collection of key-value pairs
- Key values can be any kind of objects and serve as a unique identifiers of pairs



- Extremely scalable in compare to other storage models
- Allows horizontal scaling
- Used to store sessions in web applications, shopping carts in e-commerce solutions, gaming industry

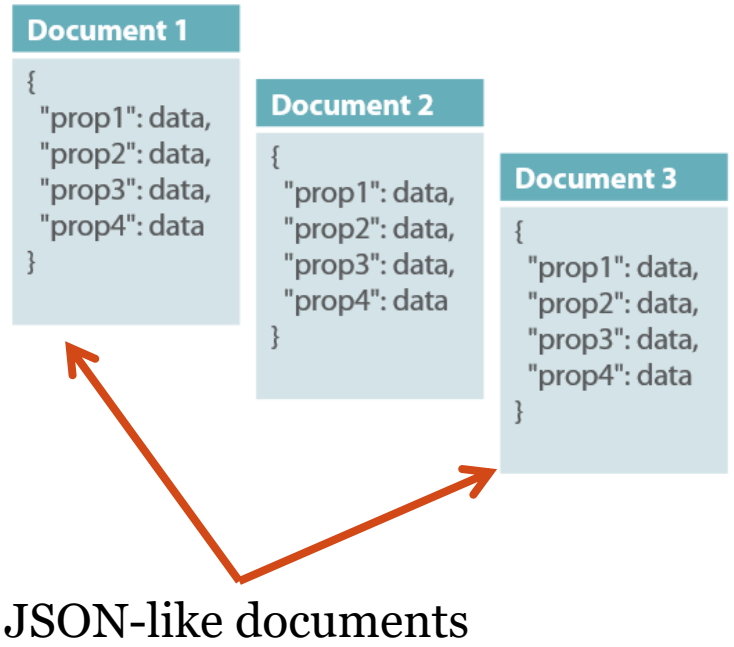
https://aws.amazon.com/nosql/key-value/

Other Data Storage Models and DBMS

Document-Oriented Databases

- Designed to store and query data as JSON-like documents (XML documents for earlier systems)
- Easier to use the same document-model format that is used in applications code
- Easy to support flexible, semi-structured, and hierarchical documents

Col1	Col2	Col3	Col4
Data	Data	Data	Data
Data	Data	Data	Data
Data	Data	Data	Data

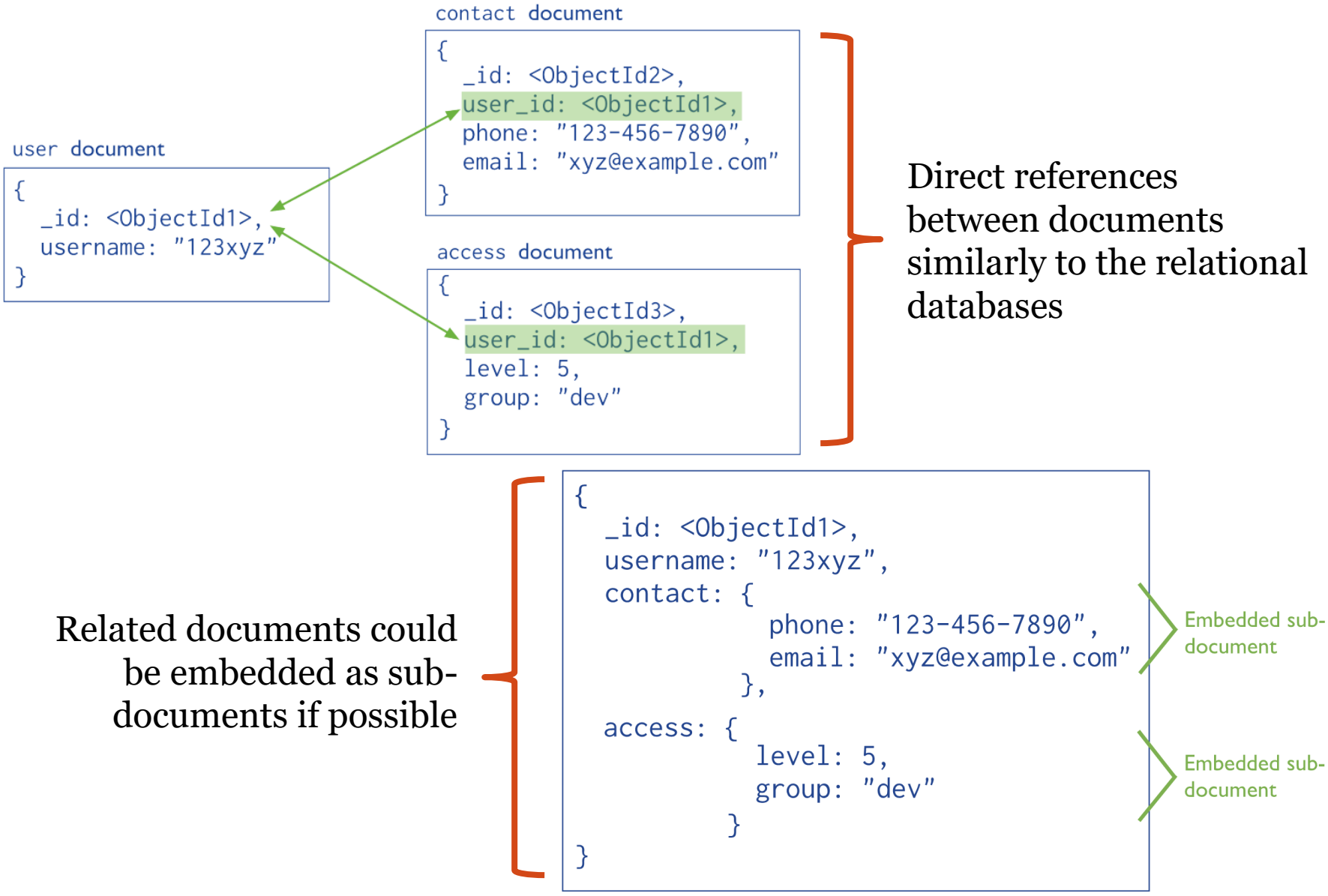


JSON-like documents

Used in:

- Content-management systems
- User profiles
- Catalogs for E-Commerce

https://lennilobel.wordpress.com/2015/06/01/relational-databases-vs-nosql-document-databases/

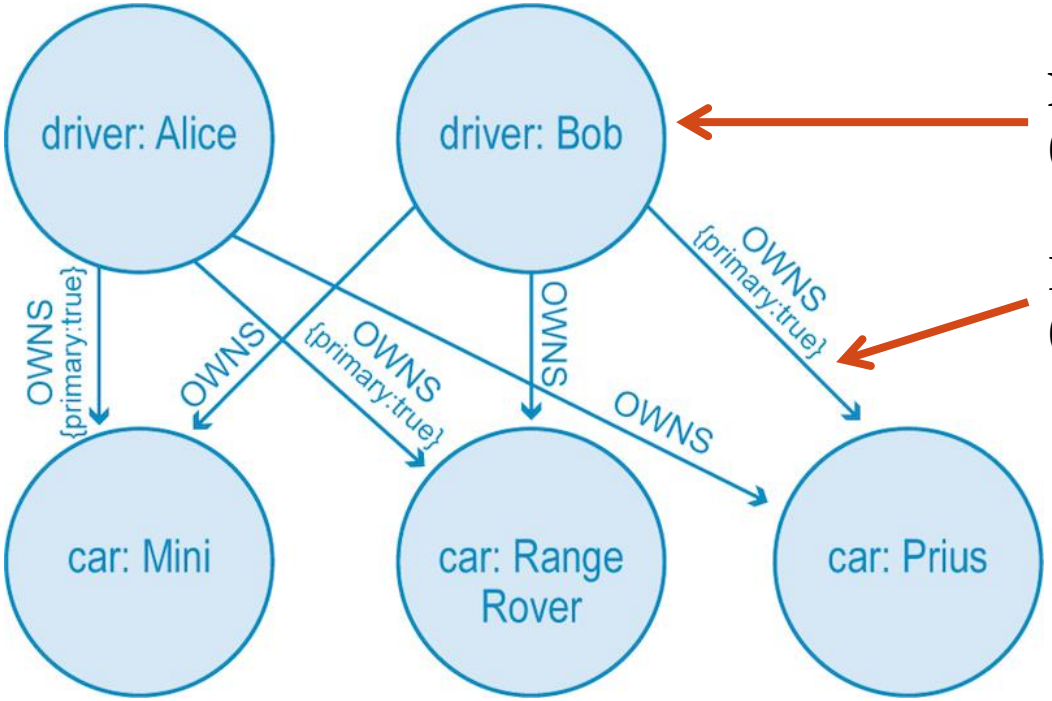


https://dzone.com/articles/design-decisions-the-data-model-mongodb

Other Data Storage Models and DBMS

Graph-Based Databases

- Built to store and navigate relationships
- Use nodes to store data entities and edges to store relationships
- Very fast traversing since relationships are persisted in the database and do not need to be calculated at query times



Node
(data entity)

Edge
(relationship between entities)

Used in:

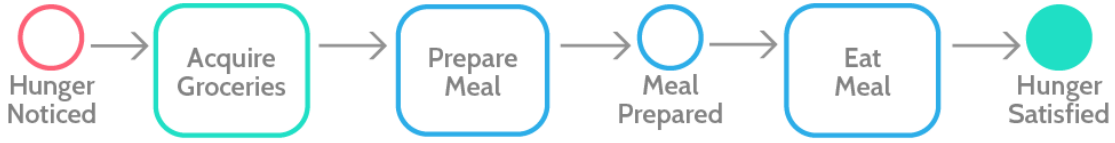
- Recommendation engines
- Fraud detection systems
- Social networks
- E-Commerce systems

https://neo4j.com/blog/other-graph-database-technologies/

BPM and CEP

Business Process Management (BPM)
modeling, automation, execution of person-to-person or system-to-system tasks linked together based on conditions (business processes)

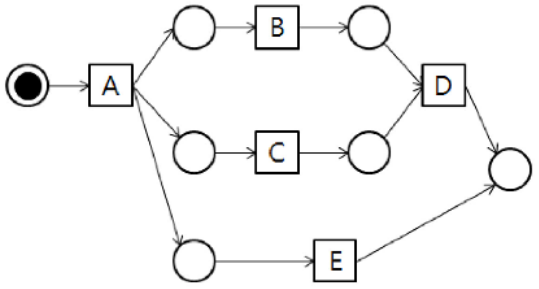
Complex Event Processing (CEP)
real-time capturing and analyzing data streams (mostly event logs of enterprise information systems) to identify opportunities or threats



Process "traces" or "footprints"

Case ID	Activity ID	Resource
Case 1	A	John
Case 2	A	Sue
Case 3	A	Sue
Case 1	B	Mike
Case 1	C	John
Case 1	D	Pete
Case 2	C	Carol
Case 3	E	Sue
Case 2	B	Sue
Case 2	D	Pete

Event Log L1



Process Model M1

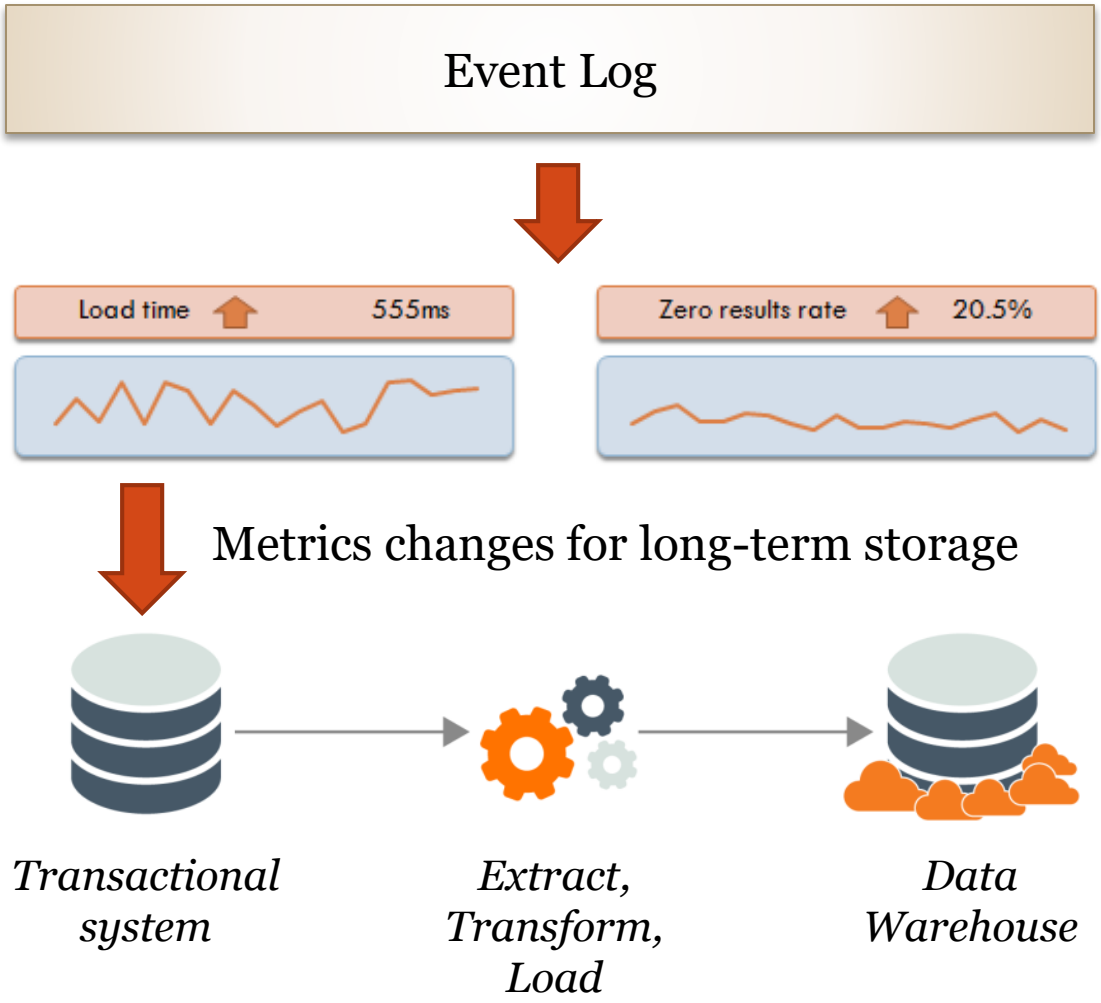
Process Mining
detect activities sequences and process behavior patterns in event logs to build business process model "as-is"

https://www.process.st/bpmn-tutorial/
https://www.researchgate.net/figure/A-process-model-M1-discovered-from-an-event-log-L1-by-a-process-mining-algorithm_fig4_264065543

BAM and ETL

Business Activity Monitoring (BAM)
real-time business process monitoring, indicating and alerting about deviations of metrics from target values

Extract, Transform, Load (ETL)
data extraction from external sources, its transformation, cleaning, and loading to the warehouse

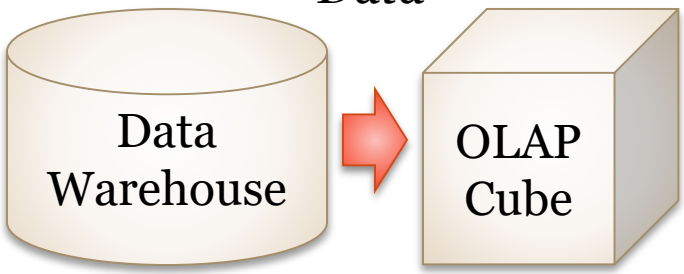


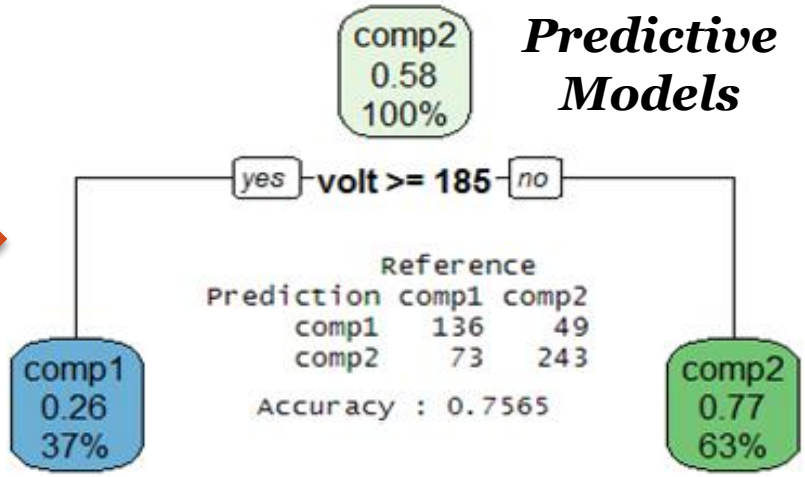
https://www.researchgate.net/figure/A-process-model-M1-discovered-from-an-event-log-L1-by-a-process-mining-algorithm_fig4_264065543

Predictive Analytics

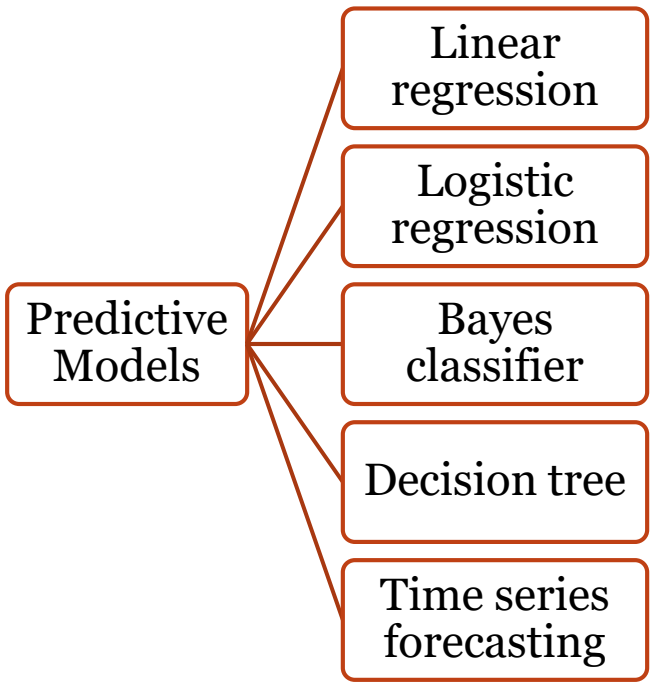
Aggregated Data

Insights & Intelligence

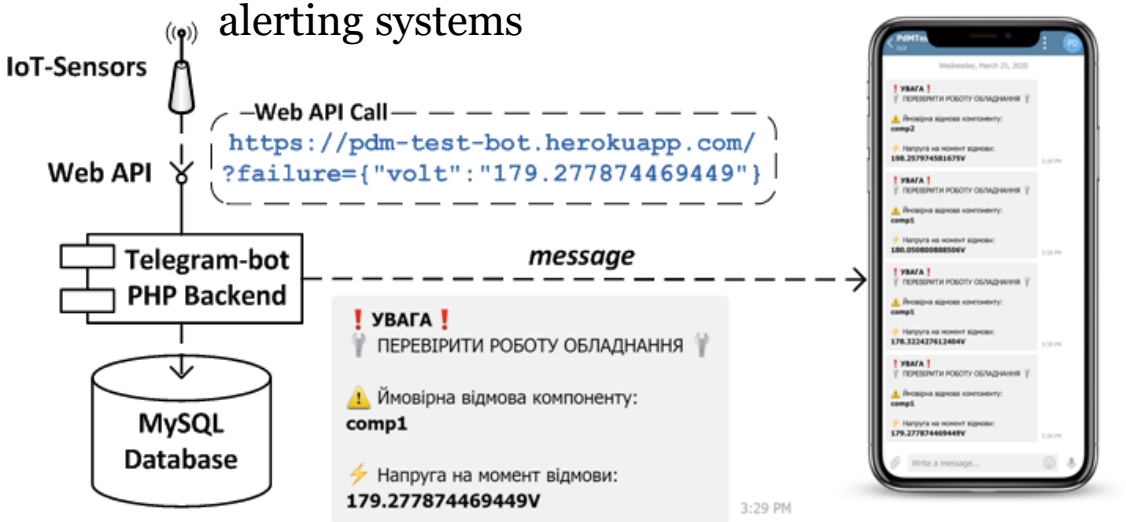




Predictive Models



Predictive models are used to build advanced alerting systems

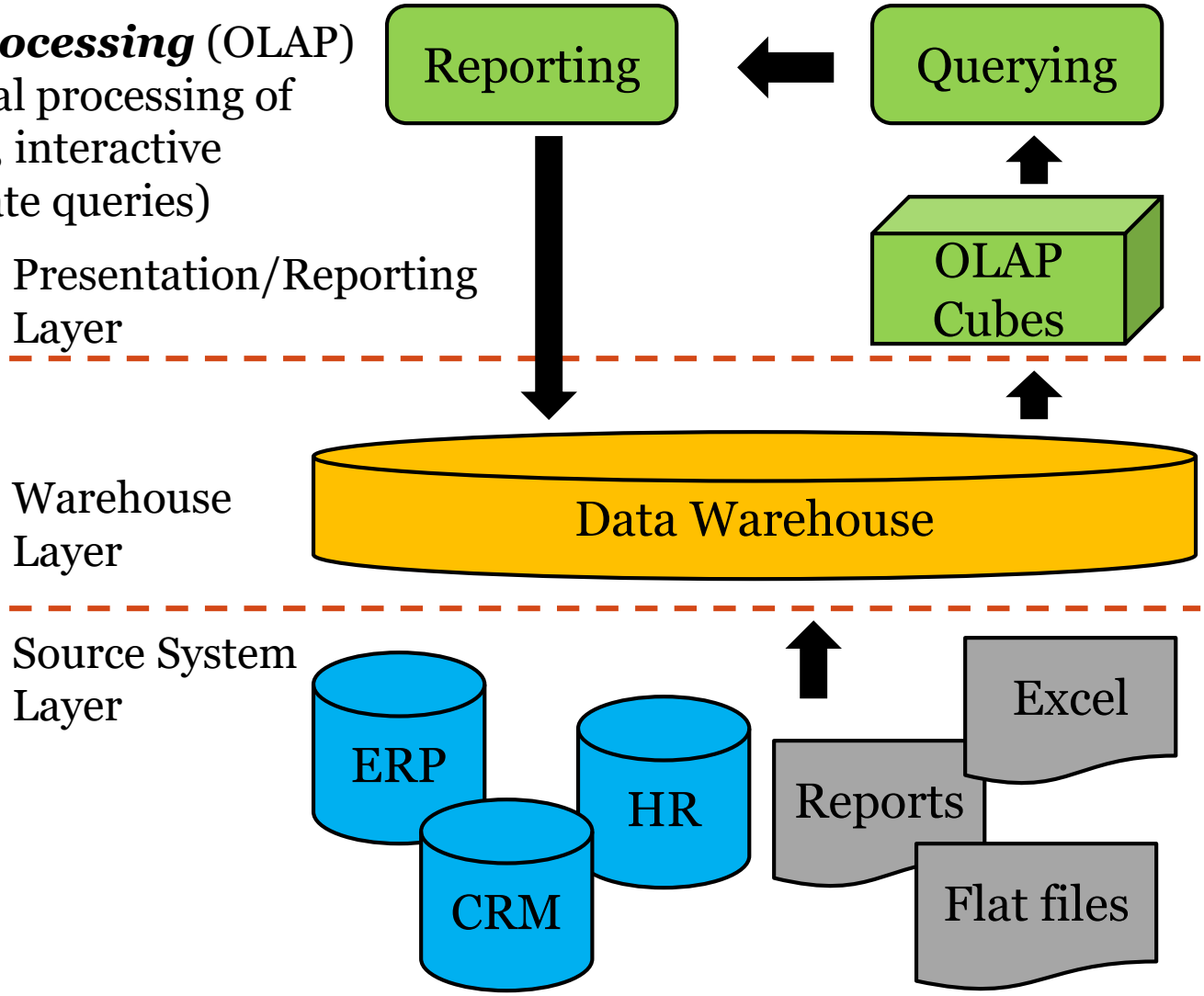


OLAP

Online Analytical Processing (OLAP) aims at multidimensional processing of large data volumes (fast, interactive answers to large aggregate queries)

Querying (ad-hoc) helps to explore business questions while looking through the data (online)

Reporting is turning raw data into information that can be used to make intelligent business decisions



FASMI Test for OLAP Systems

Fast Analysis of Shared Multidimensional Information

- Fast

(responses query in 1-20 seconds)

- Analysis

(supports any complex logical or statistical analysis for business applications)

- Shared

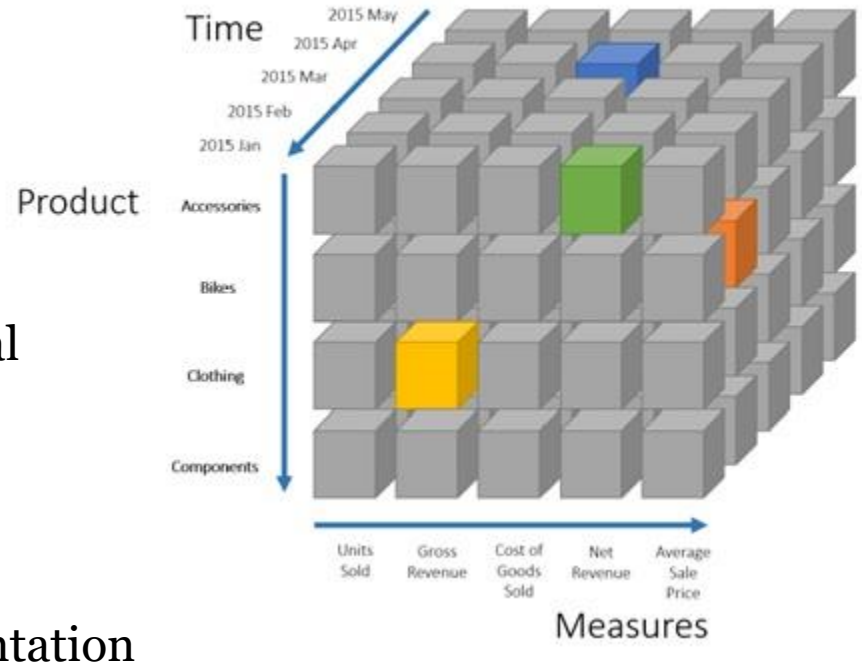
(provides secured multiuser access)

- Multidimensional

(supports multidimensional data representation including hierarchies)

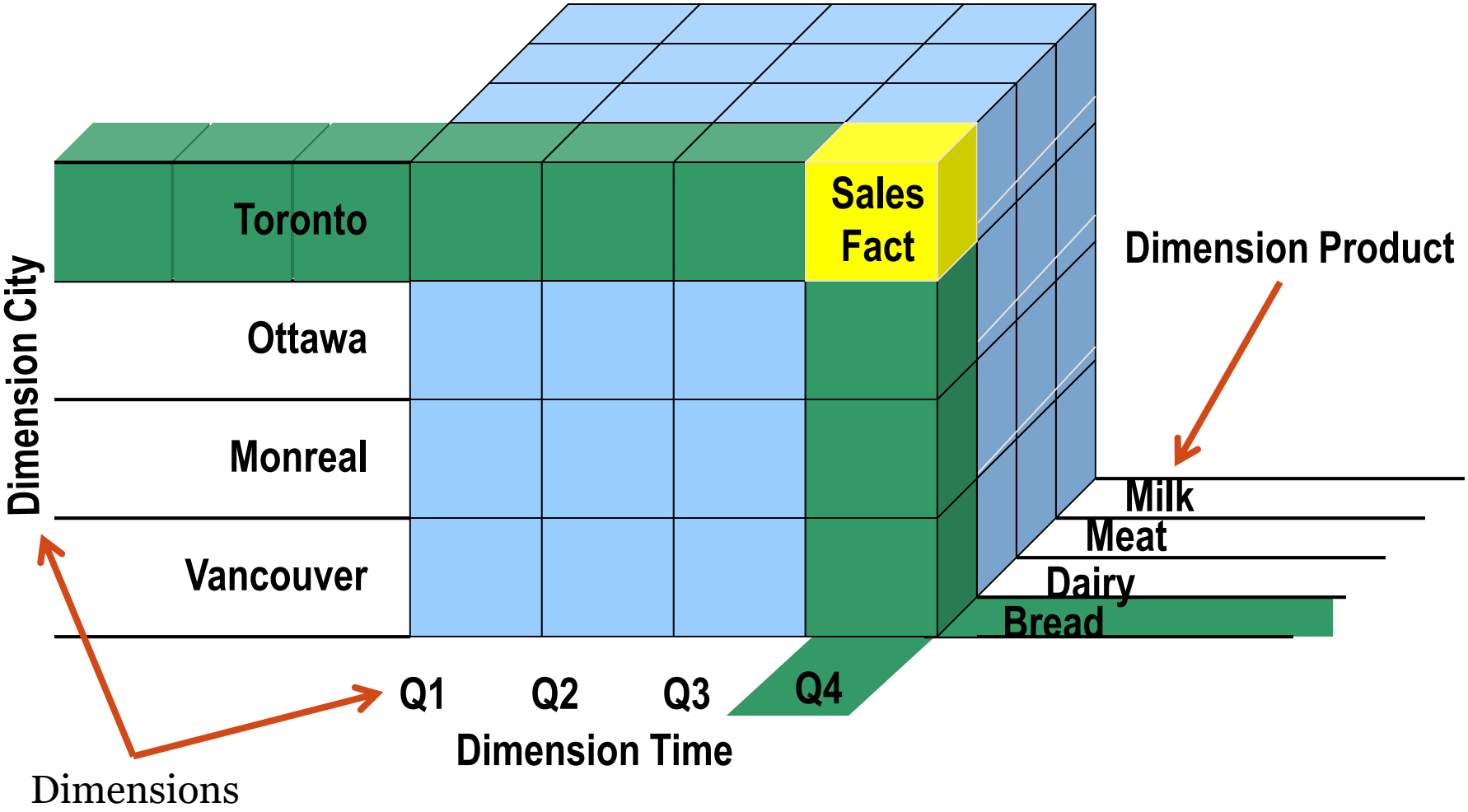
- Information

(processes huge volume of data and information)



https://asperasolutions.com/to-bi-or-not-to-bi/

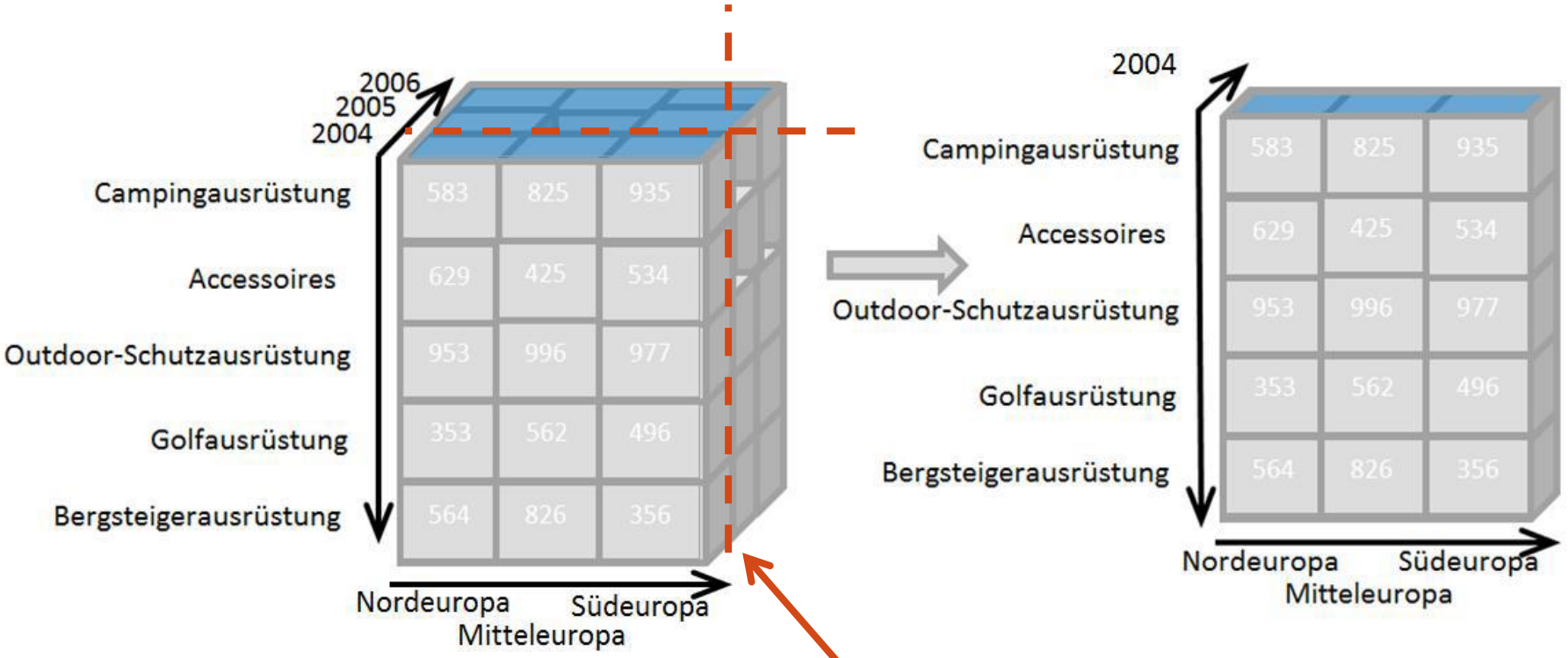
OLAP: Why Multidimensional?



Codd Rules for OLAP: Basic Features

Rule	Description
Multidimensional Conceptual View	Support slice and dice, drill down, roll up, and pivot operations
Intuitive Data Manipulation	Multidimensional manipulations should be intuitive for users
Accessibility: OLAP as a Mediator	Wrap physical data stores with its logical model
Batch Extraction vs. Interpretive	Provide efficient access to both internal and external data warehouse
OLAP Analysis Models	Support parameterized queries, multidimensional analysis operations, “what if” analysis, goal-based simulations
Client-Server Architecture	Separate data storage and analysis
Transparency	Encapsulate data model and tools
Multi-User Support	Provide concurrent access, integrity and security

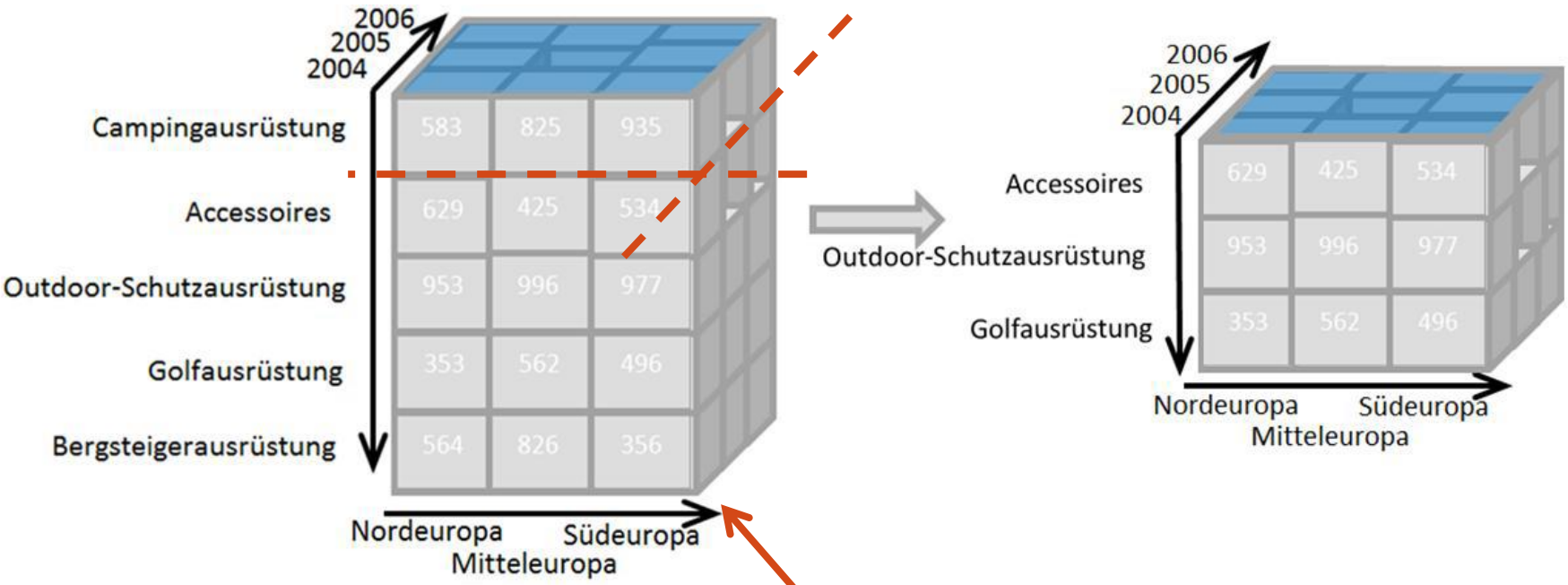
OLAP: Slice



The OLAP cube is sliced vertically by the "Year" dimension with the value 2004

https://commons.wikimedia.org/wiki/File:OLAP_slicing.png

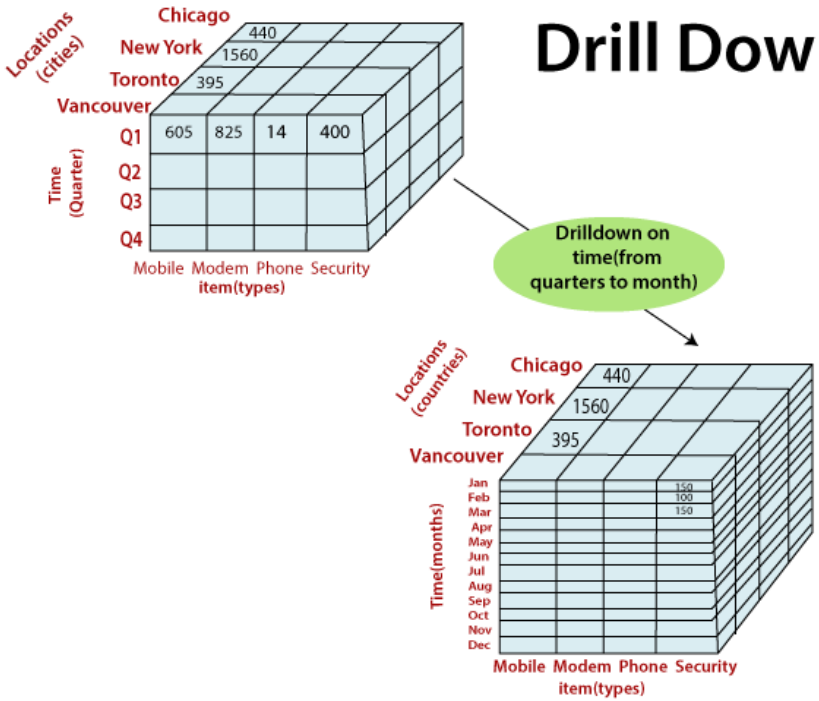
OLAP: Dice



The OLAP cube is sliced by multiple measures horizontally by the "Categories" dimension

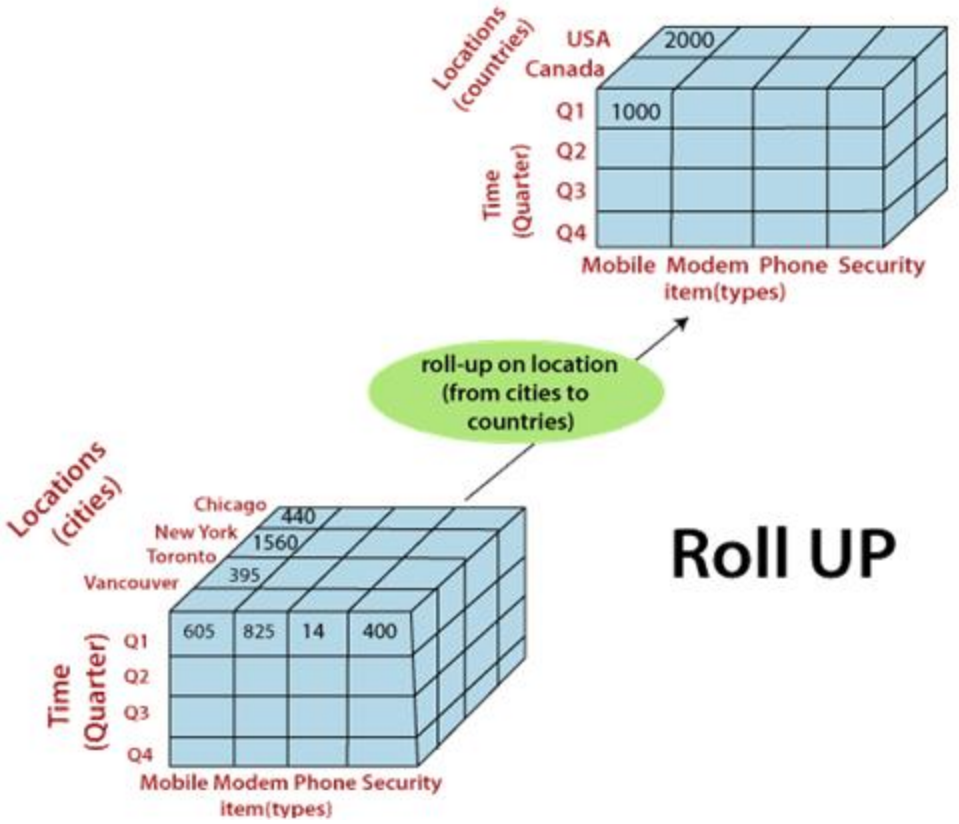
https://commons.wikimedia.org/wiki/File:OLAP_dicing.png

OLAP: Drill Down and Roll Up



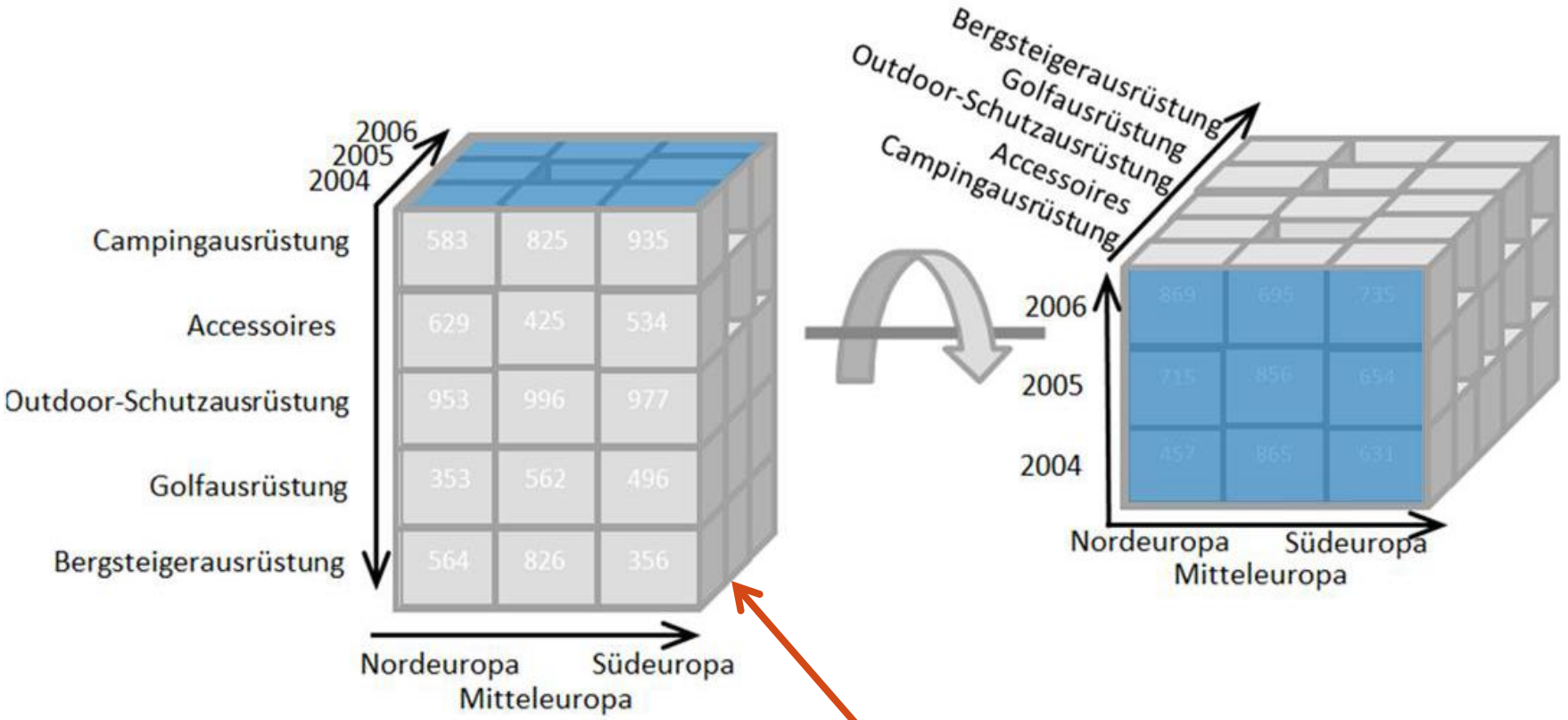
Drill Down
Transition from aggregated to detailed data

Roll Up
Transition from detailed to aggregated data



https://www.javatpoint.com/olap-operations

OLAP: Pivot



The OLAP cube is rotated by one of its measures (like matrix transposition)

https://commons.wikimedia.org/wiki/File:OLAP_pivoting.png

Codd Rules for OLAP: Special Features

Rule	Description
Treatment of Non-Normalized Data	OLAP modifications should not affect data stored in external source systems Non-normalized data sources should be integrated with OLAP systems
Storing OLAP Results: Keeping Them Separate from Source Data	OLAP applications should not work directly with processed data Data modified by OLAP should be separated from transactional data
Extraction of Missing Values	Missing values should be distinguished from empty values
Treatment of Missing Values	All missing values should be ignored by OLAP system independently from their sources

Codd Rules for OLAP: Reporting Features

Rule	Description
Flexible Reporting	OLAP systems should support various kinds of data visualization Reports should be displayed in any possible format
Consistent Reporting Performance	Increasing of dimensions and data volumes should not lead to decreasing of performance This rule assures usability and simplicity of OLAP usage
Automatic Adjustment of Physical Level (replacement of original rule Dynamic Sparse Matrix Handling)	OLAP systems should provide efficient processing of sparse matrix Access speed should remain constant despite data cells locations, number of dimensions, and data density

Codd Rules for OLAP: Dimension Control Features

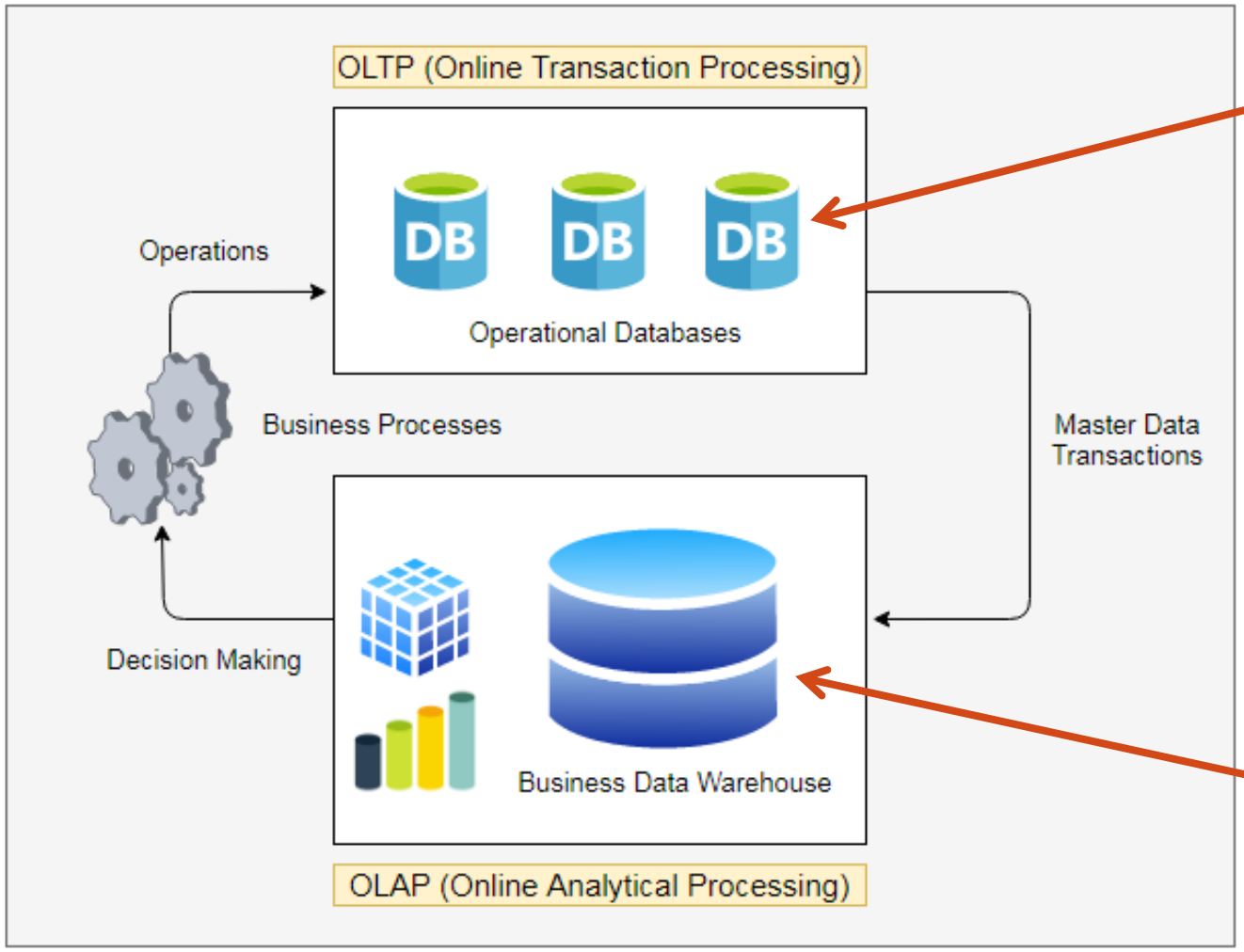
Rule	Description
Generic Dimensionality	All dimensions should be equal Since dimensions are symmetric, each dimension could be complemented with additional features Data structures, expressions, and reports should not depend on a single dimension
Unlimited Dimensions and Aggregation Levels	It is recommended to allow at least 15 or even 20 dimensions to be used in a OLAP tool Each of these dimensions should allow unlimited aggregation levels defined by users
Unrestricted Cross-Dimensional Operations	Calculations and data manipulations on any number of dimensions should not deny or restrict any relations between data cells

OLAP vs. OLTP

Feature	OLTP	OLAP
Source of data	Operational data – OLTPs are original source of data	Consolidation data – OLAP data comes from various OLTP databases
Purpose of data	Control and run business tasks	Support planning, problem solving and decision making
What the data	Snapshot of ongoing business processes	Multi-dimensional views of various business activities
Inserts and updates	Short and fast inserts and updates initiated by end users	Periodic long-running batch uploads to refresh the data
Queries	Standardized and simple queries (CRUD), returning relatively few records	Complex queries involving aggregations
Processing speed	Typically very fast	Depends on the amount of data involved, batch data updates and complex queries may take time
Space requirements	Relatively small if historical data is archived	Larger due to the aggregation structures and history data
Database design	Highly normalized with many tables	Typically de-normalized with fewer tables using star and/or snowflake schemas
Backup and recovery	Frequent backups – operational data is critical to run the business	Reloading the OLTP data as a recovery method

http://www.nicobudidarmawan.com/2014/01/business-intelligence-oltp-vs-olap.html

OLAP vs. OLTP
Source of data

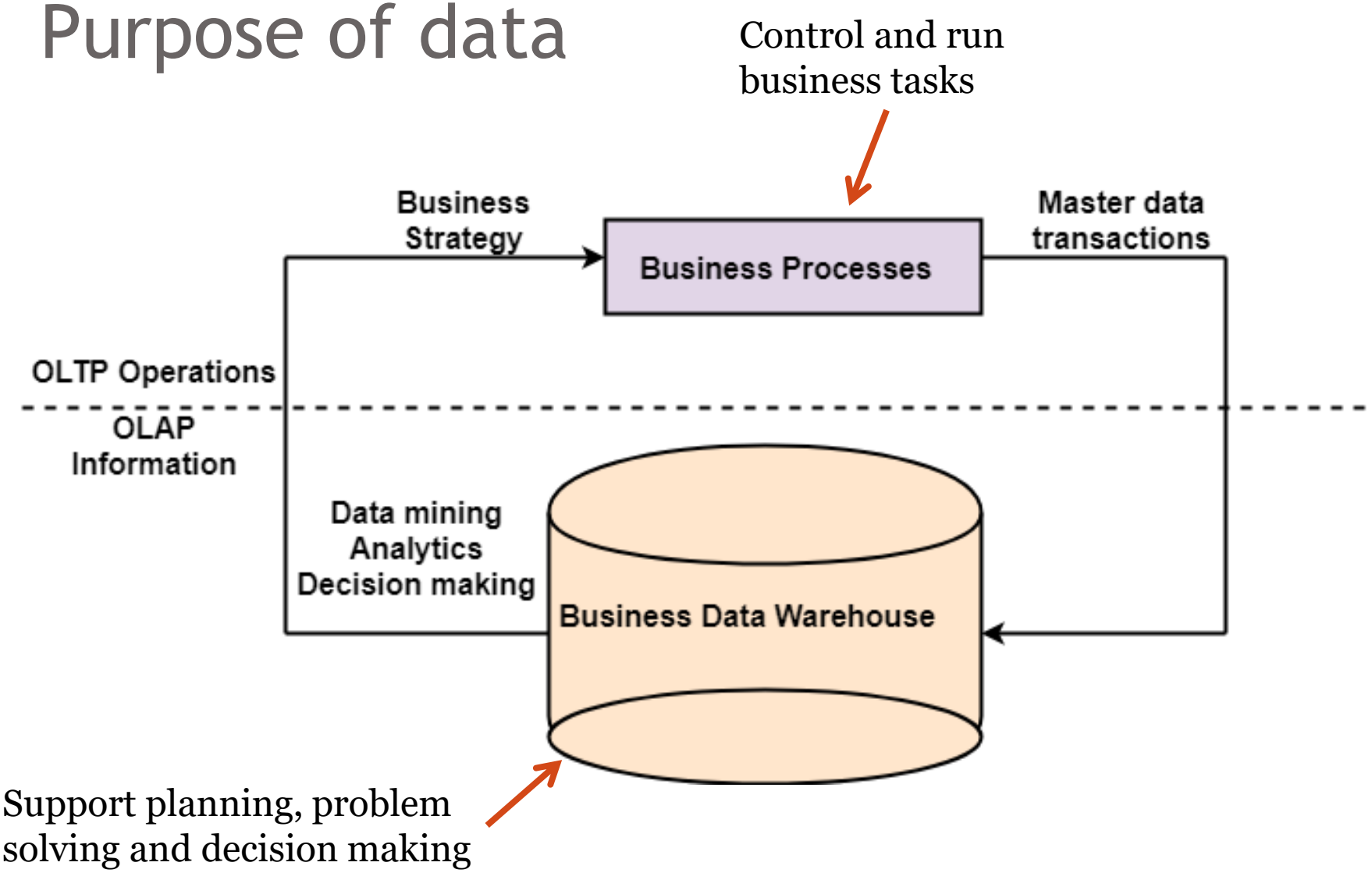


Operational data – OLTPs are original source of data

Consolidation data – OLAP data comes from various OLTP databases

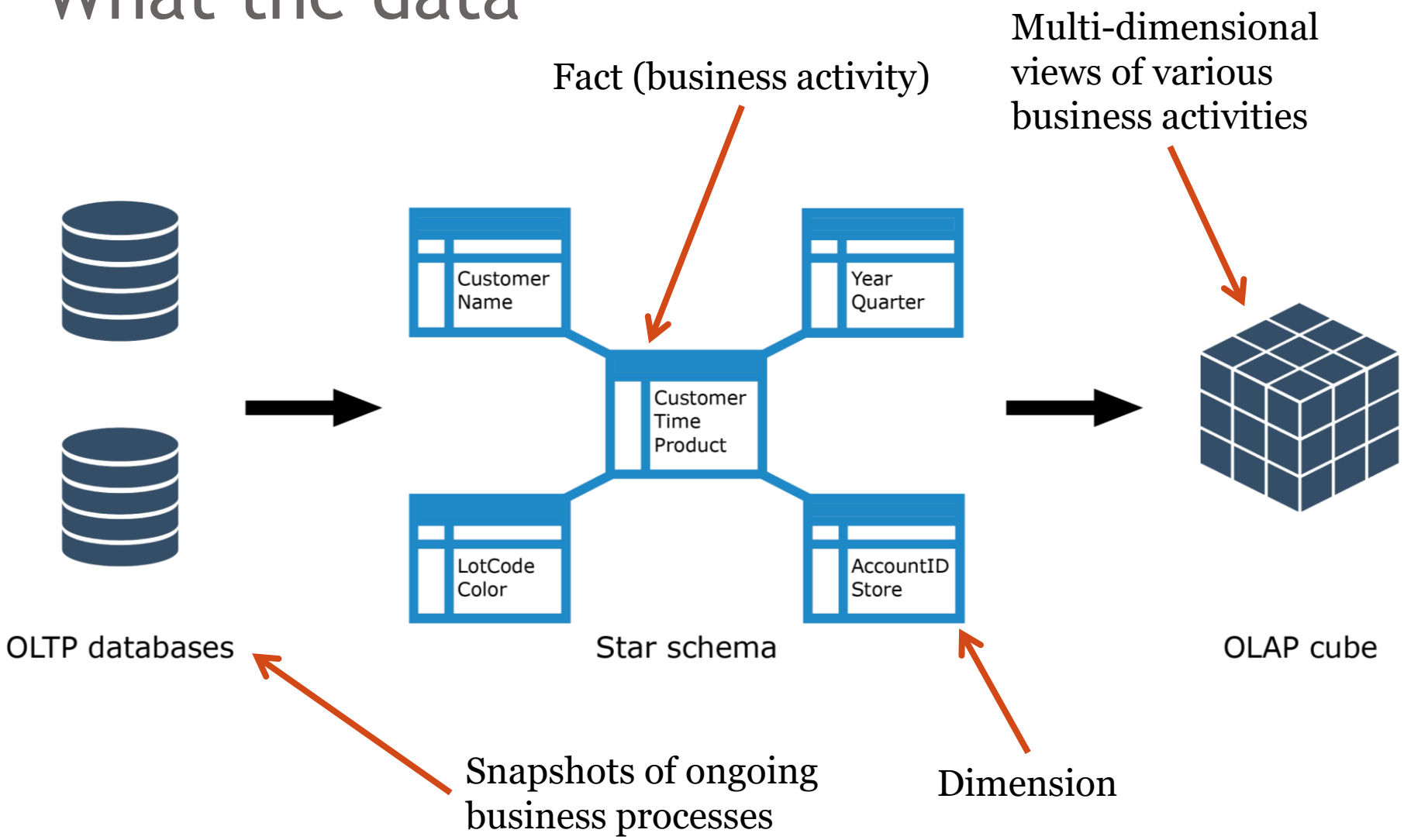
https://sqlwizard.blog/2020/03/15/sql-server-oltp-vs-olap/

OLAP vs. OLTP
Purpose of data



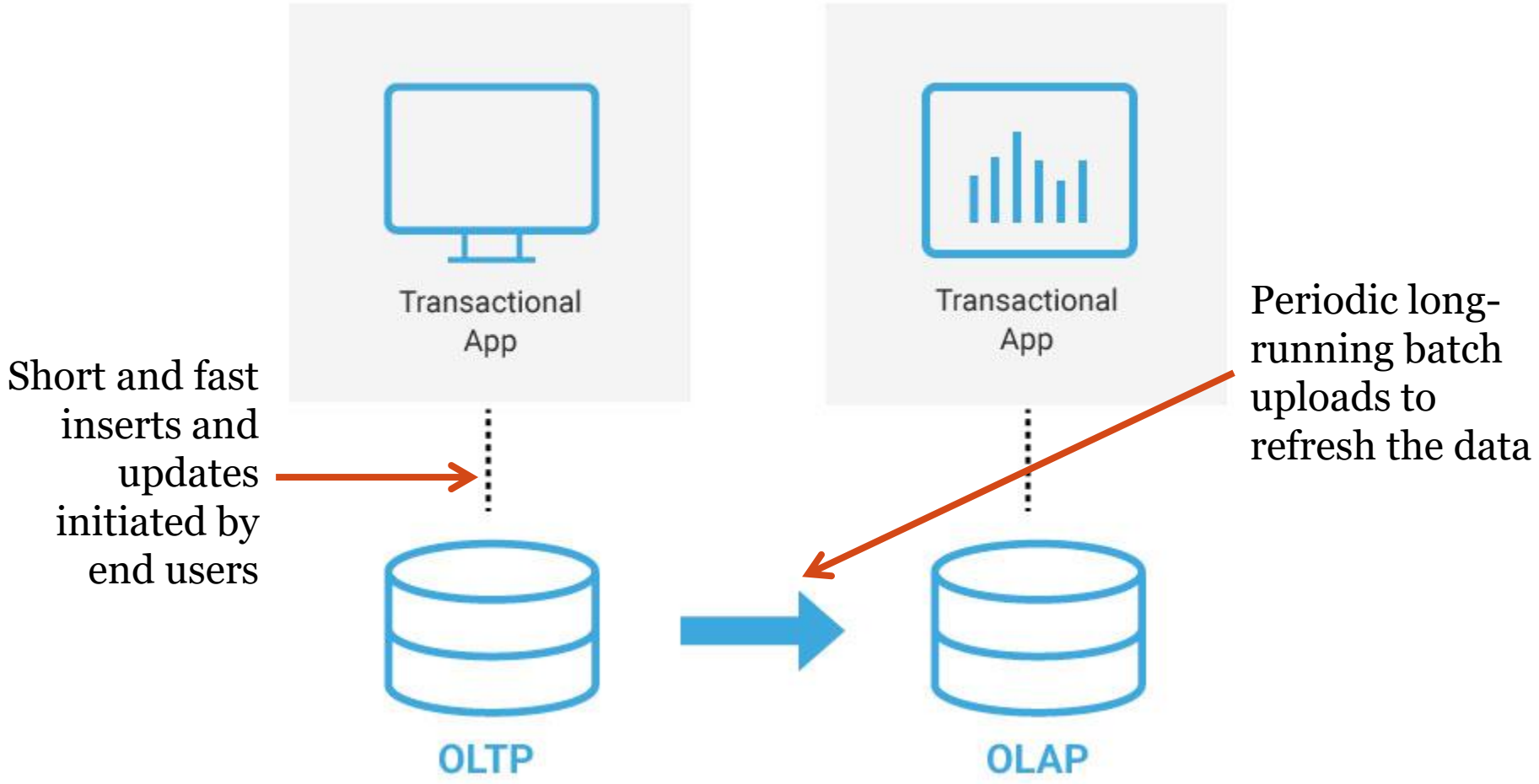
https://www.javatpoint.com/oltp-vs-olap

OLAP vs. OLTP
What the data



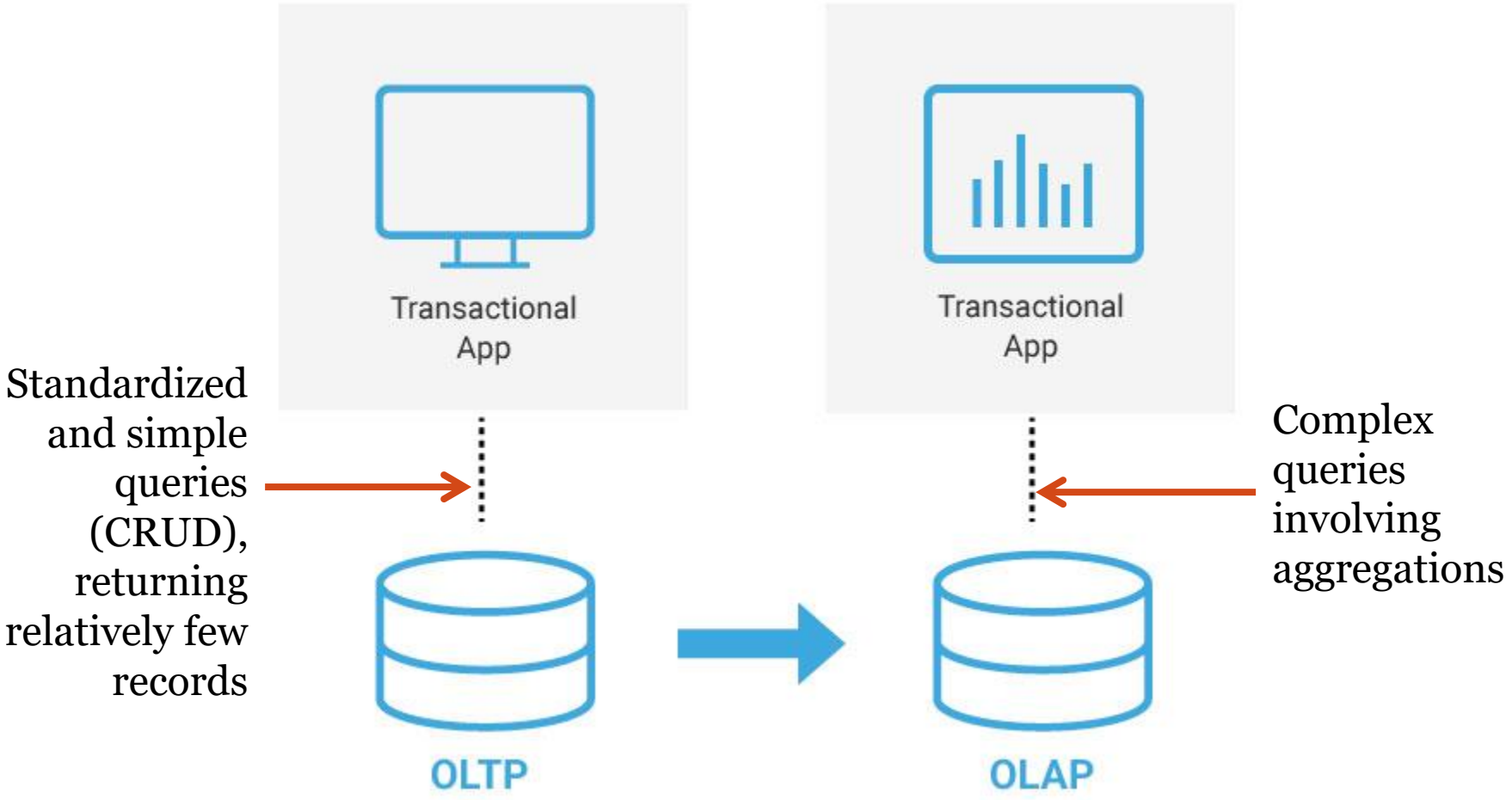
https://www.softwareadvice.com/resources/olap-data-warehouse-alternatives/

OLAP vs. OLTP Inserts and Update



https://blog.sqldbm.com/bi-data-modeling-with-sqldbm/

OLAP vs. OLTP Queries

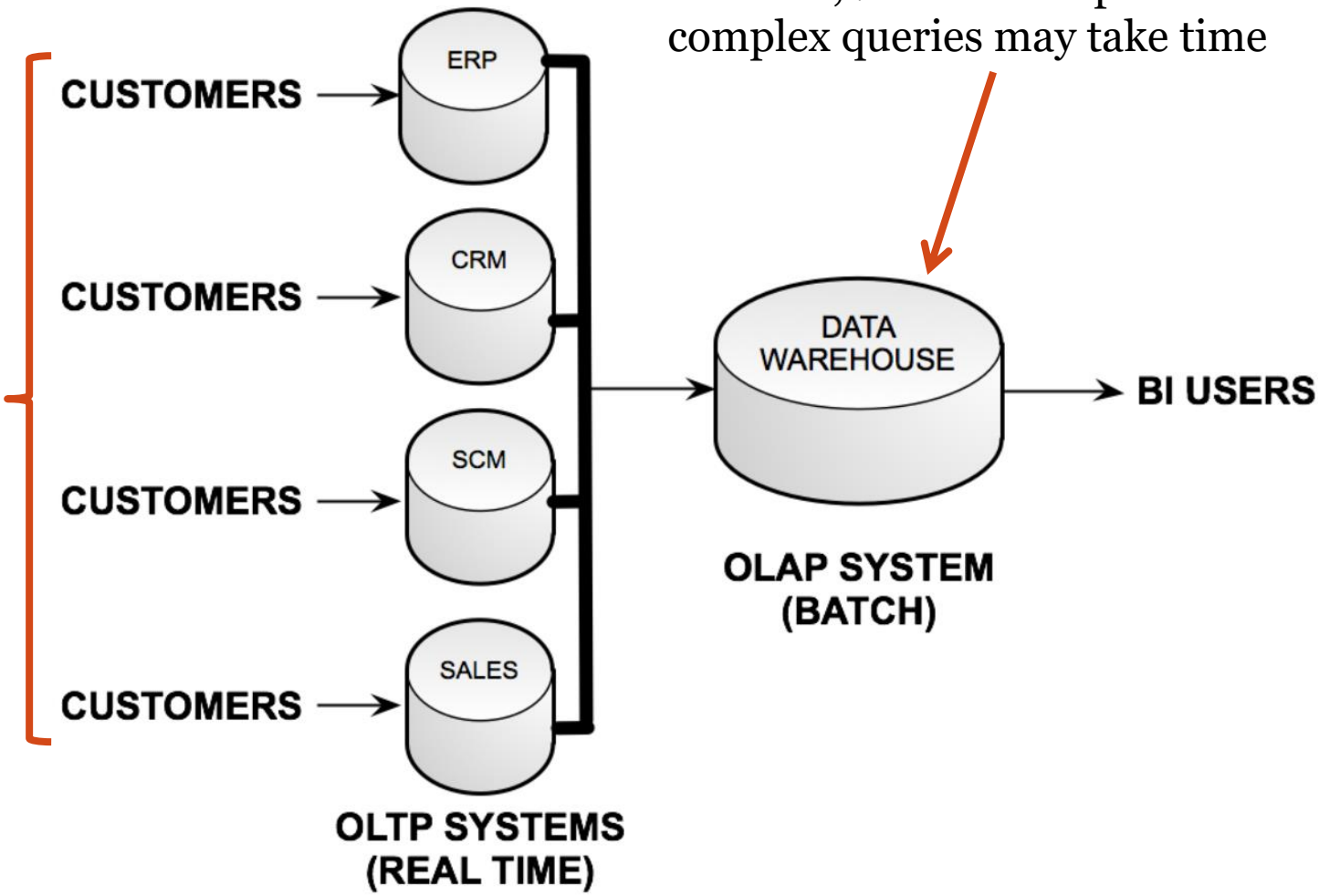


https://blog.sqldbm.com/bi-data-modeling-with-sqldbm/

OLAP vs. OLTP Processing speed

Depends on the amount of data involved, batch data updates and complex queries may take time

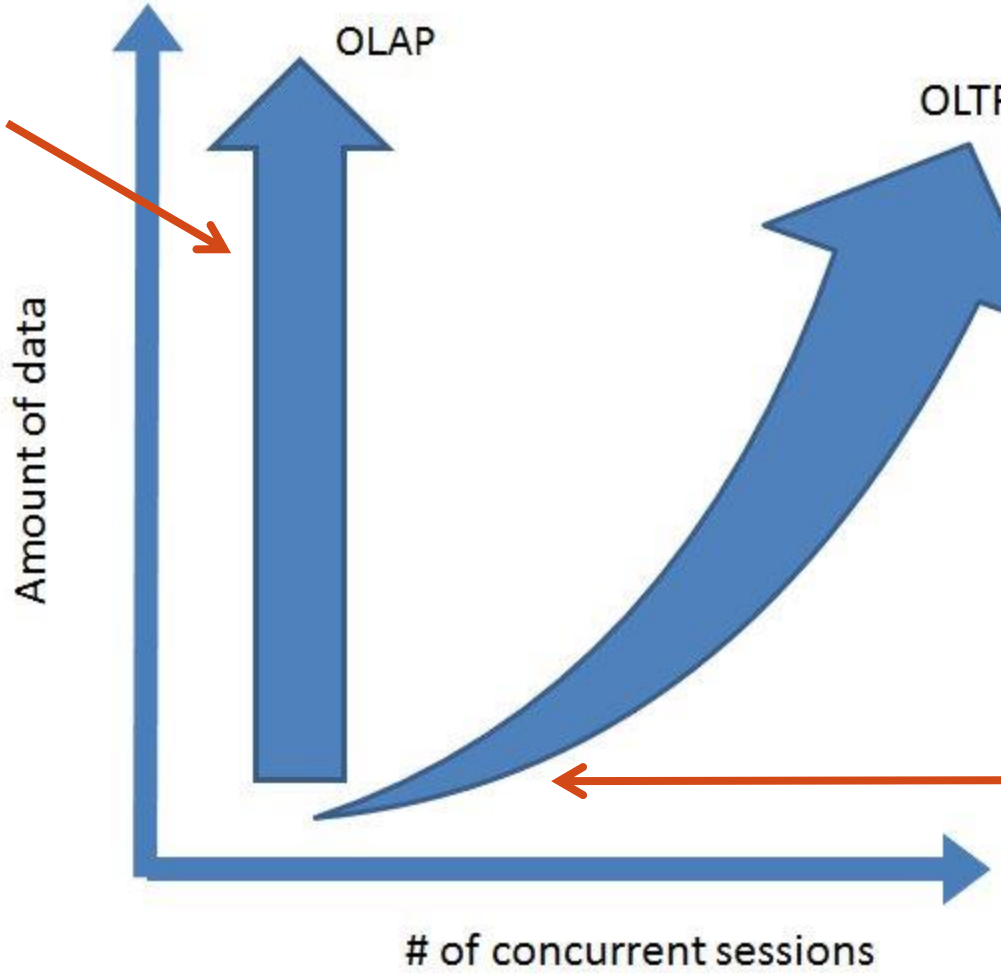
Typically very fast and short transactions



https://novacontext.com/an-introduction-to-amazon-redshift/

OLAP vs. OLTP
Space requirements

Larger due to the aggregation structures and history data



Relatively small if historical data is archived

https://www.dbbest.com/blog/extract-transform-load-etl-technologies-part-1/

OLAP vs. OLTP Database design

Highly normalized with many tables

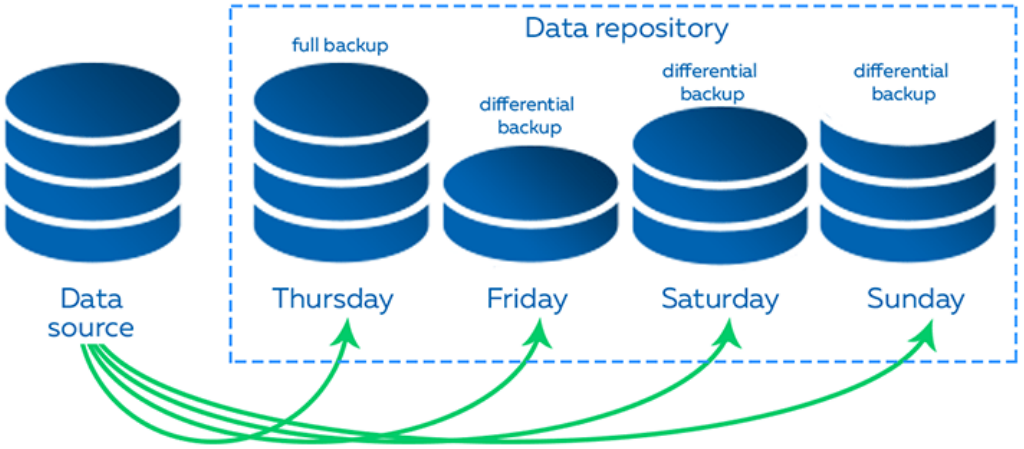
Typically de-normalized with fewer tables using star and/or snowflake schemas

https://medium.com/@shristibal1998/olap-vs-oltp-52828ff1ec93

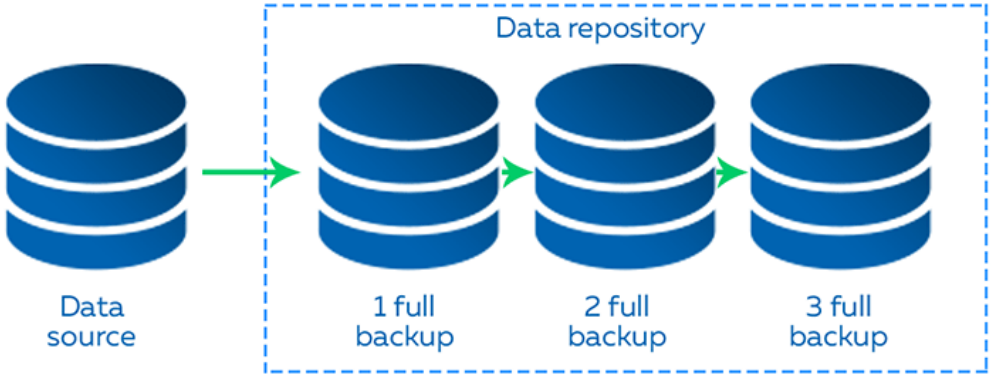
OLAP vs. OLTP
Backup and recovery

OLTP systems support frequent backups – operational data is critical to run the business

Differential



Full



OLAP data warehouses tend to support reloading the OLTP data as a recovery method

https://www.handybackup.net/full-vs-differential-sql-backup.shtml

OLAP vs. OLTP Summary

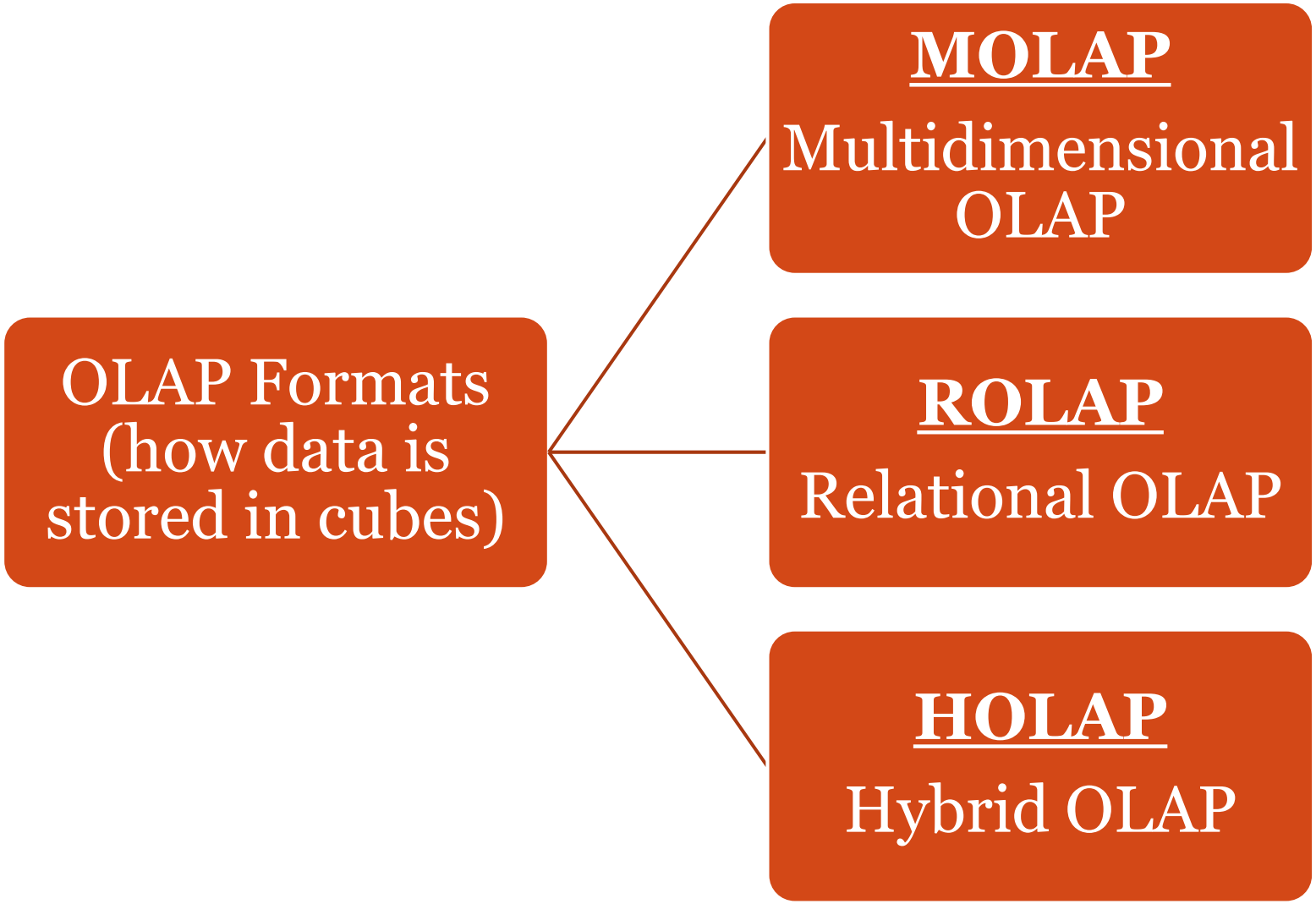
OLTP
A data modeling approach typically used to facilitate and manage usual business applications

OLAP
An approach to answer multi-dimensional queries to provide valid and useful analytics to management

Parameters	OLTP	OLAP
Application	Operational (ERP, CRM, legacy applications)	Management Information Systems Decision Support Systems
Typical users	Staff Customers	Analyst Business users
Horizon	Weeks/Month	Years
Refresh	Immediate	Periodic
Data model	Entity-relationship	Multi-dimensional
Schema	Normalized	Star/Snowflake
Emphasis	Update/Retrieval	Retrieval

https://sqlwizard.blog/2020/03/15/sql-server-oltp-vs-olap/

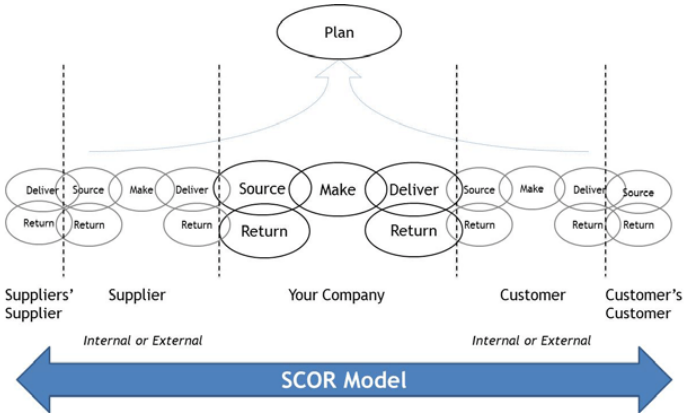
OLAP Formats

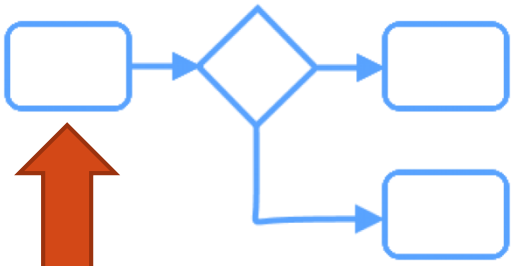


OLAP Formats: Comparison

Feature	OLTP	ROLAP	MOLAP
Typical operation	Update	Report	Analysis
Level of analytical requirements	Low	Medium	High
User interface forms	Immutable	Defined by user	Defined by user
Data volume per transaction	Small	Relatively small to large	Large
Data granularity	Detail	Detail and aggregated	Aggregated
Data storage terms	Ongoing	Historical and ongoing	Historical, ongoing, and forecasted
Structural elements	Records	Records	Arrays

Knowledge Management, Repositories, Version Control





Business Process Models

OLAP

- Obtain

Manage Knowledge

Create Knowledge

- Use
- Contribute

- Make explicit

Share Knowledge

Measure Knowledge

- Metrics
- KPIs

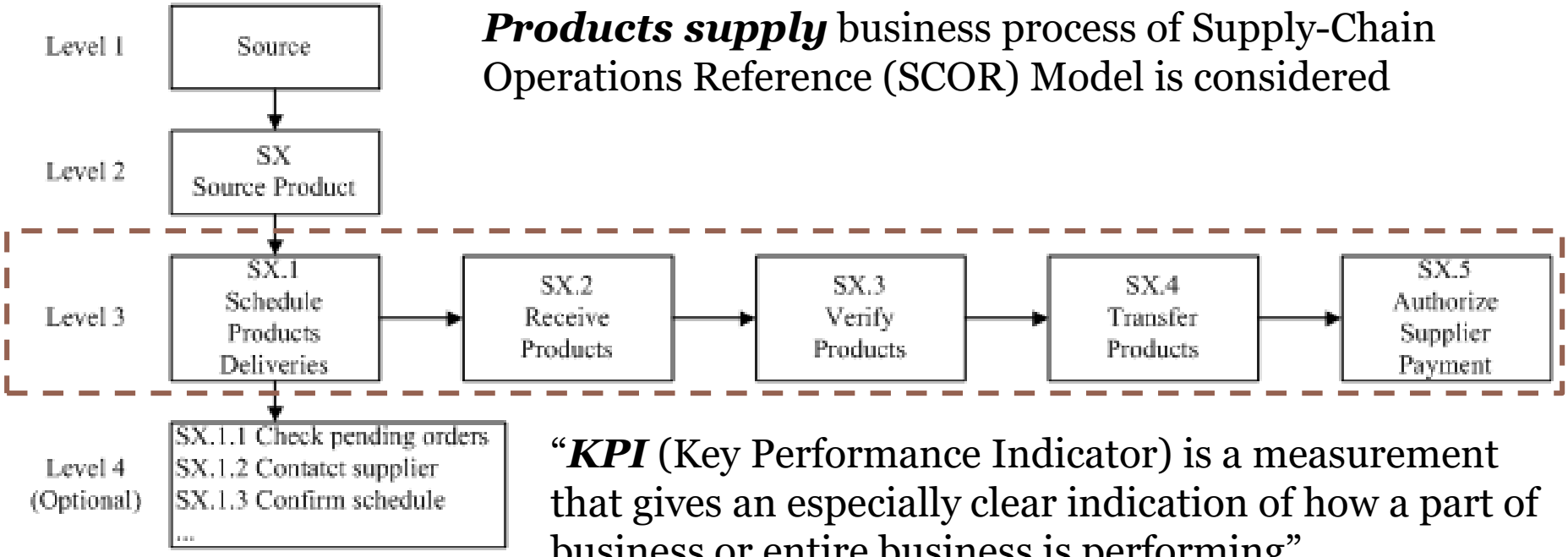
VCS

Git
- GitHub
- Bitbucket
- GitLab

SVN
- Tortoise

https://www.researchgate.net/figure/The-SCOR-R-Model-adapted-from-The-Supply-Chain-Council-20_fig2_287515269

Metrics/KPIs



Products supply business process of Supply-Chain Operations Reference (SCOR) Model is considered

"***KPI*** (Key Performance Indicator) is a measurement that gives an especially clear indication of how a part of business or entire business is performing"

Supply process metrics (KPIs):

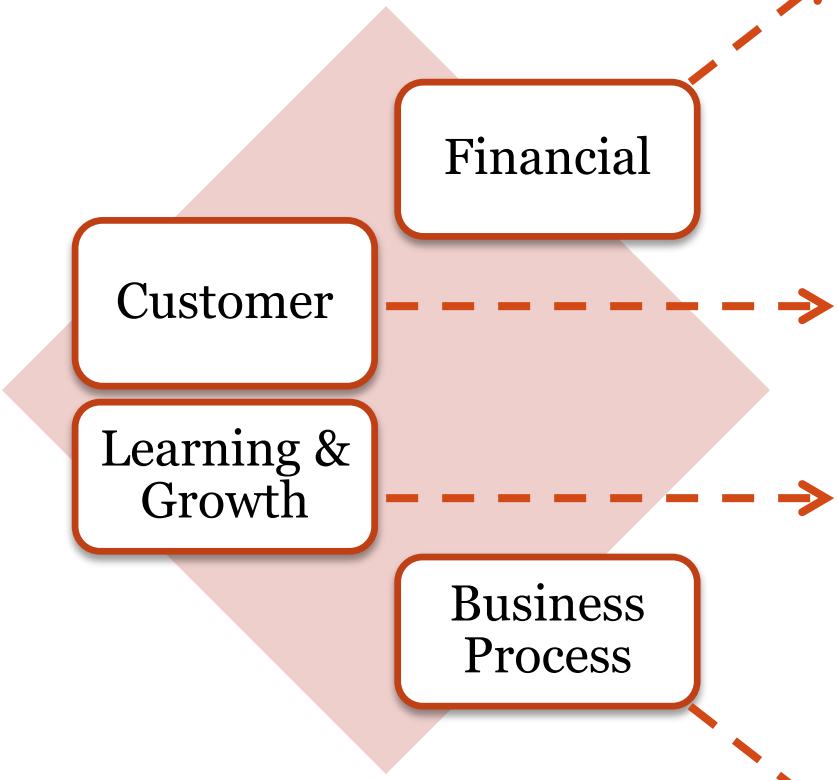
- Cost to supply, CtS
- Supply cycle time, SCT
- % orders delivered in full, $OSF_{\%}$
- % orders delivered on time, $OST_{\%}$ etc.

Specific | **M**easurable | **A**chievable

Relevant | **T**ime bounded

Scorecards

Scorecards are used to measure progress toward goals



Scorecards are similar to dashboards but used to categorize KPIs by goals

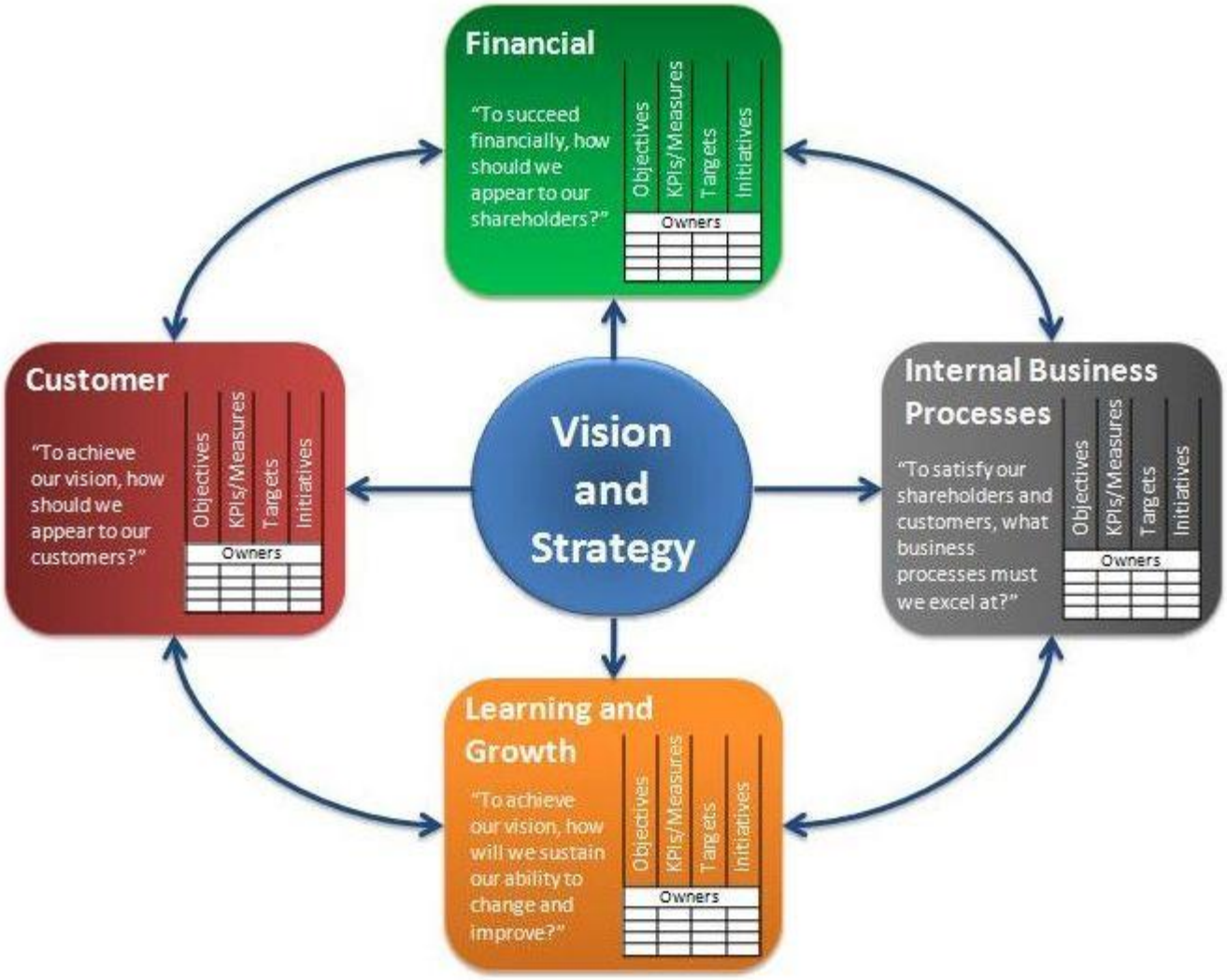
Goal	**Measure**
Reduce supply costs to X	Cost to supply (CtS)

Goal	**Measure**
Deliver Z% of fulfilled orders	% orders delivered in full ($OSF_{\%}$)
Deliver N% orders on time	% orders delivered on time ($OST_{\%}$)

Goal	**Measure**
Increase income to M	Income

Goal	**Measure**
Reduce supply cycle time to Y	Supply cycle time (SCT)

Balanced Scorecards (BSC)



Balanced ScoreCards framework is a great example of KPIs implementation

It is designed by U.S. economists:

- director of the research center Norlan Norton Institute David Norton
- professor of Harvard Business School Robert Kaplan

It was presented in 1992

Balanced Scorecards: KPIs Examples

Financial

- total assets
- total assets per employee
- income to total assets
- income per employee
- income from new products
- profit to total assets
- profit per employee, etc.

Internal Business Processes

- timely delivery
- productivity growth
- administrative expenses
- stock turnover
- production preparation time
- the cost of administrative errors
- direct contacts with clients, etc.

Vision and Strategy

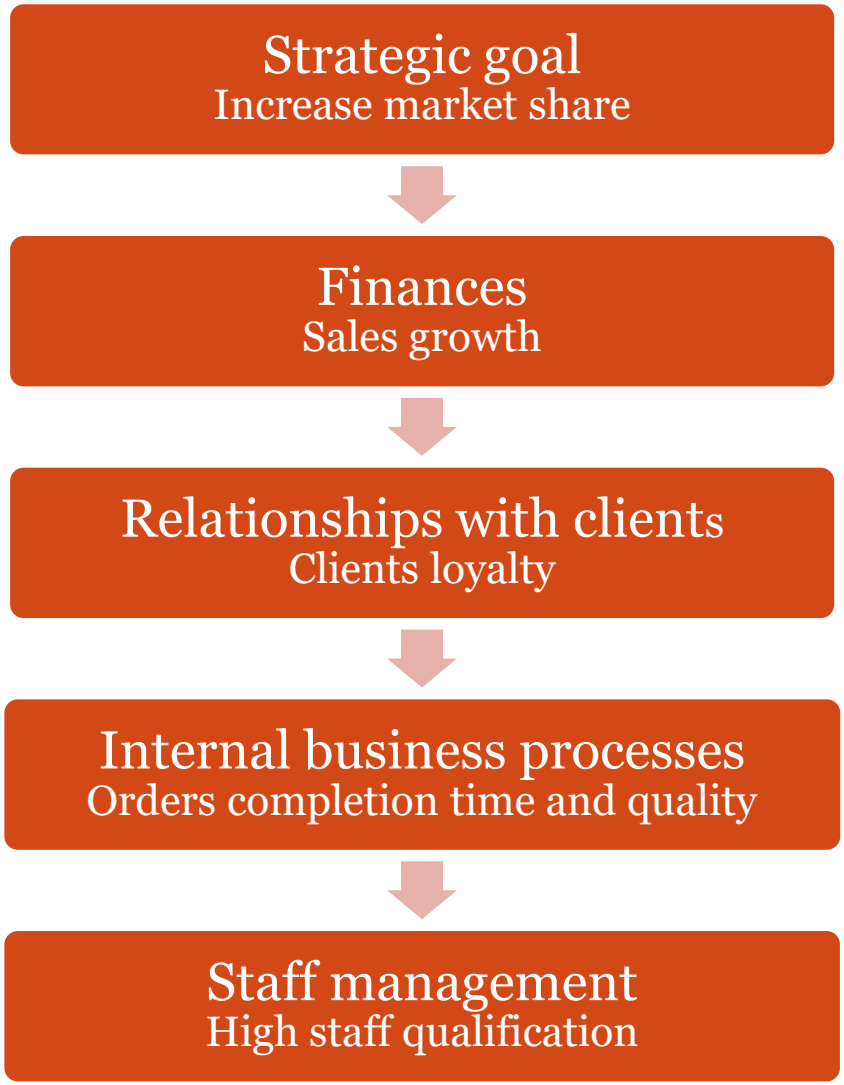
Customer

- number of clients
- market share
- average turnover per client
- average time spent on the client
- customer loyalty index
- customer satisfaction index, etc.

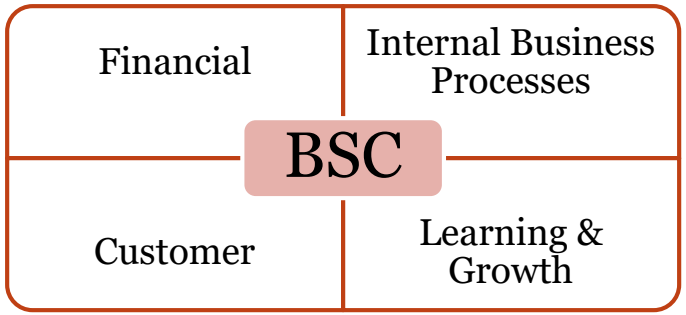
Learning and Growth

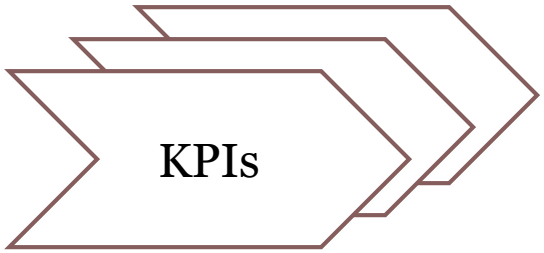
- staff turnover
- time for training
- average time of absence
- annual training costs per person
- employee satisfaction index, etc.

How KPIs Help Business?



Data Visualization and Reporting





BSC or single KPIs need to be displayed

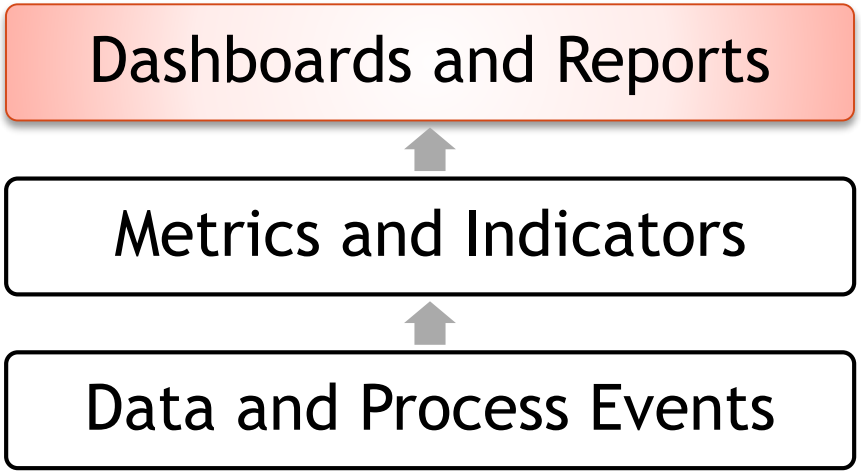
Presentation

- Advanced data visualization
- Alerts
- Dashboards
- Disconnected usage
- Geospatial data
- Mobile
- Reporting – ad hoc, analytical, production
- MS Office integration
- Visualization

http://www.dbms2.com/2008/01/14/forrester-collects-business-intelligence-buzzwords/

Dashboards

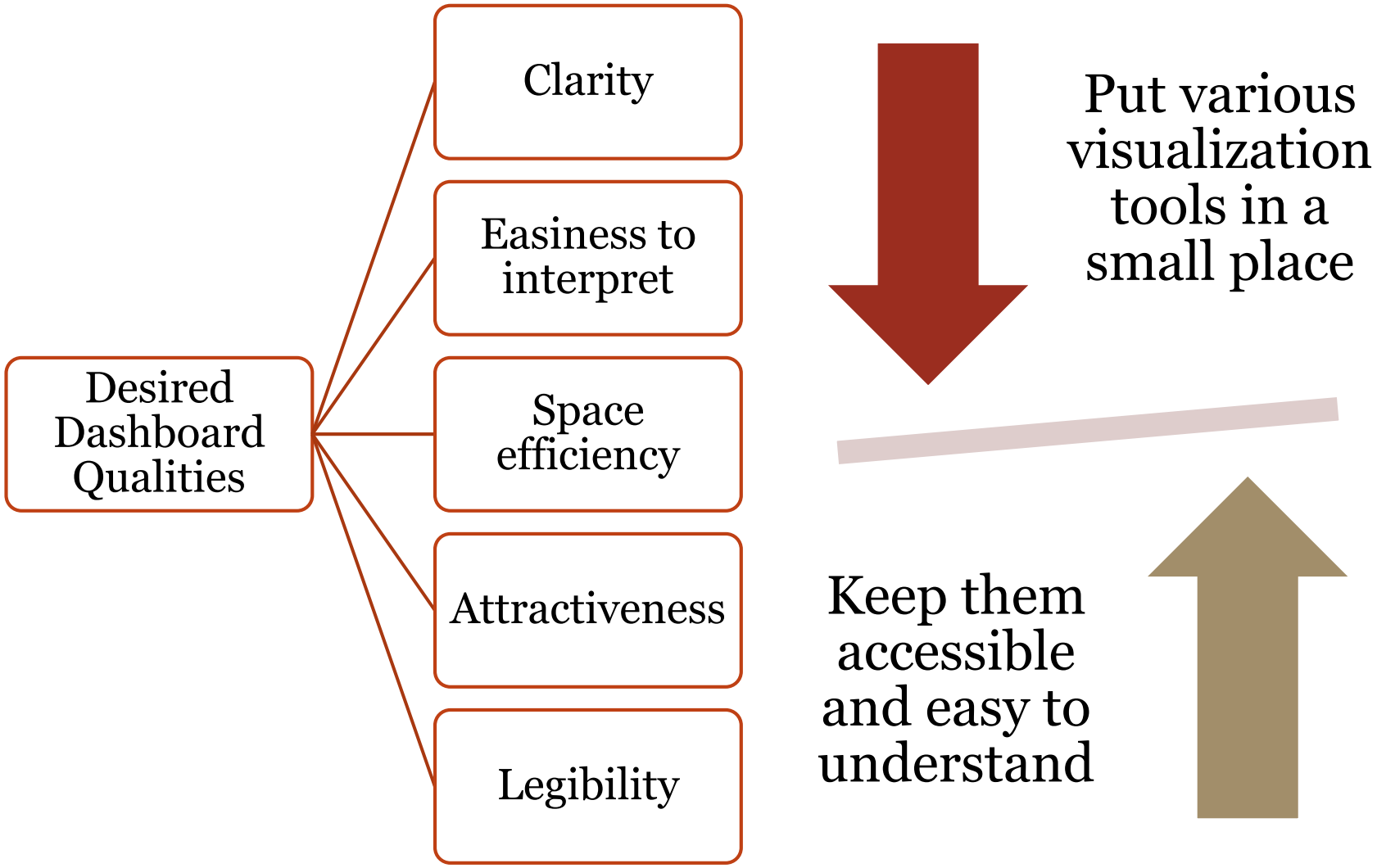
"… are just like the ones in a car or airplane"
"… are central locations where various different types of vital information show up as visual indicators"



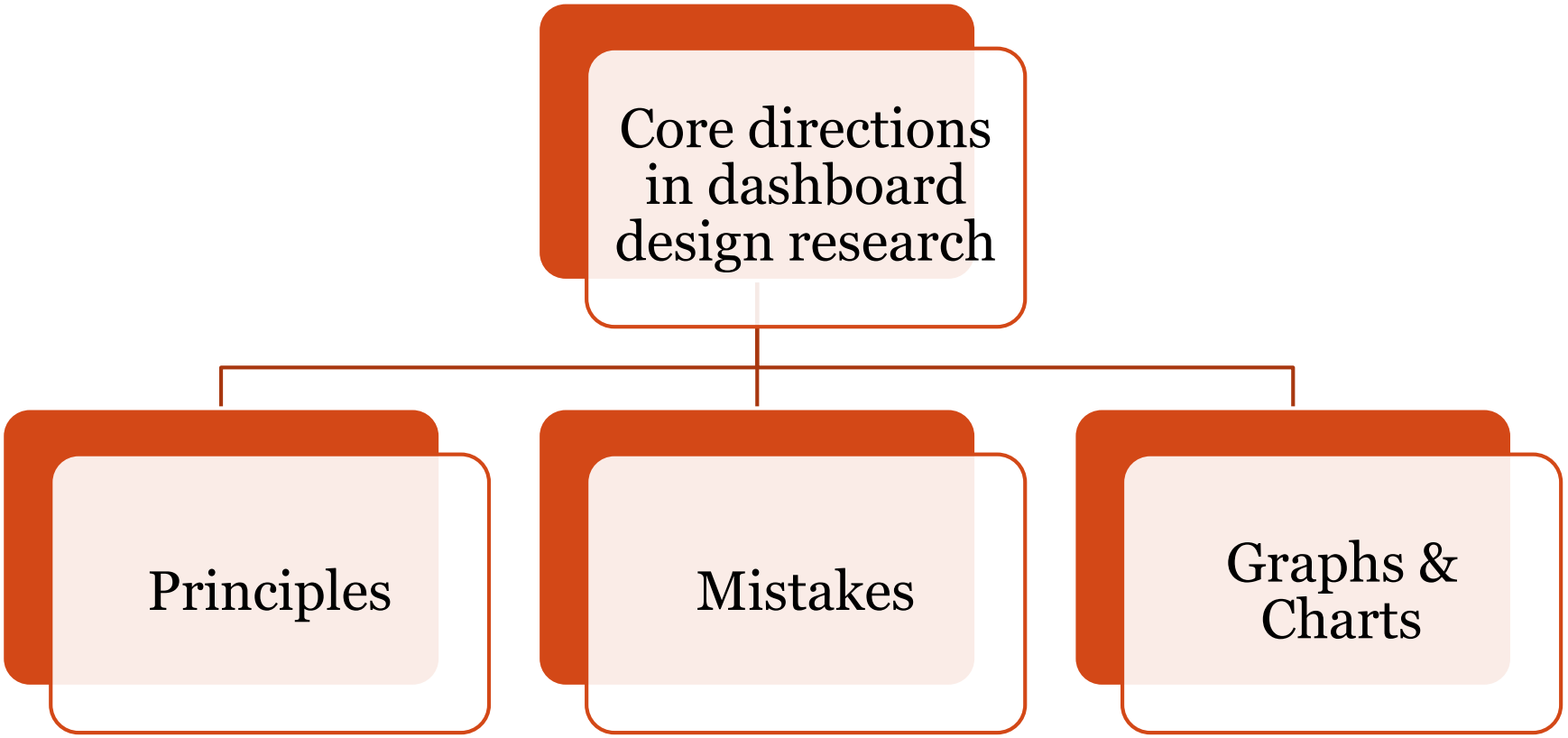
Dashboard
is a multilayer application based on business analysis and data integration infrastructure, which enables organization performance measurement, monitoring and management

https://online.visual-paradigm.com/diagrams/templates/dashboard/dashboard/

Dashboards: Design Problem



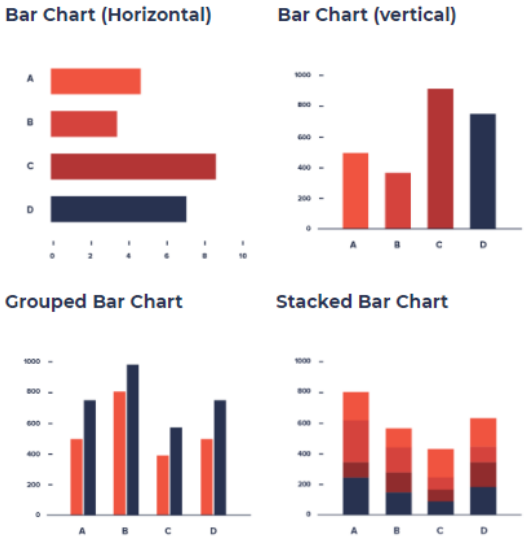
Dashboards: Research Directions



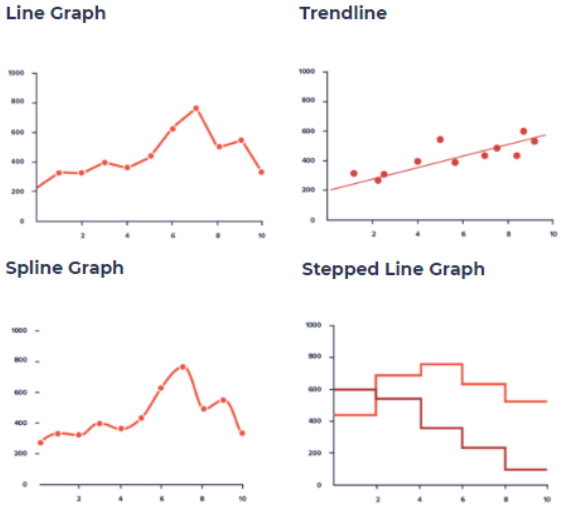
Dashboards: Graphs and Charts

Dashboard design & development assumes "placing various visualization tools in a small place, while keeping them accessible and easy to understand"

Bar Charts
comparable components of the vector value

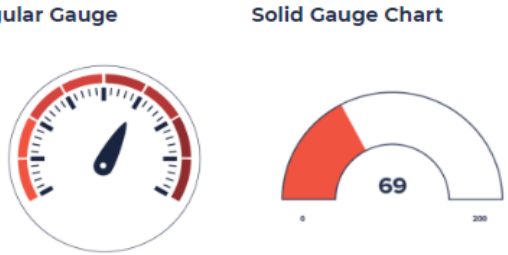


Line Charts
the direction of change of the scalar value



Pie Charts
vector values, which components represent parts of 100%

Gauges
the scalar value, which goes up or down beyond a target value

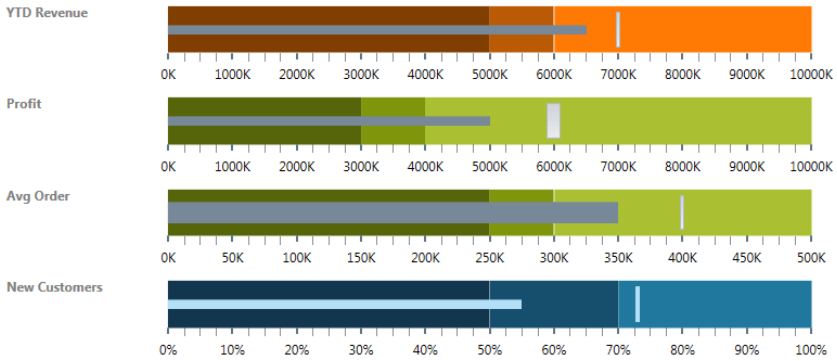


Few, S.: Information dashboard design. (2006).

Dashboards: Other Charts

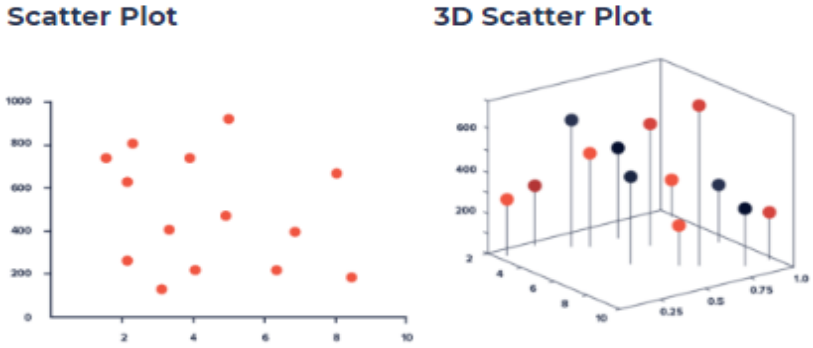
Bullet Graph

space-efficient alternative of gauges



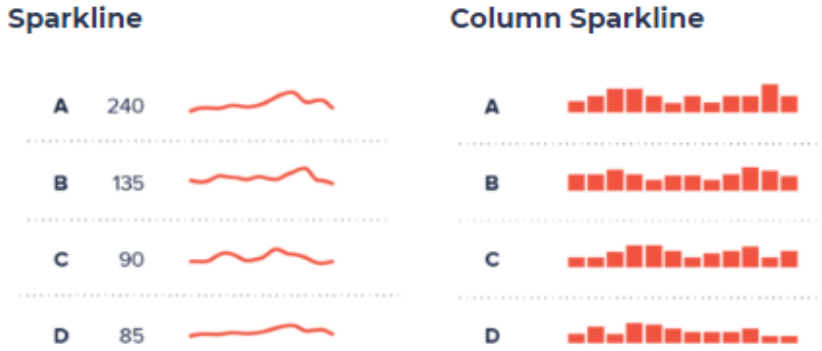
Scatter Chart

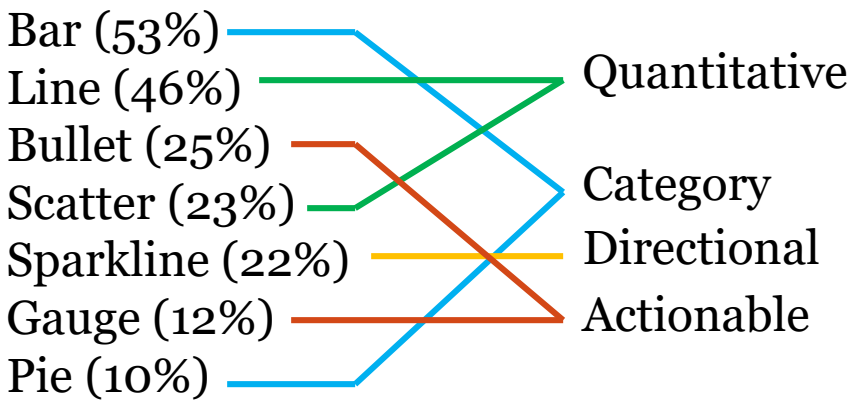
displays three-dimensional data



Sparkline

represents trends of scalar values





http://cdnlarge.tableausoftware.com/sites/default/files/whitepapers/tdwi_bpreport_q111_vra_tableau.pdf

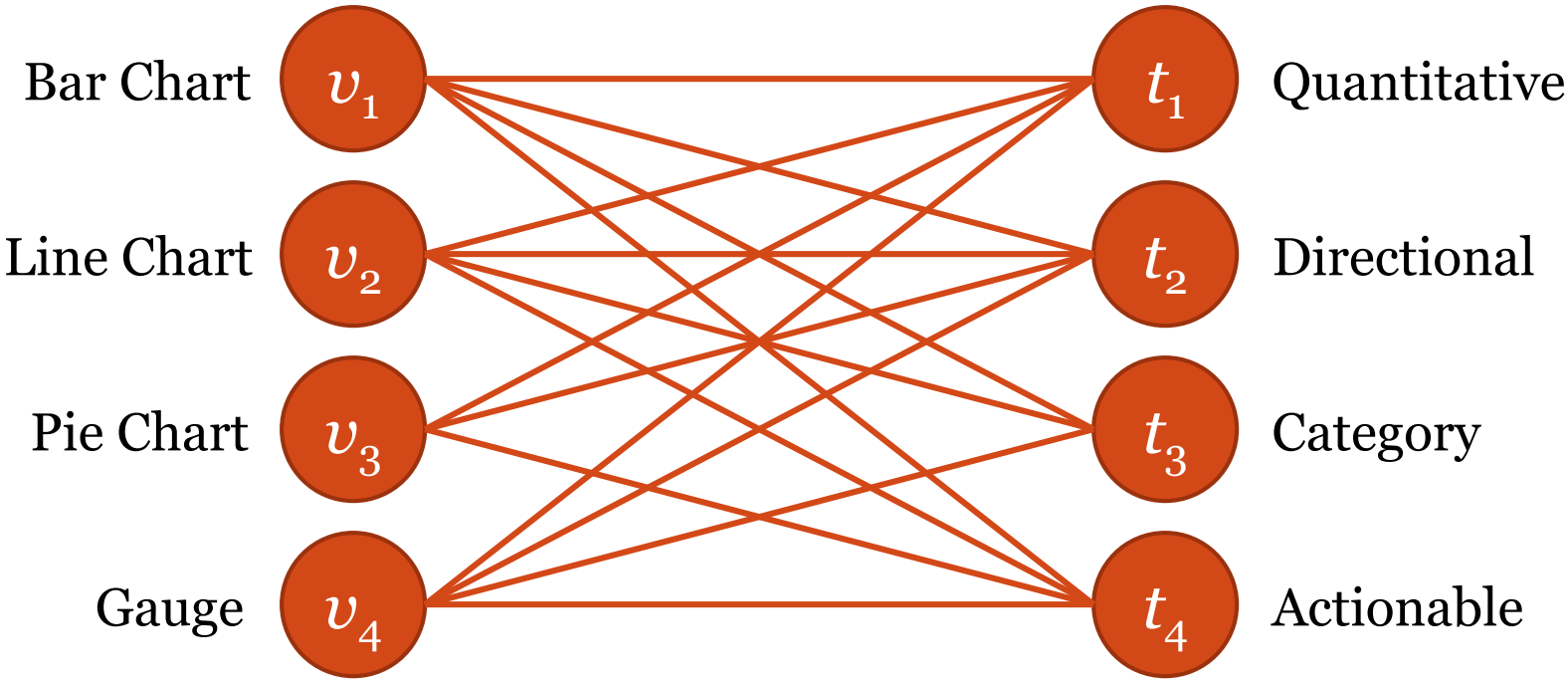
Dashboards: Mistakes

Choosing ***inappropriate*** (that does not correspond to the nature of visualized data) visualization tools (graphs and charts) is the common mistake in dashboard design

Actionable	• Above or below its target value • Outside the tolerance
Directional	• Direction that value is trending • Within the tolerance
Quantitative	• Value is more important • Exactly matches its target value
Category	• Distribution within an entire value • Compare values for each category

Dashboards: Principles

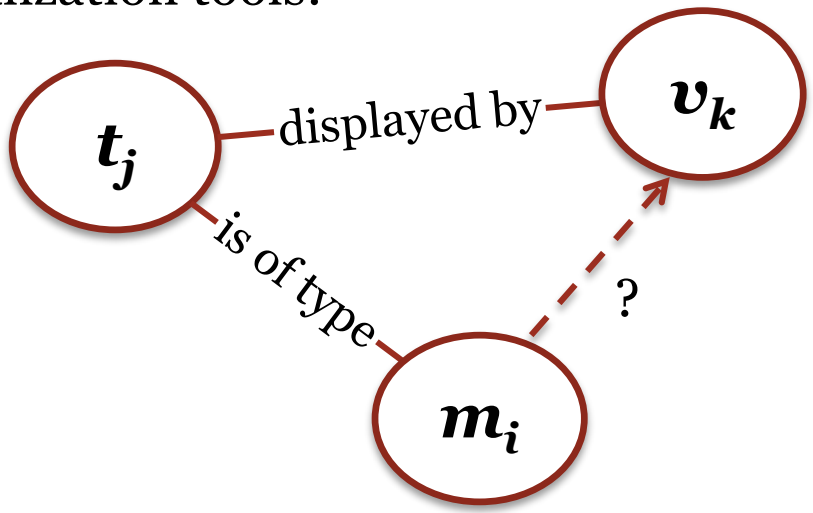
- Selected chart should fit the best a data type of a certain dataset displayed on a dashboard
- Selected chart should serve its purpose even if it is resized in order to be placed into a small place on a dashboard



Dashboards: Design Problem Formalization

1. Introduce the fuzzy relation "is of type" between measures and data types: $\mu(m_i, t_j)$
2. Introduce the fuzzy relation "is displayed by" between data types and visualization tools: $\mu(t_j, v_k)$
3. Find relation between measures and visualization tools:
$$\mu(m_i, v_k) = \max\{\min\{\mu(m_i, t_j), \mu(t_j, v_k)\}\}$$
4. Estimate informativeness of visualization tools:

Chart	Informativeness	w_i
Bar	53%	1.00
Line	46%	0.87
Bullet	25%	0.47
Scatter	23%	0.43
Sparkline	22%	0.42
Gauge	12%	0.23
Pie	10%	0.19



http://cdnlarge.tableausoftware.com/sites/default/files/whitepapers/tdwi_bpreport_q111_vra_tableau.pdf

Dashboards: Design Problem Formalization

5. Obtain recommendations on charts to be used in order to visualize data: $r_i = \arg\max_k \{w_k \cdot \mu(m_i, v_k)\}$
6. Estimate sizes of visualization tools:
   - We propose to use 12-column grid of the Bootstrap UI framework in order to provide the adaptive dashboard
   - The size of a chart assigned to i-th metric is denoted as l_i
7. Then compromise between informative and space-efficient data visualization could be achieved by solving the following optimization problem:

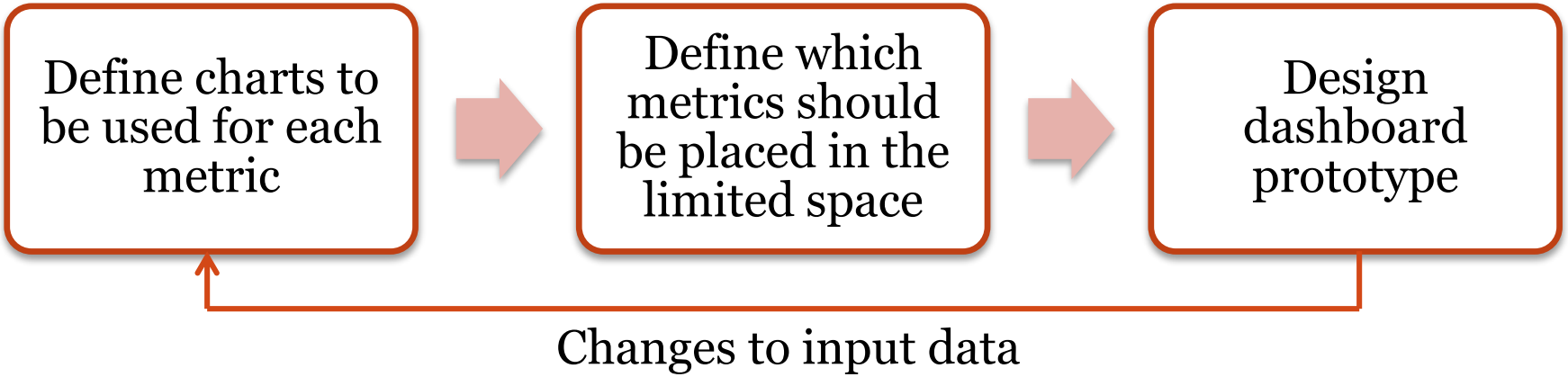
Optimization model is inspired by 0-1 Knapsack Problem

$$\sum_i w_i \cdot x_i \rightarrow \max$$
$$l_i \cdot x_i \leq 12 \cdot \alpha$$
$$x_i \in \{0, 1\}$$

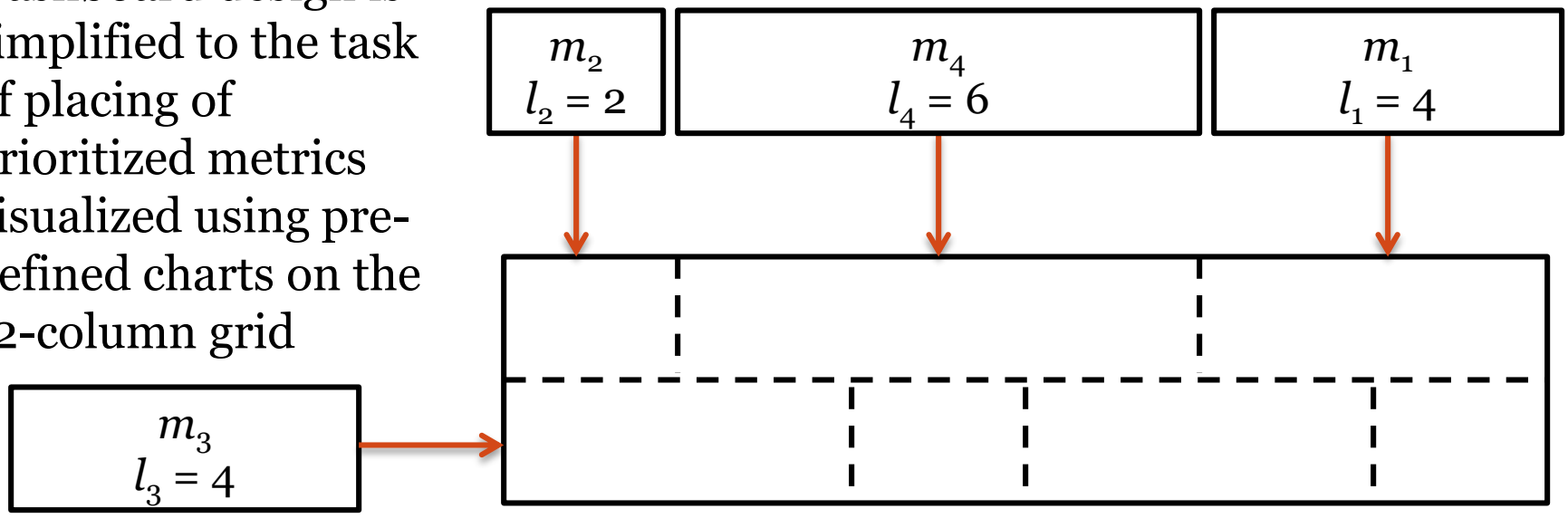
This problem could be solved using the Greedy algorithm: sort metrics by w_i from large to small and put them to the dashboard until the all space is taken

Here α is the number of rows to be included into the dashboard, while x_i indicates whether i-th chart should be included into the dashboard or not

Dashboards: Design Problem Formalization



Dashboard design is simplified to the task of placing of prioritized metrics visualized using pre-defined charts on the 12-column grid



Dashboards: Design Problem Example

Some KPIs of ***products supply*** process of the SCOR model are considered:

- Cost to supply, CtS
- Supply cycle time, SCT
- % orders delivered in full, $OSF_{\%}$
- % orders delivered on time, $OST_{\%}$

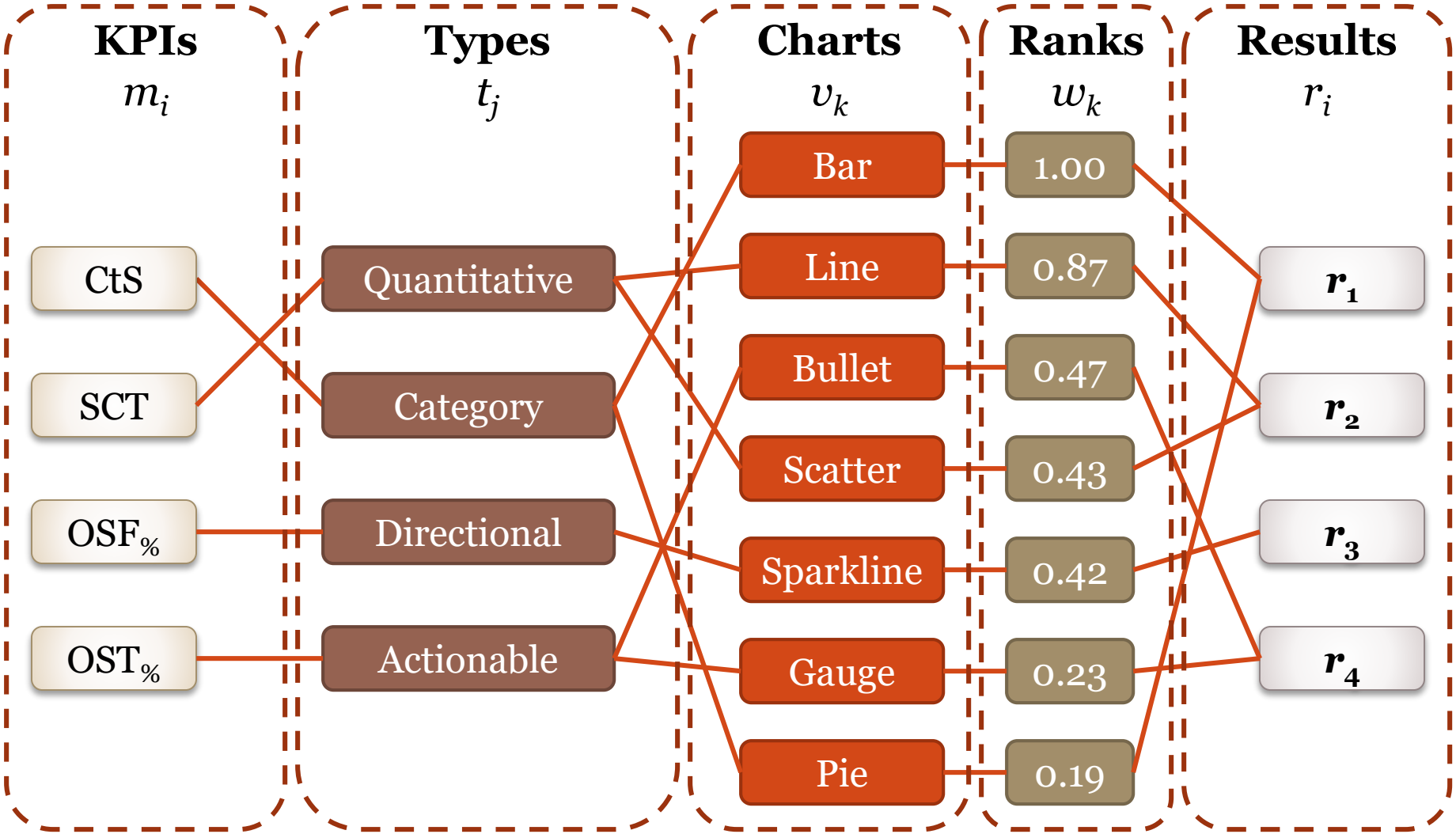
"is of type" binary relation between KPIs and data types

KPI x Type	Quantitative	Category	Directional	Actionable
CtS	0	1	0	0
SCT	1	0	0	0
OSF%	0	0	1	0
OST%	0	0	0	1

"is displayed by" binary relation between data types and charts:

Type x Chart	Bar	Line	Bullet	Scatter	Sparkline	Gauge	Pie
Quantitative	0	1	0	1	0	0	0
Category	1	0	0	0	0	0	1
Directional	0	0	0	0	1	0	0
Actionable	0	0	1	0	0	1	0

Dashboards: Design Problem Example



Dashboards: Design Problem Example

Find relations between KPIs and visualization tools:

KPI x Chart	Bar	Line	Bullet	Scatter	Sparkline	Gauge	Pie
CtS	1	0	0	0	0	0	1
SCT	0	1	0	1	0	0	0
OSF%	0	0	0	0	1	0	0
OST%	0	0	1	0	0	1	0

Estimate informativeness of charts:

Chart	Bar	Line	Bullet	Scatter	Sparkline	Gauge	Pie
Weight	1	0.87	0.47	0.43	0.42	0.23	0.19

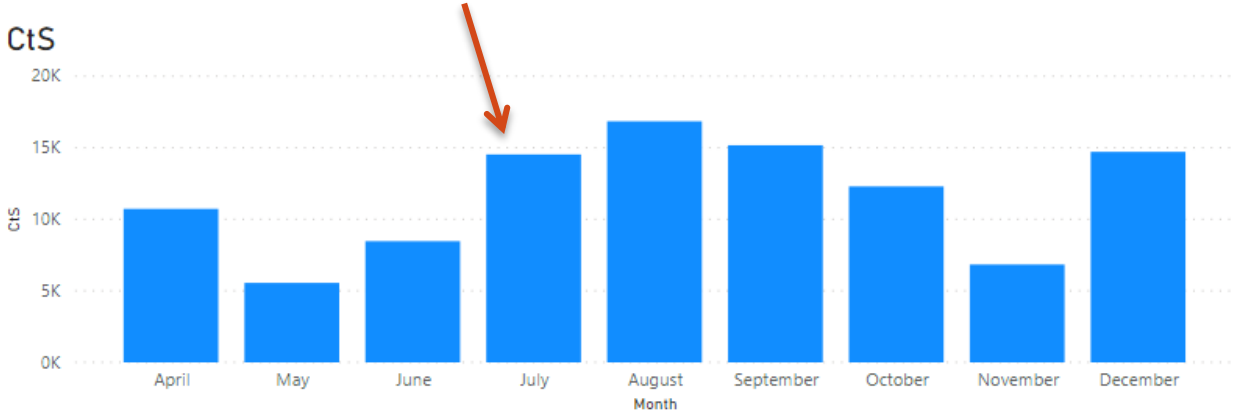
Obtain recommendations:

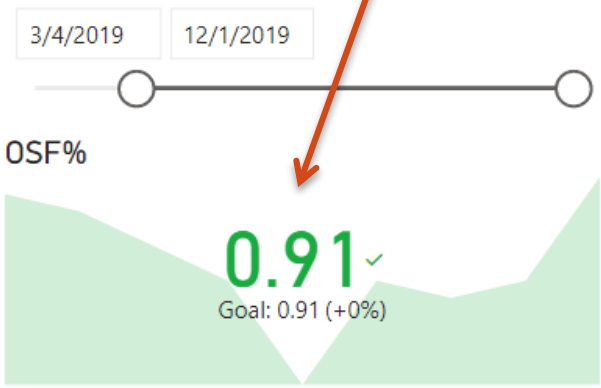
Weighted	Bar	Line	Bullet	Scatter	Sparkline	Gauge	Pie	*max*	*arg max*
CtS	1	0	0	0	0	0	0.19	1	**Bar**
SCT	0	0.87	0	0.43	0	0	0	0.87	**Line**
OSF%	0	0	0	0	0.42	0	0	0.42	**Sparkline**
OST%	0	0	0.47	0	0	0.23	0	0.47	**Bullet**

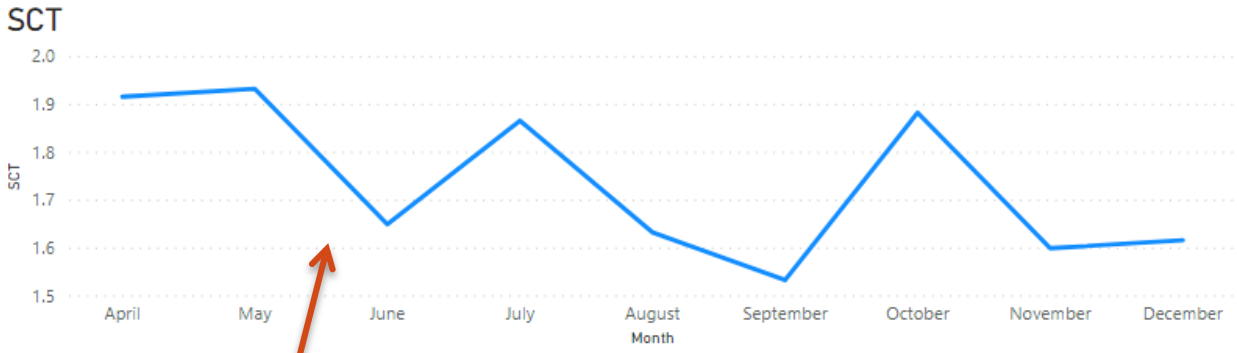
Dashboards: Design Problem Example

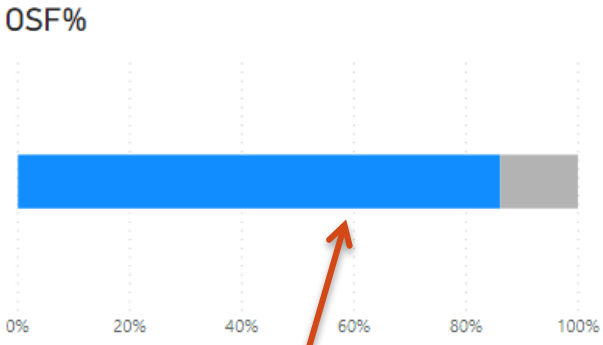
Cost to supply, CtS

% orders delivered in full, $OSF_{\%}$









Dashboard prototype designed in Microsoft Power BI

Supply cycle time, SCT

% orders delivered on time, $OST_{\%}$

Dashboards: Proposed Method Summary

Define the nature of KPIs data

Rank charts or use default weights

Use suggested charts to build the dashboard

The nature of KPIs has changed?

yes

no

Deploy designed dashboard

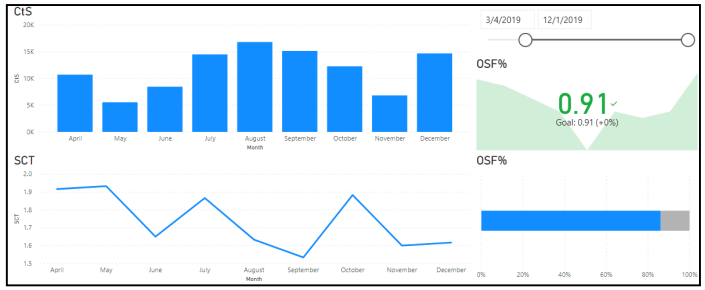
<is of type> relation

KPI x Type	Quantitative	Category	Directional	Actionable
CtS	0	1	0	0
SCT	1	0	0	0
OSF%	0	0	1	0
OST%	0	0	0	1

ranked charts

Chart	Bar	Line	Bullet	Scatter	Sparkline	Gauge	Pie
Weight	1	0.87	0.47	0.43	0.42	0.23	0.19

suggested dashboard content



Related Authors Work on Data Analysis and Visualization

1. Andrii Kopp, Dmytro Orlovskyi, Danylo Kuka. An approach to forming dashboards for business processes state analysis. Bulletin of NTU "KhPI". Series: System analysis, control and information technology, no. 51, pp. 44–52. NTU “KhPI”, Kharkiv (2017). http://samit.khpi.edu.ua/article/view/120762
2. Andrii Kopp, Dmytro Orlovskyi. An Approach to Forming Dashboards for Business Process Indicators Analysis using Fuzzy and Semantic Technologies. Proceedings of the PhD Symposium at 14th International Conference on ICT in Education, Research, and Industrial Applications ICTERI 2018. Vol. 2122. pp. 1–7. CEUR Workshop Proceedings, Kyiv (2018). http://ceur-ws.org/Vol-2122/paper_11.pdf
3. Dmytro Orlovskyi, Andrii Kopp, Vitalii Kondratiev. Using dashboards for the business processes status analysis. Information technologies and automation 2019, pp. 31-33. ONAFT, Odesa (2019). http://repository.kpi.kharkov.ua/handle/KhPI-Press/47103
4. Dmytro Orlovskyi, Andrii Kopp, Vitalii Kondratiev. Development of a model and a software solution to support the analytical dashboards design problem. Bulletin of National Technical University "KhPI". Series: System Analysis, Control and Information Technologies, no. 1 (3), pp. 58-67. NTU “KhPI”, Kharkiv (2020). http://samit.khpi.edu.ua/article/view/2079-0023.2020.01.11

JOURNAL OF THE NATIONAL TECHNICAL UNIVERSITY "KHPI" SERIES: SYSTEM ANALYSIS, CONTROL AND INFORMATION TECHNOLOGY (SACIT)

Journal of the NTU "KhPI". Series: SACIT is an open-access peer-reviewed edition that publishes scientific works in the fields of system analysis, computer science, information systems and technology, software engineering, automation and computer-integrated technology, applied mathematics.

Indexing: DOAJ, Index Copernicus (ICV 2018: 83.21), OCLC WorldCat, Google Scholar, Crossref, and others.

International Editorial Board: Ukraine, France, Australia, Georgia, Germany

e-ISSN: 2410-2857
p-ISSN: 2079-0023
DOI: 10.20998 (assigned to each paper)

Email original research paper:
andrii.kopp@khpi.edu.ua
Visit website:
http://samit.khpi.edu.ua/

- Peer-reviewed
- Issued twice a year
- Accepts papers in English
- Papers published online as well as print versions (hard copy)
- Publication fee is about 20 USD

This journal is practicing and supporting a policy of open access according to the **Budapest Open Access Initiative (BOAI)**. The editorial board is adheres to the international standards of publication ethics and recommendations of **Committee on Publication Ethics (COPE)**.

Thank you for your attention!