

# Big Data infrastructure and skills from a consulting point of view

„The Bottom up View“



EDISON – **E**ducation for **D**ata Intensive  
**S**cience to **O**pen **N**ew science frontiers

Grant 675419 (INFRASUPP-4-2015: CSA)

**Kevin Berwind, M.Sc.**  
University of Hagen

EGI Community Forum  
November 12<sup>th</sup> 2015, Bari, Italy



# Background and CV

## Education:

- **University of Ludwigshafen am Rhein**
  - Business Informatics (B.Sc.)
  - Business Informatics with the specialization “Information Management & Consulting” (M.Sc.)

## Working Experience:

- **Accenture Technology Solutions GmbH, Kronberg im Taunus**
  - Consulting: Relationship Management and CRM Analytics
- **BTEXX GmbH, Mainz**
  - Consulting: Big Data, Predictive Analytics and Predictive Maintenance



# Agenda

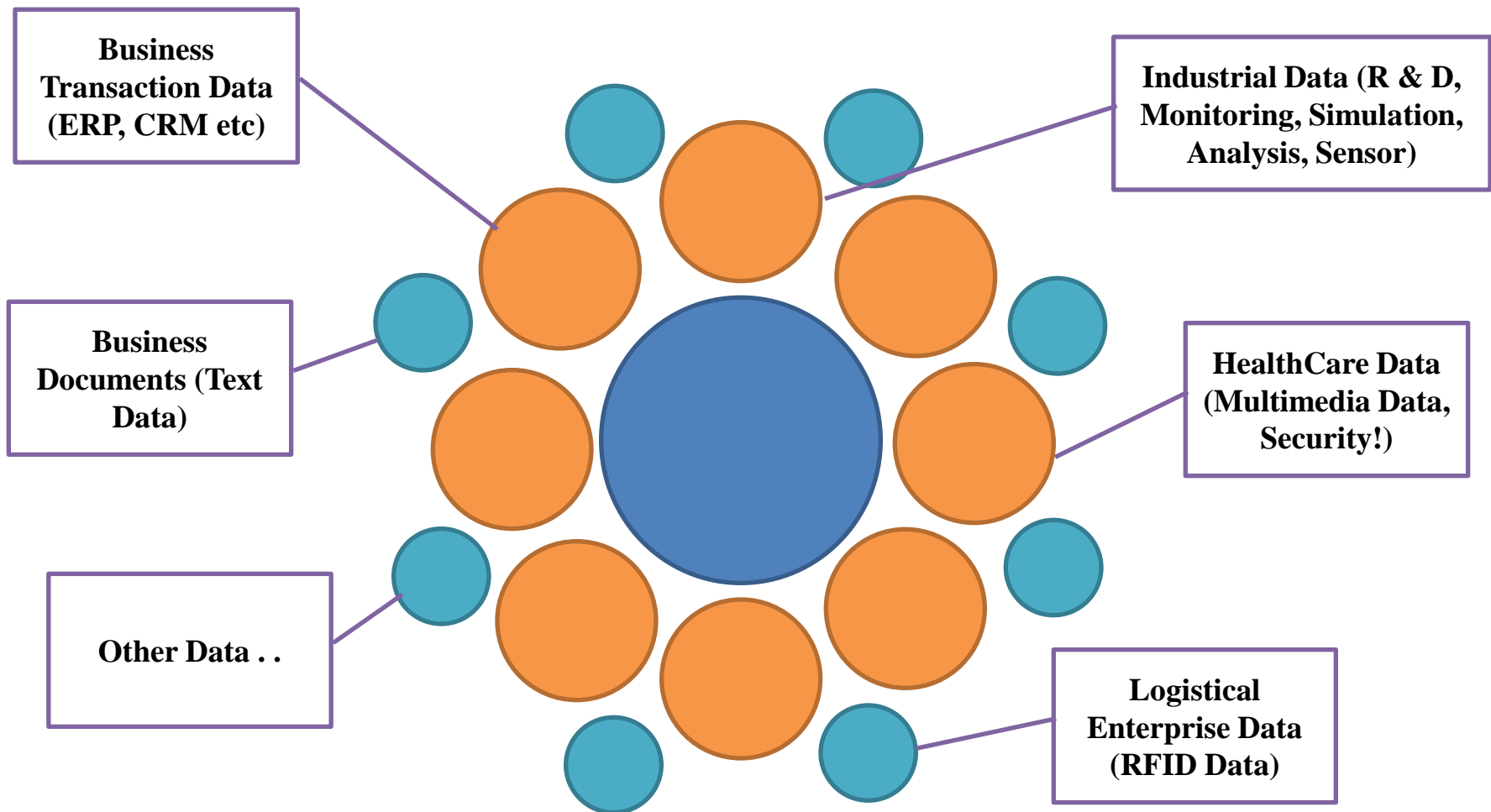
---

- Big Data in Enterprises
- Big Data Reference Infrastructures
- Needed Big Data Skills For Consulting Projects



# Big Data in Enterprises I

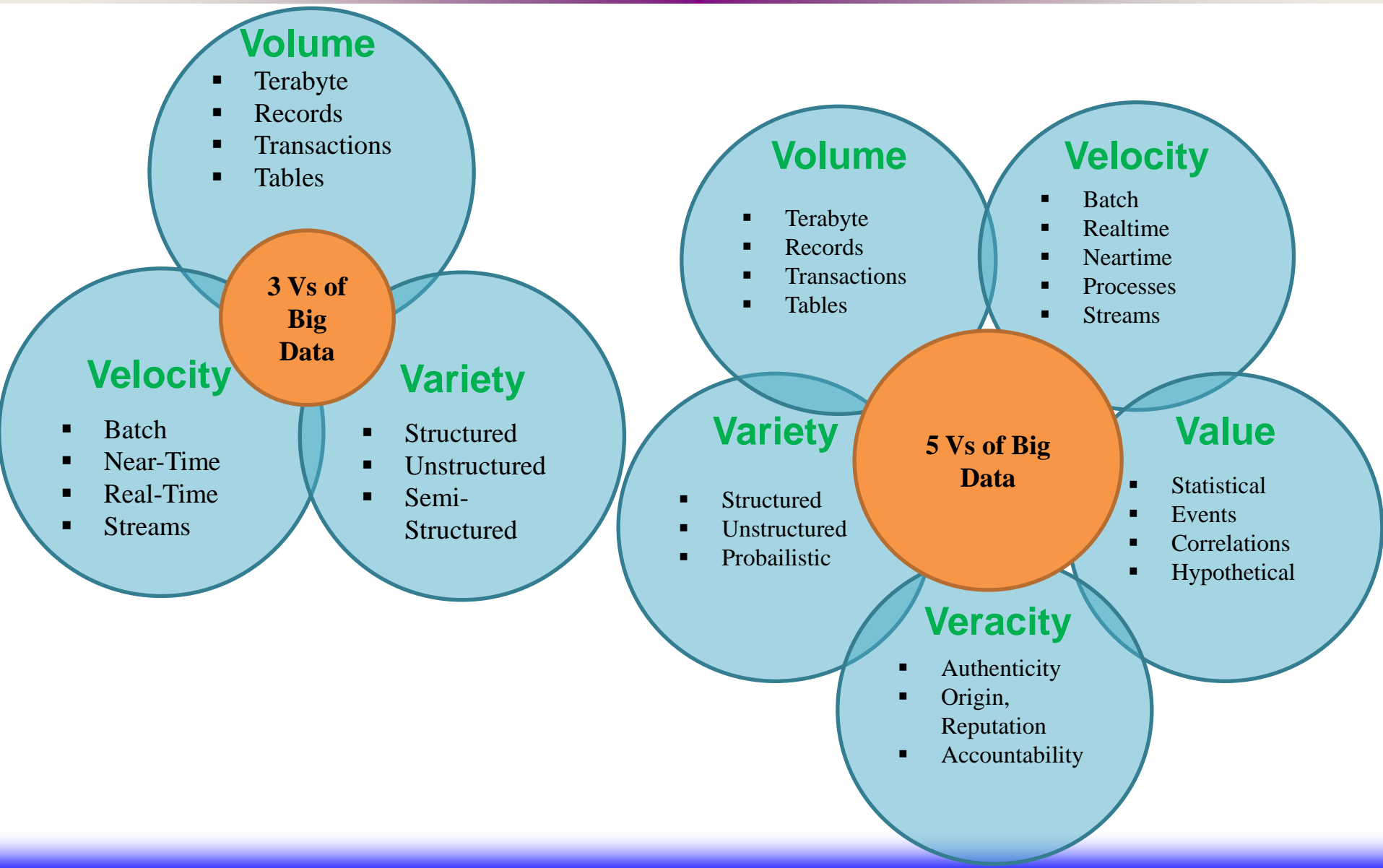
## Industrial Application Domains





# Big Data in Enterprises II

## Industrial Big Data Characteristics





# Big Data in Enterprises II

## Big Data as immeterial Assetmanagement



### Data Value:

Detect risks, information about customers, competitors, supplier, processes, new opportunities

### Data Volume:

„Increasing of data volume up to 65% each year“

### Data Types:

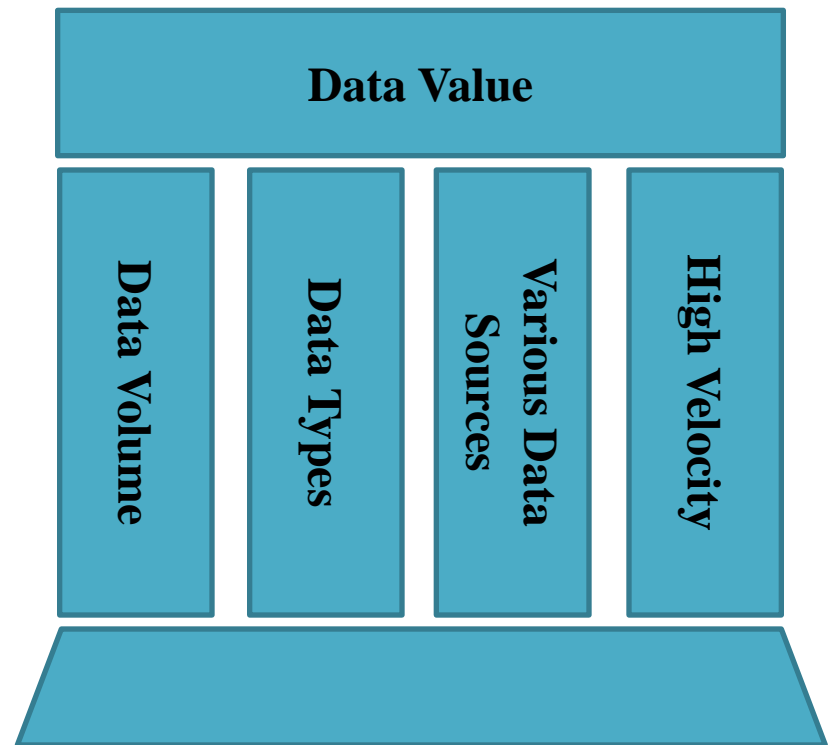
Structured, unstructured, semi-structured

### Various Data Sources:

Data bases, In-memory data bases, geographical distributed

### High Velocity:

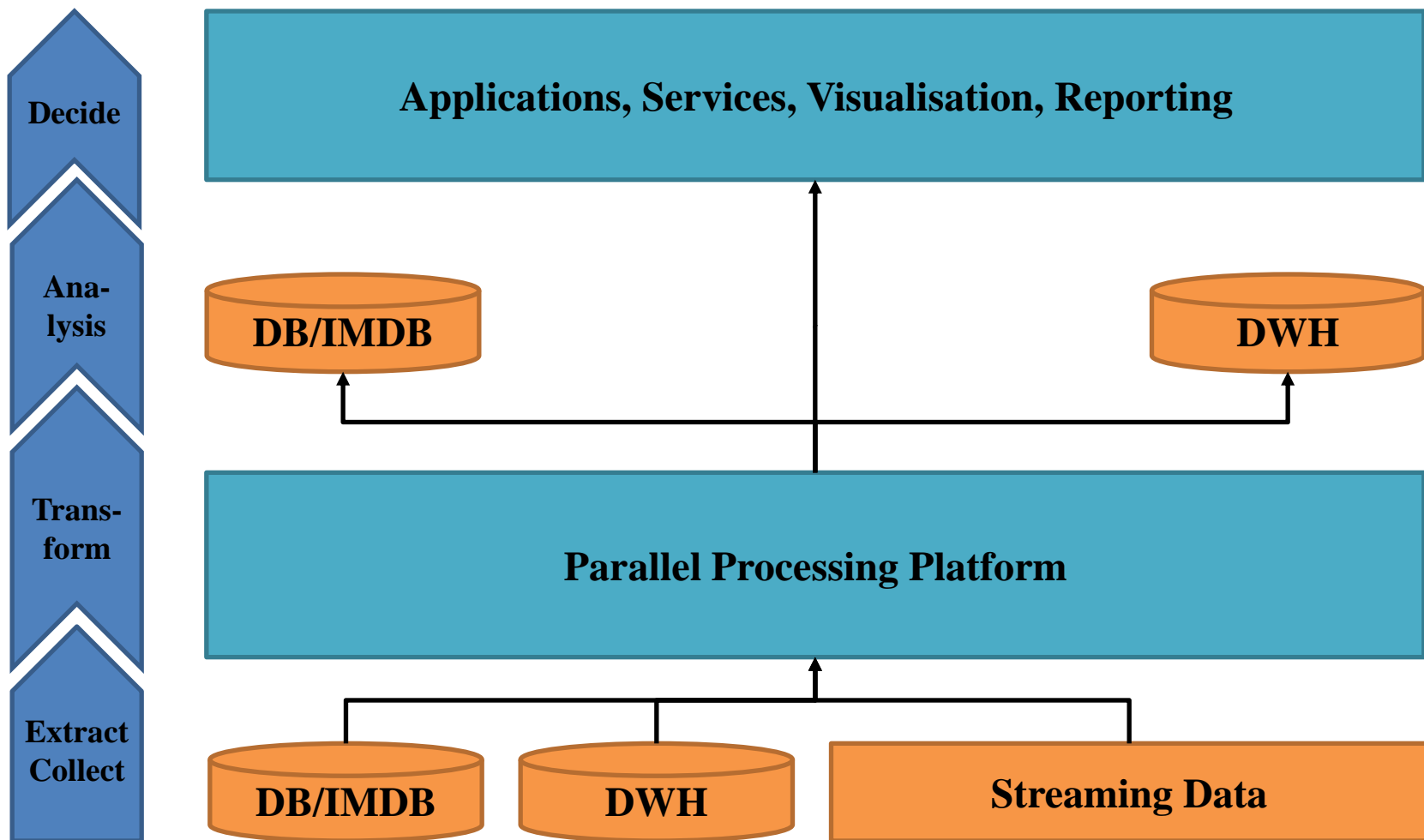
Near-time, real-time, batch





# Big Data Reference Infrastructures II

## Architecture and Processes





# Big Data Reference Infrastructures II

## Architecture and Processes



### Data sources:

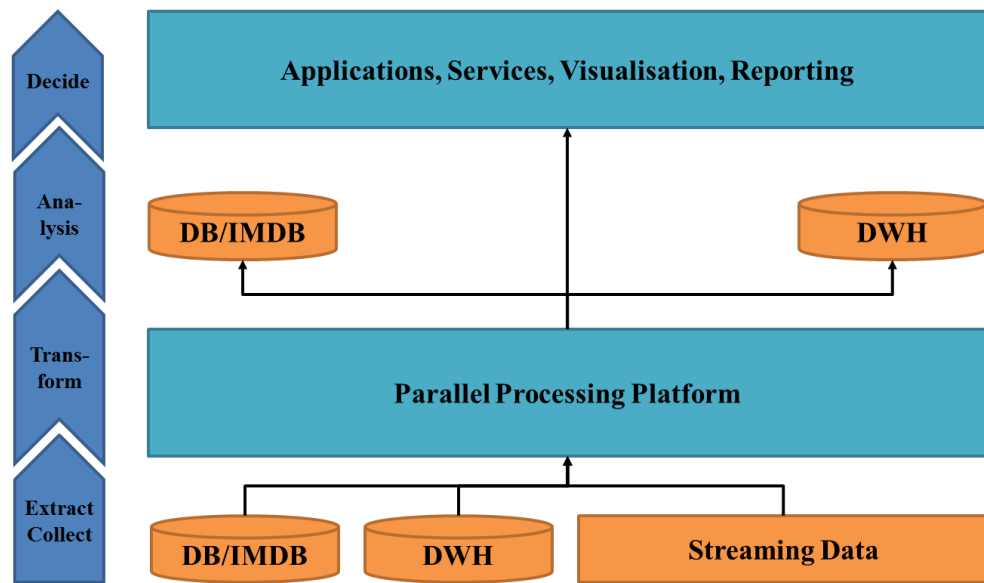
Transactional systems, data warehouses, protocol data, streaming data, social media

### Parallel Processing Platform:

Hadoop Distributed File System (HDFS) and other software products like SAP HANA

### Analysis and Visualization:

Data analysis within the PPP, Data forwarding to other systems, Data analysis and reporting via applications or services







# Big Data Reference Infrastructures III

## Big Data Distribution Models



**SaaS – analysis- and visualization tools**

**Paas – middleware (example Hadoop system)**

**IaaS – server, storage, network componets**



# Big Data reference infrastructures IV



## Enterprise Big Data infrastructure

- Well-matched hard- and software.
- Special staffing to operate, develop and manage the infrastructure.
- Cost reduction concerning missing staff and infrastructure, investments in new technology and infrastructure.
- Useful if you are using high volume or sensible data or data streams for example from machines, sensors, processes.

## External Big Data Cloud infrastructure

- Cloud provider is taking care of well-matched hard- and software.
- Staffing is done by the cloud provider.
- Various price models like pay-per-use, staffing and investments are done by cloud provider.
- You may use storage from the cloud provider for non-sensible data or data from the internet to keep your data and information within the company and to prevent high network workload.



# Needed Big Data Skills For Projects



## Successful Big Data Projects

### Technical Data Competences

- Statistics
- Operations research
- Data mining
- Machine learning
- Visualization
- Computer science

### Technical Software Competences

- Software and product overview
- Software architecture
- Service architecture

### Management Competences

- Coordination and planning
- Reporting and budgeting
- Leadership and staffing
- Quality management



Thank You for your kind attention

---

**EGI Community Forum 2015**

November 12<sup>th</sup>, 2015 - Bari, Italy

**Presentation: „Big Data infrastructure and skills from a consulting point of view“**

**Kevin Berwind, M.Sc.**



# Backup

---



# Big Data Methods And Processes

---

- **SEMMA**
- **CRISP\_DM**
- **OWN**
- **Organisations**
- **KDD Process**