IN4020 (only) – Database Systems (2026)
Lecture 28: Database Research Trends

---

4 May

Egor V. Kostylev
IFI, University of Oslo
egork@ifi.uio.no



Lecture plan

Today we look at:

1. Trends in database research
2. Related courses at IFI
3. Related research initiatives at UiO

The Beckman report (2016)

Available online[1]

Report from meeting(s) of top database researchers

From a series of meetings every 5 years (approximately) since 80's

Identified Big Data as the defining challenge

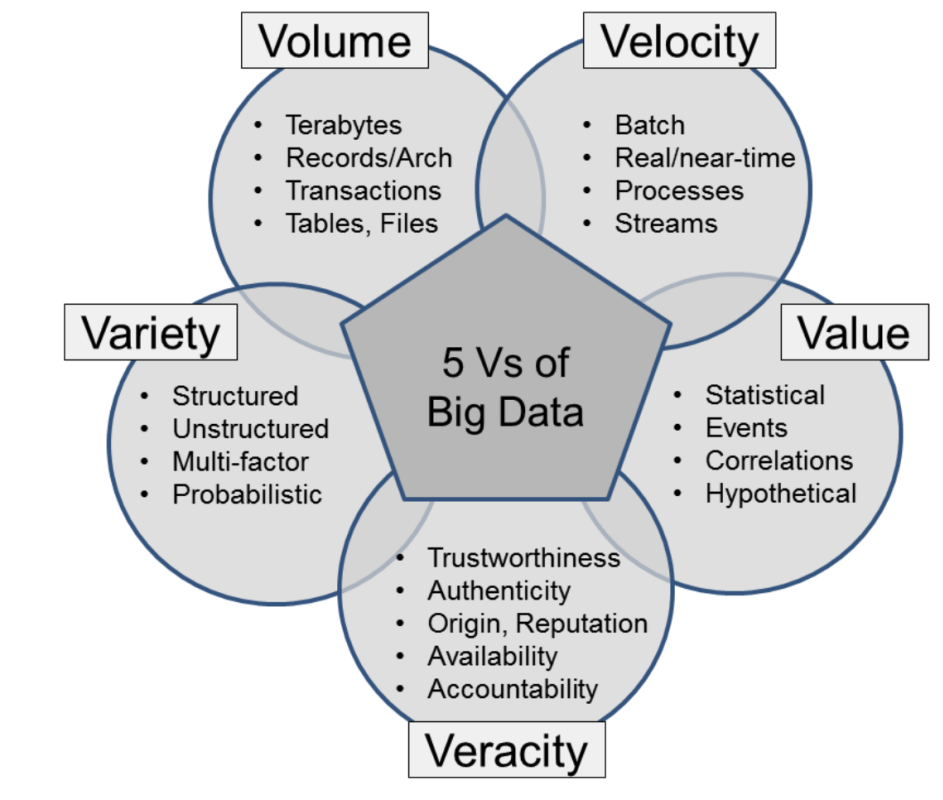
Identified the following trends:

- developing scalable data infrastructures
- coping with increased diversity in both data and data management
- addressing the end-to-end data-to-knowledge pipeline
- responding to the adoption of cloud-based computing
- accommodating the many and changing roles of individuals in the data life cycle

[1] https://cacm.acm.org/research/the-beckman-report-on-database-research/

Main challenge: Big Data

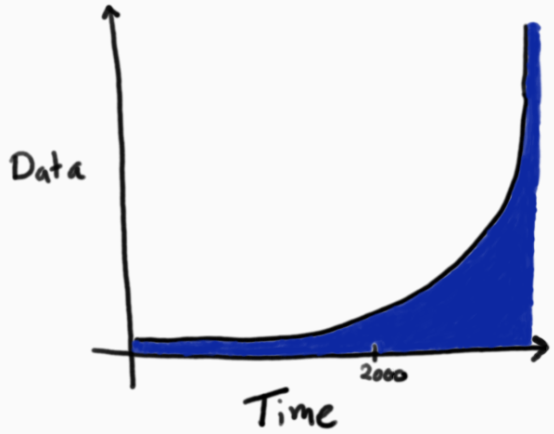
- Main characteristics:
  Volume, Velocity and Variety
  (+ Veracity, Value, etc.)
- Need to adapt current solutions and develop new solutions and techniques
- Many systems move to cloud



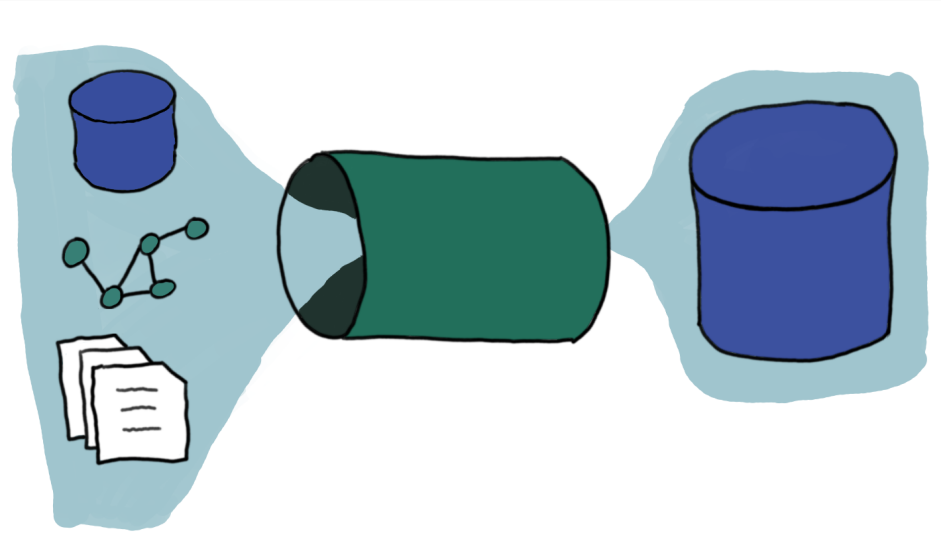
Trend: Scalable big/fast data infrastructures

Need to develop:

- more cost-aware query processing and optimisation (multiple cores, large clusters)
- more specialised hardware (database machine)
- more cost-efficient storage (new types of memory)
- high-speed data-stream processing (e.g., PipelineDB extension of PostgreSQL)
- late-bound schemas (for un-/rarely-used data)
- new models for consistency and concurrency (NoSQL)
- new types of benchmarks (cost of ownership)



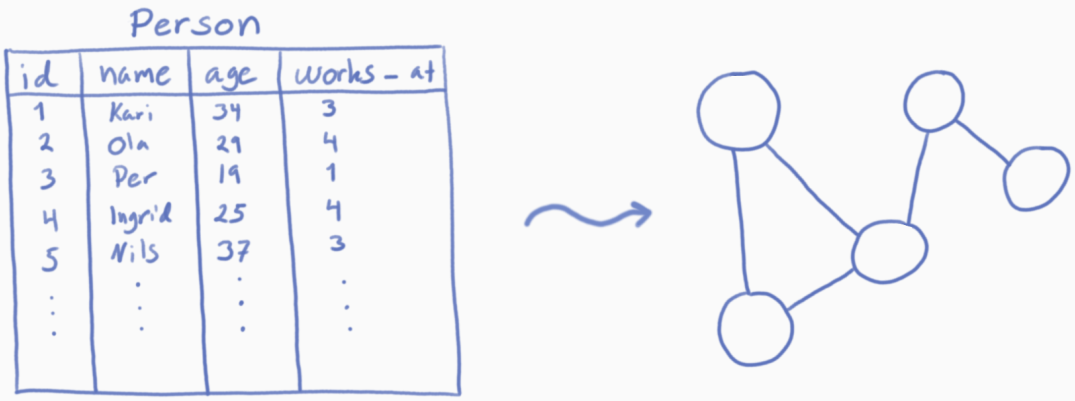
Trend: Coping with diversity in data management



Need to develop:

- better cross-platform integration (performance, abstraction)
- programming models suitable for operations on very large datasets (R, Python, etc.)
- better data processing workflows (lazy "Lego bricks")

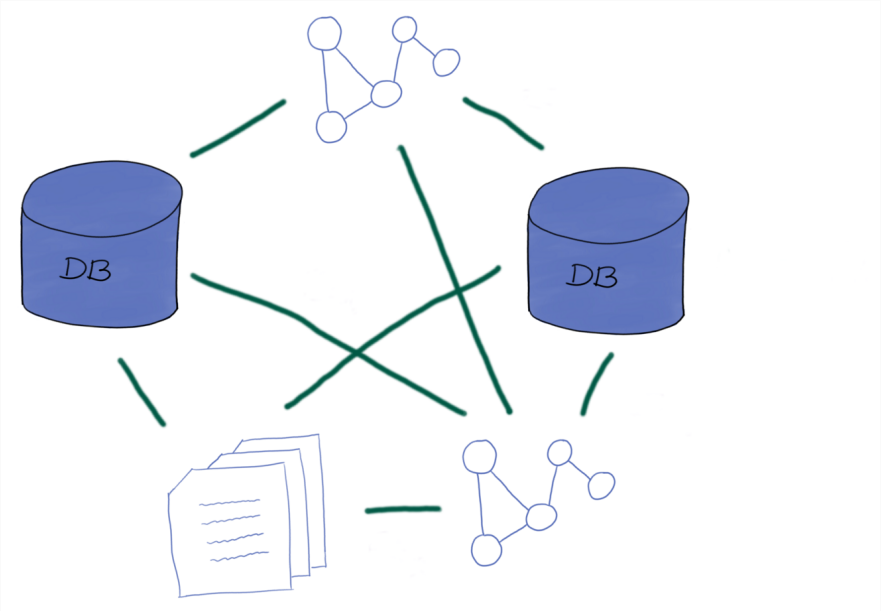
Trend: End-to-end processing of data



Need to develop:

- data-to-knowledge pipeline adapted to scale and diversity
- more diverse, (knowledge-)customisable and open source tools
- more tools for understanding data (visualisation, explanation, etc.)
- knowledge bases capturing more of domain knowledge

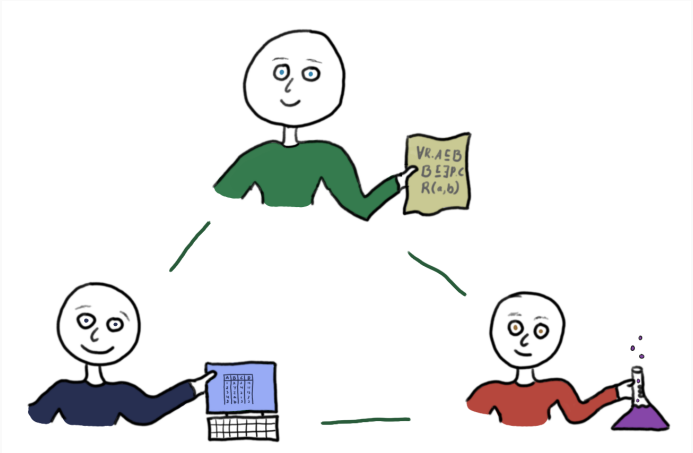
Trend: Cloud services



Need to improve:

- data elasticity (make data easier to move)
- protocols for data replication (master-master vs. master-slave)
- automatic system administration
- multitenancy (multiple users of same physical resource)
- data curation and provenance in the cloud
- support for hybrid clouds (local + cloud)

Trend: Roles of people in the data life cycle



New roles:

- Data producers: Make it easier to add data and metadata
  (e.g., annotations in Google maps)
- Data curators: Make it easier to crowd-source curation
  (e.g., Wikipedia reviewing, like/dislike)
- Data consumers: Make it easier to pose complex queries
  (e.g., graph algorithms, ML)
- Online communities: Make it easier to share, mine, exploit data
  (e.g., transfer from Facebook to Mastodon)

The Seattle Report (2022)

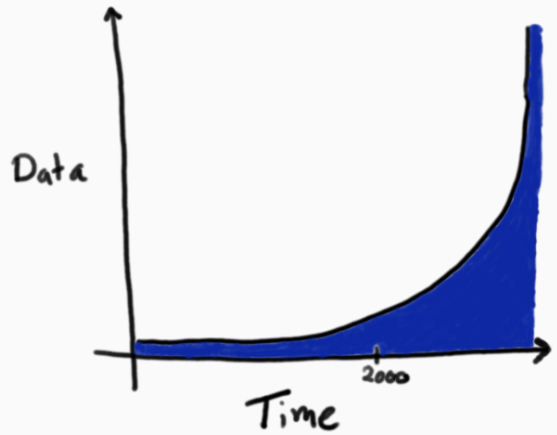
- Available online[2]
- Noted that we have made progress on 2016-trends,
  but still lots to do
- Many new developments (AI/ML, data science)
- Trends:
  - Data science as a field
  - Data governance
  - Cloud data services
  - Database engines

[2] https://dl.acm.org/doi/pdf/10.1145/3524284

Trend: Data science as a field

Need to improve:

- data-to-insight pipelines
  (cleaning, integration, transformation takes 80-90% of the time)
- data provenance
  (track data when integrating, reproduce)
- interactive data exploration at scale
  (not just visualisation)
- declarative programming for data
  (what, not how)
- metadata management for improved ML



Trend: Data governance

Need to improve:

- data use policy (e.g. GDPR)
- data privacy (e.g. in analysis/aggregation)
- ethical considerations
  (e.g. bias, discrimination)



All with automatisation

Trend: Cloud services

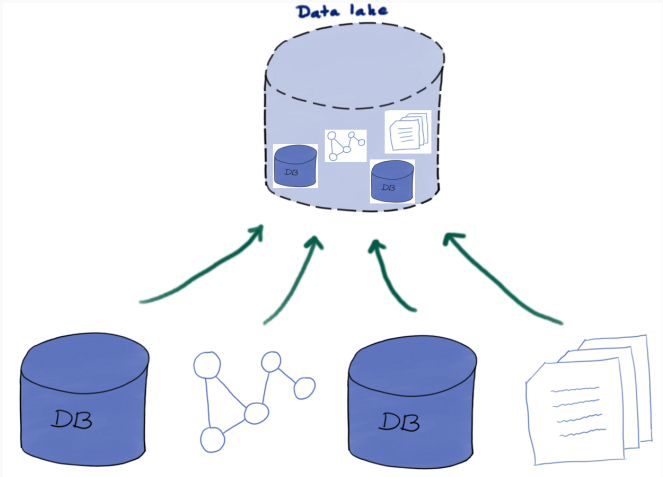
Need to improve:

- systems for serverless cloud database services
  (autoscaling, pay for quick allocation)
- support for disaggregation
  (separate storage and compute)
- support for multitenancy
- support for edge computing
  (compute as much as possible at the source/client)
- support for hybrid/multi cloud
  (several clouds add another layer)
- auto-tuning and configuration
  (e.g., depending on load)
- SaaS (software as a service) cloud database applications
  (study tradeoffs of multitenancy)

Trend: Database engines

Need to:

- support heterogeneous computing (GPU, FPGA)
- improve algorithms for distributed transactions
- make better systems for 'data lakes'
- improve algorithms for query approximation
- improve in-database ML-inferencing
- investigate ML for auto-tuning
- improve benchmarking and reproducibility



Other relevant courses at IFI

- IN3060/4060: Semantic Web Technologies
- IN5040: Advanced Database Systems for Big Data
- IN5800: Declarative Data Engineering
  (not this year, unfortunately)

Database reasearch at UiO: Examples of centers and projects

- SIRIUS: *Centre for Scalable Data Access*[3] (SFI, 2015-2023)
- Respire: *Responsible Explainable Machine Learning for Sleep-related Respiratory Disorders*[4] (2022-2027)
- Integreat: *Norwegian Centre for Knowledge-driven Machine Learning*[5] (SFF, 2023-2033)
- TRUST: *The Norwegian Centre for Trustworthy AI*[6] (AI Billion project, 2025-...)

[3] https://sirius-labs.no/

[4] https://www.mn.uio.no/ifi/forskning/prosjekter/respire/index.html

[5] https://www.integreat.no/

[6] https://www.trust-aicentre.no/english//