



**Data-Intensive systemer:  
kritisk masse, langsigtet forskning  
og værdiskabelse**

Christian S. Jensen

Center for Data-intensive Systems

daisy

**Setting:**  
**Digitalization, Big Data**

# Instrumentation and Digitalization

---

- Instrumentation of reality
  - E.g., smartphones
- Digitization of processes
  - Societal processes, including commerce, public services, social interactions, transportation
    - ◆ E-boks, cf. PostNord AB
    - ◆ Audio and video at home, Amazon Alexa, Apple Siri, Google Assistant, Microsoft Cortana, Smart TVs, ...
    - ◆ Facebook, Twitter, ..., fake news
    - ◆ Transportation apps, rejsekortet, ...
  - Industrial processes
  - Research processes

# Data Explosion: The Digital Universe

---

- [...] the digital universe is **doubling in size every two years** and will multiply 10-fold between 2013 and 2020 – from 4.4 trillion gigabytes to 44 trillion gigabytes.

<http://www.infodocket.com/2014/04/16/how-large-is-the-digital-universe-how-fast-is-it-growing-2014-emc-digital-universe-study-now-available/>

# Big Data

---

- Massive data and unprecedented computing infrastructure combine to offer potentials for obtaining insight and creating value from data.
- To be competitive, society and businesses must be able to create value from data.
- Data-based decisions and data-driven processes
  - Decisions based on good data beat decisions based on feelings or opinions.
- Enables a finer granularity of services and entirely new services



## News &amp; Media

- › News-Archive
- › Publications

09.12.2015

## “Society has to find a balance between sharing data and protecting it”



### Christian S. Jensen

Christian S. Jensen is professor at the Department of Computer Science at Aalborg University in Denmark. His research focuses on the management of spatio-temporal data including modelling, database design and indexing. He previously worked at the universities of Aarhus, Arizona and Maryland as well as at Google's headquarters in Mountain View, CA. He presides over the Steering Committee of NRP 75 “Big Data”.

Privacy issues need to be debated publicly and openly, says Christian S. Jensen, president of the Steering Committee of NRP 75 “Big Data”.

### How do you explain the buzz around Big Data in the media and the business world?

It's the confluence of two developments. First, the amount of available data has exploded:

# **Requirements of a researcher and how to meet them**

# Requirements of a Researcher...

---

- World-class teaching
- World-class dissemination
- World-class research administration
- World-class foundational research
- World-class research-based societal value creation
  
- No single researcher can do it all!
  
- A well-functioning research group with critical mass can.
- Such a group maintains synergistic activities with *pipelines* that include research ranging from curiosity-driven research to agenda-driven research.
- Enter Daisy – Center for Data Intensive Systems...



# **Daisy and its spatio-temporal pipeline**

# Daisy

---

- Some 40 members of staff
  - Some 25 Ph.D. students plus full, associate, and assistant professors and postdocs
- The Ph.D. students and postdocs are funded externally
  - Calls for a sizable portfolio of grants
- Spatio-temporal pipeline
  - Data structures and indexes, algorithms, languages, tools, systems
  - Invention, prototyping, and empirical study
  - A key application area is *intelligent transport systems*, including mobility, traffic, and transportation analytics
  - Much of the research requires evaluation using real data

# Data, Software and Hardware

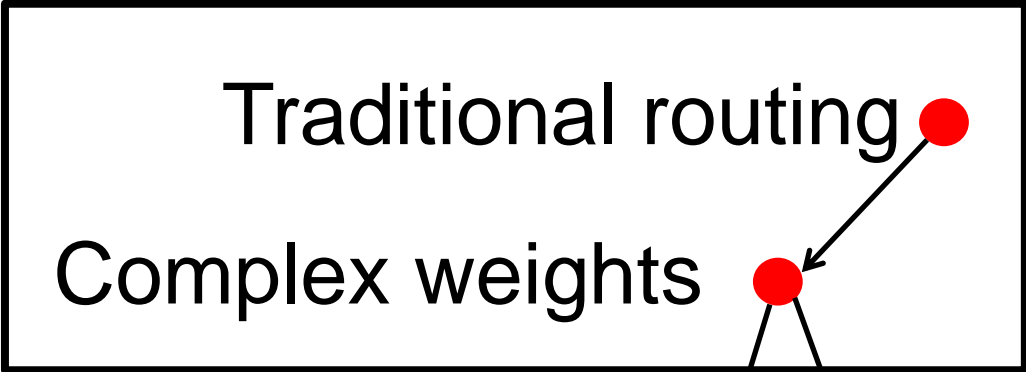
---

- Data
  - 60+ billion GPS records from 200.000+ vehicles, live
  - 10 billion GPS records from bicycles
  - 8 billion GPS records from buses, more available
  - 300+ million GPS and fuel consumption records
  - 250 million GPS and energy consumption records electric vehicles
  - ~20 data sources
- Software and hardware
  - A complete software stack, including data cleaning, warehousing, multiple map support, maps-based analytics
  - Modern servers, up to 2 TB main memory

# New Routing Paradigms

---

Dijkstra's paradigm



Path-centric paradigm



On-the-fly paradigm

Non-cost based routing ●

# Scientific Excellence

---

- Study of publication performance 2000-2009

## Affiliation analysis of database publications

David Aumüller, Erhard Rahm  
University of Leipzig, Leipzig, Germany  
{aumueller, rahm}@informatik.uni-leipzig.de

- Publications in the top-4 most prestigious outlets

Institution	pubs	<i>frac</i>	inst/p	1 <sup>st</sup>	2 <sup>nd</sup>	aut
ETH Zurich	29	<i>23</i>	1.52	11	18	39
Aalborg University	26	<i>14</i>	1.96	11	15	18
Univ. of Edinburgh	25	<i>15</i>	2.16	5	20	16
INRIA Le Chesnay	23	<i>12</i>	2.35	20	3	33
Univ. of Athens	22	<i>12</i>	2.32	6	16	17
MPI Saarbrücken	17	<i>12</i>	1.71	2	15	9
CWI Amsterdam	14	<i>12</i>	1.43	5	9	15
Univ. of Munich	12	<i>9</i>	1.67	10	2	20

Table 6: Top European research institutions

# Two Value Creation Examples

---

- FlexDanmark
  - Organizes public-sector paid transportation by taxi companies
  - Annually, DKK 1.3 billion is paid according to estimated travel time
  - Daisy delivers the data foundation that controls these payments
- Region Nordjylland, Region Midtjylland, a few companies
  - Speed and congestion maps
  - See <http://www.daisy.aau.dk/its> for demos

# Summary

---

- No single researcher can deliver on all parameters.
- Research groups with critical mass are needed.
- Research groups should facilitate synergistic activities by organizing their research into pipelines.
- Example: Daisy's spatio-temporal pipeline
  - A proven track record for transitioning research advances to real systems, bridging long-term and applied research
  - Top academic performance in Europe in its area
  - A healthy portfolio of value-creating applied research activities