Distributed Data Management
Introduction

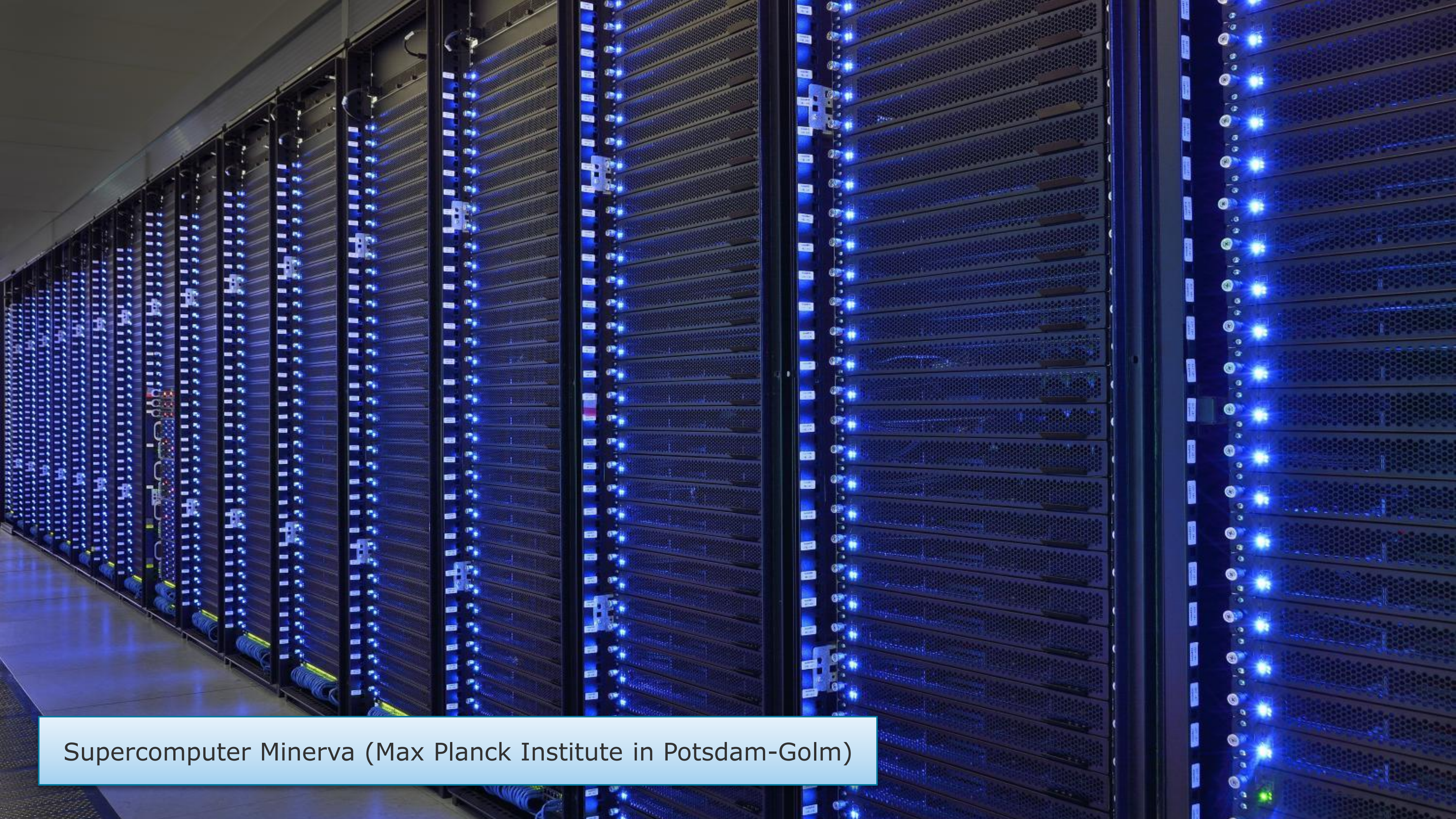
Thorsten Papenbrock

F-2.04, Campus II

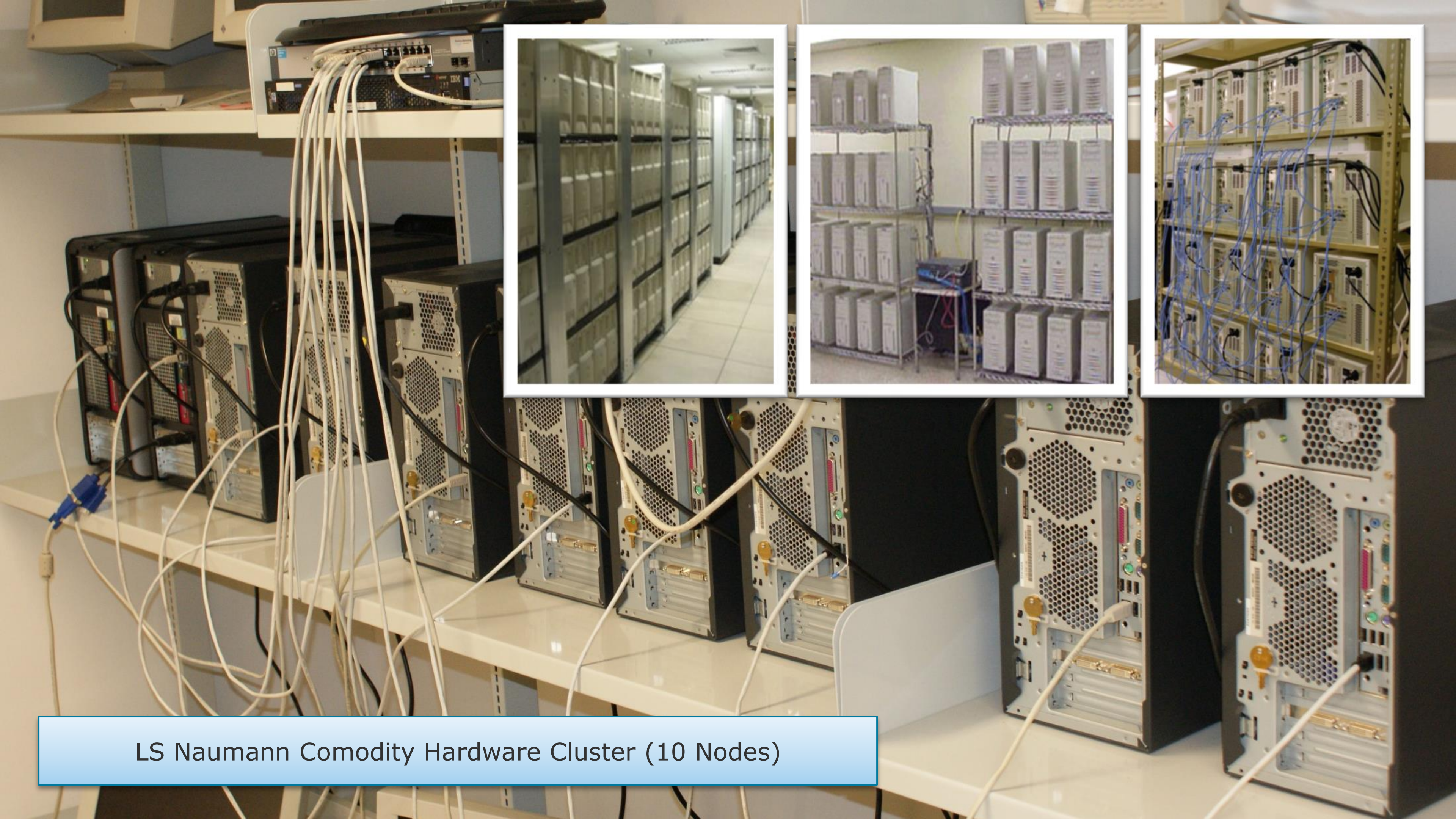
Hasso Plattner Institut

O'REILLY
Martin Kleppmann: *Designing Data-Intensive Applications*
O'Reilly Media, 2017
ISBN 978-1-4493-7332-0

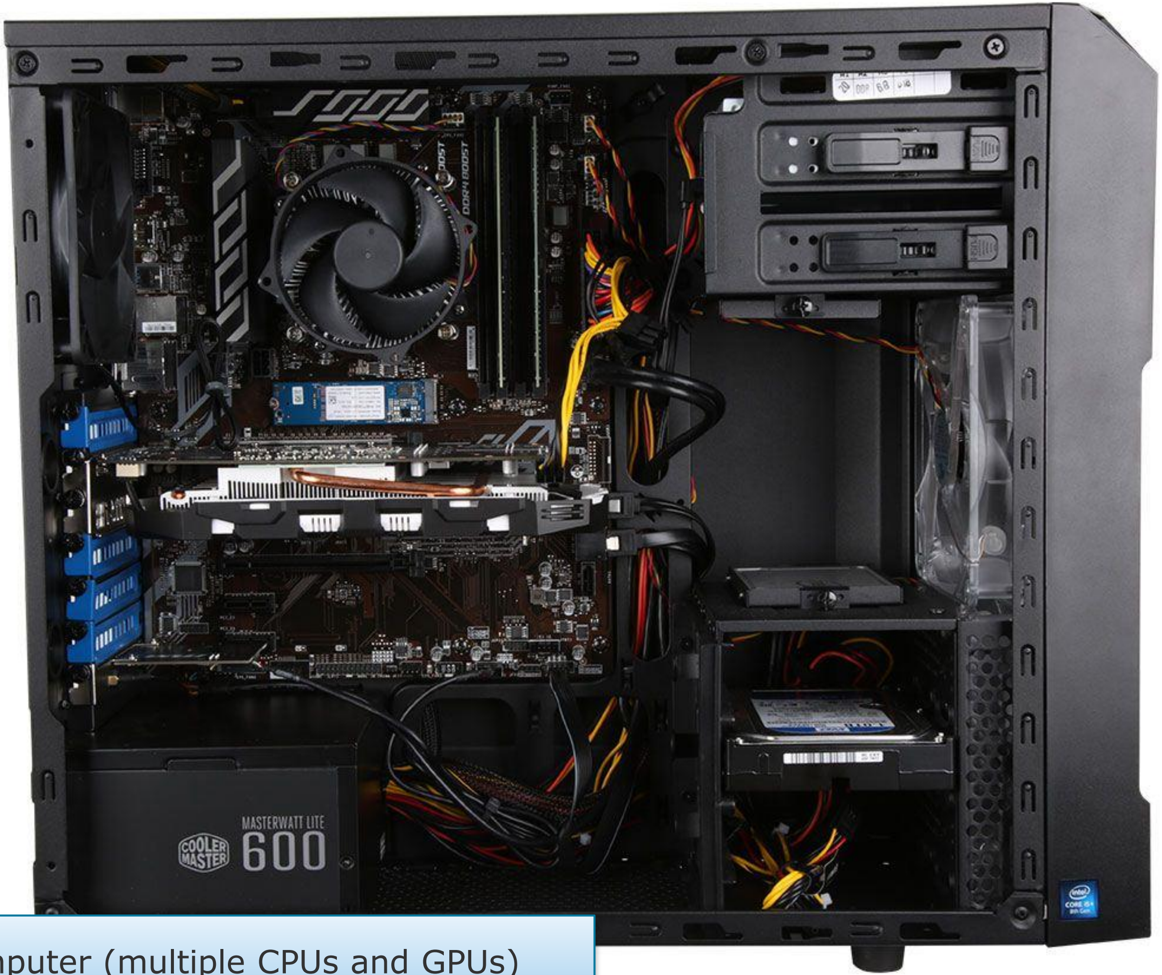
HPI Hasso Plattner Institut
IT Systems Engineering | Universität Potsdam



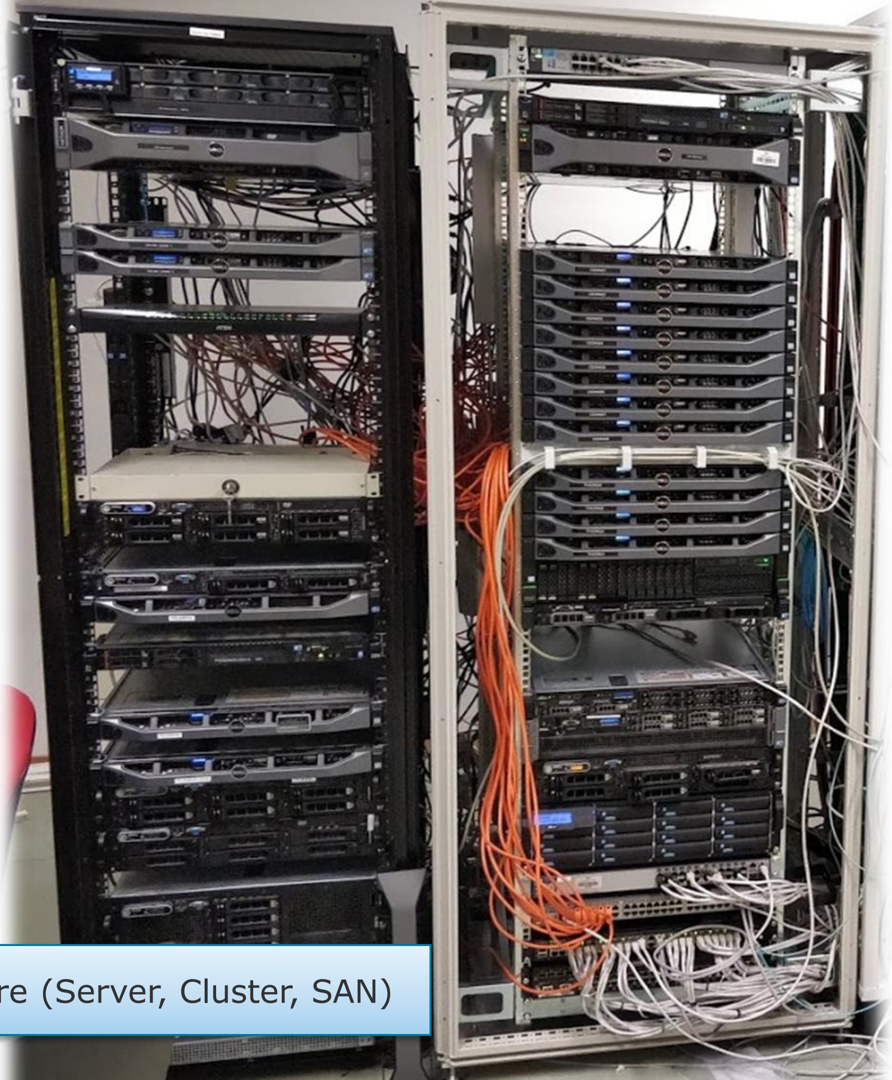
Supercomputer Minerva (Max Planck Institute in Potsdam-Golm)



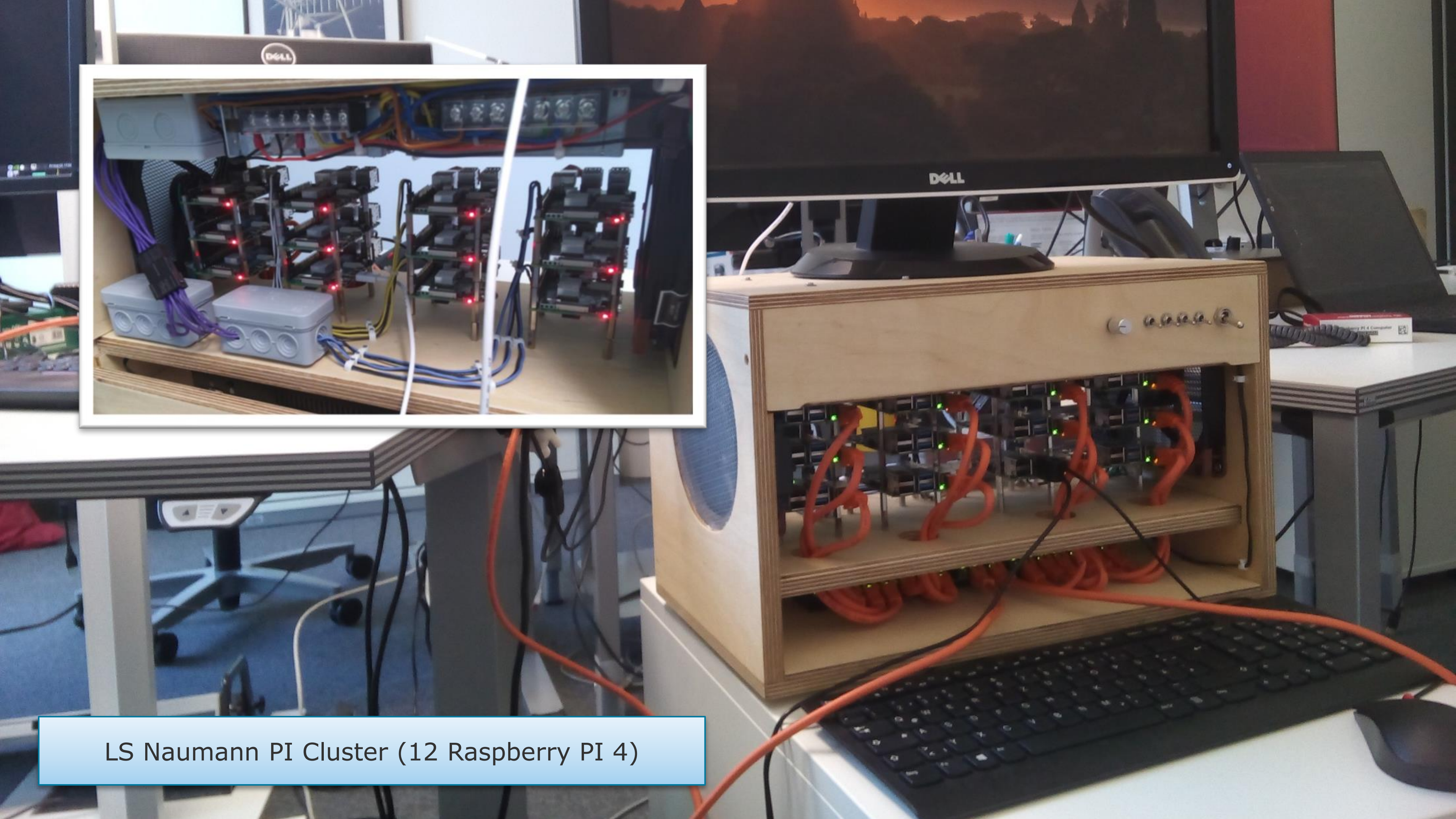
LS Naumann Comodity Hardware Cluster (10 Nodes)



Desktop Computer (multiple CPUs and GPUs)

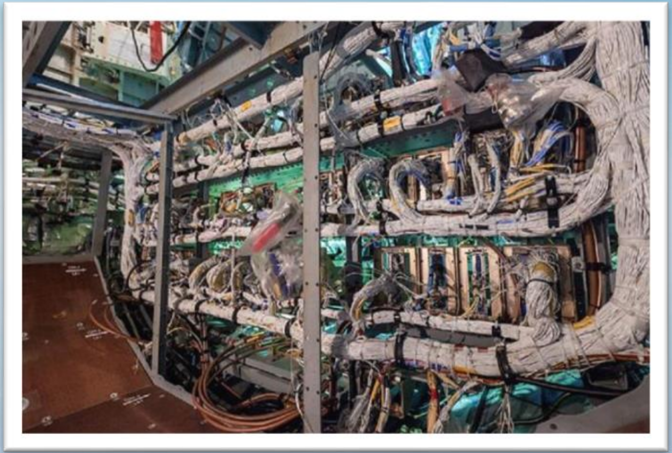


LS Naumann Infrastructure (Server, Cluster, SAN)

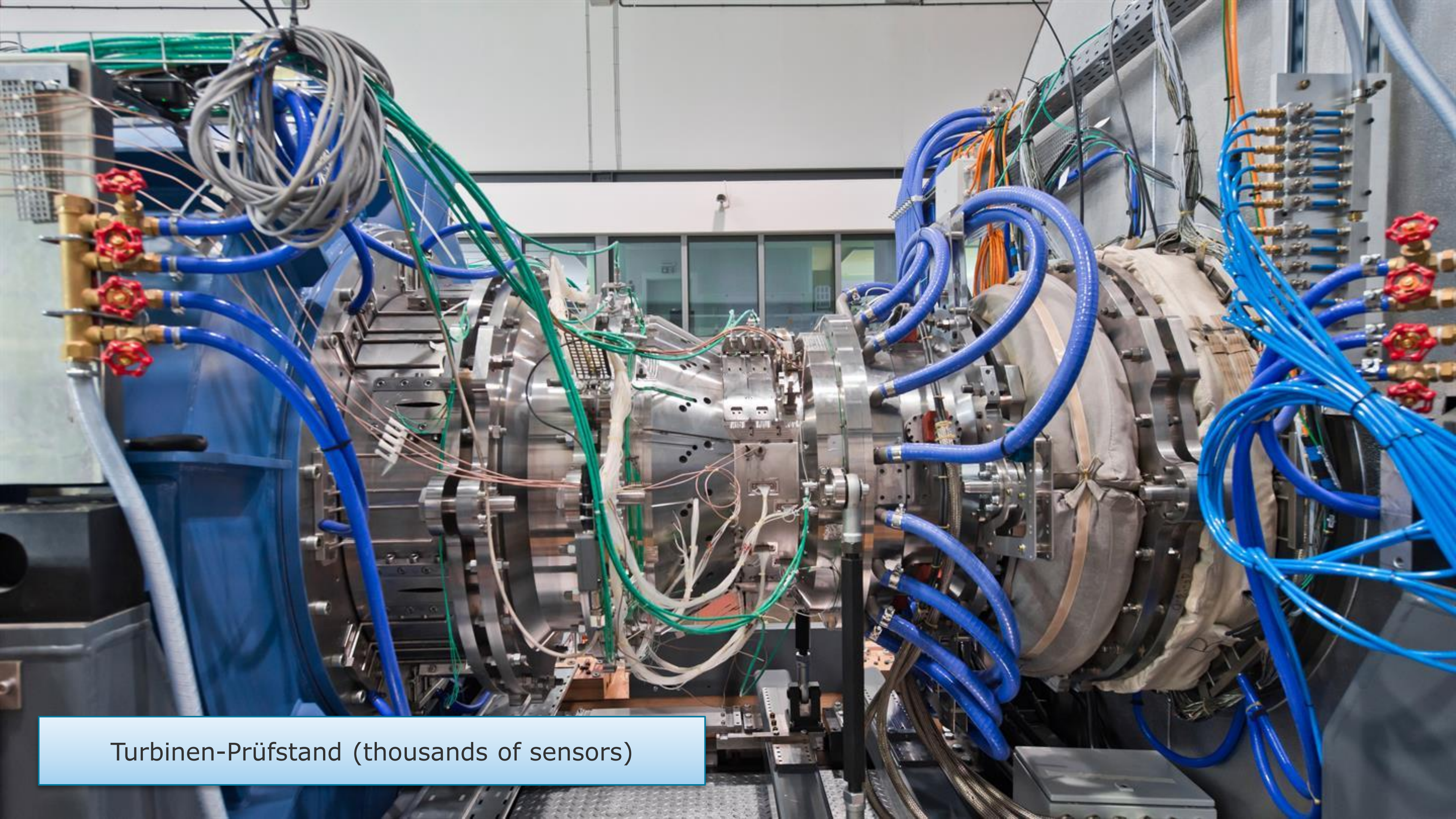


LS Naumann PI Cluster (12 Raspberry PI 4)

DreamHack (12,000-computer LAN party)



Boing 747 (thousands of computers)



Turbinen-Prüfstand (thousands of sensors)

Startpage.com

The world's **most private** search engine

Startpage (search engine backed by other search engines)


Outdated Data
Lost & Invalid Messages
Concurrency
Consensus
Termination

Introduction

- Examples Distributed Systems
- **Lecture Organization**
- Motivation “Distributed”
- Motivation “Data”
- Motivation “Management”

Prof. Felix **Naumann**

Dr. Thorsten **Papenbrock**

Diana **Stephan**

Dr. Ralf **Krestel**

Tim **Repke**

Julian **Risch**

project **DuDe**

Service-Oriented Systems

Data Fusion

Data Scrubbing

project **Stratosphere**

Entity Search

Distributed Computing

Information Integration

Data as a Service

Web Science

project **Metanome**

Duplicate Detection

Data Cleansing

Web Data

Text Mining

Agile Systems

Data Profiling

Dependency Detection

Linked Open Data

RDF Data Mining

Entity Recognition

Data Change

Opinion Mining

ETL Management

project **DataChEx**

Change Exploration

Data Preparation

Phillip **Wenig**

Nitisha **Jain**

Michael **Loster**

Hazar **Harmouch**

John **Koumarelas**

Tobias **Bleifuß**

Leon **Bornemann**

Lan **Jiang**


Which semester?
ITSE, DE, DH?
English?
Database knowledge?
HPI or Guest?
Distributed experience?
Other related lectures?

Courses 2019/2020

- Datenbanksysteme II (VL, Bachelor)
- Information Integration (VL, Master)
- Distributed Data Management (VL, Master)
- Data Engineering in Practice (VL, Bachelor/Master)
- Paint it Black: Ethik in der Datenanalyse (S, Bachelor)
- Paint it Black: Ethical Data Analytics (S, Master)
- Machine Learning for Data Streams (S, Master)
- Genealogy of Natural Language (Masterproject)
- Data Analytics – Museumserlebnisse mit Datenanalyse optimieren (Bachelorprojekt)
  - In Kooperation mit dem Museum Barberini

https://hpi.de/naumann/teaching/current-courses.html

This Lecture

Lecture

- For master students
  (IT-Systems Engineering,
  Digital Health, Data Engineering)
- 6 credit points, 4 SWS
- Mondays 13:30 – 15:00
  Tuesdays 15:15 – 16:45

Exercises

- Interleaved with lectures

Slides

- On website

Prerequisites

- To participate:
  A little background and interest in databases (e.g. DBS I lecture); object oriented programming skills
- For exam:
  Attending lectures, participation in exercises, and completion of exercise homework tasks

Exam

- Written exam
- Probably first week after lectures

Website

- https://hpi.de/naumann/teaching/teaching/ws-1920/distributed-data-management-vl-master.html

Feedback

Hasso Plattner Institut

Question any time please!

- During lectures
- Visit us: Campus II, Room F-2.04
- Email:
  - thorsten.papenbrock@hpi.de

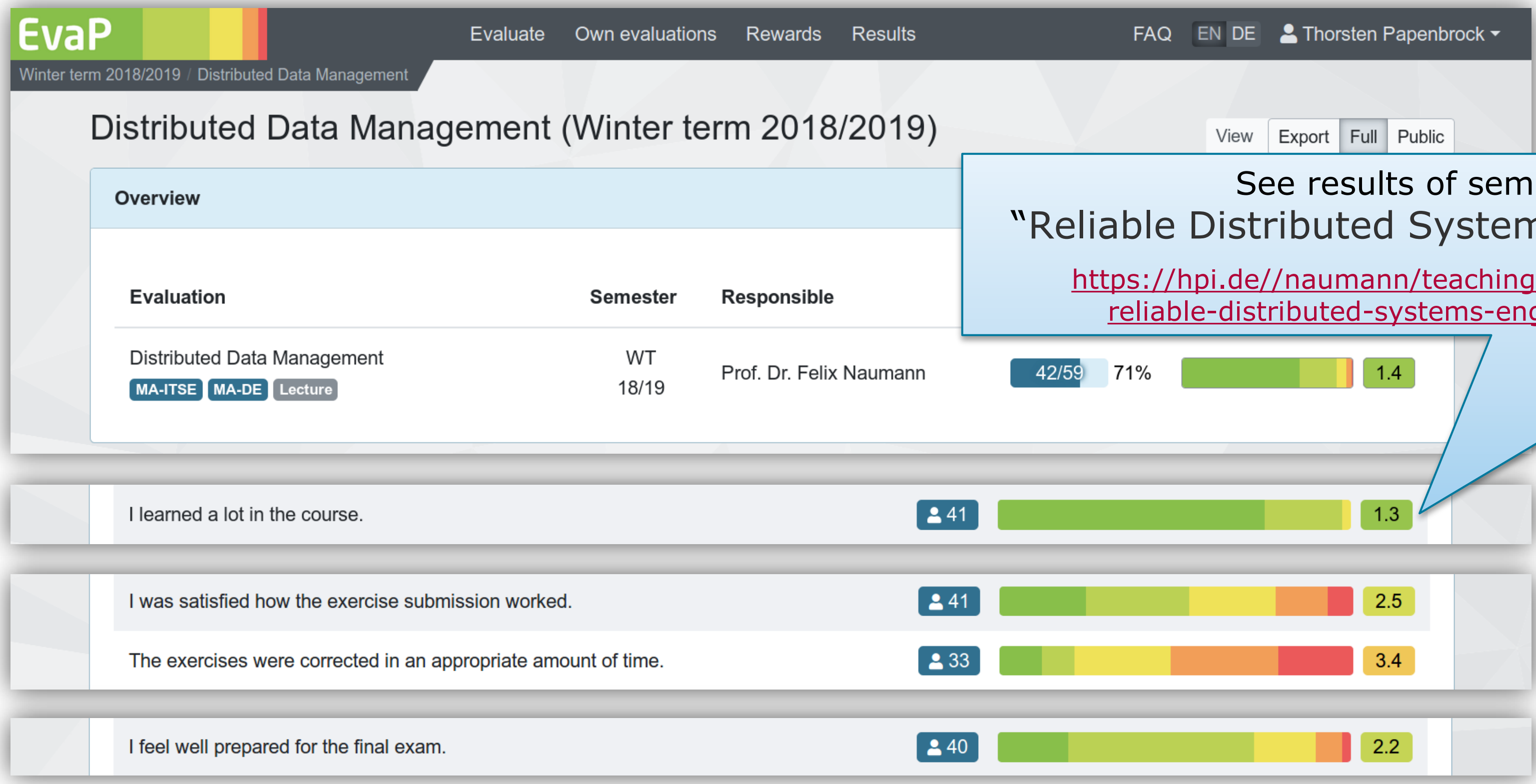
Also: Give feedback about …

- improving lectures
- informational material
- organization

Official evaluation

- At the end of this semester
- … too late for important feedback!

Feedback



See results of seminar
“Reliable Distributed Systems Engineering”
https://hpi.de//naumann/teaching/teaching/ss-19/reliable-distributed-systems-engineering.html

How could the course be further improved?

- Teilweise sollte der Stoff entschlackt werden. Es wurde wirklich sehr viel behandelt. Teilweise merkte man Thorsten zu Ende einer Vorlesung an, dass er schneller und schneller wurde, um bloß noch den Stoff dieser Vorlesung durchzubringen. Ebenso ist die Klausurvorbereitung somit extrem zeitintensiv.

  Leider lief das Check-Yourself-Teil sehr schlecht. Tobias stand zu Fragen vial Mail nicht zur verfügung. Lösungen kamen ab ca. der Mitte des Semesters nur noch sehr sporadisch und zuletzt überhaupt nicht mehr. Schade!
- Time spend on Flink. Structure in which concepts are explained. Make connections to other already known concepts and describe the differences. Or give a quick overview at first and then dive into the -important- details.
- - Lösungen der Check yourselfs rechtzeitig zusenden
  - Übung frühzeitig kontrollieren
  - wenn man eine Übung vorstellen soll, bescheid geben, damit man auch anwesen sein kann
- Teilweise wenig technisch. Außerdem sind die Folien verbesserungswürdig. Sie haben zu viele Überschriften, zu viele Bilder die mit dem Thema nur im übertragenen Sinne zu tun haben. Zu viele Schriftfarben. Zu wenig leicht ersichtliche Gliederung. Sie sind oft nach den Schema: Lösung1, Lösung2, Lösung3, LösungX, ... aufgebaut. Ideal wäre aber eine manchmal deutlichere Motivation des Problems, ein kurzes heads up, dass es drei Lösungen gibt und dann eine deutlich abgegrenzte Besprechung der drei Lösungen. Diese gehen teilweise etwas ineinander über.
- - Especially with the DDM+ slides at the end, it might be worth thinking about adding that content and then splitting it into two lectures? The current lecture already has a ton of content, so it might make sense to go deeper on slightly fewer topics
- Es ist sooo viel. Ich würde mir vielleicht eine kleinere Klausur wünschen, mündlich zum Beispiel (in einer Gruppe?). Dann müsste man nicht nochmal coden und irgendwelche query languages auswendig lernen. :)
- Keine Klausur
- - Zwischenklausur halten, weil es wirklich sehr viel Inhalt war.
  - Übungsevaluierung ist etwas unklar, gibt es nur bestanden oder nicht, wo ist die Grenze?
- The content of the course are very broad, missed possibilities to dive deeper into a topic, since the workload was already quite high.
- Die Vorlesung war für meinen Geschmack zu umfangreich.
- Die Lösungen der Check Yourself Aufgaben haben zum Teil sehr lange auf sich warten lassen, sodass einem die Thematik der zu bearbeitenden Aufgaben nicht mehr genau im Kopf war, wenn es die Lösung gab.
- Point out the motivation more. Like, tell us in the beginning of each set of slides why we are talking about this topic in respect to the scope of the lecture. That would help a lot to know where we are and why we should learn und understand this topic.
- - fragen für die teletask aufzeichnung wiederholen

Lecture Outline (2018 !)

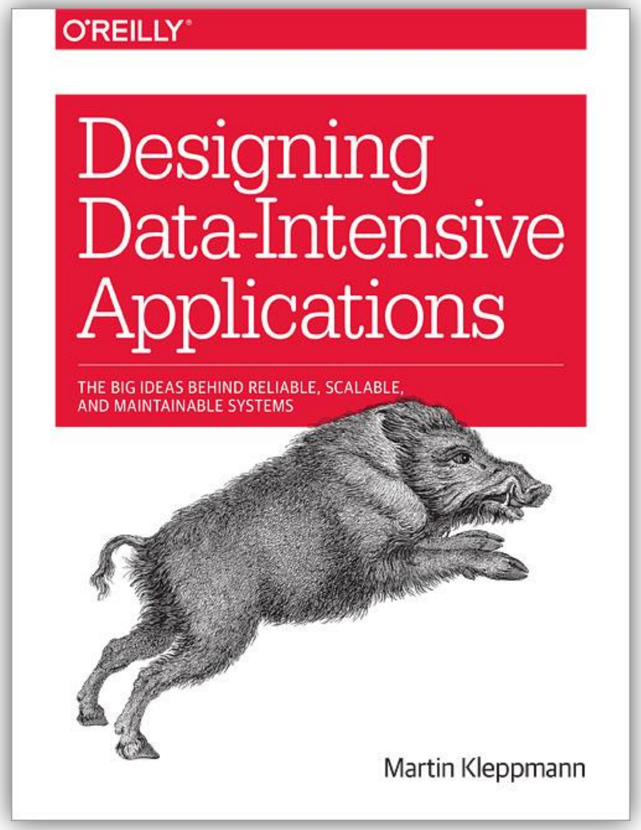
1. Introduction
2. Foundations
3. OLAP and OLTP
4. Encoding and Evolution
5. Hands-On: Akka
6. Data Models and Query Languages
7. Storage and Retrieval
8. Replication
9. Partitioning
10. Batch Processing
11. Hands-On: Spark
12. Distributed Systems
13. Consistency and Consensus
14. Transactions
15. Stream Processing
16. Hands on: Flink
17. Mining Data Streams
18. Distributed Algorithms
19. Services and Containerization
20. Cloud-based Data Systems
21. Lecture Summary and Exam Preparation

Lecture Outline (2018 !) – Homework

1. Introduction
2. Foundations
3. OLAP and OLTP
4. Encoding and Evolution
5. Hands-On: Akka
6. Data Models and Query Languages
7. Storage and Retrieval
8. Replication
9. Partitioning
10. Batch Processing
11. Hands-On: Spark

12. Distributed Systems
13. Consistency and Consensus
14. Transactions
15. Stream Processing
16. Hands on: Flink
17. Mining Data Streams
18. Distributed Algorithms
19. Services and Containerization
20. Cloud-based Data Systems
21. Lecture Summary and Exam Preparation

Literature: Course Book



Designing Data-Intensive Applications

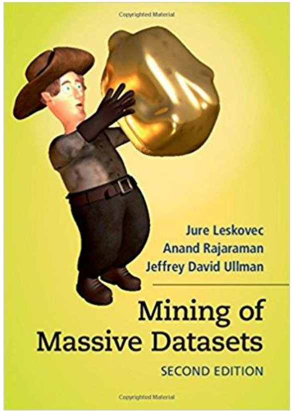
- Author: Martin Klappmann
- Date: March 2017
- Publisher: O'Reilly Media, Inc
- ISBN: 978-1-449-37332-0
- References: https://github.com/ept/ddia-references

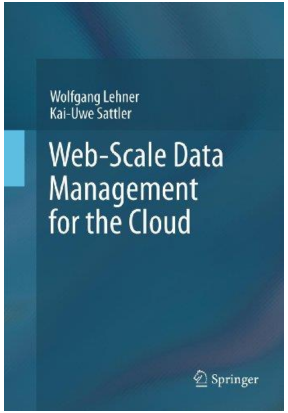
Scope for this lecture

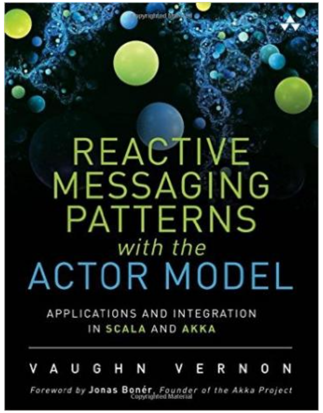
- Distributed and parallel systems
- Big data storage
- Batch and stream processing

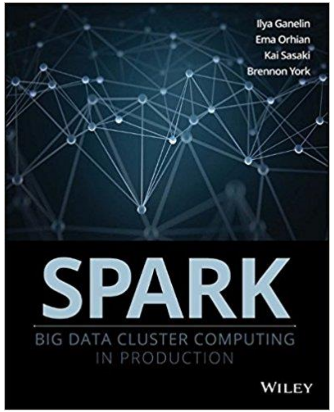
Distributed Data Management

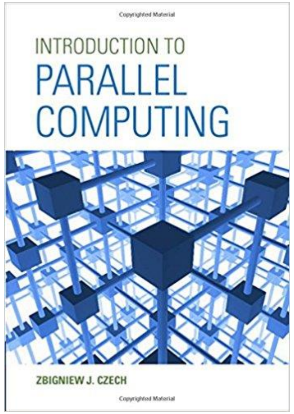
Literature: Further Reading

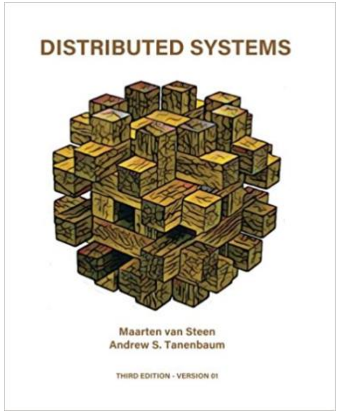


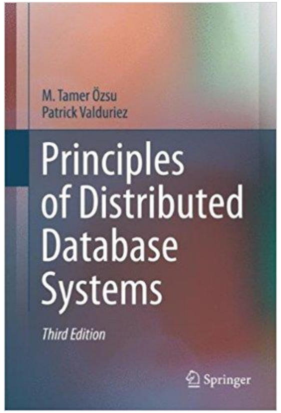












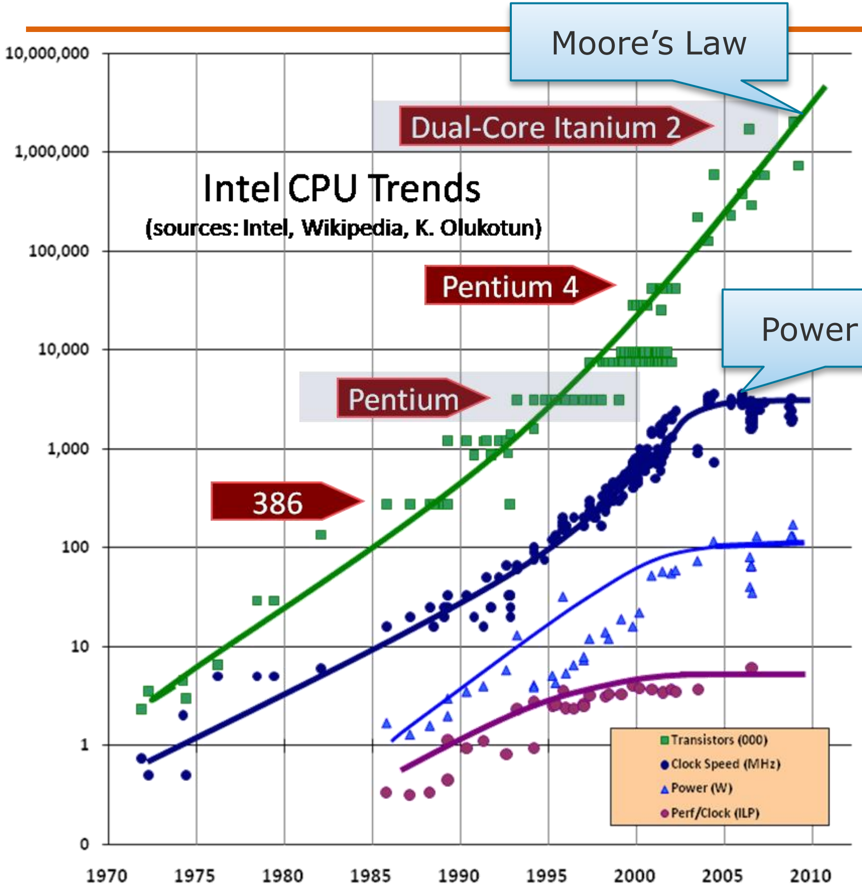
And Web-links that are given on the slides during the lecture.

Introduction

- Examples Distributed Systems
- Lecture Organization
- **Motivation “Distributed”**
- Motivation “Data”
- Motivation “Management”

Motivation: "Distributed"

Paradigm Shift in Software-Writing



The free lunch is over!

- Clock speeds stall
- Transistor numbers still increase
  - Cores in CPUs/GPUs
    CPUs/GPUs in compute nodes,
    compute nodes in clusters
- Paradigm Shift:
  - Earlier: optimize code for a single thread
  - Now: solve tasks in parallel

Distributed computing

"Distribution of work on (potentially) physically isolated compute nodes"

http://www.gotw.ca/publications/concurrency-ddj.htm

Motivation: “Distributed”

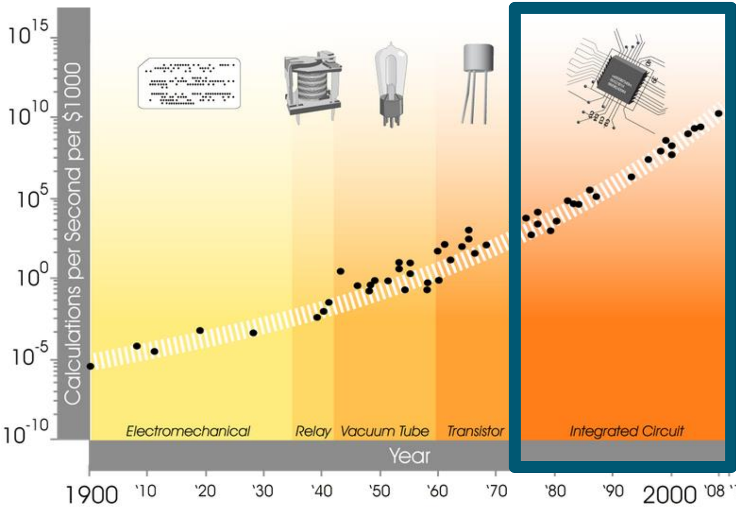
Surpassing Moor’s Law

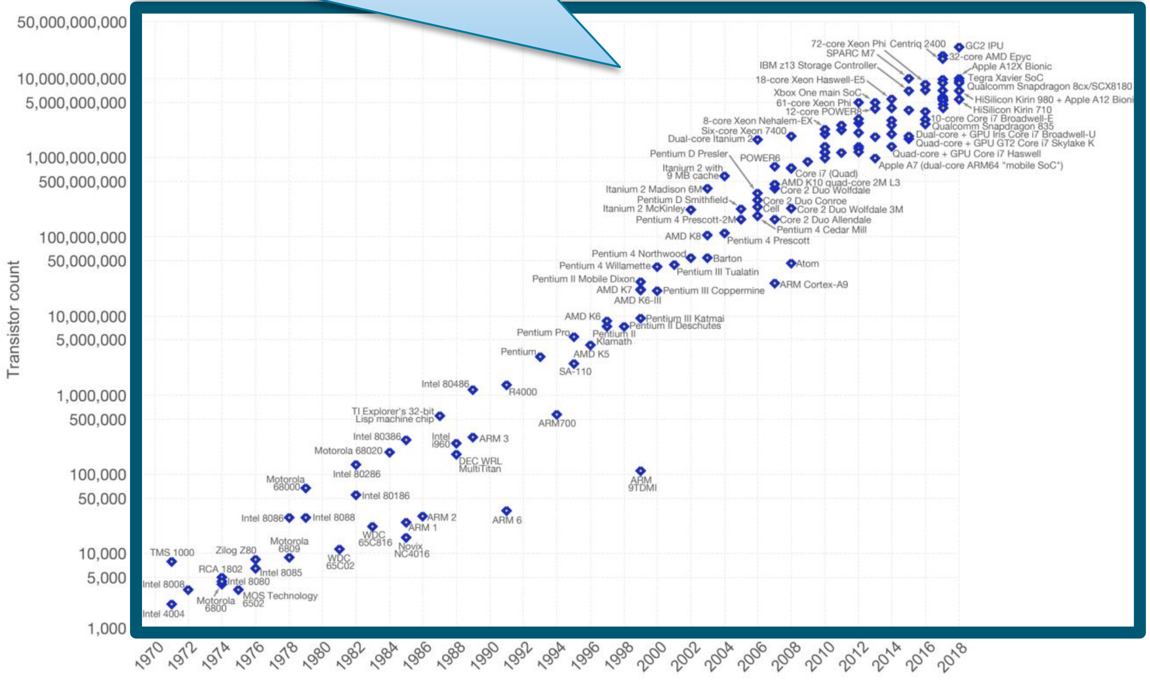
Hyperscale: With clusters of **distributed machines**, we can already build systems with any number of transistors!

(don’t even need to wait for a new processors)

Moore’s Law (Observation)

“The number of transistors on integrated circuit chips doubles approximately every two years”





Data source: Wikipedia (https://en.wikipedia.org/wiki/Transistor_count)
The data visualization is available at OurWorldinData.org. There you find more visualizations and research on this topic.

Motivation: "Distributed"

High Performance and Hyperscale Computing

- High Performance Computing (HPC)
  - Super computers
    - Specialized hardware (NUMA systems)
    - Heterogeneous hardware (FPGAs, GPUs, etc.)
  - Precision matters
    - Floating points per second (FLOPS)
  - Scientific and analytical use cases
    - OLAP, simulations, forecasts, machine learning, data mining, …
- Hyperscale Computing
  - Standard computers
    - Fast commodity servers
  - Response time, availability and throughput matters
    - X-percentile response time, queries-per-second, …
  - Scalable systems (and analytical) use cases
    - OLTP, web services, application hosting, cloud, data transformation, …

Both use distributed computing!

A Rule to Acknowledge

Even **distributed parallelization** cannot work around this law!

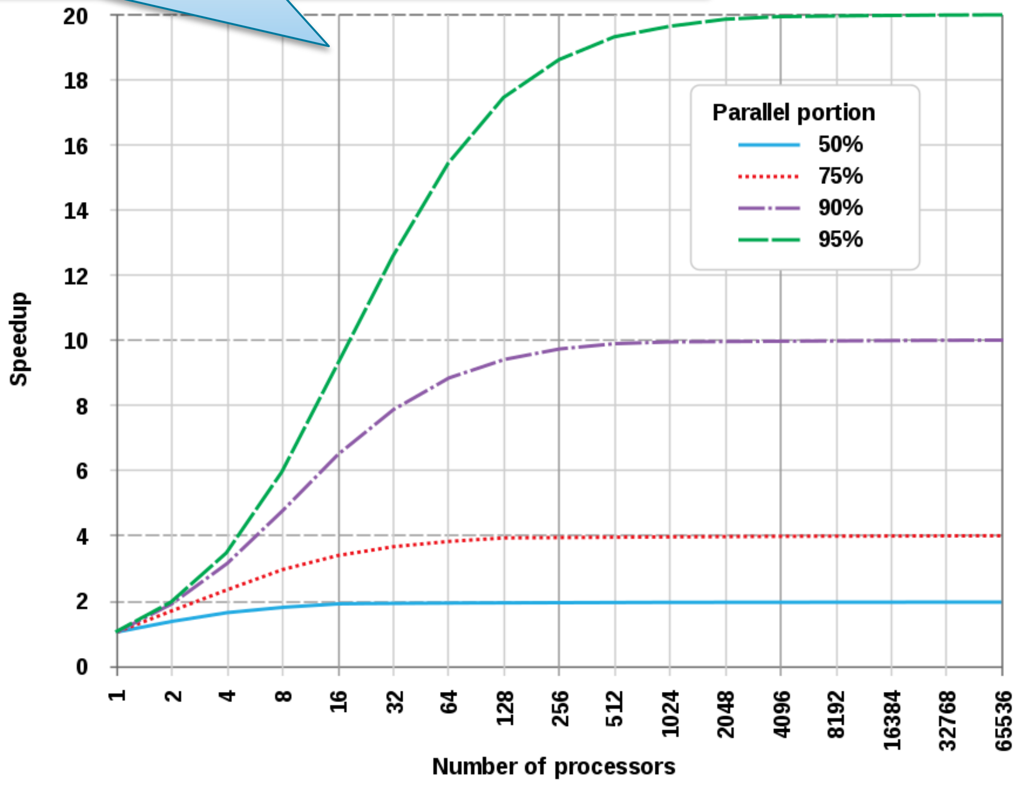
Amdahl's Law

"The speedup of a program using multiple processors for parallel computing is limited by the sequential fraction of the program"

$$Speedup(s) = \frac{1}{(1-p) + \frac{p}{s}}$$

s: degree of parallelization (e.g. #cores)

p: percentage of the algorithm that profits from parallelization



New Technologies

Distributed Computing

hadoop MapReduce

akka

Spark

Flink

STORM

...

Distributed Storage

hadoop

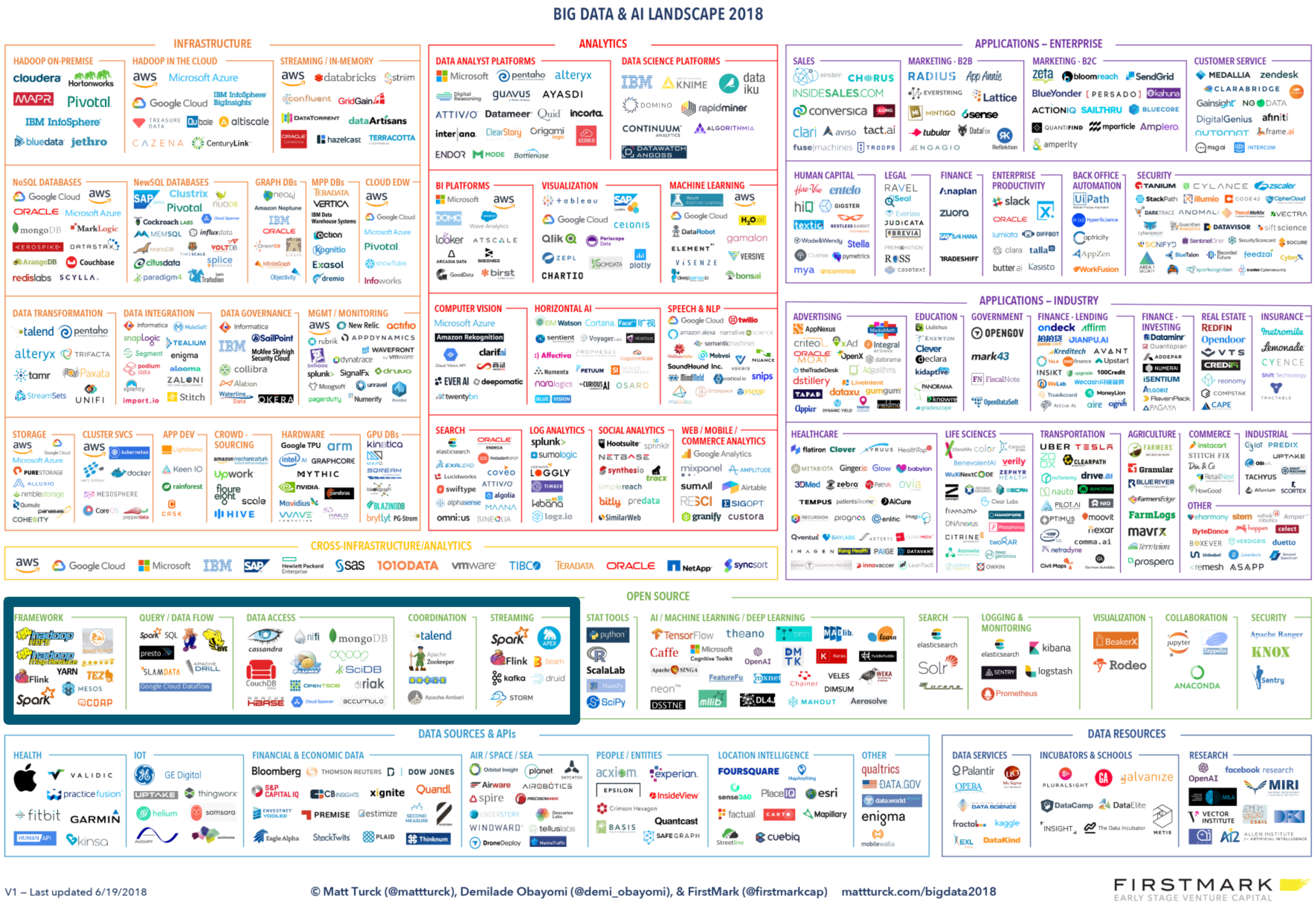
redis

CouchDB relax

cassandra

mongoDB

...



Hasso Plattner Institut

Distributed Data Management

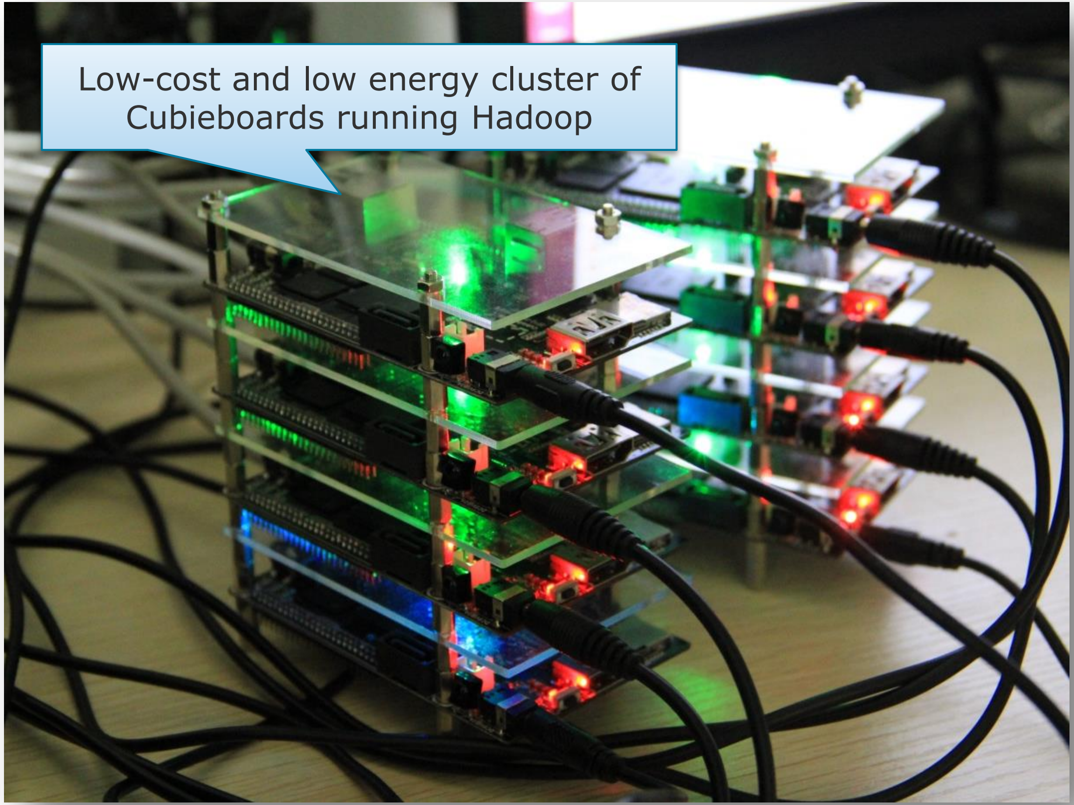
Introduction

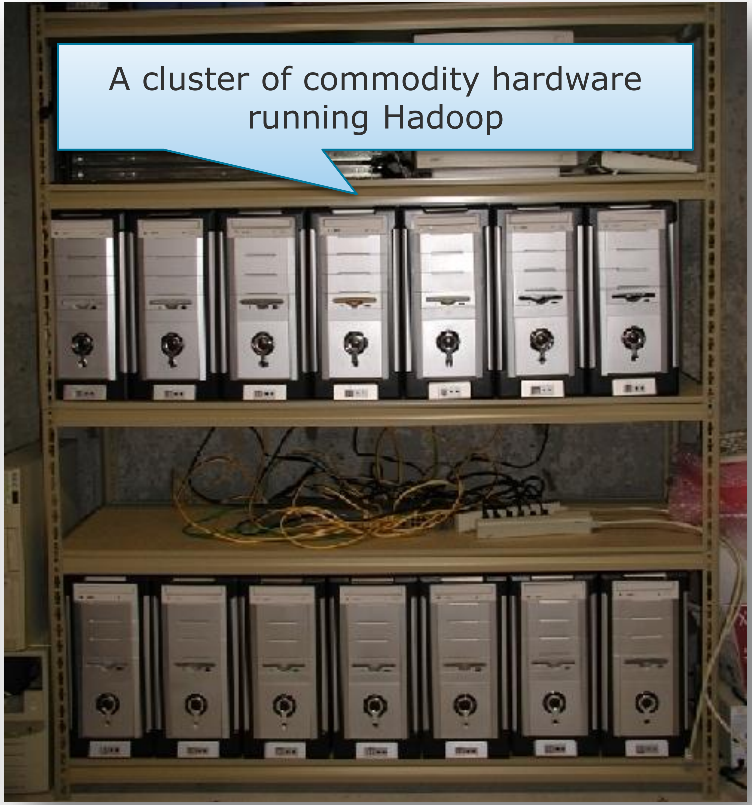
Driving Forces

- Data volumes increase:
  business data, sensor data, social media data, …
- Data analytics gains importance:
  downtime-less, real-time, predictive
- Parallelization paradigm shifts:
  multi-core and network speeds increase while CPU clock speeds stall
- Computation resources become more available:
  IaaS, PaaS, SaaS
- Free and open source software gains popularity:
  setting standards, utilizing external development resources, improving software quality, avoiding vendor locks …

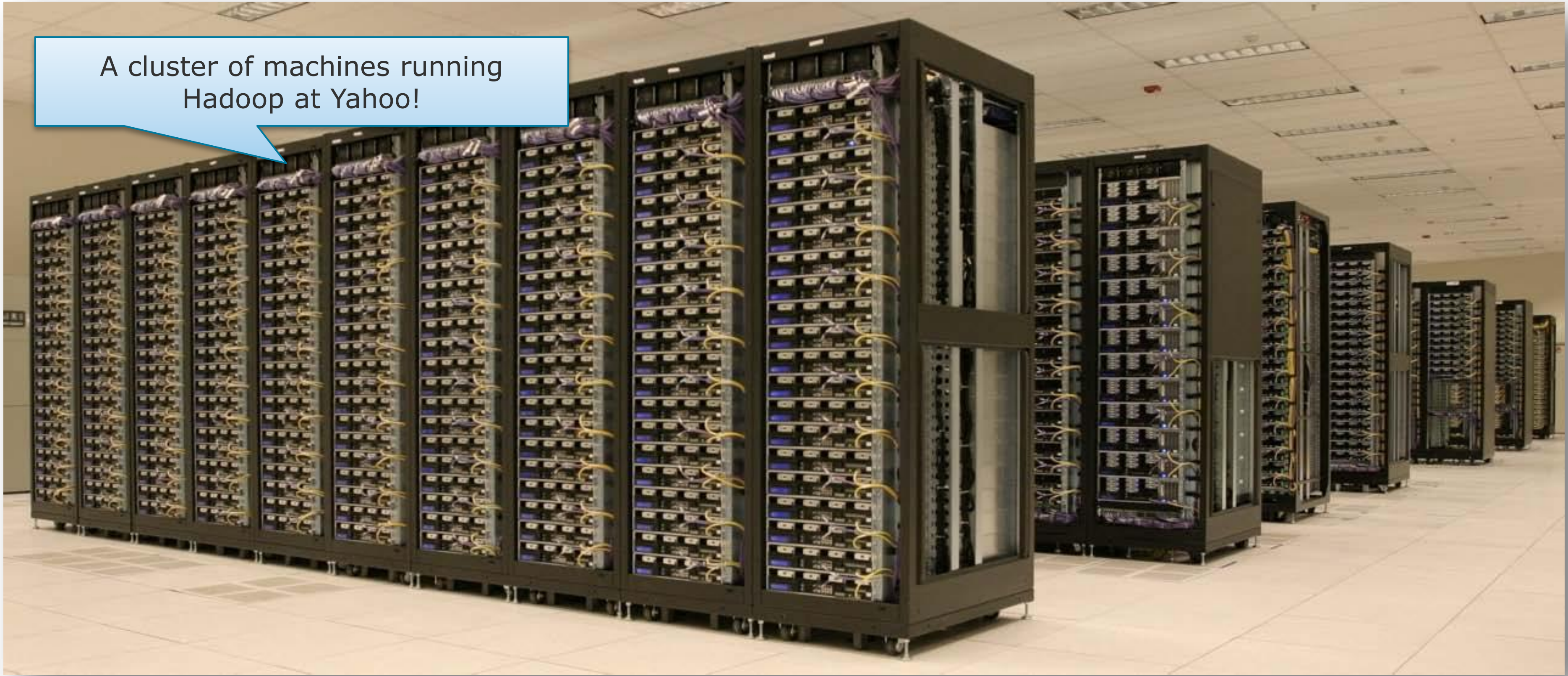
Motivation: "Distributed"

Small and Medium Scale





Large Scale

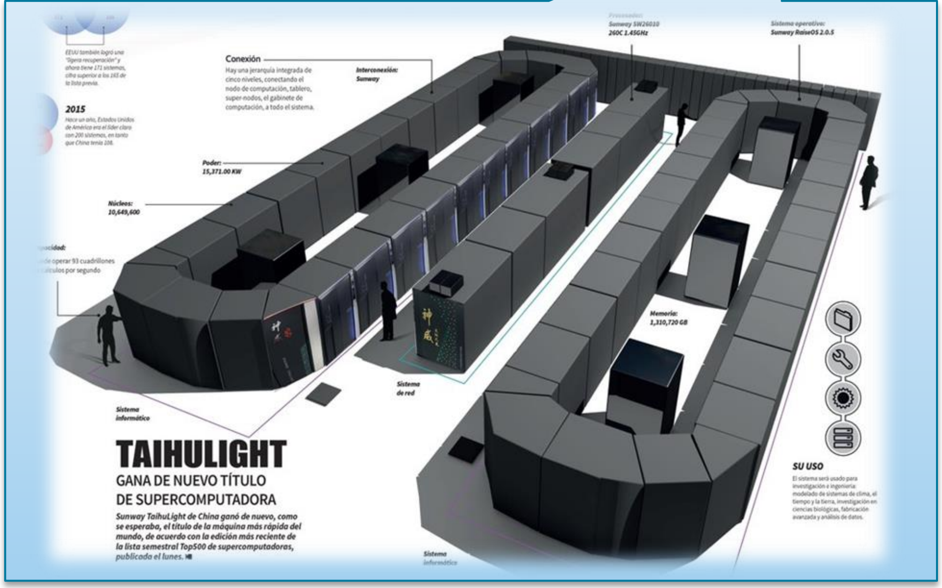


A cluster of machines running Hadoop at Yahoo!

Motivation: "Distributed"

Super Large Scale

Top 10 Super Computers 2017



All **distributed systems**!

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway , NRCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	125,435.9	15,371
2	**Tianhe-2 (MilkyWay-2)** - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P , NUDT National Super Computer Center in Guangzhou China	3,120,000	33,862.7	54,902.4	17,808
3	**Piz Daint** - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 , Cray Inc. Swiss National Supercomputing Centre (CSCS) Switzerland	361,760	19,590.0	25,326.3	2,272
4	**Titan** - Cray XK7, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x , Cray Inc. DOE/SC/Oak Ridge National Laboratory United States	560,640	17,590.0	27,112.5	8,209
5	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom , IBM DOE/NNSA/LLNL United States	1,572,864	17,173.2	20,132.7	7,890
6	**Cori** - Cray XC40, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect , Cray Inc. DOE/SC/LBNL/NERSC United States	622,336	14,014.7	27,880.7	3,939
7	**Oakforest-PACS** - PRIMERGY CX1640 M1, Intel Xeon Phi 7250 68C 1.4GHz, Intel Omni-Path , Fujitsu Joint Center for Advanced High Performance Computing Japan	556,104	13,554.6	24,913.5	2,719
8	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect , Fujitsu RIKEN Advanced Institute for Computational Science (AICS) Japan	705,024	10,510.0	11,280.4	12,660
9	**Mira** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom , IBM DOE/SC/Argonne National Laboratory United States	786,432	8,586.6	10,066.3	3,945
10	**Trinity** - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Aries interconnect , Cray Inc. DOE/NNSA/LANL/SNL United States	301,056	8,100.9	11,078.9	4,233

https://www.top500.org/lists/2017/06/

Motivation: "Distributed"

Super Large Scale

Top 10 Super Computers 2017

All **distributed systems**!

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway , NRCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	125,435.9	15,371
2	**Tianhe-2 (MilkyWay-2)** - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P , NUDT National Super Computer Center in Guangzhou China	3,120,000	33,862.7	54,902.4	17,808
3	**Piz Daint** - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 , Cray Inc. Swiss National Supercomputing Centre (CSCS) Switzerland	361,760	19,590.0	25,326.3	2,272
4	**Titan** - Cray XK7, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x , Cray Inc. DOE/SC/Oak Ridge National Laboratory United States	560,640	17,590.0	27,112.5	8,209
5	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom , IBM DOE/NNSA/LLNL United States	1,572,864	17,173.2	20,132.7	7,890
6	**Cori** - Cray XC40, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect , Cray Inc. DOE/SC/LBNL/NERSC United States	622,336	14,014.7	27,880.7	3,939
7	**Oakforest-PACS** - PRIMERGY CX1640 M1, Intel Xeon Phi 7250 68C 1.4GHz, Intel Omni-Path , Fujitsu Joint Center for Advanced High Performance Computing Japan	556,104	13,554.6	24,913.5	2,719
8	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect , Fujitsu RIKEN Advanced Institute for Computational Science (AICS) Japan	705,024	10,510.0	11,280.4	12,660
9	**Mira** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom , IBM DOE/SC/Argonne National Laboratory United States	786,432	8,586.6	10,066.3	3,945
10	**Trinity** - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Aries interconnect , Cray Inc. DOE/NNSA/LANL/SNL United States	301,056	8,100.9	11,078.9	4,233

https://www.top500.org/lists/2017/06/

Motivation: “Distributed”

Super Large Scale

Top 10 Super Computers 2017

All **distributed systems**!

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway , NRCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	125,435.9	15,371
2	**Tianhe-2 (MilkyWay-2)** - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P , NUDT National Super Computer Center in Guangzhou China	3,120,000	33,862.7	54,902.4	17,808
3	**Piz Daint** - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 , Cray Inc. Swiss National Supercomputing Centre (CSCS) Switzerland	361,760	19,590.0	25,326.3	2,272
4	**Titan** - Cray XK7, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x , Cray Inc. DOE/SC/Oak Ridge National Laboratory United States	560,640	17,590.0	27,112.5	8,209
5	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom , IBM DOE/NNSA/LLNL United States	1,572,864	17,173.2	20,132.7	7,890
6	**Cori** - Cray XC40, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect , Cray Inc. DOE/SC/LBNL/NERSC United States	622,336	14,014.7	27,880.7	3,939
7	**Oakforest-PACS** - PRIMERGY CX1640 M1, Intel Xeon Phi 7250 68C 1.4GHz, Intel Omni-Path , Fujitsu Joint Center for Advanced High Performance Computing Japan	556,104	13,554.6	24,913.5	2,719
8	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect , Fujitsu RIKEN Advanced Institute for Computational Science (AICS) Japan	705,024	10,510.0	11,280.4	12,660
9	**Mira** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom , IBM DOE/SC/Argonne National Laboratory United States	786,432	8,586.6	10,066.3	3,945
10	**Trinity** - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Aries interconnect , Cray Inc. DOE/NNSA/LANL/SNL United States	301,056	8,100.9	11,078.9	4,233

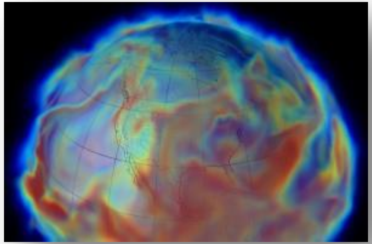
https://www.top500.org/lists/2017/06/

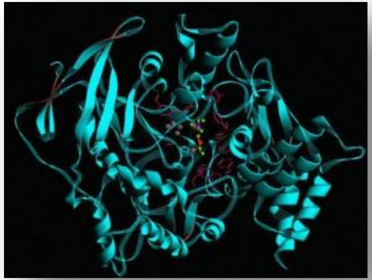
Motivation: “Distributed”

Super Large Scale

Use cases

- Weather forecasting
- Market analysis
- Crash simulation
- Disaster simulation
- Brute force decryption
- Molecular dynamics modeling
- …





Data-intensive analytics tasks!

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway , NRCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	125,435.9	15,371
2	**Tianhe-2 (MilkyWay-2)** - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P , NUDT National Super Computer Center in Guangzhou China	3,120,000	33,862.7	54,902.4	17,808
3	**Piz Daint** - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 , Cray Inc. Swiss National Supercomputing Centre (CSCS) Switzerland	361,760	19,590.0	25,326.3	2,272
4	**Titan** - Cray XK7, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x , Cray Inc. DOE/SC/Oak Ridge National Laboratory United States	560,640	17,590.0	27,112.5	8,209
5	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom , IBM DOE/NNSA/LLNL United States	1,572,864	17,173.2	20,132.7	7,890
6	**Cori** - Cray XC40, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect , Cray Inc. DOE/SC/LBNL/NERSC United States	622,336	14,014.7	27,880.7	3,939
7	**Oakforest-PACS** - PRIMERGY CX1640 M1, Intel Xeon Phi 7250 68C 1.4GHz, Intel Omni-Path , Fujitsu Joint Center for Advanced High Performance Computing Japan	556,104	13,554.6	24,913.5	2,719
8	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect , Fujitsu RIKEN Advanced Institute for Computational Science (AICS) Japan	705,024	10,510.0	11,280.4	12,660
9	**Mira** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom , IBM DOE/SC/Argonne National Laboratory United States	786,432	8,586.6	10,066.3	3,945
10	**Trinity** - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Aries interconnect , Cray Inc. DOE/NNSA/LANL/SNL United States	301,056	8,100.9	11,078.9	4,233

Introduction

- Examples Distributed Systems
- Lecture Organization
- Motivation “Distributed”
- **Motivation “Data”**
- Motivation “Management”

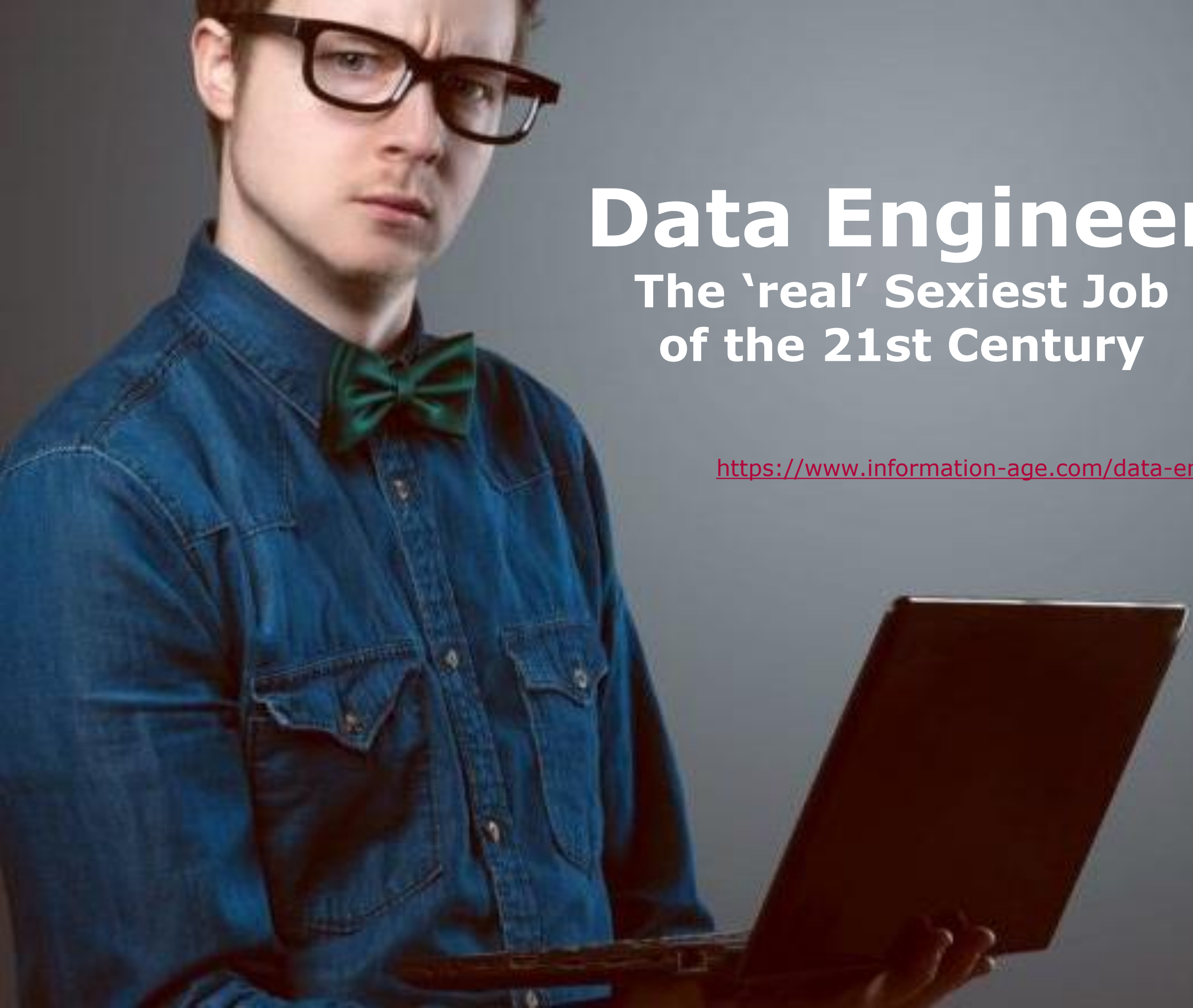
Data Scientist

The Sexiest Job of the 21st Century

https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

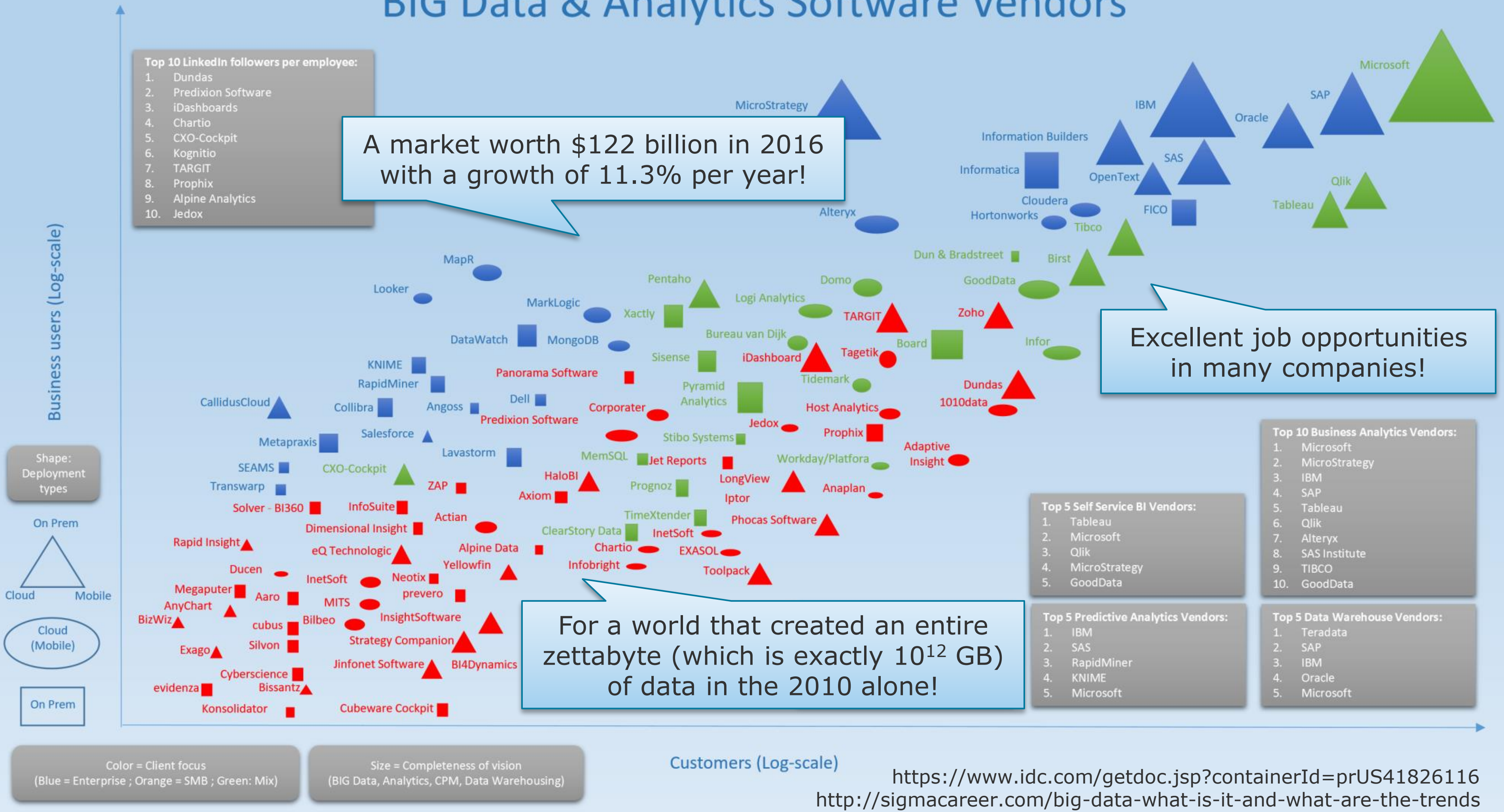
Data Engineer

The ‘real’ Sexiest Job of the 21st Century



https://www.information-age.com/data-engineer-sexiest-job-21st-century-123480578/

BIG Data & Analytics Software Vendors



https://www.idc.com/getdoc.jsp?containerId=prUS41826116
http://sigmacareer.com/big-data-what-is-it-and-what-are-the-trends

VLDB 2017 Program

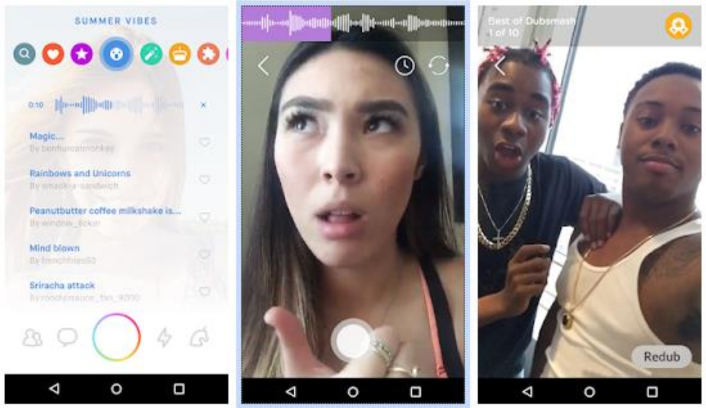
International conference "Very Large Data Bases"

All data processing and analytics tasks that are more and more based on distributed computing.

29 Aug (Tue)								
8:30-10:00		Welcome Messages + Wolfgang Lehner keynote						
10:00-10:30	break							
10:30-12:00		Panel		DB Engines 1	Data Cleaning	Spatial Data Management 1	Graphs and Networks 1	Tutorial: Caching at the Web Scale
12:00-13:30	lunch							
13:30-15:00		Query Processing and Optimization		Stream Processing 1	Data Formats	Privacy and Security	Event Processing	Tutorial: Summarizing Static and Dynamic Big Graphs
15:00-15:30	break							
7:00		Transactions		Spatial Data management 2	Graphs and Networks 2	Information Integration	Applications	Tutorial: Blockchains and Databases
:00			poster reception					
(Wed)								
8:30-10:00		Endowment Update + VLDB Awards + Michael Franklin keynote						
10:00-10:30	break							
10:30-12:00		High Performance Query Processing		Recommendations and Skylines	Approximation Structures	Stream Processing 2		Tutorial: New Trends on Exploratory Methods for Data Analytics
12:00-13:30	lunch							
5:00		Scalable Storage		Crowdsourcing	Stream Processing 3	Graphs and Networks 3	Social Network Analysis	Tutorial: Geometric Approaches for Top-k Queries
5:30	break							
00		Concurrency Control		Query Optimization	Specialized Data Management	Graphs and Networks 4	Data Mining and Analytics	Tutorial: Spatial Crowdsourcing: Challenges, Techniques, and Applications
:00			poster reception					
..	**"VLDB Octoberfest" banquet at the Hofbräuhaus**							
(Thu)								
30		Endowment Awards + Demo Award + Jens Dittrich plenary						
10:30-10:55	break							
10:55-12:00		Transactions and Persistence		Data Access	Data Statistics	Potpourri		
12:00-13:30	lunch							
13:30-15:00		Data Partitioning		Graphs and Networks 5	Visualization	Distributed Systems and Cloud 1	Estimation and Approximation	Tutorial: The Era of Big Spatial Data
15:00-15:30	break							
15:30-17:00		DB Engines 2		Query Processing	Text and Semi-Structured	Spatial Data Management 3	Distributed Systems and Cloud 2	Tutorial: Complex Event Recognition in the Big Data Era
17:00-18:00			poster reception					

Motivation: “Data”

Successful IT Startups



Example: Mobile Motion GmbH

Dubsmash

- An HPI-Startup of 2013
- Founders:
  - Jonas Drüppel, Roland Grenke, Daniel Taschik

November 19, 2014:
- **Launch** of the Dubsmash app

November 26, 2014:
- Dubsmash reached the **number one** downloaded app in Germany

June 1, 2015:
- Dubsmash had been downloaded over **50 million times in 192 countries**

Motivation: "Data"

Successful IT Startups

Successful IT Startups

Successful IT-Startups in recent years are masters of data:

1. **AirBnB**
2. **Instagram**
3. **Pinterest**
4. **Angry Birds**
5. **Linkedin**
6. **Uber**
7. **Snapchat**
8. **WhatsApp**
9. **Twitter**
10. **Facebook**
11. **…**

Peta- to Exabytes of …

- profile data (names, addresses, friends, …)
- content data (images, videos, messages, …)
- event data (logins, interactions, games, …)
- …

Challenged with …

- streaming
- persistence
- analytics
- load-balancing
- …

Introduction

- Examples Distributed Systems
- Lecture Organization
- Motivation “Distributed”
- Motivation “Data”
- **Motivation “Management”**

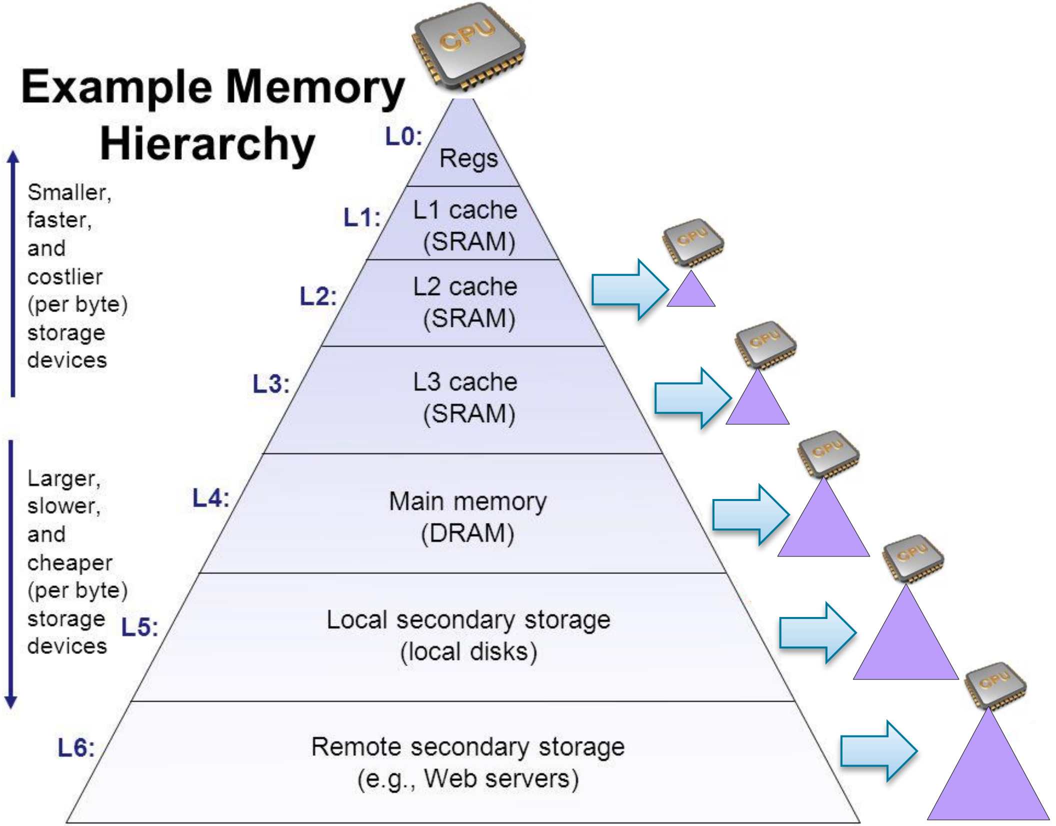
Motivation: "Management"

Rethinking Data Management

Data is distributed and replicated!

- Data needs to reach a processor to be computed.
- Processor memory is very small but data is usually large.

- Data is stored distributed and replicated in memory hierarchies.
- Data needs to be fetched, i.e., copied to a processor before it can be computed.
- Data needs to be flushed, i.e., copied to higher memory levels to become visible to other processors.



Rethinking Data Management

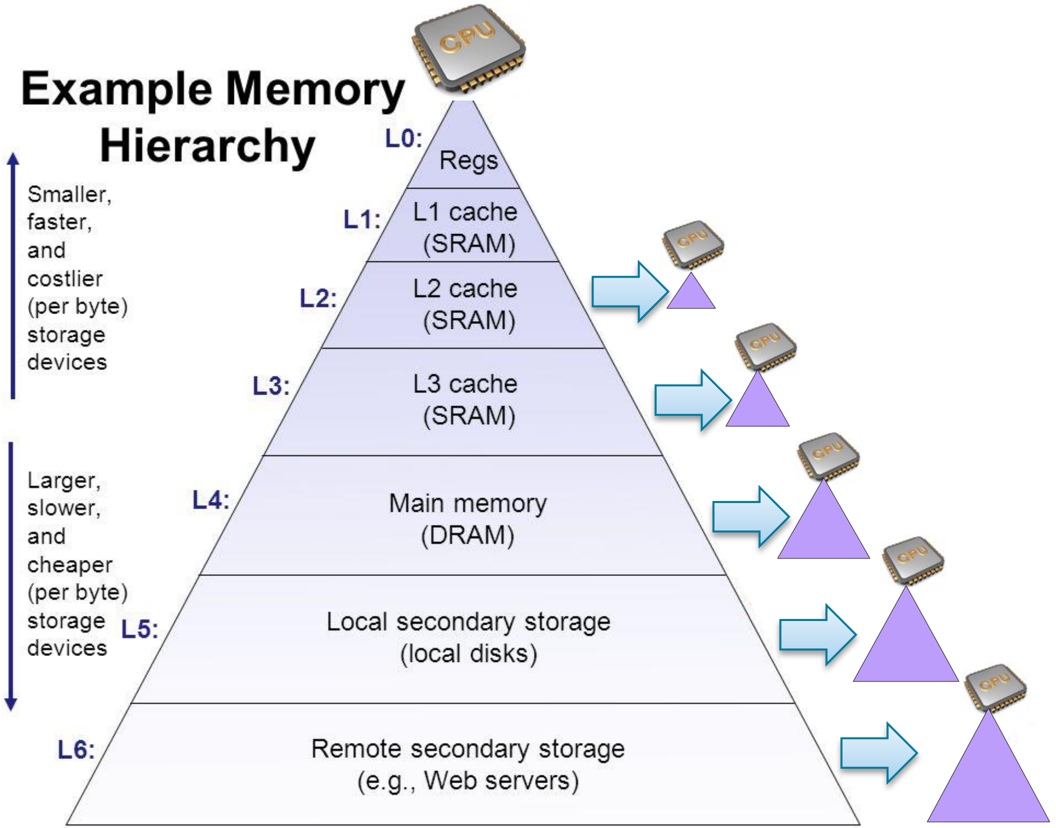
Moving data costs magnitudes more time and energy than computing data!

- Copying data costs time and energy.
- Stalled processors during data copying consume energy.

Operation	Operation Energy Cost (nJ)	Equivalent ADD
ADD	0.64	-
L1->REG	1.11	1.8x
L2->REG	2.21	3.5x
L3->REG	9.80	15.4x
MEM->REG	63.64	99.7x
Stall	1.43	-
Prefetching	65.08	-

https://hpc.pnl.gov//modsim/2014/Presentations/Kestor.pdf

➢ Push computation to the data not data to the computation.



Rethinking Data Management

Hasso Plattner Institut

Moving data costs magnitudes more time and energy than computing data!

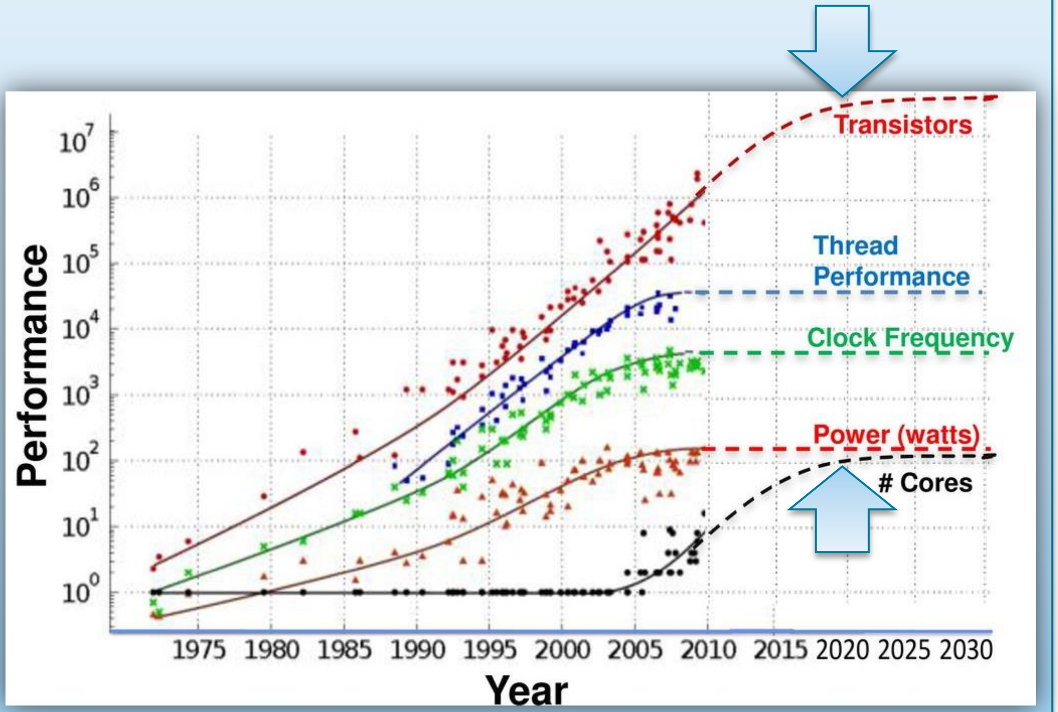
- Copying data costs time and energy.
- Stalled processors during data copying consume energy.

Operation	Operation Energy Cost (nJ)	Equivalent ADD
ADD	0.64	-
L1->REG	1.11	1.8x
L2->REG	2.21	3.5x
L3->REG	9.80	15.4x
MEM->REG	63.64	99.7x
Stall	1.43	-
Prefetching	65.08	-

https://hpc.pnl.gov//modsim/2014/Presentations/Kestor.pdf

➢ Push computation to the data not data to the computation.

Why energy is a concern:



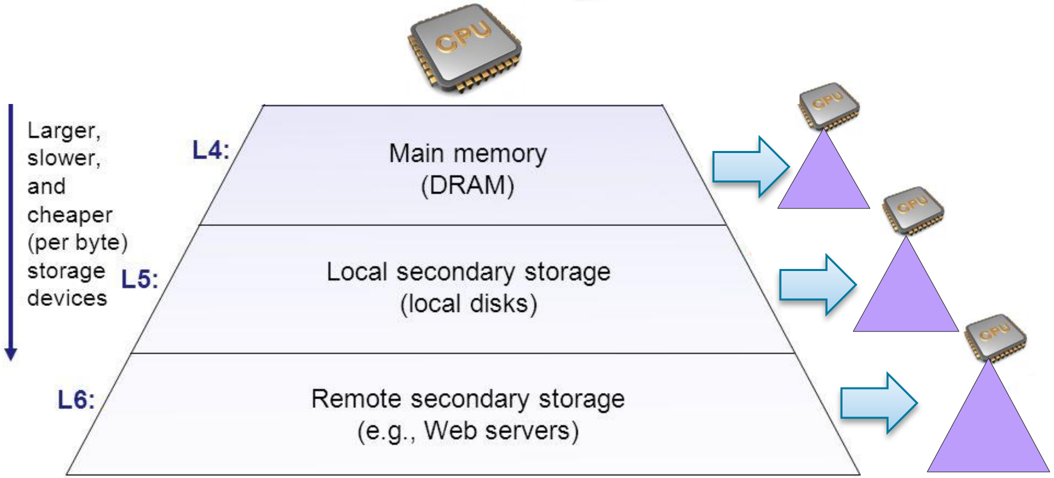
Motivation: "Management"

Rethinking Data Management

- Data engineers and data scientists need to be good data manager!
  - Data encoding
  - Data transmission
  - Data replication
  - Data partitioning
  - Data consistency management
  - Load scheduling
  - Load balancing

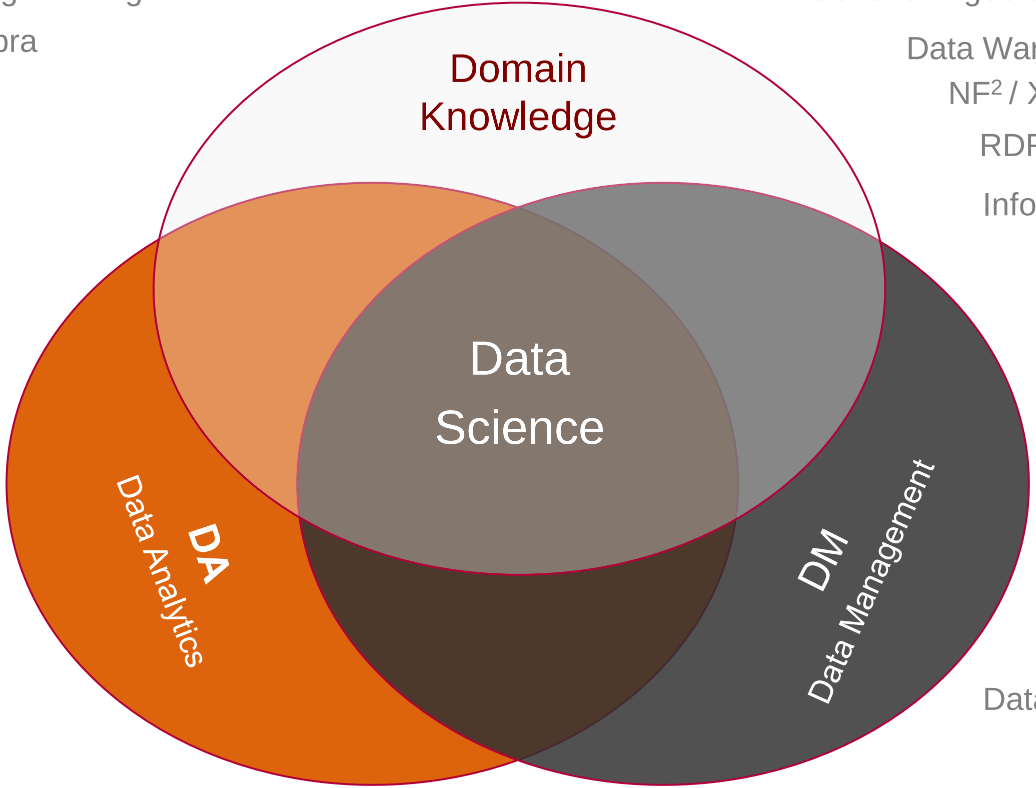
We do not consider L0-L3 in this lecture, but this is super relevant for High Performance Computing!

I recommend: https://www.youtube.com/watch?v=3PjNgRWmv90&list=LLbLaqsrSDDURdv_ZV75-AMQ&index=6&t=0s



Domain Expertise (e.g., Industry 4.0, Medicine, Physics, Engineering, Energy, Logistics)

Mathematical Programming
Linear Algebra
Statistics
Text Mining
Graph Mining
Signal Processing
Stochastic Gradient Descent
Machine Learning
Error Estimation
Active Sampling
Monte Carlo
Regression
Predictive Analytics
Sketches
Data Obfuscation
Convergence
Decoupling
Iterative Algorithms
Curse of Dimensionality



Relational Algebra / SQL
Data Warehouse/OLAP
NF^2 / XQuery
RDF / SparQL
Information Integration
Information Extraction
Visual Analytics
Privacy
Memory Management
Parallelization
Scalability
Memory Hierarchy
Fault Tolerance
Security
Data Analysis Languages
Query Optimization
Real-Time
Indexing

Legal Aspects
Business Models
Control Flow
Data Flow

Data Management

Data Analytics

"The ability to effectively extract and calculate various kinds of information from data!"

- Structural information
- Explicit information
- Implicit/derived information

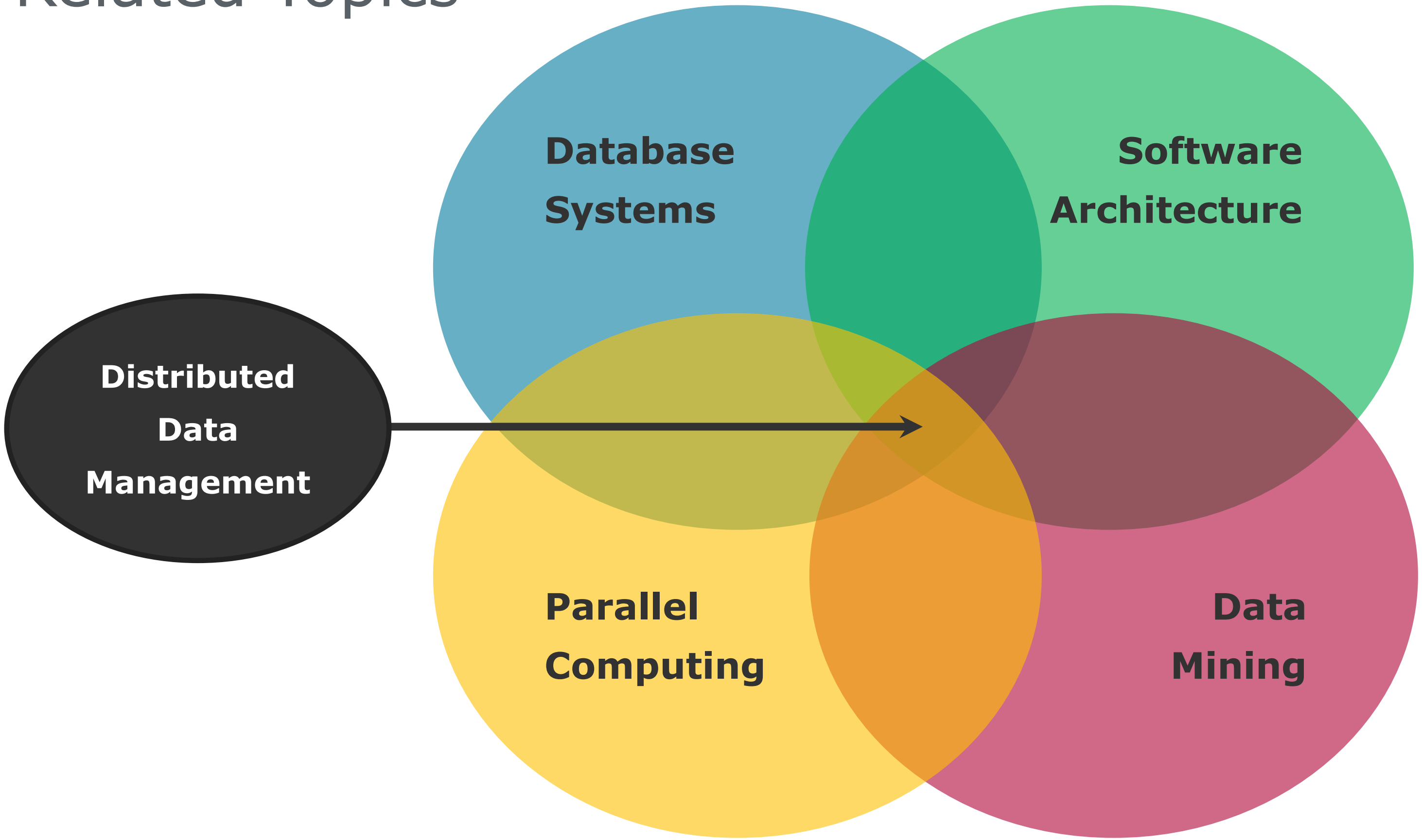
Data Management

"The ability to efficiently read, transform, and store large amounts of data!"

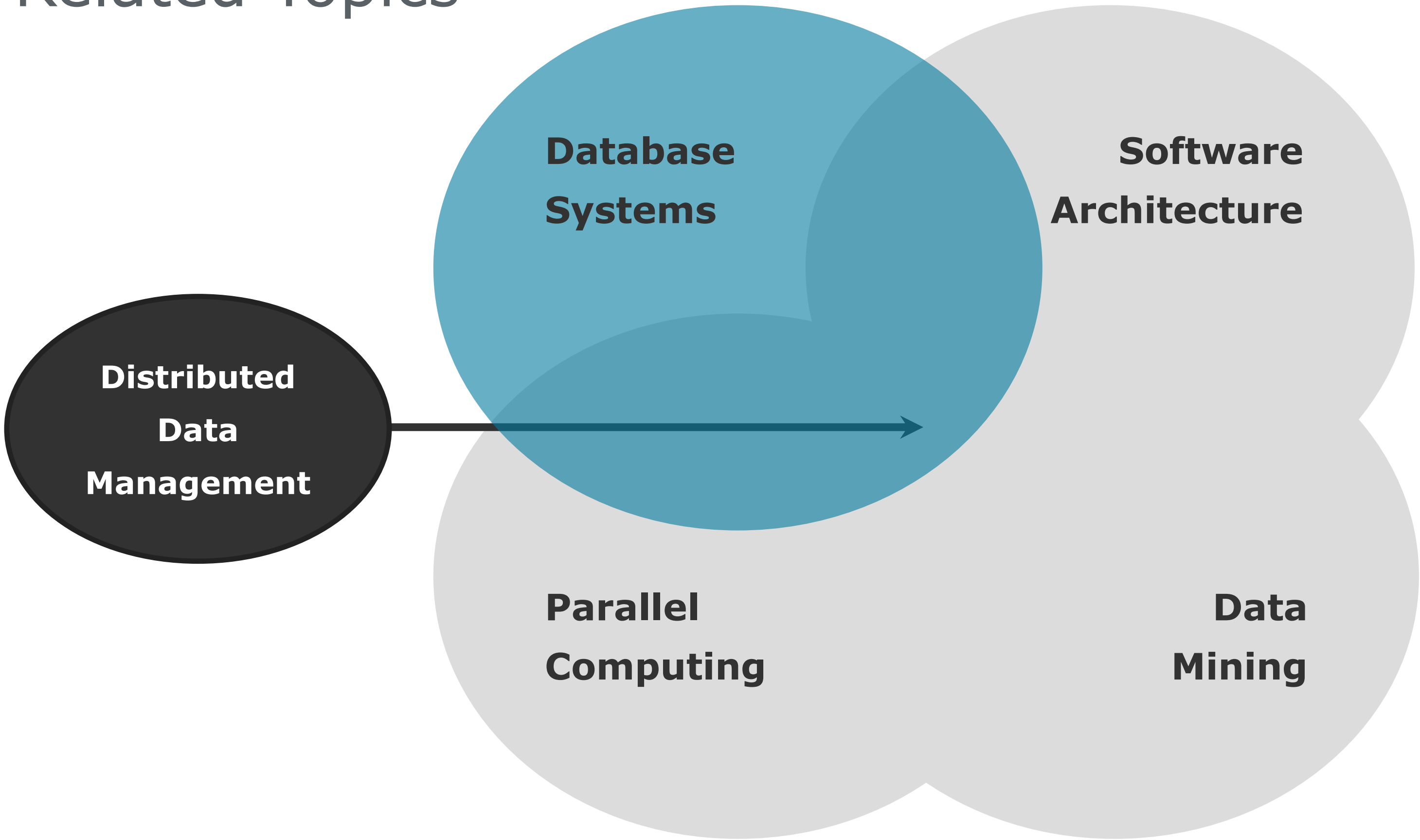
- Static (block) data
- Volatile (streaming) data

Motivation: "Management"

Related Topics



Related Topics



Motivation: “Management”

Database Systems

Touch points

- Data models, query languages, and consistency guarantees
- Distributed storage and retrieval of data
- Index structures

Not in this lecture

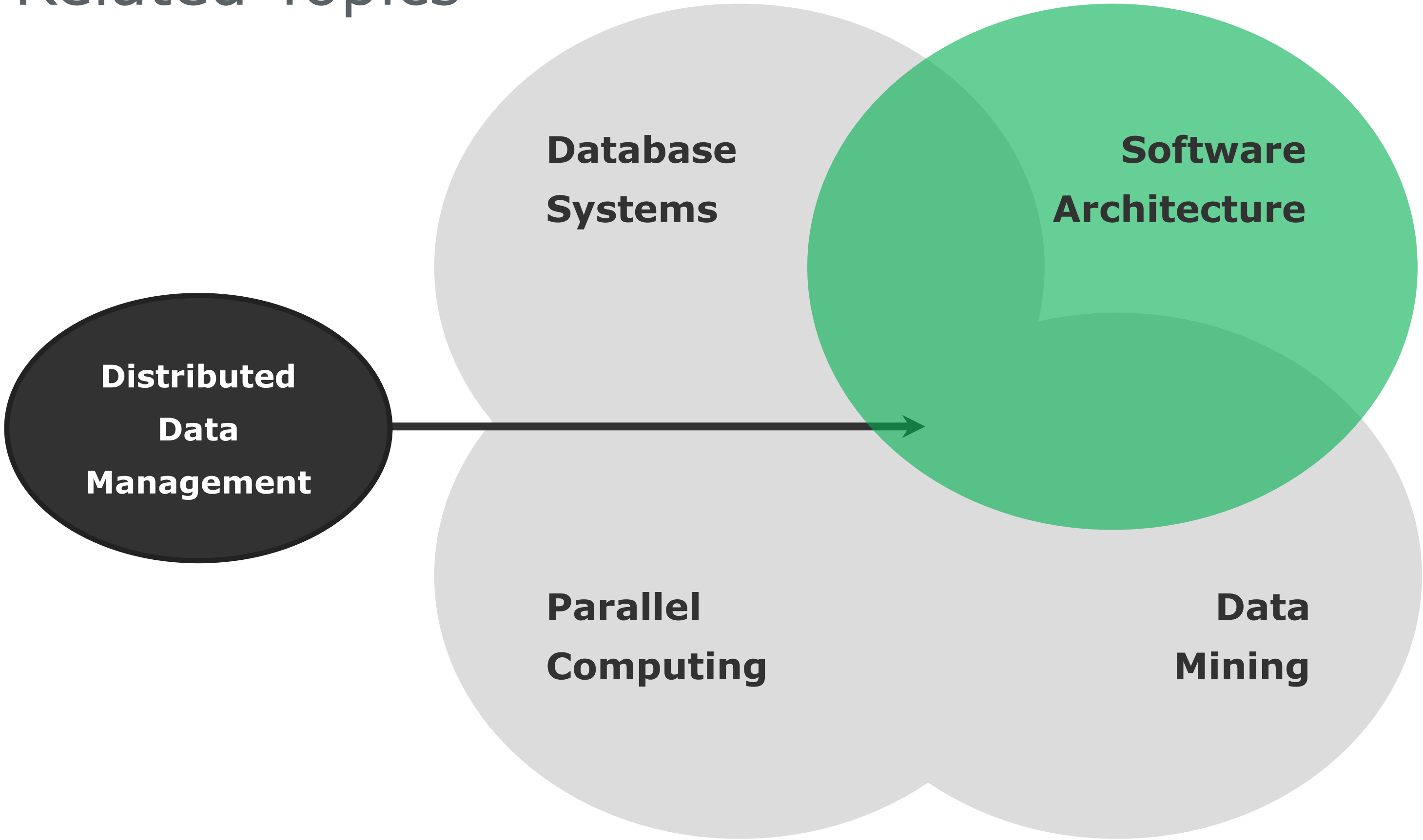
- Physical data storage
- Foundations on transaction management and logging
- Core database technology, e.g., query optimizer

More focused lectures

- Database Systems I + II (Prof. Naumann)
- Trends and Concepts in Software Industry (Prof. Plattner)

Motivation: "Management"

Related Topics



Motivation: "Management"

Software Architectures

Touch points

- Requirements, design, and architecture of distributed systems
- Pros and cons of different technologies for distributed systems

Not in this lecture

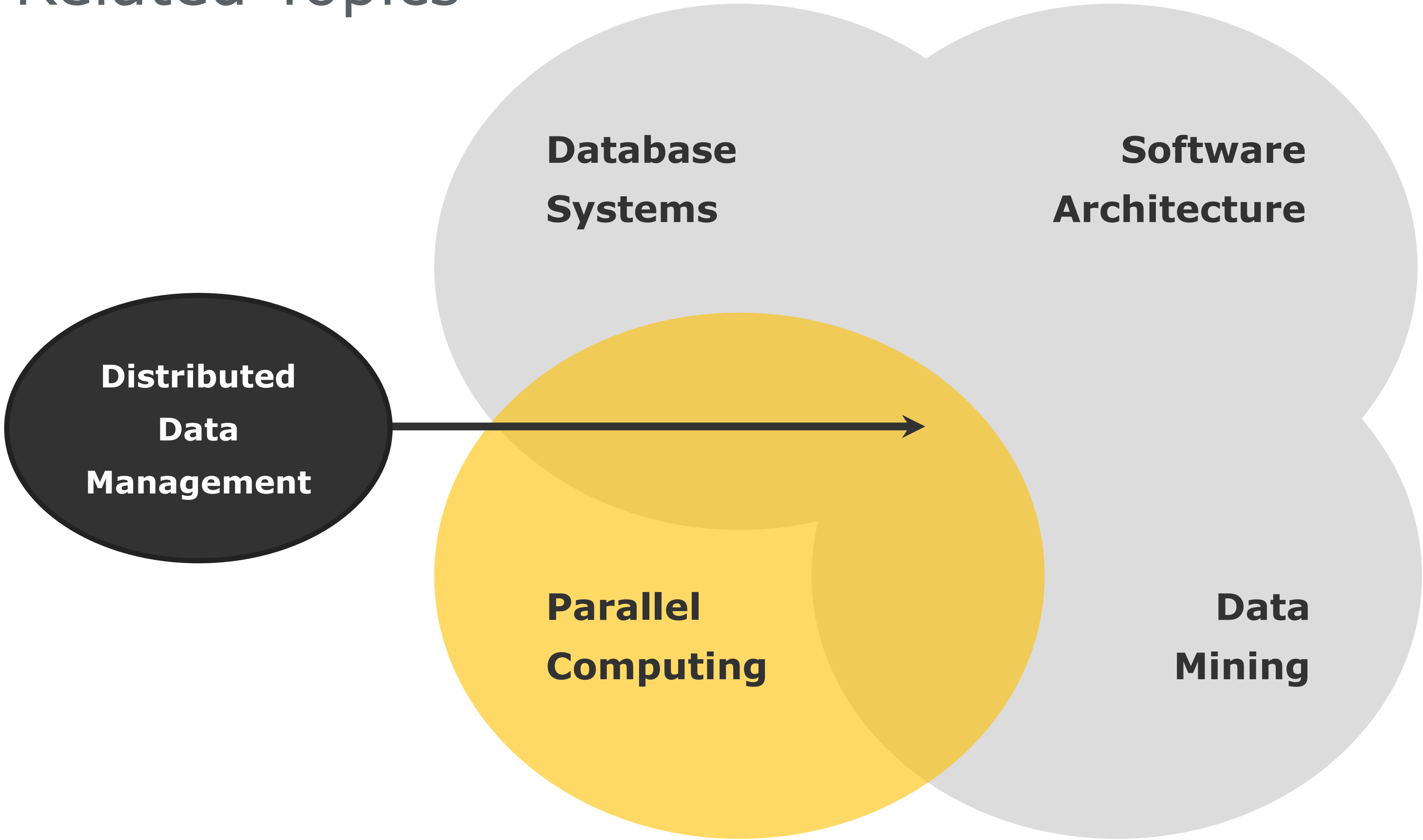
- Non-distributed systems
- Agile software development techniques
- Software patterns

More focused lectures

- Software Architecture (Dr. Uflacker)
- Software Technique (Dr. Uflacker)

Motivation: "Management"

Related Topics



Motivation: "Management"

Parallel Computing

Touch points

- Distributed data storage concepts
- Distributed programming models, e.g., actor programming and MapReduce

Not in this lecture

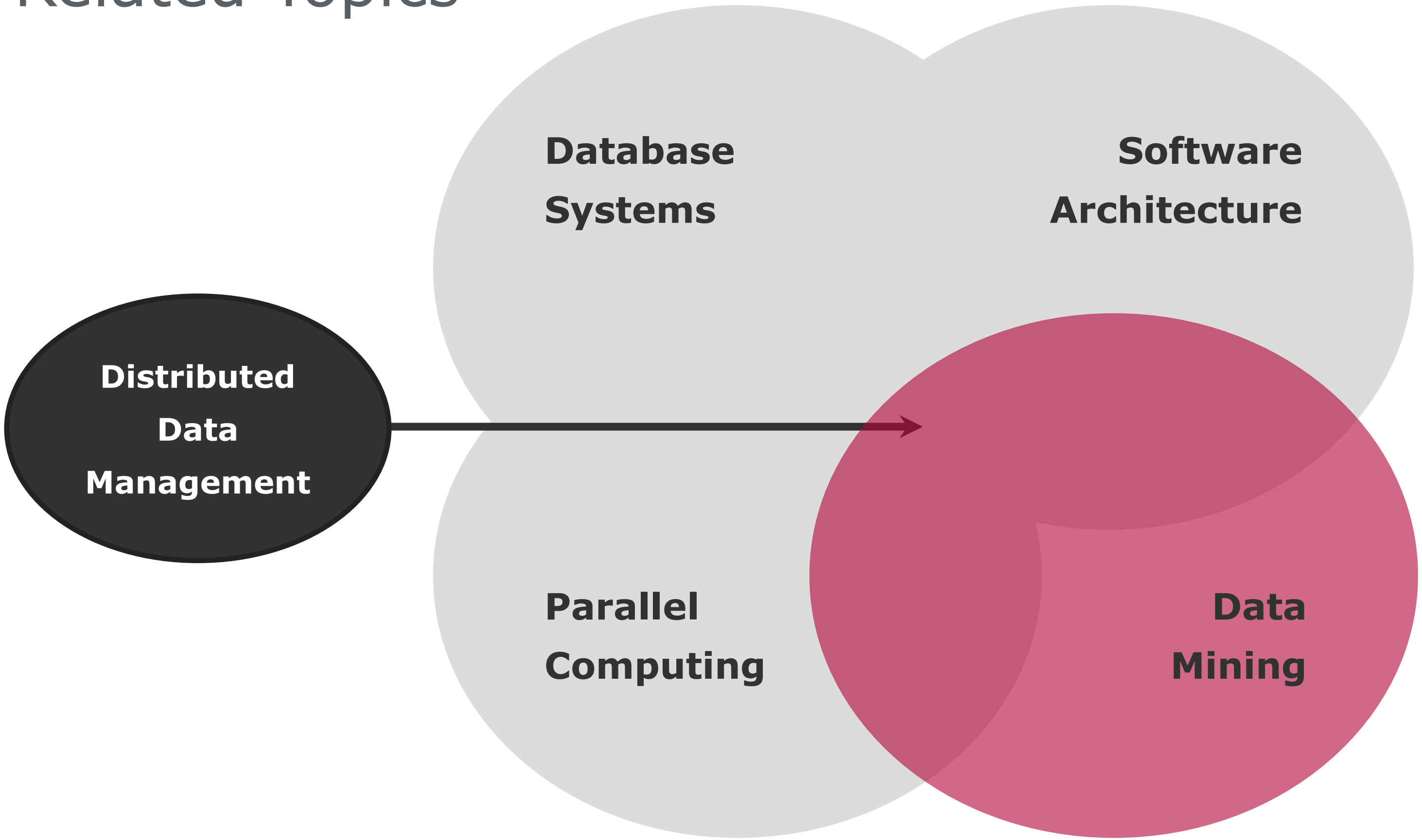
- Parallel, non-distributed programming languages, e.g., CUDA or OpenMP
- Core parallel computing concepts, e.g., scheduling or shared memory
- Processor architectures, cache hierarchies, GPU programming, …

More focused lectures

- Parallel Programming (Dr. Tröger)
- Programmierung paralleler und verteilter Systeme (Dr. Feinbube)

Motivation: "Management"

Related Topics



Motivation: "Management"

Data Mining

Touch points

- Data analytics: aggregation queries and basic data mining algorithms

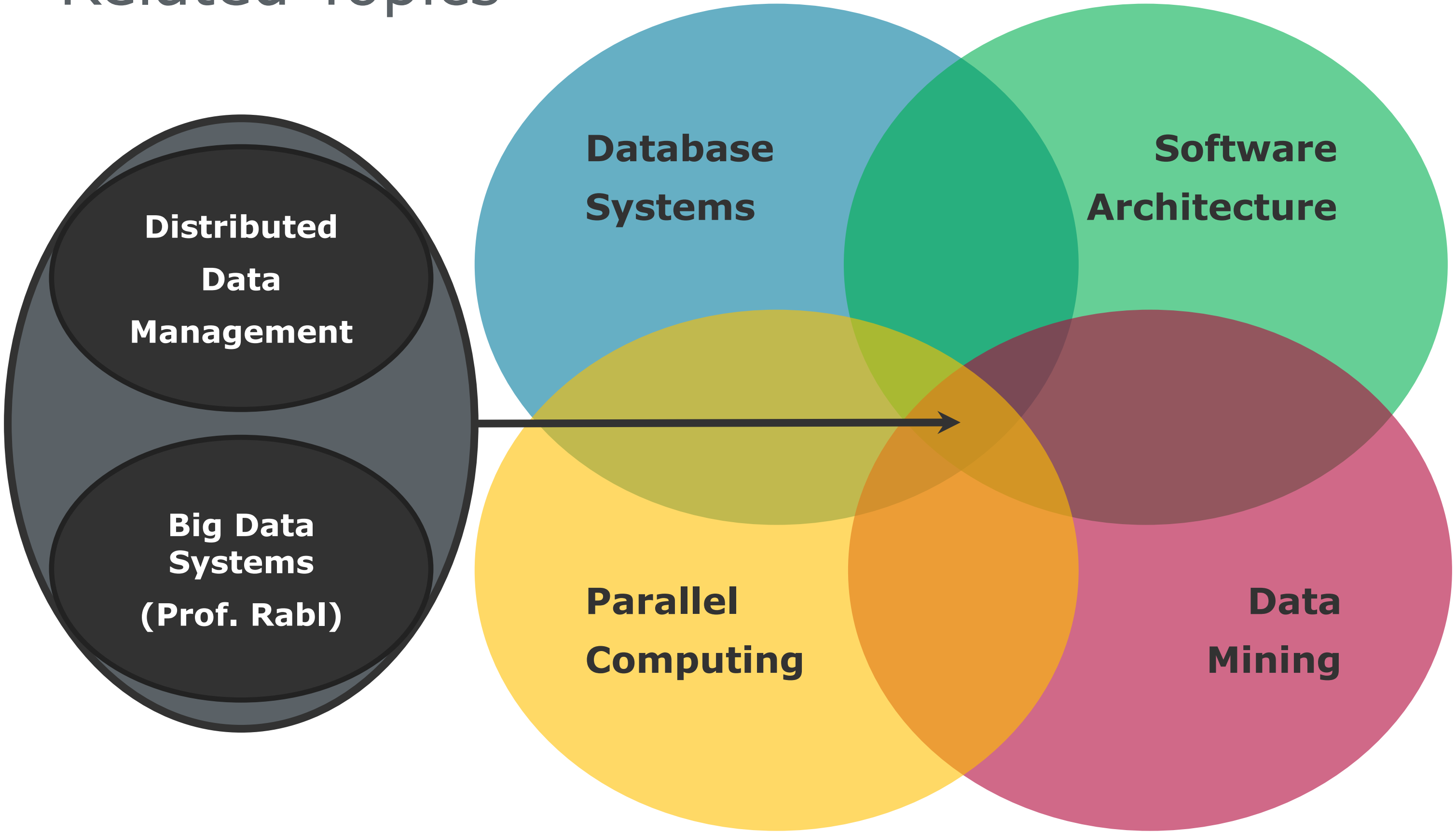
Not in this lecture

- Detailed introduction to machine learning, e.g., neuronal networks, (un)supervised learning, or Bayesian classification
- Statistics, linear algebra, and most sophisticated mining algorithms

More focused lectures/seminars

- Data Analysis in R (Lippert, Konigorski, Schurmann)
- Selected Topics in Data Analytics (Döllner, Hagedorn)
- Machine Learning for Data Steams (Albrecht)
- Neuro Design (Von Thienen)

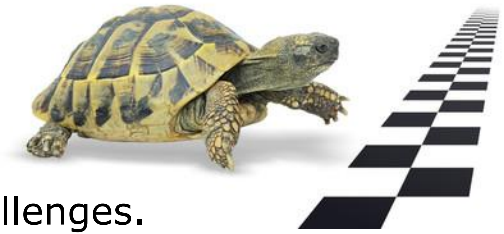
Related Topics



Lecture Goals

Sorting the buzzwords

- NoSQL, Big Data, OLAP, Web-scale, ACID, Sharding, MapReduce, Scale-out…



Understanding distributed systems

- You know how state-of-the-art distributed systems work.
- You know core technologies and techniques to solve distributed challenges.
- You know the advantages and disadvantages of important systems.
- You know how to handle data in distributed settungs.

Exercising in distributed data management and analytics

- You can implement distributed algorithms and applications.
- You can solve problems that arise in distributed setups.
- You can write data-parallel and task-parallel jobs.

“Dark Magic”

- With distributed computing we can utilize incredible amounts of compute power!

  - At the cost of harder programming (e.g. fault tolerance, testing and protocols)

  - At the cost of additional energy (e.g. communication and redundancy)

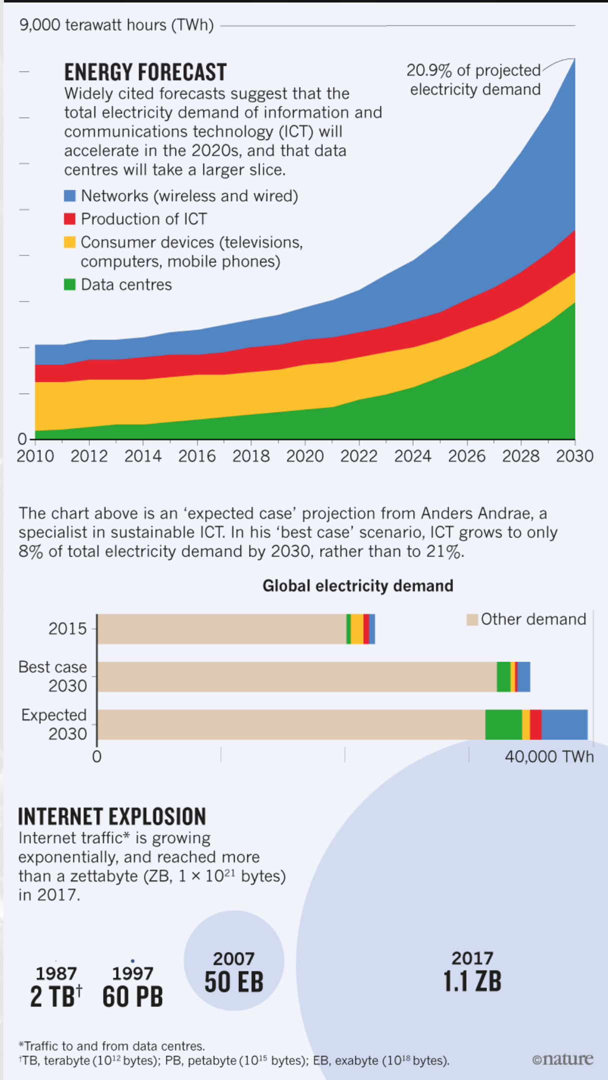
➢ Efficient, fault resistant code matters all the more, because inefficiency and failures scale, too!

“Dark Magic”

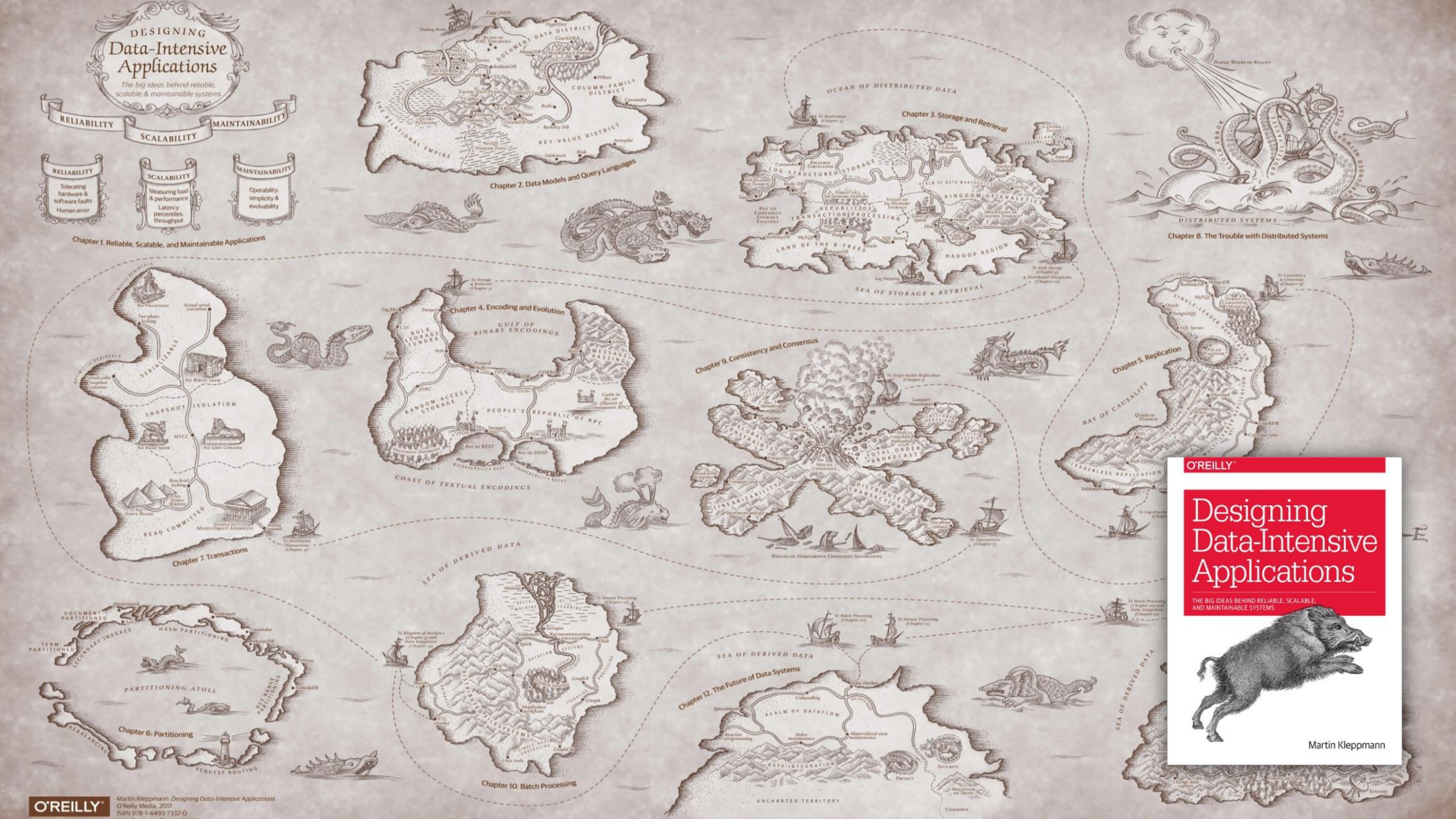
- “Around 10% of the world’s total electricity consumption is being used by the internet.”
  Swedish KTH
  https://www.insidescandinavianbusiness.com/article.php?id=356
  https://www.sciencedirect.com/science/article/pii/S2214629618301051

- “The Internet’s data centers alone may already have the same CO_2 footprint as global air travel.”
  Global e-Sustainability Initiative
  https://internethealthreport.org/2018/the-internet-uses-more-electricity-than/

- “Data centres […] consume about 3% of the global electricity supply […] accounting for about 2% of total greenhouse gas emissions” in 2016.
  Independent
  https://www.independent.co.uk/environment/global-warming-data-centres-to-consume-three-times-as-much-energy-in-next-decade-experts-warn-a6830086.html



https://www.nature.com/articles/d41586-018-06610-y


DESIGNING
Data-Intensive Applications
The big ideas behind reliable, scalable & maintainable systems
RELIABILITY
SCALABILITY
MAINTAINABILITY
RELIABILITY
Tolerating hardware & software faults
Human error
SCALABILITY
Measuring load & performance
Latency percentiles, throughput
MAINTAINABILITY
Operability, simplicity & evolvability
Chapter 1. Reliable, Scalable, and Maintainable Applications
Chapter 2. Data Models and Query Languages
DOCUMENT DATA DISTRICT
COLUMN-FAMILY DISTRICT
KEY-VALUE DISTRICT
THE RELATIONAL EMPIRE
NoSQL DANGER ZONE
OCEAN OF DISTRIBUTED DATA
Chapter 3. Storage and Retrieval
LOG-STRUCTURED STORAGE
REPUBLIC OF TRANSACTION PROCESSING
KINGDOM OF ANALYTICS
LAND OF THE B-TREES
HADOOP REGION
SEA OF STORAGE & RETRIEVAL
DISTRIBUTED SYSTEMS
Chapter 8. The Trouble with Distributed Systems
Chapter 4. Encoding and Evolution
GULF OF BINARY ENCODINGS
BULK STORAGE TUNDRA
MESSAGE PASSING
RANDOM-ACCESS STORAGE
PEOPLE'S REPUBLIC OF RPC
BAY OF REST
BAY OF SOAP
COAST OF TEXTUAL ENCODINGS
SERIALIZABLE
SNAPSHOT ISOLATION
READ COMMITTED
Chapter 7. Transactions
Chapter 9. Consistency and Consensus
LINEARIZABILITY
TOTAL ORDER BROADCAST
DISTRIBUTED TRANSACTIONS
Wrecks of Homegrown Consensus Algorithms
Chapter 5. Replication
SINGLE-LEADER REPLICATION
BAY OF CAUSALITY
LEADERLESS REPLICATION
SEA OF DERIVED DATA
PARTITIONING ATOLL
HASH PARTITIONING
SECONDARY INDEXES
REBALANCING
REQUEST ROUTING
Chapter 6: Partitioning
DISTRIBUTED FILESYSTEMS
Chapter 10. Batch Processing
SEA OF DERIVED DATA
Chapter 12. The Future of Data Systems
REALM OF DATAFLOW
DATA INTEGRATION
UNCHARTED TERRITORY
O'REILLY
Martin Kleppmann: *Designing Data-Intensive Applications*
O'Reilly Media, 2017
ISBN 978-1-4493-7332-0
O'REILLY
Designing Data-Intensive Applications
THE BIG IDEAS BEHIND RELIABLE, SCALABLE, AND MAINTAINABLE SYSTEMS
Martin Kleppmann