

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 decreases throughput
- MAINTAINABILITY**
Operability simplicity & evolvability

Chapter 1. Reliable, Scalable, and Maintainable Applications



Chapter 2. Data Models and Query Languages



Chapter 3. Storage and Retrieval



Chapter 8. The Trouble with Distributed Systems



Chapter 7. Transactions



Chapter 4. Encoding and Evolution



Chapter 9. Consistency and Consensus



Chapter 5. Replication

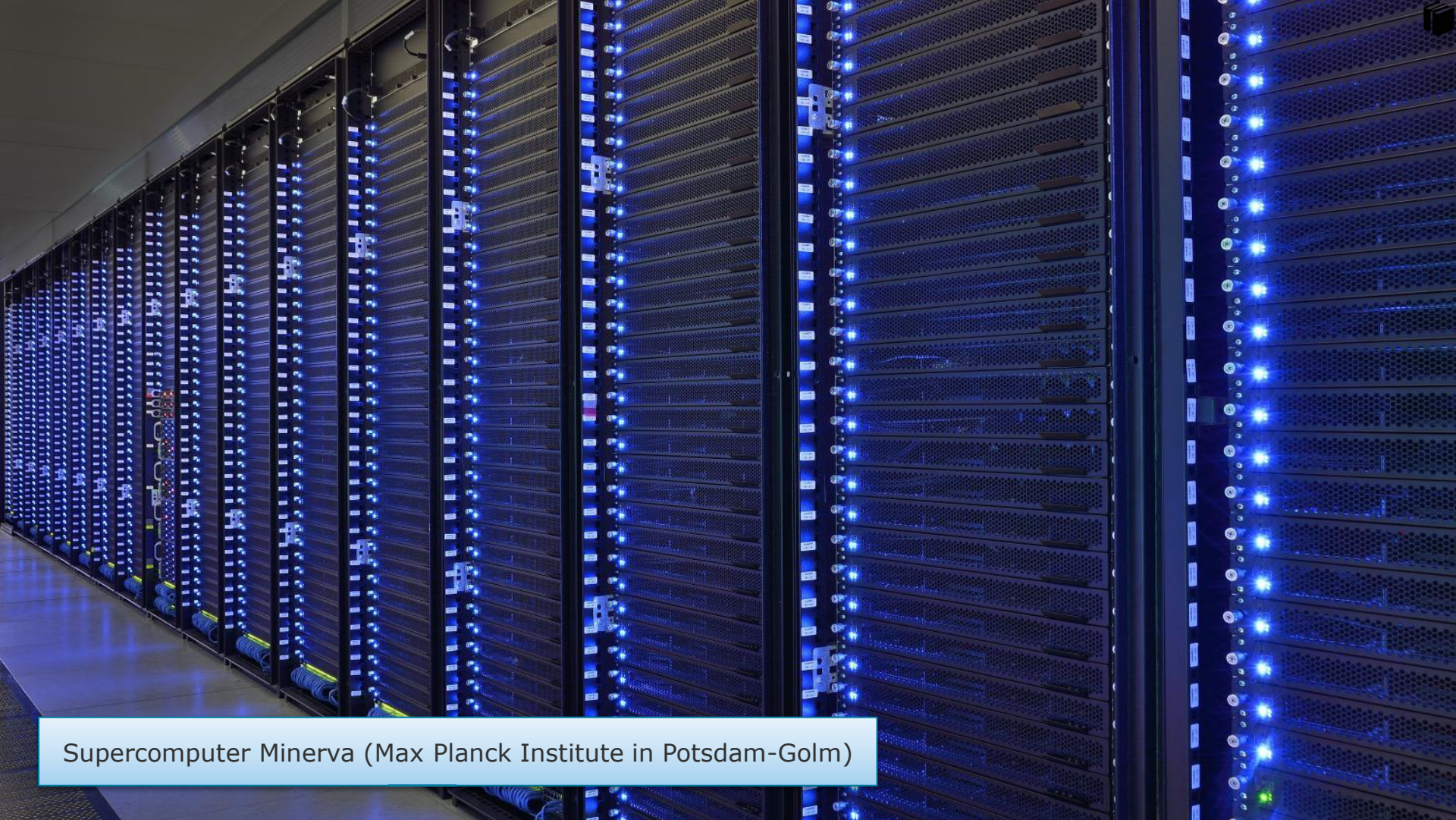


Distributed Data Management Introduction

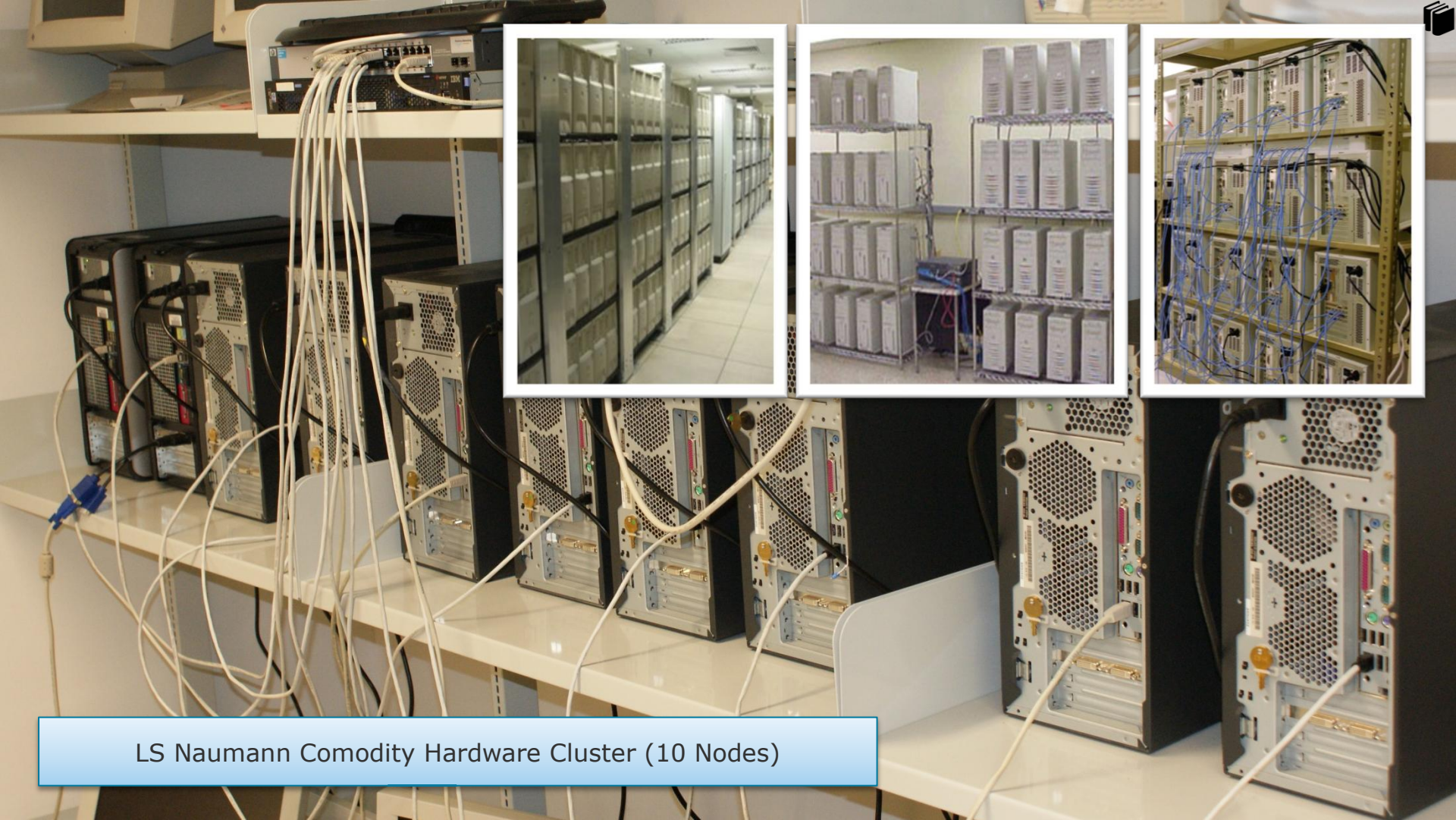
Thorsten Papenbrock

F-2.04, Campus II

Hasso Plattner Institut



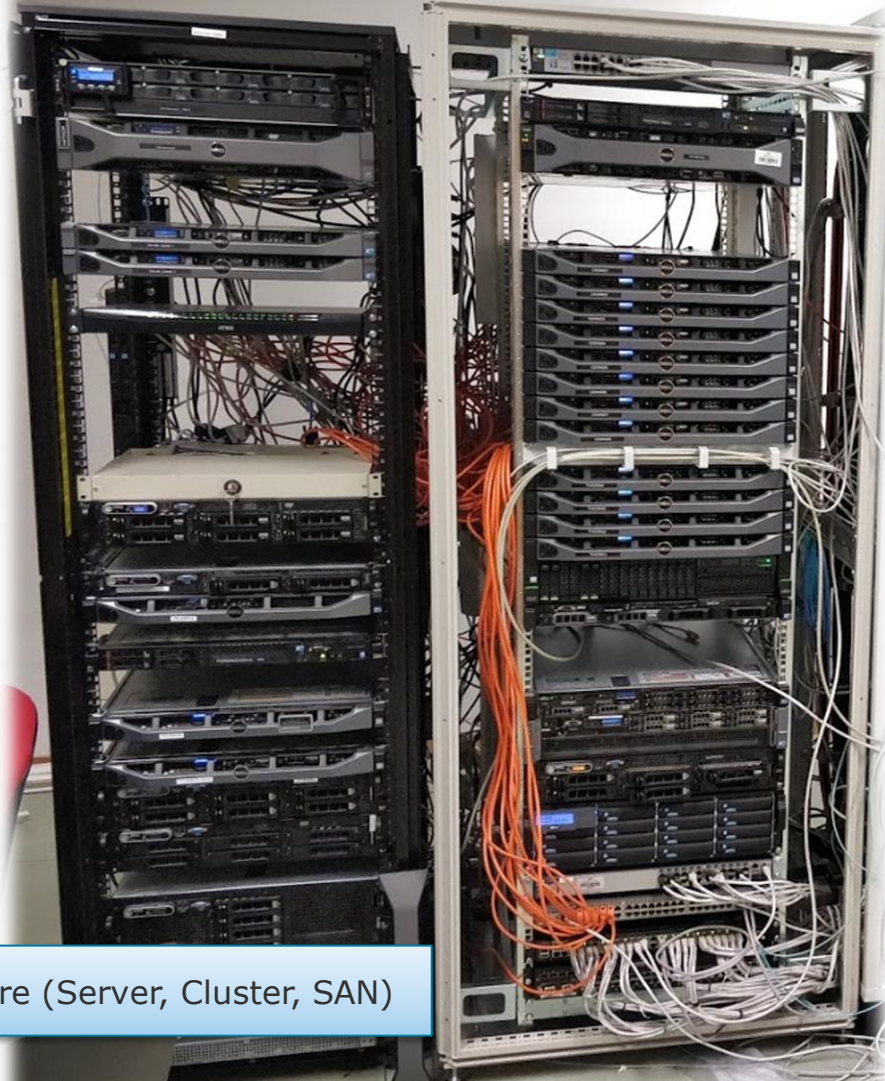
Supercomputer Minerva (Max Planck Institute in Potsdam-Golm)



LS Naumann Comodity Hardware Cluster (10 Nodes)

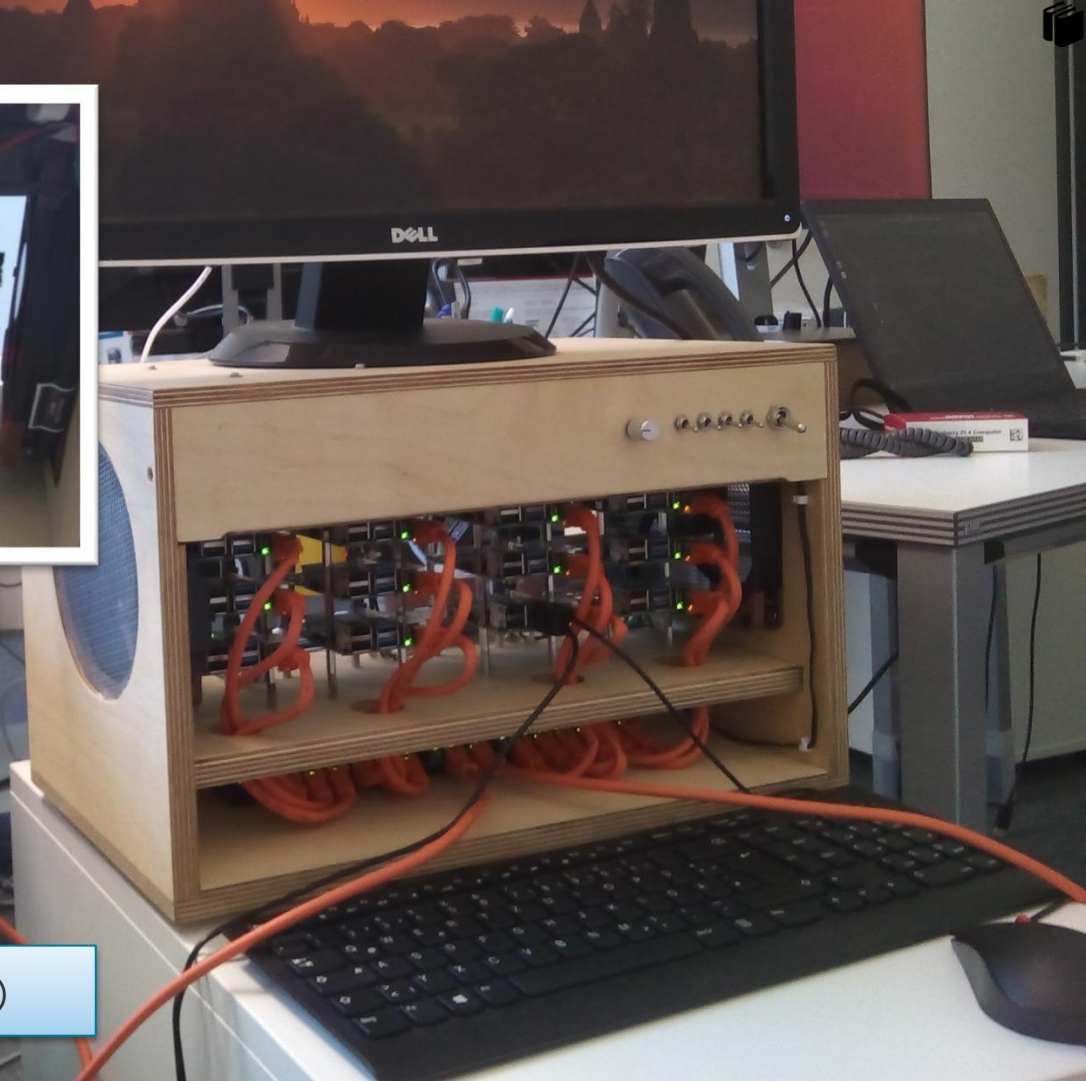


Desktop Computer (multiple CPUs and GPUs)



LS Naumann Infrastructure (Server, Cluster, SAN)





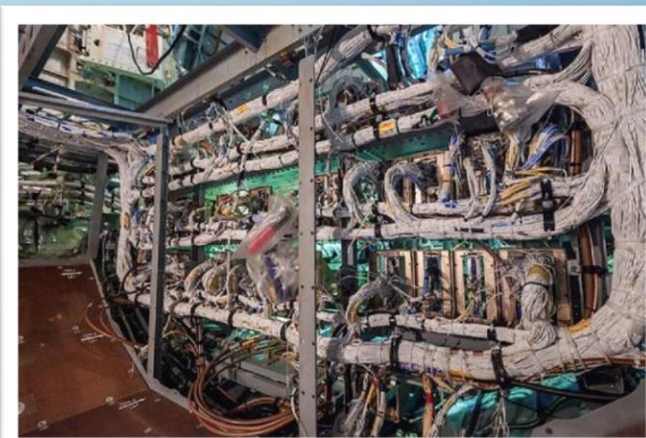
LS Naumann PI Cluster (12 Raspberry PI 4)

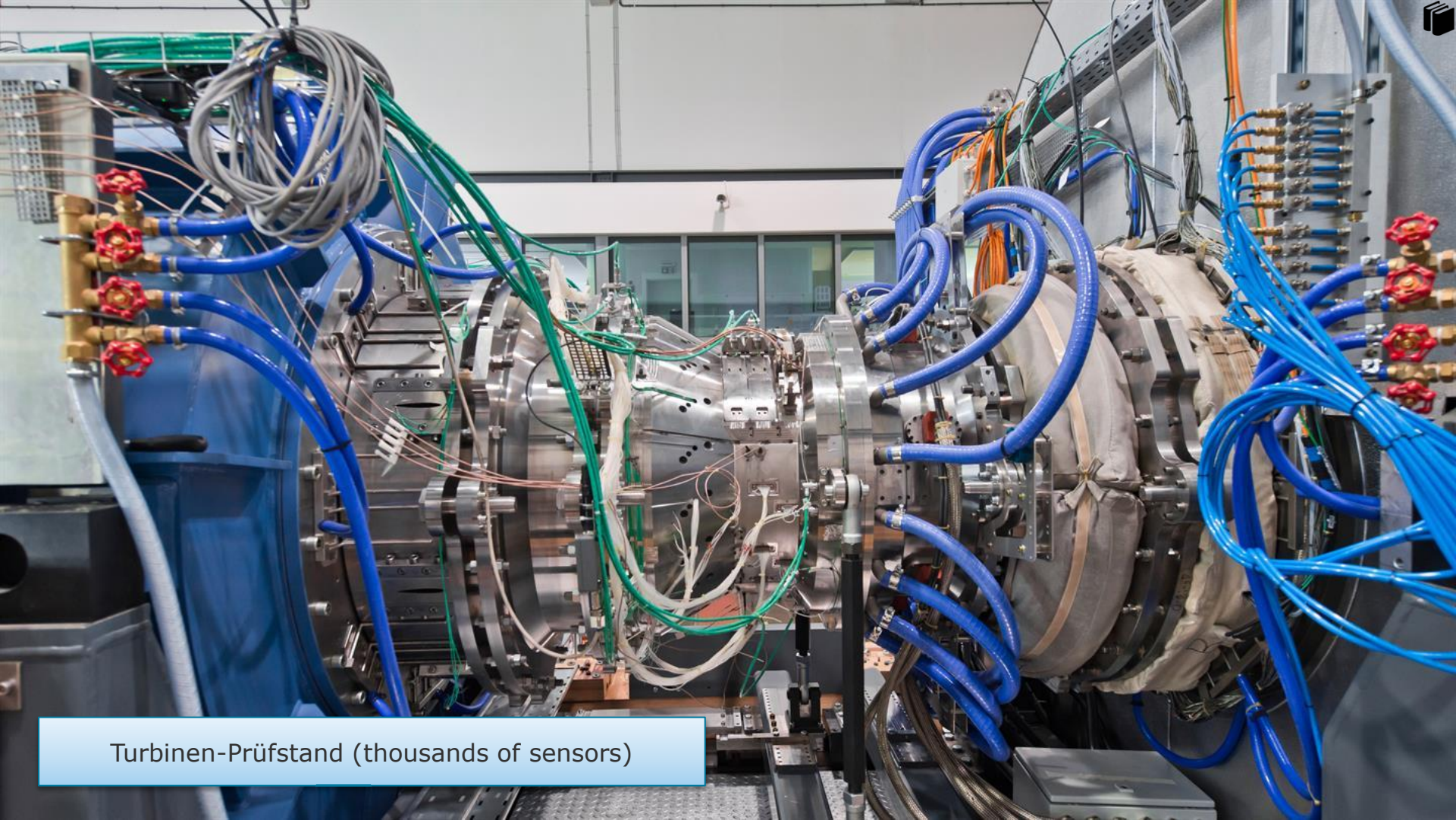


DreamHack (12,000-computer LAN party)



Boeing 747 (thousands of computers)





Turbinen-Prüfstand (thousands of sensors)

The logo for Startpage.com, featuring the word "Startpage" in a dark blue serif font with a blue underline under the "a", followed by ".com" in a smaller, dark blue sans-serif font.

Startpage.com

A white search bar with rounded ends and a thin grey border. On the right side, there is a blue circular button containing a white magnifying glass icon.

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

Startpage (search engine backed by other search engines)

A photograph of a group of people in a meeting. A woman with long blonde hair is on the left, looking towards the center. A man with glasses is in the foreground, seen from the back. A woman on the right is holding a yellow pencil and gesturing. A laptop is on the table. Overlaid on the image are five white arrows with blue outlines and text: 'Outdated Data' (top left), 'Lost & Invalid Messages' (middle right), 'Concurrency' (bottom left), 'Consensus' (center), and 'Termination' (bottom right).

Outdated Data

Lost & Invalid Messages

Concurrency

Consensus

Termination



Dr. Thorsten Papenbrock



Diana Stephan

Service-Oriented Systems



Prof. Felix Naumann

project DuDe
Data Fusion



Dr. Ralf Krestel

Data Scrubbing



Tim Repke

Entity Search



Nitisha Jain



Phillip Wenig

Distributed Computing

project Metanome

Agile Systems

project AKITA

Anomaly Detection

Information Integration

Duplicate Detection

Data Profiling

Dependency Detection

Data Change

project DataChEx

Change Exploration



Sebastian Schmidl



Dr. Hazar Harmouch



Tobias Bleifuß



Leon Bornemann



Lan Jiang

project Stratosphere

Data as a Service

Data Cleansing

Entity Recognition

Opinion Mining

RDF Data Mining

ETL Management

Data Preparation

Web Science

Web Data

Text Mining



Michael Loster



Gerardo Vitagliano





English?

ITSE, DE, DH?

Which semester?

HPI or Guest?

Database knowledge?

Distributed experience?

Other related lectures?

Distributed Data Management

Courses 2021 - Information Systems Group

- > **Datenbanksysteme I** (VL, Bachelor)
- > **Einführung in die Programmier technik II** (VL, Bachelor)
- > **Distributed Data Management** (VL, Master)
- > **Methoden der Forschung** (SE, Master)
- > **Table Recognition** (PS, Master)
- > **Building Machine Learning Applications** (PS, Master)
- > **Knowledge Graphs** (SE, Master)
- > **UltraMine**: Skalierbare Analyse von Messdatenströmen (Bachelorprojekt)
- > **Data Matching Benchmark** (Bachelorprojekt)

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

**Distributed Data
Management**

Introduction

Thorsten Papenbrock
Slide **15**

Distributed Data Management

This Lecture

Lecture

- For master students
(IT-Systems Engineering,
Digital Health, Data Engineering)
- 6 credit points, 4 SWS
- Mondays 11:00 – 12:30
Wednesdays 15:15 – 16:45

Exercises

- Interleaved with lectures

Slides

- On website

Website

- <https://hpi.de/naumann/teaching/current-courses/ss-21/distributed-data-management.html>

Prerequisites

- To participate:
Interest and a little background 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

Question at any time please!

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

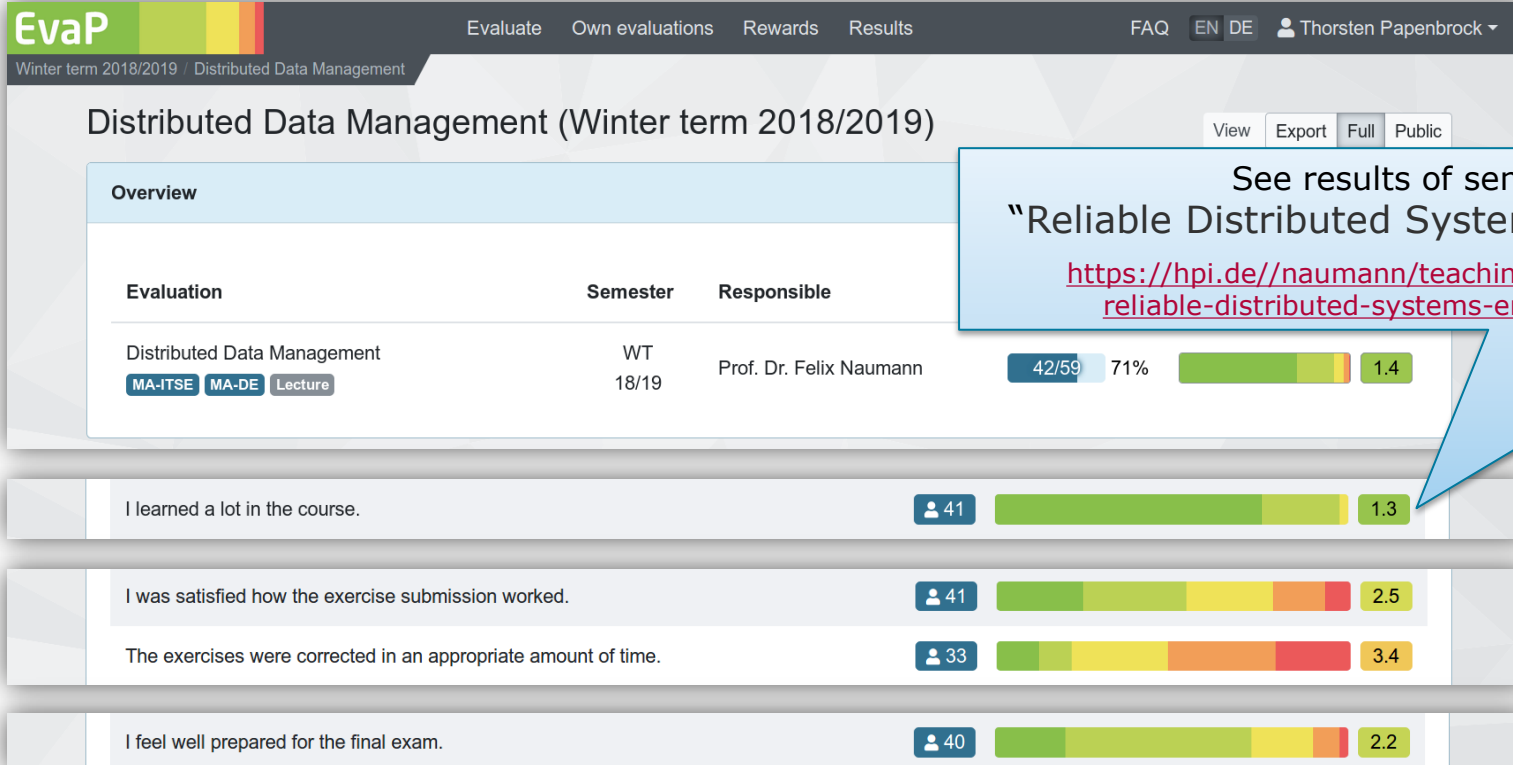
Also: Give feedback about ...

- improving lectures
- informational material
- organization

Official evaluation

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

Distributed Data Management Feedback



See results of seminar
"Reliable Distributed Systems Engineering"

[https://hpi.de//naumann/teaching/teaching/ss-19/
reliable-distributed-systems-engineering.html](https://hpi.de//naumann/teaching/teaching/ss-19/reliable-distributed-systems-engineering.html)

**Distributed Data
Management**

Introduction

ThorstenPapenbrock
Slide **18**

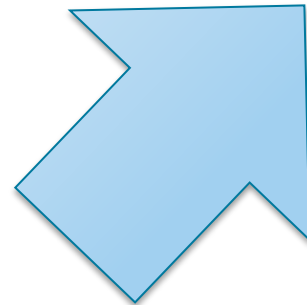


How could the course be further improved?

- 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 via 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 yourselves rechtzeitig zusenden
 - Übung frühzeitig kontrollieren
 - wenn man eine Übung vorstellen soll, bescheid geben, damit man auch anwesend 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 dem 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



Distributed Data Management

Introduction

ThorstenPapenbrock
Slide 19

Distributed Data Management

Lecture Outline

1. Introduction
2. Foundations
3. Encoding
4. Communication
5. Hands-On: Akka 
6. Data Models and Query Languages
7. Storage and Retrieval
8. Replication
9. Partitioning
10. Distributed Systems
11. Consistency and Consensus
12. Transactions
13. Batch Processing
14. Hands-On: Spark 
15. Stream Processing
16. Distributed DBMSs
17. Distributed Query Optimization
18. Lecture Summary and Exam Preparation

Distributed Data Management

Introduction

Distributed Data Management

Lecture Outline

1. Introduction
2. Foundations
3. Encoding
4. Communication
5. Hands-On: Akka
6. Data Models and Query Languages
7. Storage and Retrieval
8. Replication
9. Partitioning



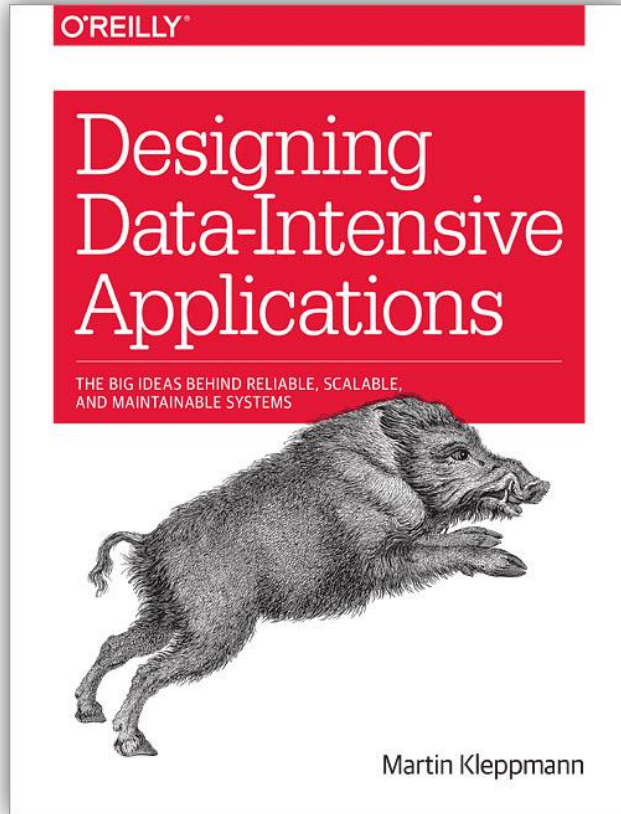
10. Distributed Systems
11. Consistency and Consensus
12. Transactions
13. Batch Processing
14. Hands-On: Spark
15. Stream Processing
16. Distributed DBMSs
17. Distributed Query Optimization
18. Lecture Summary and Exam Preparation



Distributed Data Management

Introduction

ThorstenPapenbrock
Slide **21**



Designing Data-Intensive Applications

- Author: Martin Kleppmann
- 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

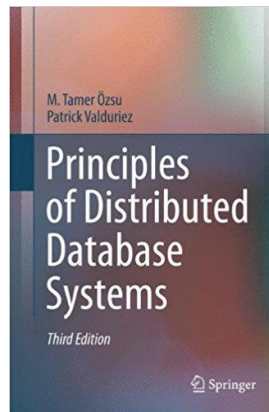
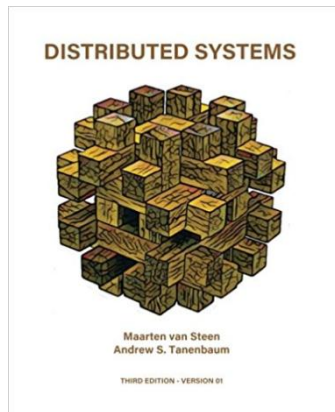
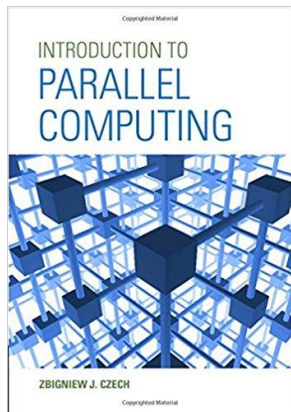
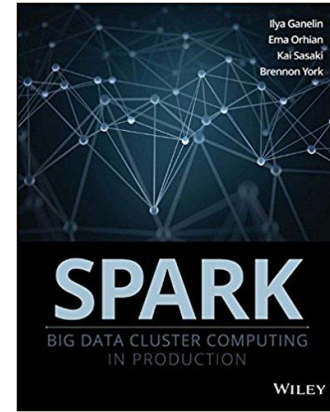
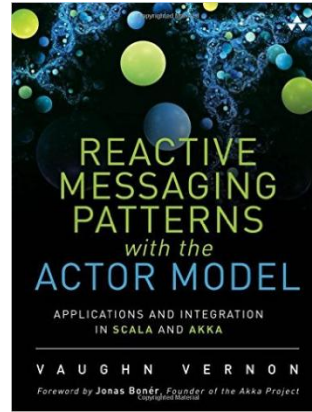
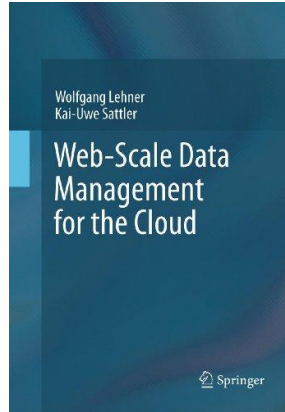
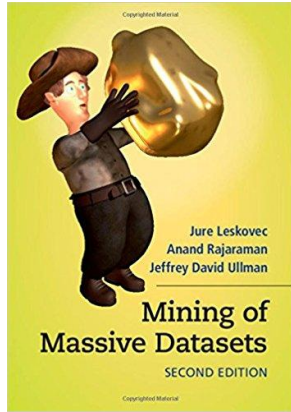
- Distributed and parallel systems
- Big data storage
- Batch and stream processing

Distributed Data Management

Introduction

Thorsten Papenbrock
Slide **22**

Distributed Data Management Literature: Further Reading



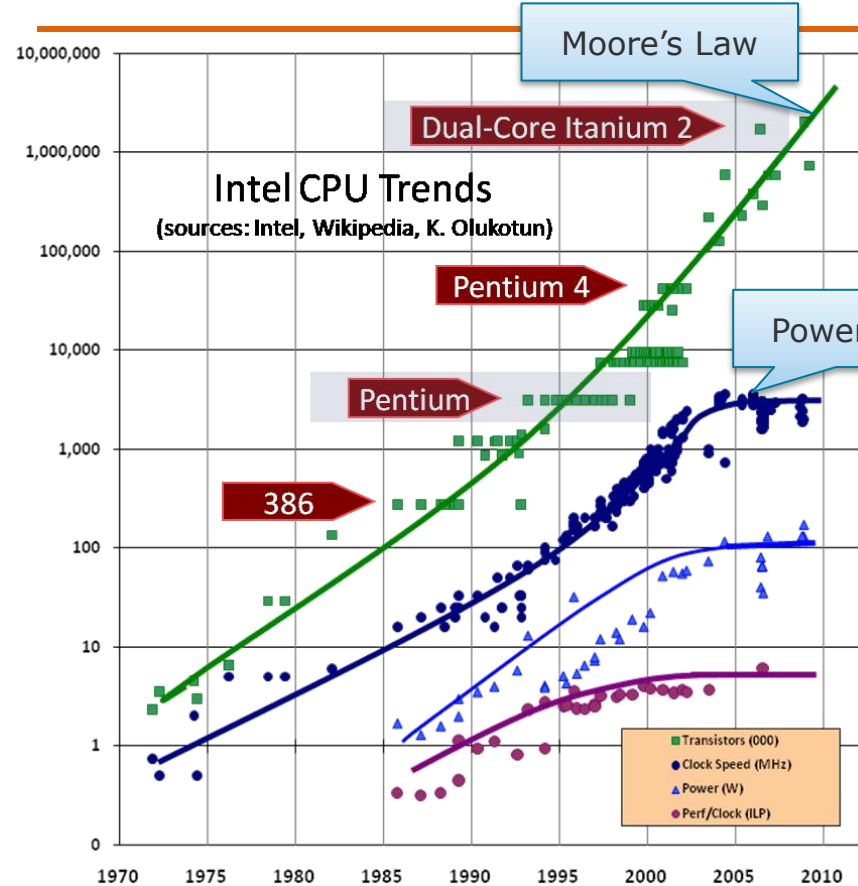
And Web-links that are given on the slides during the lecture.

Distributed Data Management

Introduction

ThorstenPapenbrock
Slide **23**

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

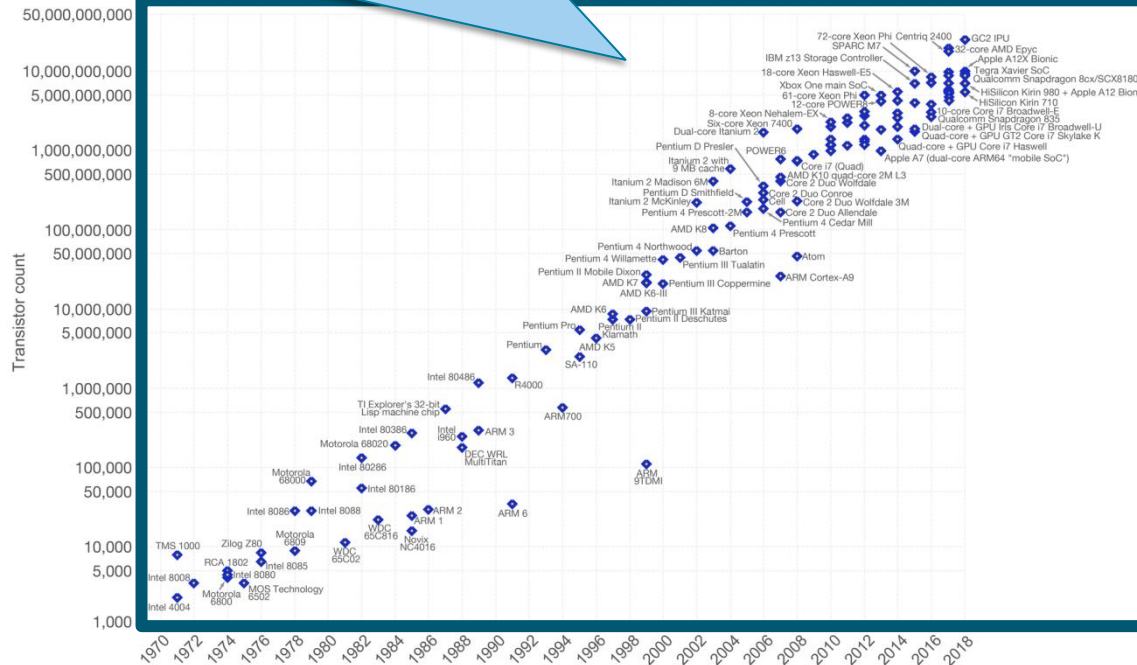
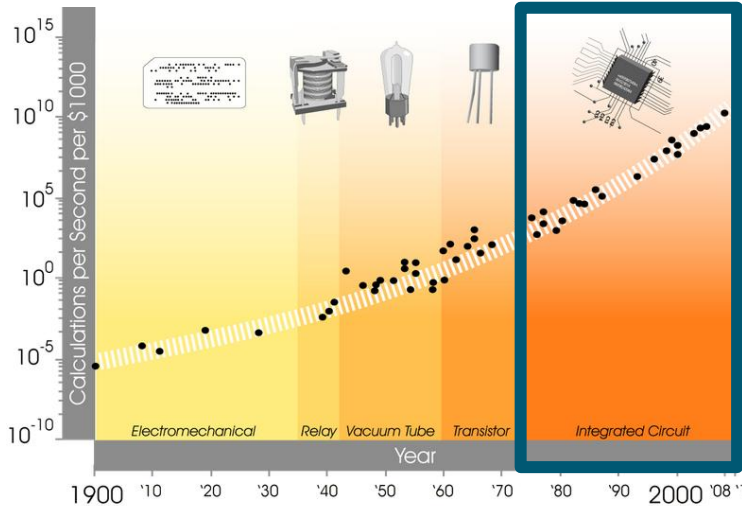
"Distribution of work on (potentially) physically isolated compute nodes"

Motivation: "Distributed" Surpassing Moore's Law

Moore's Law (Observation)

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

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)



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.

- 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, ...

Both use distributed computing!

- 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, ...

Distributed Data Management

Introduction

ThorstenPapenbrock
Slide **27**

Motivation: "Distributed" A Rule to Acknowledge

Amdahl's Law

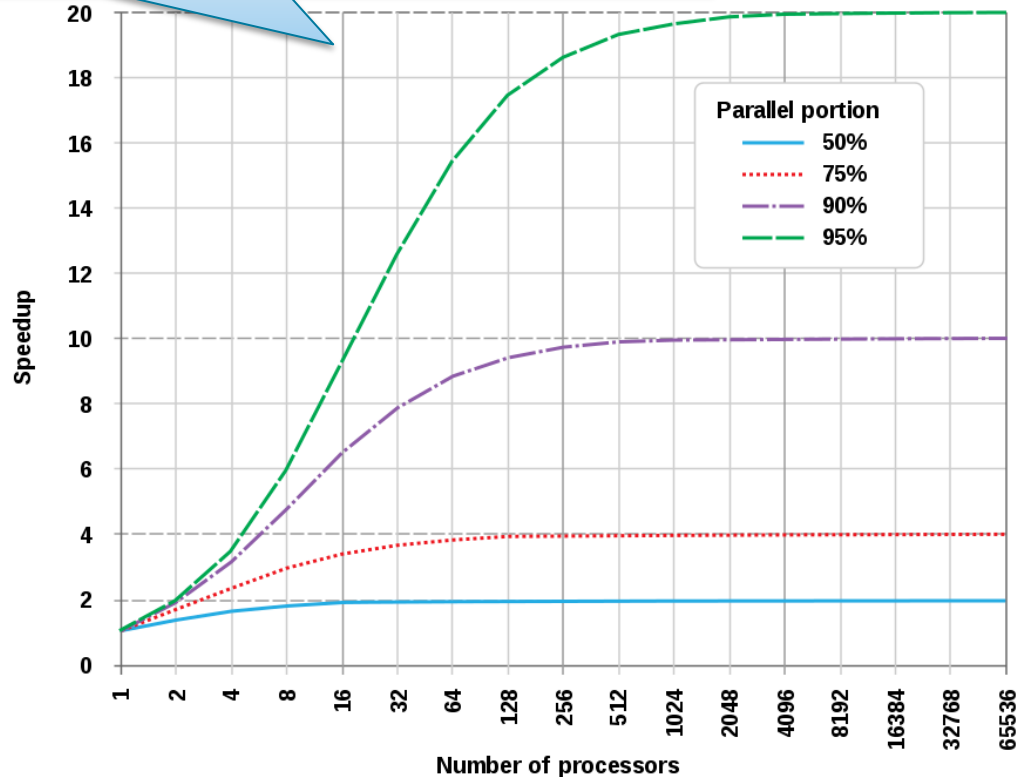
"The speedup of a program using multiple processors for parallel computing is limited by the sequential fraction of the program"

$$Speed^d(s) = \frac{1}{(1-p) + \frac{p}{s}}$$

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

p: percentage of the algorithm that profits from parallelization

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

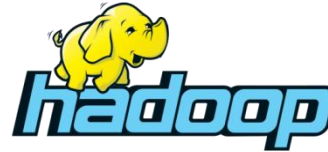


Motivation: "Distributed" New Technologies

Distributed Computing



Distributed Storage



Distributed Data Management

Introduction

ThorstenPapenbrock
Slide 29



INFRASTRUCTURE

HADOOP ON-PREMISE
 cloudera Hortonworks
 MAPR Pivotal
 IBM InfoSphere
 bluedata jethro

HADOOP IN THE CLOUD
 aws Microsoft Azure
 Google Cloud IBM InfoSphere
 CAZENA CenturyLink

STREAMING / IN-MEMORY
 aws databricks strim
 confluent GIGASCALE
 DATATORRY dataArtisans
 CHACRA hazelcast TERRACOTTA

NoSQL DATABASES
 Google Cloud aws
 ORACLE Microsoft Azure
 mongoDB MarkLogic
 REDISCLUB DATASTAX
 ArangoDB Couchbase
 redislabs SCYLLA

NewSQL DATABASES
 SAP Clustring Pivotal
 Cockroach Labs
 MEMSQL InfluxData
 CLOUDATA SPLOTE
 paradigms

GRAPH DBs
 neo4j Amazon Neptune
 IBM
 GraphSense
 Infoblox
 Vertica
 IBM
 Microsoft Azure
 Oracle
 Pivotal
 Snowflake
 InfoWorks

MPP DBs
 AWS
 IBM
 Microsoft Azure
 Oracle
 Pivotal
 Snowflake
 InfoWorks

CLOUD EDW
 AWS
 Microsoft Azure
 Oracle
 Pivotal
 Snowflake
 InfoWorks

DATA TRANSFORMATION
 talend pentaho
 alteryx TRIFORCE
 tami PAKALA
 StreamSets UNIFI

DATA INTEGRATION
 Informatica MapInfo
 segment enigma
 odium xipity
 import.io Stich

DATA GOVERNANCE
 Informatica
 IBM
 Collibra
 Alation
 WATERGATE
 KERA

MGMT / MONITORING
 AWS New Relic acinfo
 rubrik APPDYNAMICS
 IBM
 Splunk
 SignalFx
 Datadog
 PagerDuty Humio

STORAGE
 AWS Google Cloud
 Microsoft Azure
 PURE STORAGE
 ALIENEX
 Qumulo
 COHESTY

CLUSTER SVCS
 AWS
 Docker
 Keen IO
 MESOSPHERE
 Core OS
 CRSK

APP DEV
 Amazon
 Keen IO
 rainforest
 four
 scode
 IHIVE

CROWD-SOURCING
 amazonmechanics
 upwork
 crowd
 scode
 IHIVE

HARDWARE
 Google TPU
 intel AI GRAPH CORE
 MYTHIC
 NVIDIA
 Movidius
 PLAINJOB
 WAVE

GPU DBs
 kineticio
 IBM
 NVIDIA
 PLAINJOB
 bytlyt PLIN

CROSS-INFRASTRUCTURE/ANALYTICS
 AWS Google Cloud Microsoft IBM SAP Hewlett Packard Enterprise SAS 1010DATA VMWare TIBCO TERADATA ORACLE NetApp syncsort

ANALYTICS

DATA ANALYST PLATFORMS
 Microsoft pentaho alteryx
 QV
 QUAVUS AYASDI
 ATTIVO Datameer Quid Incofa
 inter:ona ClearStory Origami
 ENDOR HODR Borflinose

DATA SCIENCE PLATFORMS
 IBM KNIME dataiku
 DOMINGO rapidminer
 CONTINUUM ANALYTICS ALGORITHMIA
 DATAVIA ANALYTICS

BI PLATFORMS
 Microsoft AWS
 SAP
 Looker ATSCALE
 Google Data
 birst

VISUALIZATION
 +tableau
 Google Cloud
 Qlik
 ZEPL
 CHARTIO

MACHINE LEARNING
 AWS
 Google Cloud H2O
 DataRobot gamalon
 ELEMENT VISENZE
 bonasai

COMPUTER VISION
 Microsoft Azure
 Amazon Rekognition
 clarifai
 EVER AI deepomatic
 twenty9

HORIZONTAL AI
 IBM Watson Cortana
 sentiment Voyageur
 Afffective
 Humana FETUM
 nonlogics OSAR

SPEECH & NLP
 Google Cloud
 Amazon Alexa
 IBM Watson
 Microsoft Azure
 SoundCloud Inc.
 SoundCloud Inc.
 snips

SEARCH
 ORACLE
 ELASTICSEARCH
 SOLR
 SWIFTTYPE
 ATTIVO
 alphaSense
 omnius

LOG ANALYTICS
 splunk
 SUMOLOGGY
 IBM
 bitly predata
 Logzio

SOCIAL ANALYTICS
 Hootsuite sprinklr
 NETBASE
 synthesio
 simplereach
 libana
 SimilarWeb

WEB / MOBILE / COMMERCE ANALYTICS
 Google Analytics
 mixpanel AMPUTRUE
 sumall
 RECCI
 SIGOPT
 granify custora

APPLICATIONS - ENTERPRISE

SALES
 Oracle CHORUS
 INSIDESALES.COM
 conversica
 clari aviso tact.ai
 fuse:machines TRADERS

MARKETING - B2B
 RADIUS
 EVERSTRING
 HINTIGO sense
 tubular Datafat
 JENGA IO

MARKETING - B2C
 bloomreach SendGrid
 BlueYonder (PARSAO) Dixa
 ACTIONIQ SALUTRU BLUECORE
 QUANTINIA repartice Amperio
 amperity

CUSTOMER SERVICE
 MEDALLIA zendesk
 CLARABRIDGE
 Gainsight NOC DATA
 DigitalGenius afnrit
 AUTOMATON frame.ai
 imago INTERCOM

HUMAN CAPITAL
 entelo
 hiQ
 LEXIP
 WUOLU
 mya

LEGAL
 RAVEL
 JUDICATA
 R:SS
 Casetext

FINANCE
 Anaplan
 ZUORA
 TRADESHIFF

ENTERPRISE PRODUCTIVITY
 slack
 ZUORA
 ORACLE
 JIRA
 clara talla
 butter Kasisto

BACK OFFICE AUTOMATION
 UiPath
 WorkFusion

SECURITY
 TANUKI
 BlackRub
 BANKTRACE ANDMAL
 SANS
 DATASOFT
 SCYFIO
 BlueTrust
 imago

APPLICATIONS - INDUSTRY

ADVERTISING
 AppNexus
 Criteo
 Oracle
 DoubleClick
 Distillery
 Taboola

EDUCATION
 Edmentum
 Clever
 Edmentum
 K12
 Edmentum

GOVERNMENT
 OPENGOV
 mark43
 FiscalNote
 OpentixSoft

FINANCE - LENDING
 ondeck Affirm
 JUANPAI
 Kreditech AVANT
 INSEKT
 Lendix
 MoneyLion
 airc ognit

REAL ESTATE
 REDFIN
 Opendoor
 VTS
 CREDDOR
 COMSTOCK
 CAPE

INSURANCE
 natomials
 Lemnatec
 CYNCE
 SHR Technology
 TRAVELER

HEALTHCARE
 Fatiron Clever
 Metabion
 Gingeo
 Med
 Tempus
 Proton

LIFE SCIENCES
 color
 BenevolentAI
 Verity
 Clear Labs
 Cytiva
 Cytiva

TRANSPORTATION
 UBER TESLA
 CLEARPATH
 drive.ai
 nauto
 Pegasus
 Pegasus

AGRICULTURE
 FARMERS
 Granular
 Blueberry
 FarmLogs
 mavrx
 Prospera

COMMERCE
 instacart
 STITCH FIX
 TACHYUS
 TevGood
 other
 vithomy stem
 BytDance
 BIKEEVER
 remesh ASAPP

INDUSTRIAL
 GIGOT PREDIX
 UPTAKE
 TACHYUS
 TevGood
 other
 vithomy stem
 BytDance
 BIKEEVER
 remesh ASAPP

FRAMEWORK
 Apache Spark
 Flink
 YARN
 MESOS
 CDAP

QUERY / DATA FLOW
 Spark SQL presto
 SLAMDATA
 Google Cloud DataFlow

DATA ACCESS
 nifi mongoDB
 cassandra
 SOiDB
 CouchDB
 HBASE

COORDINATION
 talend
 Apache Zookeeper
 Apache Ambari

STREAMING
 Spark
 Flink
 Beam
 kafka
 druid
 STORM

OPEN SOURCE

STAT TOOLS
 Python
 ScalaLab
 SciPy

AI / MACHINE LEARNING / DEEP LEARNING
 TensorFlow theano
 Caffe
 Microsoft Cognitive Toolkit
 OpenAI
 FeatureFusion
 Chainer
 VESLS
 DIMSUM
 MAHOUT
 Aerospike

SEARCH
 elasticsearch
 Solr

LOGGING & MONITORING
 elasticsearch kibana
 logstash
 Prometheus

VISUALIZATION
 Tableau
 Rodeo

COLLABORATION
 Jupyter
 Anaconda

SECURITY
 Apache Ranger
 Knox
 Sentry

DATA SOURCES & APIs

HEALTH
 Apple
 VALIDIC
 practice fusion
 UPTAKE
 GE Digital
 thingwork
 helium
 somasora
 fitbit GARMIN
 kinsco

FINANCIAL & ECONOMIC DATA
 Bloomberg
 THOMSON REUTERS
 DOW JONES
 S&P CAPITAL IQ
 CBRIGHT
 xignite
 Quandl
 ENVENTRY
 PREMIER
 estimate
 Single Alpha
 StockWits
 PLAID
 Thinknum

AIR / SPACE / SEA
 Orbital Insights
 Airwave
 spire
 INDESTRY
 WINDWARD
 DroneDeploy

PEOPLE / ENTITIES
 acxiom Experian
 EPFLON
 InsideView
 Crimson Hexagon
 BASIS
 SAFEGRAPH

LOCATION INTELLIGENCE
 FOURSQUARE
 sense360
 PlaceIQ
 factical
 esri
 Mapillary
 StreetView
 cuebiq

DATA RESOURCES

DATA SERVICES
 Palantir
 OPERA
 fractal
 kaggle
 DataKind

INCUBATORS & SCHOOLS
 PLURALIGHT
 galvanize
 DataCamp
 DataMentor
 INSIGHT
 The Data Incubator

RESEARCH
 facebook research
 OpenAI
 MIRI
 VECTOR INSTITUTE
 A.I. RESEARCH LABS
 A.I. RESEARCH LABS

Distributed Data Management Introduction

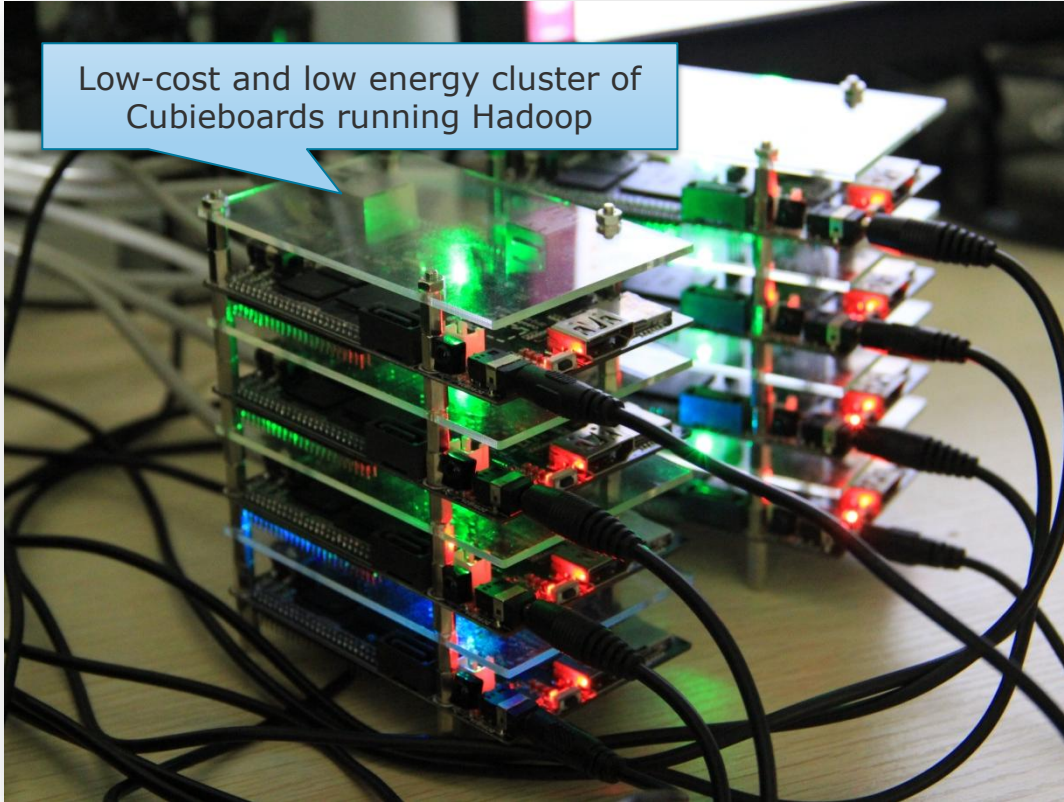
ThorstenPapenbrock Slide 30

Motivation: “Distributed” Driving Forces

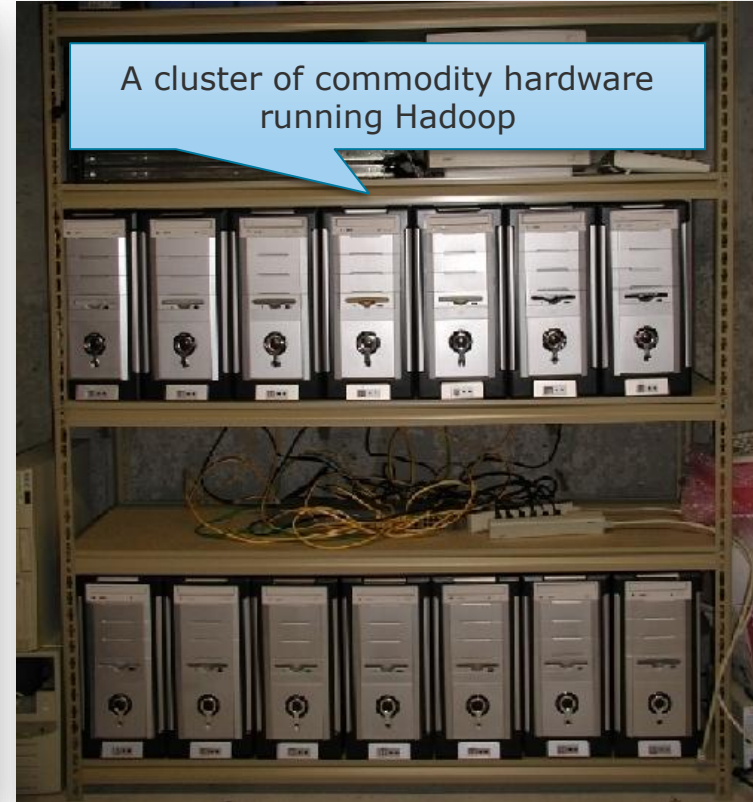
- **Data volumes increase:**
business data, sensor data, social media data, ...
- **Data analytics gains importance:**
downtime-less, real-time, predictive
- **Parallelization paradigm shifts:**
multi-core and network speeds increase while CPU clock speeds stall
- **Computation resources become more 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

Low-cost and low energy cluster of
Cubieboards running Hadoop



A cluster of commodity hardware
running Hadoop



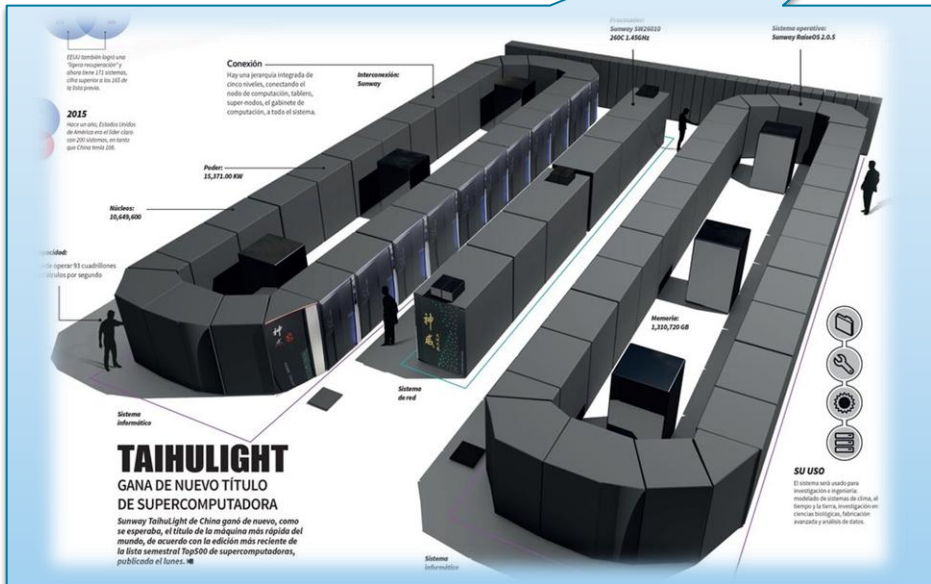
Motivation: "Distributed" Large Scale

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)	Pow (kW)
1	Sunway TaihuLight - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway , NRPCPC 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 VIIIff 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

Motivation: "Distributed" Super Large Scale

Top 10 Super Computers 2017



All distributed systems!

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Pow (kW)
1	Sunway TaihuLight - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway , NRPCPC 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

Motivation: "Distributed" Super Large Scale

Top 10 Super Computers 2017



All distributed systems!

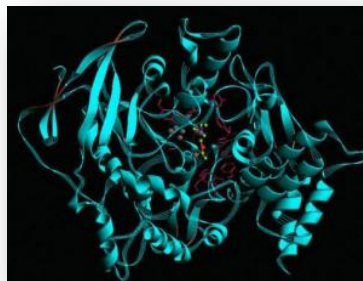
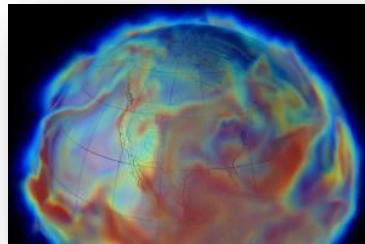
Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Pow (kW)
1	Sunway TaihuLight - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway , NRPCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	125,435.9	15,371
2	Tianhe-2 (MilkyWay-2) - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH 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

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)	Pow (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”



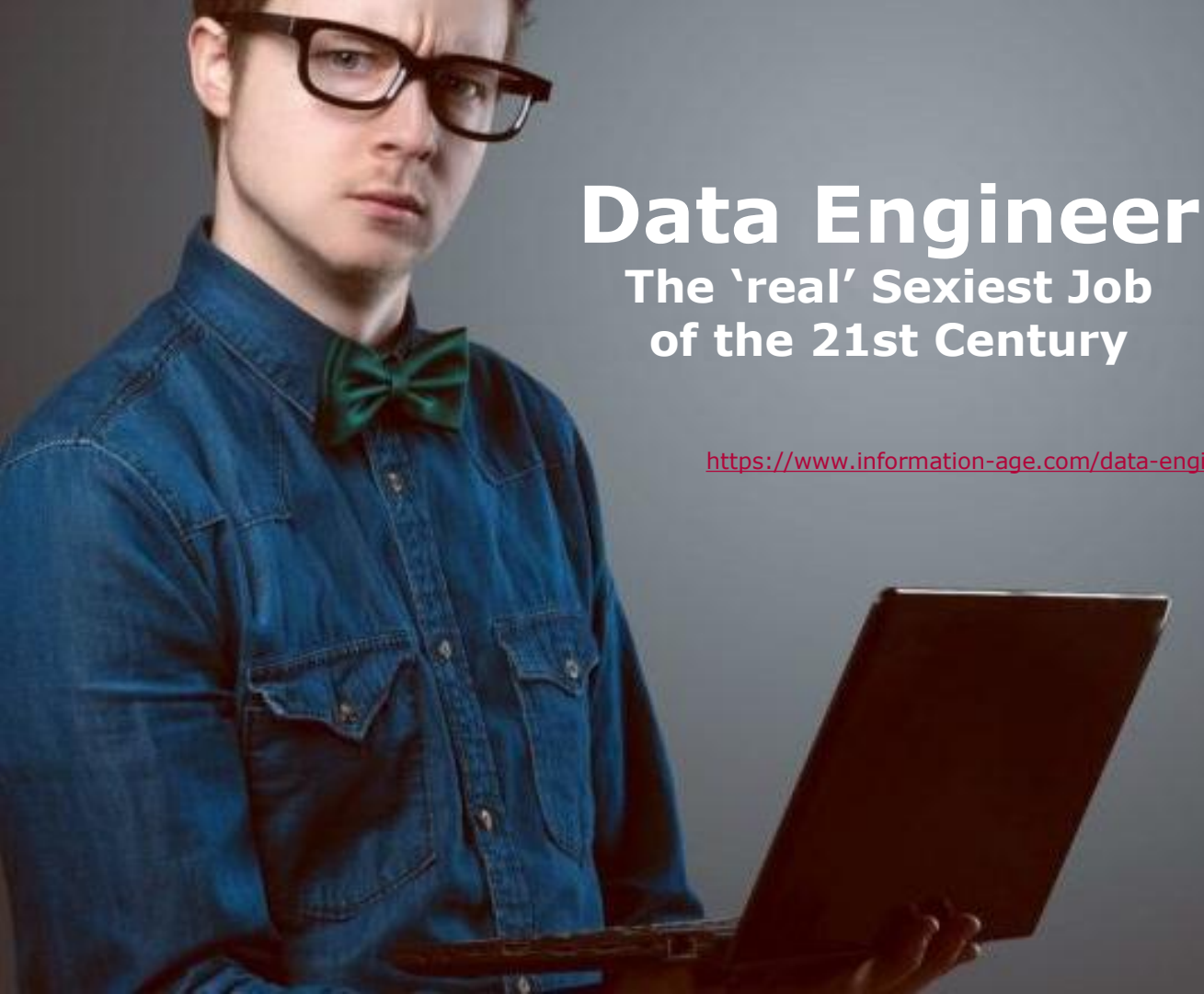


Data Scientist

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



- Top 10 LinkedIn followers per employee:**
1. Dundas
 2. Predixion Software
 3. iDashboards
 4. Chartio
 5. CXO-Cockpit
 6. Kognitio
 7. TARGIT
 8. Prophix
 9. Alpine Analytics
 10. Jedox

A market worth \$122 billion in 2016 with a growth of 11.3% per year!

Excellent job opportunities in many companies!

For a world that created an entire zettabyte (which is exactly 10^{12} GB) of data in the 2010 alone!

Business users (Log-scale)

Customers (Log-scale)

Shape: Deployment types

On Prem



Cloud (Mobile)

On Prem

Color = Client focus
(Blue = Enterprise ; Orange = SMB ; Green: Mix)

Size = Completeness of vision
(BIG Data, Analytics, CPM, Data Warehousing)



- Top 5 Self Service BI Vendors:**
1. Tableau
 2. Microsoft
 3. Qlik
 4. MicroStrategy
 5. GoodData

- Top 5 Predictive Analytics Vendors:**
1. IBM
 2. SAS
 3. RapidMiner
 4. KNIME
 5. Microsoft

- Top 10 Business Analytics Vendors:**
1. Microsoft
 2. MicroStrategy
 3. IBM
 4. SAP
 5. Tableau
 6. Qlik
 7. Alteryx
 8. SAS Institute
 9. TIBCO
 10. GoodData

- Top 5 Data Warehouse Vendors:**
1. Teradata
 2. SAP
 3. IBM
 4. Oracle
 5. Microsoft

Motivation: "Data"

Successful IT Startups

Example: Mobile Motion GmbH



Dubsplash

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

November 19, 2014:

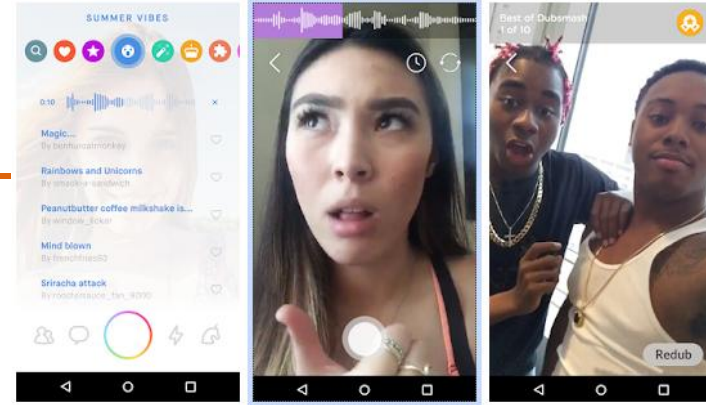
- **Launch** of the Dubsplash app

November 26, 2014:

- Dubsplash reached the **number one** downloaded app in Germany

June 1, 2015:

- Dubsplash had been downloaded over **50 million times in 192 countries**

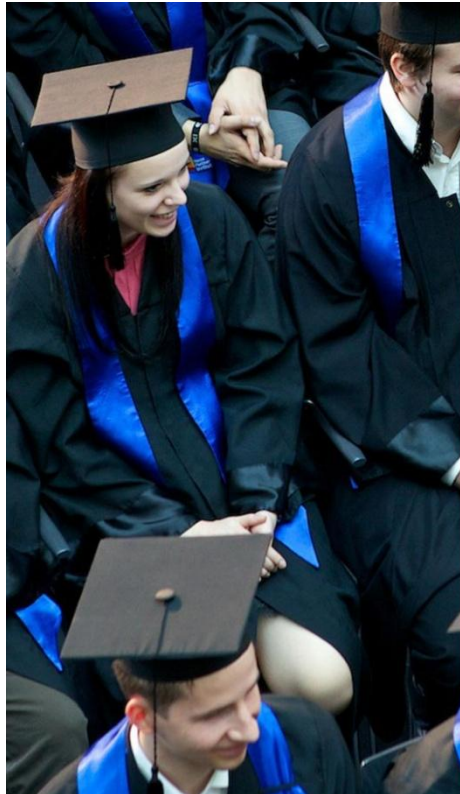


Distributed Data Management

Introduction

ThorstenPapenbrock
Slide **43**

Motivation: "Data" Successful IT Startups



Many further HPI Startups!

Distributed Data Management

Introduction

ThorstenPapenbrock
Slide 44

Motivation: "Data"

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
- ...

**Distributed Data
Management**

Introduction

ThorstenPapenbrock
Slide **45**

Introduction

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

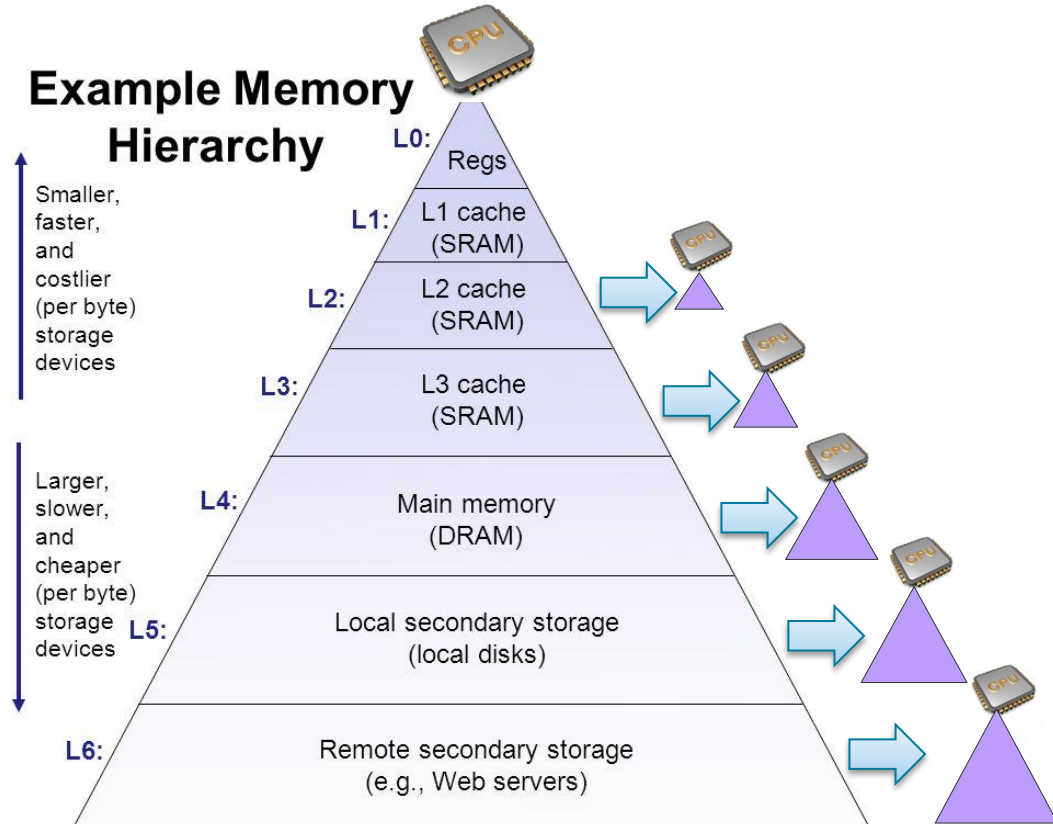


Motivation: "Management"

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.



Motivation: "Management"

Rethinking Data Management

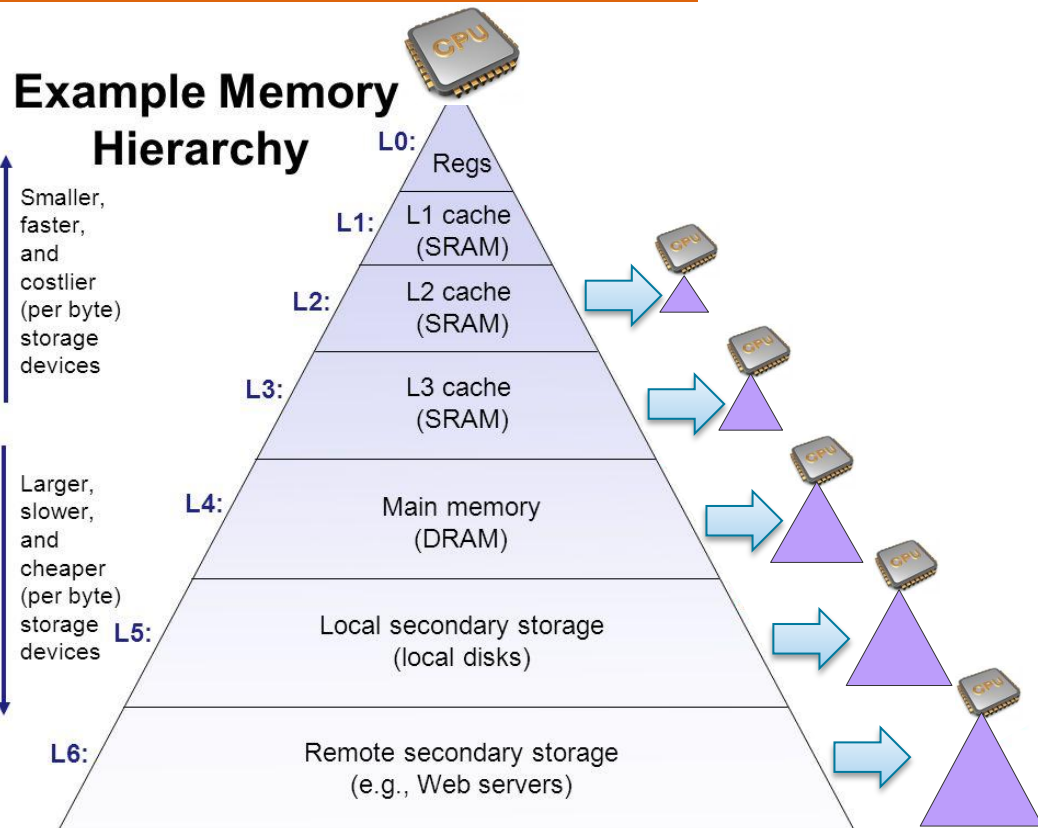
Moving data costs magnitudes more time and energy than computing data!

- 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.



Motivation: "Management"

Rethinking Data Management

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

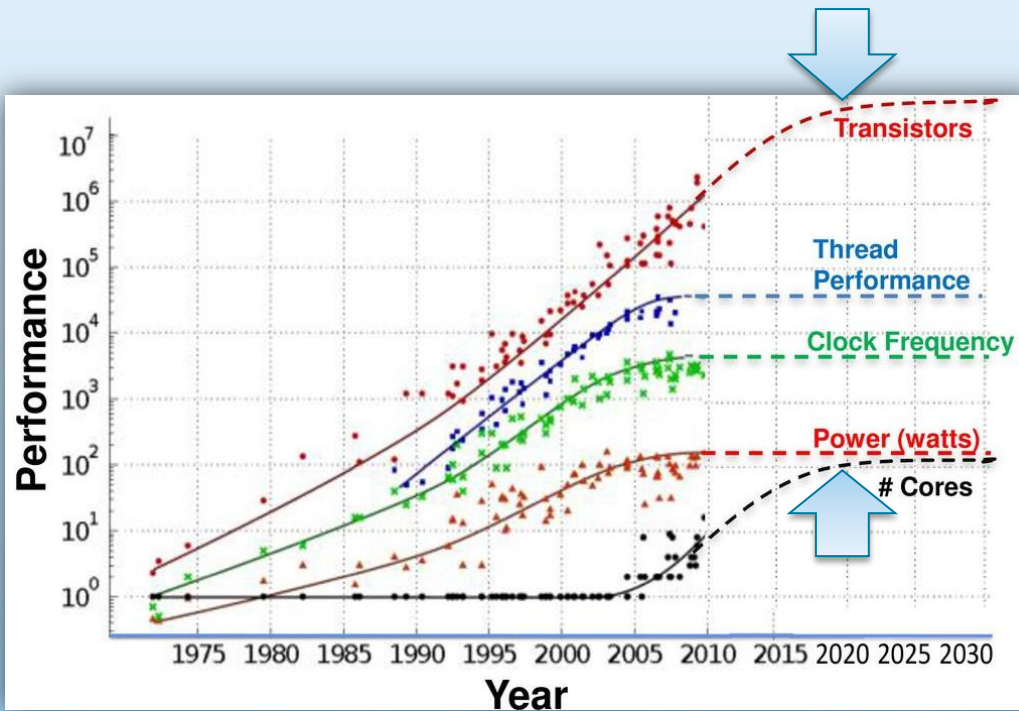
- Copying data costs time and energy
- Stalled processors during data copying consume energy.

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.

Why energy is a concern:



Motivation: "Management"

Rethinking Data Management

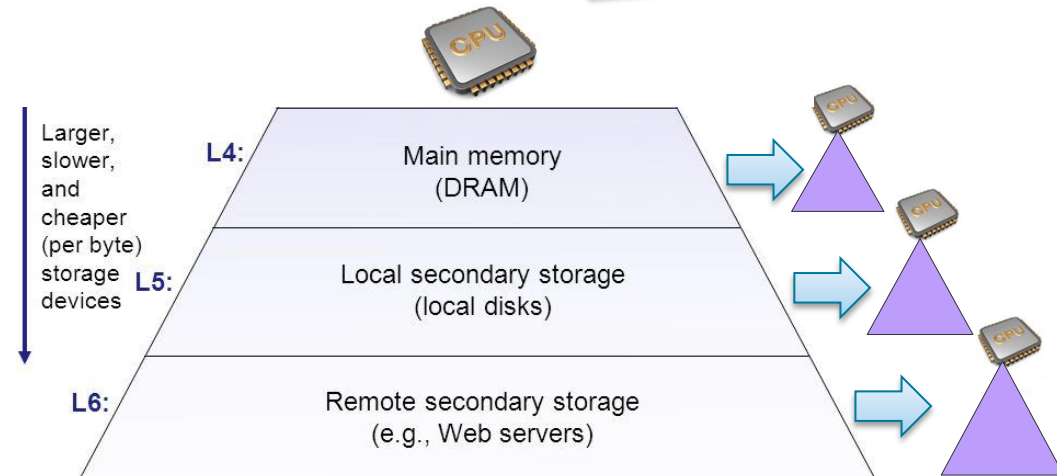
➤ Data engineers and data scientists need to be good data manager!

- 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

Relational Algebra / SQL

Linear Algebra
Statistics

Data Warehouse/OLAP
NF² / XQuery

Text Mining
Graph Mining

RDF / SparQL

Signal Processing

Information Integration

Stochastic Gradient Descent

Information Extraction

Machine Learning

Visual Analytics

Error Estimation

Privacy

Active Sampling

Memory Management

Monte Carlo

Parallelization

Regression

Scalability

Predictive Analytics

Memory Hierarchy

Sketches

Fault Tolerance

Data Obfuscation

Security

Convergence

Data Analysis Languages

Decoupling

Query Optimization

Iterative Algorithms

Real-Time

Curse of Dimensionality

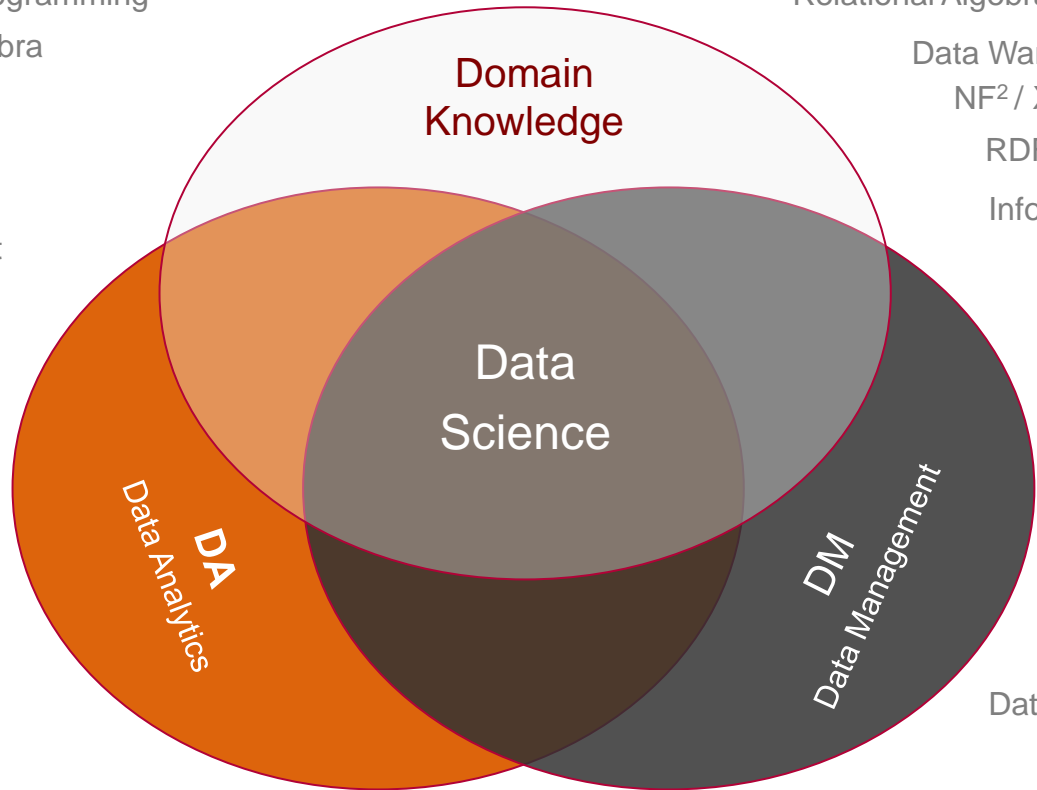
Legal Aspects

Control Flow

Indexing

Business Models

Data Flow



Motivation: “Management” 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!”

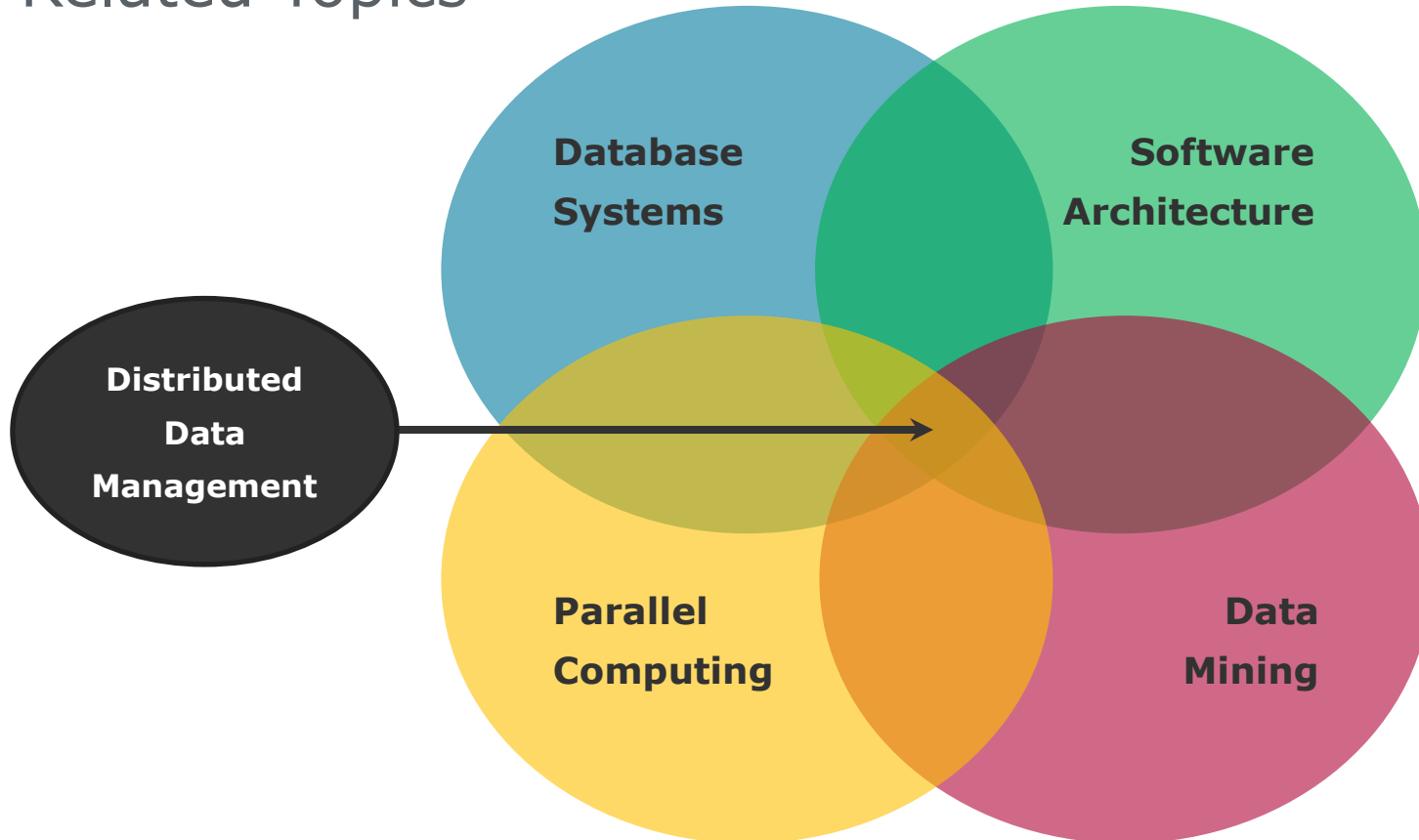
- Static (block) data
- Volatile (streaming) data

Distributed Data Management

Introduction

Motivation: "Management"

Related Topics

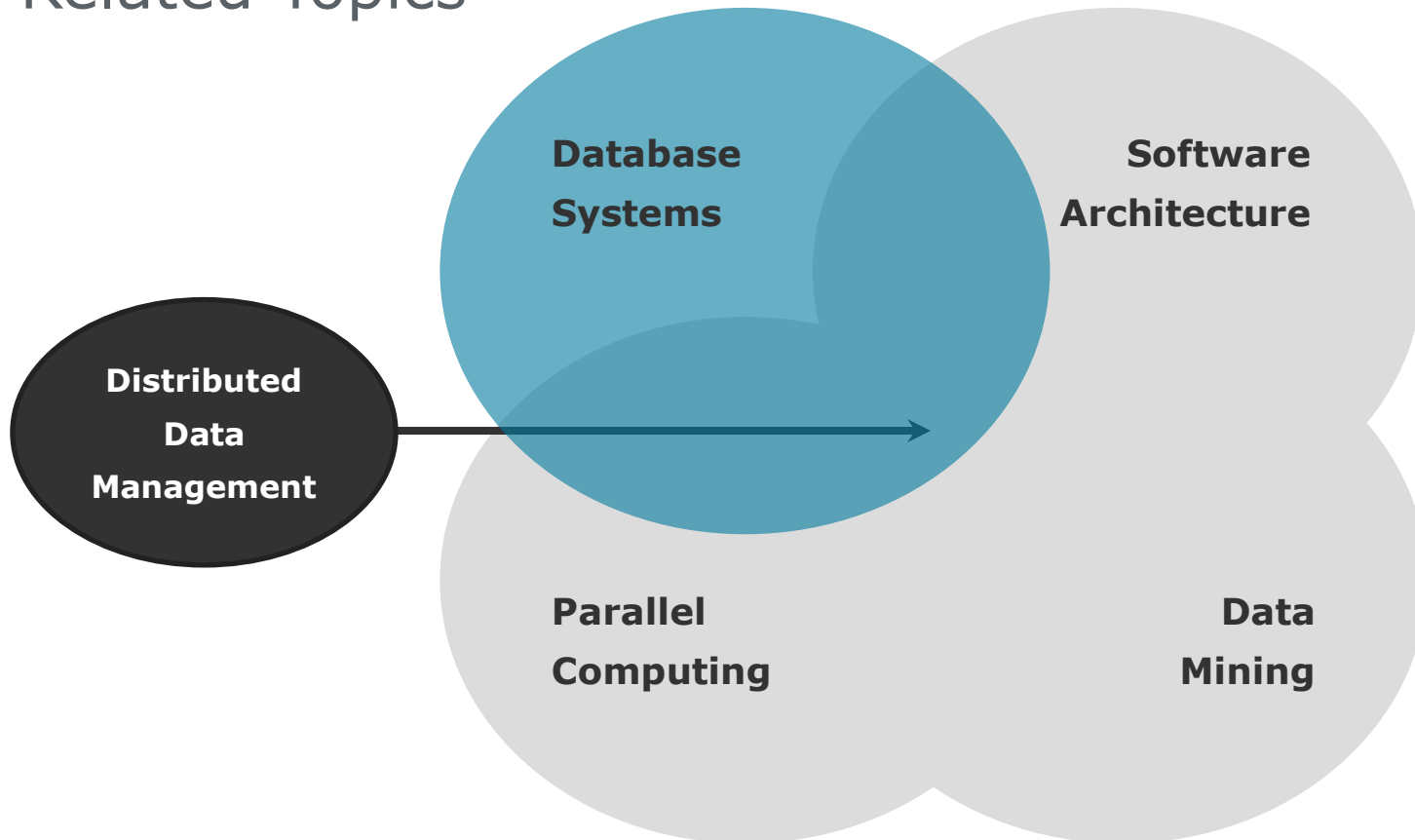


Distributed Data Management
Introduction

ThorstenPapenbrock
Slide **53**

Motivation: "Management"

Related Topics



**Distributed Data
Management**

Introduction

ThorstenPapenbrock
Slide **54**

Motivation: “Management” Database Systems

Touch points

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

Not in this lecture

- Physical data storage
- Foundations on transaction management and logging

More focused lectures/seminars

- Database Systems I + II (Prof. Naumann)
- Develop Your Own Database (Dr. Perscheid)

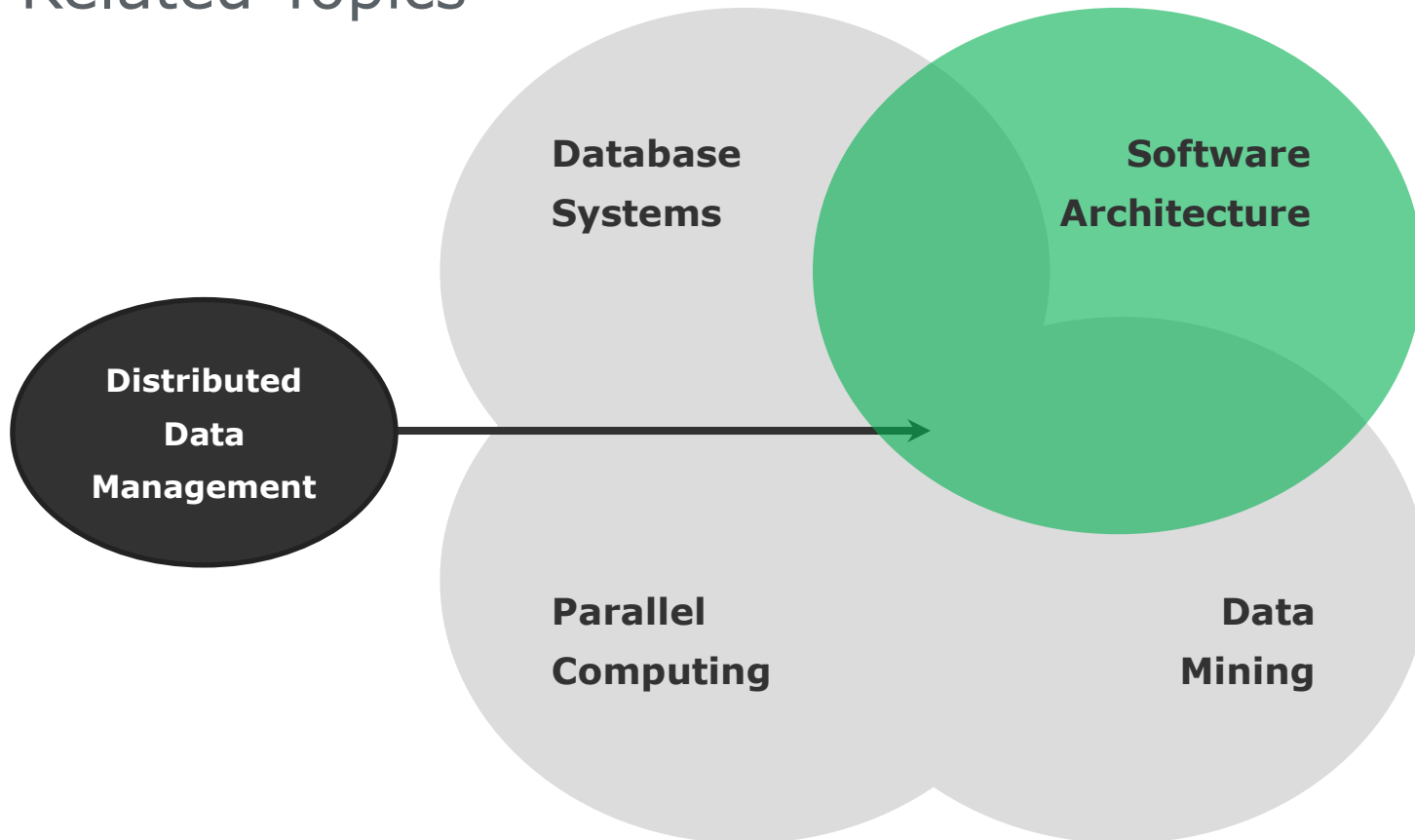
**Distributed Data
Management**

Introduction

Thorsten Papenbrock
Slide **55**

Motivation: "Management"

Related Topics



**Distributed Data
Management**

Introduction

ThorstenPapenbrock
Slide **56**

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 (Prof. Hirschfeld)
- Software Technique (Prof. Hirschfeld)

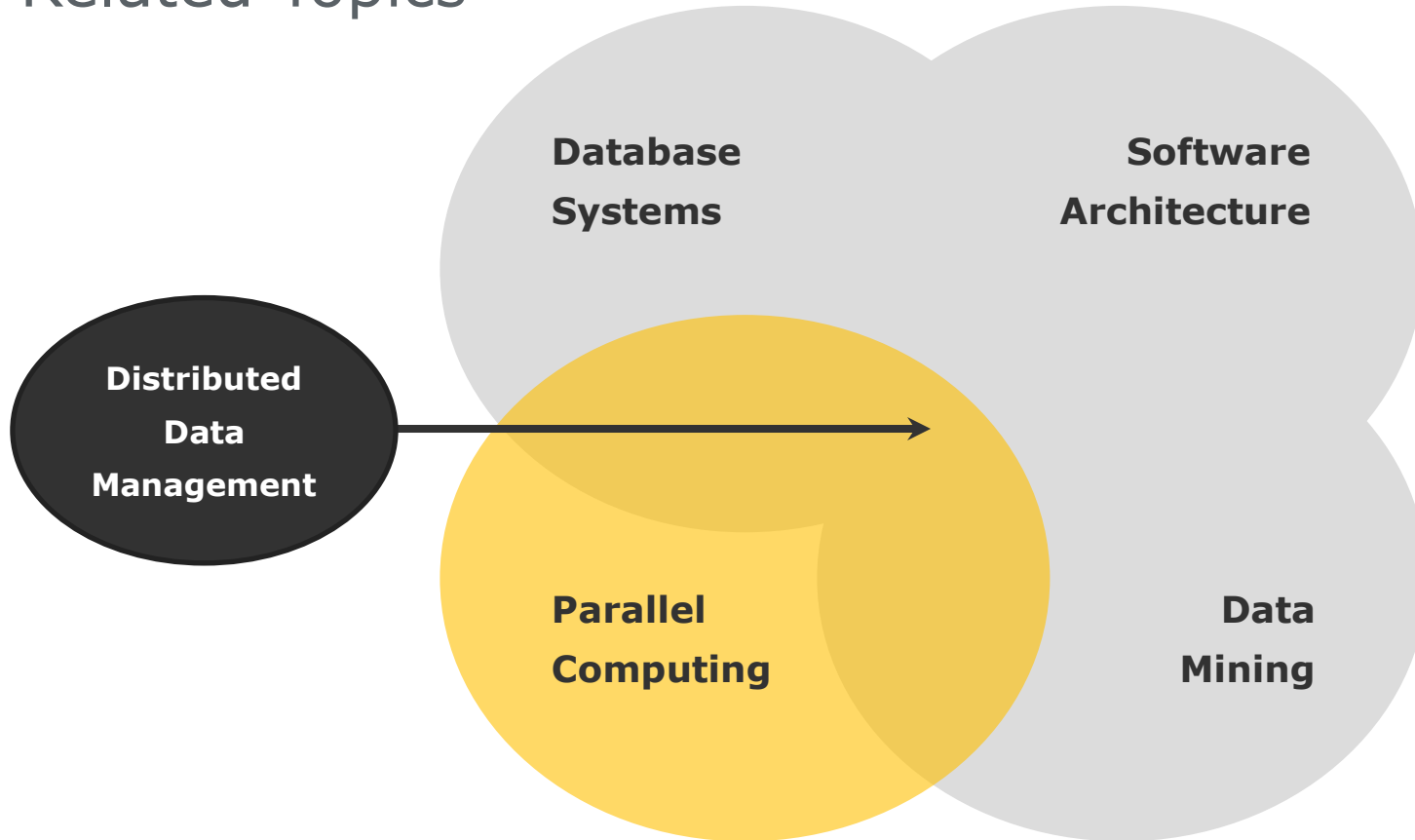
Distributed Data Management

Introduction

ThorstenPapenbrock
Slide **57**

Motivation: "Management"

Related Topics



**Distributed Data
Management**

Introduction

ThorstenPapenbrock
Slide **58**

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

- Parallele Programmierung und Hetergenes Rechnen (Prof. Polze)
- Programmierung paralleler und verteilter Systeme (Dr. Feinbube)

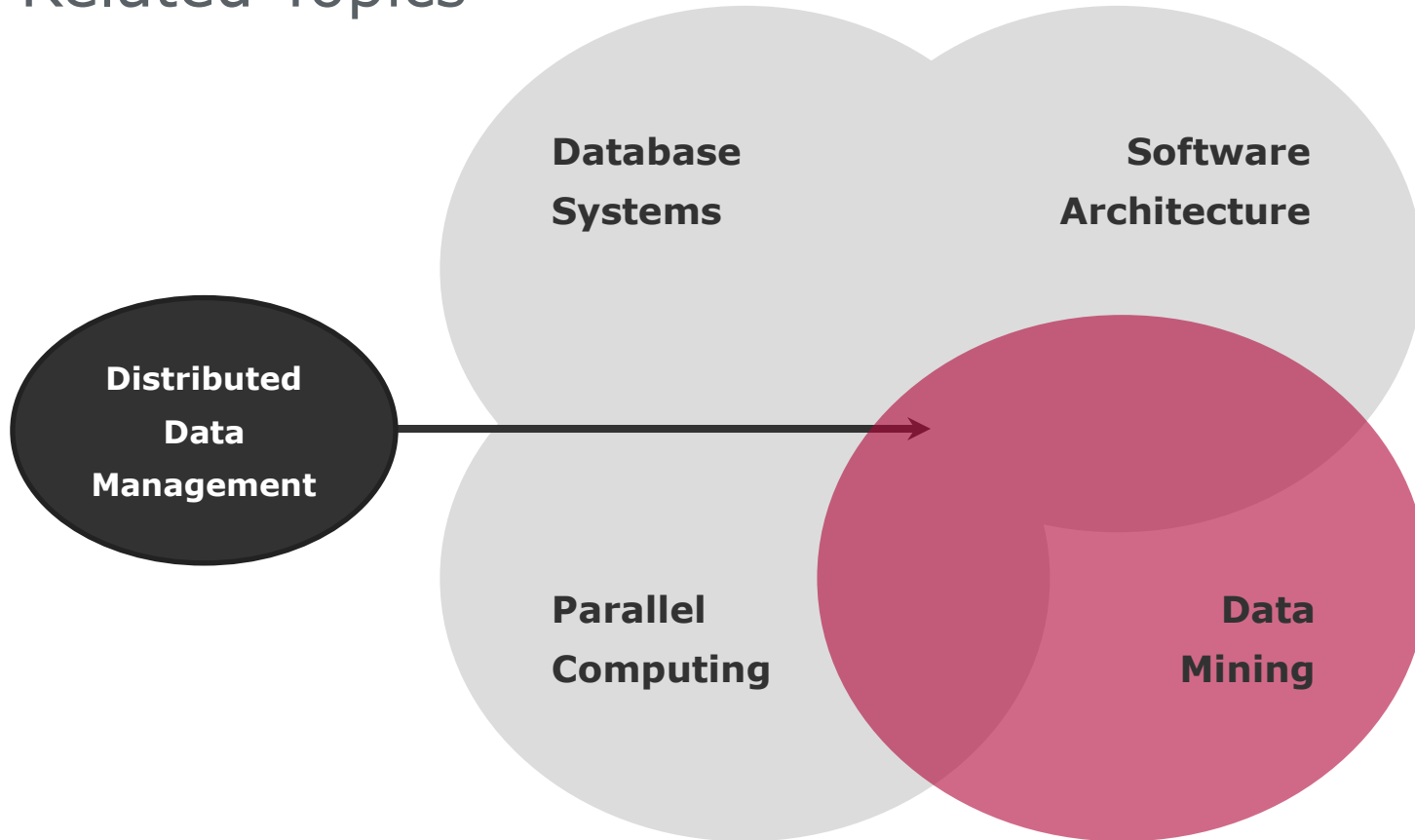
Distributed Data Management

Introduction

ThorstenPapenbrock
Slide **59**

Motivation: "Management"

Related Topics



**Distributed Data
Management**

Introduction

ThorstenPapenbrock
Slide **60**

Motivation: “Management”

Data Mining

Touch points

- Efficient data querying and aggregation queries
- Data engineering for data analysis
- Scalable computing

Not in this lecture

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

More focused lectures/seminars

- Deep Learning (Prof. Lippert), Knowledge Graphs (Dr. Krestel), Statistical Models (Prof. Renard), Building Machine Learning Applications (Dr. Albrecht), Machine Learning-based Control of Dynamical Systems (Prof. Giese) ...

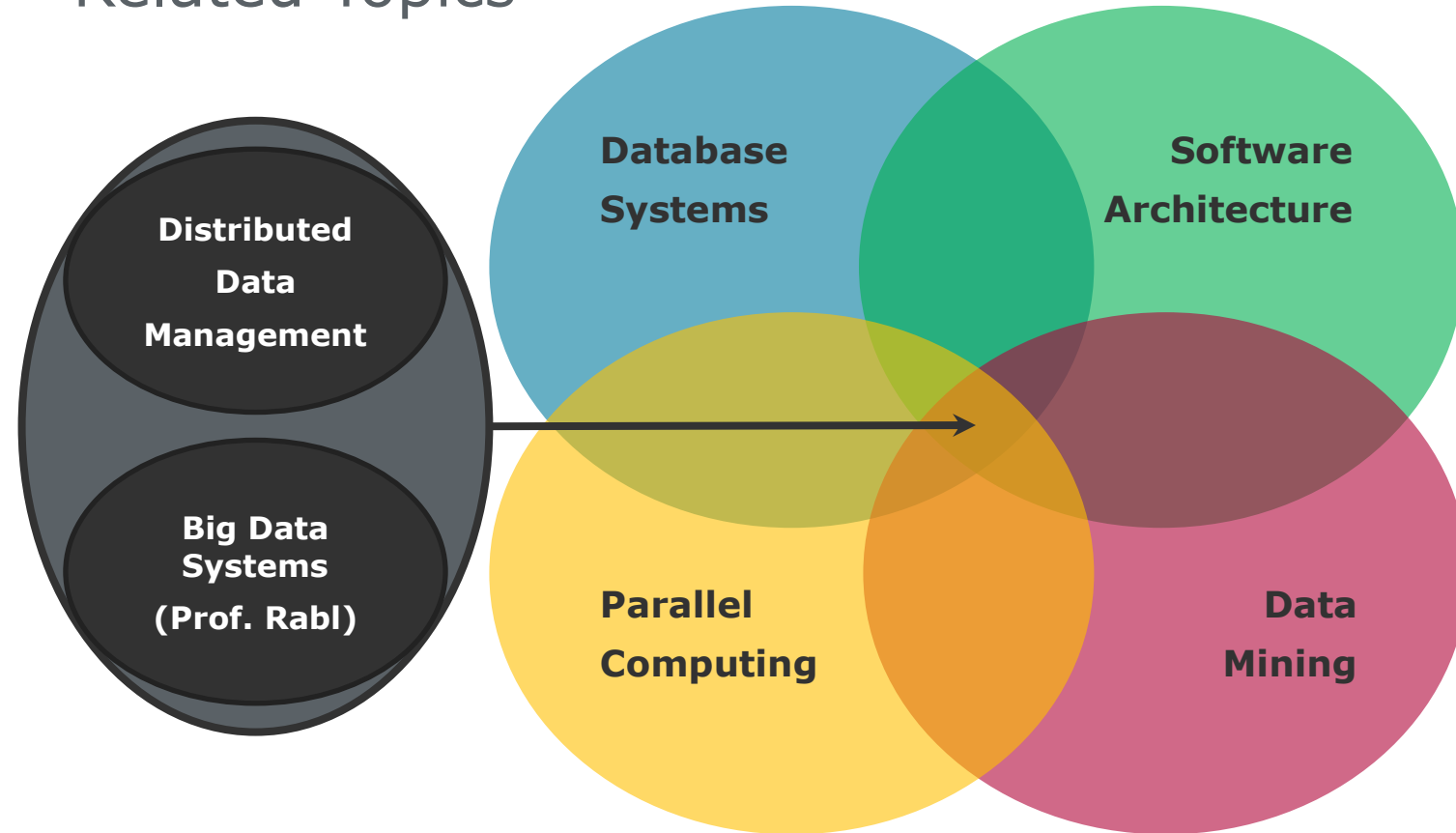
**Distributed Data
Management**

Introduction

Thorsten Papenbrock
Slide 61

Motivation: "Management"

Related Topics



Distributed Data Management
Introduction

ThorstenPapenbrock
Slide **62**

Motivation: "Management"

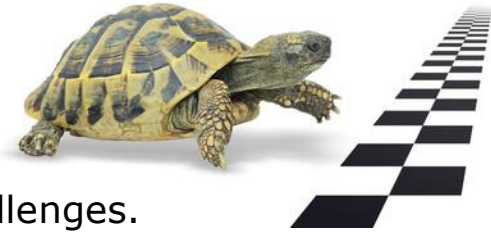
Lecture Goals

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 settings.



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 applications.

**Distributed Data
Management**

Introduction

ThorstenPapenbrock
Slide **63**

“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 CO2 footprint as global air travel.”

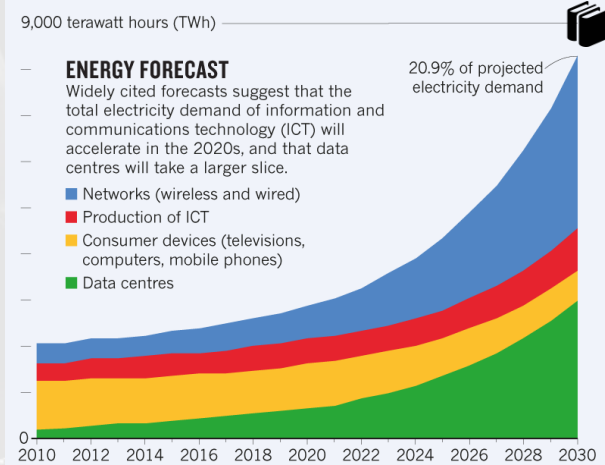
Global e-Sustainability Initiative

<https://internethealthreport.org/2018/the-internet-uses-more-electricity-than/>

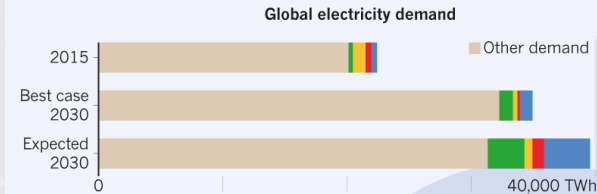
- “Data centres [...] consume about 3% of the global electricity supply [...] accounting for about 2% of total greenhouse gas emissions” in 2016.

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>



The chart above is an ‘expected case’ projection from Anders Andrae, a specialist in sustainable ICT. In his ‘best case’ scenario, ICT grows to only 8% of total electricity demand by 2030, rather than to 21%.



INTERNET EXPLOSION

Internet traffic* is growing exponentially, and reached more than a zettabyte (ZB, 1×10^{21} bytes) in 2017.



*Traffic to and from data centres.

†TB, terabyte (10^{12} bytes); PB, petabyte (10^{15} bytes); EB, exabyte (10^{18} bytes).

“Dark Magic”

- “In kürzester Zeit wird die Digitalisierung zum Klimaproblem Nummer eins werden.”

HPI, clean-IT Initiative

<https://hpi.de/open-campus/hpi-initiativen/clean-it-initiative.html>

CO2-footprint



1 AI model training



= 300 round-trip flights



= 5 car life cycles

<https://www.aclweb.org/anthology/P19-1355.pdf>

https://theoutline.com/post/8186/artificial-intelligence-destroy-environment?utm_campaign=Artificial%2BIntelligence%2BWeekly&utm_medium=email&utm_source=Artificial_Intelligence_Weekly_131&zd=1&zi=cayumttz

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
Throughput

MAINTAINABILITY
Operability
Simplicity & evolvability

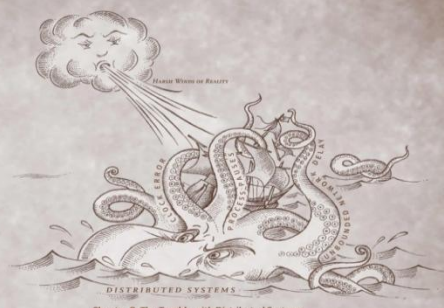
Chapter 1. Reliable, Scalable, and Maintainable Applications



Chapter 2. Data Models and Query Languages



Chapter 3. Storage and Retrieval



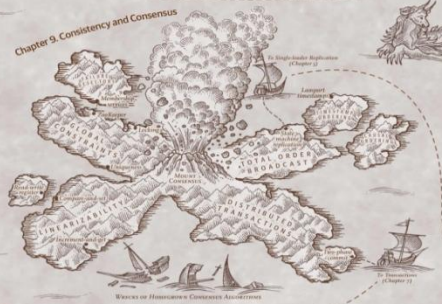
Chapter 8. The Trouble with Distributed Systems



Chapter 7. Transactions



Chapter 4. Encoding and Evolution



Chapter 9. Consistency and Consensus



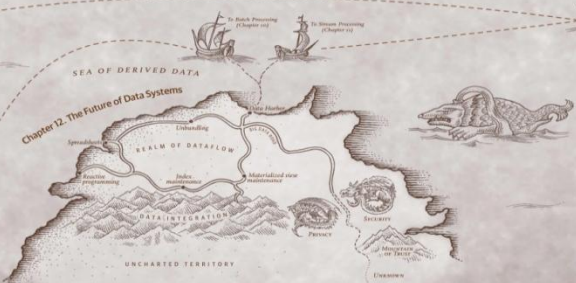
Chapter 5. Replication



Chapter 6. Partitioning



Chapter 10. Batch Processing



Chapter 12. The Future of Data Systems

O'REILLY

Designing Data-Intensive Applications

THE BIG IDEAS BEHIND RELIABLE, SCALABLE, AND MAINTAINABLE SYSTEMS

Martin Kleppmann