

# Big Data - Introduction

*Winter Semester 2018,  
Cooperative State University Baden-Wuerttemberg*



# Agenda – 01.10.2018

## 01 About This Lecture

Audience, Schedule, Materials (Slides, Script, Exercises, Books), Prerequisites, Exam, Feedback, Questions

## 02 Introduction To Big Data

Distributed Computing, Big Data V's, Big Data and Consistency, Big Data in Business, Definition of Big Data, Big Data Science

## 03 HandsOn - Hadoop

Introduction To The Hadoop Ecosystem, HDFS, Hive, Cluster Components

## 04 Exercises

Lorem Ipsum  
Sir dolor



# About This Lecture

Scope, Schedule, Materials, Prerequisites, Exam,  
Feedback, Questions



# About Big Data

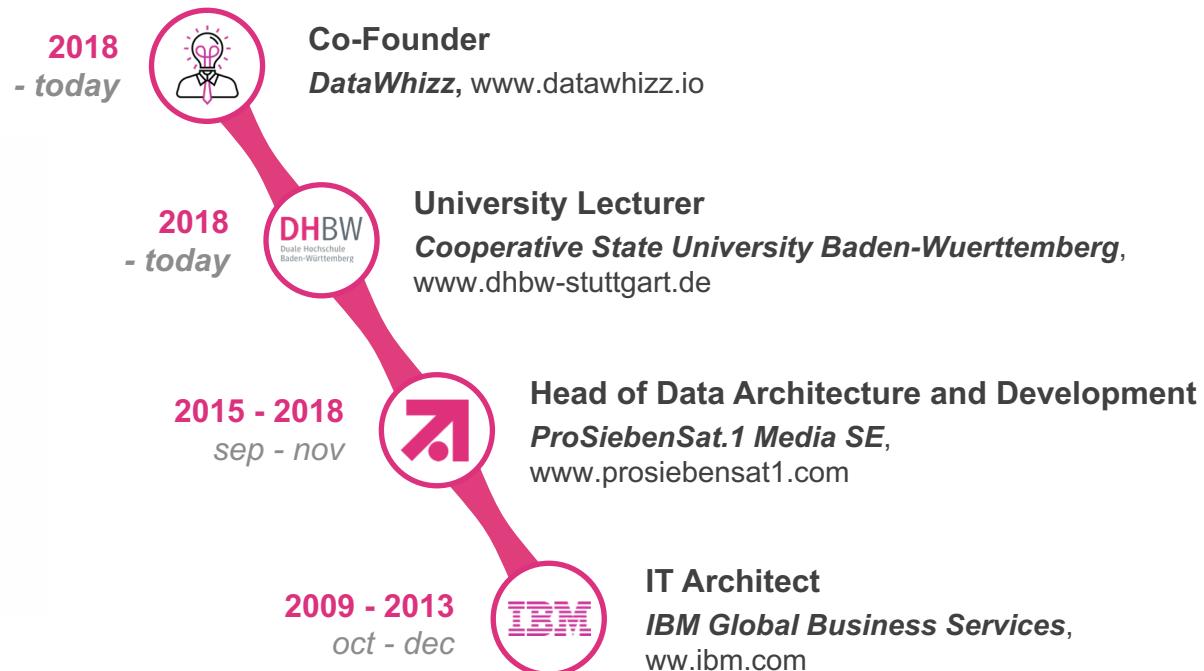
***“Big Data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it.”***

— Dan Ariely, Professor of Psychology and Behavioral Economics, Duke University

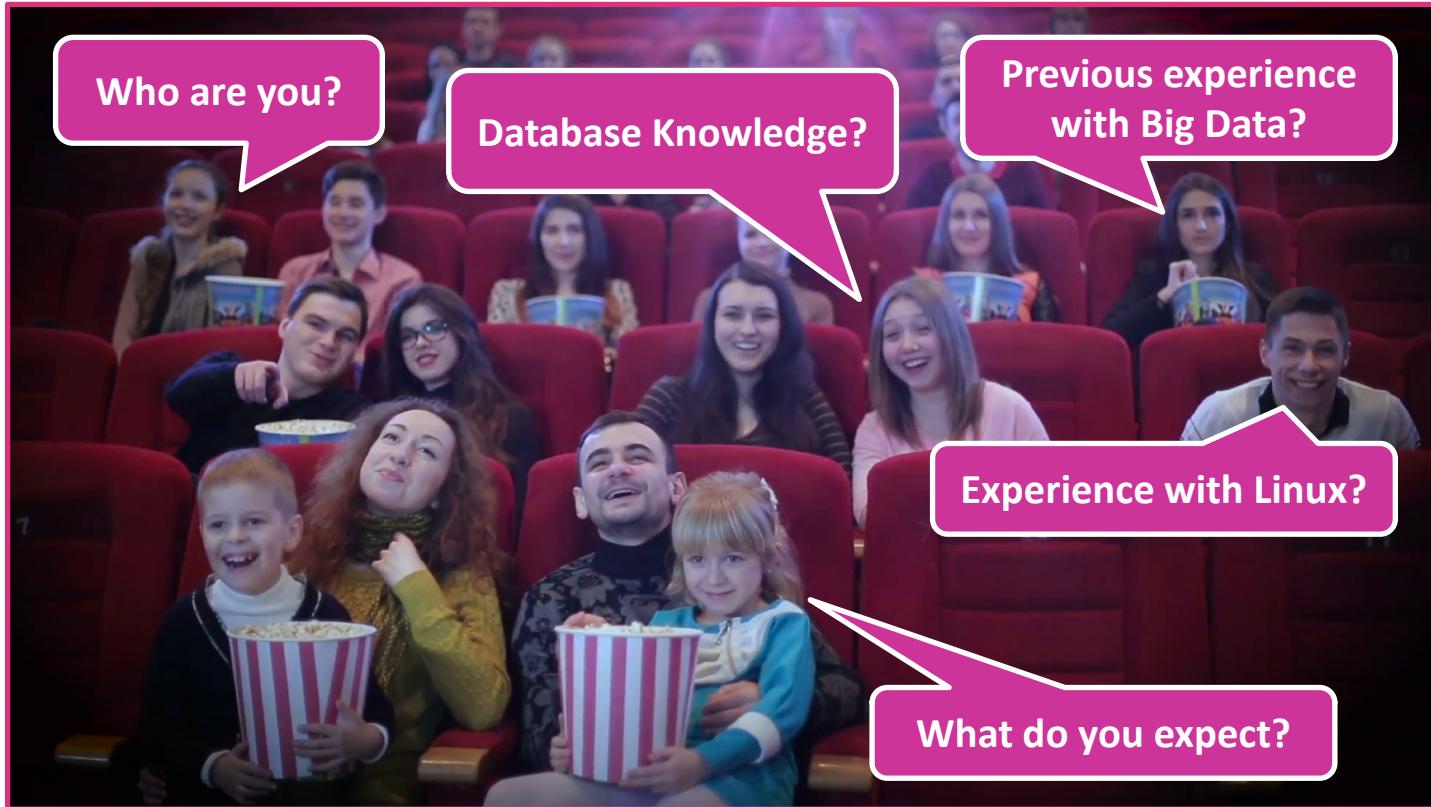
# About This Lecture



# About Me



# About You



# Schedule

01.10.2018	16:00-18:30	Ro. 0.11	<b>About This Lecture, Introduction to Big Data</b>	HandsOn Hadoop
08.10.2018	16:00-18:30	Ro. 0.11	Non-Functional Requirements Of Distributed Data-Systems	HandsOn MapReduce
15.10.2018	16:00-18:30	Ro. 0.11	Data Models and Access	HandsOn Hive and UDF's
22.10.2018	16:00-18:30	Ro. 0.11	Challenges Of Distributed Data Systems: Replication	HandsOn Spark
29.10.2018	16:00-18:30	Ro. 0.11	Challenges Of Distributed Data Systems: Partitioning	HandsOn MongoDB
05.11.2018	16:00-18:30	Ro. 0.11	Batch Processing	HandsOn Cassandra
12.11.2018	16:00-18:30	Ro. 0.11	Stream Processing	HandsOn Spark Streaming
19.11.2018	16:00-18:30	Ro. 0.11	ETL Workflow And Automation	HandsOn PDI/Airflow
26.11.2018	16:00-18:30	Ro. 0.11	Software And Frameworks	HandsOn tbd
03.12.2018	16:00-18:30	Ro. 0.11	Exam Preparation / Data Science	HandsOn PySpark
10.12.2018	tbd	tbd	<b>Exam</b>	



# Schedule – Lecture Composition

01.10.2018 16:00-18:30 Ro. 0.11

08.10.2018 16:00-18:30 Ro. 0.11

15.10.2018 16:00-18:30 Ro. 0.11

22.10.2018 16:00-18:30 Ro. 0.11

29.10.2018 16:00-18:30 Ro. 0.11

05.11.2018 16:00-18:30 Ro. 0.11

12.11.2018 16:00-18:30 Ro. 0.11

19.11.2018 16:00-18:30 Ro. 0.11

26.11.2018 16:00-18:30 Ro. 0.11

03.12.2018 16:00-18:30 Ro. 0.11

10.12.2018 tbd tbd

1

**Presentation, Review and Discussion of exercises** from previous lecture

2

**Lecture** of current topic

3

**HandsOn** to Software&Frameworks

4

**Exercise** execution

# Materials

## Script

<https://github.com/marcelmittelstaedt/BigData/tree/master/script>



## Slides

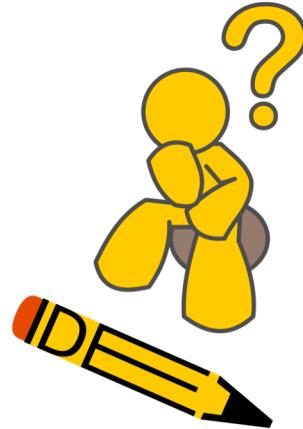
<https://github.com/marcelmittelstaedt/BigData/tree/master/slides>



# Materials

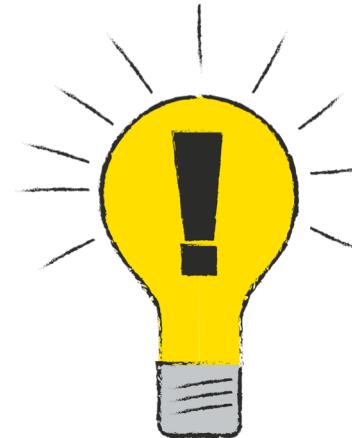
## Exercises

[https://github.com/marcelmittelstaedt/BigData/  
tree/master/exercises](https://github.com/marcelmittelstaedt/BigData/tree/master/exercises)

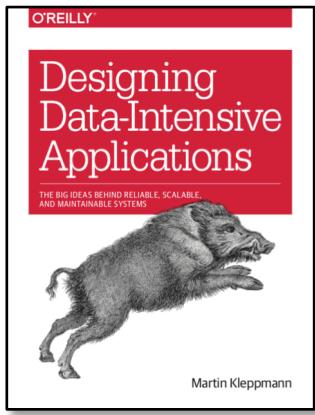


## Solutions

[https://github.com/marcelmittelstaedt/BigData/  
tree/master/solutions](https://github.com/marcelmittelstaedt/BigData/tree/master/solutions)

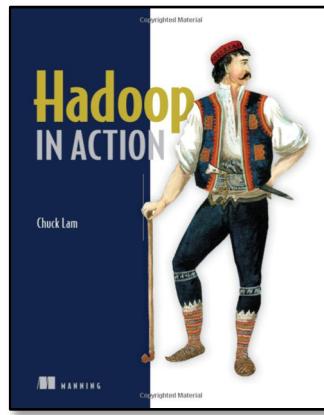


# Literature / Course Books



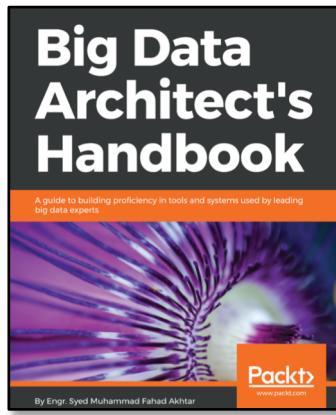
**Designing Data-  
Intensive Applications**  
Martin Kleppmann

[Amazon Link](#)



**Hadoop In Action**  
Chuck Lam

[Amazon Link](#)



**Big Data Architect's  
Handbook**  
Syed Muhammad  
Fahad Akhtar

[Amazon Link](#)

[PacktPub Link](#)



**Tech Ebooks from  
developer for developer**

[www.packtpub.com](http://www.packtpub.com)



[www.marcel-mittelstaedt.com](http://www.marcel-mittelstaedt.com)

# Prerequisites

To efficiently participate:

For exam:



# Feedback

Questions: **anytime**



This is a **new** and my **first lecture** ;)

If you find any **mistakes** or **misspellings** in the script or slides:

- Mail: [contact@marcel-mittelstaedt.com](mailto:contact@marcel-mittelstaedt.com)
- commit a push request (git)

**Feedback:** after lecture or via mail

# Questions



# Introduction To Big Data

Distributed Computing, Big Data V's, Big Data and Consistency, Big Data in Business, Definition of Big Data, Big Data Science



# HandsOn - Hadoop

Introduction To The Hadoop Ecosystem, HDFS,  
Hive, Cluster Components



[www.marcel-mittelstaedt.com](http://www.marcel-mittelstaedt.com)



# Exercises

Lorem Ipsum

