

Big Data SoSe 2025

Module Overview

Lecture Materials

Exercise Materials

Session 1: Introduction

What is Big Data?

Quiz 1: Introduction

Session 2: Storage Infras

Session 3: Column Stores

Session 4: Distributed Da

Guest Lecture Task

Session 5: Processing of

Session 6: Processing Lar

Session 7: Link Analysis

Forum

Quiz 1: Introduction

Tutorial on 8.05.2025

Performance summary

Assessed

Success status

Undefined

Score

48 of 54 points

Attempts

2

Results

Course

Big Data SoSe 2025

ID: 4853531345 / 107573654097444

Test

Test 1 Big Data 2025

ID: 4886921307

This are your test results

Duration

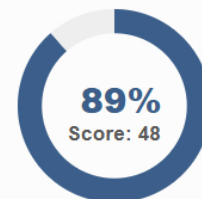
0h 32m 57s 5/7/2025, 3:01 PM - 5/7/2025, 3:34 PM

Answered

16 of 16 questions (100%)

Your score

48 of 54 points (89%)



Recall

Big Data definition

Status

Answered

Your score

4 / 4

100%

Response

Determine if the following statments are True or False

True

False



Lack of structure often the bigger problem then the data volume



"Big" is the only distinctive aspect of new forms of data



"Big" is a moving target. Only when the size becomes a challenge is it worth referring to it as big



Big data is data, that is too big, moves too fast, or doesn't fit the structures of your database architectures

Vs of Big data

Status

Answered

Your score

2 / 2

100%

Response

What are the initial 3 Vs of Big Data?



volume, velocity, and veracity



volume, velocity, and variety

- ☐ volume, velocity, and value
- ☐ volume, value, and variety

Additional Vs

Status	Answered
Your score	4 / 4 <div style="width: 100%;"></div> 100%

Response

Match Additional Vs of Big data with their definition

Validity

Means data are appropriate for the intended use

Volatility

Refers to how long is data valid and how long should it be stored

Variability

Refers to the data constantly changing meaning

Visualisation

Representation of the data in comprehensive form

Big Data vs Traditional analytics

Status	Answered
Your score	7 / 7 <div style="width: 100%;"></div> 100%

Response

Match given characteristics with Traditional analytics or Big Data

	Big Data	Traditional analytics
Data formatted in rows and columns	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Constant flow of data	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Focus on statistical and mathematical analysis	<input type="checkbox"/>	<input checked="" type="checkbox"/>
"Data-first" approach	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Unstructured, fast-moving data	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Hypothesis-based approach	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Those methods typically operate at a faster velocity	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Variety Aspect of Big Data

Status	Answered
Your score	3 / 3 <div style="width: 100%;"></div> 100%

Response

Which type of data source contributes to the variety aspect of Big Data?

- ☒ Unstructured data from social media
- ☒ Semi-structured data from IoT devices
- ☒ Structured data from databases

Definition

Status	Answered		
Your score	1 / 1	<div><div></div></div>	100%

Response

What is the term used to describe turning everyday actions or items into data?

- ☐ Information Encoding
- ☒ Datafication
- ☐ Digitization
- ☐ Data Migration

Hadoop Ecosystem

Status	Answered		
Your score	3 / 3	<div><div></div></div>	100%

Response

Which of the following technologies are part of the Hadoop ecosystem?

- ☐ Tensor Flow
- ☒ Pig
- ☒ Hive
- ☒ Spark

More Big data

Status	Answered		
Your score	2 / 4	<div><div></div></div>	50%

Response

True False

- | | | |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | Big Data is irrelevant to small businesses. |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Scaling up in Big Data infrastructure involves using many standard computers and distributing data and computation across them |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Big Data challenges include storage, processing, and analysis. |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | Hadoop is a proprietary system developed by Facebook. |

Hadoop Core Modules

Status	Answered		
Your score	4 / 4	<div><div></div></div>	100%

Response

Hadoop Distributed File System

A distributed file system that provides high-throughput access to application data.

Hadoop Common

The utilities that support the other Hadoop modules.

Hadoop YARN

A framework for job scheduling and cluster resource management.

Hadoop MapReduce

A YARN-based system for parallel processing of large data sets.

Other Hadoop-related projects

Status Answered

Your score 5 / 5 100%

Response

Match the Term with the definitions:

	HBase	Hive	Spark	ZooKeeper	Pig
A high-level data-flow language and execution framework for parallel computation.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
A high-performance coordination service for distributed applications.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
A fast and general compute engine for Hadoop data.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
A scalable, distributed database that supports structured data storage for large tables.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
A data warehouse infrastructure that provides data summarization and ad hoc querying.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Knowledge Questions

To complete this section, you need to use the material from further reading and external sources.

Netflix Algorithms

Status Answered

Your score 4 / 4 100%

Response

Carefully study the further reading case study on Netflix. Match the algorithms used at Netflix with their purpose.

	Sims	Top N	PVR	Evidence
decides which image for the same video depending on the user	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
forms the Because you Watched row	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
forms the Top Picks row	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
orders the entire catalog in personalized way	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

☒ A/B tests

Status Answered

Your score 1 / 1 100%

Response

What is a common practice when A/B test results show retention wins without supporting engagement metrics increases?

- ☒ Assuming random variation and rerunning the tests.
- ☐ Accepting the results without further analysis.
- ☐ Implementing the changes immediately based on retention wins.
- ☐ Ignoring the results and moving on to other tests.

☒ A/B test & Netflix

Status Answered

Your score 1 / 1 100%

Response

1. 1. What challenge in interpreting A/B test results did Netflix identify in the paper?

- ☒ The presence of uncertainty and subjectivity
- ☐ The lack of statistical sophistication in test design
- ☐ The absence of alternative metrics
- ☐ The overreliance on retention metrics

☒ Amazon prediction model

Status Answered

Your score 3 / 3 100%

Response

Which Amazon model predicts which products customers are likely to purchase and when, distributing them to local distribution centres?

- ☐ AI assistant provider
- ☒ Anticipatory Shipping model
- ☐ Personalised recommendation system
- ☐ Supply Chain optimisation

☒ Analytics at DHL

Status Answered

Your score 0 / 4 0%

Response

How are the 4 types of big data analytics applied across the supply chains in DHL? Match the types of analytics with the examples of their application.

	Descriptive	Prescriptive	Predictive	Diagnostic
revealing if roller cages are broken based on data from sensors	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
helping logistics leaders find patterns	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
calculating the risk of lane disruption	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
ensuring a better price point	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

☒ IBM Watson

Status Answered

Your score 4 / 4 100%

Response

Carefully study the further reading case study on IBM Watson.

	Cognitive System	Deep NLP	Corpus	Confidence Score
Helps systems simulate human-like understanding by decomposing ideas and applying reasoning.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Analyzes multiple layers of language and meaning, focusing on accuracy over surface precision.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Assigns a probability-based value to how likely a candidate answer is correct.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

A knowledge base of unstructured data Watson uses to generate responses.



Test execution

Information

↻ Max. attempts: Unlimited

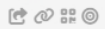
👁 Results of this test are visible to administrators and tutors of this course.

Start test

► Change log

[^ Go to top](#)

Logged in as *Ravi Himmatbhai Ramani* (1443 People are online)



Imprint
Datenschutzerklärung

OpenOlat 19.1.14

