

Syllabus

Data Engineering 2: Cloud Computing

- **Instructor:**
 - Zoltan C. Toth
 - TothZ@ceu.edu
 - +36 20 254 4934
 - Office: by appointment
- **Credits:** 2 (4 ECTS)
- **Term:** Fall 2022-2023
- **Course level:** Master's
- **Prerequisites:**
 - Data Analysis 1: Exploration – Business Analysis track
 - SQL Knowledge (Data Engineering 1: SQL & Different Shapes of Data Course)
 - Linux command-line knowledge (Data Engineering 1: SQL & Different Shapes of Data Course)
- **Course drop:** Course can be dropped free of charge 24 hours after the first session. After this date drop is possible until the course is halfway over (late drop fee applies). No changes are allowed past that date.

1. COURSE DESCRIPTION

This is a technology-focused course on cloud computing and cloud-based data analytics tools.

Current Data Analytics Architectures often work with an amount of data that cannot be fit on a single computer; furthermore, companies tend to avoid large upfront investments in hardware. Even companies that work with reasonably small datasets expect rapid growth, so they prefer to use data analytics solutions that scale when needed. In this course, you will get an overview and hands-on experience with Amazon Web Services, one of the leading cloud solutions. You will also see how cloud computing can help you quickly iterate and scale your data analytics infrastructure and how it can help you reduce operational costs.

The course includes a starter lecture where students can familiarize themselves with client-server architectures, secure internet communication, and digital signatures.

2. LEARNING OUTCOMES

Key outcomes:

At the end of this course, you will have an overview of Cloud technologies applied in modern businesses. You will have a general understanding of how these technologies work, and you will be able to reason about when to use or not to use them. You will be hands-on with Amazon Web Services.

Once you complete the assignments for this course, you will be hands-on with the following technologies:

- Internet Basics
- Public Key Encryption, Digital Signatures
- Fundamental Cloud Computing concepts: Storage, Virtual Machines
- Serverless Services for image and text recognition
- Amazon Athena: An SQL-based analytics tool in the cloud.

Other outcomes. The course will also help develop skills in the following areas:

Learning Area	Learning Outcome
Critical Thinking	You will be able to reason about the do's and don'ts of Cloud systems.
Hands-on Technology Skills	SSH, Amazon Web Services, Amazon S3, Amazon EC2, Amazon Athena

3. READING LIST

There is no compulsory reading for the class. Presentations and the source code for the exercises will be provided.

Databases: The CEU Library boasts a range of databases covering financial and company data, market and industry reports, global news and more. For a full list of databases visit the [CEU Library](#).

- Refinitiv (Thomson Reuters) Eikon for Students + Datastream/Thomson ONE
 - Eikon: Platform used by finance practitioners including market traders to monitor and analyze financial information. Information, analytics and news on all major financial markets including real-time pricing data, financial research, global financial news and commentary, financial estimates, fundamentals analysis, visual analysis through charting. Import/export from Excel.
 - Datastream: Range of economic, securities and company financial data. Excel add-in.
 - Thomson ONE: Global overviews on 55,000 public companies, one million private companies. Reuters News, ownership, deals, private equity, key ratios, company filings, officers and

directors. Investext analyst reports, active and historical research from 1,600 independent research firms, brokerages, investment banks.

- Standard & Poor's Capital IQ
 - Web and Excel-based platform combining deep global company information, credit ratings and research, and market research with powerful tools for risk assessments. Real-time and historical information on markets, industries, companies, transactions and people. Tearsheet data.
- Lexis Nexis Academic
 - Global database of news, business, legal and other sources. Full text of 350 newspapers, 300 magazines and journals, 600 newsletters. Wire services including Associated Press, Business Wire, and PR Newswire. Company financial information, market research, industry reports.

4. TEACHING METHODS AND LEARNING ACTIVITIES

The course will involve a mix of presentations, discussions, and practical sessions.

Learning objectives will be achieved through in-class discussions, reviewing the course materials, and solving homework.

5. ASSESSMENT

- Assignments (60%)
- Exam (40%)

Grading Policy

Students shall not miss more than two lectures. Failing to do so will yield an administrative fail grade.

To pass, students will need to get at least 60% of the overall grade. Failure to do so will yield a Fail grade.

6. TECHNICAL/LAPTOP REQUIREMENT

Laptops with a working browser are required for this class.

7. TOPIC OUTLINE AND SCHEDULE

Session	Topics	Readings
1	Basics of Internet Communication, Encryption on the Internet, Digital Signatures	

2	Basics of Cloud Computing using Amazon Web Services (AWS): Storage and Virtual Machines	
3	Using AWS programmatically. Serverless solutions in the cloud	
4	Amazon Athena: Scalable Data Lakes and SQL in the Cloud	

8. SHORT BIO OF THE INSTRUCTOR

Zoltan Toth holds computer science and math master's degrees from Vrije Universiteit, Amsterdam and ELTE, Budapest. He used to be the founder and CTO of a Data Infrastructure Service and Training company, Datapao. These days he works as a contract data engineering and machine learning operations architect. Zoltan worked in Data Engineering and Management roles in global startup companies like Prezi and RapidMiner.