Training Schedule

on

Fundamentals and Technologies for Big Data Analytics

Priv.Doz. Dr. Hong-Linh Truong
Email: truong@rdsea.io
Web: http://truong.rdsea.io/

Overall:
- fundamentals and technologies for data science, including architectures, communications, data sources, and data management.
- fundamentals and technologies for big data processing for *data at rest* and *data in transit/motion.*

Topic	Content	Time
Lecture 1: Basic key concepts of data science/big data analytics in distributed systems	Understand key concepts of data science, big data and their relation to distributed systems (clocks, synchronization, scalability, load-balancing, asynchronous communication, virtual machines and containers)	1 hour
Lecture 2: Distributed architectures models for data analytics	Understand different models of distributed architectures (client-server, services, P2P), types of distributed systems (clusters, clouds, and mobile-edge-computing); key characteristics of cloud computing; IoT and cloud platforms for data analytics	2 hours
Lecture 3: Programming Models and Languages for Big Data Analytics	Understand main programming models and languages for big data analytics, including client-server, web services, workflows and parallel programs in Java, Python, R, and NodeJS	1 hour
Lecture 4: Communications	Overview middleware, protocols and connectors. Understand existing middleware, their supporting protocols (e.g., MQTT, AMQP, Kafka, REST, etc.) and the connector/integration concepts	2 hour
Lecture 5: Scalable data models and database technologies	- Overview data models and database technologies for big data, e.g., document-based, key-value based, columnar-based, and graph-based databases - Discuss examples/scenarios where such data models and database technologies are used.	1 hour
Lecture 6: Data Ingestion	- Understand data source characteristics (IoT, enterprises, logs, size of data items, rate of data ingestion, interfaces, etc.) - Understand architectures of data ingestion systems and ingestion policy - Understand scalability models for architecture for ingestion/sub/storage	1 hour
Lecture 7: Complex	Study complex event processing models	2 hours

Event Processing	• Overview complex event processing technologies • Work with examples using WSO2/Flink framework	
Lecture 8: Streaming Data Analytics	• Understand streaming data processing models • Study existing streaming data processing technologies • Work with examples from Apache Storm	2 hours
Lecture 9: MapReduce/Hadoop	• Study fundamental concepts of MapReduce for big data analytics • Examine MapReduce tools and technologies • Develop simple MapReduce applications with Hadoop	2 hours
Lecture 10: Workflows	• Study workflow models used for offline/batch processing. • Examine workflow tools and engines. • Develop workflows to analyze data using Airbnb/Apache Airflow	2 hours

Key requirements

It is a general course in the training series on IoT Cloud Platforms and Big Data Analytics. We aim at providing a general concept and foundations. Therefore, the minimum requirement for a participant would be that she/he has a university/college-level background on computer science, software engineering and networks. For those, who have not dealt with data analytics programming in a daily basis but need to work with IoT and data analytics, we do not need other skills to understand the core concepts in order to take these concepts into their work. For those, who need to develop analytics software, we expect them to have programming skills with Java (additional programming languages could be NodeJS/Python/R).

Instruction Languages

Slides will be provided in English (samples of code will be provided for certain parts of the course) the instruction language will be in English or Vietnamese.

Other notes:

Course participants are encouraged to bring their own laptop with Internet access. Course participants should send their access information (username/email) so that they can access supplement materials.