Training Concept & Schedule

Big Data Analytics for Communication Network Operations

Priv.Doz. Dr. Hong-Linh Truong
Email: truong@rdsea.io
Web: http://truong.rdsea.io/
Course supplement materials (code and data) in bitbucket.org

Overall: focusing on cases and applying suitable solutions to perform big data analytics.

Topic	Content
Lecture 1: Big data analytics with Batch Processing Models, infrastructures and platforms	• Big data processing with batching processing models • Overview of big data infrastructures with Hadoop ecosystems • Preparing Hadoop ecosystem systems
Lecture 2: Scenarios and Overall Concept of Big Data Analytics for Mobile Network Performance	• Scenarios with Mobile Communication Networks Management • Data sources and Requirements
Lecture 3: Tools for Managing Big Data	• Hadoop Ecosystems & Tools • Spark Ecosystems & Tools • Hive and BigQuery
On-site working group and self-study	• *Working with real Hadoop/Spark systems in the cloud*
Lecture 4: Data Governance for Large-scale Datasets	• Datalake view and Data Storage Strategies • Data Governance Techniques
Lecture 5: Data Extraction and Preparation	• Data wrangling • Data ingestion for data mining • Automatic data ingestion with LogStash & Apache NiFi
Onsite working group & Self study	• *Working on specific examples of data governance for telco operational data* • *Working with data extraction tool*
Lecture 6: Data Mining Overview	• Big data mining overview

Lecture 7: Analytics Case Development	• A set of small lectures
	  ○ Analytics problem and requirements
	  ○ Data mining techniques to be used
	  ○ Data processing models to be used
	  ○ Solution development and analytics

Onsite working group and self-study	• *Working on specific examples of data mining with mobile data for network operations*

Lecture 8: Analytics Services	• Develop Web services Interacting with Data Analytics
Onsite working group and self-study	• *Review cases & Discussions* • *Review the whole proof-of-concept big data analytics system for mobile network performance* • *Putting things together/Review*

Key requirements

We should have Hadoop, Spark and Python and suitable tools for the case development and exercise. Lectures 3, 5, 7 and 8 require technical people who are key members of the big data development team.

Instruction Languages

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

System Preparation for Training

1 Overview

For this training, we will use Hadoop and Spark systems for data processing as well as other tools for data wrangling and management. Our main programming language are Python and Java. In this tutorial, we will use a pre-installed Hadoop/spark systems using the Google cloud platform for practical. Furthermore, we also use BigQuery, Apach Nifi, and many other tools.

2 Requirements

- Internet **access MUST be available**. It is expected to have strong network to the internet.
- Participants need **to send their google email address to the instructor**, if they want to run examples with the real system. A group of participants can use one account to do work together.
- Participants **need to setup Google Cloud SDK** (https://cloud.google.com/sdk/docs/) which provides useful tools to do the work with google cloud.
- It highly recommends to install Apache NIFI (https://nifi.apache.org/) so that course participants can play around with Apache NIFI by her/himself (a Nifi server is provided in Google cloud for all participants)
- Python and data mining packages, pandas (http://pandas.pydata.org/) and scikit-learn (http://scikit-learn.org/stable/), should be installed so that course participants can do some self-study of data mining with her/his own machine.

3 Notes

Linux-based laptops are highly recommended for doing the course work.

