Data Research Centre Campus Fryslân and Frankfurt Big Data Lab presents:

Big Data and Data Science

A hands-on introduction

Monday 25 September till Thursday 28 September

Tresoar
Boterhoek 1
8911 DH Leeuwarden
This short course will teach some of the main technologies for Big Data processing (data analytics) and storage. It will present a mix of theoretical lectures and practical hands-on labs. The main topics that will be covered are:

- Introduction to Big Data,
- Hadoop Ecosystem (HDFS, MapReduce, Hive, Sqoop, Flume, etc.),
- Apache Spark,
- Introduction to IPython Notebook.

With knowledge acquired the participants will be grouped in small teams with the goal to implement a real use case, using open transportation data.

**Participant Skills Requirements:**
The course is intended for professionals who have a basic practical knowledge of IT, in particular:

- Basic knowledge in any programming languages such as Java, Python, Scala, C# or similar.
- Familiarity with Linux/Unix shell commands.

**Hardware and Software Requirements for the course:**
VMware Player or Oracle VirtualBox. (VM image with prepared exercises will be provided for the practical sessions.)

- Computer with at least 8 GB of main memory (RAM) and 200 GB of storage.
- One computer per participant is optimal.

---

**Introduction to Big Data**

- What is Big Data?
- Overview of the Big Data Ecosystem

**Hadoop Ecosystem**

- HDFS & MapReduce
- Sqoop & Flume
- Pig & Hive
- Practical Exercises

**Apache Spark**

- Basics and Spark Libraries
- Practical Exercises

**Introduction to Data Science**

- Acquiring and profiling data, tools and community
- Motivational use cases using Ipython Notebook
  - Clustering IRIS flowers data set
  - Spam Email detection
  - Energy consumption prediction
- Hands-on projects using pySpark
  - Kmeans - visualization taxi on a map
  - Classification - taxi tip classification
  - Regression - traffic planning
- Presentations of the projects by the participants

**Instructors:** Todor Ivanov, Kim Hee. Frankfurt Big Data Lab, Goethe University, Frankfurt, www.bigdata.uni-frankfurt.de

---

**Todor Ivanov**

Todor Ivanov is a research assistant and PhD student at the Frankfurt Big Data Lab, Goethe University Frankfurt. He received his BSc. in Computational Engineering and MSc in Distributed Software Systems from Technical University of Darmstadt. His main interests are in the design and benchmarking of complex distributed software systems for big data and data-intensive applications. Prior to that, he has worked as a senior software engineer developing Flight Information Display Systems (FIDS) for different international airports, research assistant in the FlashyDB project at the department of Databases and Distributed Systems, TU Darmstadt and IT consultant in business intelligence and database related projects in the bank sector.

---

**Kim Hee**

Kim, Hee is a research assistant and PhD student at the Frankfurt Big Data Lab, Goethe University Frankfurt. She received her BSc. in Media Engineering from Ajou University in South Korea and MSc in Telecommunication and Media Engineering from University of Applied Science Offenburg in Germany. In general, her main interests are data analytics, particularly data blending. Prior to that, she has worked as a software engineer in Samsung in Manufacturing Execution System (MES). She has received an Oracle Certified Professional (OCP) in 2012.
Directions and parking

Address (A)
Tresoar
Boterhoek 1
8911 DH Leeuwarden

Parking (B)
Parking Garage Oldenhove
Oldehoofsterkerkhof 3
8911 DH Leeuwarden