Lowering the Barriers to Big Data Solutions
Big Data Workshop

Ivan Ermilov

Leipzig University
Institut für Angewandte Informatik
October 9, 2017

Abstract

"Lowering the Barriers to Big Data Solutions" workshop will present the results achieved throughout the BigDataEurope (BDE) project, which is the part of the European Union's Horizon 2020 research Europe flag and innovation program. The workshop will concentrate on the practical aspects of developing and maintaining Big Data applications. The applications will be deployed on the Big Data Integrator platform, a modular platform for deploying Big Data applications, which is developed in the BDE project. The workshop will cover the theory behind Big Data applications and will provide the explanation of the key decisions for the Big Data Integrator platform architecture. Each theorethical part will be followed by a hands-on session. The participants will be invited to accomplish assignments, which will teach them how to solve Big Data problems using Spark and BDI platform.

Keywords: Big Data, Big Data Applications, Software Lifecycle, DevOps

Description

The "Lowering the Barriers to Big Data Solutions" workshop will present the practical results of the BigDataEurope project[1], which aims to simplify the usage of Big Data solutions such as Hadoop, Spark, Flink etc. In particular, it will be shown how to:

---

[1] https://www.big-data-europe.eu/

- develop Big Data clusters (BDI Stacks) on Big Data Integrator platform from scratch,

- deploy BDI Stacks locally and on a remote server,

- modify existing BDI Stacks to suit your needs (e.g. adding MongoDB to a BDI Stack),

- develop Spark applications,

- run Spark applications locally and on a remote cluster.

Requirements for the participants

To be able to participate in the hands-on sessions during the workshop you will need a PC (preferably Ubuntu/Debian) with installed Docker engine[2]and docker-compose[3]. You will also be required to download a number of docker images from Docker Hub, the full list will be announced two weeks before the workshop.

Provided materials

The presentations, all the instructions and code will be publicly available on a dedicated Github repository. The repository will be announced two weeks before the workshop. You are invited to contribute, open issues and ask questions there after the workshop. The final materials will be published after the workshop day.

Timetable

- **9:00-10:00** Introduction: Big Data Europe Project, Architecture, Components and Interfaces

- **10:00-11:00** Planning BDI Stack for Your Application

- **11:00-11:15** Coffee Break

- **11:15-11:45** Introduction to Docker

- **11:45-12:15** Hands-on: Configuring Docker and Docker Swarm Cluster

---

[2] https://docs.docker.com/engine/installation/linux/docker-ce/ubuntu/

[3] https://docs.docker.com/compose/install/

- **12:15-13:00** Hands-on: Deploying BDI Stack Locally and on Remote Server[4]

- **13:00-14:00** Lunch

- **14:00-15:00** Methodologies for Developing a Big Data Applications

- **15:00-16:30** Hands-on: Developing and Deploying a Big Data Application

- **16:30-17:00** Conclusions, Closing Remarks, Q&A Session

Outcomes

By the end of this workshop attendees will be able to:

- install and configure Docker locally and as Docker Swarm cluster,

- plan deployment environment for Big Data Applications (i.e. BDI Stack),

- deploy BDI Stacks on Docker and Docker Swarm cluster,

- understand the lifecycle of Big Data Applications,

- develop Spark applications and package them into Docker images,

- deploy Big Data Applications as part of BDI platform.

About Presenter

Ivan Ermilov is a doctoral candidate working on automatic conversion of Web Tables to semantically rich representations. Contributes to open source Semantic Web community and making Big Data simpler to use.

Ivan Ermilov has been part of Agile Knowledge and Semantic Web (AKSW) research group for the last six years contributing to the projects related to data integration, analytics, semantification of Big Data, benchmarking and education. For the last two years, Ivan has been working on Big Data Europe (BDE) Horizon2020 project. In this project, he contributed to the development of platform architecture and adoption of the Big Data components for the platform as well as development of user interfaces and integrations. The complete list of publications and projects can be found on the AKSW research group website[5].

---

[4] There should be 20 minutes break for praying during this session.

[5] http://aksw.org/IvanErmilov