

METADATA (*)

TOPIC B – Training Unit 2: Digital Platform for Process Industries

Source:

Partner: **SCCH** – Lesson 1 and 2

Partner: **SYXIS** – Lesson 3 and 4

Project: TRINEFLEX - Transformation of energy intensive process industries through integration of energy, process, and feedstock flexibility, Grant agreement ID: 101058174

The contents of lessons 1 and 2 were developed using the following sources:

- S., Mayr, M., Chasparis, G. C., & Pichler, M. (2021). Avubdi: A versatile usable big data infrastructure and its monitoring approaches for process industry. *Frontiers in Chemical Engineering*, 3. <https://doi.org/10.3389/fceng.2021.665545>
- Himmelbauer, J., Mayr, M., Luftensteiner, S. (2022). From Data to Decisions - Developing Data Analytics Use-Cases in Process Industry. In *Database and Expert Systems Applications - DEXA 2022 Workshops*. DEXA 2022. *Communications in Computer and Information Science*, vol 1633. Springer, Cham. https://doi.org/10.1007/978-3-031-14343-4_8
- <https://www.alibabacloud.com/knowledge/what-is-containerization>
- <https://www.cometchat.com/blog/cloud-vs-on-premise-deployment>
- https://dev.to/ezinne_anne/containers-what-is-containerization-and-container-orchestration-38pe
- <https://kafka.apache.org/>
- <https://spark.apache.org/>
- <https://www.influxdata.com/>
- <https://grafana.com/>
- <https://kubernetes.io/>
- <https://www.docker.com/>

The contents of lessons 3 and 4 were developed using the following sources:

- Cassina J, Critelli I, Coscia E, Borgia S., Tools and Procedures to Embed and Retrieve Product-Service Lifecycle Knowledge. In: *Models, Methods and Tools for Product Service Design: The Manutelligence Project*, 63-81
- T.Koc, E.Bozdog (2017). Measuring the degree of novelty of innovation based on Porter's value chain approach . <https://doi.org/10.1016/j.ejor.2016.07.049>
- <https://kafka.apache.org/>
- <https://www.ibm.com/storage>
- <https://www.hpe.com>
- <https://redhat.com>
- <https://kubernetes.io/>

Ownership

Michael Mayr (SCCH) - Lesson 2.1 (State-of-the-art in Big-Data Technologies) and 2.2 (Deployment and configuration of Big-Data Solutions)

Paolo Perillo (SYXIS) - Lesson 2.3: (Interfacing Data Infrastructure) and 2.4: (Data Storage)

The training material of lessons 2.1 and 2.2 is provided under Creative Commons Attribution Share-Alike License

<https://creativecommons.org/licenses/by-sa/4.0/deed.en>

The Lessons 3 and 4 are NOT provided under Creative Commons Attribution Share-Alike License

Abstract

It has recently been recognized that machine learning and data analytics play a critical role in realizing long-term sustainability goals in process industry. Indeed, industrial digitization is expanding at a relatively fast pace, given that sensors are getting faster, better and cheaper. Such a transformation requires substantial innovation in the context of Big Data management and analytics. In this training unit we will introduce the state-of-the-art Big Data technologies, interfacing data infrastructures and data storage solutions.

Structure

- Lesson 1: State-of-the-art in Big-Data Technologies
This lesson provides an overview of the State-of-the-Art Big Data technologies in process industry.
- Lesson 2: Deployment and configuration of Big Data Solutions
This lesson provides an overview of general practices in the deployment and configuration of Big Data solutions.
- Lesson 3: Interfacing Data Infrastructure
General Notes on Data Infrastructures, Project scenarios, Why interfacing data infrastructures, architectures of a system for gathering, streaming, storing and dispatching data from different infrastructures
- *Lesson 4: Data Storage*
Definitions, Types of Storage solutions, Storage needs in a project, Storage models and architectures in a project

Learning Outcomes

For Lesson 1, participants will learn about the general concept of Big-Data in the context of industrial processes and the challenges associated. In addition, state-of-the-art technologies are discussed for retrieving, processing, persisting, and visualizing such data, and a potential infrastructure is shown.

For Lesson 2, participants will learn about on-premises and cloud deployments, as well as the importance of containerization and orchestration methodologies. In addition, the deployment and configuration of such systems is discussed.

For Lesson 3, participants will learn about Data infrastructure basic components, about key needs and requirements for interfacing infrastructure in projects and in production, and about architectures of systems for interfacing infrastructures.

For Lesson 4, participants will learn about Storage basic components and their evolution to Cloud Storage, about storage models used in projects and about enabling software to manage data storage infrastructures.

Intended Audience

This training unit is designed for researchers and decision-makers and should serve as a basis for further discussions on technological-, infrastructural- and modelling-related decisions.

The training for Lessons 3 and 4 is also conceived to provide the consortium partners and use cases an overall knowledge base.

Pre-requisites

The training is high-level and requires no in-depth knowledge.

Language: English

Format: Video mp4, PDF

Expected workload: approximately 10 minutes for each lesson.

Complementary additional training material:

Additional in-depth information relevant for Lesson 1 and 2 can be found in the following two paper:

- Luftensteiner, S., Mayr, M., Chasparis, G. C., & Pichler, M. (2021). Avubdi: A versatile usable big data infrastructure and its monitoring approaches for process industry. *Frontiers in Chemical Engineering*, 3. <https://doi.org/10.3389/fceng.2021.665545>
- Himmelbauer, J., Mayr, M., Luftensteiner, S. (2022). From Data to Decisions - Developing Data Analytics Use-Cases in Process Industry. In *Database and Expert Systems Applications - DEXA 2022 Workshops. DEXA 2022. Communications in Computer and Information Science*, vol 1633. Springer, Cham. https://doi.org/10.1007/978-3-031-14343-4_8

(*) The structure of the Metadata for the Training Units derives from the training Metadata model developed within the Leonardo da Vinci project LINKVIT (2013-15, GA N. 2013-IT1-LEO05-04046)