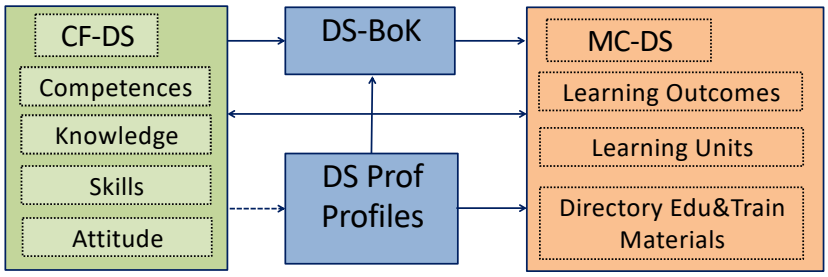
Distributed and Parallel Programming Curriculum Design using EDISON Data Science Framework Methodology

EDISON Data Science Framework (EDSF) Release 4



EDISON Data Science Framework (EDSF) provides conceptual basis for the Data Science profession

- CF-DS – Data Science Competence Framework
- DS-BoK – Data Science Body of Knowledge
- MC-DS – Data Science Model Curriculum
- DSPP - Data Science Professional Profiles
- Data Science Taxonomy and Scientific Disciplines Classification

Data Science Body Of Knowledge (DS-BoK)

DS-BoK Knowledge Area Groups (KAG) are defined in compliance with the CF-DS

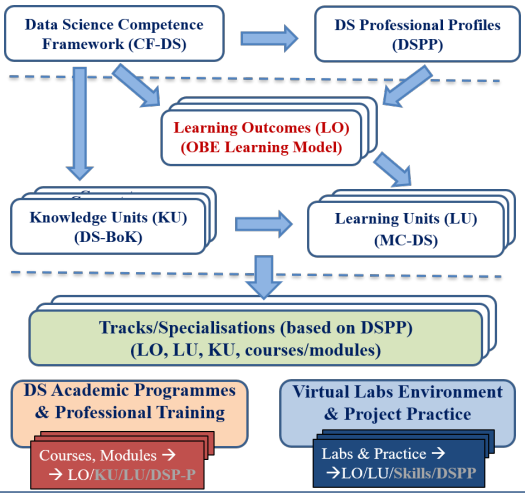
- KAG1-DSDA: Data Analytics group including Machine Learning, statistical methods, and Business Analytics
- KAG2-DSENG: Data Science Engineering group including Software and infrastructure engineering
- KAG3-DSDM: *Data Management group including data curation, Research Data Management, Open Data, Open Access*
- KAG4-DSRMP: *Research Methods and Project Management, use cases/practices*
- KAG5-DSBPM: Business Process Management (data driven)

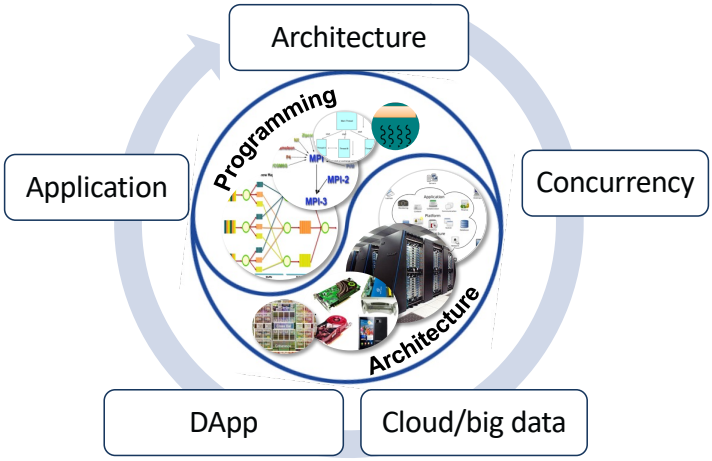
Includes selected KAG and KU defined in IEEE/ACM CCS (2012), PMI-BoK, BABOK, SWEBOK, DM-BoK, ACM CS-BoK

Distributed and Parallel Programming (DPP) is a course for the second-year bachelor of Informatica (computer science) at the college of computer science, University of Amsterdam. During the past years, DPP (called Concurrent and Parallel Programming before 2021) delivered a spectrum of up-to-date technical topics across parallel and distributed domains and attracted lots of student interest. In DPP, students will learn to write programs in various emerging architectures and programming paradigms, including multi-core, cluster, Cloud, and decentralized environments. Students will learn to reflect, design, and use parallel and distributed systems.

Outcome Based Educations and Training Model:

From Competences and DSP Profiles to Learning Outcomes (LO) and to Knowledge Units (KU) and Learning Units (LU)





The course is structured into **five chapters**: 1) Systems Architecture and Performance, 2) Parallel and distributed programming (Multi-threading, GPU Programming, and MPI), 3) cloud and Big Data systems, 4) Decentralized systems, and 5) Applications. The course features several assignments: 1) Literature study, 2) Multi-threading with Pthreads and OpenMP, 3) GPU programming with Cuda, 4) Distributed Programming with MPI, 5) Cloud and Big data, and 5) Distributed Ledgers and smart contract.

Competence-Driven Curriculum Design outcome

- Data Science curriculum for different types of learners and professional profiles
- Virtual Labs and Project Development Environment that can combine local university facilities and use cloud-based Big Data and Data analytics facilities and services on demand

Distributed and Parallel Programming course at UvA
Contact: Zhiming Zhao <z.zhao@uva.nl>