

Big data

1. **Course number and name:** 020BIGES4 Big data
2. **Credits and contact hours:** 4 ECTS credits, 2x1:15 contact hours
3. **Instructor's or course coordinator's name:** Dany Mezher
4. **Text book:** Mining of Massive Datasets, Jure Leskovec, Anand Rajaraman, Jeffrey Ullman
 - a. **Other supplemental materials:**
Course slides, E-learning support using Moodle, MOOC on coursera.org
5. **Specific course information**
 - a. **Catalog description:**
Introduction to massive data challenges, High Performance File System and MapReduce, Link Analysis in Graphs, Similar Sets, Similar Item sets, Community detection in graphs, mining data streams, recommender systems, Clustering and classifiers.
 - b. **Prerequisites or co-requisites:**
 - c. **Required:** Elective for CCE students
6. **Specific goals for the course**
 - a. **Specific outcomes of instruction:**
 - The student will be able to analyze big data problems and propose adequate solutions
 - The student will be able to apply the appropriate techniques and skills to solve big data problems.

b. KPI:

KPI	a2	b1	b2	b3	c1	c3	I2	j1	k1	k2	k3
Covered	x	x	x	x	x	x	x	x		x	x
Assessed	x				x	x			x	x	x

7. Brief list of topics to be covered

#lectures	Description
1	Introduction: Massive Data Challenges
3	HPFS and MapReduce
1	Lab. MapReduce
2	PageRank and link analysis in graphs

1	Lab. PageRank
3	Similar sets & similar item sets
1	Lab. Similar sets
1	Lab. Similar item sets
3	Community detection in graphs (Clustering and BigClam)
1	Lab. Community detection (clustering)
1	Lab. BigClam
2	Mining data streams
1	Lab. Mining data streams
3	Recommender systems
1	Lab. Recommender systems
2	Clustering
1	Lab. Clustering