Big Data Management

**General Instruction**: Each student reads one paper in advance and discuss the material in groups during the meeting. In the study group meeting, all students who read the same paper first sit together to discuss the questions of each paper. Then the students are shuffled into different groups to present those papers to other students who do not read those papers. The assignment of paper reading and presentation group is shown in the following table. All five papers and questions are also described as follows.

No.	Name	Reading Paper No.	Presentation Group
1	Alcantara Beltran, Jose	1	1
2	Bukonte, Laura	1	2
3	De Leydet, Rémy	1	3
4	Denis, Devin	1	4
5	Gafurova, Lina	1	5
6	Ghasemi, Mandana	1	1
7	Goetsch, Peter	1	2
8	Hamberg, Jonatan	1	3
9	Harju, Esa	1	4
10	Huang, Biyun	1	5
11	Ivanova, Sardana	1	1
12	Jaakkola, Kasperi	1	2
13	Jitta, Aditya	2	3
14	Kangassalo, Lauri	2	4
15	Kapoor, Shubham	2	5
16	Karvanen, Jami	2	1
17	Khan, Nazmul	2	2
18	Kieloaho, Antti-Jussi	2	3
19	Korpinen, Kari	2	4
20	Kronser, Andrew	2	5
21	Laakso, Antti	2	1
22	Lehikoinen, Ilkka	2	2
23	Li, Yilin	2	3
24	Liljeblad, Nina	2	4
25	Maristo, Ilkka	3	5
26	Martikainen, Jussi-Pekka	3	1
27	Mesimäki, Jerry	3	2
28	Murtaza, Adnan	3	3
29	Myllyoja, Aleksi	3	4
30	Mäklin, Tommi	3	5
31	Noykova, Neli	3	1
32	Osmani, Lirim	3	2
33	Panchamukhi, Sandeep	3	3
34	Pereira, Patricia	3	4
35	Pollak, Clemens	3	5
36	Rantanen, Cedric	3	1
37	Raut, Bidur	4	2
38	Roy, Suravi	4	3
39	Saitkulov, Marat	4	4

40	Salmi, Joni	4	5
41	Salosensaari, Aaro	4	1
42	Santana Vega, Carlos	4	2
43	Soisalon-Soininen, Eliel	4	3
44	Sore, Shewangizaw	4	4
45	Soyoye, Fiyinfoluwa	4	5
46	Szkalisity, Ábel	4	1
47	Tiirikainen, Suvi	4	2
48	Ture, Tsegaye	4	3
49	Tähtinen, Sara	5	4
50	Wahlroos, Kristian	5	5
51	Wahlroos, Mika	5	1
52	Wallenius, Otto	5	2
53	Viiri, Kalle	5	3
54	Vähämaa, Ilkka	5	4
55	Xu, Pengfei	5	5
56	Zhang, Chao	5	1
57	Zhou, Ziye	5	2
58	Zhukova, Bella	5	3
59	Zuñiga Corrales, Wladimir	5	4
60	He, Chen	5	5
61	Wen, Guo	5	1
62	Mikael Snellman	5	2
63	Jonne Airaksinen	5	3

Paper 1. Cheikh Kacfah Emani, Nadine Cullot, Christophe Nicolle: Understandable Big Data: A survey. Computer Science Review 17: 70-81 (2015)

Download: https://www.cs.helsinki.fi/u/jilu/paper/bigdatasurvey01.pdf

Advance reading: Read at least Sections 1–4 of the above paper before the study group session.

**Topics for Discussion**: Discuss at least the following topics in your group. Prepare to summarize the discussion for members of the other groups.

- What are 4V's of big data?
- What are 4A's of big data analysis?
- What is the Named entity resolution (NER) ? Give several examples.
- What is coreference resolution?
- What is billion triple challenge?
- What are three main steps in data integration? Explain the meaning of each step using an example.

Paper 2. *H. V. Jagadish: Big Data and Science: Myths and Reality. Big Data Research 2(2): 49-52 (2015)* Download: <u>https://www.cs.helsinki.fi/u/jilu/paper/bigdatasurvey02.pdf</u>

Advance reading: Read the above paper before the study group session.

Topics for Discussion: Discuss at least the following topics in your group. Prepare to summarize the discussion for members of the other groups.

- Why are Variety and Veracity more challenge than Volume and Velocity?
- Why does big data pose huge challenge for human interaction?
- Why are the steps in the big data analysis pipeline?
- What are the challenges for data reuse?
- What are the primary differences between "data science" and "big data"?

Paper 3. Ruogu Fang, Samira Pouyanfar, Yimin Yang, Shu-Ching Chen, S. S. Iyengar: Computational Health Informatics in the Big Data Age: A Survey. ACM Comput. Surv. 49(1): 12 (2016)

Download: https://www.cs.helsinki.fi/u/jilu/paper/HealthBigdata.pdf

Advance reading: Read at least Sections 1–3 of the above paper before the study group session.

**Topics for Discussion**: Discuss at least the following topics in your group. Prepare to summarize the discussion for members of the other groups.

- What are three scenarios of big data in health informatics?
- Explain 4V's of big data in health informatics.
- Give several concrete examples for big data analytics in health informatics.
- Give several examples for healthcare data types and sources.
- What data mining algorithms can be used to discover knowledge from large datasets?
- What is Computer-assisted intervention (CAI)? How to apply big data technology on CAI?

Paper 4. Stéphane Marchand-Maillet, Birgit Hofreiter: Big Data Management and Analysis for Business Informatics - A Survey. Enterprise Modelling and Information Systems Architectures 9(1): 90-105 (2014)

Download: https://www.cs.helsinki.fi/u/jilu/paper/businessbigdata.pdf

Advance reading: Read at least Sections 1–4 of the above paper before the study group session.

**Topics for Discussion**: Discuss the following topics in your group. Prepare to summarize the discussion for members of the other groups.

- What is business informatics? Why can big data technology be used for business informatics?
- What are the five waves of information security?
- What is "query -by -example"?
- How to model trust and reputation over communication networks?
- What is crowdsourcing? Describe the main purpose of Amazon Mechanical Turk.

Paper 5. Renzo Angles, Claudio Gutiérrez: Survey of graph database models. ACM Comput. Surv. 40(1) (2008)

Download: https://www.cs.helsinki.fi/u/jilu/paper/graphDataModel.pdf

Advance reading: Read at least Sections 1–4 of the above paper before the study group session.

**Topics for Discussion**: Discuss the following topics in your group. Prepare to summarize the discussion for members of the other groups.

- What is db-model?
- Describe the evolution of db-model in Figure 1.
- Describe the graph db-model from three aspects: data structures, transformation language and integrity constraints.

- Compare graph db-model with other models, such as physical, OO, semi-structured db models.
- How to present relations in graph db-model? Describe examples in LDM, Groovy and GOOD.
- How to present integrity constraints? Give several examples.
- Give several examples for graph query and manipulation languages.