Autumn 2016

Big Data Management

Study group on 23.11

NoSQL databases

Department of Computer Science
University of Helsinki

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	1	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	2	5
26	Martikainen, Jussi-Pekka	2	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	3	2
38	Roy, Suravi	3	3

39	Saitkulov, Marat	3	4
40	Salmi, Joni	3	5
41	Salosensaari, Aaro	3	1
42	Santana Vega, Carlos	3	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	4	4
50	Wahlroos, Kristian	4	5
51	Wahlroos, Mika	4	1
52	Wallenius, Otto	4	2
53	Viiri, Kalle	4	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. *Jignesh M. Patel: Operational NoSQL Systems: What's New and What's Next? IEEE Computer 49(4): 23-30 (2016)*

Download: https://www.cs.helsinki.fi/u/jilu/paper/NoSQL01.pdf

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.

- *What are operational and analytical NoSQL systems?*
- *What are malleable schemas? How do NoSQL systems handle malleable schemas?*
- *What is flexible querying? Why do NoSQL systems need to support flexible querying?*
- *What are four new truths about database products according to companies like Google and Yahoo?*
- *Why are NoSQL databases better than RDBMS to handle rich data structures?*
- *What are ACID and BASE? What is the potential future work on consistency model in NoSQL systems?*
- *Why do NoSQL systems support "one size fits many"? What do "many" refer here?*
- *Is it possible to design a single benchmark that covers all users' cases of NoSQL databases? Why?*

Paper 2. Anita Brigit Mathew, S. D. Madhu Kumar: Analysis of data management and query handling in social networks using NoSQL databases. ICACCI 2015: 800-806

Download: https://www.cs.helsinki.fi/u/jilu/paper/NoSQL02.pdf

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 do social networks company need to use NoSQL databases?*
- *NoSQL databases can support unstructured, semi-structured and structured data. What are the meanings of unstructured, semi-structured and structured data?*
- *According to this paper, what are NoSQL databases used in Myspace, Facebook, Flickr and LinedIn and Foursquare?*
- *What are four categories of NoSQL databases? Describe the storage model of each type.*
- *What is concurrency control? Compare the different concurrency control strategies in MongoDB, CouchDB, DynamoDB, Riak, Neo4j and OrientDB.*
- *According to their experimental results, which graph database is the best in most cases?*

Paper 3. Alejandro Corbellini, Cristian Mateos, Alejandro Zunino, Daniela Godoy, Silvia N. Schiaffino:

Persisting big-data: The NoSQL landscape. Inf. Syst. 63: 1-23 (2017)

Download: https://www.cs.helsinki.fi/u/jilu/paper/NoSQL03.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.

- *When was the term "NoSQL" first coined? By whom, and what meaning?*
- *In NoSQL databases, what are two mechanisms used to achieve horizontal scaling? Give examples to illustrate these two mechanisms.*
- *What are four categories of NoSQL databases? List some database products for each category.*
- *What is the philosophy of "let it crash"? How does this philosophy affect the design of NoSQL databases?*
- *What is the meaning of "data affinity" in this paper?*
- *What is the CAP theorem? What are three properties which cannot be simultaneously guaranteed?*
- *What are two phases in the 2PC protocol? Why does 2PC negatively impact on system availability?*
- *What are "Read-repair", "Write-repair" and "Asynchronous-repair" in Cassandra?*
- *What is consistent hashing? Why Dynamo uses virtual nodes in consistent hashing?*
- *Briefly describe the main properties of the following key-value databases: Riak and Voldemort, Redis, Infinispan, Hazelcast and Membase.*

Paper 4. Joohyoung Jeon, Minjeong An, Hongchul Lee: NoSQL Database Modeling for End-of-Life Vehicle Monitoring System. JSW 10(10): 1160-1169 (2015)

Download: https://www.cs.helsinki.fi/u/jilu/paper/NoSQL04.pdf

Advance reading: Read 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 the PACELC theory?*
- *What are two requirements of databases in the end of life vehicle monitoring system?*
- *Give a brief introduction to MongoDB system.*
- *What are the main functions of the Infrastructure Layer and the Service layer?*
- *MongoDB does not support join operation. How is MongoDB used to process query in this paper?*
- *This paper compared the performance between MongoDB and MySQL. Which one has better performance for the operations of insert, find, join, and update?*

Paper 5. *Sugam Sharma: An Extended Classification and Comparison of NoSQL Big Data Models. CoRR abs/1509.08035 (2015)*

Download: https://www.cs.helsinki.fi/u/jilu/paper/NoSQL05.pdf

Advance reading: Read 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 the White House "Big data" initiative?*
- *Give a brief introduction to CouchDB.*
- *Give the comparison between Neo4j and OreintDB.*
- *Give the comparison between Riak and Oracle NoSQL.*
- *What is the main property of wide column store?*
- *Compare Cassandra, HBase and BigTable.*
- *Although NoSQL data models are gaining popularity, they are still far behind than the major relational data models in terms of the number of users (organizations). Why does it happen?*