CS4225/CS5425 BIG DATA SYSTEMS FOR DATA SCIENCE

Tutorial 2: NoSQL

**1.** Explain the pros and cons of NoSQL systems in the form of the BASE properties which they (often) satisfy.

**2.** The following question relates to the paper (not required reading for the class, but still a useful summary if you are interested):

*Rick Cattell. 2011. Scalable SQL and NoSQL data stores. SIGMOD Rec. 39, 4 (May 2011), 12-27.*

In the paper, they have shared suitable applications for key-value stores and document stores:

Application of key-value store: Application of document store:

Text

Description automatically generatedText

Description automatically generated

Discuss some factors that make these applications suitable for key-value stores and document stores respectively.

**3.** Assume that your e-commerce company’s webpage has a NoSQL document store database containing data about users visiting the webpage (e.g. IP address, country, time spent browsing the webpage, number of products they have bought, etc.) Your CEO suggests using the Range Partitioning scheme with the number of products they have bought as a partition key. State and explain a possible benefit of this choice, and a possible disadvantage of this choice.

**4.** Imagine you are developing a digital platform for NUS students to manage their courses and extracurricular activities they are taking, which also manages information about courses (e.g., professors, prerequisites, time-tables, etc), and aims to recommend courses for students.

Consider the NoSQL databases we have covered in the class, and choose one which is suitable for satisfying a specific use-case within the above digital platform. (You do not have to explain how to implement the entire system – you can choose a single use-case to focus on). Explain why your choice is appropriate.