Data Intensive Computing - Review Questions 6

Deadline: October 9, 2020

1. **1 point.** Assume you are designing a NoSQL database to store students profiles. Each student has a unique ID and some extra information, which are not fixed among students. What data model do you use in your database?

---

2. **1 point.** When do we need to use `cache` in Spark?

---

3. **1 point.** Explain how Kafka provides scalability and fault tolerance?

---

4. **2 points.** Assume we have two types of resources in the system, i.e., CPU and Memory. In total we have 28 CPU and 56GB RAM (e.g., 1 CPU = 2 GB). There are two users in the systems. User 1 needs $\langle 1CPU, 2GB \rangle$ per task, and user 2 needs $\langle 1CPU, 4GB \rangle$ per task. How do you share the resources fairly among these two users, considering (i) the asset fairness, and (ii) DRF.