

Data Intensive Computing - Review Questions 5

Deadline: October 7, 2023

1. 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 $\langle 2CPU, 2GB \rangle$ 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.
-

2. What are the similarities and differences among Mesos, YARN, and Borg?
-

3. What are the differences between Warehouse and Datalake? What is Lakehouse?
-

4. Briefly explain how Deltlake handle concurrent writing on the same file.