INFO8002 Large-scale data systems

Instructions

The exam is in English.

For the main question, be ready to:

- Describe the architecture of your solution and its operations.
- Define all its components and their interactions.
- Make diagrams whenever necessary.
- Motivate your design decisions.
- State clearly your choices and assumptions.

For all sub-questions:

- Give *short* but precise answers. Explain how they fit in your architecture.
- Be ready to give details if asked.

Question 1.

Define a sequence of MapReduce operations to compute an approximation of Pi (tip: geometric probability).

- Identify the map and reduce operations.
- Design a mechanism to prevent recomputing of the complete sequence of operations in the case a worker node fails.

Question 2.

Given a dataset containing the graph representation of a social network, define a sequence of MapReduce operations to compute the degree of every node (user). Define the data-format.

- Can you come up with a mechanism to mitigate the effect of *stragglers*? What negative effect do they have on the computation?