



Assignment-1 (10% weight)

In this assessment, you must not use generative artificial intelligence (AI) to generate any materials or content in relation to the assessment task.

This is an individual-based assignment.

Scope: Big Data Volume.

Weight: 10% of total unit marks.

Release: Week 3

Due: Tuesday, Week 7

This assignment consists of **four questions** that are, taken together, worth 10% of the final marks. **Please treat this Assignment as a take-home test. This is an individual assignment and you should complete the questions on your own.** You should not attempt to post questions on EdForum seeking solutions to the answers. If you require clarification on the Assignment questions, you can post a private post on EdForum or seek consultation from the tutors.

1. The first question is related to **Parallel Search Algorithms (2 Mark)**.
2. The second question is related to **Parallel Join Algorithms (4 Marks)**.
3. The third question is related to **Parallel Sort Algorithms (2 Marks)**.
4. The fourth question is related to **Parallel GroupBy Algorithms (2 Marks)**.

Instructions:

- You will be using Python 3. Answer all questions inside the Jupyter Notebook provided on this Assignment page. Please use the provided Docker to load the Jupyter Notebook.
- Read the instructions, skeleton code, and comments carefully.
- There are **code blocks that you need to complete** yourself as a part of the assignment.
- You are also required to **answer the questions below**.
- **Comment each line of code properly such that the tutor can easily understand what you are trying to do in the code. Marks may be deducted for insufficient or unclear comments.**
- Once completed, please rename the Jupyter Notebook to include your Student ID at the beginning of the filename (e.g., 12345678_FIT3182_Assignment_1.ipynb). Submit this Jupyter Notebook via the Assignment 1 submission link in Moodle.

Feedback

Grade	8.00 / 10.00
Graded on	Thursday, 2 May 2024, 8:48 PM

Question 1 - Parallel Search Algorithms	
Description for students	Completeness of the implementation of the algorithms
score	1.5 / 2
Remarks	-0.5 for incorrect explanation and code snippet

Question 2 - Parallel Join Algorithms	
Description for students	Completeness of the implementation of the algorithms
score	2.75 / 4
Remarks	<p>-0.25 for 2.3 for unwanted element in join output</p> <p>-0.5 for 2.4 as the outer join should not required check case to add 'nan'</p> <p>-0.5 for incorrect explanation, the answer is: We cannot use round-robin (random-equal) or random-unequal partitioning because we can miss joins due to matching records (on their join attribute) being on different partitions (processors). We can use hash partitioning because matching records will be in the same partition (processor).</p>

Question 3 - Parallel Sort Algorithms	
Description for students	Completeness of the implementation of the algorithms
score	1.75 / 2
Remarks	-0.25 for partially incomplete explanation in 3.4, the answer is: The sort phase uses all available processors because a local sort is applied to each partition. The merge phase uses half the available processors for the initial merge, and then halves it until a single processor merges the entire result.

Question 4 - Parallel GroupBy Algorithms	
Description for students	Completeness of the implementation of the algorithms
score	2 / 2
Remarks	