ST446 DISTRIBUTED COMPUTING FOR BIG DATA
COURSE PROJECT – MARKING CRITERIA (*r.* WT2024)

Rubric and Weight	Fail (0-49%)	Pass (50-59%)	Merit (60-69%)	Distinction (70% and over)
1. Introduction: problem formulation and summary of results (10%)	Poor or not well-defined problem formulation and poor summary of the results	Clear explanation of the problem formulation and clear summary of the results	Clear and well-motivated problem formulation based on literature that has some challenging elements; clear summary of the results	Clear explanation of the problem formulation; challenging and well-motivated problem formulation based on literature; clear summary of the results; evidence of critical thinking
2. Solution methodology (e.g. choice of data model; computation model; storage systems, machine learning algorithms) (20%)	Poor or inadequate application of distributed computing methods; poor explanation of underlying principles; application of basic methods	Satisfactory application of distributed computing methods; satisfactory explanation of principles; application of basic methods	Well identified and applied appropriate distributed computing methods; well explained methods; application of more advanced methods	Well identified and applied distributed computing methods; well explained underlying methods; show deep understanding of the solution concepts; application of advanced methods; evidence of critical thinking
3. Implementation (30%)	Poor implementation	Basic implementation using a distributed data processing system; performed simple data processing using an API	Good implementation using a system for distributed data processing; good exploration of the data processing API; well-structured and well commented code	Excellent implementation using a system for distributed data processing; show full understanding of the underlying distributed computing concepts; performed data processing using an API; developed own functions when appropriate; well-structured and commented code
4. Choice and description of data (10%)	Poor choice of data and poor description of data properties	Satisfactory choice of data for the given problem; dataset has some characteristics of big data (volume, variety, velocity, veracity) and is of moderate volume; good description of data properties	Good choice of data to address the given problem; dataset has some characteristics of big data and is of sizeable volume; clear description of data properties	Excellent choice of data to address the given problem; data has substantial complexity with respect to characteristics of big data and is clearly of volume that justifies the use of distributed computing; clear description of data properties

5. Numerical evaluation (20%)	Use of inappropriate metrics to evaluate the performance of the solution	Use of standard performance metrics	Good use of standard performance metrics; reported some performance gains obtained by distributed computing on multiple worker nodes	Use of standard performance metrics following best practices to evaluate and compare different solutions; clearly demonstrated and explained performance gains obtained by distributed computing on multiple worker nodes; obtained conclusive results; evidence of critical thinking
6. Conclusion (5%)	No conclusion or not well justified concluding remarks	Overall sound conclusion	Sound conclusion; summarised main results of the study based on obtained results; identified directions for future research	Excellent conclusion; well summarised main results of the study, strongly supported by obtained results; identified interesting directions for future research; evidence of critical thinking
7. Presentation quality (5%)	Poor structure with few explanations	Good structure of the report; cited references used in research	Clear presentation, well-structured report; cited references used in research	Clear presentation; well-structured report; effective communication of the results up to a professional standard; cited references used in research; cited references indicate a thorough research; identified and discussed key references