Department of Computing

Infrastructure for Big Data – Assessment 3

Assignment Brief - March 2020

Module Title:	Infrastructure for Big Data
Module Code:	DATAH5R04
Module Leader:	Michael Gleeson
Stage:	Award
Assessment Title:	Hadoop Cluster Technologies
Assessment Number:	3 of 3
Assessment Type:	Report
Restrictions on Length:	N/A
Individual/Group:	Individual
Assessment Weighting:	60%
Issue Date:	25th March 2019
Hand In Date:	30th April 2019
Planned Feedback Date:	Submission plus 4 weeks
Mode of Submission:	Video demo, document online via Blackboard

Assessment Details:

The purpose of this assessment is to allow the learner to independently research, implement and/or evaluate from a range of tools and techniques available in a Big Data Infrastructure environment and provide a summary report on your work. It is composed of a number of scenarios, all of which satisfy the narrative above.

You are required to utilise the IT Carlow Hadoop cluster to implement from an array of the technologies available. The assignment brief is intentionally left open to different approaches to allow learners to satisfy individual areas of interest or technologies, which ideally are appropriate for use in your dissertation while also delivering a body of work appropriate to a Level 9 MSc in Data Science. What this means is that you can choose your own path, suitable to your goals. Select **one** from the three scenarios on the following page.

What this essentially enables is that, as a Big Data Infrastructure specialist, you should have a broad knowledge of a range of tools that can be used. You are not expected to be an expert in everything but depending on particular situations, you must be able to critique available services, software and technology and determine which one (or combination) to use, backing up your decision with qualitative data (reason, inferred) or quantitative data (implementations, use cases or experiments).

Over the next few weeks we will use ~~lab time~~ online tools to work on this. There maybe technical issues that can arise, I would hope to have any technical issues resolved with a quick turnaround however it will be up to you to inform me of any technical issues, upon which can be investigated and determined if it is a cluster issue or an implementation issue.

This is an individual assignment and while we do work in a collaborative lab environment you must ensure that the work you submit is your own work and you own words. Any submissions deemed to be similar or lacking in appropriate references will be subject to IT Carlow's policy on plagiarism.

The Institutes policy is clearly laid out in Part 10 of the Academic Policies and Procedures available here: http://www.itcarlow.ie/public/userfiles/files/ITCarlow-Academic-Policies-Procedures.pdf

Scenario 1

You might choose to concentrate you efforts on one particular technology/service available on the cluster. This can be beneficial in the case where you are planning to use a specific Hadoop technology (for example, in your dissertation). In this scenario, there will be a significant depth of work required for your deliverable as you have decided to focus your work on one particular area.

Scenario 1 Deliverables:

1. A significant implementation and demo.
2. An accompanying technical 'How to' report.
3. Critical analysis of the specific technology and implementation (<500 words).

Scenario 2

This might involve you selecting a breath of technologies available across the cluster to explain, test and implement at a use case level and subsequently critiquing the service or use case based on the implementation. This can be beneficial in the instance where you are unsure of specific technologies and you wish to get hands on experience with a broader range of technologies available via Hadoop.

Scenario 2 Deliverables:

1. A series of use case implementations and demo.
2. An accompanying technical 'How to' report.
3. Critical analysis of each technology and implementation (<500 words).

Scenario 3

In this instance you may choose to identify two related or similar technologies in the Hadoop ecosphere and compare and contrast these. Here you will be required to deliver an implementation of each technology (to an intermediate level) and offer an evaluation of both. This can be beneficial in the case where you are aware to technologies with could match your dissertation requirements but need to evaluate further the suitability and pros and cons of each.

Scenario 3 Deliverables:

1. Two intermediate implementations and demo.
2. An accompanying technical 'How to' report.
3. Critical analysis of both technologies and implementations (<500 words).

Marks Breakdown

Implementation Dataset selected, appropriate degree of depth or breath Functionality, completeness and overall quality Video and presentation of the implementation with authority	50 Marks
How to Describe your implementation clearly and concisely Professionally presented documentation	30 Marks
Critical Analysis Critically analyse strengths and weaknesses of technology and usage. Discuss any contribution, innovation or optimization you added.	20 Marks

Total: 100 Marks

Adjectives to Classify Grades

%	**Quality**	**Description**
90-100	Supreme	Optimal critical capacity, highly original thought and expression, sophisticated mastery of subject matter. Adds significantly to understanding of the subject.
80-89	Exceptional	Outstanding critical capacity, significant originality of thought and expression, comprehensive and sophisticated command of subject matter. Provides new perspective of subject matter.
70-79	Excellent	Highly developed critical capacity, originality of thought and expression, comprehensive command of subject matter. Provides original perspective on subject matter.
60-69	Very Good	Developed critical capacity, some originality of thought and expression, competent command of subject matter. Displays capability of extended learning at higher levels.
50-59	Good	Some critical awareness evident, logical thought and expression, capable grasp of subject matter. Effective submission without expanding boundaries.
40-49	Acceptable	Little critical awareness, hesitant and uncertain thought and expression, basic grasp of subject matter, compromised by limited and unclear focus.
35-39	Insufficient	Hesitant and uncertain though, poorly organised expression, insufficient grasp of question or activity. Some knowledge of subject matter but compromise by inaccuracies, omissions and errors.
25-34	Poor	Unclear thought, inappropriate and disorderly expression. Little grasp of question or task, misunderstanding of subject matter or activity.
15-24	Unacceptable	Unstructured arguments with no support or premise, with evidence of cursory knowledge of subject matter and obvious irrelevancies and inaccuracies.
10-14	Irrelevant	Use or repetition and/or material not relevant to marking scheme. Incorrect and inaccuracies in subject matter evident.
0-9	Minimal	Rewriting of question, inappropriate commentary, not addressing the question in any meaningful way.