DEGREE: MSc Data Analytics

Module: Big Data Analytics

Assignment Title: Scalable Big Data Solutions with Hadoop Architecture and Map Reduce Abstraction.

Assignment Type: Report
Word Limit: 3000 words (+/- 300)

Weighting: 100%

Issue Date: 03-12-2025

Submission Date: 07-02-2026

Feedback Date: 10-03-2026

Issued by: Dr Balaganesh

Plagiarism:

When submitting work for assessment, students should be aware of the InterActive/Canvas guidance and regulations in concerning plagiarism. All submissions should be your own, original work.

You must submit an electronic copy of your work. Your submission will be electronically checked.

Harvard Referencing:

The Harvard Referencing System must be used. The Wikipedia, UKEssays.com or similar websites must **not** be used or referenced in your work.

Learning Outcomes:

LO1. Demonstrate the understanding of basic concepts of Big Data, its importance and need in business context.

LO2. Explain the various components of Hadoop and HFDC along with their role in the Big Data ecosystem.

LO3. Summarize the learning on Big Data analytics using Yarn, HDFC and MapReduce

Assessment Criteria: Weighting 100% 3000 words

Overview:

This project-based assignment is designed for master's students to demonstrate their understanding of Big Data concepts and practical skills in the Hadoop ecosystem. Students are required to implement Hadoop-based distributed processing using Map Reduce abstraction components (Hive / Hue / etc.).

Tasks (All tasks are equally weighted):

Task 1: Problem Definition and Business Context
- Identify a real-world business problem.
- Explain business context and need for Big Data.
- Choose dataset generating at least 3–4 HDFS blocks (128MB each).

Task 2: Environment Setup & Data Ingestion
- Configure Name Node, Data Node, Secondary Name Node.
- Explain FSImage, EditLog, check pointing.
- Ingest data using Hue or CLI and show block distribution.

Task 3: Data Processing using MapReduce Abstraction, (Hive / Hue..)
Perform **one distributed processing task** :
- Implement one task using MapReduce abstraction or Hive.
- Compare MapReduce abstraction with any other component.

Task 4: Machine Learning & Visualisation
- Perform ML using Mahout, Python, Hive, or MapReduce logic.
- Include pre-processing, training, evaluation, and visualisation.
- Provide business insights.

Data Source:

- Dataset must generate minimum 3–4 HDFS blocks (≥128MB each). Recommended options:
    1. Retail Store Transaction Dataset
        - Suitable for sales trend, forecasting, segmentation.
    2. Amazon Customer Reviews Dataset (subset)
        - Download enough records to exceed 3 HDFS blocks.
    3. FuturexSkill Big Data Retail Dataset

Submission Instructions:

- Compile a comprehensive project report or presentation that addresses each task outlined in the preceding section. This report includes: Steps to setup the Hadoop cluster
- Steps to ingest data into HDFS
- Source code of the MapReduce job (Mapper and Reducer)
- Instructions on how to submit the job to YARN
- Analysis of the results obtained from the MapReduce job

- Ensure that your report is clear, well-organized, and visually appealing
- Prepare a document using the BSBI assignment template available on Canvas.
- User Harvard referencing style for your bibliography.
- Refer to the Essay-Guide available on Canvas for further instructions.
- Submit your assignment electronically by the specified deadline.

Note:

Students are encouraged to explore additional reputable sources in the digital library beyond the list provided to strengthen their individual research and critical analysis skills. The unit lecturer can help with recommending additional literature.

The criteria for successfully passing the units with two assessment components require passing both components successfully. Passing marks for BSc and MA/MSc levels are different and are stated at the end of the assignment brief. Please find the examples below:

· **Example 1** (e.g., pass mark: 50)
· Component 1 (50% weighting) - 55
· Component 2 (50% weighting) - 55
· Final average: 52.5
· Result: **Pass (both components passed and therefore the unit is passed)**

· **Example 2** (e.g., pass mark: 50)
· Component 1 (50% weighting) – 80
· Component 2 (50% weighting) – 30
· Final average: 55
· Result: **Fail (because although the average is above 50, one of the components is a failure**)

GRADING DESCRIPTORS: LEVEL 7

EXPERIMENTATION & INNOVATION

Threshold Criteria	FAIL			PASS				
	0-29%	30-39%	40-49%	50-59%	60-69%	70-79%	80-89%	90-100%
Deals with complex issues both systematically and creatively demonstrating self- direction and originality in tackling and solving problems	Little to no ability to use techniques to deal with complex issues systematically (including those of ethics and sustainability) and creatively to solve problems and/or make decisions.	Low utilisation of established techniques to deal with complex issues systematically (including those of ethics and sustainability) and creatively to solve problems and/or make decisions, but with limitations in techniques or approach.	Limited research or advanced scholarship to their area of study by using a range of information and established and advanced techniques	Competent understanding of solving problems, through own research or advanced scholarship displaying a comprehensive understanding of established and advanced techniques	Good understanding of solving problems through own research and advanced scholarship critically selecting and displaying a comprehensive understanding of established and advanced techniques.	Very Good problem-solving skills displaying a comprehensive understanding of techniques applicable to their own research or advanced scholarship	Excellent range of extremely well- developed problem- solving displaying an understanding of techniques applicable to their own research or advanced scholarship beyond which is taught.	Exceptional problem-solving skills with sophisticated evaluation and application of a wide range of advanced information and techniques to undertake projects.
Comprehensive understanding of techniques applicable to their own research or	Little to no understanding of techniques applicable to their own research or advanced	Low understanding of techniques applicable to their own research or advanced	Limited understanding of key techniques applicable to their own research or	Competent understanding of techniques applicable to their own research or advanced	Good understanding of techniques applicable to their own research or advanced scholarship	Very good understanding of techniques applicable to their own research or advanced	Excellent understanding of techniques applicable to their own research or advanced	Exceptional understanding of techniques applicable to their own research or

advanced scholarship	scholarship or their limitations and ambiguities.	scholarship including their limitations and ambiguities.	advanced scholarship including their limitations and ambiguities.	scholarship including their limitations and ambiguities	and a some understanding of more specialised techniques.	scholarship and a some understanding of more specialised techniques.	scholarship and mastery of some more specialised areas.	advanced scholarship and mastery of some more specialised areas.

GRADING DESCRIPTORS: LEVEL 7

RESEARCH & ANALYSIS								
	FAIL			**PASS**				
Threshold Criteria	**0-29%**	**30-39%**	**40-49%**	**50-59%**	**60-69%**	**70-79%**	**80-89%**	**90-100%**
Systematic understanding of knowledge, and a critical awareness of current problems and/or new insights, much of which is at, or informed by, the forefront of their academic discipline, field of study or area of professional practice	Little to no knowledge of the subject with limited breadth or depth or deficiencies in major areas or currency.	Low knowledge of the subject lacking coherence, breadth, or detail with only some reference to ideas or arguments at the forefront of any part of the subject.	Limited knowledge to deal with terminology, facts and concepts some of which is informed by the forefront of defined areas of the subject.	Competent knowledge of ideas or arguments at the forefront of any part of the subject sufficient to deal with current issues in the discipline, generally more descriptive than critical or analytical.	Good knowledge of ideas or arguments at the forefront of any part of the subject showing a clear, critical insight into the discipline as whole and current issues/problems.	Very good knowledge of ideas or arguments at the forefront of the subject some of which are significantly beyond what has been taught and show a critical insight into the discipline and current issues/problems.	Excellent knowledge of ideas or arguments at the forefront of the subject many of which are significantly beyond what has been taught and show a critical insight into the discipline and current issues/problems.	Exceptional knowledge of ideas or arguments at the forefront of the subject most of which are significantly beyond what has been taught and show a critical insight into the discipline and current issues/problems.
Conceptual understanding that enables the student to display originality in the application of knowledge	Little to no conceptual understanding or argument and a focus on descriptive explanations which do not comment on arguments of others or alternative views.	Low conceptual understanding and arguments are weak or poorly constructed, and the work does not critically evaluate the arguments of others or consider alternative views.	Limited conceptual understanding and argument construction with critical evaluation of alternative views or comment on advanced scholarship.	Competent conceptual understanding and argument construction with critical evaluation of a range of views and consistent engagement with advanced scholarship.	Good conceptual understanding which critically evaluate and synthesise other views and information with a thoughtful interpretation of advanced scholarship.	Very good conceptual understanding which systematically synthesises a wide range of views with a critical insight into advanced scholarship.	Excellent conceptual understanding which critically apply a wide range of views through a perceptive use of advanced scholarship.	Exceptional conceptual understanding of publishable quality with systematic engagement and usage of advanced scholarship.

GRADING DESCRIPTORS: LEVEL 7

ENGAGING WITH PRACTICE								
	FAIL			**PASS**				
Threshold Criteria	0-29%	30-39%	40-49%	50-59%	60-69%	70-79%	80-89%	90-100%
Practical understanding of how established techniques of research and enquiry are used to create and interpret knowledge in the discipline	Little to no evidence of background investigation, analysis, research, enquiry, ethical awareness, and/or study.	Low evidence of background investigation, analysis, research, enquiry, ethical awareness, and/or study.	Limited background investigation, analysis, research, enquiry , ethical awareness, and/or study using established techniques, with the ability to extract relevant points.	Competent investigation, analysis, research, enquiry, ethical awareness, and/or study using established techniques accurately, and can critically appraise and use academic sources.	Good background investigation, analysis, research, enquiry, ethical awareness, and/or study using established techniques accurately, and possesses a well-developed ability to critically appraise a wide range of sources.	Very good, independent, extensive and appropriate investigation, analysis, research, enquiry, ethical awareness, and/or study beyond the usual range, and critically evaluates this to advance the work and/or direct arguments.	Excellent independent, extensive and appropriate investigation, analysis, research, enquiry, ethical awareness, and/or study well beyond the usual range, and critically evaluates this to advance the work and/or direct arguments.	Exceptional investigation, analysis, research, enquiry, ethical awareness, and/or study which demonstrates carefully considered depth and breadth and critically synthesises this to advance the work and/or direct arguments.
Originality in the application of knowledge	Little to no technical, creative or artistic skills related to their area of study.	Low technical, creative or artistic skills related to their area of study.	Limited technical, creative or artistic skills required for area of study.	Competent technical, creative or artistic skills required for area of study.	Good technical, creative or artistic skills required for area of study.	Very good range of technical, creative or artistic skills.	Excellent range of technical, creative or artistic skills	Exceptional range of technical, creative or artistic skills
Independently advance your own knowledge and understanding, and to develop new skills to a high level.	Little to no contribution to group activity and/or undertaking further training at a high/advanced level.	Low contribution to group activity and/or undertaking further training at a high/advanced level.	Limited contribution to group activity and/or undertaking further training at a high/advanced level.	Competent contribution to group activity and/or independently undertakes further training at a high/advanced level.	Good contribution to group activity and/or independently undertakes further training at a high/advanced level with an understanding of team roles	Very good contribution to group activity and/or independently undertakes further training at a high/advanced level with an understanding of team roles	Excellent contribution to group activity and/or independently undertakes further training at a high/advanced level with teamwork and leadership	Exceptional contribution to group activity and/or independently undertakes further training at a high/advanced level with teamwork and strong leadership.

REALISATION & COMMUNICATION

Threshold Criteria	FAIL			PASS				
	0-29%	30-39%	40-49%	50-59%	60-69%	70-79%	80-89%	90-100%
Communicate information, ideas, problems and solutions to both specialist and non-specialist audiences.	Little to no clarity in the communication of ideas, problems and solutions to audiences.	Low clarity in the communication of ideas, problems and solutions to audiences.	Limited clarity in the communication of ideas, problems and solutions to audiences.	Competent communication of ideas, problems and solutions to audiences.	Good, confident and clear communication of ideas, problems and solutions to audiences in a range of means / media.	Very good, confident and clear communication of ideas, problems and solutions to audiences in a range of means / media.	Excellent communication of ideas, problems and solutions to audiences in a range of means / media.	Exceptional communication of ideas, problems and solutions to audiences in a range of means / media.