LEVEL 4 & 5 DIPLOMA IN DATA SCIENCE

CURRICULUM FOR DATA SCIENCE BASED ON CREDIT SYSTEM

PROGRAMME LEARNING OUTCOMES (PLO):

I. Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

II. Problem analysis Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

III. Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

IV. Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions Manage Construction Projects for Planning, Analyzing, Costing, Scheduling, Predicting and complete within the stipulated period and fund.

V. Modern tool usage Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations

VI. Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

VII. Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development, Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

VIII. Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design

documentation, make effective presentations, and give and receive clear instructions.

IX. Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments

X. Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change

PROGRAMME GUIDELINES	
PROGRAMME TITLE	**Level 4 & 5 Data Science**
QUALIFICATION CODE	**701/1923/4**
LEVEL	**LEVEL – 4 & 5**
TOTAL CREDITS	**240**
TOTAL LEARNING HOURS	**2400 HOURS**
GUIDED LEARNING HOURS	**960 HOURS**

Total learning hour 2400 Hours

1 Credit = 10 hours of effort (10 hours of learning time which includes everything a learner has to do to achieve the outcomes in a qualification including the assessment procedures and practical's).

Guided Learning Hour for first year is 480 hours and second year is 480 hours.
Total Guided Learning Hours for Diploma in Data Science is 960 hours.

DIPLOMA IN DATA SCIENCE

COURSE STRUCTURE

YEAR	SEMESTER	UNIT SPECIFICATION	NO. OF. UNITS	UNIT CREDIT	CREDIT/YEAR
I	SEMESTER 1	Essential unit	1	20	120
		Essential unit	1	20	
		Essential unit	1	20	
	SEMESTER 2	Essential unit	1	20	
		Essential unit	1	20	
		Essential unit	1	20	
II	SEMESTER 3	Essential unit	1	20	120
		Essential unit	1	20	
		Essential unit	1	20	
	SEMESTER 4	Essential unit	1	20	
		Essential unit	1	20	
		Essential unit	1	20	
				TOTAL	**240**

FIRST YEAR	Essential unit carries	20 credit
	Essential unit carries	20 credit
	Essential unit carries	20 credit
SECOND YEAR	Essential unit carries	20 credit
	Essential unit carries	20 credit
	Essential unit carries	30 credit

LIST OF UNITS

S. No.	Subject Code	UNIT	UNIT SPECIFICATION	CREDIT
1	I/719/2021	Foundations of Data Management	Essential unit	20
2	I/719/2022	Data Management in Machine Learning Workflow	Essential unit	20
3	I/719/2023	Distributed Data Processing and Machine Learning Experimentation	Essential unit	20
4	I/719/2024	Data Privacy and anonymity	Essential unit	20
5	I/719/2025	Streaming Data Systems Architecture	Essential unit	20
6	I/719/2026	Streaming Data Frameworks	Essential unit	20
7	I/719/2027	Data Pipelines and Data Models	Essential unit	20
8	I/719/2028	Kafka Fundamentals and Programming	Essential unit	20
9	I/719/2029	Advanced Streaming Applications	Essential unit	20
10	I/719/2030	Streaming Analytics with Cloud	Essential unit	20
11	I/719/2031	Systems for Data Analytics	Essential unit	20
12	I/719/2032	Storytelling with Data and Ethics for Data Science	Essential unit	20

Semester : I
Year : 1
Credit : 60

UNIT CODE	UNIT	UNIT SPECIFICATION	CREDIT
I/719/2021	Foundations of Data Management	Essential unit	20
I/719/2022	Data Management in Machine Learning Workflow	Essential unit	20
I/719/2023	Distributed Data Processing and Machine Learning Experimentation	Essential unit	20

Semester : II
Year : 1
Credit : 60

UNIT CODE	UNIT	UNIT SPECIFICATION	CREDIT
I/719/2024	Data Privacy and anonymity	Essential unit	20
I/719/2025	Streaming Data Systems Architecture	Essential unit	20
I/719/2026	Streaming Data Frameworks	Essential unit	20

Semester : III
Year : 2
Credit : 60

UNIT CODE	UNIT	UNIT SPECIFICATION	CREDIT
I/719/2027	Data Pipelines and Data Models	Essential unit	20
I/719/2028	Kafka Fundamentals and Programming	Essential unit	20
I/719/2029	Advanced Streaming Applications	Elective Unit	20

Semester : IV
Year : 2
Credit : 60

UNIT CODE	UNIT	UNIT SPECIFICATION	CREDIT
I/719/2030	Streaming Analytics with Cloud	Essential unit	20
I/719/2031	Systems for Data Analytics	Essential unit	20
I/719/2032	Storytelling with Data and Ethics for Data Science	Essential unit	20

UNIT CODE : I/719/2021

UNIT TITLE : Foundations of Data Management

CREDIT : 20

SPECIFICATION : Essential Unit

UNIT DESCRIPTION

This unit aims developer to walk through the data engineering lifecycle and show you how to stitch together a variety of cloud technologies to serve the needs of downstream data consumers. You'll understand how to apply the concepts of data generation, ingestion, orchestration, transformation, storage, and governance that are critical in any data environment regardless of the underlying technology.It get a concise overview of the entire data engineering landscape and assess data engineering problems using an end-to-end framework of best practice. This unit helps in understand the marketing hype when choosing data technologies, architecture, and processes. It aims the data engineering lifecycle to design and build a robust architecture and incorporate data governance and security across the data engineering lifecycle.

UNIT LEARNING OUTCOMES

ULO1 - Understand the necessity of data management in data science.

ULO2 - Understand the concept of data ingestion, flow, storage and processing on data platforms

MAPPING

	PLO1	PLO2	PLO3	PLO4	PLO5	PLO6	PLO7	PLO8	PLO9	PLO10
ULO1	M	M		M	M		M	M		M
ULO2	M			M		M		M		

UNIT CODE : I/719/2022

UNIT TITLE : Data Management in Machine Learning Workflow

CREDIT : 20

SPECIFICATION : Essential Unit

UNIT DESCRIPTION

This unit aims to develop learners' knowledge in Data management for machine learning. Learners understand the processes and practices involved in organizing, storing, preprocessing, and maintaining data to support machine learning tasks. It encompasses various activities to ensure that data is high quality, accessible, and suitable for training and validating machine learning models. Effective data management is crucial for successful machine learning projects as the performance and accuracy of models heavily rely on the quality and availability of data. By effectively managing data for machine learning, organizations can enhance the accuracy and reliability of their models, improve decision-making processes, and derive valuable insights from data-driven solutions.

UNIT LEARNING OUTCOME

ULO1 – Gain experience in designing and handling the dataflow.
ULO2 – Apply the acquired conceptual data management knowledge and practices over a real-world machine learning workflow addressing the model metadata, privacy and monitoring aspects

MAPPING

	PLO1	PLO2	PLO3	PLO4	PLO5	PLO6	PLO7	PLO8	PLO9	PLO10
ULO1	M	M			M	M		M	M	M
ULO2			M	M			M	M		

UNIT CODE : I/719/2023

UNIT TITLE : Distributed Data Processing and ML Experimentation

CREDIT : 20

SPECIFICATION : Essential Unit

UNIT DESCRIPTION

This unit aims to build big data systems using an architecture that takes advantage of clustered hardware along with new tools designed specifically to capture and analyse web-scale data. It describes a scalable, easy-to-understand approach to big data systems that can be built and run by a small team. This unit guides learner through the theory of big data systems, how to implement them in practice, and how to deploy and operate them once they're built. In addition to discovering a general framework for processing big data, you'll learn specific technologies like Hadoop, Storm, and NoSQL databases.

UNIT LEARNING OUTCOME

ULO1 - Understand the Real-time processing of web-scale data using tools like Hadoop, Cassandra, and Storm

ULO2 - Deploy Machine Learning models in production

MAPPING

	PLO1	PLO2	PLO3	PLO4	PLO5	PLO6	PLO7	PLO8	PLO9	PLO10
ULO1	M	M		M	M		M	M	M	M
ULO2	M						M		M	
ULO3	M	M	M	M	M	M			M	

UNIT CODE : I/719/2024

UNIT TITLE : Data Privacy and Anonymity

CREDIT : 20

SPECIFICATION : Essential Unit

UNIT DESCRIPTION

This unit aims to understand the data privacy and anonymity on top search engines, social media and other powerful Internet players from tracking and profiling your online activities, gain unrestricted access to all the content and downloads the Internet has to offer use social media to stay connected with friends in ways that don't compromise your privacy or safety.This units helps to use the best privacy, anonymity and security apps that really work mask your IP address with a proxy, The Onion Router (Tor) or a virtual private network (VPN). Use encryption to keep your digital items, downloads and personal information completely hidden and safe. Prevent surveillance and the monitoring of your activities by Internet service providers (ISP), governments, adversaries and other unwelcome snoops

UNIT LEARNING OUTCOMES

ULO1 - Understand the data privacy and anonymity on top search engines

ULO2 - Understand the Causes of ML System failure and Problems with ML Production Monitoring

MAPPING

	PLO1	PLO2	PLO3	PLO4	PLO5	PLO6	PLO7	PLO8	PLO9	PLO10
ULO1	M	M		M	M		M	M		M
ULO2	M			M		M		M		

UNIT CODE	:	I/719/2025
UNIT TITLE	:	Streaming Data Systems Architecture
CREDIT	:	20
SPECIFICATION	:	Essential Unit

UNIT DESCRIPTION

This unit aims understand the components of streaming data systems with their capabilities and characteristics. Learn the relevant architecture and best practices for processing and analysis of streaming data. Gain knowledge about the development of system for data aggregation, delivery and storage using Open-source tools. Get familiarity with the advance streaming applications like Streaming SQL, Streaming machine learning.Learner can understand streaming of data systems with the relevant examples and illustrated use cases, you'll explore designs for applications that read, analyze, share, and store streaming data. Along the way, you'll discover the roles of key technologies like Spark, Storm, Kafka, and more.

UNIT LEARNING OUTCOMES

ULO1 - Understand the streaming of data systems and various data platforms
ULO2 - Understand the algorithms for streaming data application for data aggregation, delivery and storage

MAPPING

	PLO1	PLO2	PLO3	PLO4	PLO5	PLO6	PLO7	PLO8	PLO9	PLO10
ULO1	M	M		M	M		M	M		M
ULO2	M			M		M		M		

UNIT CODE : I/719/2026

UNIT TITLE : Streaming Data Frameworks

CREDIT : 20

SPECIFICATION : Essential Unit

UNIT DESCRIPTION

The unit aims to understand stream processing framework that can process data in real-time from multiple sources, including Apache Kafka, which Samza was developed in conjunction with. It is written in Java and Scala, uses Apache YARN for resource management, and provides exactly-once processing semantics. This unit provides core principles and concepts behind robust out-of-order data processing. It provides watermarks track progress and completeness in infinite datasets and data processing techniques ensure correctness. The concepts of streams and tables form the foundations of both batch and streaming data processing.

UNIT LEARNING OUTCOMES

ULO1- Understand streaming systems and Streaming data framework.

ULO2- Understand the Compositional engines such as Apache Storm, Samza, Apex the coding

MAPPING

	PLO1	PLO2	PLO3	PLO4	PLO5	PLO6	PLO7	PLO8	PLO9	PLO10
ULO1	M	M		M	M		M	M		M
ULO2	M			M		M		M		

UNIT CODE : I/719/2027

UNIT TITLE : Data Pipelines and Data Models

CREDIT : 20

SPECIFICATION : Essential Unit

UNIT DESCRIPTION

This unit aims the learner to understand about Data pipelines are the foundation for success in data analytics. Moving data from numerous diverse sources and transforming it to provide context is the difference between having data and actually gaining value from it. This pocket reference defines data pipelines and explains how they work in today's modern data stack.Learner will learn common considerations and key decision points when implementing pipelines, such as batch versus streaming data ingestion and build versus buy. This book addresses the most common decisions made by data professionals and discusses foundational concepts that apply to open-source frameworks, commercial products, and homegrown solutions.

UNIT LEARNING OUTCOMES

ULO1- Understand the products and design for building data pipelines

ULO2- Understand the Data Ingestion from Extracted Data

MAPPING

	PLO1	PLO2	PLO3	PLO4	PLO5	PLO6	PLO7	PLO8	PLO9	PLO10
ULO1	M	M		M	M		M	M		M
ULO2	M			M		M		M		

UNIT CODE : I/719/2028

UNIT TITLE : Kafka Fundamentals and Programming

CREDIT : 20

SPECIFICATION : Essential Unit

UNIT DESCRIPTION

This units understand the learners to create enterprise application data, whether it consists of log messages, metrics, user activity, or outgoing messages. Moving all this data is just as important as the data itself. With this updated edition, application architects, developers, and production engineers new to the Kafka streaming platform will learn how to handle data in motion. Additional chapters cover Kafka's Admin Client API, transactions, new security features, and tooling changes. Learner learn to working with Kafka and how to deploy production Kafka clusters, write reliable event-driven microservices, and build scalable stream processing applications with this platform. Through detailed examples, you'll learn Kafka's design principles, reliability guarantees, key APIs, and architecture details, including the replication protocol, the controller, and the storage layer.

UNIT LEARNING OUTCOMES

ULO1- Understand the Kafka deploying and configuring messages

ULO2- Understand the Kafka Kafka's delivery capabilities for stream processing systems

MAPPING

	PLO1	PLO2	PLO3	PLO4	PLO5	PLO6	PLO7	PLO8	PLO9	PLO10
ULO1	M	M		M	M		M	M		M
ULO2	M			M		M		M		

UNIT CODE : I/719/2029

UNIT TITLE : Advanced Streaming Applications

CREDIT : 20

SPECIFICATION : Essential Unit

UNIT DESCRIPTION

This unit aims the learner to understand the unbounded and fast-moving data streams has historically been difficult. But with Kafka Streams and ksq lDB, building stream processing applications is easy and fun. This practical guide shows data engineers how to use these tools to build highly scalable stream processing applications for moving, enriching, and transforming large amounts of data in real time. This unit provide learners to understand important stream processing concepts against a backdrop of several interesting business problems. You'll learn the strengths of both Kafka Streams and ksq lDB to help you choose the best tool for each unique stream processing project. Non-Java developers will find the ksq lDB path to be an especially gentle introduction to stream processing.

UNIT LEARNING OUTCOMES

ULO1- Understand Kafka communication pattern and build stateless and stateful stream processing applications.

ULO2- Understand Kafka Streams and ksqlDB applications to production environment.

MAPPING

	PLO1	PLO2	PLO3	PLO4	PLO5	PLO6	PLO7	PLO8	PLO9	PLO10
ULO1	M	M		M	M		M	M		M
ULO2	M			M		M		M		

UNIT CODE : I/719/2030

UNIT TITLE : Streaming Analytics with Cloud

CREDIT : 20

SPECIFICATION : Essential Unit

UNIT DESCRIPTION

This unit introduced to Azure Stream Analytics, and see how you can use the tools and functions in Azure to develop your own Streaming Analytics. Over the course, you will be given comparative analytic guidance on using Azure Streaming with other Microsoft Data Platform resources such as Big Data Lambda Architecture integration for real time data analysis and differences of scenarios for architecture designing with Azure HDInsight Hadoop clusters with Storm or Stream Analytics. The unit also provides you how you can manage, monitor, and scale your solution for optimal performance. You will be well-versed in using Azure Stream Analytics to develop an efficient analytics solution that can work with any type of data. Style and approach A comprehensive guidance on developing real-time event processing with Azure Stream Analysis.

UNIT LEARNING OUTCOMES

ULO1- Understand the Azure Stream Analytics to develop an efficient analytics solution

ULO2- Understand the serverless streaming data service that makes it easy to capture, process, and store data streams at any scale.

MAPPING

	PLO1	PLO2	PLO3	PLO4	PLO5	PLO6	PLO7	PLO8	PLO9	PLO10
ULO1	M	M		M	M		M	M		M
ULO2	M			M		M		M		

UNIT CODE	:	I/719/2031
UNIT TITLE	:	Systems for Data Analytics
CREDIT	:	20
SPECIFICATION	:	Essential Unit

UNIT DESCRIPTION

The unit introduce the systems perspective of data analytics: to leverage systems effectively, understand, measure, and improve performance while performing data analytics tasks. Enable learner to develop a working knowledge of how to use parallel and distributed system for data analytics. Enable learner to apply best practices in storing and retrieving data for analytics. Enable learner to leverage commodity infrastructure (such as scale-out clusters, distributed datastores, and the cloud) for data analytics. The units extend the basis for effective decisions. Whoever has the data has the ability to extract information promptly and effectively to make pertinent decisions. The premise of this unit is to empower users and tool developers with the appropriate collection of formulas and techniques for data analytics and to serve as a quick reference to keep pertinent formulas within fingertip reach of learner.

UNIT LEARNING OUTCOME

ULO1 - Understand the Systems Attributes and Data Storage for Data Analytics

ULO2 - Understand the Strategies for data access: Partition, Replication, and Messaging

MAPPING

	PLO1	PLO2	PLO3	PLO4	PLO5	PLO6	PLO7	PLO8	PLO9	PLO10
ULO1	M	M	M	M			M	M	M	M
ULO2		M				M			M	

UNIT CODE : I/719/2032

UNIT TITLE : Storytelling with Data and Ethics for Data Science

CREDIT : 20

SPECIFICATION : Essential Unit

UNIT DESCRIPTION

This unit Storytelling with Data teaches the learner about the fundamentals of data visualization and how to communicate effectively with data. Learner will discover the power of storytelling and the way to make data a pivotal point in your story. The content in this illuminative text are grounded in theory, but made accessible through numerous real-world examples—ready for immediate application to your next graph or presentation. Storytelling is not an inherent skill, especially when it comes to data visualization, and the tools at our disposal don't make it any easier. It provides impact of data science continues to grow on society there is an increased need to discuss how data is appropriately used and how to address misuse. Yet, ethical principles for working with data have been available for decades. The real issue today is how to put those principles into action.

UNIT LEARNING OUTCOMES

ULO1- Understand the importance of context and audience.

ULO2- Understand the deliberate practice of data ethics for better products, better teams, and better outcomes

MAPPING

	PLO1	PLO2	PLO3	PLO4	PLO5	PLO6	PLO7	PLO8	PLO9	PLO10
ULO1	M	M		M	M		M	M		M
ULO2	M			M		M		M		

ASSESSMENT METHODS AND TECHNIQUES FOR DIPLOMA IN DATA SCIENCE

Assessment technique	Type of Assessment	Description	Formative or Summative
Case studies	Oral/ Problem based/ Practical	Students are required to work through a case study to identify the problem(s) and to offer potential solutions; useful for assessing students' understanding and for encouraging students to see links between theory and practice. Case studies could be provided in advance of a time-constrained assessment.	**Formative**
Concept maps	Written/ Oral	Students map out their understanding of a particular concept. This is a useful (and potentially quick) exercise to provide feedback to staff on students' understanding.	**Formative**
'Doing it' exam	Written	An exam which requires students to do something, like read an article, analyze and interpret data etc.	**Formative / Summative**
Field report	Written/ Oral	Students are required to produce a written/ oral report relating to a field/ site visit.	**Formative**
Laboratory books / Reports	Practical/ Written	Students are required to write a report for all (or a designated sample) of practical's in a single lab book. A sample of lab books will be collected each week to mark any reports of labs done in previous weeks; this encourages students to keep their lab books up to date. Each student should be sampled the same number of times throughout the module with a designated number contributing to the assessment mark.	**Summative**
Multiple choice questions (MCQs)	Written	Can be useful for diagnostic, formative assessment, in addition to summative assessment. Well-designed questions can assess more than factual recall of information, but do take time to design.	**Formative / Summative**
Online discussion boards	Written	Students are assessed on the basis of their contributions to an online discussion for example, with their peers; this could be hosted on a virtual learning environment (VLE).	**Formative**
Open book exams	Written	Students have the opportunity to use any or specified resources to help them answer set questions under time constraints. This method removes the over-reliance on memory and recall and models the way that professionals manage information.	**Summative**
Oral presentations	Oral / Written	Students are asked to give an oral presentation on a particular topic for a specified length of time and could also be asked to prepare associated	**Summative**

		handout(s). Can usefully be combined with self- and peer-assessment.	
Problem sheets	Written	Students complete problem sheets, e.g. on a weekly basis. This can be a useful way of providing students with regular formative feedback on their work and/or involving elements of self- and peer assessment.	**Formative**
Research projects / Group projects	Written/ Practical/ Oral/ Performance/ Problem based/ Work placement	Potential for sampling wide range of practical, analytical and interpretative skills. Can assess wide application of knowledge, understanding and skills.	**Formative / Summative**
Short answer questions	Written	Useful to assess a wide range of knowledge/skills across a module.	**Summative**
Simulations	Practical/ Written/ Oral/ Problem-based	Text or virtual computer-based simulations are provided for students, who are then required to answer questions, resolve problems, perform tasks and take actions etc. according to changing circumstances within the simulation. Useful for assessing a wide range of skills, knowledge and competencies.	**Formative**
Viva voce	Oral	Often used for assessing 'borderline' degree classifications but also useful to explore students' understanding of a wide range of topics. Depending on class size however, they can be time consuming for staff.	**Summative**