Module Synopsis

- **Specialist Diploma in Data Science (Data Engineering) (NSDEG)**
- **Work-Study Post-Diploma (Specialist Diploma in Data Science (Data Engineering)) (WSP-NSDEG)**

PDC 1:	**Post-Diploma Certificate in Fundamentals of Data Science**	**(120 hours)**
1.	**MSA9001: Introduction to Statistics for Data Science**	**(60 hours)**
	This module provides students with an introduction to elementary probability theory and statistical concepts and principles that lay the foundation to understand and learn the statistical procedures and methods in the subsequent modules. The topics covered include descriptive statistics, rules of probability, probability distributions of discrete and continuous random variables, sampling distributions, statistical estimation, hypothesis testing and linear regression.	
2.	**IT8701: Introduction to Programming for Data Science**	**(60 hours)**
	This module equips students with the fundamental programming skills required to retrieve, manipulate, process, analyse, and visualise data using the Python programming language. Students learn key concepts of structured, and unstructured data, and develop applications to work with relational and NoSQL databases. The module also introduces basic text and document processing techniques to extract insights from diverse data sources. Throughout the module, students are exposed to Generative AI-assisted coding and debugging workflows to enhance productivity while developing a sound understanding of programming fundamentals and responsible AI usage.	

PDC 2:	**Post-Diploma Certificate in Data Engineering**	**FQ: (150 hours)/ WSP: (165 hours)**
1	**IT8704: Data Management Systems***	**FQ: (75 hours)**
1	**IT8706: Data Management Systems Practicum^**	**WSP: (90 hours)**
	This module provides students with knowledge of database concepts, database architecture, and data warehousing solutions used in modern data engineering. Topics include data modelling, relational databases, OLTP and OLAP concepts, SQL development and optimisation, data validation, data integration, and Extract, Transform, and Load (ETL) processes. Students will also gain exposure to modern analytical data processing concepts and contemporary data architectures used to support analytical workloads. Generative AI-assisted workflows are introduced to support database design, SQL development, and query optimisation while reinforcing students' ability to validate and evaluate AI-generated outputs.	
2.	**IT8705: Big Data Technologies**	**(75 hours)**
	This module provides students with an understanding of Big Data technologies and the design and implementation of scalable data pipelines for processing large volumes of structured and unstructured data. Topics include the Hadoop ecosystem, Apache Spark, distributed data processing, Spark SQL, Spark MLlib, and Spark Streaming. Students will develop practical skills in building modern data pipelines, analysing large-scale datasets,	

Module Synopsis

- **Specialist Diploma in Data Science (Data Engineering) (NSDEG)**
- **Work-Study Post-Diploma (Specialist Diploma in Data Science (Data Engineering)) (WSP-NSDEG)**

and implementing machine learning and streaming analytics applications. The module also introduces Generative AI-assisted and agentic workflows to enhance productivity in data engineering tasks while maintaining emphasis on core technical competencies, responsible AI usage, and current industry practices.

***For NSDEG:**

- Student must complete PDC 1: MS9001 and IT8701 + PDC2: **IT8704** and IT8705

^For WSP-NSDEG:

- Student must complete PDC 1: MS9001 and IT8701 + PDC2: **IT8706** and IT8705