CCEK – NSQF ALIGNED PROGRAM

COURSE SYLLABUS

FOR

Data Analyst

CCEK - NATIONAL SKILL DEVELOPMENT TRAINING PROGRAM

Data Analyst

CCEK – NSDC course package covers the following Qualification Packs and leads to the following NSDC certifications. The students who successfully completed the course programs are entitled to get NSDC certification after undergoing the assessment process of NSDC as per the rules and regulations stipulated by NSDC from time to time.

SL. NO.	QUALIFICATIONS PACK	QUALIFICATIONS PACK CODE	NSQF LEVEL
1	**AI - Data Engineer** **Brief Job Description:** Individuals at this job are responsible for working on different aspects of data architecture. They will be responsible for developing data integrations such as data warehouses or data lakes for storing data, or data pipelines for performing operations on data in real-time. They will need to have strong communication skills to work with stakeholders and decision makers across multiple teams to ensure timely release of applications.	**SSC/Q8106**	**6**

COURSE DETAILS

Data Analyst

EXAMINATION DETAILS

COURSE NAME	COURSE CODE	ELIGIBILITY	DURATION
Data Analyst	G10	Degree	370

SL. NO.	EXAM	EXAM CODE	MAXIMUM MARK	INTERNAL	TOTAL MARK
THEORY PAPERS					
1	Exploring the Role of Machine Learning in Data Analytics	T001	100	50	150
2	Artificial Intelligence in Cyber Security	T002	100	50	150
PRACTICAL PAPERS					
1	Exploratory Data Analysis (EDA) on a Public Dataset	L001	100	50	150
2	Predictive Modeling with Machine Learning	L002	100	50	150
TOTAL MARKS					
1	Total Examination Marks (THEORY Online + PRACTICAL Examination)				400
2	Total Internal Marks				200
3	**Total Marks (Total Internal Marks + Total Examination Marks)**				**600**

Data Analyst

INTERNAL MARK CRITERIA FOR EACH

SL NO.	MODULE	MODULE CODE	MAXIMUM MARK	INTERNAL MARK	TOTAL MARK
1	Exploring the Role of Machine Learning in Data Analytics	T001	100	50	150
2	Artificial Intelligence in Cyber Security	T002	100	50	150
3	Exploratory Data Analysis (EDA) on a Public Dataset	L001	100	50	150
4	Predictive Modeling with Machine Learning	L002	100	50	150
	TOTAL		400	200	600

ATTENDANCE	GENERAL PERFORMANCE	INTERNAL EXAMINATIONS/ PROJECTS/ ASSIGNMENTS	TOTAL MARKS
5	5	40	50

COURSE SYLLABUS

FOR

Data Analyst

COURSE	Data Analyst	
TOTAL MARKS	Mark: 600	Internal Mark: 200
TOTAL HOURS	370 Hrs	

DEFENITION OF CREDIT

1 Credit	15Hrs Theory/ 30Hrs Practical
Skill Components	60 – 70 % of Total Credit

MODULES INCLUDED IN THIS SUBJECT

SL NO	MODULE NAME	CREDIT BREAKUP
1	Module 1: Artificial Intelligence & Big Data Analytics – An Introduction	**2**
2	Module 2: Basics of Architecture	
3	Module 3: Global Data Regulations and Standards	
4	Module 4: Data Tools and Usage	**1.5**
5	Module 5: Data Ingestion and Synchronization	**5**
6	Module 6: Data Transformation and Processing	
7	Module 7: Inclusive and environmentally sustainable workplaces	**1**
8	Module 8 Introduction to Employability Skills	
9	Module 9: Constitutional values – Citizenship	

10	Module 10: Becoming a Professional in the 21st Century	
11	Module 11: Basic English Skills	1.5
12	Module 12: Career Development and Goal Setting	
13	Module 13: Communication skills	
14	Module 14: Diversity and Inclusion	
15	Module 15: Financial and Digital Literacy	
16	Module 16: Essential Digital Skills	
17	Module 17: Entrepreneurship	
18	Module 18: Customer Service	0.5
19	Module 19: Getting Ready for Apprenticeship and Jobs	1
	Total	12.5

Training Outcomes

- Describe the use cases of AI & Big Data Analytics in various industries and define the various roles under this occupation
- Apply principles and methods of enterprise architecture which conform to industry standards and frameworks
- Identify general principles and basic concepts of data management standards across the globe
- Evaluate the uses of different data management tools and platforms
- Develop data integrations for storage and access such as data warehouses and data lakes, or processing such as data pipelines
- Design and implement ETL (Extract-Transform-Load) processes
- Perform different integrations for batch and real-time data
- Develop, test and troubleshoot data pipelines
- Plan their schedules and timelines based on the nature of work
- Demonstrate ways to communicate and work effectively with colleagues
- Use different approaches to effectively manage and share data and information
- Develop strong relationships at the workplace through effective communication and conflict management
- Identify best practices to maintain an inclusive, environmentally sustainable workplace

MODULES

Module 1: Artificial Intelligence & Big Data Analytics – An Introduction

THEORY

- Explain the relevance of AI & Big Data Analytics for the society
- Explain the various use-cases of AI & Big Data in the industry
- Define "general" and "narrow" AI
- Describe the fields of AI such as image processing, computer vision, robotics, NLP, etc.
- Create a career map for roles in Information/Cyber Security
- Demonstrate the working mechanism of malicious codes such as virus, malware, logic bomb, ransomware, spyware, phishing, trojan, etc.

PRACTICAL

- Outline a career map for roles in AI & Big Data Analytics
- Analyse the differences between key terms such as Supervised Learning, Unsupervised Learning and Deep Learning

Module 2: Basics of Architecture

THEORY

- Explain enterprise architecture principles, components and their PRACTICAL application
- Identify critical success factors for common enterprise architect approaches Demonstrate the usage of basic methods/tools in preventing cyber- attacks.
- Demonstrate system architecture of sample information systems.
- Demonstrate the basic functionalities of the applications, their components and security features.
- Demonstrate the utility of application and database layer IPS/IDS.

Module 3: Global Data Regulations and Standards

THEORY

- Discuss the need for data regulations and standards
- Analyse commonly used global data regulation policies (such as GDPR)
- Discuss the roles and responsibilities of key actors involved in enforcing data regulations and standards

- Identify best practices used by various organizations in the enforcement of data regulations and standards

Module 4: Data Tools and Usage

THEORY

- Identify the various data integration and management tools and platforms
- Discuss the fundamentals of using the data integration and management tools and platforms
- Discuss fundamentals of various infrastructure components such as storage devices, networking hardware, server-storage connectivity
- Explain fundamental tools used for distributed computing services such as data storage, integration and processing
- Assess the role of scripts to extract and understand data

PRACTICAL

- Demonstrate the ways to use data integration and management tools
- Demonstrate the ways to use different infrastructure components such as storage devices, networking hardware, server-storage connectivity, virtualization technologies
- Configure various microservices, frameworks, libraries, packages
- Create scripts for data extraction

Module 5: Data Ingestion and Synchronization

THEORY

- Prepare a tracker in prescribed format to capture vulnerabilities and risk exposure data of sample applications.
- Demonstrate categorization of vulnerabilities based on level of weakness, sensitivity of information, relevance, root causes, risk criticality, and mitigation methods.
- Perform root cause analysis of identified vulnerabilities in sample applications.
- Prepare a report on vulnerability analysis including security requirements, vulnerabilities identified and recommended solutions.
- Demonstrate the procedure to securely store data collected during the assessment, vulnerabilities, analysis results, and mitigation recommendations.

PRACTICAL

- Demonstrate the process of performing batch or real-time integrations
- Estimate the effort required to integrate a given dataset

- Demonstrate the methods to create a data warehouse design
- Demonstrate the method to select the right data format (AVRO, Parquet) and compression for a given application
- Illustrate ways to retrieve, partition and ingest data into a data lake (such as Azure, AWS)
- Demonstrate ways to handle errors and faults in a Data Lake
- Identify ingestion, throughput and latency of a Data Lake and optimise it based on the results
- Demonstrate the implementation of an ingestion policy with different custom plugins or tools for the interfaces between source system and data lake (using batch processing and real-time tools)

Module 6: Data Transformation and Processing

THEORY

- Evaluate the different tools and frameworks to develop data pipelines
- Discuss the methods to develop, test and troubleshoot data pipelines
- Explain the methods to transform and load data to a data warehouse
- Explain tools and concepts for in- memory models and OLAP cubes

PRACTICAL

- Demonstrate the features of various tools (such as Airflow, Luigi) to develop data pipelines
- Demonstrate method(s) to transform a given data into another format using data processing systems such as Spark, Hadoop
- Build a data pipeline that can run in an experimental sandbox, containers or virtual machine
- Demonstrate a Proof of Concept (PoC) for the designed Data Lake
- Demonstrate ways to optimize the performance of an individual data pipeline and the overall system
- Demonstrate ways to implement query execution in a Data Lake
- Demonstrate the ways to program and execute OLAP queries using SQL or other query tools on a data warehouse
- Perform error and fault handling in data pipelines

Module 7: Inclusive and environmentally sustainable workplaces

THEORY

- Describe different approaches for efficient energy resource utilisation and waste management
- Describe the importance of following the diversity policies
- Identify stereotypes and prejudices associated with people with disabilities and the

negative consequences of prejudice and stereotypes
- Discuss the importance of promoting, sharing and implementing gender equality and PwD sensitivity guidelines at organization level

PRACTICAL

- Practice the segregation of recyclable, non-recyclable and hazardous waste generated
- Demonstrate different methods of energy resource use optimization and conservation
- Demonstrate essential communication methods in line with gender inclusiveness and PwD sensitivity

Module 8: Introduction to Employability Skills

THEORY

- Discuss the Employability Skills required for jobs in various industries
- List different learning and employability related GOI and private portals and their usage

Module 9: Constitutional values – Citizenship

THEORY

- Explain the constitutional values, including civic rights and duties, citizenship, responsibility towards society and personal values and ethics such as honesty, integrity, caring and respecting others that are required to become a responsible citizen
- Show how to practice different environmentally sustainable practices

Module 10: Becoming a Professional in the 21st Century

THEORY

- Discuss importance of relevant 21st century skills.
- Exhibit 21st century skills like Self-Awareness, Behaviour Skills, time management, critical and adaptive thinking, problem-solving, creative thinking, social and cultural awareness, emotional awareness, learning to learn etc. in personal or professional life.
- Describe the benefits of continuous learning

Module 11: Basic English Skills

THEORY

- Show how to use basic English sentences for everyday conversation in different contexts, in person and over the telephone
- Read and interpret text written in basic English

Module 12: Career Development and Goal Setting

THEORY

- Create a career development plan with well-defined short- and long-term goals

Module 13: Monitoring and Logging of Application Events and Alarms

THEORY

- Demonstrate how to communicate effectively using verbal and nonverbal communication etiquette.
- Explain the importance of active listening for effective communication
- Discuss the significance of working collaboratively with others in a team

Module 14: Diversity and Inclusion

THEORY

- Demonstrate how to behave, communicate, and conduct oneself appropriately with all genders and PwD
- Discuss the significance of escalating sexual harassment issues as per POSH

Module 15: Financial and Digital Literacy

THEORY

- Outline the importance of selecting the right financial institution, product, and service
- Demonstrate how to carry out offline and online financial transactions, safely and securely

Module 16: Essential Digital Skills

THEORY

- Describe the role of digital technology in today's life
- Demonstrate how to operate digital devices and use the associated

applications and features, safely and securely
- Discuss the significance of displaying responsible online behaviour while browsing, using various social media platforms, e-mails, etc., safely and securely
- Create sample word documents, excel sheets and presentations using basic features
- utilize virtual collaboration tools to work effectively

Module 17: Entrepreneurship

THEORY

- Explain the types of entrepreneurship and enterprises
- Discuss how to identify opportunities for potential business, sources of funding and associated financial and legal risks with its mitigation plan
- Describe the 4Ps of Marketing-Product, Price, Place and Promotion and apply them as per requirement
- Create a sample business plan, for the selected business opportunity

Module 18: Customer Service

THEORY

- Describe the significance of analysing different types and needs of customers
- Explain the significance of identifying customer needs and responding to them in a professional manner.
- Discuss the significance of maintaining hygiene and dressing appropriately

Module 19: Getting Ready for Apprenticeship and Jobs

THEORY

- Create a professional Curriculum Vitae (CV)
- Use various offline and online job search sources such as employment exchanges, recruitment agencies, and job portals respectively
- Discuss the significance of maintaining hygiene and confidence during an interview
- Perform a mock interview
- List the steps for searching and registering for apprenticeship opportunities