

NSQF Qualification File
Approved in 17th NSQC, dated- 31/03/22

Post Graduate Program in Data Engineering

QUALIFICATION FILE - CONTACT DETAILS OF THE SUBMITTING BODY

Name and address of submitting body:

NATIONAL INSTITUTE OF ELECTRONICS AND INFORMATION TECHNOLOGY
NIELIT Bhawan, Plot No. 3, PSP Pocket, Sector-8,
Dwarka, New Delhi-110077

Name and contact details of individual dealing with the submission

Name	:	Dr. Sanjeev Kumar Jha
Position in the organization	:	Joint Director
Address if different from above	:	NA
Tel number(s)	:	044-2442 1445 ext. 210, 7765803105
E-mail address	:	skjha@nielit.gov.in

List of documents submitted in support of the Qualifications File

1. Annexure I – Evidence of Job /Requirement in Industry
2. Annexure II – Detailed Model Curriculum/Syllabus
3. Annexure III –Candidates Trained – 250

Post Graduate Program in Data Engineering

SUMMARY

1	Qualification Title	Post Graduate Program in Data Engineering
2	Qualification Code, if any	--
3	NCO code and occupation	2512.0300, Programmer Analyst
4	Nature and purpose of the qualification (Please specify whether qualification is short term or long term)	<p>Nature:</p> <ul style="list-style-type: none"> ❖ The Course was started in 2019 and got NSQF aligned in 2020. The course is a unique 6-month (840 Hours) program which is an excellent blend of knowledge and practice in the field of Data Science and its industrial applications. The program is targeted for creating qualified Data Science professionals. The course progresses through the Operating System, concepts of Data and its storage, programming for data science, Big Data Technology and its implementation. Various advanced tools like Python and R, along with MySQL, Apache Cassandra and Hadoop Framework are used for achieving the goal of solving critical business and Data Analytic problems. The course has been designed after proper industry survey and consultation with multiple industry leaders to ensure that participants learn exactly what employers need. ❖ With a strong emphasis on 'learning by doing', our programs are developed with the goal of creating well-rounded, job-ready professionals that can add immediate value to any organization. ❖ On completion of the Course, the Participants will learn the concept of Data

Post Graduate Program in Data Engineering

		<p>Analytics using open source statistical tools like R, Python and some very good visualization tools and techniques. They will be able to implement industry-oriented Data Science Projects.</p> <ul style="list-style-type: none"> ❖ The Course is targeted for creating qualified Data Science professionals which will help in employment and entrepreneur development. ❖ We had already trained around 250 candidates and certified around 100 candidates in this course. ❖ Almost 90% of certified candidates are placed. <p>Purpose:</p> <ul style="list-style-type: none"> ❖ To make Data Analyst, Data Scientists and Data Engineers. ❖ This course is meant to sensitize students for computational statistics applications and usage as well as provide hands-on experience for solving real world data science issues. ❖ Qualification is a short-term course
5	Body/bodies which will award the qualification	<ul style="list-style-type: none"> ❖ National Institute of Electronics and Information Technology NIELIT Bhawan, Plot No. 3, PSP Pocket, Sector-8, Dwarka, New Delhi-110077
6	Body which will accredit providers to offer courses leading to the qualification	<ul style="list-style-type: none"> ❖ National Institute of Electronics and Information Technology NIELIT Bhawan, Plot No. 3, PSP Pocket, Sector-8, Dwarka, New Delhi-110077
7	Whether accreditation/affiliation norms are already in place or not , if applicable (if yes, attach a copy)	<ul style="list-style-type: none"> ❖ The Handbook for TP accreditation norm is available at: https://www.nielit.gov.in/content/nsqf

Post Graduate Program in Data Engineering

8	Occupation(s) to which the qualification gives access	❖ Statistical Analysts, Data Analysts, Data Scientists, Big Data Engineer, Hadoop Developer.
9	Job description of the occupation	<ul style="list-style-type: none"> ❖ Statistical Analysts: Statistical Analysis ❖ Data Analysts: Collecting and interpreting data, analysing results, Reporting the results back to the relevant members of the business, identifying patterns and trends in data sets, working alongside teams within the business or the management team to establish business needs, defining new data collection and analysis processes. ❖ Data Scientists: Applying data mining techniques, doing statistical analysis, and building high quality prediction systems. ❖ Big Data Engineer: Development of Hadoop Environment, loading data from raw data sets, pre-processing using various techniques, designing solutions independently based on high-level architecture, Manage the technical communication between the survey vendor and internal systems, Maintain the production systems, collaborate with other development and research teams, building a cloud-based platform that allows easy development of new applications. ❖ Hadoop Developer: Design and develop Hadoop system with strong documentation skills. The job of a Hadoop developer is almost similar to the software developer but in the Big Data domain.
10	Licensing requirements	❖ NA

Post Graduate Program in Data Engineering

11	Statutory and Regulatory requirement of the relevant sector (documentary evidence to be provided)	❖ NA
12	Level of the qualification in the NSQF	❖ Level 6
13	Anticipated volume of training/learning required to complete the qualification	❖ 840 Hours
14	Indicative list of training tools required to deliver this qualification	<ul style="list-style-type: none"> ❖ Hardware: A rack mounted server, Desktop as per batch size, One Projector ❖ Software: Windows with MS Office and Virtual Box with Linux Operating System.
15	Entry requirements and/or recommendations and minimum age	<ul style="list-style-type: none"> ❖ B.E./B.Tech in any stream Or ❖ DOEACC 'B' Level/ Or ❖ M.Sc. Or ❖ MCA Or ❖ Master Degree in Mathematics or Statistics or Operations Research/Economics or Econometrics/Applied Mathematics/Applied Statistics/ Or ❖ M.B.A. Or ❖ M.S. <p>Minimum Age: 20 Years</p>
16	Progression from the qualification (Please show Professional and academic progression)	<p>Academic:</p> <ul style="list-style-type: none"> i) Horizontal Progression: Specialized courses in the area of Data Analytics like working with Tableau, Business Intelligence, Power BI etc. ii) Vertical Progression: After completion of this course, students can go for higher studies in specialized courses of

Post Graduate Program in Data Engineering

		<p>NoSQL database like Apache Cassandra Administrator, Apache Cassandra Developer, and Hadoop Security etc. Candidates can also go for research in Data Science from reputed institutes.</p> <p>Professional: Statistical Analysts-> Data Analysts-> Data Scientists->Big Data Engineer->Hadoop Engineer->Hadoop Developer->Hadoop Administrator ->Business Analyst</p>
17	Arrangements for the Recognition of Prior learning (RPL)	<ul style="list-style-type: none"> ❖ Presently only candidates who undergo training shall be assessed. ❖ It will be incorporated once RPL strategy is finalized
18	International comparability Where known (research evidence to be provided)	❖ NA
19	Date of planned review of the Qualification.	❖ After Every 5 years
20	Formal structure of qualification (Table below)	

Module Code	Module Name	Mandatory/ Optional	Estimated Size (Learning Hours)	Level
DS601	Configuring Platform for Data Engineering	Mandatory	240	6
DS602	Data Analytics & Machine Learning	Mandatory	240	6
DS603	Big Data Analytics	Mandatory	240	6
DS604	Project	Mandatory	120	6

Detail Curriculum attached at **Annexure II**.

Post Graduate Program in Data Engineering
SECTION 1

ASSESSMENT

21	Body/Bodies which will carry out assessment: The Examination Section National Institute of Electronics and Information Technology NIELIT Bhawan, Plot No. 3, PSP Pocket, Sector-8, Dwarka, New Delhi-110077
22	How will RPL assessment be managed and who will carry it out? RPL Policy will be described as and when available.
23	Describe the overall assessment strategy and specific arrangements which have been put in place to ensure that assessment is always valid, reliable and fair and show that these are in line with the requirements of the NSQF. The emphasis is on practical demonstration of skills & knowledge based on the performance criteria. Student is required to pass in all OUTCOMES individually and marks are allotted. The Following assessment methodologies are used. A. Written Assessment (Multiple Choice Questions) B. Practical Assessment C. Internal Assessments D. Assignments E. Major Project The assessment results are backed by following evidences. 1. The assessor collects a copy of the attendance/online attendance for the training done under the scheme. The attendance sheets are signed and stamped by the course coordinator of the Training Centre. 2. The assessor verifies the authenticity of the candidate by checking the photo ID card issued by the institute as well as any one Photo ID card issued by the Central/Government. The same is mentioned in the attendance sheet. 3. The assessor assigns roll number.

Post Graduate Program in Data Engineering

	4. The assessor takes signature of all the students along with the assessor in a prescribed attendance sheet.
--	---

24. ASSESSMENT EVIDENCE

Title of Unit/Component: Configuring Platform for Data Engineering					
Outcomes to be assessed	Assessment Criteria for the outcome	Means of Assessment			
		Total Marks	Written	Practical	
Understand Linux Environment.	Able to understand architecture & different terminology used in Linux.	01	01	00	
Acquire Skills on Basic Linux Commands.	Understand CLI of Linux and able to run basic Linux commands & visual editor.	03	03	00	
Acquire Skills on Adding, removing Software's/packages.	Able to install and configure various packages/software's using Linux commands and changing mode of the operating system.	01	01	00	
Understand Important Folders and File Hierarchy-structure in Linux.	Understanding File Hierarchy and use of Important files and folders with their permissions in Linux Environment.	01	01	00	
Understanding BASH & writing Basic BASH Scripts.	<ul style="list-style-type: none"> ❖ Able to understand BASH, Variables and other constructs used in BASH Script. ❖ Able to write simple 	05	05	00	

Post Graduate Program in Data Engineering

	BASH Scripts.			
Understanding & Automating basic statistical calculations using BASH.	Able to understand advance concepts in BASH and implement basic statistical computations in a file using BASH.	10	05	05
Automating basic file handling using BASH.	Able to implement file handling operations using BASH.	10	05	05
Acquire Skills of Configuring Secures Shell (server and client) and password less connection between different servers.	Configuring LAN & Password less Connection between different servers using SSH.	01	01	00
Acquire knowledge on User and Group Management under Linux Operating System.	<ul style="list-style-type: none"> ❖ Managing multiple users & Groups in Linux Operating System. ❖ Understanding configuration files of Users and Groups. ❖ Managing Files i.r.o permissions and ownerships. 	04	04	00
Acquire skills on virtualization technique & creating virtual machines independently.	Ability to create virtual machines with Linux Operating System and allocate resources on this machine.	04	04	00
Acquire Skills on Basic Cloud & Information Security.	Understanding concepts of cloud and basic information security concepts.	01	01	00
Acquire skills of Java basics & configuring JAVA variables.	Able to understand JAVA programming environment and write simple JAVA	05	05	00

Post Graduate Program in Data Engineering

	programs.			
Acquire skills on Inheritance & exception handling and creating a JAR file.	<ul style="list-style-type: none"> ❖ Implementation of Exception Handling and Custom Exception handling in a Java Program. ❖ Writing multiple modules in JAVA and creating JAR file for the same. 	18	08	10
Acquire Skills on File handling & Serialization in Java.	Implementing File Handling and Serialization	04	04	00
Understand Basic concepts of data warehousing & data storage technique.	Understand concepts of how a database stores information via tables.	02	02	00
Understand basic concepts of MySQL and data base design.	<ul style="list-style-type: none"> ❖ Ability of create appropriate database and table. ❖ Loading Data in Table from various sources. 	07	05	02
Retrieve and manipulate data from one or more tables.	<ul style="list-style-type: none"> ❖ Identification of various keys. ❖ Skills of writing SQL query based on various conditions. ❖ Skills of writing SQL query by joining more than one table. ❖ Skills of writing Complex SQL query by using various functions of MySQL. 	16	08	08

Post Graduate Program in Data Engineering

Acquire Skills of Data Base backup & Recovery.	Implementation of backup and restore as and when required.	02	02	00
Understand Concepts of NoSQL Eco System & acquire knowledge on when to use NoSQL.	Able to understand NoSQL.	01	01	00
Understand Apache Cassandra and its architecture.	Able to understand various components of Apache Cassandra.	01	01	00
Acquire Skills on Installation & Configuration of apache Cassandra.	Able to install Apache Cassandra and configure it with appropriate parameters.	05	05	00
Understand Data Model used in Cassandra.	Understand Cassandra Data Model and Cassandra API.	01	01	00
Acquire Skills on creating Cassandra cluster & managing various Cassandra Services using available Monitoring Tools.	<ul style="list-style-type: none"> ❖ Able to understand Cluster and its layer. ❖ Able to Use Cluster Builder and monitor cluster statistics. 	04	04	00
Understand Cassandra Query Language & acquire skills on of executing commands in Cassandra.	<ul style="list-style-type: none"> ❖ Able to run Document Shell Commands in Cassandra. ❖ Able to run Data Definition Commands in Cassandra. ❖ Able to run Data Manipulation Commands in Cassandra. 	26	08	08
Acquire Skills on Data Types in Cassandra.	Use available and user defined data types in	07	05	02

Post Graduate Program in Data Engineering

	Cassandra			
Acquire Skills on running Complex Queries in Cassandra.	Implement Complex Queries in Apache Cassandra.	07	05	02
Implementing indexes in Distributed Database in Cassandra.	Understanding Distributed Data and implementing various types of Indexes.	02	02	00
Understand the Composite Column Concepts and its implementation in Cassandra.	Understanding and Implementation of Composite Columns.	01	01	00
Acquire Skills of Data Partitioning in Cassandra.	Understanding Partition Key and implementing Partition in Cassandra Cluster.	04	01	03
Acquire Skills on Cassandra Interfaces.	Implementation of Cassandra Interface	01	01	00
	Total	145	100	45

Post Graduate Program in Data Engineering

Title of Unit/Component: Data Analytics & Machine Learning					
Outcomes to be assessed	Assessment Criteria for the outcome	Means of Assessment			
		Total Marks	Written	Practical	
Understanding of Data Analytics.	Able to understand concepts of Data Analytics.	01	01	00	
Acquire basic skills on working with R.	Able to use Basic Syntax of R Programming.	01	01	00	
Understanding on Data Frames and Data handling in R.	<ul style="list-style-type: none"> ❖ Working knowledge of basic Data Sets. ❖ Able to use various R Packages and functions for reshaping data and its manipulation. ❖ Data Reshaping, Data Pipelines, and Data Manipulation. 	11	11	00	
Acquire Skills of writing simple python scripts.	Able to write simple Python Scripts using Conditional Statement, Loops, Functions, File Handling etc. in Python.	02	02	00	
Acquire Knowledge on Object Oriented Concept & implementation of Exception Handling in Python.	<ul style="list-style-type: none"> ❖ Writing Python script by using Object Oriented Approach. ❖ Able to handle Exceptions properly. 	02	02	00	
Understanding Pandas Data structures.	Able to use Python Data Analysis library (PANDAS) and explore various features Series, Data Frames,	01	01	00	

Post Graduate Program in Data Engineering

	Indexing etc..			
Acquire Skills on importing & exporting data from different data sources.	<ul style="list-style-type: none"> ❖ Able to Import from and Export to .csv/Excel Files. ❖ Manipulating Data. 	01	01	00
Acquire skills to handle missing value.	Handling Missing Value.	01	01	00
Acquire skills of Data aggregation.	<ul style="list-style-type: none"> ❖ Working with Data Aggregation and filtering Data. ❖ Reshaping of Data. 	01	01	00
Acquire Skills on Visualization of Data.	<ul style="list-style-type: none"> ❖ Using Python Library for graphical representation of Data & representation of various aspects of data using visualization techniques. 	02	01	01
Acquire Skills of Basic Statistical techniques used in Data Analytics.	Knowledge on basic Inferential statistics and statistical computations like random variable, Mathematical Expectations, Important Statistical Distributions used in Data Analytics.	07	05	02
Acquire Skills of Using SciPy Package for Inferential Statistics.	Able to understand various statistical functions available in SciPy library and perform statistical computations using these functions.	05	05	00
Acquire skills to understand the concept of Hypothesis testing and Goodness of fit Test and its implementation.	<ul style="list-style-type: none"> ❖ Understanding and identifying Null & Alternate Hypothesis for a given data. ❖ Identification of Level of 	25	15	10

Post Graduate Program in Data Engineering

	<p>Significance and Critical Region.</p> <ul style="list-style-type: none"> ❖ Finding P Value. ❖ Making Decision based on Data. 			
Acquire Skills of understanding time series data.	<ul style="list-style-type: none"> ❖ Understanding whether a data is time series or not? ❖ Importing Time Series Data into Python and working with Time Series libraries like autots, tsfresh, dart, atspy etc. 	10	05	05
Acquire Mathematical skills for analysing Time Series Data.	Understanding various mathematical functions & its implementation in Time Series analysis.	15	10	05
Acquire Skills of visualisation of Time Series Data.	Graphical representation of time series data.	05	05	00
Acquire Skills for forecasting a time series data.	Implementing forecasting methods in Time Series Data.	05	05	00
Understand Machine Learning.	<ul style="list-style-type: none"> ❖ Understanding on Machine Learning Concept and its application in different domains. ❖ Working Knowledge of Supervised and Unsupervised machine learning techniques. 	10	05	05

Post Graduate Program in Data Engineering

Implementation of Machine Learning algorithms on given data.	<ul style="list-style-type: none"> ❖ Implementation Knowledge of Machine Learning Algorithms. ❖ Able to identify Outlier from given data & implementation of Association rule learning. ❖ Able to implement Text mining technique. 	10	05	05
Understanding of Neural Network & building Simple Neural Networks.	<ul style="list-style-type: none"> ❖ Implementation Knowledge of building simple Neural Network in Python. ❖ Technical Skills on Error Analysis. 	06	04	02
Acquire Knowledge of “how to train or design the neural networks”.	Technical Skills to train a neural network by identifying different layers and corresponding weight etc.	04	02	02
Acquire Skills on “How to Use Neural network” in Data Analytics.	Technical Skills to use Neural Network for pattern extracting from different sources in Data Analytics.	02	01	01
Understand Artificial Neural network & Deep Learning.	Basic Knowledge of Artificial Neural Network and Deep Learning.	04	02	02
Acquire knowledge to Implement deep learning algorithms and solve real world problems.	Working Knowledge to use KERAS and Tensor flow.	08	03	05
	Total	145	100	45

Post Graduate Program in Data Engineering

Title of Unit/Component: Big Data analytics				
Outcomes to be assessed	Assessment Criteria for the outcome	Means of Assessment		
		Total Marks	Written	Practical
Understanding Big Data.	Concepts of Big Data & processing bigdata.	02	02	00
Acquire in-depth knowledge of Configuring Basic and Multi-node Hadoop Cluster.	<ul style="list-style-type: none"> ❖ Understanding of Hadoop architecture & it's working. ❖ Technical Skills on configuring Basic & Multi-Node Hadoop Cluster. ❖ Able to run Basic Hadoop Commands. 	11	06	05
Acquire knowledge of Hadoop Administration for managing Hadoop Server.	Technical Skills of running various Hadoop administration commands.	15	10	05
Understand the concepts of MapReduce and Big Data.	<ul style="list-style-type: none"> ❖ Skills on HDFS Java API with Map Reduce Framework. ❖ Understanding on YARN MR Application flow. 	08	08	00
Acquire knowledge on writing & deploy Map reduce programs in Hadoop Cluster.	<ul style="list-style-type: none"> ❖ Technical skills of writing MapReduce Codes in Python. ❖ Deployment of Map Reduce Technology in Hadoop Cluster. ❖ Skills on Cassandra 	13	08	05

Post Graduate Program in Data Engineering

	<p>Hadoop Integration.</p> <ul style="list-style-type: none"> ❖ Technical skills of using Hadoop Map reduce feature in Cassandra. 			
Acquire Knowledge on Configuring Pig in Hadoop Eco System & write Pig Scripts for unstructured and unorganized data.	Technical Skills on configuring Pig and writing Pig Scripts.	03	03	00
Acquire Knowledge of implementing SQL on huge volume of unstructured & Un-organized data in Hadoop environment using Map reduce.	Writing SQL & executing SQL using Map Reduce technique.	10	05	05
Acquire Knowledge on Configuring Hive & implement Data warehousing on large volume of data without using Map Reduce Technique.	<p>Skills on Configuring Hive & understanding its architecture.</p> <p>Skills of writing Hive Query and storing query results.</p>	10	05	05
Implementation of Statistical Techniques in Data warehousing.	Skills of implementing Statistical function in Hive Query.	09	04	05
Comprehensive understanding of HBase architecture and its configuration in Hadoop ecosystem.	<ul style="list-style-type: none"> ❖ Understanding Architecture of HBase, various modes of HBase and its Data Model. ❖ Working knowledge of Zookeeper, Master and Region servers. 	06	06	00
Technical Skills on understanding usage of HBase Commands and	❖ Technical Skills of executing HBase Commands,	14	09	05

Post Graduate Program in Data Engineering

implementation of MapReduce technique in HBase.	<p>understanding HBase Constraints.</p> <ul style="list-style-type: none"> ❖ Technical Skills of implementing MapReduce technique with HBase. 			
Acquire Skills of writing program in Scala.	<ul style="list-style-type: none"> ❖ Technical Skills on understanding basic Scala operations, Variable types and Control Structures. ❖ Technical Skills on Functional Scala for Data Science. ❖ Technical Skills on Implementing Object Oriented Programming & understanding Scala Collection API. 	11	06	05
Acquire Skills on Configuring Spark.	<ul style="list-style-type: none"> ❖ Technical Skills of Spark Installation and Configuration. ❖ Understanding Spark Cluster Modes on YARN. 	02	02	00
Acquire skills of working with RDDs.	Understand RDD and various operations of RDD.	14	04	10
Acquire skills on SparkSQL & deploying Spark applications.	<ul style="list-style-type: none"> ❖ Technical Skills on writing Spark applications using Data Frames and SparkSQL. ❖ Configuring Spark Properties & Deployment of Spark 	16	06	10

Post Graduate Program in Data Engineering

	Applications.			
Acquire skills on different Machine Learning Techniques/algorithms, and Spark-MLlib.	<ul style="list-style-type: none"> ❖ Technical skills on using MLlib & its various features. ❖ Implementing Supervised & Un-supervised Machine Learning in Spark. 	12	02	10
Acquire Skills on understanding architecture of Apache Kafka.	<ul style="list-style-type: none"> ❖ Understanding of Apache Kafka architecture and components of Kafka Cluster. ❖ Technical Skills on Configuring Kafka Cluster. 	07	02	05
Acquire Skills on working with Apache Kafka.	<ul style="list-style-type: none"> ❖ Technical Skills on understanding concept of messages. ❖ Technical Skills on implementing Kafka Producer and Kafka Consumer. ❖ Technical Skills on implementing Serialization using Apache Avro. ❖ Technical Skills on implementing Stream processing using Apache Kafka. ❖ Technical Skills on integration of Apache 	09	04	05

Post Graduate Program in Data Engineering

	Spark and Kafka for stream processing.			
Understanding Apache Flume Concept and its purpose.	<ul style="list-style-type: none"> ❖ Technical Skills on configuration of Apache Flume. ❖ Will be able to understand purpose of Flume and Kafka. 	07	02	05
Acquire Skills on working with Apache Flume and its integration with Kafka.	<ul style="list-style-type: none"> ❖ Technical Skills of using Flume to move large amounts of unstructured data from various sources to HDFS. ❖ Technical Skills of using Flume Sources to write to Kafka and Write to Flume Sinks reading from Kafka. 	04	04	00
Acquire Skills of Spark GraphX programming.	Technical Skills of implementing graph-parallel computation by using Spark-GraphX component.	07	02	05
	Total	190	100	90

Means of assessment

S. No	Examination Pattern	Modules Covered	Duration in Minutes	Maximum Marks
1	Theory 1: Basic Linux, Java & Data Warehousing	Module 1	90	100
2	Theory 2: Data Analytics & Machine Learning	Module 2	90	100
3	Theory 3: Big Data Analytics	Module 3	90	100
4	Practical 1: Basic Linux, Java, Data Warehousing & Data Analytics	Module 1 & Module 2	180	90
5	Practical 2: Big Data Analytics	Module 3	180	90
6	Internal Assessment	Module 1,2,3	-	60
7	Assignment	Module 1,2,3	-	60
8	Major Project	Module 1,2,3	-	100
Total				700

Theory Papers:

1. Theory 1: Basic Linux, Java & Data Warehousing
2. Theory 2: Data Analytics & Machine Learning
3. Theory 3: Big Data Analytics

Practical Papers:

1. Practical 1: Basic Linux, Java, Data Warehousing & Data Analytics.
2. Practical 2: Big Data Analytics.

Note:

1. Pass percentage would be 50% marks in each component, with aggregate pass percentage of 50% and above.
2. Grading will be as under:

Grade	S	A	B	C	D
Marks Range (in %)	$\geq 85\%$	$\geq 75\%$ and $< 85\%$	$\geq 65\%$ and $< 75\%$	$\geq 55\%$ and $< 65\%$	$\geq 50\%$ and $< 55\%$

3. Theory examination would be conducted online and the paper comprises of MCQ and each question will carry 1 mark.
4. Practical examination/Internal Assessment/ Project/Presentation/Assignment would be evaluated internally.
5. Major Project/Dissertation would be evaluated preferably by External / Subject Expert including NIELIT Officials.

NSQF Qualification File Approved in 17th NSQC, dated- 31/03/22

6. Candidate may apply for re-examination within the validity of registration.
7. The examinations would be conducted in English Language only.

SECTION 2

25. EVIDENCE OF LEVEL

Title: Post Graduate Program in Data Engineering / NIELIT Certified Data Scientist			Level : 6
Unit 1: Linux , Java & Data Ware Housing			
NSQF Domain	Outcomes of the Qualification/Component	How the Job role relates to the NSQF level descriptor	NSQF Level
Process	Acquiring Skills on Bash Scripting & Data Warehousing	<ul style="list-style-type: none"> ● In Linux, candidate will acquire skills on wide range of commands for configuring Linux for Data Analytics. ● They will acquire specialized skills of writing BASH scripts for preprocessing the data and simple descriptive analytics. ● With the Java programming, they acquire the special skills used in configuration of Hadoop Server and working with Apache Spark. ● In Data Warehousing participants will learn theoretical and practical skills on structured and NoSQL Databases. 	6
Professional Knowledge		<ul style="list-style-type: none"> ● As a professional knowledge, the candidates will understand Linux Environment, they will also acquire knowledge on adding, removing software's/packages. ● They will be able to work with various Linux Commands, understand Directory and File Hierarchy-structure in Linux. They acquire knowledge on User and Group Management under Linux Operating System. 	6

		<ul style="list-style-type: none">● In the BASH scripting, they will understand factual and theoretical knowledge of writing scripts for automating basic statistical calculations and automating basic file handling.● IN SSH, they will learn theoretical and practical skills to configuring Secures Shell (server and client) and password less connection between different servers.● They acquire skills on virtualization technique & creating and managing virtual machines independently.● In Java, candidate will acquire skills on defining class and its relationships, working on arrays and strings, creating jar files for the java programs, usage of collection, serialization, input into files, reading and writing into the files.● They will learn comprehensive skills to write program in Java language which will be used in map reducing technique. They will acquire practical skills for dealing with the exceptions while configuring the Hadoop cluster and working with Apache Spark.● In Data Warehousing, participants will acquire industry relevant Knowledge for Structured and NoSQL Databases.● Apart from this they will learn theoretical & practical skills to configure Cassandra and its different components. They will acquire practical skills for configuration of standalone Cassandra Server and server in clustered mode with replication.● After acquiring knowledge of Data warehousing, they will also learn Comprehensive knowledge to Export and import of archived data bases and their tables for analysis.● They will have practical skills on relating database tables using keys and view the resultant data & perform data analytics on retrieved result.	
--	--	---	--

Professional Skill		<ul style="list-style-type: none"> • After learning Linux, participants will have wide range of cognitive and practical skills to configure Linux for Data Analytics & develop creative solution for pre-processing the data & basic Descriptive Analytics using BASH. • With Database, they will be able to bring up the distributed setup in the local environment on their own. Make the connectivity with the databases and write queries in Java for retrieving and manipulation of data for data-warehousing. • They will have professional skills of writing optimized queries using both (SQL and NoSQL) query languages for data analysis. • They will also have theoretical knowledge and practical skills to develop Cassandra cluster and use monitoring tools to monitor Cassandra. 	6
Core Skill		<ul style="list-style-type: none"> • After acquiring Core skills in BASH Scripting and Data warehousing, candidates will learn logical and mathematical skills for descriptive analytics and data preprocessing. • The participants will have in-depth knowledge in a broad context for Linux servers, programming language and structured and non-structured databases. They will also learn wide ranging theoretical and practical skills in configuration of cloud techniques, information security and virtualization techniques for data analytics platform. They acquire skill to connect to database using Java database connection. 	7
Responsibility		<ul style="list-style-type: none"> • Study of these skills will be helping the candidates in choosing the suitable technology for the deployment of any problem solution in the industry. They will be able to choose between the SQL and NoSQL databases. 	6

		<ul style="list-style-type: none"> • In other situations, as per present day, data come from different forms for analytics, hence they will also be able to work in the combination of both the databases. • Will be able to create multiple server machine through virtualization, work in the latest technology of cloud computing. Will have the knowledge to do replication of data, thus by addressing the failure management and risk areas without redundancy in data management. In the field, one will have the knowledge to address the application of geo-redundant data also. 	
Unit 2: Data Analytics and Machine Learning			
Process	Acquiring Skills on Data Analytics and Machine Learning using R and Python.	<ul style="list-style-type: none"> • After acquiring knowledge of Data Analytics using Python & R candidate will have advanced knowledge of statistical analysis. • They will be able to understand the various statistical concepts used in Data Analytics. By doing this they will have advanced knowledge and critical understanding of statistical distributions and their fitting in different scenarios. They will also learn the implementation techniques for these distributions using Python. • After completion of the course they will able to demonstrate the hypothesis to be used for analysis. • They can develop their own statistical model for a specific analysis using existing statistical distributions. • After acquiring the skill they will acquire theoretical and practical skills of implementing Machine Learning techniques in Data Analytics. 	6
Professional Knowledge		<ul style="list-style-type: none"> • As a professional knowledge, the candidates will acquire theoretical knowledge and practical skills of R programming and will have knowledge of constructing Data frames and using various packages related to data science. 	7

		<ul style="list-style-type: none"> • In Python programming, they will acquire knowledge in OOPS concept and features of python, knowledge in Panda data structures, series, data frames. They will be able to access data through indexing, slicing, data aggregation, group by, indexing concepts, will work with JSON and XLS objects importing and manipulation, they will be able to work with Data Visualization using Python libraries. They will also acquire skills to work with inferential statistical analysis, understanding of mathematical expectation, distribution function. Skills to work with the hypothesis techniques Null and Alternative hypothesis. Understand the critical region, level of significance using Python programming • In time series data analysis, they gain theoretical knowledge and practical skills to develop creative solutions in time series domain by understanding various topics in Time Series. • In Machine Learning, they will acquire theoretical knowledge and practical skills in the basic Concepts of Machine Learning, application of Machine Learning, types of Machine Learning using supervised and UN supervised types, data mining on the data with the help of trees, decision tree, splits, entropy and etc. • By acquiring skills of Neural Network and deep learning participants will acquire theoretical knowledge and practical skills of implementing AI techniques in analytics. 	
Professional Skill		<ul style="list-style-type: none"> • The participants will acquire skills to create various Data Sets independently in R. • They will have theoretical and practical skills of Reshaping Data in desired format. • They will learn about the usage of various analytical libraries and will be able to create their own library. They will acquire skill on various visualisation techniques. • They will acquire mathematical and logical skills to construct their own statistical model and 	6

		<p>understand the fitting of model for the given data.</p> <ul style="list-style-type: none"> • They can import Time Series Data in Python and test for Stationary of Time Series using the techniques, forecasting of the time series data, handling missing values in time series data analysis, lag plots and perform causality test for time series data. • They will acquire skills to apply various AI Techniques in Data Analytics. 	
Core Skill		<ul style="list-style-type: none"> • After acquiring the skills on data analytics using Python, candidates will enhance their skills related to distribution fitting. They will be able create a data frame from existing data sources. They will also be able to identify critical region and acceptance region for a particular analysis. They will learn various packages in Python for making decisions more efficiently. • Acquire knowledge to Implement deep learning algorithms and solve real world problems 	6
Responsibility		<ul style="list-style-type: none"> • In the real world, the candidate can do Data Analytics using Python, they can apply their skills to identify a proper test method depending upon the current situations. • They will get confidence of handling complex situations related to technical activities in Data Analytics. They can create their own neural network and apply the analytics technologies to arrive at a problem solution. • Understand and Implement Artificial Neural network. • Identify deep learning algorithms which are more appropriate for various types of learning tasks in various domains. 	6
Unit 3: Big Data Analytics			
Process	Implementation of Big Data Technology	<ul style="list-style-type: none"> • In this Module, candidate will acquire theoretical knowledge and Practical Skills of implementing Big Data Technology using Hadoop Eco System. Candidate will learn the various 	6

	<p>using Hadoop Eco System</p>	<p>components of Hadoop Eco System, Hadoop architecture, usage and implementation of the components in Big Data.</p> <ul style="list-style-type: none"> • After completion of the course they will be able to demonstrate the working of Hadoop Eco System. They will be able to work with various Big data technologies and programming languages like Python, Scala, Apache Spark, Kafka & Flume, in details about Spark. 	
<p>Professional Knowledge</p>		<p>The candidates will acquire professional knowledge in Big Data techniques like Hadoop ecosystem HDFS file system, working with Pig, Hive, HBase Data ware houses, Apache Spark, Apache Kafka, and streaming technologies like Apache Flume.</p> <ul style="list-style-type: none"> • In the Hadoop framework, the candidate will acquire knowledge in core components of the Hadoop ecosystem. They will be able to set up the Hadoop cluster and monitor the health of the cluster. They will acquire knowledge in the HDFS file system, its architecture, its design, and able to define its role in Hadoop. They will be able to work with Hdfs shell using CLI. • In the MapReduce concept of Hadoop, candidates will understand theoretical knowledge and practical skills to implement map reduce using Python. They will also acquire knowledge to integrate Apache Cassandra with Hadoop, and perform MapReduce operation with Cassandra. • They will acquire professional knowledge of performing parallel processing using Hadoop Components. • Like Hadoop, candidates will also acquire knowledge to work with Apache Spark [Scala & Python]. In Spark, they will learn the installation, configuration of Spark, cluster modes on YARN, loading, and saving of data, Spark shell, lambda operations, associative property, and caching of data. They will also understand the backbone of Spark RDD and the operations of RDD. In RDD, the candidate will understand different types of 	<p>6</p>

		<p>key-value pairs. They will also acquire knowledge to work with data frames, Spark SQL architecture and statements, integration of Spark and Hive. They will learn the deployment of SparkContext, building Spark application with Scala and Web UI with a case study.</p> <ul style="list-style-type: none"> • Candidates will be able to implement machine learning techniques using Spark MLLib. • In Apache Kafka, candidates will acquire the basic knowledge in introduction with Kafka, its architecture, and components, they will also acquire knowledge to use Kafka for website activity tracking, providing operational tracking data, log aggregation, stream processing, event sourcing, as a replacement for a message broker, and as an external commit log for distributed systems. 	
<p>Professional Skill</p>		<ul style="list-style-type: none"> • In each area of Big data analytics, after the completion of the course, the candidates will acquire skills specialized skills in related technology. • On completion of the module, candidates will acquire skills of configuring the Hadoop environment, in standalone, pseudo-distributed mode, and in fully distributed mode, will be able to set up the Hadoop environment to work in Big Data Analysis. They will also acquire knowledge on writing & deploy MapReduce programs and concepts in different use-cases. • In the data warehouse, the candidate will acquire skills to configure Pig, Hive, and Hbase in the Hadoop environment & write Pig Scripts for unstructured and unorganized data. They will be able to write SQL scripts, use Hive and perform data warehousing for large volumes of data. They will also be able to do statistical analysis on streaming data and in data warehouses. • The candidates can develop case studies - Log Data and Timeseries Data for the 	<p>7</p>

		<p>Customer/Order data. They will be able to use Hbase and MapReduce for that purpose.</p> <ul style="list-style-type: none"> • After understanding Spark, they can apply Spark algorithms for machine learning, deploy the spark streaming applications, integrate with Hadoop Scala/PySpark-Environment. In addition, they will be able to implement various algorithms supported by MLlib such as Linear Regression, Decision Tree, Random Forest, and so on. • Candidates will also acquire skills to configure different types of Kafka Clusters, Flume-Kafka integration for event processing. Thus, acquire skills, in processing large-scale data using apache-spark, apply techniques of manipulating data with Spark and Scala, acquire technical skills on expressing algorithms for data analysis in a functional style. • They will acquire skills in ingest streaming data using flume, implement different GraphX algorithms and their implementations. 	
Core Skill		<ul style="list-style-type: none"> • With acquiring the skills, candidates will enhance their cores skills related to analytics for big data. They will be able to create a Hadoop eco system with all related tools for big data analytics. They will be able to create a data base server for unstructured data and will be skilled with analysis of unstructured data. • They will also be able to identify the suitability of various Hadoop eco system tools and work on Pig Hive, Hbase databases in combined form or apply as per the design proposal. • They will be able to use/write Deep Learning Algorithms to optimize Data Analytics. They will get confidence of handling complex situations related to technical activities in different data science environment. 	6

Responsibility		<ul style="list-style-type: none"> • After completion of the module, the candidates will be able to work with full responsibilities on the technologies for big data analysis from the available frameworks. They will be able to find the solutions using the suitable frameworks of Hadoop and Spark ecosystem application in Data science. • They will be able to work independently or as a team's contribute more on the analysis and decision making process. The broader knowledge of the Big data analysis techniques will provide confidence for the candidates to involve to finding the suitable solutions for any problem description. 	6
Unit 3: Mini Project			
Process		As a part of the course completion of the course, the participants will have to complete a Mini industry-oriented project of their choice. The output of the project will be analysed based on the data analytics techniques, framework used and the analysis done on the data. They will also use the suitable data warehouses for their projects and show cast the analysis done. The project output will be analysed based on the application of their knowledge in big data analytics and data science and evaluated accordingly.	7
Professional Knowledge		By executing this mini project, the candidates' level of understanding and learning of all the tools and technologies for big data analysis will be evaluated on their own.	6
Professional Skill		<ul style="list-style-type: none"> • Candidates will able to do mini project in Linux machine installing Sql or no-sql databases, perform statistical analysis of data using R, use Python Pandas for machine learning, install Hadoop ecosystem, deploy HDFS. They will also be able to work with large volume of data using Pig, Hbase, Hive. They can create neural 	6

		<p>networks or artificial neural networks with the skills obtained from the course.</p> <ul style="list-style-type: none"> • By presenting the project, they will gain confidence of appearing in various job interview and it will increase their interview skills. 	
Core Skill		<p>They will be able to choose between the data warehouses, implement the Hadoop environment, configure the databases, perform the data analysis, by plotting graph and charts. They will also be able to clean up the data from different resources using pandas or map reducing techniques. Use the formatted data to arrive at a decision for the business needs.</p>	6
Responsibility		<p>Acquiring skills of developing a live project in Big Data domain, candidate will enhance their skills of decision making in various complex scenarios. They will be able to take an appropriate decision in any unpredictable situations. Their decision making and analytical capabilities will be enhanced manifold.</p>	6

SECTION 3

EVIDENCE OF NEED

26	<p>What evidence is there that the qualification is needed?</p> <p>This course has been designed to meet the increasing manpower requirements in Data science industry after discussion with our alumni working in various Top Industries across in India and also various reports, survey from Industry bodies, government think tanks were referred.</p> <ol style="list-style-type: none"> Edge Analytics in 2021: What it is, Why it matters & Use Cases - https://research.aimultiple.com/edge-analytics/ NASSCOM Reskilling Series: The Journey of CGI-2020 - https://futureskills.nasscom.in/research.html Market Survey – https://www.grandviewresearch.com/industry-analysis/machine-learning-market IET Skills and Demand in Industry 2019 Survey-
----	---

	<p>https://www.theiet.org/media/4812/skills-survey2019.pdf</p> <p>e) Towards a Reskilling Revolution-World Economic Forum-2019 - http://www3.weforum.org/docs/WEF_FOW_Reskilling_Revolution.pdf</p> <p>f) Evidence of Requirement in the Industry https://insidebigdata.com/2020/12/21/big-data-industry-predictions-for-2021/</p>
27	<p>What is the estimated uptake of this qualification and what is the basis of this estimate?</p> <p>Estimated uptake is 60 students / Batch / centre with 2 Batches / Year and on the basis of Facilities and Infrastructure in respective NIELIT Centre.</p> <p>Program may also run using Virtual Lab setup for cloud software's and remote access facility available with NIELIT Chennai.</p>
28	<p>What steps were taken to ensure that the qualification(s) does (do) not duplicate already existing or planned qualifications in the NSQF? Give justification for presenting a duplicate qualification</p> <p>Similar Qualification does not exist as per information available in NQR portal.</p>
29	<p>What arrangements are in place to monitor and review the qualification(s)? What data will be used and at what point will the qualification(s) be revised or updated? Specify the review process here</p> <p>Based on feedback by participants, employers and based on market survey the qualification will be reviewed in every 5 years.</p>

SECTION 4

EVIDENCE OF PROGRESSION

30	<p>What steps have been taken in the design of this or other qualifications to ensure that there is a clear path to other qualifications in this sector?</p> <p>This qualification has been designed in consultation with industry and domain expert keeping in mind today's need. Evaluation criteria have been added to ensure progression to related path ways identified as per career path.</p>
-----------	--

SECTION 5

EVIDENCE OF INTERNATIONAL COMPARABILITY

List any Comparisons which have been established

1. Course: Coursera – IBM Data Science Professional Certificate
Source: <https://www.coursera.org/professional-certificates/ibm-data-science>
2. Course : M.Sc. Data Science (University of London)
Source: <https://london.ac.uk/msc-data-science-structure>
3. Course: PG Program in Data Science and Business Analytics (Great Learning)
Source: <https://www.greatlearning.in/>

Annexures

Annexure-I: Evidence of Job

Annexure-II: Detailed Syllabus

Annexure-III: Industry validations

Annexure-O

NSQF Approved