QUALIFICATION FILE–Standalone NOS

Essentials of Big Data

☐ Horizontal/Generic ☐Vertical/Specialization

☒ Upskilling ☐ Dual/Flexi Qualification ☐ For ToT ☐ For ToA

☐General ☐ Multi-skill (MS) ☐ Cross Sectoral (CS) ☒ Future Skills ☒ OEM

NCrF/NSQF Level: 5

Submitted By:

National Institute of Electronics and Information Technology (NIELIT)

NIELIT Bhawan,
Plot No. 3, PSP Pocket, Sector-8,
Dwarka, New Delhi-110077,
Phone: - 91-11-25308300
e-mail: - contact@nielit.gov.in

Table of Contents

Section 1: Basic Details

1.	**NOS-Qualification Name**	**Essentials of Big Data**	
2.	**Sector/s**	**IT-ITeS**	
3.	**Type of Qualification ☒ New ☐ Revised**	**NQR Code & version of the existing /previous qualification: NA**	**Qualification Name of the existing/previous version: NA**
4.	**National Qualification Register (NQR) Code & Version**	NG-05-IT-01427-2023-V1-NIELIT & Version 1	**5. NCrF/NSQF Level:** 5
6.	**Brief Description of the Standalone NOS**	The "Essentials of Big Data" upskilling course offers a comprehensive exploration of fundamental concepts and technologies in the realm of big data. Participants delve into key components such as Hadoop, Spark, and distributed computing, gaining a profound understanding of how to manage and analyze vast datasets efficiently. Through hands-on exercises and real-world case studies, learners develop the skills necessary to harness the power of big data technologies, enabling them to extract valuable insights and make informed decisions. This course is ideal for professionals seeking to enhance their expertise in the dynamic field of big data, equipping them with the knowledge to navigate large-scale data challenges and contribute to innovative solutions in various industries. Throughout the program, participants engage with practical scenarios, implementing big data technologies to solve complex problems. The curriculum covers topics such as data storage, processing, and analysis, ensuring a well-rounded understanding of the tools and techniques essential for working with big data. By the course's conclusion, participants emerge with a solid foundation in big data essentials, positioning them to excel in roles that require proficiency in managing and extracting meaningful insights from massive datasets.	

<table>
<tr><td>7.</td><td>Eligibility Criteria for Entry for a Student/Trainee/Learner/Employee</td><td colspan="2">a. Entry Qualification &Relevant Experience:
<table>
<tr><th>S. No.</th><th>Academic/Skill Qualification (with Specialization - if applicable)</th><th>Relevant Experience (with Specialization - if applicable)</th></tr>
<tr><td>1</td><td>Pursuing^ Final Year B.Tech in any branch of Engineering*

Or

Pursuing^ Final Year MCA
Or

Pursuing^ Final Year B.Sc. in any branch of Sciences*
Or

Pursuing^ Final Year B.Sc. in IT/CS/Electronics/allied subjects

*Students should have relevant knowledge of data analysis and programming concepts.

#Students with the above entry requirements are eligible to take the course subject to clearing the written test comprising of Analytical Reasoning, Mathematics and English

^Passout students in the above entry requirements are also eligible for the course.</td><td>NA</td></tr>
</table>
b. Age: No bar</td></tr>
<tr><td>8.</td><td>Credits Assigned to this NOS-Qualification, Subject to Assessment (as per National Credit Framework (NCrF))</td><td>3</td><td>9. Common Cost Norm Category (I/II/III) (wherever applicable): Category-II</td></tr>
<tr><td>10.</td><td>Any Licensing Requirements for Undertaking Training on This Qualification (wherever applicable)</td><td colspan="2">Not Applicable</td></tr>
</table>

<table>
<tr><td>11.</td><td>Training Duration by Modes of Training Delivery(Specify Total Duration as per selected training delivery modes and as per requirement of the qualification)</td><td>☒ Offline ☐ Online ☐ Blended

<table>
<tr><th>Training Delivery Mode</th><th>Theory (Hours)</th><th>Practical (Hours)</th><th>Total (Hours)</th></tr>
<tr><td>Classroom (offline)</td><td>40</td><td>50</td><td>90</td></tr>
</table>

The mode of delivery shall be based on the regional demand and can be offered in any of the above modes mentioned.</td></tr>
<tr><td>12.</td><td>Assessment Criteria</td><td>
<table>
<tr><th>Theory (Marks)</th><th>Practical (Marks)</th><th>Project (Marks)</th><th>Viva (Marks)</th><th>Total (Marks)</th><th>Passing %age</th></tr>
<tr><td>100</td><td>0</td><td>0</td><td>0</td><td>100</td><td>50</td></tr>
</table>

The centralised online assessment is conducted by the Examination Wing, NIELIT Headquarters.</td></tr>
<tr><td>13.</td><td>Is the NOS Amenable to Persons with Disability</td><td>☒ Yes ☐ No
If “Yes”, specify applicable type of Disability:

a. Locomotor Disability: Leprosy Cured Person, Dwarfism, Muscular Dystrophy and Acid Attack Victims
b. Visual Impairment: Low Vision</td></tr>
<tr><td>14.</td><td>Progression Path After Attaining the Qualification, wherever applicable</td><td>Big Data Analyst
Business Intelligence Analyst
Data Architect
Data Analyst
Big Data Consultant
Big Data Developer</td></tr>
<tr><td>15.</td><td>How will the participation of women be encouraged?</td><td>Participation by women can be ensured through Government Schemes. Occasionally, exclusive batches for women would be run for the proposed courses. Funding is available for women’s participation under other schemes launched by the Government from time to time.</td></tr>
</table>

16.	**Other Indian languages in which the Qualification & Model Curriculum are being submitted**	Only English	
17.	**Is similar NOS available on NQR-if yes, justification for this qualification**	☐**Yes** ☒ **No**	
18.	**Name and Contact Details Submitting / Awarding Body SPOC** *(In the case of CS or MS, provide details of both Lead AB & Supporting ABs)*	A. Name: SHRI NILADRI DAS Position in the organization: Scientist E Address: NIELIT Agartala Tel number(s): 8794028299 E-mail address: niladridas@nielit.gov.in B. Name: SHRI BINOY DAS Position in the organization: Senior Technical Officer Address: NIELIT Agartala Tel number(s): 8794822459 E-mail address: erbinoy@nielit.gov.in	
19.	**Final Approval Date by NSQC: 30/11/2023**	**20. Validity Duration: 3 years**	**21. Next Review Date: 30/11/2026**

Section 2: Training Related

1.	**Trainer's Qualification and experience in the relevant sector (in years) (as per NCVET guidelines)**	B.Tech or Equivalent as per NCrF with 15+ years of experience
2.	**Master Trainer's Qualification and experience in the relevant sector (in years) (as per NCVET guidelines)**	B.Tech or Equivalent as per NCrF with 15+ years of experience
3.	**Tools and Equipment Required for the Training**	☒ Yes ☐No Available at Annexure-II
4.	**In Case of Revised NOS, details of Any Upskilling Required for Trainer**	Not Applicable

Section 3: Assessment Related

1.	**Assessor's Qualification and experience in relevant sector (in years)** *(as per NCVET guidelines)*	B.Tech or Equivalent as per NCrF with 15+ years of experience
2.	**Proctor's Qualification and experience in relevant sector (in years)** *(as per NCVET guidelines), (wherever applicable)*	The assessor carries out theory online assessments through the remote proctoring methodology. Theory examination would be conducted online and the paper comprises MCQ. Conduct of assessment is through trained proctors. Once the test begins, remote proctors have full access to the candidate's video feeds and computer screens. Proctors authenticate the candidate based on registration details, pre-test image captured and I-card in possession of the candidate. Proctors can chat with candidates or give warnings to candidates. Proctors can also take screenshots, terminate a specific user's test session, or re-authenticate candidates based on video feeds.
3.	**Lead Assessor's/Proctor's Qualification and experience in relevant sector (in years)** *(as per NCVET guidelines)*	External Examiners/ Observers (Subject matter experts) are deployed including NIELIT scientific officers who are subject experts for evaluation of Practical examination/ internal assessment / Project/ Presentation/ assignment and Major Project (if applicable). Qualification is generally B.Tech.
4.	**Assessment Mode***(Specify the assessment mode)*	Centralized online examination will be conducted
5.	**Tools and Equipment Required for Assessment**	Same as for training ☒ Yes ☐ No

Section 4: Evidence of the Need for the Standalone NOS

1.	Government /Industry initiatives/ requirement (Yes/No): Yes.
2.	Number of Industry validation provided: 4
3.	Estimated number of people to be trained: 1000 persons per year shall be trained.
4.	Evidence of Concurrence/Consultation with Line/State Departments (In case of regulated sectors): No NIELIT is recognised as AB and AA under the Government Category. NIELIT is the HRD arm of MeitY, GoI.

Section 5: Annexure & Supporting Documents Check List

Specify Annexure Name / Supporting document file name

1.	**Annexure:** NCrF/NSQF level justification based on NCrF/NSQF descriptors *(Mandatory)*	*Available at* Annexure-I: Evidence of Level
2.	**Annexure:** List of tools and equipment relevant for NOS *(Mandatory, except in case of online course)*	*Available at* Annexure-II: Tools and Equipment
3.	**Annexure: Industry Validation**	*Available at* Annexure-III: Industry Validation
4.	**Annexure: Training Details**	*Available at* Annexure-IV: Training Details
5.	**Annexure:** Blended Learning *(Mandatory, in case the selected Mode of delivery is Blended Learning)*	*Available at* Annexure-V: Blended Learning
6.	**Annexure/Supporting Document:** Standalone NOS- Performance Criteria Details Annexure/Document with PC-wise detailing as per NOS format (Mandatory- Public view)	*Available at* Annexure-VI: Performance Criteria
7.	**Annexure:** Performance and Assessment Criteria *(Mandatory)*	*Available at* Annexure-VII: Detailed Assessment Criteria
8.	**Annexure:** Assessment Strategy *(Mandatory)*	*Available at* Annexure-VIII: Assessment Strategy
9.	**Annexure:** Acronym and Glossary *(Optional)*	*Available at* Annexure-IX: Acronym and Glossary
10.	**Supporting Document:** Model Curriculum	*Available at* Annexure-A: Model Curriculum

Annexure-I: Evidence of Level

NCrF/NSQF Level Descriptors	Key requirements of the job role/ outcome of the qualification	How the job role/ outcomes relate to the NCrF/NSQF level descriptor	NCrF/NSQF Level
Professional Theoretical Knowledge/Process	This course demands a solid theoretical foundation in several key areas. Professionals must acquire a deep understanding of distributed computing principles and technologies, including Hadoop and Spark frameworks. Theoretical knowledge of big data storage, processing, and analysis techniques is essential, as well as familiarity with tools for managing and extracting insights from large datasets. Proficiency in concepts like data parallelism and fault tolerance is crucial, enabling professionals to navigate the complexities of distributed systems effectively. Additionally, a theoretical grasp of real-time processing and streaming technologies further enhances the capabilities of professionals in harnessing the power of big data for informed decision-making and innovative solutions.	This course is aligned with higher NSQF/NCrF levels, indicating a need for advanced theoretical knowledge and practical application in the specialized field of Big Data.	5
Professional and Technical Skills/ Expertise/ Professional Knowledge	Participants gain technical skills in distributed computing, Hadoop, Spark, and big data storage and processing tools. Additionally, they will develop professional skills in navigating complex distributed systems, ensuring fault tolerance, and applying real-time processing techniques, enabling them to adeptly manage and extract insights from large datasets in diverse professional settings.	Individuals completing this qualification are likely to possess the expertise required for roles demanding advanced and specialized knowledge in the field of Big Data.	5
Employment Readiness & Entrepreneurship Skills & Mind-set/Professional Skill	Professional skills collectively enable individuals to contribute effectively to organizations leveraging Big Data technologies, adapt to changing industry demands, and excel in roles such as Big Data Analyst, Business Intelligence Analyst, Data Architect Data Analyst, Big Data Consultant, Big Data Developer or other specialized positions within the cloud computing domain.	Candidates will be learning effective communications which will make them smart in communicating with various companies and people.	5

Broad Learning Outcomes/Core Skill	This course provides broad learning outcomes, equipping professionals with technical proficiency in distributed computing, Hadoop, Spark, and big data storage and processing. Participants also develop a comprehensive understanding of real-time processing and streaming technologies, enabling them to contribute effectively to the strategic management and utilization of large-scale datasets for informed decision-making.	Candidate can perform well under supervision of team lead	5
Responsibility	Imparting a deep understanding of key big data technologies and principles, ensuring professionals can effectively manage and analyze large datasets. It also equips participants with the skills necessary to apply big data tools and frameworks, fostering their ability to derive valuable insights and contribute to data-driven decision-making within their respective professional domains.	Takes complete responsibility for delivery and quality of own work and output as also the subordinates. Shares responsibility for the group tasks.	5

Annexure II: Tools and Equipment (lab set-up)

List of Tools and Equipment

Batch Size: 30

S. No.	Tool/ Equipment Name	Specification	Quantity for specified Batch size
1	Classroom	1 (30 Sq.m)	30
2	Students Chair	30	30
3	Students Table	30	30
4	Desktop computer with accessories	GUI based Operating System, Hadoop, Apache Spark, Hive, Pig, Apache Kafka, Jupyter Notebooks, Data Visualization Tools	30
5	Deskjet printer	1 No.	Paper-A4

Classroom Aids for offline and blended mode of training:

The aids required to conduct sessions in the classroom are:

1. LCD Projector/Smart Board
2. Pin-up Board
3. WhiteBoard, Markers

Annexure III: Industry Validations/ Government Recognition Summary

S. No	Organization Name	Representative Name	Designation	Contact Address	Contact Phone No	E-mail ID
1	Software World	Amrita Saha	Proprietor	Ujan Abhoynagar, Manipuripara, Agartala, Tripura(West)	7005261744	support@softwareworld.co.in
2	Bada Biplab Power Solution LLP	Iduli Debbarma	Designated Partner	Agartala West Tripura, Pin: 799003	9436740983	bbpsllp@gmail.com
3	Krishna Industrial Services	Debajit Dey	Proprietor	Badharghat Chowmuhani Agartala, Pin: 799003	9862770077	jbyacademy@gmail.com
4	JB Youth Computer Solution & Educational Society	Nishi Kanta Das	Project Coordinator	Badharghat Chowmuhani, Siddi Ashram, Agartala	9436740983	jbyacademy@gmail.com

Annexure IV : Training Details

Training Projections:

Year	Estimated Training # of Total Candidates	Estimated training# of Women	Estimated training# of People with Disability
2023-24	1000	200	20
2024-25	1000	200	20
2025-26	1000	200	20

Data to be provided year-wise for next 3 years.

Annexure V: Blended Learning

Blended Learning Estimated Ratio &Recommended Tools:

Refer NCVET "Guidelines for Blended Learning for Vocational Education, Training & Skilling" available on:

S. No.	Select the Components of the NOS	List Recommended Tools – for all Selected Components	Offline: Online Ratio
1	Theory/ Lectures - Imparting theoretical and conceptual knowledge	Online interaction platforms like JitSi Meet, Bharat VC, Google Meet, MS Teams, etc.	60:40
2	Imparting Soft Skills, Life Skills and Employability Skills /Mentorship to Learners	NA	NA
3	Showing Practical Demonstrations to the learners	Online interaction platforms like JitSi Meet, Bharat VC, Google Meet, MS Teams, etc.	60:40
4	Imparting Practical Hands-on Skills/ Lab Work/ workshop/ shop floor training	PCs/Laptops	100:0
5	Tutorials/ Assignments/ Drill/ Practice	Online interaction platforms like JitSi Meet, Bharat VC, Google Meet, MS Teams, etc.	50:50
6	Proctored Monitoring/ Assessment/ Evaluation/ Examinations	NIELIT Online Examination	Online: 100% Theory
7	On the Job Training (OJT)/ Project Work Internship/ Candidate Training	NA	NA

Annexure VI : Standalone NOS- Performance Criteria details

1. Description:

This upskilling course provides a comprehensive exploration of fundamental concepts and technologies in the field, covering key tools such as Hadoop and Spark. Participants gain hands-on experience in managing and analyzing large datasets, acquiring the skills needed to harness the power of big data for informed decision-making and innovative solutions in various industries.

2. Scope:

The scope covers the following:

- Enhances career prospects by preparing individuals for roles in data engineering, analytics, and business intelligence.
- capable of effectively handling and extracting insights from massive datasets, making participants valuable assets across diverse industries.
- Extends to opportunities in data-driven decision-making, machine learning, and emerging fields that require proficiency in big data technologies.

3. Elements and Performance Criteria

Elements	Performance Criteria
Big Data Hadoop Ecosystem	PC1: Efficiently store, process, and analyze large volumes of data while ensuring data integrity, security, and governance. PC2: Scalable and distributed computing capabilities for handling diverse data formats and enabling real-time analytics
Modeling Big Data	PC1: Effectiveness in storing and managing vast amounts of raw data in a scalable and cost-effective manner PC2: Capability of supporting complex analytical queries and extracting valuable insights from diverse data formats
Processing Big Data	PC1: Processing frameworks should efficiently handle large datasets, enabling rapid data ingestion, transformation, and analysis. PC2: Processing frameworks should be scalable and flexible to accommodate diverse data formats, sources, and processing needs.

Streaming and Advanced Topics	PC1: Effectively manage and process large-scale data using Spark on Databricks, leveraging notebook integration and automation for streamlined workflows. PC2: Implement robust data pipelines and streaming applications using Apache Airflow, employing scheduling workflows for efficient task execution.

4. Knowledge and Understanding (KU):

The individual on the job needs to know and understand:

KU1. Participants gain insight into key technologies, including distributed computing, Hadoop, and Spark, fostering a comprehensive understanding of the tools essential for big data processing.

KU2. Proficiency in managing distributed systems, fault tolerance, and real-time processing is developed, addressing the multifaceted challenges inherent in working with large-scale datasets.

KU3. Knowledge in data storage, processing, and analysis techniques is emphasized, ensuring a well-rounded grasp of fundamental concepts in the big data ecosystem.

KU4. Equips professionals with the theoretical knowledge required to navigate complex big data scenarios, fostering expertise in strategic decision-making and innovation within data-rich environments.

5. Generic Skills (GS):

User/individual on the job needs to know how to:

GS1. Follow instructions, guidelines and procedures

GS2. Listen effectively and communicate information accurately

GS3. Apply formatting features to achieve the desired results

Annexure VII: Assessment Criteria

Detailed PC-wise assessment criteria and assessment marks for the NOS are as follows:

S. No.	Assessment Criteria for Performance Criteria	Theory Marks	Practical Marks	Project Marks	Viva Marks
Big Data Hadoop Ecosystem	PC1: Efficiently store, process, and analyze large volumes of data while ensuring data integrity, security, and governance. PC2: Scalable and distributed computing capabilities for handling diverse data formats and enabling real-time analytics	25	-	-	-
Modeling Big Data	PC1: Effectiveness in storing and managing vast amounts of raw data in a scalable and cost-effective manner PC2: Capability of supporting complex analytical queries and extracting valuable insights from diverse data formats	25	-	-	-
Processing Big Data	PC1: Processing frameworks should efficiently handle large datasets, enabling rapid data ingestion, transformation, and analysis. PC2: Processing frameworks should be scalable and flexible to accommodate diverse data formats, sources, and processing needs.	25	-	-	-
Streaming and Advanced Topics	PC1: Effectively manage and process large-scale data using Spark on Databricks, leveraging notebook integration and automation for streamlined workflows. PC2: Implement robust data pipelines and streaming applications using Apache Airflow, employing scheduling workflows for efficient task execution.	25	-	-	-
Total Marks		**100**	**-**	**-**	**-**

Annexure VIII: Assessment Strategy

This section includes the processes involved in identifying, gathering, and interpreting information to evaluate the Candidate on the required competencies of the program.

Assessment of the qualification evaluates candidates to ascertain that they can integrate knowledge, skills and values for carrying out relevant tasks as per the defined learning outcomes and assessment criteria.

The underlying principle of assessment is fairness and transparency. The evidence of the outcomes and assessment criteria. competence acquired by the candidate can be obtained by conducting Theory (Online) examination.

About Examination Pattern:

1. The question papers for the theory exams are set by the Examination wing (assessor) of NIELIT HQS.
2. The assessor assigns roll number.
3. The assessor carries out theory online assessments. Theory examination would be conducted online and the paper comprise of MCQ
4. Pass percentage would be 50% marks.
5. The examination will be conducted in English language only.

Quality assurance activities: A pool of questions is created by a subject matter expert and moderated by other SME. Test rules are set beforehand. Random set of questions which are according to syllabus appears which may differ from candidate to candidate. Confidentiality and impartiality are maintained during all the examination and evaluation processes.

Annexure IX : Acronym and Glossary

Acronym

Acronym	Description
AA	Assessment Agency
AB	Awarding Body
NCrF	National Credit Framework
NOS	National Occupational Standard(s)
NQR	National Qualification Register
NSQF	National Skills Qualifications Framework

Glossary

Term	Description
National Occupational Standards (NOS)	NOS define the measurable performance outcomes required from an individual engaged in a particular task. They list down what an individual performing that task should know and also do.
Qualification	A formal outcome of an assessment and validation process which is obtained when a competent body determines that an individual has achieved learning outcomes to given standards
Qualification File	A Qualification File is a template designed to capture necessary information of a Qualification from the perspective of NSQF compliance. The Qualification File will be normally submitted by the awarding body for the qualification.

Sector	A grouping of professional activities on the basis of their main economic function, product, service or technology.