



# Model Curriculum

**QP Name: Big Data - Fundamentals of Big Data**  
**OEM Name - OEM Name: SIC – Big Data**

**QP Code: ELE/N1418**

**QP Version: 1.0**

**NSQF Level: 4.5**

**Model Curriculum Version: 1.0**

Electronics Sector Skills Council of India || 155, 2nd Floor, ESC House, Okhla Industrial Area - Phase 3, New Delhi – 110020

## Table of Contents

Training Parameters.....	3
Program Overview .....	4
Training Outcomes.....	4
Compulsory Modules.....	4
Module Details.....	5
Module 1: Understanding Big Data Ecosystem and Tools.....	5
Module 2: Ingestion Frameworks and Architectures .....	6
Module 3: Big Data Analytical Tools and Techniques.....	7
Module 4: Data Modeling Concepts .....	9
Module 5: Data Security Fundamentals.....	10
Module 6: Big Data Project and Capstone Tutorial.....	11
Annexure.....	13
Trainer Requirements .....	13
Assessor Requirements.....	14
Assessment Strategy .....	14
References .....	15
Glossary.....	15
Acronyms and Abbreviations .....	15

## Training Parameters

<b>Sector</b>	Electronics
<b>Sub-Sector</b>	Semiconductor & Components
<b>Occupation</b>	Product Design S&C
<b>Country</b>	India
<b>NSQF Level</b>	4.5
<b>Aligned to NCO/ISCO/ISIC Code</b>	NCO-2015/ 2523.0200
<b>Minimum Educational Qualification and Experience</b>	UG Certificate or Equivalent No Experience required OR 12th 1.5 years relevant experience OR 3 year diploma after 10th No Experience required OR 10th Grade pass 3 years relevant experience OR Previous relevant Qualification of NSQF Level 43 years relevant experience
<b>Pre-Requisite License or Training</b>	NA
<b>Minimum Job Entry Age</b>	18 years
<b>Last Reviewed On</b>	27.08.2024
<b>Next Review Date</b>	27.08.2027
<b>NSQC Approval Date</b>	27.08.2024
<b>QP Version</b>	1.0
<b>Model Curriculum Creation Date</b>	27.08.2024
<b>Model Curriculum Valid Up to Date</b>	27.08.2027
<b>Model Curriculum Version</b>	1.0
<b>Maximum Duration of the Course</b>	240 Hours

## Program Overview

This section summarizes the end objectives of the program along with its duration.

### Training Outcomes

At the end of the program, the learner should have acquired the listed knowledge and skills.

- Project Planning and Execution
- Data Acquisition and Preparation
- Big Data Solutions Development
- AI and Data Security Implementation
- Project Evaluation and Reporting

### Compulsory Modules

The table lists the modules and their duration corresponding to the Compulsory NOS of the QP.

NOS and Module Details	Theory Duration	Practical Duration	On-the-Job Training Duration (Mandatory)	On-the-Job Training Duration (Recommended)	Total Duration
<b>ELE/N1418: Introduction to Big Data and Ecosystem</b>	<b>160:00</b>	<b>80:00</b>	<b>00:00</b>	<b>00:00</b>	<b>240:00</b>
Module 1: Understanding Big Data Ecosystem and Tools	30:00	00:00	00:00	00:00	30:00
Module 2: Ingestion Frameworks and Architectures	30:00	00:00	00:00	00:00	30:00
Module 3: Big Data Analytical Tools and Techniques	35:00	00:00	00:00	00:00	35:00
Module 4: Data Modeling Concepts	35:00	00:00	00:00	00:00	35:00
Module 5: Data Security Fundamentals	30:00	00:00	00:00	00:00	30:00
Module 6: Big Data Project and Capstone Tutorial	00:00	80:00	00:00	00:00	80:00
<b>Total Duration</b>	<b>160:00</b>	<b>80:00</b>	<b>00:00</b>	<b>00:00</b>	<b>240:00</b>

## Module Details

### Module 1: Understanding Big Data Ecosystem and Tools

#### Mapped to ELE/N1418: Fundamentals of Big Data

#### Terminal Outcomes:

- Mastery in defining and explaining Big Data concepts, emphasizing its critical role in managing large datasets.
- Expertise in describing the components of the Big Data ecosystem and major technologies like Hadoop and Spark.
- Proficiency in addressing challenges, trends, ethical considerations, and privacy issues in Big Data applications.

<b>Duration:</b> 30:00	
Theory – Key Learning Outcomes	Practical – Key Learning Outcomes
<ul style="list-style-type: none"> <li>• Understand the significance and impact of Big Data in various industries and applications.</li> <li>• Learn about the characteristics of Big Data: Volume, Velocity, Variety, and Veracity.</li> <li>• Explore the history and evolution of Big Data technologies and key milestones.</li> <li>• Gain knowledge of the Big Data ecosystem components and their roles.</li> <li>• Compare traditional data management systems with Big Data systems.</li> </ul>	<ul style="list-style-type: none"> <li>• Analyze real-world Big Data applications in healthcare, finance, and social media.</li> <li>• Identify and categorize data sources and methods of data generation.</li> <li>• Use tools like Hadoop and Spark in hands-on exercises.</li> <li>• Implement data ingestion using frameworks like Apache NiFi and Flume.</li> <li>• Develop simple data processing workflows to experience the Big Data lifecycle.</li> </ul>
<b>Classroom Aids:</b>	
Training Kit - Trainer Guide, Presentations, Whiteboard, Marker, Projector, Laptop	
<b>Tools, Equipment and Other Requirements</b>	
<ul style="list-style-type: none"> <li>• <b>Big Data Software</b></li> <li>• <b>Data Sources</b></li> </ul>	

## Module 2: Ingestion Frameworks and Architectures

### Mapped to ELE/N1418: Fundamentals of Big Data

#### Terminal Outcomes:

- Proficiency in defining and distinguishing between structured, semi-structured, and unstructured data types.
- Expertise in implementing data ingestion frameworks like Apache NiFi and Flume for diverse use cases.

**Duration:** 30:00

Theory – Key Learning Outcomes	Practical – Key Learning Outcomes
<ul style="list-style-type: none"> <li>• Define and distinguish structured, semi-structured, unstructured data with relevant examples.</li> <li>• Explain data formats like CSV, JSON, XML, their significance in Big Data.</li> <li>• Discuss data ingestion frameworks, their functionalities, and appropriate use cases.</li> <li>• Compare batch and real-time ingestion methods, highlighting their advantages and challenges.</li> <li>• Describe various data sources for ingestion, including IoT devices and social media platforms.</li> <li>• Explain the architecture and benefits of distributed file systems like HDFS and S3.</li> <li>• Discuss NoSQL databases like Cassandra, MongoDB, and their suitability for Big Data.</li> <li>• Compare on-premises and cloud storage solutions, evaluating cost and scalability.</li> <li>• Define the role and structure of data warehouses in Big Data.</li> <li>• Explore ETL processes for data warehousing, including extraction, transformation, and loading techniques.</li> </ul>	<ul style="list-style-type: none"> <li>• Identify and classify different data types and their storage requirements.</li> <li>• Implement data ingestion using frameworks like Apache NiFi and Flume.</li> <li>• Set up batch and real-time data ingestion pipelines, optimizing for efficiency.</li> <li>• Connect to various data sources and enable seamless data ingestion.</li> <li>• Implement real-time data ingestion tools like Kafka, detailing their key features.</li> <li>• Develop data pipelines ensuring efficient data flow and minimal latency.</li> <li>• Implement distributed file systems like HDFS for scalable data storage.</li> <li>• Set up and configure NoSQL databases for Big Data applications.</li> <li>• Use cloud storage solutions like AWS S3 for Big Data projects.</li> <li>• Implement ETL processes for data warehouses, enhancing data integration.</li> </ul>
<b>Classroom Aids:</b>	
Training Kit - Trainer Guide, Presentations, Whiteboard, Marker, Projector, Laptop	
<b>Tools, Equipment and Other Requirements</b>	
<ul style="list-style-type: none"> <li>• <b>Data Ingestion Tools</b></li> <li>• <b>Storage Solutions</b></li> <li>• <b>NoSQL Database</b></li> </ul>	

### Module 3: Big Data Analytical Tools and Techniques

#### Mapped to ELE/N1418: Fundamentals of Big Data

##### Terminal Outcomes:

- Mastery in defining and emphasizing the significance of Big Data Analytics for extracting insights.
- Proficiency in performing various types of analytics (descriptive, diagnostic, predictive, prescriptive) with practical use cases.

**Duration: 35:00**

Theory – Key Learning Outcomes	Practical – Key Learning Outcomes
<ul style="list-style-type: none"> <li>• Define Big Data Analytics and its significance in extracting insights from large datasets.</li> <li>• Explain types of analytics: descriptive, diagnostic, predictive, prescriptive with relevant use cases.</li> <li>• Outline the Big Data analytics lifecycle: data collection, processing, analysis, visualization.</li> <li>• Discuss common industry applications of Big Data Analytics in sectors like healthcare, finance, and marketing.</li> <li>• Introduce tools and technologies: Hadoop, Spark, R, Python for data analysis and machine learning.</li> <li>• Highlight the importance of data visualization in making complex data understandable and actionable.</li> <li>• Explore data visualization tools such as Tableau, Power BI, and D3.js and their capabilities.</li> <li>• Discuss best practices in data visualization including design principles and effective communication.</li> <li>• Address challenges and requirements of real-time data processing, emphasizing low latency and high throughput.               <ul style="list-style-type: none"> <li>• Introduce Spark Streaming and its architecture for real-time data processing.</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Use Hadoop, Spark, R, Python for data manipulation, analysis, and machine learning.</li> <li>• Implement regression, classification, clustering algorithms for data analytics.</li> <li>• Apply Spark MLlib for advanced analytics on large datasets.</li> <li>• Create interactive dashboards and reports using Tableau, Power BI, D3.js.</li> <li>• Develop Spark Streaming applications for real-time data processing.</li> <li>• Integrate Spark Streaming with MLlib for real-time analytics applications.               <ul style="list-style-type: none"> <li>• Address practical challenges in real-time data processing and analytics.</li> </ul> </li> </ul>
<p><b>Classroom Aids:</b></p>	
<p>Training Kit - Trainer Guide, Presentations, Whiteboard, Marker, Projector, Laptop</p>	
<p><b>Tools, Equipment and Other Requirements</b></p>	
<p>Hadoop, Spark, R, Python, Tableau, Power BI, D3.js.</p>	

## Module 4: Data Modeling Concepts

### Mapped to ELE/N1418: Fundamentals of Big Data

#### Terminal Outcomes:

- Master data modeling principles for various applications.
- Expertise in relational, dimensional, hierarchical data models.
- Design scalable data models for Big Data, ensuring efficiency.

Duration: 35:00	
Theory – Key Learning Outcomes	Practical – Key Learning Outcomes
<ul style="list-style-type: none"> <li>• Explain data modeling principles: conceptual, logical, physical models.</li> <li>• Differentiate relational, dimensional, hierarchical data models and their applications.</li> <li>• Implement scalable data models for Big Data, considering performance.</li> <li>• Discuss techniques optimizing data models for query performance, data retrieval efficiency.</li> <li>• Define machine learning concepts: supervised, unsupervised, reinforcement learning.</li> <li>• Outline AI lifecycle: data collection, preprocessing, model training, evaluation, deployment.</li> </ul>	<ul style="list-style-type: none"> <li>• Develop data models tailored for Big Data environments.</li> <li>• Optimize data models to enhance query performance.</li> <li>• Implement basic AI models using Python or R on sample datasets.</li> <li>• Develop deep learning models using TensorFlow, Keras for tasks like image classification.</li> <li>• Analyze AI use cases in Big Data: predictive analytics, recommendation systems, fraud detection.</li> </ul>
Classroom Aids	
Training Kit (Trainer Guide, Presentations). Whiteboard, Marker, Projector, Laptop	
Tools, Equipment and Other Requirements	
Python, R, TensorFlow, Keras	



## Module 5: Data Security Fundamentals

### Mapped to ELE/N1418: Fundamentals of Big Data

#### Terminal Outcomes:

- Expertise in implementing robust data security measures to protect sensitive information effectively.
- Proficiency in designing and managing access control mechanisms to safeguard data integrity and confidentiality.
- Comprehensive understanding of compliance requirements and implementation strategies for data privacy regulations.

<b>Duration: 30:00</b>	
<b>Theory – Key Learning Outcomes</b>	<b>Practical – Key Learning Outcomes</b>
<ul style="list-style-type: none"> <li>• Emphasize the importance of data security in preventing unauthorized access and breaches.</li> <li>• Identify and analyze common data security threats like malware, phishing, and insider threats.</li> <li>• Implement fundamental security measures such as firewalls, antivirus software, and secure authentication protocols.</li> <li>• Discuss data encryption techniques (symmetric, asymmetric) and their role in data protection.</li> <li>• Define access control and explore models like RBAC and ABAC for managing data access.</li> <li>• Implement access control policies and monitor systems to prevent unauthorized access attempts.</li> </ul>	<ul style="list-style-type: none"> <li>• Implement basic security measures (firewalls, antivirus) to protect against common threats.</li> <li>• Configure and manage access control mechanisms (RBAC, ABAC) to enforce data access policies.</li> <li>• Utilize tools like Kerberos for authentication and LDAP for directory services in security implementations.</li> <li>• Apply data masking and anonymization techniques to protect sensitive information in datasets.</li> <li>• Conduct regular audits and vulnerability assessments to evaluate and improve security measures.</li> </ul>
<b>Classroom Aids</b>	
Training kit (Trainer guide, Presentations), White board, Marker, projector, laptop, flipchart.	
<b>Tools, Equipment and Other Requirements</b>	
<ul style="list-style-type: none"> <li>• Kerberos, LDAP for authentication and directory services.</li> </ul>	

## Module 6: Big Data Project and Capstone Tutorial

### Mapped to ELE/N1418: Fundamentals of Big Data

#### Terminal Outcomes:

- Proficiency in defining and aligning Big Data project scope with business goals and stakeholder expectations.
- Expertise in implementing data acquisition, preparation, and deployment strategies for Big Data solutions.
- Ability to evaluate project success, analyze outcomes, and effectively present findings to stakeholders.

**Duration:** 80:00

Theory – Key Learning Outcomes	Practical – Key Learning Outcomes
<ul style="list-style-type: none"> <li>• Define scope, objectives, and alignment of Big Data projects with business goals and stakeholder expectations.</li> <li>• Explore Agile and Waterfall methodologies, selecting appropriate approaches for project management.</li> <li>• Develop detailed project plans with timelines, milestones, and resource allocation strategies.</li> <li>• Identify project stakeholders, establish roles, and ensure effective communication and collaboration.</li> <li>• Define data requirements, sources, and quality considerations for Big Data projects.</li> <li>• Implement data collection methods (APIs, web scraping, database queries) and preprocessing workflows (cleaning, normalization, transformation).</li> </ul>	<ul style="list-style-type: none"> <li>• Develop and test Big Data solutions using appropriate tools and technologies.</li> <li>• Implement data preprocessing workflows to ensure data quality and consistency.</li> <li>• Deploy Big Data applications using on-premises, cloud, or hybrid strategies.</li> <li>• Utilize CI/CD practices to automate deployment and ensure rapid delivery.</li> <li>• Monitor and manage deployed solutions for optimal performance, scalability, and reliability.</li> </ul>
<b>Classroom Aids</b>	
Training kit (Trainer guide, Presentations), White board, Marker, projector, laptop, flipchart.	
<b>Tools, Equipment and Other Requirements</b>	
<ul style="list-style-type: none"> <li>• Agile, Waterfall methodologies; APIs, web scraping, database queries for data collection.</li> </ul>	

## Annexure

### Trainer Requirements

Trainer Prerequisites						
Minimum Educational Qualification	Specialization	Relevant Industry Experience		Training Experience		Remarks
		Years	Specialization	Years	Specialization	
BE/B.Tech	Electrical/ Mechanical/ Electronics/Co mputer science	1	Big Data field	1	Training	
Diploma/ITI	Electrical/ Mechanical/ Electronics/Co mputer science	2	Big Data field	1	Training	
Certified in relevant CITS Trade						

Trainer Certification	
Domain Certification	Platform Certification
<p><b>“ELE/N1418: Fundamentals of Big Data, version 1.0”.</b> Minimum accepted score is 80%.</p>	<p>Recommended that the Trainer is certified for the <b>ELE/N1418: Fundamentals of Big Data</b> “Trainer (VET and Skills)”, mapped to the Qualification Pack: “MEP/Q2601, V2.0”, with minimum score of 80%</p>

## Assessor Requirements

Assessor Prerequisites						
Minimum Educational Qualification	Specialization	Relevant Industry Experience		Training/Assessment Experience		Remarks
		Years	Specialization	Years	Specialization	
BE/B.Tech	Electrical/ Mechanical/ Electronics/Co mputer science	2	Big Data field	1	Training	
Diploma/ITI	Electrical/ Mechanical/ Electronics/Co mputer science	3	Big Data field	1	Training	
Certified in relevant CITS Trade						

Assessor Certification	
Domain Certification	Platform Certification
<p><b><i>"ELE/N1418: Fundamentals of Big Data, version 1.0"</i></b>. Minimum accepted score is 80%.</p>	<p>Recommended that the Assessor is certified for the <b><i>"ELE/N1418: Fundamentals of Big Data "Assessor (VET and Skills)"</i></b>, mapped to the Qualification Pack: <b><i>"MEP/Q2701, V2.0"</i></b>, with minimum score of 80%</p>

## Assessment Strategy

1. Assessment System Overview:
  - Batches assigned to the assessment agencies for conducting the assessment on SDMS/SIP or email
  - Assessment agencies send the assessment confirmation to VTP/TC looping SSC
  - Assessment agency deploys the ToA certified Assessor for executing the assessment
  - SSC monitors the assessment process & records
2. Testing Environment:
  - Confirm that the centre is available at the same address as mentioned on SDMS or SIP
  - Check the duration of the training.
  - Check the Assessment Start and End time to be as 10 a.m. and 5 p.m.
  - If the batch size is more than 30, then there should be 2 Assessors.
  - Check that the allotted time to the candidates to complete Theory & Practical Assessment is correct.
  - Check the mode of assessment—Online (TAB/Computer) or Offline (OMR/PP).
  - Confirm the number of TABs on the ground are correct to execute the Assessment smoothly.
  - Check the availability of the Lab Equipment for the particular Job Role.
3. Assessment Quality Assurance levels / Framework:
  - Question papers created by the Subject Matter Experts (SME)
  - Question papers created by the SME verified by the other subject Matter Experts
  - Questions are mapped with NOS and PC
  - Question papers are prepared considering that level 1 to 3 are for the unskilled & semi-skilled individuals, and level 4 and above are for the skilled, supervisor & higher management
  - Assessor must be ToA certified & trainer must be ToT Certified
  - Assessment agency must follow the assessment guidelines to conduct the assessment
4. Types of evidence or evidence-gathering protocol:
  - Time-stamped & geotagged reporting of the assessor from assessment location
  - Centre photographs with signboards and scheme specific branding
  - Biometric or manual attendance sheet (stamped by TP) of the trainees during the training period
  - Time-stamped & geotagged assessment (Theory + Viva + Practical) photographs & videos
5. Method of verification or validation:
  - Surprise visit to the assessment location
  - Random audit of the batch
  - Random audit of any candidate
6. Method for assessment documentation, archiving, and access
  - Hard copies of the documents are stored
  - Soft copies of the documents & photographs of the assessment are uploaded / accessed from Cloud Storage
  - Soft copies of the documents & photographs of the assessment are stored in the Hard Drives

## References

## Glossary

<b>Sector</b>	Sector is a conglomeration of different business operations having similar business and interests. It may also be defined as a distinct subset of the economy whose components share similar characteristics and interests.
<b>Sub-sector</b>	Sub-sector is derived from a further breakdown based on the characteristics and interests of its components.
<b>Occupation</b>	Occupation is a set of job roles, which perform similar/ related set of functions in an industry.
<b>Job role</b>	Job role defines a unique set of functions that together form a unique employment opportunity in an organisation.
<b>Occupational Standards (OS)</b>	OS specify the standards of performance an individual must achieve when carrying out a function in the workplace, together with the Knowledge and Understanding (KU) they need to meet that standard consistently. Occupational Standards are applicable both in the Indian and global contexts.
<b>Performance Criteria (PC)</b>	Performance Criteria (PC) are statements that together specify the standard of performance required when carrying out a task.
<b>National Occupational Standards (NOS)</b>	NOS are occupational standards which apply uniquely in the Indian context.

<b>Qualifications Pack (QP)</b>	QP comprises the set of OS, together with the educational, training and other criteria required to perform a job role. A QP is assigned a unique qualifications pack code.
<b>Unit Code</b>	Unit code is a unique identifier for an Occupational Standard, which is denoted by an 'N'
<b>Unit Title</b>	Unit title gives a clear overall statement about what the incumbent should be able to do.
<b>Description</b>	Description gives a short summary of the unit content. This would be helpful to anyone searching on a database to verify that this is the appropriate OS they are looking for.
<b>Scope</b>	Scope is a set of statements specifying the range of variables that an individual may have to deal with in carrying out the function which have a critical impact on quality of performance required.
<b>Knowledge and Understanding (KU)</b>	Knowledge and Understanding (KU) are statements which together specify the technical, generic, professional and organisational specific knowledge that an individual needs in order to perform to the required standard.
<b>Organisational Context</b>	Organisational context includes the way the organisation is structured and how it operates, including the extent of operative knowledge managers have of their relevant areas of responsibility.
<b>Technical Knowledge</b>	Technical knowledge is the specific knowledge needed to accomplish specific designated responsibilities.
<b>Core Skills/ Generic Skills (GS)</b>	Core skills or Generic Skills (GS) are a group of skills that are the key to learning and working in today's world. These skills are typically needed in any work environment in today's world. These skills are typically needed in any work environment. In the context of the OS, these include communication related skills that are applicable to most job roles.
<b>Electives</b>	Electives are NOS/set of NOS that are identified by the sector as contributive to specialization in a job role. There may be multiple electives within a QP for each specialized job role. Trainees must select at least one elective for the successful completion of a QP with Electives.
<b>Options</b>	Options are NOS/set of NOS that are identified by the sector as additional skills. There may be multiple options within a QP. It is not mandatory to select any of the options to complete a QP with Options.

## Acronyms and Abbreviations

NOS	National Occupational Standard(s)
NSQF	National Skills Qualifications Framework
QP	Qualifications Pack
TVET	Technical and Vocational Education and Training
IPR	Intellectual Property Rights



