

National Occupational Standards



Fundamentals of Big Data (OEM Name-SIC - Big Data)

Unit Code: ELE/N1418

Version: 1.0

NSQF Level: 4.5

Electronics Sector Skills Council of India || 155, 2nd Floor, ESC House Okhla Industrial Area-Phase 3
New Delhi- 110020 || email:ceo@essc-india.org

National Occupational Standards

Description

The " SIC - Big Data" upskilling course offers an in-depth look at fundamental big data concepts and technologies, including Hadoop, Spark, and distributed computing. Participants engage in hands-on exercises and real-world case studies to master efficient data management and analysis. Ideal for professionals, this course provides practical skills to navigate and solve large-scale data challenges across various industries. The curriculum covers data storage, processing, and analysis, ensuring participants gain a well-rounded expertise. By the end, learners will be equipped to excel in roles requiring proficiency in big data technologies and insights extraction.

Scope

The scope covers the following :

- The scope covers the following:
- Identify and explain the characteristics of Big Data.
- Explore the history and evolution of Big Data technologies.
- Describe the components and tools of the Big Data ecosystem.
- Compare traditional data systems with Big Data systems.
- Identify challenges in managing Big Data and future trends.
- Discuss ethical and privacy considerations in Big Data.

Elements and Performance Criteria

Big Data Fundamentals and Ecosystem

To be competent, the user/individual on the job must be able to:

- PC1.** Collect and organize large datasets from multiple sources, including structured and unstructured data.
- PC2.** Set up and configure components in the Hadoop ecosystem (e.g., HDFS, YARN).
- PC3.** Ingest real-time data streams using tools like Apache Kafka, ensuring minimal data latency.
- PC4.** Build a distributed storage system using HDFS, managing data replication and fault tolerance.
- PC5.** Compare traditional data management systems with Big Data systems, highlighting differences in architecture and processing.

Data Ingestion, Storage, and Processing

To be competent, the user/individual on the job must be able to:

- PC6.** Implement a NoSQL database such as MongoDB, and perform CRUD operations on large datasets.
- PC7.** Design and implement a data pipeline using Apache Spark for batch processing tasks.
- PC8.** Build and deploy real-time data streaming pipelines using Spark Streaming and integrate with external systems.
- PC9.** Create a data warehousing solution using cloud technologies like Redshift or BigQuery, and manage data extraction and loading processes.
- PC10.** Manage real-time data ingestion using tools like Apache NiFi or Apache Flume, integrating IoT and social media data streams.

Big Data Security and Privacy

National Occupational Standards

To be competent, the user/individual on the job must be able to:

- PC11.** Implement security measures in Big Data systems, including encryption, firewalls, and secure access controls.
- PC12.** Set up and manage access control mechanisms using Kerberos or LDAP to ensure data integrity and prevent unauthorized access.
- PC13.** Conduct vulnerability assessments and audits on Big Data platforms to identify and mitigate security risks.
- PC14.** Implement data masking and anonymization techniques to secure sensitive information and ensure compliance with GDPR and CCPA.
- PC15.** PC 15. Monitor and optimize Big Data system performance, ensuring scalability, high availability, and compliance with privacy regulations.

Big Data Analytics and Machine Learning

To be competent, the user/individual on the job must be able to:

- PC16.** Perform exploratory data analysis on large datasets using tools like R or Python.
- PC17.** Apply machine learning algorithms such as clustering, regression, and classification to datasets using Spark MLlib.
- PC18.** Create real-time analytics pipelines using Spark Streaming, showcasing the ability to analyze data as it is ingested.
- PC19.** Implement deep learning models using TensorFlow or Keras for tasks like image recognition and natural language processing.
- PC20.** Build interactive visualizations and dashboards using tools like Tableau or Power BI to present insights from Big Data analytics.

Project Development, Deployment, and Evaluation (Capstone Project)

To be competent, the user/individual on the job must be able to:

- PC21.** Design and implement an end-to-end Big Data project, including data collection, processing, analysis, and reporting.
- PC22.** Develop and execute an ETL pipeline for moving data between relational databases and Big Data storage systems.
- PC23.** Deploy a full-scale Big Data solution on cloud platforms (AWS, Azure) and monitor its performance and scalability.
- PC24.** Conduct project evaluation by presenting insights, challenges, and lessons learned from the deployment.
- PC25.** Present the results of your Big Data project using comprehensive reports, dashboards, and interactive visualizations.

Knowledge and Understanding (KU)

The individual on the job needs to know and understand:

- KU1.** Definition and significance of Big Data, highlighting its role in handling large, complex datasets that traditional systems cannot manage efficiently.
- KU2.** Detailed understanding of the four Vs of Big Data: Volume, Velocity, Variety, and Veracity, with real-world examples.
- KU3.** Differences between structured, semi-structured, and unstructured data, and examples of each.

National Occupational Standards

- KU4.** Common data formats like CSV, JSON, and XML, and their relevance.
- KU5.** Definition and significance of Big Data Analytics for extracting insights from large datasets.
- KU6.** Types of analytics (descriptive, diagnostic, predictive, prescriptive) and their applications.
- KU7.** Data modeling principles, including conceptual, logical, and physical models.
- KU8.** Different types of data models (relational, dimensional, hierarchical) and their specific applications.
- KU9.** Different types of data (structured, semi-structured, unstructured) and common data formats (CSV, JSON, XML).
- KU10.** Data ingestion frameworks like Apache NiFi and Flume.
- KU11.** Understanding of business goals and stakeholder expectations for project alignment.
- KU12.** Familiarity with project management methodologies such as Agile and Waterfall.
- KU13.** Need for employability skills and different learning and employability related portals
- KU14.** Various constitutional and personal values

Generic Skills (GS)

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

- GS1.** Analytical skills to interpret and understand large datasets and their characteristics.
- GS2.** Research skills to explore historical advancements and current trends in Big Data technologies.
- GS3.** Analytical skills to differentiate between various data types and ingestion frameworks.
- GS4.** Technical skills to implement real-time data ingestion tools and manage data pipelines.
- GS5.** Analytical thinking to extract meaningful insights from large datasets.
- GS6.** Technical proficiency in tools like Hadoop, Spark, R, and Python.
- GS7.** Analytical thinking and problem-solving skills for data modeling and AI.
- GS8.** Technical proficiency in using tools like Python, R, TensorFlow, and Keras.
- GS9.** Analytical skills for differentiating data types and ingestion methods.
- GS10.** Technical proficiency in using data ingestion frameworks.
- GS11.** Strategic thinking and alignment with business goals.
- GS12.** Project management and planning skills.
- GS13.** Read and write different types of documents/instructions/correspondence
- GS14.** Communicate effectively using appropriate language in formal and informal settings

National Occupational Standards

Assessment Criteria

Assessment Criteria for Outcomes	Theory Marks	Practical Marks	Project Marks	Viva Marks
<i>Big Data Fundamentals and Ecosystem</i>	13	-	-	-
PC1. Collect and organize large datasets from multiple sources, including structured and unstructured data.	2	-	-	-
PC2. Set up and configure components in the Hadoop ecosystem (e.g., HDFS, YARN).	2	-	-	-
PC3. Ingest real-time data streams using tools like Apache Kafka, ensuring minimal data latency.	3	-	-	-
PC4. Build a distributed storage system using HDFS, managing data replication and fault tolerance.	3	-	-	-
PC5. Compare traditional data management systems with Big Data systems, highlighting differences in architecture and processing.	3	-	-	-
<i>Data Ingestion, Storage, and Processing</i>	13	-	-	-
PC6. Implement a NoSQL database such as MongoDB, and perform CRUD operations on large datasets.	2	-	-	-
PC7. Design and implement a data pipeline using Apache Spark for batch processing tasks.	2	-	-	-
PC8. Build and deploy real-time data streaming pipelines using Spark Streaming and integrate with external systems.	3	-	-	-
PC9. Create a data warehousing solution using cloud technologies like Redshift or BigQuery, and manage data extraction and loading processes.	3	-	-	-
PC10. Manage real-time data ingestion using tools like Apache NiFi or Apache Flume, integrating IoT and social media data streams.	3	-	-	-
<i>Big Data Security and Privacy</i>	14	-	-	-

National Occupational Standards

Assessment Criteria for Outcomes	Theory Marks	Practical Marks	Project Marks	Viva Marks
PC11. Implement security measures in Big Data systems, including encryption, firewalls, and secure access controls.	2	-	-	-
PC12. Set up and manage access control mechanisms using Kerberos or LDAP to ensure data integrity and prevent unauthorized access.	3	-	-	-
PC13. Conduct vulnerability assessments and audits on Big Data platforms to identify and mitigate security risks.	3	-	-	-
PC14. Implement data masking and anonymization techniques to secure sensitive information and ensure compliance with GDPR and CCPA.	3	-	-	-
PC15. PC 15. Monitor and optimize Big Data system performance, ensuring scalability, high availability, and compliance with privacy regulations.	3	-	-	-
<i>Big Data Analytics and Machine Learning</i>	14	-	-	-
PC16. Perform exploratory data analysis on large datasets using tools like R or Python.	2	-	-	-
PC17. Apply machine learning algorithms such as clustering, regression, and classification to datasets using Spark MLlib.	3	-	-	-
PC18. Create real-time analytics pipelines using Spark Streaming, showcasing the ability to analyze data as it is ingested.	3	-	-	-
PC19. Implement deep learning models using TensorFlow or Keras for tasks like image recognition and natural language processing.	3	-	-	-
PC20. Build interactive visualizations and dashboards using tools like Tableau or Power BI to present insights from Big Data analytics.	3	-	-	-
<i>Project Development, Deployment, and Evaluation (Capstone Project)</i>	14	-	32	-
PC21. Design and implement an end-to-end Big Data project, including data collection, processing, analysis, and reporting.	2	-	6	-

National Occupational Standards

Assessment Criteria for Outcomes	Theory Marks	Practical Marks	Project Marks	Viva Marks
PC22. Develop and execute an ETL pipeline for moving data between relational databases and Big Data storage systems.	3	-	6	-
PC23. Deploy a full-scale Big Data solution on cloud platforms (AWS, Azure) and monitor its performance and scalability.	3	-	6	-
PC24. Conduct project evaluation by presenting insights, challenges, and lessons learned from the deployment.	3	-	7	-
PC25. Present the results of your Big Data project using comprehensive reports, dashboards, and interactive visualizations.	3	-	7	-
NOS Total	68	-	32	-



National Occupational Standards

National Occupational Standards (NOS) Parameters

NOS Code	ELE/N1418
NOS Name	Fundamentals of Big Data (OEM Name- SIC - Big Data)
Sector	Electronics
Sub-Sector	
Occupation	Product Design-S&C
NSQF Level	4.5
Credits	8
Minimum Educational Qualification & Experience	UG in relevant field (Or Equivalent) with NA of experience OR 12th Class with 1.5 years of experience Relevant Experience OR Completed 3 year diploma after 10th with NA of experience OR 10th Class with 3 Years of experience Relevant Experience OR Previous relevant Qualification of NSQF Level (Level 4) with 3 Years of experience
Version	1.0
Last Reviewed Date	27/08/2024
Next Review Date	27/08/2027
NSQC Clearance Date	27/08/2024
Reference code on NQR	NG-4.5-EH-02978-2024-V1-ESSC
NQR Version	1
CCN Category	1