Government of Pakistan

National Vocational and Technical Training Commission

Prime Minister's Hunarmand Pakistan Program

"Skills for All"

Course Contents / Lesson Plan
Course Title: BIG DATA ANALYTICS

Duration: 3 Months

Revised Edition

Trainer Name	
Course Title	**BIG DATA ANALYTICS**
Objectives and Expectations	**(i) Employable skills and hands-on practice for Big Data Analytics** This is a special course designed to address unemployment in the youth. Thecourse aims to empower students with the right skillset that would help them get Big Data Analyst jobs in the industry. The course offers a broad, cross- disciplinary learning experience for students looking to pursue careers in relevant industry. In this course, students are introduced to key aspects of the design process, from research/strategy, creative brief development, and campaign development to teamwork and presentation and content creation so that they can enter the relevant market as strong candidates for beginner to intermediate level jobs. **Main Expectations:** In short, the course under reference should be delivered by professional instructors in such a robust hands-on manner that the trainees are comfortably able to employ their skills for earning money (through wage/self-employment) at its conclusion. This course thus clearly goes beyond the domain of the traditional training practices in vogue and underscores an expectation that a market-centric approach will be adopted as the main driving force while delivering it. The instructors should therefore be experienced enough to be able to identify the training needs for the possible market roles available out there. Moreover, they should also know the strengths and weaknesses of each trainee to prepare them for such market roles during/after the training. **i.** Specially designed practical tasks to be performed by the trainees have been included in the Annexure-I to this document. The record of all tasks performed individually or in groups must be preserved by the management of the training Institute clearly labelling name, trade, session, etc so that these are ready to be physically inspected/verified through monitoring visits from time to time. The weekly distribution of tasks has also been indicated in the weekly lesson plan given in this document. **ii.** To materialize the main expectations, a special module on **Job Search & Entrepreneurial Skills** has been included in the latter part of this course (5^{th} & 6^{th} month) through which, the trainees will be made aware of the Job search techniques in the local as well as international job markets (Gulf countries). Awareness around the visa process and immigration laws of the most favoured labour destination countries also form a part of this module. Moreover, the trainees would also be encouraged to venture into self-employment and exposed to the main requirements in this regard. It is also expected that a sense of civic duties/roles and responsibilities will also be inculcated in the trainees to make them responsible citizens of the country.

iii. A module on **Work Place Ethics** has also been included to highlight the importance of good and positive behaviour in the workplace in the line with the best practices elsewhere in the world. An outline of such qualities have been given in the Appendix to this document. Its importance should be conveyed in a format that is attractive and interesting for the trainees such as through PPT slides +short video documentaries. Needless to say that if the training provider puts his heart and soul into these otherwise non-technical components, the image of the Pakistani workforce would undergo a positive transformation in the local as well as international job markets.

To maintain interest and motivation of the trainees throughout the course,modern techniques such as:

- Motivational Lectures
- Success Stories
- Case Studies

These techniques would be employed as an additional training tool wherever possible (these are explained in the subsequent section on Training Methodology). Lastly, evaluation of the competencies acquired by the trainees will be done objectively at various stages of the training and a proper record of the same will be maintained. Suffice to say that for such evaluations, practical tasks would be designed by the training providers to gauge the problem-solving abilities of the trainees.

(ii) Success Stories

Another effective way of motivating the trainees is using Success Stories. Its inclusion in the weekly lesson plan at regular intervals has been recommendedtill the end of the training.

A success story may be disseminated orally, through a presentation, or using a video/documentary of someone that has risen to fortune, acclaim, or brilliant achievement. A success story shows how a person achieved his goal through hard work, dedication, and devotion. An inspiring success story contains compelling and significant facts articulated clearly and easily comprehendible words. Moreover, it is helpful if it is assumed that the reader/listener knows nothing of what is being revealed. The optimum impact is created when the story is revealed in the form of:-

- Directly in person (At least 2-3 cases must be arranged by the training institute)
- Through an audio/ videotaped message (2-3 high-quality videos must be arranged by the training institute)

It is expected that the training provider would collect relevant high-quality success stories for inclusion in the training as suggested in the weekly lesson plan given in this document. Suggestive structure and sequence of a sample success story.

Case Studies

Where a situation allows, case studies can also be presented to the trainees to widen their understanding of the real-life specific problem/situation and to explore the solutions.

In simple terms, the case study method of teaching uses a real-life case example/a typical case to demonstrate a phenomenon in action and explain theoretical as well as practical aspects of the knowledge related to the same. It is an effective way to help the trainees comprehend in depth both the

	theoretical and practical aspects of the complex phenomenon in depth with ease. Case teaching can also stimulate the trainees to participate in discussions and thereby boost their confidence. It also makes the classroom atmosphere interesting thus maintaining the trainee interest in training till the end of the course. Depending on suitability to the trade, the weekly lesson plan in this document may suggest case studies be presented to the trainees. The trainer may adopt a PowerPoint presentation or video format for such case studies whichever is deemed suitable but only those cases must be selected that are relevant andof a learning value. The Trainees should be required and supervised to carefully analyze the cases. For this purpose, they must be encouraged to inquire and collect specific information/data, actively participate in the discussions, and intended solutions to the problem/situation. Case studies can be implemented in the following ways:- i. A good quality trade-specific documentary (At least 2-3 documentaries must be arranged by the training institute) ii. Health &Safety case studies (2 cases regarding safety and industrial accidents must be arranged by the training institute) iii. Field visits (At least one visit to a trade-specific major industry/site must be arranged by the training institute)
Entry-level of trainees	For an advanced course of Big Data Analytics proposed entry level is minimum bachelors in relevant subject, so expectations from the trainees are: • Have knowledge of Programming Concepts • Have studied languages such as C, C++, Python • Have concept of Computer system
Learning Outcomes of the course	By the end of this course, students will be able to develop skills to convert bulk information into knowledge, and to assist the business managers in taking data driven decisions.
Course Execution Plan	The total duration of the course: **3 months (12 Weeks)** Class hours: **4 hours per day** Theory: **20%** Practical: **80%** Weekly hours: **20 hours per week (5 days a week)** Total contact hours: **240 hours**
Companies offering jobs in the respective trade	Every company nowadays has huge amounts of Data, and they are in need of good analyst that can help them shape their business future.

Job Opportunities	• Big Data Engineer • Big Data Architect • Business & Data Analyst
No of Students	25
Learning Place	Classroom / Lab
Instructional Resources	• https://www.w3schools.com/ • https://www.coursera.com/ • https://www.towardsdatascience..com/ • https://www.codingbat.com/ • https://www.pythonforeverybody.com/ • https://www.edx.org/course/big-data-analytics-2 • https://online-learning.harvard.edu/subject/big-data • https://www.theknowledgeacademy.com/pk/courses/big-data-and-analytics-training/#showmoreoverview50339330

MODULES

Scheduled Weeks	Module Title	Days	Hours	Learning Units	Home Assignment
Week 1	Introduction to Big Data and Big Data Analytics	Day 1	Hour 1	Course Introduction	• **Task 1** *Details may be seen at Annexure-I*
			Hour 2	Job market	
			Hour 3	Course Applications	
			Hour 4	• Institute/work ethics • Success stories	
		Day 2	Hour 1	History of Analytics	
			Hour 2		
			Hour 3	Definitions of Big Data	
			Hour 4		
		Day 3	Hour 1	Big Data Characteristics	
			Hour 2		
			Hour 3	Use Cases	
			Hour 4		
		Day 4	Hour 1	**Motivational Lecture** (*For further detail please see Annexure: II)*	
			Hour 2		
			Hour 3	10 Vs of Big Data	
			Hour 4		
		Day 5	Hour 1	10 Vs of Big Data	
			Hour 2		
			Hour 3	Why Big Data Matters	
			Hour 4		
Week 2	Types of Big Data and Data Lakes	Day 1	Hour 1	Success stories (For further detail please see Annexure:	• **Task 2**

			Hour 2	III)	*Details may be seen at Annexure-I*
			Hour 3	Types of Big Data	
			Hour 4		
		Day 2	**Hour 1**	Types of Data Lakes	
			Hour 2		
			Hour 3	Big Data Landscapes	
			Hour 4		
		Day 3	**Hour 1**	Categorization of Big Data Analytics	
			Hour 2		
			Hour 3		
			Hour 4		
		Day 4	**Hour 1**	Overview of NoSQL databases	
			Hour 2		
			Hour 3		
			Hour 4		

		Day 5	**Hour 1**	Case study/visit to a software house/data setup etc.	
			Hour 2		
			Hour 3		
			Hour 4		
Week 3	• **NoSQL databases** • **Apache Hadoop Ecosystem**	**Day 1**	**Hour 1**	Success stories	• **Task 3** *<u>Details may be seen at Annexure-I</u>*
			Hour 2	Hands on NoSQL Databases	
			Hour 3		
			Hour 4		
		Day 2	**Hour 1**	Overview of Apache Hadoop Ecosystem	
			Hour 2		
			Hour 3		
			Hour 4		
		Day 3	**Hour 1**	Hadoop 2	
			Hour 2		
			Hour 3	Hands on Hadoop 2	

			Hour 4		
		Day 4	**Hour 1**	YARN	
			Hour 2		
			Hour 3	Hands on YARN	
			Hour 4		
		Day 5	**Hour 1**	HDFS	
			Hour 2		
			Hour 3	Setting up Hadoop clusters	
			Hour 4		
Week 4	• **MapReduce: Theory and Hands-on** • **MapReduce**	**Day 1**	**Hour 1**	Success Stories of Big Data	• **Task 4** *Details may be seen at Annexure-I*
			Hour 2	MapReduce: Theory and Hands-on	
			Hour 3		
			Hour 4		
		Day 2	**Hour 1**	Hands on MapReduce	
			Hour 2		
			Hour 3		
			Hour 4		

		Day 3	**Hour 1**	Apache Spark with Apache Kafka	
			Hour 2		
			Hour 3		
			Hour 4		
		Day 4	**Hour 1**	Hands-on Practice with Apache Spark	
			Hour 2		
			Hour 3		
			Hour 4		
		Day 5	**Hour 1**	Apache Hive	
			Hour 2		
			Hour 3		
			Hour 4		
Week 5	• **Apache Spark with Apache Kafka** • **Apache Hive, Apache HBase and Apache Cassandra**	**Day 1**	**Hour 1**	Apache HBase	• **Task 5** *<u>Details may be seen at Annexure-I</u>*
			Hour 2		
			Hour 3		

			Hour 4		
		Day 2	**Hour 1**	Apache Cassandra	
			Hour 2		
			Hour 3		
			Hour 4		
		Day 3	**Hour 1**	Hands-on Activity	
			Hour 2		
			Hour 3		
			Hour 4		
		Day 4	**Hour 1**	Browse the following website and create an account on each website • Bayt.com – The Middle East Leading Job Site • Monster Gulf – The International Job Portal • Gulf Talent – Jobs in Dubai and the Middle East Find the handy 'search' option at the top of your homepage to search for the jobs that best suit your skills. • Select the job type from the first 'Job Type' drop-down menu, next,	
			Hour 2		
			Hour 3		
			Hour 4		

				select the location from the second drop- down menu. • Enter any keywords you want to use to find suitable job vacancies. • On the results page you can search for part-time jobs only, full-time jobs only, employers only, or agencies only. Tick the boxes as appropriate to your search. • Search for jobs by: • Company • Category • Location • All jobs • Agency • Industry	
		Day 5	**Hour 1**	Motivational Lecture	
			Hour 2		
			Hour 3		
			Hour 4		
Week 6	**Apache Presto and Apache Drill**	**Day 1**	**Hour 1**	Apache Presto	• **Task 6** *Details may be seen at Annexure-I*
			Hour 2		
			Hour 3		
			Hour 4		

		Day 2	**Hour 1**	Apache Drill	
			Hour 2		
			Hour 3		
			Hour 4		
		Day 3	**Hour 1**	Hands on Apache Presto and Apache Drill	
			Hour 2		
			Hour 3		
			Hour 4		
		Day 4	**Hour 1**	Hands on Apache Presto and Apache Drill	
			Hour 2		
			Hour 3		
			Hour 4		
		Day 5	**Hour 1**	Motivational Lecture	
			Hour 2		
			Hour 3		

			Hour 4		
Week 7	• **Document NoSQL with MongoDB** • **Graph NoSQL with Neo4J**	**Day 1**	**Hour 1**	NoSQL	• **Task 7** *<u>Details may be seen at Annexure-I</u>*
			Hour 2		
			Hour 3	Hands on NoSQL	
			Hour 4		
		Day 2	**Hour 1**	NoSQL with MongoDB	
			Hour 2		
			Hour 3	Hands on	
			Hour 4		
		Day 3	**Hour 1**	Graph NoSQL with Neo4J	
			Hour 2		
			Hour 3		
			Hour 4		
		Day 4	**Hour 1**	Hands on Graph NoSQL with Neo4J	
			Hour 2		

			Hour 3		
			Hour 4		
		Day 5	**Hour 1**	Case study/visit to a software house/data setup etc.	
			Hour 2		
			Hour 3		
			Hour 4		
Week 8	**Key Value Stores with Redis**	**Day 1**	**Hour 1**	Client Connection	• **Task 8** *Details may be seen at Annexure-I*
			Hour 2		
			Hour 3	Cluster Initialization	
			Hour 4		
		Day 2	**Hour 1**	Cluster Maintenance	
			Hour 2		
			Hour 3	Database Usage	
			Hour 4		
		Day 3	**Hour 1**	CURL Command	

			Hour 2	Data Manipulation	
			Hour 3		
			Hour 4		
		Day 4	**Hour 1**	Data Manipulation	
			Hour 2		
			Hour 3	Getting Started with Redis	
			Hour 4	Basic Commands of Redis	
		Day 5	**Hour 1**	Assignment on Redis	
			Hour 2		
			Hour 3		
			Hour 4		
Week 9	**Large-Scale Supervised Learning**	**Day 1**	**Hour 1**	Introduction to Supervised learning	• **Task 9** *Details may be seen at Annexure-I*
			Hour 2		
			Hour 3		
			Hour 4		
		Day 2	**Hour 1**	Generalized Linear Models and Logistic Regression	

			Hour 2		
			Hour 3		
			Hour 4		
		Day 3	**Hour 1**	Regularization	
			Hour 2		
			Hour 3		
			Hour 4		
		Day 4	**Hour 1**	Support Vector Machine (SVM) and the kernel trick	
			Hour 2		
			Hour 3		
			Hour 4		
		Day 5	**Hour 1**	Outlier Detection	
			Hour 2		
			Hour 3	Spark ML library	
			Hour 4		

Week 10	**Large-Scale Unsupervised Learning**	**Day 1**	**Hour 1**	Introduction to Unsupervised learning	• **Task 10** *Details may be seen at Annexure-I*
			Hour 2		
			Hour 3		
			Hour 4		
		Day 2	**Hour 1**	K-means / K-medoids	
			Hour 2		
			Hour 3		
			Hour 4		
		Day 3	**Hour 1**	Gaussian Mixture Models	
			Hour 2		
			Hour 3		
			Hour 4		
		Day 4	**Hour 1**	Dimensionality Reduction	
			Hour 2		
			Hour 3		

			Hour 4		
		Day 5	**Hour 1**	Spark MLlib for Unsupervised Learning	
			Hour 2		
			Hour 3		
			Hour 4		
Week 11	**Large Scale Text Mining**	**Day 1**	**Hour 1**	Latent Semantic Indexing	• **Task 11** *Details may be seen at Annexure-I*
			Hour 2		
			Hour 3		
			Hour 4		
		Day 2	**Hour 1**	Topic models	
			Hour 2		
			Hour 3		
			Hour 4		
		Day 3	**Hour 1**	Latent Dirichlet Allocation	
			Hour 2		

			Hour 3		
			Hour 4		
		Day 4	**Hour 1**	Spark ML library for NLP	
			Hour 2		
			Hour 3		
			Hour 4		
		Day 5	**Hour 1**	Projects	
			Hour 2		
			Hour 3		
			Hour 4		
Week 12	**Final Project**	**Day 1**	**Hour 1**	Final Project	• **Task 12** *Details may be seen at Annexure-I* **Final Project**
			Hour 2		
			Hour 3		
			Hour 4		
		Day 2	**Hour 1**	Final Project	

			Hour 2		
			Hour 3		
			Hour 4		
		Day 3	**Hour 1**	Final Project	
			Hour 2		
			Hour 3		
			Hour 4		
		Day 4	**Hour 1**	Final Project Presentation	
			Hour 2		
			Hour 3		
			Hour 4		
		Day 5	**Hour 1**	Final Project Presentation	
			Hour 2		
			Hour 3		
			Hour 4		

Annexure-I

Tasks for Certificate in Big Data Analytics

Task No.	Task	Description	Week
1.	**Explore Job Market**	Make presentation on Job market for Big Data profession	1
2.	**Data Ingestion**	Ingest data from various sources such as CSV files, databases, or streaming data sources into Hadoop HDFS using tools like Apache Sqoop or Apache Kafka.	2
3.	**Data Processing**	Write a MapReduce program to process the ingested data, such as performing data cleaning, filtering, aggregation, or transformation tasks. Alternatively, use Apache Spark to process the data using RDDs (Resilient Distributed Datasets) or DataFrames.	3
4.	**Data Analysis**	Use Apache Hive or Apache Pig to write SQL-like queries or data processing scripts for analyzing the data.	4
5.	**Machine Learning**	Train a machine learning model on the processed data using libraries like Apache Mahout or Apache Spark MLlib. Implement a recommendation system, classification, regression, or clustering algorithm depending on the nature of the data and the problem statement.	5
6.	**Data Visualization**	Visualize the analyzed data using tools like Apache Zeppelin or Jupyter Notebooks. Generate charts, graphs, or interactive dashboards to present the insights derived from the data analysis.	6

Task No.	Task	Description	Week
7.	**Optimization**	Optimize the performance of data processing jobs by tuning parameters such as block size, replication factor, or JVM settings. Implement partitioning, caching, or indexing strategies to improve query performance in Apache Hive or Apache Spark SQL.	**7**
8.	**Real-time Processing**	Implement real-time data processing using Apache Storm or Apache Flink to analyze streaming data as it arrives. Perform continuous computations, windowing, or event processing on the streaming data.	**8**
9.	**Data Security**	Ensure data security by implementing authentication, authorization, and encryption mechanisms in the Hadoop cluster. Configure role-based access control (RBAC) and audit logging to monitor and control access to sensitive data.	**9**
10.	**Scalability and Fault Tolerance**	Test the scalability of the Hadoop cluster by running data processing jobs with varying data volumes. Evaluate fault tolerance mechanisms such as data replication and job recovery to ensure data integrity and reliability.	**10**
11.	**Documentation and Reporting**	Document the entire data analytics workflow, including data sources, processing steps, analysis techniques, and insights obtained. Prepare reports or presentations summarizing the findings and recommendations derived from the data analysis for stakeholders.	**11**
12.	**Final project**	Final project Assessment	**12**

Annexure-II

Workplace/Institute Ethics Guide

Work ethic is a standard of conduct and values for job performance. The modern definition of what constitutes good work ethics often varies. Different businesses have different expectations. Work ethic is a belief that hard work and diligence have a moral benefit and an inherent ability, virtue, or

value to strengthen character and individual abilities. It is a set of values-centered on the importance of work and manifested by determination or desire to work hard.

The following ten work ethics are defined as essential for student success:

1. **Attendance:**
   Be at work every day possible, plan your absences don't abuse leave time. Be punctual every day.
2. **Character:**
   Honesty is the single most important factor having a direct bearing on the final success of an individual, corporation, or product. Complete assigned tasks correctly and promptly. Look to improve your skills.
3. **Team Work:**
   The ability to get along with others including those you don't necessarily like. The ability to carry your weight and help others who are struggling. Recognize when to speak up with an idea and when to compromise by blend ideas together.
4. **Appearance:**
   Dress for success set your best foot forward, personal hygiene, good manner, remember that the first impression of who you are can last a lifetime
5. **Attitude:**
   Listen to suggestions and be positive, accept responsibility. If you make a mistake, admit it. Values workplace safety rules and precautions for personal and co-worker safety. Avoids unnecessary risks. Willing to learn new processes, systems, and procedures in light of changing responsibilities.
6. **Productivity:**
   Do the work correctly, quality and timelines are prized. Get along with fellows, cooperation is the key to productivity. Help out whenever asked, do extra without being asked. Take pride in your work, do things the best you know-how. Eagerly focuses energy on accomplishing tasks, also referred to as demonstrating ownership. Takes pride in work.
7. **Organizational Skills:**
   Make an effort to improve, learn ways to better yourself. Time management; utilize time and resources to get the most out of both. Take an appropriate approach to social interactions at work. Maintains focus on work responsibilities.
8. **Communication:**
   Written communication, being able to correctly write reports and memos. Verbal communications, being able to communicate one on one or to a group.

9. **Cooperation:**

   Follow institute rules and regulations, learn and follow expectations. Get along with fellows, cooperation is the key to productivity. Able to welcome and adapt to changing work situations and the application of new or different skills.

10. **Respect:**

    Work hard, work to the best of your ability. Carry out orders, do what's asked the first time. Show respect, accept, and acknowledge an individual's talents and knowledge. Respects diversity in the workplace, including showing due respect for different perspectives, opinions, and suggestions.