

D+ Applited	
Data Analytics	* Need of Data Analytics:
Unit - 1 [One - Shot]	· Informed decision - making :
Most important topics:	
. Data analytics (definition, need and application)	Organizations seely on data analytics to make data - driven decisions.
2. Big data platforms, sources of data.	· Competative advantage:
3. Brocess and tools of analytics	Companies use data analytics to
4. Structured vs Semi-structured vs Unstructured data.	Companies use data analytics to gain insights into market tounds customer poujuences, which can give them an edge over competitors.
s. Data analytics life cycle.	· Cost reduction:
- Obrevilerning	By analyzing data, organizations can identify inefficiencies, optimize resource use, and cut costs.
Data analytics:	The second second
Definition: DA is the science of	· Risk management:
Definition: DA is the science of examining year data with the purpose of drawing conclusions about that information. It involves various techniques and process for inspecting and analysing the data.	DA helps in predicting and mitigation of visks by identifying batterns that could lead to financial, operational or strategic pit falls.
information. It involves various	financial, operational or strategic
inspecting and analysing the data.	pit falls
	SUBSCRIBE Obrevilearning

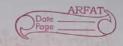
		uo o
	Date Proge	Date ARFAT Page
	Applications of DA:	* Key features of Big Data platforms: • Scalability: Can add more nodes
	Risk assessment	data volumes increased
	Fraud detection Healthcare	· Flexibility: Flexible to multiple types
01-	Personalized advertisement	Real-time Brocessing: Cambility to handle process, and analyze data in real-time fox timely insights.
	@ brevilenstning	insights.
(#)	Big data platforms: These one comprehensive frameworks that integrate various tools	Distributed Computing: Utilizes distributed computing architectures to parallelize the data processing tasks across
1	and technologies to manage, process and analyze large	multiple nodes.
	traditional data processing software cannot handle efficiently.	· Data integration: Combines data from various sources into an analytical format.
•	these platforms provides the infrastructure pecenary for storing vast amount of data	* Big Data platforms:
	managing it across distributed eystems and performing complex analysis at high speed.	Hadoop: An open-source framework that allows for the distributed processing of large

3 brongle BigQuery:

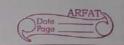
Data:

Apache Kajka, Flume, Grangle analytics, data APIs. Cleaning (data preparation):

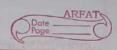
Tools: OpenRefine, Trifacta, Python (pandas library

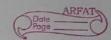


@brevilearning



DI PIII.	6 Data Visualization:
2. Data Exploration: • Exploring the data to understands its structure, patterns, and relationships.	· Greating visual representations of the analyzed data to communicate findings effectively
Tools: Python, R, Tableau, Power Bl, etc.	Tools: Tableau, Pouver Bl, R (Shiny), etc.
Applying statistical to and machine learnings models to the data to extract insights and make predictions. Tools: Apache spark MLlib, SAS, R, python, etc.	Deployment and Monitoring: Deploying the models and insights into production systems where they can be used for decision making. Monitoring involves tracking the performance of deployed models and ensuring they continue to provide accurate predictions.
Data analysis: Analyzing the data using various techniques such as duscriptive, diagnostic, predictive and pruscriptive analytics to derive meaningful insights. Jools: SQL, Python (Numby, Scily), R, Excel.	Jools: Docker, Kubernetes, Apache airflow. 8. Decision making and reporting: • Using the insights derived from the data to make informed decision This stage also involves creating detailed reports to share insights with relevant stakeholders.





Structured us Semi-structured semi-structured integrity, and integrity integrity. Structured Semi-structured Unstructured integrity, and integrity integrity. Structured Semi-structured Unstructured integrity, and integrity integrity. Structured Semi-structured Unstructured integrity, and integrity integrity. Some financial methods of the semi-series and information methods of the semi-series and with tags or predictional format information, cle. This etc. constructure. Southwest of the semi-structure of structure. South semi-structure of structure of semi-structure. South semi-structure of semi-structure of semi-structure. South semi-structure of semi-structure. South semi-structure of semi-structure of semi-structure. South semi-structure. Sout					
Simulated Semi-structured data data level for level for level to data data data store financial metadata e-rail coa several in Partially organized Den't have a customer contiguration data multi-result and with lags or predicted format information, etc. files, etc. content etc. files etc. firmages etc. files etc. f		us Semi-structured	vs Unstructured	High data Moderate data low integrity and integrity integrity	date
Dignized in Partially organized Don't have a customer configuration data multi- view and with tage or predicting from information, ctr. files, etc. content, etc. Have pe pre- Elexible schema, No pre-diffined Excel Addinal schema was tags. schema Excel SQL baud XQuery for XIII. No SQL based. SQL baud XQuery for XIII. No SQL based. SQL baud TSON Bath for JSON Data Analytics life Cycle (DALC) Easy to Comparitively Difficult to Hat quides the process of secured transforming to the secured transforming transforming secured transforming transforming transforming transforming transforming transforming transforming to the secure of the se				· Used to Used for Used to	con
SQL band XQuery for XML. NoSQL based. JSON Path for JSON. Data Analytics life Cycle (DALC) Easy to Compositively Difficult to watch and more complicated server and subscribe and that quides the process of transforming reasy data into actionable origins. SQL based XQuery for XML. NoSQL based. Data Analytics life Cycle (DALC) That quides the process of transforming reasy data into actionable origins. Solidary Highly scalable. Special Discovery: This phase involves understanding the business original problem, defining objective describing for the problem. Jest Moderately Highly flexible. Jest Pooblem, defining objective and identifying the documents.	columns.	танкия.	or sourcible.	secords, management, social customer configuration data, no information, etc. files, etc. content,	etc
Easy to Compositively Difficult to search and more complicated search and netwine data search and netweive. Limited Moderately Highly scalable. scalability. scalable. limited Moderately Highly scalable. scalability. I his phase involves understanding the business flexibility. least Moderately Highly flexible. limited scalable involves understanding the business flexibility. least Moderately Highly flexible. least Moderately Highly flexible. sources I his phase involves understanding the business flexibility. least Moderately Highly flexible. least Moderately Highly flexible. least Moderately Highly flexible. sources I have business and identifying the day. sources I have business and identifying the day.		a uses tags.	1804	Show, etc.	
· East Moderately Highly scalable. Least Moderately Highly scalable. Least Moderately Highly flexible. Least Moderately Mighly flexible Mighly flexible. Least Mighly flexible Mig	UVL Dasia.	JSON.	A Ye II	A CONTRACTOR OF THE PARTY OF TH	
limited Moderately Highly scalable. Scalability. Scalable. Pleast Moderately Highly flexible. Pleast Moderately Highly flexible. Siscovery: This phase involves understanding the business problem, defining objective flexibility. Jessible and identifying the do sources required to	search and	more complicated search and setrewal.	search and	DALC consists of several	of
flexibility flexible and identifying the sources grequired to	scalability.			i) Discovery: This phase involves	inus Livy
	flexibility.	1 ()	ougray perme.	sources desquired to	

