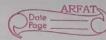


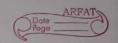
Data Analytics Unit-1 one shot notes by brevilearning YT compressed copy

B.tech (Dr. A.P.J. Abdul Kalam Technical University)



Scan to open on Studocu





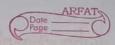
Data Analytics	Need of Data Analytics:
	Informed decision - making:
	0
, Data analytics (definition, need and application)	Organizations rely on data analytics to make data - driven decisions.
2. 000	Competative advantage:
3. Bioces and tools of analytics.	Companies use data analytics to
4. Structured vs Semi-structured vs Unstructured data.	Companies use data analytics to gain insights into market tounds, customer poulvences, which can give them an edge over competitors.
3. State	Cost reduction:
- Obravilaning	By analyzing data organizations can identify inefficiencies optimize resource use and cut costs
	and other than the same
Definition: DA is the science of	Risk management:
Definition: DA is the science of examining your 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 risks by identifying patterns that could lead to financial, operational or strategic pit falls.
information. It involves various	patterns that could lead to
inspecting and analysing	pit falls.
	SURSCRIBE Obrevilearning

This document is available on

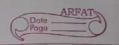


			QIO.
	Date Proge		Dole ARFAT
	Applications of DA: Customer analysis		Key features of Big Data platforms: Scalability: Can add more nodes
	Risk assessment Exaud detection		data volumes
ч	Healthcare	Lale	of aaia.
5	Personalized advertisement @brevilearning		Real - time Processing: (apphility to handle process, and analyze data in neal - time fox timely insights.
	Big data platforms: These are comprehensive frameworks	Anna .	Distributed Computing: Utilizes distributed computing architectures to parallelize the data
	that integrate various tools and technologies to manage, process and analyse large	hear	processing tasks across multiple nodes.
	process and gnalyze large volumes of data that traditional data processing software cannot handle efficiently.	3/1	Data integration: Combines data from various sources into an analytical format.
	These platforms provides the infrastructure pecenary for storing wast amount of data	*	
	managing it across distributed yestern and performing complex analysis at high speed	Asset A	Hadoop: An open-source framework that allows for the distributed processing of large
	Downloaded by Anuj Singh (anujsir	ngh73.anuj	@gmail.com)

	Date Page	Obrevileasing YT
	Spark: An opensource unified analytics engine for large-scale data processing known for its	Brocess and took of DA: The data analytics process involves several stages, each critical for transforming raw data into actionable insights. Data Collection: Crathering raw data from various
3.	Grosgle BigQuery: A serverless, highly scalable, and cost - effective multi-cloud data	Crathering raw data from various sources, such as databases, sensors, social media, and transactional records. Tools: Apache Kajka, Flume Grangle analytics, data APIs. Data Cleaning (data preparation): Cleaning the data to remove everous, handle missing values, and ensure consistency.
*	Social media Social media To T devices Scientific research Streaming data Streaming data Downloaded by Anuj Singh (anujsingh73.anuj	This step also involves transforming data into a suitable format for analysis. docu: OpenRefine, Trifacta, Python (pandas library.



@brevilearning



	Market Ma	ODIEWIE LAND
3.	1 AMIL	Data Visualization: Greating visual propregentations of the analyzed data to communicate findings effectively
	Tools: Python, R., Tableau, Power BI, etc.	Tools: Tableau, Pouver Bl, R (Shiny), etc.
- 1	Into modelling: Applying statistical to and machine earnings models to the data o extract insights and	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.
	ata analysis:	Tools: Docker, Kubernetis, Apache airflow.
· A	Tools: SQL, Python (Numby Scila) R. Excel	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 relivant stakeholders.
42 14	Downloaded by Anuj Singh (anujsingh73.anu	

			Date ARFAT			(Date Proge
	Structured data:	us Semi-structured	ve Unstructured	•	High data integrity and consistency	Moderate data integrity	- low data integrity
	Structured	Semi-structured	Unstructured		Used to store financial	- Used for metadata +	- Used to store e-mail content,
	Organized in rous and columns.	- Britially organized with tags or markers.	- Don't have a predictioned format		customer information, etc.	management, configuration files, etc.	social media data, multimedia content, etc.
•	Have pe pre- -defined schema.	Legg II	ANTINAL COLT .	•	e.g. SQL Dotabases, Excel Sheets, etc.	JSON files, etc.	e.g. text, images, videos, etc.
•	SQL band	X Dury for XML, JSON Path for JSON.	No SQL based.	(#)	Data Analyt		@bzevileazning Cycle (DALC):
•	Easy to search and sectreive data	more complicated search and	Difficult to search and retrieve		DALC construction of the pulse	0	sexural phases
	limited scalability.	Moderately scalable Moderately	Highly scalable.		Discovery: This	exitanding the	învolves e busines g objectives,
	east flexibility.	flexible	This document is available on Downloaded by Anuj Singh (an	Sstu	and some	dentifying ecus equir	the data peoblem.

