**DATA ANALYTICS PRE-ASSESSMENT**

1. Main reason for using handheld scanner:

--

1. A company that makes exercise equipment has been giving out special codes that customers use to get a discount on the product price. What can these codes tell the company?

--

3. Administrative/Institutional data refers to information collected for non-statistical reasons by governments and other organizations to give an overview of registration, transaction and record-keeping.

4.Each organization has their own inventory management system. Why should they consolidate inventory management operations in one database?

--

5.The terms: volume, variety and velocity apply to which concept?

--

6. Big data, social media data, accounting data, sales data.

7. Data architecture is **a framework for how IT infrastructure supports your data strategy**.

8.Automatic log out users after period of inactivity: to keep data secure.

9.Which of the following are intended to protect and secure credit card transactions?

--

10.People need to analyses for multiple set of data across many servers. Which tool?

--

11.Most important reason for keeping critical company data-inhouse, instead of using a cloud based system?

--

12. Migration to cloud base system concern:

13.Best database to track customers and appointments:

14.What kind of data should be considered to provide a report identifying the best geographical location for a new plant:

15.What kind of data to decide of the management should reduce the store hours:

--

16. What kind of data determine the location and facilities for senior citizens?

--

17.How to check if the company’s product quality is declining and affecting its profitability?

--

18.Data management system types:

19.An agent upgrades the customer’s cable box and also additionally offers a free router. This is an example of:

--

20. Why would a company implement customer Relationship management (CRM) system?

--

21. Which data transformation tools I the best to convert image and text files to into easily searchable format?

--

22.Which aspect of R project will help to know the condition of water?

--

23.Which aspect of R project for ocean temperature?

--

24.Which part of Hadoop framework is responsible for managing the system resources?

--

25.To collaborate biology and computer program analysis. Tool:

26. Openrefine:

GoogleBigQuery:

The R project:

TableuPublic:

27.Why the patients are not using patient portal service? How to find out:

--Through social media surverys.

28.Key performance indicator:

29. Upcoming quarterly performance review seminar should include;-:

--

30.Data Security concepts

**UNIT 1: SUMMARY**

--A data analyst gathers, manages, and carries out statistical data analysis. They must take the data and use it to aid businesses in making better decisions. This includes finding out how to price new materials for the market, how to reduce the shipping cost, solve problems that cost money to the company, and so on.

--Organizations also use data in a variety of ways to set benchmarks and targets, Key Performance Indicators (KPIs) that measure performance, study trends over time, and formulate predictive analyses.

Types of data:

* Centralized data is data that is housed, stored and managed from a single location. One of the main tenets of centralized data is integrity reliability. Storing data in a single location helps to ensure accuracy and consistency. Additionally, this will lower data redundancy by holding one main record of the data.
* Administrative data refers to information collected for non-statistical reasons by governments and other organizations to give an overview of registration, transaction and record-keeping.
* Open data refers to the type of data that can be accessed, updated, reused and exchanged by anyone and everyone.
* Public data refers to information that can be freely accessed, reused and redistributed by anyone who has no existing local, national or international access or use restrictions.
* Accounting data gives company information on the operations and transactions of the company related to revenues, payrolls, expenses, costs and assets.

--Structured data refers to data that is coded in a manner that makes it easily converted into a form usable for data analysis. Whereas, Unstructured data refers to data that is more complex and possibly stored in a format that is not easily decoded.

--Marketing data can be valuable in understanding the effectiveness of your own company's advertisement campaigns as well as those of your competitors. It can be obtained by conducting market surveys, trade shows, as well as utilizing existing research.

--The General Data Protection Regulation (GDPR) (EU) 2016/679 is a regulation in the European Union (EU) on data protection and privacy for all individuals within the EU and the European Economic Area (EEA).

--One of the largest areas in which business data and data analysis can be used concurrently to increase profits is in the area of Search Engine Optimization (SEO).

--Data life cycle management (DLM) is a policy-based approach for regulating the flow of an information system's data throughout its life cycle: from creation and initial storage to being obsolete and deleted. It provides a good infrastructure for data protection which helps in the safety of an organization in the event of danger or emergency.

--Data analysis is a process of data collection, transformation, cleaning and modeling with the aim of discovering the information required.

**UNIT 2: SUMMARY**

--Big data is a term used to describe the explosive growth in various types of data in our everyday lives. Examples: visual images, video, audio, biological data, health records, financial transactions etc. Big data is often defined in terms of three V's: volume, velocity, and variety.

--Data architecture is the sum of the models, rules, and processes for managing data within a problem domain. To build data architecture, a business must understand its data requirements and IT assets.

--Cloud computing, also known as "on-demand computing," uses converged infrastructure and shared information services to benefit the customer. Cloud providers recognize that privacy is a critical customer concern. Cryptographic encryption, authentication and integrity protection tools provide assurance that data is only sent to the intended destination and it is not modified inappropriately.

Tools used in data analytics:

* One of the tools used in analyzing big data is Hadoop. Hadoop provides for distributed processing with clusters of computers. They describe Hadoop as "a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models."
* R Project, or R, is also very popular with data analytics. R is a programming language useful in statistical computing and graphics, supported by a non-profit organization, the R Foundation for Statistical Computing. "R", as it is known for short, is used by a wide variety of workers including data scientists, statisticians, quantitative traders at investment funds, academic researchers, and other data-oriented jobs for the purposes of data analysis and visualization.

--The advantages of in-house data may discourage some companies from migrating to the cloud. While cloud services provide a great storage solution, there is still likely a need to hold some of the most important or urgent data onsite. Managing an in-house data center can get very expensive because of all the costs that come with purchasing and maintaining servers, racks, and systems.

--Shifting to the cloud has brought benefits and challenges. Cloud-based architectures reduced the required amount of start-up capital for software deployment, with corresponding reductions in ongoing maintenance costs. Methods used with structured in-house data such as clustered sessions and shared file systems in vertical architectures do not scale well in cloud-based systems.

--An organizational strategy is a dynamic long-term plan that sums up all of the actions you intend to take in order to achieve your business goals. It helps you in determining where to spend your money, human capital and time.

--Data modeling is the process that is used to create a data model. It helps in the visual data presentation and implements business rules, regulatory compliances, and government policies on the data.

**UNIT 3: SUMMARY**

--Customer relationship management (CRM) is essentially a business strategy aimed at managing customers to optimize long-term value. It relates to the process of acquiring, retaining and enhancing relationships with selected customers to build long-term value for the company and its customers.

--Social media: User-generated data from social media has also provided fresh opportunities to study consumer behavior. In social media campaigns, posts or videos put out by an organization are usually categorized by tags or subjects, thus providing additional insights about user preferences.

--Supply chain management (SCM) refers to the designing, planning, executing and monitoring of products, information, and finances as they flow from the initial supplier of raw materials all the way to final consumer.

--Inventory management refers to the supervision of the amount of each item in the inventory list. The goal of inventory management is to provide optimum levels of production and service in a timely manner at a minimal cost.

**UNIT 4: SUMMARY**

DATA ANALYTICS TOOLS:

* **Microsoft Excel:** Microsoft Excel is the most commonly used tool in data analytics. It's capable of cleaning, organizing, manipulating and visualizing data.
* **Structured Query Language (SQL):** SQL is a special-purpose programming language used to interact and maintain a database, especially in a relational database management system (RDBMS) or relational data stream management system (RDSMS) framework.
* **SAS:** SAS is a statistical software suite developed by SAS Institute for data management, advanced analytics, multivariate analysis, business intelligence, criminal investigations, and predictive analytics.
* **Tableau:** Tableau is a strong and fast-growing tool for data visualization used in the business intelligence sector. With null coding knowledge and technical skills, this effective tool lets you transform uncooked data into an easily logical format.
* **Google Analytics:** Google Analytics is a free analytics tool provided by Google, which tracks Website traffic and records it. It helps you measure the performance of your Website.
* **Google Tag Manager:** Google Tag Manager is a user-friendly, web-based interface tool that helps you manage tags on your site. It assists you in adding, editing and disabling tags without having to touch your site code.
* **Google Ads:** Google Ads (formerly Google AdWords, before July 24, 2018) is Google's online advertising platform which enables you to create online ads to reach audiences that are interested in the products and services you offer. It works on ads with pay-per-click (PPC).

--BigQuery is a Google Developers tool that is a serverless, highly scalable, and cost-effective cloud data warehouse designed to help in making informed decisions quickly, so you can transform your business with ease.When you export data to BigQuery, you own that data, and you can use BigQuery ACLs to handle projects and dataset permissions.

--OpenRefine is a tool that can take disorganized data and transform it from one format to another. In addition, OpenRefine also has the ability to extend that data with Web services and external data.

--Hadoop: To process huge amount of data without a supercomputer requires that the task be split into bite-size pieces that most computers could handle. This is where the environment of Apache's Hadoop really shines. Apache Hadoop, or simply Hadoop, makes it possible for companies to manipulate data to allow the dataset to be processed faster and more efficiently than it would in a more conventional format.

**UNIT 5: SUMMARY**

--Network traffic is defined as anything where there is commerce or communication between a sender and a receiver. For example, a business that is marketing a product is a sender, and the receiver is whom they are trying to persuade into buying that product.

-- A Web log is a list of what users are looking at on a Website and how long they look at it. A popular tool that many companies use to analyze Web log data is Google Analytics. With the use of Google Analytics, companies have been able to know their customers better, know how to communicate with their customers better, and know how to create products for their customers.

--Data integration is the method of integrating data from various source systems for creating unified sets of information for both operational and analytical use. Data integration provides a complete picture of key performance indicators (KPIs), customers, production and supply chain activities, regulatory compliance initiatives, financial risks and other facets of business processes to data analysts, corporate executives and business managers. As a result, they have better analytical information available for uses such as tracking business performance, managing operations, and planning advertising and marketing campaigns.

--ETL stands for Extract-Transform-Load and it is a process of how data is loaded from the source system to the data warehouse. The purpose of ETL testing is to ensure that the data which is loaded from a source to the destination after business transformation is correct.

--Statistical computing can be defined as the interaction between computer science, numerical analysis, and statistics.

--[Tenable Network Security](https://www.tenable.com/sc-dashboards/iavm-executive-summary-dashboard) is a powerful company that specializes in continuous network management and vulnerability management

--Another example of how to use Excel is to show what is called a Gantt chart. A Gantt chart is calendar-based type of spreadsheet that is used to show a short- or long-term analysis of data. These charts are used in project management at many organizations. Creating a Gantt chart is a great way to plan out and visualize the time required to complete a project. Gantt charts also help keep all personnel involved on task and aware of what is expected of them and when.

--Prezi.com is quickly becoming a popular tool for making presentations in the academic and business environment. Numerous students are using it in classrooms for presentations and many businesses are applying it to meetings.

--