

Final Exam Review

Matt Warner

1 Chapter 17

1.1 Data vs information vs knowledge

- **Data:** Raw facts and figures
- **Information:** Data presented in a context so that it can answer a question or support decision-making.
- **Knowledge** insight derived from experience and expertise (based on data and information).

1.2 4 V's of Big data

- **Volumn:** The extremely large amount of data that needs to be stored.
- **Velocity:** The speed and continuous nature of data acquisition.
- **Variety:** The many different forms of data being acquired and stored.
- **veracity:** The uncertainty of the quality and accuracy of data.

1.3 Understanding OLTP Vs OLAP

- **Online Transactional processing:**
 - Designed for managing and processing day-to-day operational transactions in real-time.
 - Support frequent and concurrent insert, update and delete operations, ensuring data consistency and integrity.
 - Example: Recording customer orders, updating inventory levels, processing online payments.
- **Online Analytical Processing**
 - Designed for complex querying and reporting
 - involve aggregations, calculations, and comparisons of large datasets
 - Example: Generating sales reports, analyzing market trends, performing financial forecasting

2 Chapter 18

2.1 Steps in creating LLM

1. Data collection: LLMS are trained on massive amounts of text data (books, websites, articles). This data helps the model learn about language, grammar, and facts.
2. Training:
 - A neural network is used to process the text.
 - The model predicts what comes next in a sentence, improving with each guess.
 - This process happens over many rounds to reduce errors.
3. Fine-tuning
 - The model is adjusted using specific tasks or topics (eg. conversation, coding).
 - This makes the model more accurate for particular use cases.

2.2 Supervised vs self-learning

- **supervised learning:** A type of machine learning where algorithms are trained by providing explicit examples of results sought, like recognizing products that are defective versus error-free, or stock prices.
- **Self-supervised learning:** machine learning where data is not explicitly labeled and doesn't have a predetermined result.

2.3 Risk and benefits of AI

3 Chapter 19

3.1 Zero vs first vs second vs third party data

- **zero-party data:** data that the customer explicitly shares with a firm
example: data gathered when a customer signs up for an account.
- **first-party data:** Collected by a firm through interaction, rather than data that is explicitly provided by the consumer.
 - profiling data gleaned by a news site as a customer browses articles.
- **second-party data:** data that is collected by one firm and shared with a partner organization.
 - example: if you use a branded credit card, like the Chase United Airlines Visa.
- **third-part data:** Collected by a company that has not had the customer explicitly entered a relationship with.

3.2 Cookies

A line of uniquely identifying text stored by a web browser. A cookie can only be retrieved by the web server that assigned the cookie.

How they work:

- first visit to a website will give you a cookie
- As you browse, the cookie can be matched to surfing patterns.;
- Cookie is stored on your computer or mobile device, but the website operator will create a database that matches that cookie to any other activity they choose to track.

Once a firm can track you with cookies, it can use its cookies in multiple ways:

- Advertising
- Storing persistent shopping carts that hold value even if you haven't checked out
- Holding user IDS

While most browsers have the option to completely prevent all cookies, users might be reluctant to opt out of cookie-requiring benefits like automatic log-in and persistent shopping carts.

3.3 Geotargeting

Identifying a user's physical location (sometimes called a geolocation) for the purpose of delivering tailored ads or other content.

3.4 GDPR

General Data Protection Regulation.

4 Ethics in data

4.1 5 areas we discussed in class

- Ownership: Individuals have ownership over their personal data, meaning it can only be collected or used with their explicit consent.
- Transparency: organizations must clearly communicate how they collect, store, and use data. Users should understand what they are consenting to: organizations must clearly communicate how they collect, store, and use data. Users should understand what they are consenting to.
- Privacy: Even with consent, the privacy of an individual must be protected, meaning that personally identifiable information (PII) should not be publically available or vulnerable to unauthorized accessing to.
- Intention: Data should only be collected with a clear, positive purpose, and it should be directly relevant to the task at hand. Data collection with malicious or unnecessary intent is unethical.
- Outcomes: even if data is collected with good intentions, the result of its analysis must be scrutinized to prevent unintentional harm, especially if they lead to biased or discriminatory outcomes.

5 Cloud

5.1 IaaS Vs PaaS vs SaaS - What is the difference and why would you use one over the other

- Infrastructure as a Service: Provides virtualized computing resources over the internet, such as servers, storage, and networking. It offers the foundational infrastructure, and users are responsible for managing the operating systems, applications, and data.
- Platform as a service: Provides a platform that allows developers to build, deploy and manage applications without worrying about the underlying infrastructure. The cloud provider manages the infrastructure, including servers, storage, and networking, while users focus on application development.
 - Provides development tools, middleware, database management, and OS.
 - Users focus on writing and managing the code and database.
 - The provider handles the hardware, software updates, and patching.
 - examples: snowflake, databricks.
- Software as a service: delivers fully managed software applications over the internet, users access the software through a web browser, while the cloud provider handles everything from infrastructure to application management and updates.
 - End users interact with the software, with no need to manage infrastructure or platforms the provider is responsible to maintaining the application, ensuring security and handling updates
 - Access to the software is usually subscription-based.
 - examples: google workspace, microsoft 365.

5.2 Advantages of the cloud

- Cost – pay as you go model, no need for physical storage, reduced energy costs, Reduced in house staff, built in redundancy
- Speed to market – Self Service provisioning, global reach, managed services (including databases)
- Scale – easily scale resources up or down based on demand
- Productivity –
- Performance

-
- Security
 - Reliability
 - Data Sharing

5.3 Disadvantages

6 Security

6.1 MFA

6.2 Deepfake

6.3 Phishing

6.4 Dumpster Diving

6.5 Ransomware

6.6 Pass keys

6.7 Areas of weakness in org

7 project management

7.1 Waterfall vs scrum

7.1.1 Characteristics

7.1.2 Main roles

7.1.3 Process

a b b a e b b a a b b b c b a a a b a