What does XML stand for?

Extensible Markup Language

What is XML?

* Is a file format and markup language that allows users to store, transmit, and reconstruct data
* XML is used to exchange data between databases, user desktops, and other computer systems

What could XML be used for?

* Data Exchange
  + XML is commonly used to transfer data between systems, such as between a client and a server or between different applications, as it is platform independent.
* Configuration Files
  + Many software applications use XML to store configuration settings, allowing users or systems to customize application behavior.

What are some of the key features of XML?

* Self-descriptive Structure
  + XML uses human-readable text with custom tags that describe the data, making it easier to understand the content and structure.
* Hierarchical Data Representation
  + XML organizes data in a tree-like structure with nested elements, which allows for complex, hierarchical relationships between data.
* Extensibility
  + XML allows users to define their own tags and structure, offering flexibility for different applications and data formats.

What are the benefits and drawbacks of each method of data formatting?

|  |  |  |
| --- | --- | --- |
| Format | Benefits | Drawbacks |
| ASCII | Simple, human-readable | Not suitable for non-English text, no structure, limited number of characters |
| CSV | Easy to use, widely supported, compact | Error prone if commas are in the data field, no data types |
| Fixed width | Simple parsing, no delimiters required | Padding is wasteful, hard to scale, not very flexible |
| XML | Self-descriptive, supports hierarchical data | Large file sizes, resource-intensive parsing |
| JSON | Lightweight, easy to parse, human-readable | No built-in comments |

What is a file-based structure?

A file-based structure is just a way to save and organize data in regular files on a computer without using a database system

What is a directory-based structure?

Directory-based structures are a way to organize files and folders on a device. In this system, data is grouped into directories which can contain both files and subdirectories, creating a hierarchical structure

1. What is meant by the Boolean data type?

* Boolean is a value that is either True or False (could also be written as 1 or 0, Yes or No)

1. Which data type should be used for numbers that are to be used in calculations?

* Floats (also referred to as real numbers) or Integers

1. How many characters are there in ASCII?

* There are 128 characters in ASCII

1. What is the difference between CSV and the fixed-width file formats?

* CSV has delimiters, fixed-width does not
* Fixed-width uses padding, CSV does not

1. Identify one benefit of using an XML file format

* XML supports hierarchical

1. Identify two issues that can be created by a file-based structure?

* Data Duplication – The same data can get copied into multiple files, wasting space
* Hard to Manage - Files can get messy, and finding or updating data can be lengthy

Identify 4 types of data and an example of each one:

* String – “Hello world!”
* Integer – 42
* Float/Real number – 3.14159265
* Boolean – True

What does ASCII stand for?

* American Standard Code for Information Interchange

Describe how data may be transformed into information:

* Data may be transformed into information when it is given context

Describe 2 ways in which data may be generated by the new online shopping system:

* Store transactional data when each sale is complete
* Machine learning making recommendations based on customer browsing/purchase history