Mid Term Data Pre Processing

1. What is XLS? Why we use it?

Ans.Files with XLS extension represent Excel Binary File Format. Such files can be created by Microsoft Excel as well as other similar spreadsheet programs such as OpenOffice Calc or Apple Numbers. File saved by Excel is known as Workbook where each workbook can have one or more worksheets. Data is stored and displayed to users in table format in worksheet and can span numeric values, text data, formulas, external data connections, images, and charts.

XLS was created by Microsoft for use with Microsoft Excel and is also known as Binary Interchange File Format (BIFF). This file type was introduced for the first time by making it part of Excel for Windows in 1987. XLS file format specifications were made public for the first in June 2008

XLS files are optimal for storing information that necessitates sorting in the form of rows or columns. They are capable of handling data, such as text and images. XLS files are most commonly used in a business setting.

1. What is XML?

Ans.Extensible Markup Language (XML) is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.

The design goals of XML emphasize simplicity, generality, and usability across the Internet. It is a textual data format with strong support via Unicode for different human languages. Although the design of XML focuses on documents, the language is widely used for the representation of arbitrary data structures such as those used in web services. Several schema systems exist to aid in the definition of XML-based languages, while programmers have developed many application programming interfaces (APIs) to aid the processing of XML data.

The material in this section is based on the XML Specification. This is not an exhaustive list of all the constructs that appear in XML; it provides an introduction to the key constructs most often encountered in day-to-day use.

3.Compare JSON and XML. Give 5 points.?

Ans**. JSON** (JavaScript Object Notation) is a lightweight data-interchange format and it completely language independent. It is based on the JavaScript programming language and easy to understand and generate.

**XML** (Extensible markup language) was designed to carry data, not to display data. It is a W3C recommendation. Extensible Markup Language (XML) is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable. The design goals of XML focus on simplicity, generality, and usability across the Internet. It is a textual data format with strong support via Unicode for different human languages. Although the design of XML focuses on documents, the language is widely used for the representation of arbitrary data structures such as those used in web services.

|  |  |
| --- | --- |
| **JSON** | **XML** |
| It is JavaScript Object Notation | It is Extensible markup language |
| It is based on JavaScript language. | It is derived from SGML. |
| It is a way of representing objects. | It is a markup language and uses tag structure to represent data items. |
| It does not provide any support for namespaces. | It supports namespaces. |
| It supports array. | It doesn’t supports array. |
| Its files are very easy to read as compared to XML. | Its documents are comparatively difficult to read and interpret. |

4.Give two examples of each JSON and XML.

Ans. JSON Example

var myObj = {name: "John", age: 31, city: "New York"};

var myJSON = JSON.stringify(myObj);

window.location = "demo\_json.php?x=" + myJSON;

XML Example

<?xml version="1.0" encoding="ISO-8859-1"?>

<note>

<to>Tove</to>

<from>Jani</from>

<heading>Reminder</heading>

<body>Don't forget me this weekend!</body>

</note>

5.Is XML outdated? If yes, then what else industry prefers and why?

Ans. XML is still widely used by companies whose systems are older and it would cost a fortune to update. XML is still widely used by companies whose systems are older and it would cost a fortune to update. If the idea was to change the format at a low level such as output from the database, then yes, it would be expensive

XML has many features that have made it the de facto standard for enterprise class Business to Business Integrations, Application to Application Integrations, and business interoperability requirements. These include:

* Maturity
* Robust tooling availability
* Thriving community
* Support of W3C
* Large variety of development and editing tools

For these reasons and others, XML has been embraced in many industries including Agriculture, Human Resources, Health Care, Chemical, Auto Insurance, and Vehicle Collision Repair.