CSC 223 - Advanced Scientific Programming

Data Exchange Formats

Data Exchange

- *Data exchange* is the process of transforming structured data from one format to another to facilitate data sharing between programs
- A *data exchange language* is a language that is capable of expressing general purpose data
- We will look at two data exchange languages that are commonly used to share data sets:
  - Comma Separated Values (CSV)
  - JavaScript Object Notation (JSON)

Comma-separated Values (CSV)

- A CSV file is used to represent tabular data
- The proposed specification:
  1. is plain text using a character set such as ASCII, Unicode, etc.
  2. consists of records (typically one record per line,
  3. with the records divided into fields separated by delimiters (typically a single reserved character, such as a comma),
  4. where every record has the same sequence of fields.
- Within these general constraints, many variations are used, that is, there is no standard CSV format.

JavaScript Object Notation (JSON)

- JSON is designed to be a lightweight data exchange language
- JSON is data is plain text
- The file extension for JSON files is `.json`
- The MIME type for JSON text is `application/json`

JSON Syntax

- JSON syntax is similar to the syntax of defining literal objects in JavaScript:
  - Data is in name/value pairs of the form `"name":`*value*
  - Data is separated by commas
  - Curly braces hold objects
  - Square braces hold arrays

JSON Data Types

- **Number:** a signed decimal number
- **String:** a sequence of zero or more unicode characters delimited by by double quotes
- **Boolean:** a value of `true` or `false`
- **Array:** an ordered list of zero or more values separated by commas and delimited by square brackets
- **Object:** an unordered collection of name/value pairs where pairs are separated by commas and delimited by curly braces
- **Null:** the empty value indicated by the word `null`