1. Discuss some of the problems with JSON, XML, and CSV file types. (2 pt)

JSON file types, while human-readable and has good support for Unicode character strings, do not differentiate between integer and float data types, a limitation of its JavaScript provenance. It also doesn’t natively support binary strings, which requires a workaround solution, and has complicated schema support.

XML file types are more complicated to a person to read, but like CSV, cannot differentiate between strings that contain digits and proper numbers of any format. Like the JSON file type, XML does not support binary strings, but the workaround increases the file size.

CSV files are very easily read, but can be ambiguous when examining strings that have a comma (or other delimiter) in a string. It also has no set schema, and changing the data shape requires handling it manually.

1. What is Thrift and Protocol Buffer, and how is it different than Avro? (4 pts)

Thrift, Protocol Buffer, and Avro are all binary encoding libraries from Facebook, Google, and Apache respectively, though only the first two are open-source. By employing binary encoding, the size of a data file can be drastically reduced, though there are precautions that should be taken in order to ensure proper encoding and decoding.

The main difference between Thrift/Protocol Buffer and Avro come in their treatments of schema evolution. With Thrift and Protocol Buffer, schemas are explicitly stated using field tags in each record and for each field, but with Avro, these do not exist. Instead, Avro uses a writer’s schema for each set of data it writes. While the same exact schema may not be necessary in order to decode the data, it does mean that a compatible schema must be used. This means that through iterations of schema evolution, some data may be lost in the decoding process in the case that a reader’s schema contains fields that the writer’s schema did not originally include.

While Thrift and Protocol Buffer support schema evolution using explicit field tags in their records, Avro may only evolve with default values in the fields that may be added or taken off of a schema (in the same way that new fields with Thrift and Protocol Buffer must be optional).

1. In regards to RPC, how is a network request different than a local function call (2 pts)

A local function call employs only on-machine information and is not dependent on an external network to receive a success or failure. A network request depends on a network connection, external machine availability, or bandwidth issues to receive a response. When a local function call fails, there is usually a diagnosable issue that the user can tweak in order to alter the outcome. With a network request, there are often errors that a local user cannot address.

1. What are message brokers? Give some examples. (2 pts)

Message brokers are one-way, asynchronous agents that either deliver or receive messages. This allows for a machine to send or receive several messages and have them end up in a queue, allow the machine to complete the task. The intermediary between the sender and consumer is itself the message broker. A directly available connection at the time of sending is not required for the machine to complete its task as a message broker, so long as the message is sent or received (depending on the configuration).

An example of a message broker might be AWS’ Simple Queue Service, which receives an input (configured by the user) and acts as a message broker to send a message to a subscriber at a specified endpoint. Another example of a popular message broker that acts in a similar fashion is Apache’s ActiveMQ or AWS’ similar service, Amazon MQ.