

## ASSIGNMENT 6

Lattanze

File Structure

Due: See Calendar

**Introduction:**

The purpose of this assignment is to gain some experience with designing a file structure per our discussion in class on storage systems and data management. This is a team assignment and all code will be written in Java. The assumption is that you have installed the latest version of Java SDK and the JRE on your machine.

For this assignment, a server process called DataGen (DataGen.class) has been provided. This server process simulates a remote weather buoy for the Earth Observing System (EOS) created by NASA (<http://eospsso.gsfc.nasa.gov/>). The DataGen server process provides streaming data. The data is delivered in frames that are structured as follows:

Hours: Integer data that represents the data sample time stamp hours.  
Minutes: Integer data that represents the data sample time stamp minutes.  
Seconds: Integer data that represents the data sample time stamp seconds.  
Humidity: Floating point data that represents the relative humidity  
Temperature: Floating point data that represents temperature in degrees Fahrenheit  
Pressure: Floating point data that represents pressure in kilo-Pascals (kPa)

The DataGen server is a java process that provides the above data frames at a rate of approximately 1 per second. To start the DataGen server, open a command window and type:

```
java DataGen <port number>
```

Note that <port number> is an optional argument that allows you to specify a port number. If you do not supply a port number, DataGen will use port 6789 as a default.

To help you get started, a simple client template has been provided (Client.java). This illustrates how to connect and read data frames from the DataGen server. To compile this sample program, open a command window and type:

```
javac Client.java
```

Once compiled, you can run the Client application by typing:

```
java Client <port number>
```

Note that <port number> is an optional argument that allows you to specify a port number. If you do not supply a port number, Client will use port 6789 as a default. The port number that you used for DataGen and Client, must be the same.

## Assignment Description:

Your job is to collect data for 5 minutes and store it in on disk in a format of your own design or selection. You can select/design any kind of storage structure/strategy you like that supports the following user needs and adheres to the following constraints and assumptions:

1. Assume that users are scientists who want to access the data from applications that they will write. The users need to be able to search and access specific data based on various search criteria. All data storage details should be completely hidden from the users. Users expect to be able to use your APIs to search and access data.
2. You will need to provide the following APIs:
  - *Summary data function* that tells users what measurements are in the file, the measurement units, the measurement type, how much data has been collected (in bytes), and the time span of the data.
  - *Search by finding the first-occurrence function* of a particular data value and when it occurs as well as a *next-occurrence function* that finds the next occurrence of the data value as well as when it occurs in the data. The function should return the time index of the frame where the data begins.
  - *Read data function* that reads and fills buffers of data. Users will provide the start and stop time as well as the measurements of interest. The API will return buffers of data to the user. Note that if there is more data than can be reasonably packed into a single buffer, the API should signal the programmer in some way that there is more data to read.
3. Because the users will use Java to write their applications, you must use Java to build your APIs. Assume that users will not install anything else on their systems to use your APIs other than the JRE and Java SDK that is already installed on their system.
4. Your user APIs should be configurable and support reading data streams with different measurements other than those provided by DataGen. You should assume that time in the current format will always be present.

## Deliverables:

Your deliverables include:

APIs and Documentation as follows:

- Describe the design of file structure and your rational for the design. Describe how your design meets all of the functions above. **Provide an analysis of your search algorithms in terms of performance and describe any limitations and tradeoffs you made.** Be detailed and clear in your discussion.
- Provide an API programmers guide. This should completely describe how to use the APIs to you designed to build applications. This should be a high quality user's guide. Assume that users do not have access to your source code, but that you are only distributing executable code to customers.

NOTE: Please be thorough, but as concise as possible in your discussion. Note that grammar, good style, and format counts. Any reasonable formatting style/structure is acceptable.

Provide your documents, API source code, and executables (java and class files), and data files. Put all of these materials (described above) into a zip file and turn it in electronically (blackboard or email). Use Windows file compression or Winzip to compress your files. Include a “readme” file in your archive that explains the contents of your archive. Name your archive as follows: A5+YourFamilyName. So my assignment archive would be named: A5+Lattanze.

**Grading:**

This assignment is worth 100 points as follows:

- **Documentation:** 50 points. We will evaluate the overall quality of the write up and the readability/understandability from the user’s standpoint (who is a programmer).
- **Implementation:** 50 points. We will evaluate the ease of use, the quality of the source code, and the data structure design.