Colombian School of Engineering
Enterprise Architecture, 2020-2

# INTRODUCTION TO COMPUTER SYSTEM DESIGN

Sergio Alejandro Bohórquez Alzate

21/08/2020

## 1 Introduction

In this work we specified the architecture and solution to the problem of a statistical calculator that calculates the arithmetic mean and the standard deviation, for this we built a java system built with maven and later we implemented a linked list to store the information and to be able to make the respective calculations, thanks to the implementation we generated a project that manages to solve the problem through the reading of data with and printing in the console using maven, we automated the testing process with Junit, and generated the respective documentation. The project is hosted in a GitHub repository , deployed in Heroku and continuous integration in CircleCi.

## 2 Objectives

- To understand the basic operation of Spark Framewokr, Maven, Heroku, CircleCi and Git.
- To implement a basic statistical calculator with the entire software development cycle.
- To understand how a linked list works.

## 3 Problem Description

Write a program that meets the following specification (the data structure to be used is a linked list implemented from scratch).

Input: A set of n floating numbers read from a text file

Output: Arithmetic mean and Standard Deviation of set.

**Example:**

| Column 1 | Column 2 |
|---|---|
| Estimate Proxy Size | Development Hours |
| 160 | 15.0 |
| 591 | 69.9 |
| 114 | 6.5 |
| 229 | 22.4 |
| 230 | 28.4 |
| 270 | 65.9 |
| 128 | 19.4 |
| 1657 | 198.7 |
| 624 | 38.8 |
| 1503 | 138.2 |

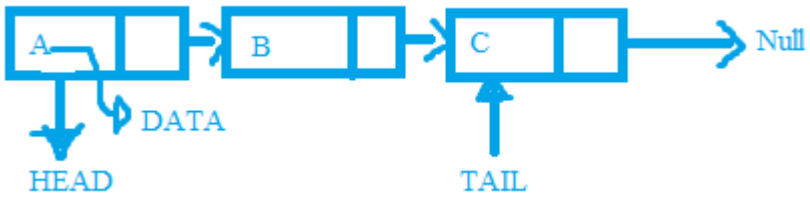
Sample input 1 (Taken from Virtual Campus-Colombian School of Engineering AREP)

| **Expected Value** | |
|---|---|
| *Mean* | *Std. Dev* |
| 550.6 | 572.03 |
| 60.32 | 62.26 |

Sample output 1 (Taken from Virtual Campus-Colombian School of Engineering AREP)

## 3.1 Basic Definitions to understand the problem

1. **LinkedList**: a linked list is a linear data structure whose elements are stored in different memory spaces. These elements are called nodes and are linked by pointers, in the following image we see an example of a linked list.[2]




LinkedList Example (Source: Self-Production).

There are some variations of a linked list, here are some of them:

- **Simple LinkedList**: in a simple linked list, each node is only linked to the following
- **Doubly LinkedList**: in a doubly linked list each node is linked to the next and previous one.

- **Circular LinkedList**: in a circular linklist, each node is linked to the next and previous one, and the head is linked to the tail.

The basic operations of a linked list are:

- remove() removes the first node of the linkedlist.
- add(Node ?) add an element to the right side of the linked list.
- getFirst() returns the first node of the linkedlist
- getLast() returns the last node of the linkedlist

2. **Arithmetic mean:** the average of a set of numerical values, as calculated by adding them together and dividing by the number of terms in the set.[1]

$$\text{Arithmetic Mean} = \sum \frac{x_i}{n}$$

Arithmetic mean formula.

3. **Standard deviation:** A quantity expressing by how much the members of a group differ from the mean value for the group.[3]

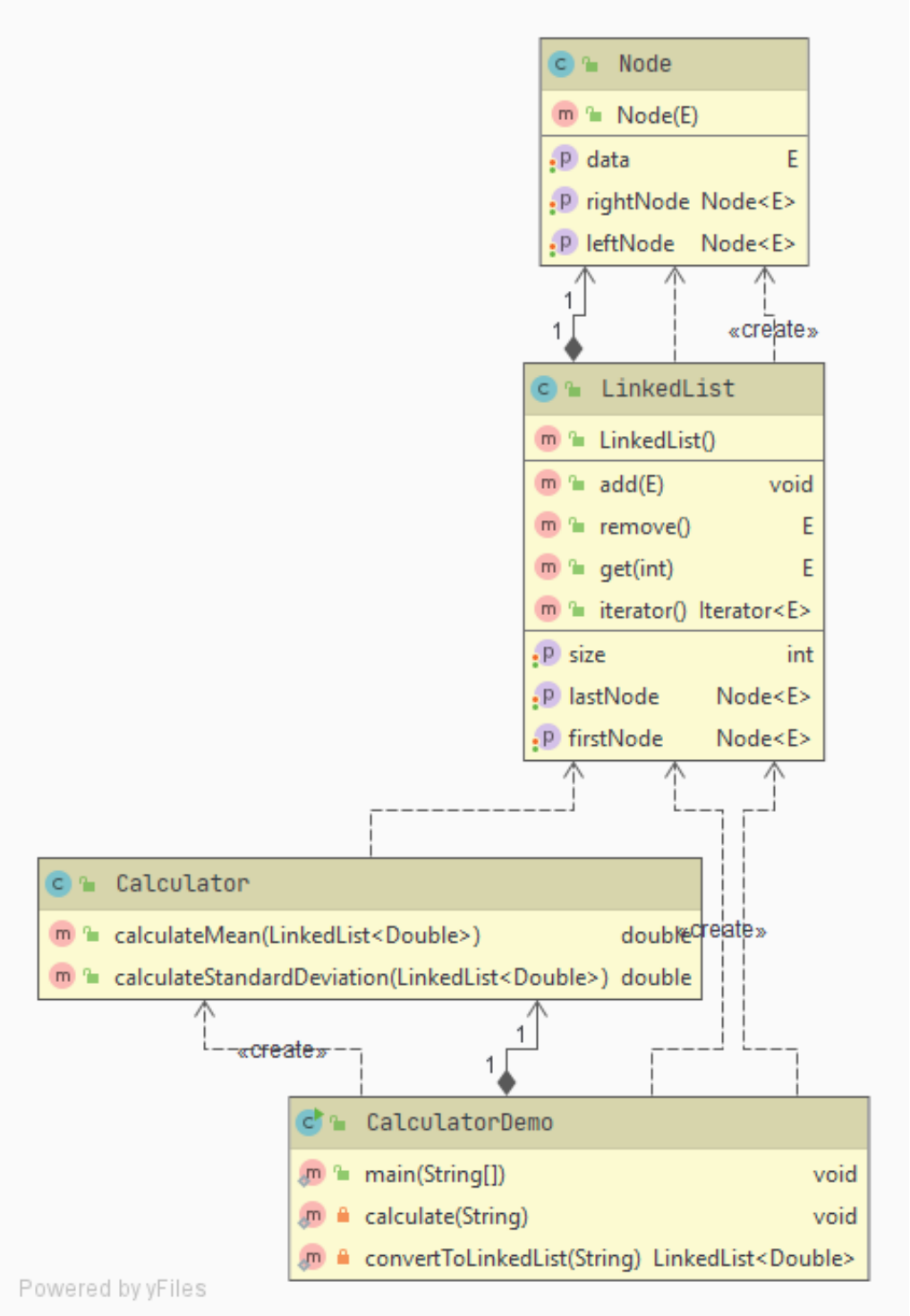
$$S = \sqrt{\frac{\sum (X - \overline{X})^2}{N}}$$

Standard deviation formula.

# 4 Architecture

## 4.1 Class Diagram

A design of the solution is modeled to be able to implement in java the system previously described.The following image shows the class diagram.



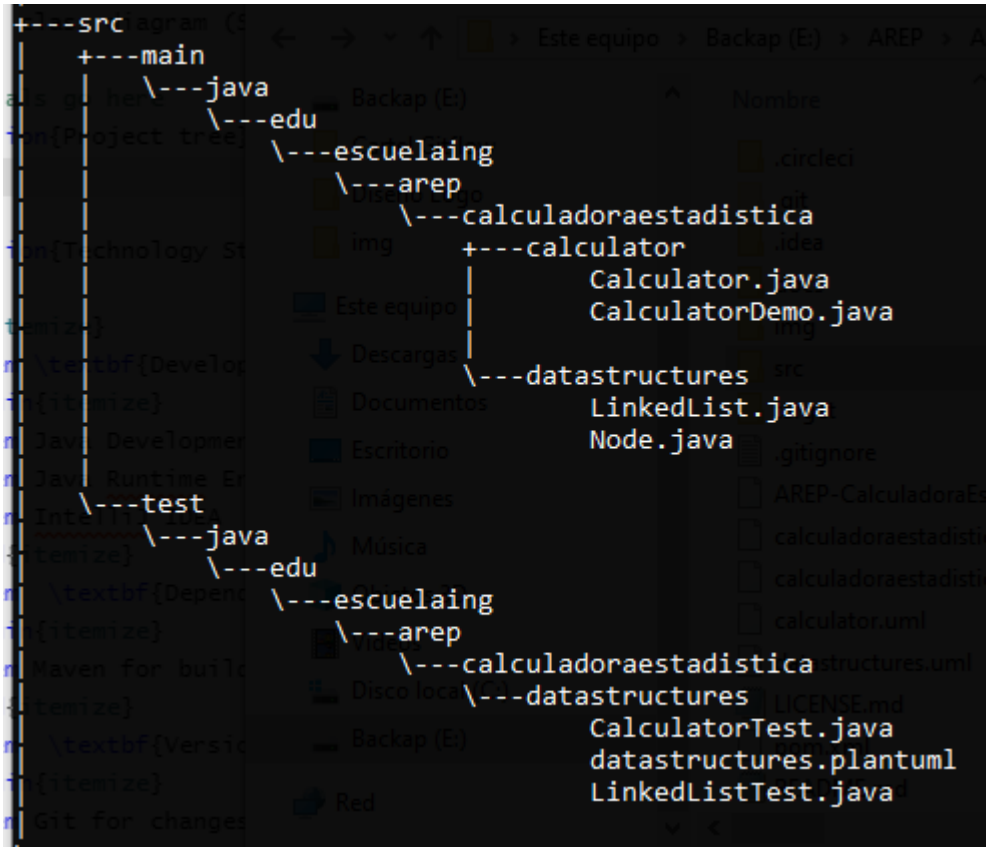

Solution class diagram (Source: Self-Production)

## 4.2 Technology Stack

- **Development environment**
  - Java Development Kit 8.0
  - Java Runtime Environment 8.2
  - IntelliJ IDEA
  - Spark framework 3.0
- **Dependency Manager**
  - Maven for build the project,
- **Version control**
  - Git for changes and version control.
- **Deployment**
  - Heroku
  - CircleCI.
- **Tests**
  - JUnit
  - PostMan

## 4.3 Project tree

Next it will be shown the structure of the project generated by the maven dependency manager, with the objective of giving more understanding to the reader about how the solution was implemented.




### 4.4 Documentation

If you want to see more information about the implementation of this program, you can go to the following link `https://alejandrobohal.github.io/AREP-CalculadoraEstadistica/apidocs/index.html` in which you will find a specification for each class found in the class diagram shown above.

## 5 Testing the application

### 5.1 How to run it

You need to have the following programs or packages installed on your operating system

- Java Development Kit 8.0
- Git
- Maven

**From a terminal:**

1. Download the project and change the directory with the following commands

```
git clone https://github.com/AlejandroBohal/StatisticalCalculatorSpark

cd Arep-CalculadoraEstadistica
```

2. Compile the project with following command

```
mvn clean install
```

3. Once the project is built, we can execute it with Maven in the following way:

```
java -cp target/classes:target/dependency/*


edu.escuelaing.arep.calculadoraestadistica.webapp.Spark
```

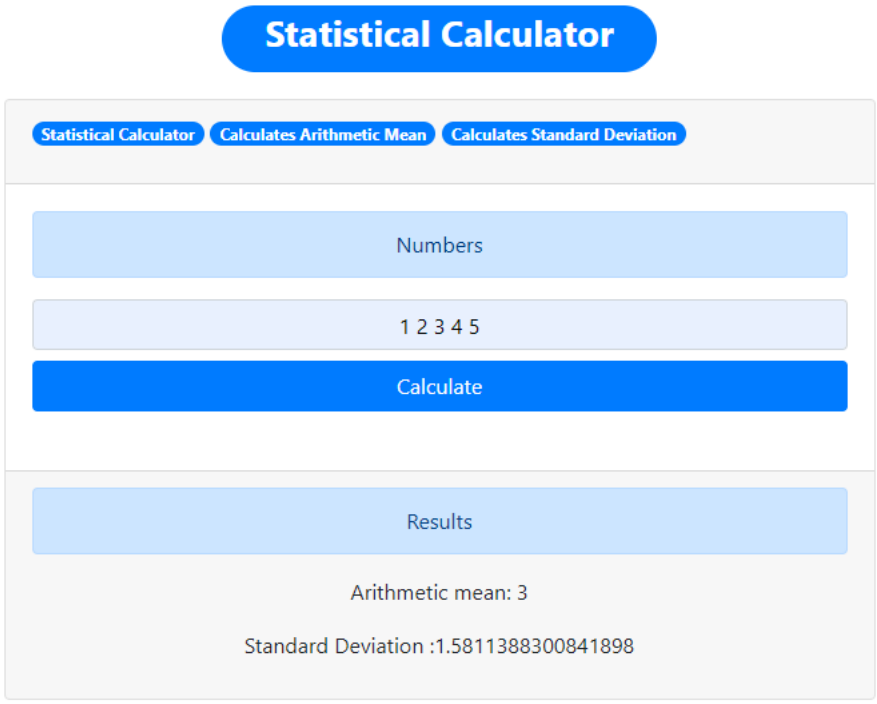
To run the program only txt files located in the root directory of the program are supported, as an example 2 files were left in the program. To test the other file we only need to change TestData.txt to TestData2.txt in the previous command. In the file we must put numbers separated by spaces, for each line will be calculated both arithmetic mean and standard deviation as seen below:

```
160 591 114 229 230 270 128 1657 624 1503
```

To run the project in a local server run the following command, remember that Heroku Cli is required.

```
heroku local web
```

Also you can see the Web App Deployed to Heroku in the following link:
'http://fathomless-bayou-96611.herokuapp.com/
and you'll see something like this:

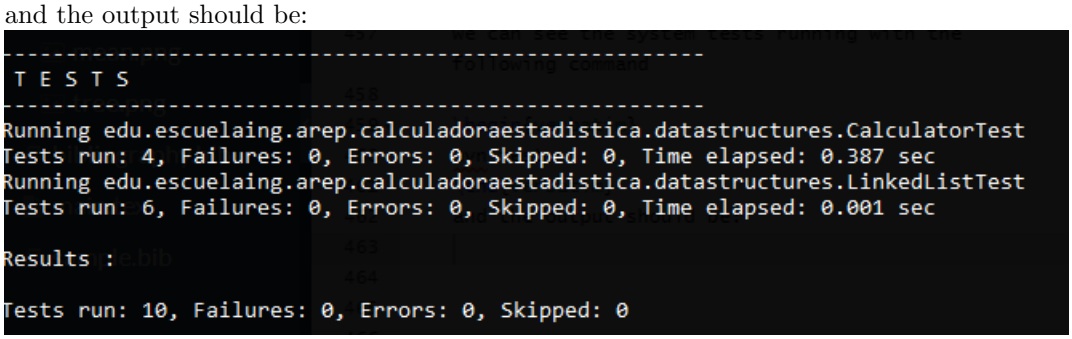



### 5.2 Testing

we can see the system tests running with the following command

```
mvn test
```

and the output should be:




Testing application (Source: Self-Production)

## 6 Conclusions

- We used maven and git in addition to java to solve a simple statistical problem, which proposes a quality implementation for the problem.

- Technologies such as Spark, Heroku and Circle Ci were used to provide a simple development environment and the great potential of these was demonstrated.
- The objectives set out at the beginning of this document were met.

# References


[1] *Arithmetic mean*. `https://en.wikipedia.org/wiki/Arithmetic_mean`. Accessed on 2020-12-08.

[2] *Linked List Data Structure*. `https://www.geeksforgeeks.org/data-structures/linked-list/`. Accessed on 2020-12-08.

[3] *Standard Deviation and Variance*. `https://www.mathsisfun.com/data/standard-deviation.html`. Accessed on 2020-12-08.