

PARADISE Explorer/Data Processing Calculator

To help demonstrate the power of PARADISE along with visualizing its components we propose the students build a calculator where a user can tune a multitude of inputs and variables that would exist in a PARADISE framework. These inputs could consist of number of records to ingest, size of the datasets, number of computers used as processors, amount of RAM and CPU power, number of additional pipelines, and number of queries coming across the DQSL just to name a few. Once the user has selected the inputs they can then calculate and see the “data” move across a visualization of PARADISE with expected throughput for each step with the estimated total amount of time it would take. The user would also be able to select other data systems like Rainmaker or a traditional ETL process to compare how much time it would take for their data to go from ingest to query. This would be useful as a demonstration of how PARADISE differs from other solutions and can help give “free” accurate estimations to potential customers of how PARADISE can help them.

The project team will meet the following objectives:

1. Understand the PARADISE data Framework
2. Identify inputs into data frameworks and quantifying the corresponding outputs
3. Develop a data model for processing inputs into the PARADISE framework Calculating time as measure of inputs into the data framework.
4. Make comparison recommendations for: PARADISE, Rainmaker, and Big Data Platform
5. Develop a functioning calculator with a GUI that can compare outputs from PARADISE and two other data processes.
6. Have the ability to process 1 TB of data from 5 sensors and measure how long will it take to extract, transform, load, and visualize it that into a data system.

Project Sponsor: AnaVation LLC