

Show Me the Data: Analyzing Production Workloads for System Design

Project in Concurrent and Distributed Systems 236371¹

Supervisor: [Dr. Gala Yadgar](mailto:gala@cs.technon.ac.il) (gala at cs.technon.ac.il)

There are many “rules of thumb” that are used to describe the characteristics of storage accesses, such as the 80-20 (Pareto’s) law, that states that approximately 80% of I/O requests access only 20% of the data. These rules are used for system design and optimization, and therefore must be validated periodically, in order to ensure that system design indeed matches requirements.

The project will consist of processing and analyzing I/O traces from production and research servers, and validating some commonly used assumptions relevant for large scale system design. The students will review the relevant literature about each assumption, determine its implications on the specified system, and evaluate its validity to the tested data.

Requirements:

Part I

The students will parse and process one or two sets of traces, and convert them to a standard, minimal, trace format to enable comparison between different traces, and use for storage simulation.

Part II

The students will implement several analysis programs, to answer a list of research questions by means of providing graph input to the analysis. After meeting and verifying with the project supervisor, the students will run a full and comprehensive set of experiments to produce the relevant graphs for all the traces.

Part III

The students will survey relevant sections of recent literature on the characteristics and analysis of I/O traces in large scale storage systems. They will compare the results of their analysis to what is described in the literature, and answer the relevant research questions accordingly.

Platform:

The data traces are stored in the [Computer Systems Lab \(CSL\)](#), where they can be used for access and processing on Linux platforms.

Prerequisites:

Operating systems course. The project is designed for a group of 1 or 2 students.

¹ Also available as Project in Operating Systems (236366)