

A1: Data Acquisition

Tyler Brown

Architecture

Performance

Run Time

Table 1: Run Times by Process

Process	Log Start Time	Log End Time	Hours Estimate
Get repos	2019-01-21 02:54:16	2019-01-21 04:40:18	1.50
Update repos	2019-01-22 03:08:06	2019-01-22 03:11:34	0.05
Extract data	2019-01-22 07:31:02	2019-01-22 09:36:33	2.00

Get repos

Disk Space

Table 2: Disk Usage by Name

Name	Disk usage
commits.csv	881M
files.csv	1.9G
messages.csv	1.3G
/repos/	98G