

Streaming Log Analytics (only Open Source Tools to be used)

Reference (https://d1.awsstatic.com/Projects/P4113850/aws-projects_build-log-analytics-solution-on-aws.pdf)

The assignment solution should achieve the following objectives:

- 1) Log records are continuously forwarded from the web server (can use <https://github.com/kiritbasu/Fake-Apache-Log-Generator> to generate fake logs) to the Log Analytics system.
- 2) System writes each log record to a persistent storage in an efficient manner.
- 3) Log records are not lost during the transit and made available to the processor whenever required.
- 4) Continuous monitoring of the streaming input data is happening.
- 5) An aggregated data is generated every minute and outputs that data through appropriate mechanism.
- 6) A view of the streaming data using Dashboards/Visualizations is possible.

Submission requirements:

1. Short note on describing (in WORD / PDF document)
2. the streaming data pipeline architecture
3. components, tools used and purposes of the same
4. data flows
5. business logic used along with commands and queries
6. dashboards / visualizations screenshots

Python programs implementing

1. Relevant data flows
2. Data monitoring and processing logic
3. Commands / Queries used with their outputs