IST-719

Topic: DEEP NETWORK VISUALIZATION

Work-In-Progress Report

Student: Sharat Sripada
(vssripad@syr.edu)

1.0 Section-1

Computer networks (on-premises or hyper-scalar environments) comprise of nodes or functions that provide applications with networking and security. Developers writing software code often brace methods to emit events or logs to a centralized log aggregator, so it is easier to troubleshoot exceptions and correlate failures across entities viz. hardware and software.

This project titled *Deep Network Visualization*, strives to build meaningful visualizations on event or log data, thus help monitor health of the network and root-cause failures to individual entities.

1.1 Description of Dataset

The dataset is exported from a centralized log aggregator in CSV format (separated by ‘,’) and loaded into a data-frame using the read.csv() function. A sample line is shown below:

```
240,"<179> 2021-06-23T19:16:02.291784+00:00 nsx-edge-1-c0 NSX 32192 LB [nsx@6876 comp=""nsx-edge"" subcomp=""nsx-edge-lb.lb"" level=""ERROR""] ""no virtual server defined in lbs ff70393e-4490-471b-8007-8e22b647fa8f."""",2ed1b9c6-7c72-4ced-9078-03401a93eaba,545,82883403,NSX,,nsx-edge,,v4_dcb8d3e0,local6,/var/log/li-syslog,nsx-edge-1-c0,ERROR,,,LB,22.0,3.0,32192,,,err,20.20.216.157,nsx-edge-lb.lb,,2021-06-23 19:16:02.291,,,,,,,,,,,,,,,,nsx-edge,,,nsx-edge-lb.lb,,,,,,,,,,,,,,,,,,,,,,,,,,,,,edg  lb  comp edg subcomp edg lb
```

NOTE: Each row of data is typically comprised of timestamp, severity, network-node, software component and a brief excerpt of an event or log that tells an exception that occurred.

1.1.1 Size of data

The CSV has ~320k rows of events or logs and 78 columns leading to a score of >> 100:

```
(78 * 4) * (323972/100) >> 100
```

1.2 Story

Event or log data from disparate network nodes and software stacks can be hard to interpret for admins owing to complexity and scale. Translating raw event or log data to useful visualizations with transitions, from a macro to micro level view would greatly help narrow down on issues.

In datacenters or cloud environments, this can translate to faster recoveries from anomalies and less impactful outages.

1.3 Audience

Any administrator or operations personnel managing large scale datacenter or cloud environments who is entrusted with the role of ensuring high reliability of service.

1.4 Questions

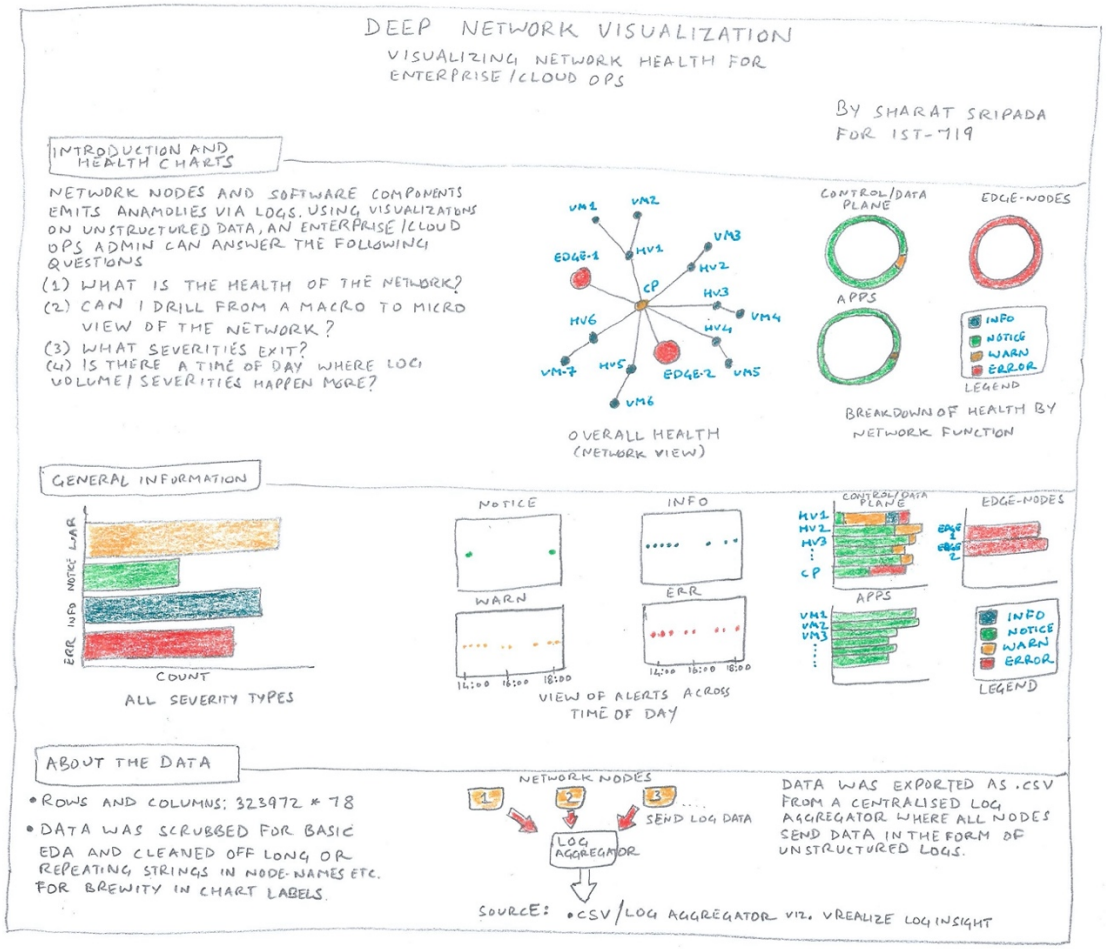
The visualizations will seek to answer in the following:

1. What error severities exist? see plot **Fig1: Log severity counts** in Section-3
2. Are there certain nodes generating higher volumes of events or logs? see plot **Fig2: Logs grouped by node** in Section-3
3. Was there a time of the day where certain severity of events or logs were high? see plot **Fig3: Log severity across time-of-day** in Section-3
4. When grouped by network function, are degraded nodes clearly visible? see plot **Fig4: Log severity grouped by function** in Section-3
5. What is the topology and health of the network? see plot **Fig5 – Network Topology and Health** and **Fig6: Network Health grouped by function** in Section-3

Overall, the goal is making charts drillable transitioning from a macro to micro level view of the network (or view-versa).

2.0 Section-2

2.1 Sketch of high-level design



3.0 Section-3

3.1 Visualizations

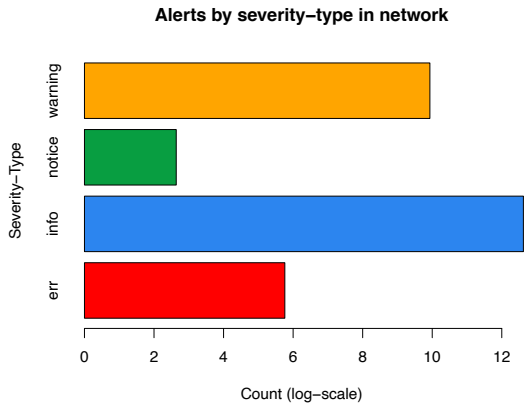


Fig1: Log severity counts

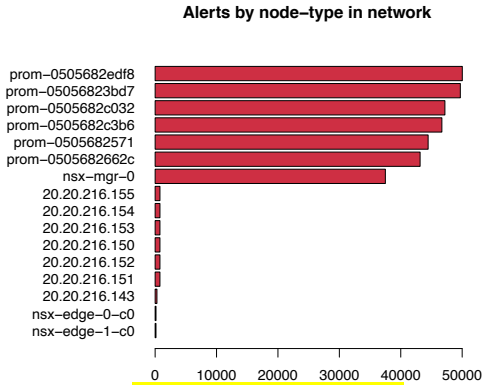


Fig2: Logs grouped by node

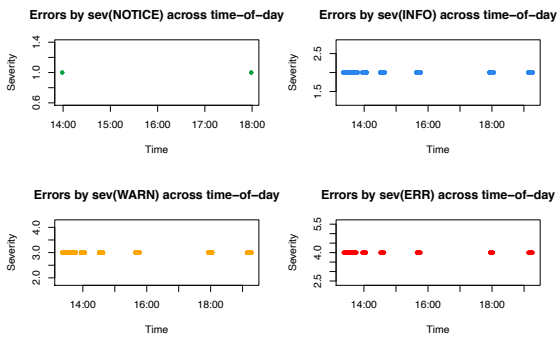


Fig3: Log severity across time-of-day

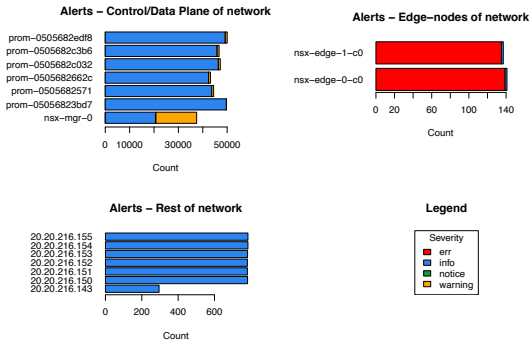


Fig4: Log severity grouped by function

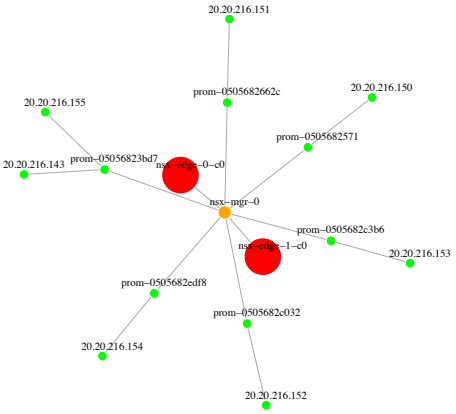


Fig5: Network topology and health

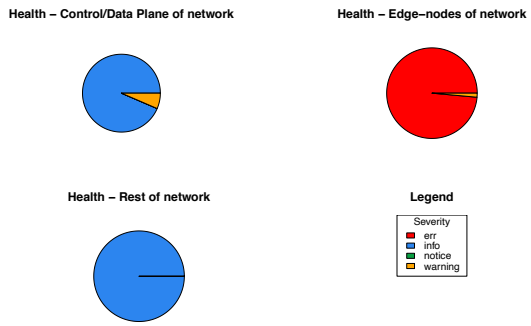


Fig6: Network Health grouped by function