## **Call for Contributions - Potential works for contributors**

- 1. Use Chaos tools to generate Data Infrastructure Data and Events.
  - a. There are many 'Chaos-Tools', such as litmus. We want to use this tool to create chaos and generate some data (infrastructure metrics). Then we can use this data and corresponding Chaos for some failure prediction works.
    - b. The main purpose is to solve the problem of 'Lack of Dataset'.
    - c. This is an experimental work we are not sure if we can 'create a usable dataset'.
- 2. Time-Varying, Load-Varying tool to emulate cnf/vnf. (Ongoing In collaboration with ViNePerf)
  - a. Enhance stress-ng to 'emulate a vnf/cnf' in terms of resource (cpu, memory, storage & N/w).
    - b. Enhance to check if the 'configured' load is same the 'actual' load. If there is a difference, then the cluster cannot support such CNF/VNF. Ex: If we configure memory I/O to be 10000 iops, but we achieve only 8000 iops, then the cluster is not suitable for that CNF (which requires 10000 iops).
    - c. We also define what 'failure of cnf' means using these differences (actual vs configured).
    - d. This enhanced stress-ng also has other usecases.
      - i. Build a custom VM/Container which tries its best to fail based on 2 and 3. (Ongoing In Collaboration with ViNePerf)
        ii. This can act as 'Noisy-Neighbor' test the impact of this noisy-neighbor on dataplane performance.
- 3. Enlist the operations (valid) that can be done from VM/Container, and that can make the VM/Container fail.
  - a. Study to understand what (kind of operations) can cause failures of CNF/VNF.
  - b. Once we have that information, we can emulate it.
- 4. Simplify deployment of Acumos
  - a. Acumos is a LF's Al/ML framework.
  - b. Just started with Kubeflow Vs Acumos comparative study ( 3 students are working on this project )
  - c. Goals: Dynamically plug&play different models, integrate with any data-source.
  - d. As of now, no framework is used we are running with jupyter notebooks, with tensorflow libraries.
- 5. Emulating Monitoring data and events GANs? (Started)
  - a. Synthetic data (metrics and logs Timeseries. Collectd-metrics, system logs, etc.) generation using GANs.
  - b. Currently the performance poor, we want to take this problem to ITU.
- 6. Data Extraction utility Given a data model, the tools should extract the required data from the large amount of data. (Ongoing Sridhar)
  - a. A tool to extract data from different sources.
  - b. Currently prometheus is done, working on Elasticsearch. Python APIs to extract the data-source.
  - c. This is ONLY required if the framework is not integrated to these data sources.
  - d. Input: Time begin and end, and filters. Filters can be used 'exclude' some columns of the data Hence data model plays importance. This also helps to achieve 'anonymization' of data.
- 7. ModelSelector Wizard tool that will ask user about the problem and data, and will suggest which ML technique to use (within Supervised, Unsupervised, and reinforced). (Ongoing - Sridhar, Kanak, Akanksha)
  - a. Completed.
  - b. https://github.com/opnfv/thoth/tree/master/tools/modelselector
  - c. Not very well tested!
  - d. Algorithm is just recommended.
  - e. This work will taken forward to MaaS.
    - i. <u>MaaS: Minimal Questions will be asked, dataset is taken, and try out with different models, and best performing is chosen and provided to the user.</u>
      - 1. Minimal: Questions about the problem and not the dataset.
      - We plan to use Kubeflow/Acumos here. Each and every model as a separate container, and we can do plug and play.
- 8. Openstack-LogAnalysis with NLP (Started)
  - a. Ongoing using Google's BERT.
    - b. Currently using China-Unicom's Dataset.

Not Started: 1, 3 and 4.

Ongoing: 2 (Kuldip Yadav ), 5 (renukananda td ), 6 (Sridhar Rao ), 8 (Rohit Singh Rathaur)

Completed: 7

Underlined: Student Volunteers already started working on these.

Meetings where some these projects are elaborated:

2021-07-23 AI/ML for NFV Meeting Minutes