SRO Data Lake

ERSAP Data Lake Implementation

Vardan Gyurjyan
JLAB EPSCI Group

gurjyan@jlab.org

Jefferson Lab

U.S. DEPARTMENT OF ENERGY | Office of Science

JSA

SRO Primary Goal

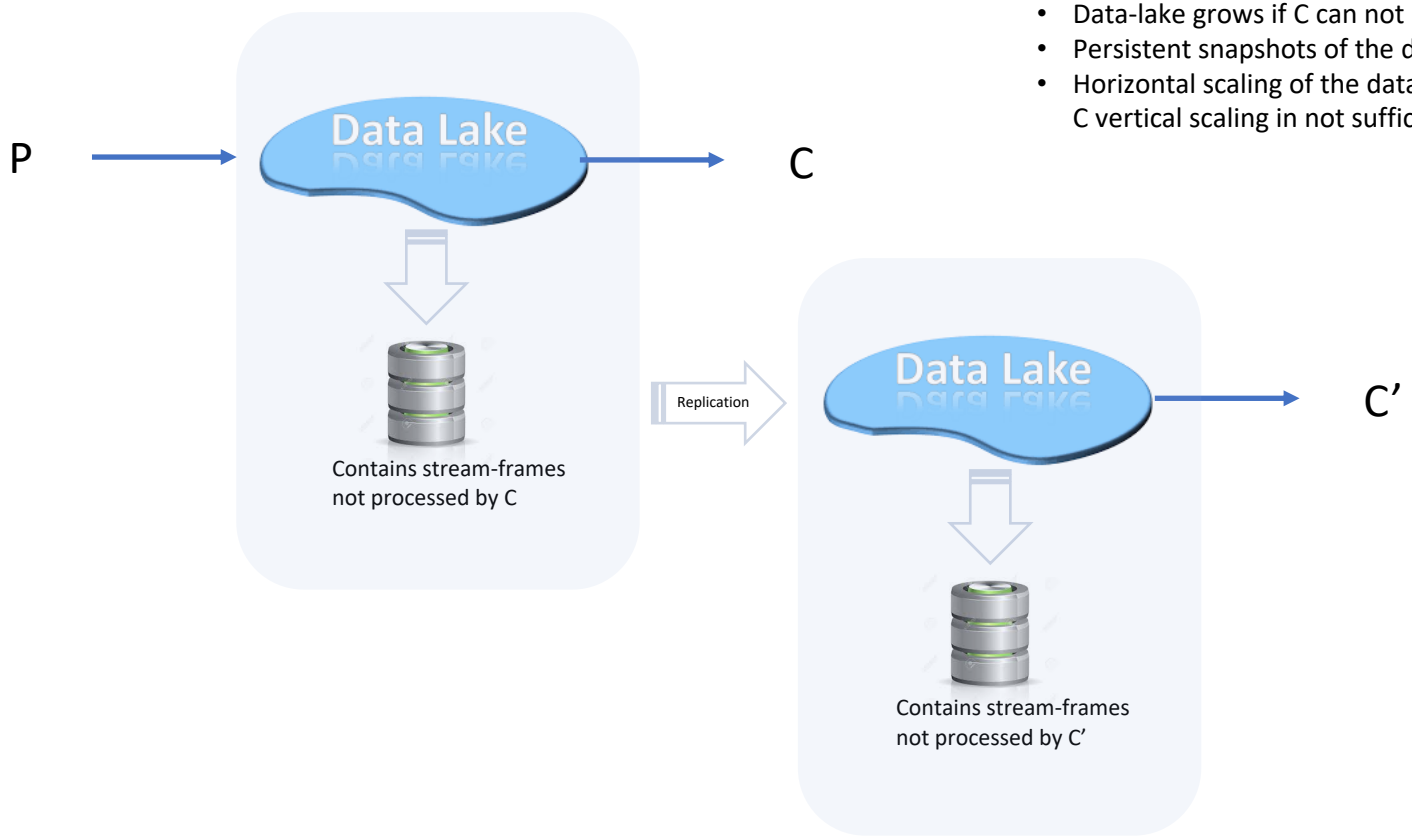
- Accept data with minimal loses
  - Minimize SRO dead time
  - Give more time for backend (services down the stream) to process data
  - Stream-strike functionality: similar to signal delay cables during the triggered readout system
- Data-Lake as a design choice for the stream-strike functionality
  - Temporary in-memory data storage with possible persistency in case of backend dyskinesia
  - Centralized storage of unprocessed frames from all streams
  - Ideal for ML applications

ERSAP Data Lake

- Low latency, in-memory storage for structured stream-frames
- Elastically scalable
- Point-in-time persistency of in-memory stream-frames
  - Contains stream-frames missed by the back-end
  - User defined horizontal scaling threshold
    - Tolerable back-end slowness after which
      - Persisted data replication on a different node
      - ”delayed” parallel processing of data-frames
- Pilot implementation based on Redis
  - In memory NoSQL database
  - Open source
  - Large user/support community

Data Lake Scaling

- C reads and removes the stream-frame
- Data-lake grows if C can not keep up with P
- Persistent snapshots of the data-lake
- Horizontal scaling of the data-lake in case C vertical scaling in not sufficient



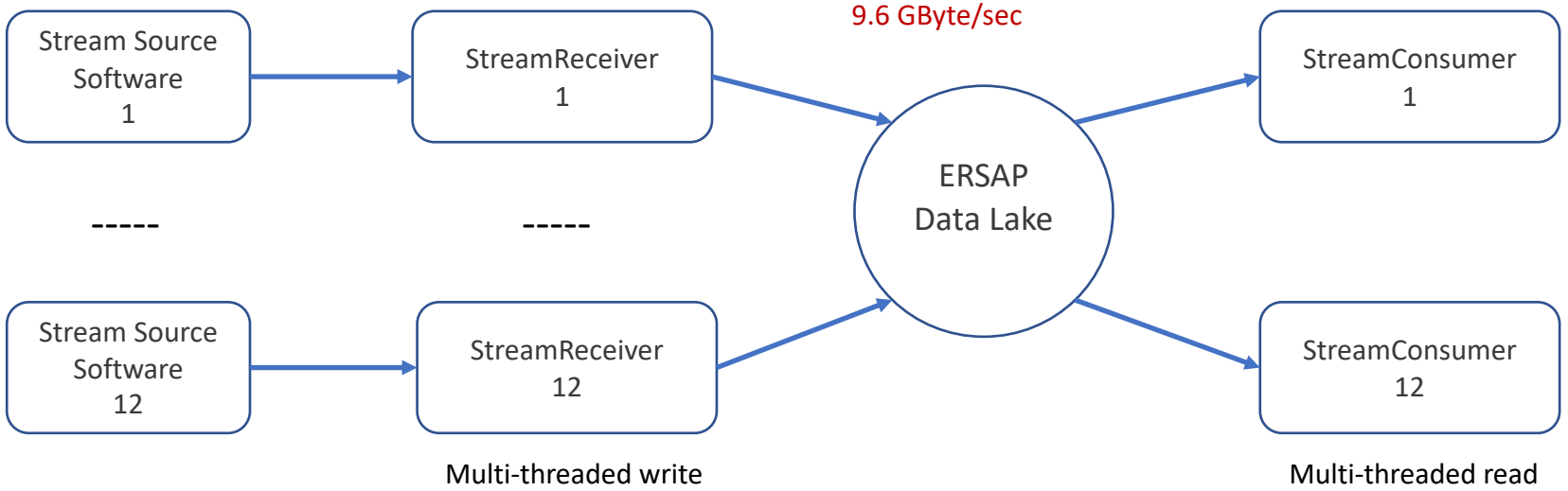
CLARA Data Lake Services

- Clara data-lake IO services
  - Raw frame read/write
  - Clara StreamReceiver service
    - Frame header parsing
    - Frame payload storage, hashed by the frame-time for easy processing
  - Clara StreamConsumer service
    - Consumes raw frames or farm payloads of multiple streams

Rapid Prototyping

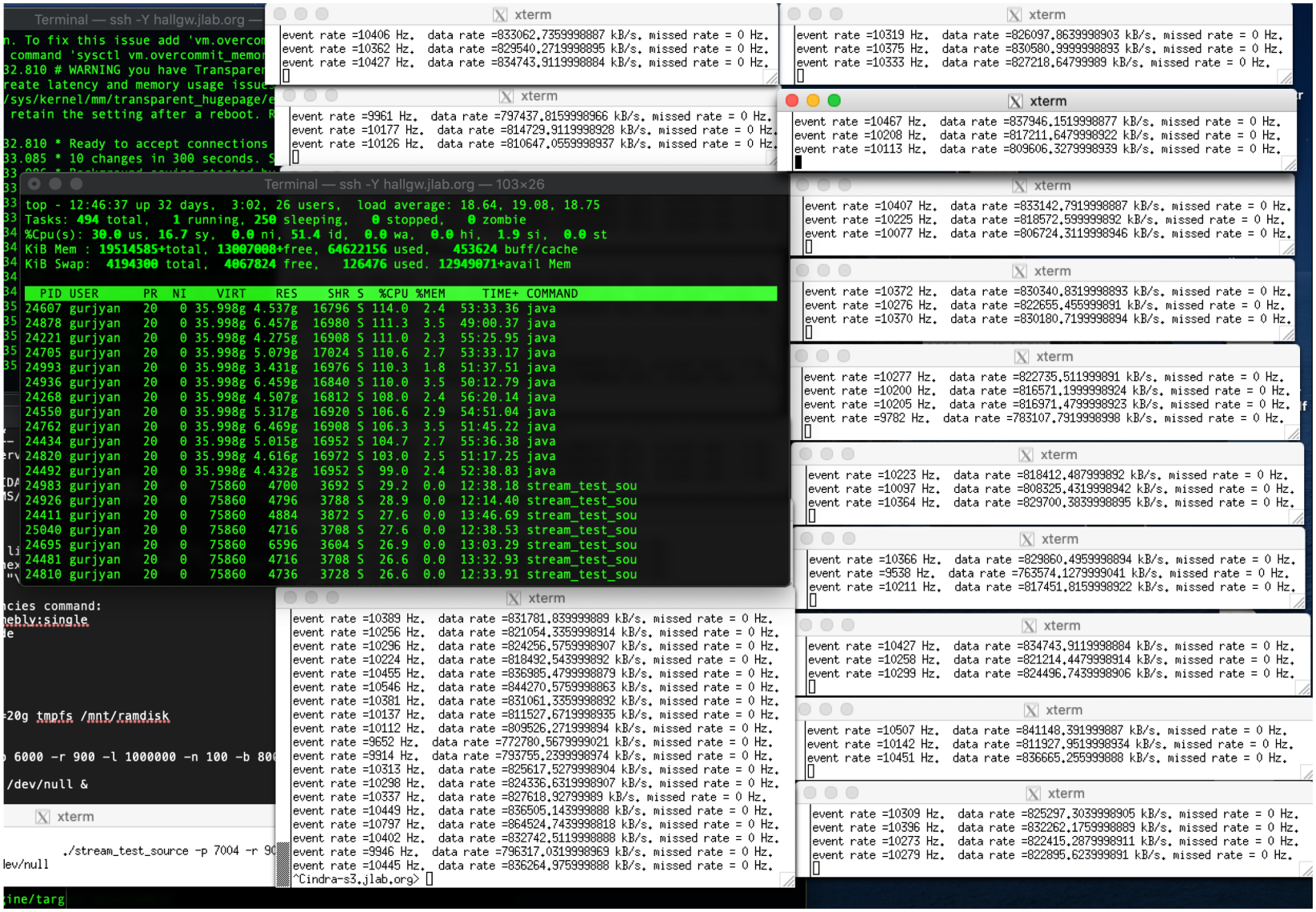
Event-rate = 10KHz
Data-rate = 800MByte/sec

Lossless data throughput
(in and out data-lake)
9.6 GByte/sec



VTP simulated Data Frame
800MByte/sec max frame size per VTP link.
65us window, 8 FADC per 10Gbps link, maximum 1 hit reported per 32ns, 4 bytes per hit:
13 (Header) + (1+65536/32)*8 = 18454 32bit words => 73816bytes

Screen Shot


Terminal — ssh -Y hallgw.jlab.org —
xterm
event rate =10406 Hz. data rate =833062.7359998887 kB/s. missed rate = 0 Hz.
event rate =10362 Hz. data rate =829540.2719998895 kB/s. missed rate = 0 Hz.
event rate =10427 Hz. data rate =834743.9119998884 kB/s. missed rate = 0 Hz.
event rate =10319 Hz. data rate =826097.8639998903 kB/s. missed rate = 0 Hz.
event rate =10375 Hz. data rate =830580.9999998893 kB/s. missed rate = 0 Hz.
event rate =10333 Hz. data rate =827218.64799989 kB/s. missed rate = 0 Hz.
event rate =9961 Hz. data rate =797437.8159998966 kB/s. missed rate = 0 Hz.
event rate =10177 Hz. data rate =814729.9119998928 kB/s. missed rate = 0 Hz.
event rate =10126 Hz. data rate =810647.0559998937 kB/s. missed rate = 0 Hz.
event rate =10467 Hz. data rate =837946.1519998877 kB/s. missed rate = 0 Hz.
event rate =10208 Hz. data rate =817211.6479998922 kB/s. missed rate = 0 Hz.
event rate =10113 Hz. data rate =809606.3279998939 kB/s. missed rate = 0 Hz.
Terminal — ssh -Y hallgw.jlab.org — 103×26
top - 12:46:37 up 32 days, 3:02, 26 users, load average: 18.64, 19.08, 18.75
Tasks: 494 total, 1 running, 250 sleeping, 0 stopped, 0 zombie
%Cpu(s): 30.0 us, 16.7 sy, 0.0 ni, 51.4 id, 0.0 wa, 0.0 hi, 1.9 si, 0.0 st
KiB Mem : 19514585+total, 13007008+free, 64622156 used, 453624 buff/cache
KiB Swap: 4194300 total, 4067824 free, 126476 used. 12949071+avail Mem
PID USER PR NI VIRT RES SHR S %CPU %MEM TIME+ COMMAND
24607 gurjyan 20 0 35.998g 4.537g 16796 S 114.0 2.4 53:33.36 java
24878 gurjyan 20 0 35.998g 6.457g 16980 S 111.3 3.5 49:00.37 java
24221 gurjyan 20 0 35.998g 4.275g 16908 S 111.0 2.3 55:25.95 java
24705 gurjyan 20 0 35.998g 5.079g 17024 S 110.6 2.7 53:33.17 java
24993 gurjyan 20 0 35.998g 3.431g 16976 S 110.3 1.8 51:37.51 java
24936 gurjyan 20 0 35.998g 6.459g 16840 S 110.0 3.5 50:12.79 java
24268 gurjyan 20 0 35.998g 4.507g 16812 S 108.0 2.4 56:20.14 java
24550 gurjyan 20 0 35.998g 5.317g 16920 S 106.6 2.9 54:51.04 java
24762 gurjyan 20 0 35.998g 6.469g 16908 S 106.3 3.5 51:45.22 java
24434 gurjyan 20 0 35.998g 5.015g 16952 S 104.7 2.7 55:36.38 java
24820 gurjyan 20 0 35.998g 4.616g 16972 S 103.0 2.5 51:17.25 java
24492 gurjyan 20 0 35.998g 4.432g 16952 S 99.0 2.4 52:38.83 java
24983 gurjyan 20 0 75860 4700 3692 S 29.2 0.0 12:38.18 stream_test_sou
24926 gurjyan 20 0 75860 4796 3788 S 28.9 0.0 12:14.40 stream_test_sou
24411 gurjyan 20 0 75860 4884 3872 S 27.6 0.0 13:46.69 stream_test_sou
25040 gurjyan 20 0 75860 4716 3708 S 27.6 0.0 12:38.53 stream_test_sou
24695 gurjyan 20 0 75860 6596 3604 S 26.9 0.0 13:03.29 stream_test_sou
24481 gurjyan 20 0 75860 4716 3708 S 26.6 0.0 13:32.93 stream_test_sou
24810 gurjyan 20 0 75860 4736 3728 S 26.6 0.0 12:33.91 stream_test_sou
event rate =10407 Hz. data rate =833142.7919998887 kB/s. missed rate = 0 Hz.
event rate =10225 Hz. data rate =818572.599999892 kB/s. missed rate = 0 Hz.
event rate =10077 Hz. data rate =806724.3119998946 kB/s. missed rate = 0 Hz.
event rate =10372 Hz. data rate =830340.8319998893 kB/s. missed rate = 0 Hz.
event rate =10276 Hz. data rate =822655.455999891 kB/s. missed rate = 0 Hz.
event rate =10370 Hz. data rate =830180.7199998894 kB/s. missed rate = 0 Hz.
event rate =10277 Hz. data rate =822735.511999891 kB/s. missed rate = 0 Hz.
event rate =10200 Hz. data rate =816571.1999998924 kB/s. missed rate = 0 Hz.
event rate =10205 Hz. data rate =816971.4799998923 kB/s. missed rate = 0 Hz.
event rate =9782 Hz. data rate =783107.7919998998 kB/s. missed rate = 0 Hz.
event rate =10223 Hz. data rate =818412.487999892 kB/s. missed rate = 0 Hz.
event rate =10097 Hz. data rate =808325.4319998942 kB/s. missed rate = 0 Hz.
event rate =10364 Hz. data rate =829700.3839998895 kB/s. missed rate = 0 Hz.
event rate =10366 Hz. data rate =829860.4959998894 kB/s. missed rate = 0 Hz.
event rate =9538 Hz. data rate =763574.1279999041 kB/s. missed rate = 0 Hz.
event rate =10211 Hz. data rate =817451.8159998922 kB/s. missed rate = 0 Hz.
event rate =10389 Hz. data rate =831781.839999889 kB/s. missed rate = 0 Hz.
event rate =10256 Hz. data rate =821054.3359998914 kB/s. missed rate = 0 Hz.
event rate =10296 Hz. data rate =824256.5759998907 kB/s. missed rate = 0 Hz.
event rate =10224 Hz. data rate =818492.543999892 kB/s. missed rate = 0 Hz.
event rate =10455 Hz. data rate =836985.4799998879 kB/s. missed rate = 0 Hz.
event rate =10546 Hz. data rate =844270.5759998863 kB/s. missed rate = 0 Hz.
event rate =10381 Hz. data rate =831061.3359998892 kB/s. missed rate = 0 Hz.
event rate =10137 Hz. data rate =811527.6719998935 kB/s. missed rate = 0 Hz.
event rate =10112 Hz. data rate =809526.271999894 kB/s. missed rate = 0 Hz.
event rate =9652 Hz. data rate =772780.5679999021 kB/s. missed rate = 0 Hz.
event rate =9914 Hz. data rate =793755.2399998974 kB/s. missed rate = 0 Hz.
event rate =10313 Hz. data rate =825617.5279998904 kB/s. missed rate = 0 Hz.
event rate =10298 Hz. data rate =824336.6319998907 kB/s. missed rate = 0 Hz.
event rate =10337 Hz. data rate =827618.92799989 kB/s. missed rate = 0 Hz.
event rate =10449 Hz. data rate =836505.143999888 kB/s. missed rate = 0 Hz.
event rate =10797 Hz. data rate =864524.7439998818 kB/s. missed rate = 0 Hz.
event rate =10402 Hz. data rate =832742.5119998888 kB/s. missed rate = 0 Hz.
event rate =9946 Hz. data rate =796317.0319998969 kB/s. missed rate = 0 Hz.
event rate =10445 Hz. data rate =836264.975999888 kB/s. missed rate = 0 Hz.
^Cindra-s3.jlab.org>
event rate =10427 Hz. data rate =834743.9119998884 kB/s. missed rate = 0 Hz.
event rate =10258 Hz. data rate =821214.4479998914 kB/s. missed rate = 0 Hz.
event rate =10299 Hz. data rate =824496.7439998906 kB/s. missed rate = 0 Hz.
event rate =10507 Hz. data rate =841148.391999887 kB/s. missed rate = 0 Hz.
event rate =10142 Hz. data rate =811927.9519998934 kB/s. missed rate = 0 Hz.
event rate =10451 Hz. data rate =836665.255999888 kB/s. missed rate = 0 Hz.
event rate =10309 Hz. data rate =825297.3039998905 kB/s. missed rate = 0 Hz.
event rate =10396 Hz. data rate =832262.1759998889 kB/s. missed rate = 0 Hz.
event rate =10273 Hz. data rate =822415.2879998911 kB/s. missed rate = 0 Hz.
event rate =10279 Hz. data rate =822895.623999891 kB/s. missed rate = 0 Hz.

Summary

- Developed and benchmarked ERSAP data-lake prototype.
- Demonstrated ~10GByte/sec lossless data throughput in and out of the data-lake.
- Work in progress

Thank you