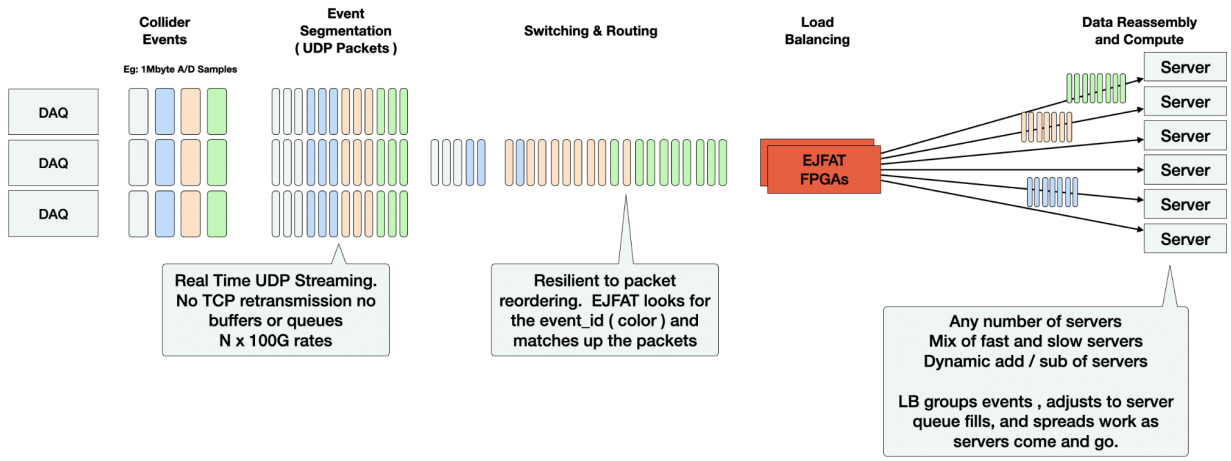
Getting Started With EJ-FAT

Introduction :

EJ-FAT (ESnet / JLab - FPGA Accelerated Transport) is a packet protocol, and implementation for transferring High Speed streaming data from Data Acquisition Systems (DAQ's) to compute facilities. EJ-FAT is generally applicable to DOE workflows that work in this mode, without any experiment specific assumptions.

Design Principles :

1. Load Balancing



Coherent grouping of events:

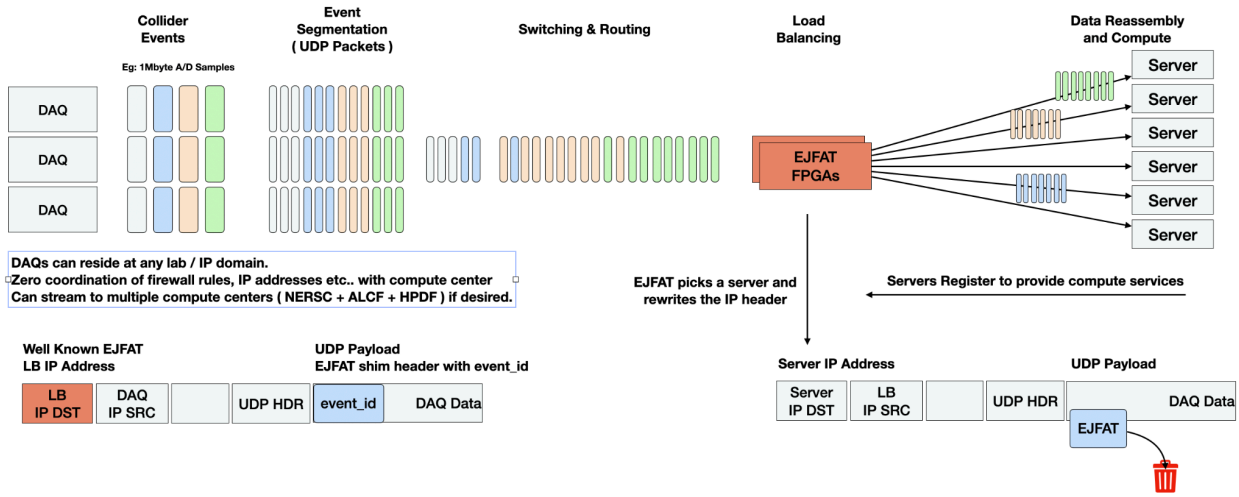
The primary purpose of the load balancer is to distribute physics event data to several compute nodes, in order to process events in parallel on a cluster of servers. DAQ systems provide the event Id as part of a special header field incorporated just after the UDP header. The load balancer parses this field and sends all events with the same event id to a single unique server that it selects from the pool of registered servers.

Dynamic assignment of work to Compute Nodes

A common scenario with real time load balancing is that servers finish work at slightly different rates. Over time the load balancer needs to increase or decrease the amount of work it assigns to a server, based on its pending work queue. The EJ-FAT load balancer receives queue updates at 1s intervals from all registered servers and updates the number of events it sends to a given server, such that its work queue does not overflow. If a server stops sending updates, the load balancer removes the server from the registry and no longer sends packets to

it. The queue status messages are currently implemented as outbound gRPC messages over TCP from the compute nodes to the load balancer.

2. Separation of IP Domains



- Pull Model For Compute
    - Compute nodes register and authenticate with EJFAT control plane (TCP)
    - EJFAT dataplane sends load balanced events to registered compute nodes (UDP)
    - Compute nodes must provide continuous (every 1 second) liveness reports (TCP) back to the control plane in order to continue receiving data from the load balancer.
- NAT
    - DAQs (data sources) send UDP packets directly to the EJFAT load balancer
    - The EJFAT load balancer rewrites the IP headers such that the UDP output packets have a source IP of the load balancer and destination IP of the selected compute node.
- Clearly established domain boundaries
    - DAQ sources only communicate with the load balancer
    - Compute Nodes only communicate with the load balancer

Implementation :

EJ-FAT is working jointly with the IRI Testbed and HPDF Hub Facility to provide a hosted solution for quick and simple pathfinding and testing exercises by different scientific workflows. Data originators that operate DAQ systems, can establish an end to end workflow on existing compute facilities. Similarly operators of compute clusters can host and integrate experimental workflows, without needing to have detailed discussions with DAQ operators. ESnet provides the FPGA based hardware and an operating Load Balancer in the network, eliminating the need

for either entity to purchase new EJ-FAT hardware, or become involved in the operation of an EJ-FAT forwarding node.

QoS Aware BGP Peering

Is implemented using Virtual Private Routed Network (VPRN) / L3VPN technology on ESnet's Nokia routers. The VPRN is terminated at the ESnet router that peers with the campus border router (BR). If it is deemed useful for the campus network the peering with the VPRN may be extended to a router inside the campus network.

The benefit of the VPRN technology, rather than the general routing table (GRT), is that ESnet is able to traffic engineer the UDP traffic. This will be essential for lossless streaming between facilities. A site may participate in EJ-FAT by establishing a BGP peering with ESnet that will land in the EJ-FAT VPRN. A peering into the EJ-FAT VPRN will provide the site with routes to the load balancer, DAQ's, and Compute at participating sites.