Big Data JET Collaboration – BiDJET Collaboration

This is intended to be a possible series of JET projects in the large scale data transfer space. The intent is to facilitate collaboration between JET networks and their attached Big Data facilities.

Year 1: Enumeration phase

- Locate big data facilities, repositories, and users connected to JET networks, share that information with other JET networks
    - Location
    - Bandwidth
    - Contact information
- Metrics
    - Each JET network will enumerate at least five facilities, or indicate that fewer than five such facilities are connected to that network
    - Information will be distributed to other JET members

Year 2: Characterization phase

- Characterize network and site performance
    - Deploy perfSONAR near three big data sites
    - Characterize performance (set up regular testing)
    - Compare with data download performance to hosts on a different JET network
- Metrics
    - Each JET network will deploy perfSONAR at the edge location as close as reasonably possible to the connection to three Big Data sites
    - Regular perfSONAR tests will be set up between the three Big Data site perfSONAR nodes and perfSONAR nodes in another JET network

Year 3: Capability phase

- Improve or sustain network performance for Big Data sites
    - Using perfSONAR data, maintain or enhance data transfer performance for users, facilities, or data repositories
    - Work with facility, repository, or user site personnel to ensure data transfer performance is scientifically useful
    - Work with sites on Science DMZ architecture for Data Transfer Nodes if applicable
- Metrics
    - Data transfer performance statistics for three sites, including performance improvement numbers (before and after performance tuning and architecture improvements) if applicable