

Give your Data the Edge: A Scalable Data Delivery Platform

Larry Peterson (PI)

NSF 1541318

Goals

- Tap multiple datasets
- Leverage commodity storage
- Scale with commodity CDNs
- Span multiple sites

Challenges

- Scientists cannot become distributed storage experts
- Unsustainable “**glue code**” binds workflows together
- End-to-end data guarantees

Solution

- Wide-area read/write volumes
- Scalable, programmable *storage fabric*
- Reusable, composable drivers
- Automatic key management

Deployment

- Compute: OpenCloud, TACC
- Storage: Amazon S3
- Datasets: iRODS users

Syndicate Gateways (SGs) give a host access to mountable volumes, and run PI-supplied drivers to apply the desired data guarantees.

Advancements

- **User-defined properties**
 - Consistency
 - Confidentiality
 - Authenticity
 - I/O side-effects
- **Sustainable compatibility**
 - Write once, use anywhere
 - Compose simple drivers to achieve complex behavior
 - Live driver updates
 - Automatic deployment
- **Minimize operations**
 - Declarative configuration
 - Automatic reconfiguration
 - Self-healing fabric

