



globus online



Delivering a scalable service

Steve Tuecke

Computation Institute

University of Chicago and Argonne National Laboratory



What is Globus Online?

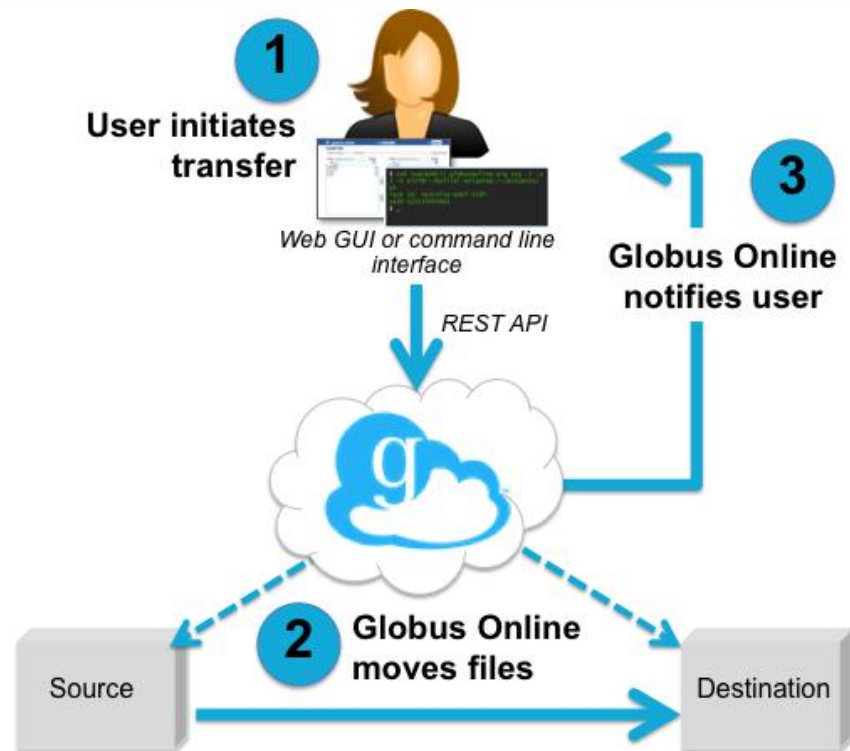
- **Reliable file transfer.**

- Easy “fire-and-forget” transfers
- Automatic fault recovery
- High performance
- Across multiple security domains

- **No IT required.**

- Software as a Service (SaaS)
 - No client software installation
 - New features automatically available
- Consolidated support & troubleshooting
- Works with existing GridFTP servers
- Globus Connect solves “last mile problem”

- **Supports Blue Waters, XSEDE, NERSC, ALCF, and many universities**





Software as a Service (SaaS) vs Traditional software delivery

- **SaaS changes assumptions and approach throughout the software lifecycle**
 - Architecture and design
 - Designed for specific environment
 - Software development
 - No porting. Focus on functionality.
 - Operations
 - Nobody else will operate
 - Focus on availability, automation, monitoring
 - Support
 - Tightly integrated with operations
- **We are delivering a *service*, not *software***



- **Product management**
- **Product development**
 - Developer-operators (dev-ops)
 - User eXperience (UX) manager
 - Web design and development
- **User services**
 - Help desk / support
 - Consulting services
- **Marketing**





- **“Continuous” service updates**
 - Globus Online updates almost every week
 - And hot fixes for critical issues
- **Independent updates of component services**
 - Nexus, Transfer (backend, CLI, REST, relay, history), Web GUI, Storage, sample endpoints, ...
- **Use Agile Scrum**
 - Backlog
 - Time-boxed development (sprints)
 - Scrums
 - Sprint reviews



Production environment

- **Uses Amazon Web Services (AWS)**
 - EC2, EBS, S3, ELB, ...
- **Many EC2 instances**
 - Each service running on 1 or more instances
 - Replication across availability zones within region
 - All services within an Amazon security group
 - Backups to S3 in another region
- **Operations services**
 - Chef based automated deployment
 - Logging to common server (rsyslog, logstash, etc.)
 - Nagios monitoring
 - OSSEC host-based intrusion detection
- **Access limited to “need to have”**
- **Zendesk based help desk w/ Globus Online user SSO**



Development and test environments

- **Dev → Test → Integration → Staging → Production**
 - Dev: AWS and laptops
 - Test: Multiple (partial) test instances on AWS (branches)
 - Integration: Full copy of production with next code to be released on AWS (trunk)
 - Staging: Full copy of production on AWS, to test updates
 - Production: AWS
- **GitHub repositories**
- **Jira w/ GreenHopper for Scrum management**
- **Python is primary development language, using**
 - PostgreSQL and Cassandra databases
 - Globus Toolkit C libraries
 - Many open source Python libraries



Platform as a Service

- **How can we enable other groups to enhance the Globus Online ecosystem without replicating everything we have done?**
- **Globus Integrate platform**
 - Globus Nexus: identity, group, and profile service
 - REST APIs to services
- **Don't constrain your implementation and hosting approaches**
 - Java, Python, Ruby, etc.
 - AWS, Google App Engine, Liferay, etc.