DRAS-TIC Fedora

Distributed Linked Data Platform w/Memento

Gregory N. Jansen
University of Maryland iSchool
IEEE Big Data 2017:: 2nd Computational Archival Science (CAS) Workshop
Boston, 13th December 2017
DRAS-TIC + Fedora

Digital Repository at Scale that Invites Computation
Cassandra NoSQL Database
CDMI & WebDAV APIs
Stateless Servers
Open to Spark, etc..

Fedora 4 Digital Repository Software
Fedora 5 is an API Specification:
Linked Data Platform
Memento
Hierarchal resources described in formal LD graphs
IMLS Grant: 2 Years of Challenges

Sustained, community-developed software

“Unlimited” horizontal scaling of capacity and performance

Schema that facilitates compute on Cassandra side

Support for Linked Data Platform / Memento

Auxiliary or convenience APIs, such as CDMI and WebDAV
Community Partners

Smithsonian Institution
University of Illinois at Urbana-Champaign
Georgetown University
University of Maryland Libraries
Test-driven Approach

Many trial schemas, software, and cluster configurations

Test battery of multi-user simulations based on partner use cases

Real data from ourselves and partners

Capture of performance metrics - RPS, request duration

Capture of cluster load metrics - i/o, cpu, memory

Unified, comparative visualization of results
DRAS-TIC Fedora Testbed

fedora-testbed
- Flask Web Portal
- Kibana and Grafana
- Elasticsearch client-only node
- Proxy for DRAS-TIC (ciber)
- FTP-over-HTTP hosted CI-BER files

logstash
- Test scheduler / runner
- Gatling test scenarios

gatling
- Gatling test scenarios

4n Elasticsearch Cluster
- ciber.umd.edu metadata
- CI-BER files inventory
- Gatling metrics
- DF node metrics

Docker Swarm

?n drastic-web cluster (test)
?n Cassandra Cluster (test)
Thank You

https://umd-drastic.github.io/

Questions?