Brookhaven National Laboratory

U.S. DEPARTMENT OF ENERGY

IT Services

Jason Smith <smithj4@bnl.gov>

June 27, 2022

@BrookhavenLab

Overview of the IT Services Group

- Group is composed of 8 staff members (7 partially funded by ATLAS)
  - FY23 (total: 2.05 FTE): J. De Stefano(0.15), J. Frith(0.25), S. Kandasamy(0.2), Z. Liu(0.3), L. Pelosi(0.4), T. Rao(0.45), J. Smith(0.3)
    - Benefit from shared staff knowledge & expertise, and services provided
- Responsibilities include
  - Central storage systems:
    - NFS (home directories)
    - Data storage (GPFS & Lustre)
    - CVMFS: repositories, caches, proxies & servers
  - General IT Services:
    - Virtualization systems (RHEV & OKD)
    - Gateways and file & data transfer services (ssh, sftp & Globus Online)
    - User accounts, VO management & IAA systems (IPA, Keycloak, PrivacyIDEA)
    - Puppet (automated centralized config mgmt) & other infrastructure (ntp, dns, syslog, mail, etc)
    - Web services (RT, Mattermost, USATLAS.org Drupal, Discourse) & NX GUI login
    - Monitoring: Nagios (non-cluster), ELK, Grafana (servers & storage), Ganglia (legacy)

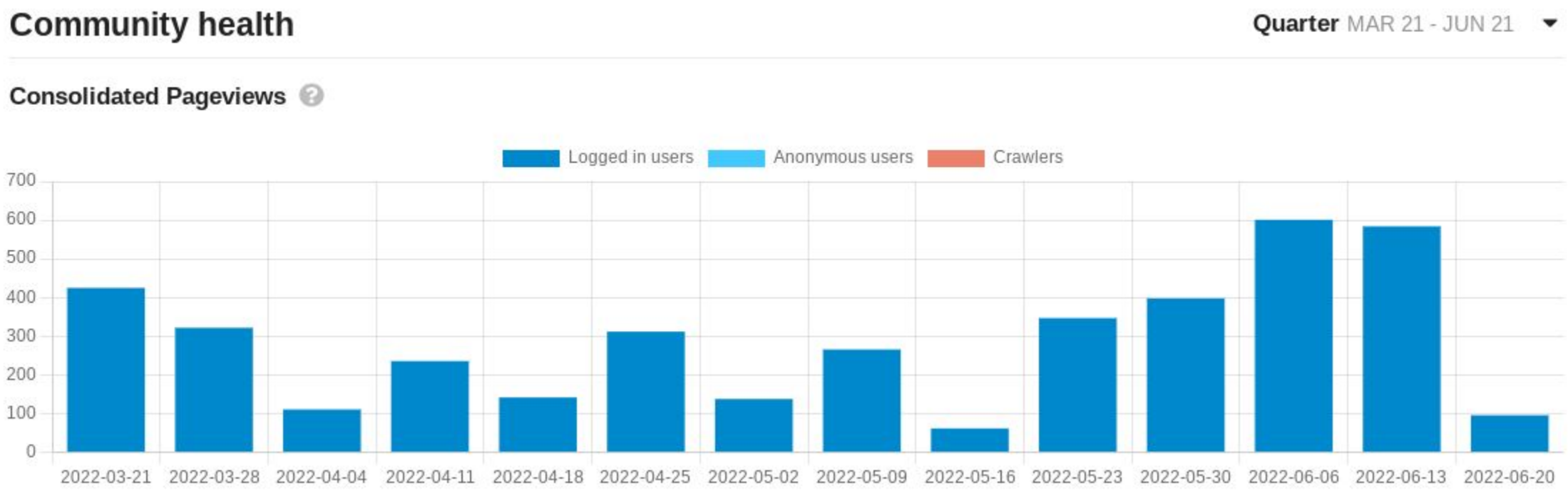
Achievements for ATLAS over last year

- USATLAS.org Drupal website completed (Pelosi and T&S group)
- RHEV infrastructure updated (hardware & software) and migrated to new data center in B725 (Frith, Rao & others)
- Globus Online updated to use CILogon AuthN (Rao)
- Continuous Mattermost upgrades & support (Rao)
  - Used for "US Cloud dCache Admins", "US Cloud Site Admins" and "US Federated Operations" support channels
- USATLAS HPC Rucio storage [BNLHPC_DATADISK] (Frith, Liu & Rao)
- OKD cluster setup for container management (Rao & Hollowell)
- Trained new staff on dCache operations & support (Liu)
- ENDIT HSM adapted and commissioned to work with BNL HPSS tape system in time for the ATLAS tape challenge. (Liu)
  - Greatly improved the performance and the scalability of BNL's dCache.
- New Discourse server used for Analysis Facility support (Rao)
- User accounts management for Federated Jupyter (Kandasamy)

Federated Identity

- Federated JupyterHub
  - Coordination across teams & Cyber Security for policies imp. (*Lauret*)
  - Collaborated with the Fabric group on implementation
  - User account management, procedures & local unix account mapping
  - See the IT Fabric presentation by Chris H. for more info on JupyterHub
- Discourse server configured to allow (non-BNL) Federated logins
- USATLAS.org Drupal web server (CERN & SDCC)
- Testing Federated login integration to Mattermost now

Discourse Server Usage Stats



Central Storage Systems

- NFS home directories (Frith & Rao)
  - NetApp A300 with 45TB usable SSD storage shared among all exp.
  - 25GB quota per user
- GPFS for ATLAS (Frith & Rao)
  - 3 servers with 300TB usable (x2 replicated) capacity
- Lustre for HPC ATLAS (Frith, Liu & Rao)
  - 3 servers+jbods with 3PB capacity
- CVMFS (De Stefano)
  - 2 server pairs
    - Internal caches for local jobs (used by local ATLAS jobs)
    - External reverse proxies for everyone to use (used by all ATLAS sites world-wide)
  - 1 stratum-0 server and 1 stratum-1 server
    - Hosting local repo for local running jobs
    - Stratum-0 disk usage: 150 GB in 2 repos, Stratum-1 disk usage: 8.1 TB in 4 repos

General IT Services

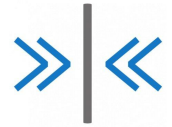
- Virtualization systems (Frith & Rao)
  - Shared RHEV: Upgraded (latest version) & migrated to new Data Center
    - 8 prod + 4 test hypervisors, RHEL8, 18TB RAM, 200TB storage: ~300+ VMs total
  - OKD: OS RedHat OpenShift - new cluster setup for container mgmt.
    - 7-nodes, version 4.10, 180 (sys+user) containers, 15TB NVMe disk, ~1TB RAM
- Gateways, file & data transfer (shared)
  - 4 ssh gateways
  - 2 sftp servers (small file transfers)
  - Globus Online (data transfers)
- User account management (Kandasamy)
  - ~250 active USATLAS accounts
  - Account creation (including new federated account policy)
  - Password resets and other general user access support
  - User account auditing & compliance

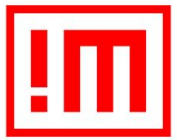
General IT Services (cont.)

- IAA (Identity, AuthN & AuthZ) systems
  - IPA (12 RedHat IPA servers - version 4.6.8)
  - Keycloak (Web AuthN - version 7)
  - PrivacyIDEA (modular AuthN server - version: 3.4.1)
    - Used as our MFA/TOTP backend store for Keycloak & IPA
- ATLAS VO management (De Stefano & Kandasamy)
  - Assisting CERN with all ATLAS VO user, group, permission requests
- Puppet (recently upgraded to support RHEL >= 8) (Rao)
  - Cobbler, Gitea repo, Puppet 7.x server & Foreman 3.4 webui
- Other shared infrastructure services
  - 3 gps ntp, 3 dns, 2 syslog, etc…

General IT Services (cont.)

- Web services
  - RT (Request Tracker) user support system (Smith)
  - Mattermost: usually latest LTS version (Rao)
  - NX version 7 (Kandasamy)
    - 1 gateway server allows NX client or web browser access (MFA required)
    - 4+1 terminal server nodes (RHEL7)
      - Testing upgrade to RHEL8 desktop (GNOME & KDE) now
  - Discourse: SDCC ATLAS Talk: Analysis Facility (Rao)
    - Federated logins via CILogon authorization from only a few sites
  - USATLAS.org website (developed by DataArt, Pelosi and T&S)
    - Federated logins (SDCC & CERN)





General IT Services (cont.)

- Monitoring (shared with all experiments)
  - Nagios version 4.4.6 (non-cluster)
  - ELK - version 7.16.2 (with Tools & Services)
    - Currently Open Source, evaluating enterprise version
  - Grafana (hosted by the IT Fabric group)
    - Server host & storage performance metrics
  - Ganglia (legacy host performance metrics)

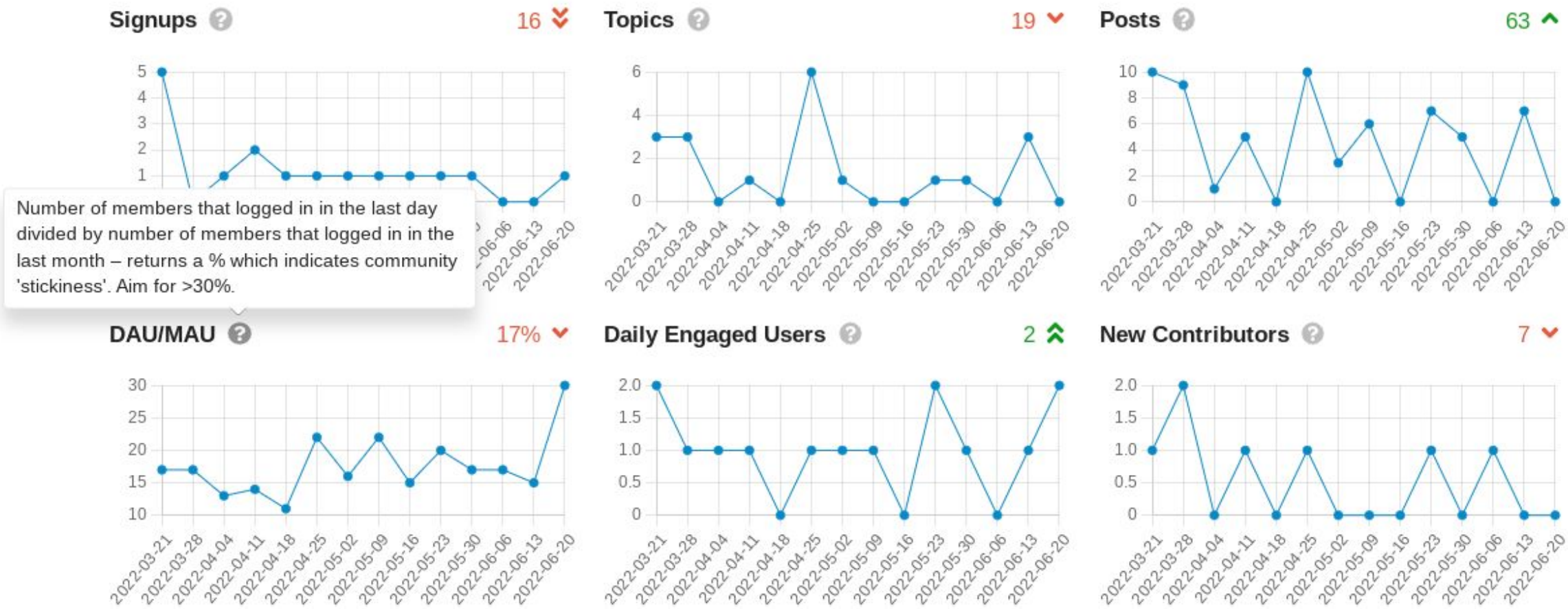
ATLAS Challenges

- Getting more difficult to compete with other IT jobs in the market
  - The pandemic exacerbated this due to the quick evolution to hybrid and full remote work opportunities available at most places now
  - Private sector more flexible with workplace arrangements + the promise of better salaries (while BNL has uncertain budget futures) causes a “skill leak” - this is a risk.
- The x509 to SciToken AuthN transition has been challenging
  - OSG pushed hard for transition without providing clear path forward
    - OSG itself delayed its 3.6 release before dropping x509 support
    - WLCG, ATLAS decided on different token provider solution (IAM)
      - Critical service, support levels still unclear

Conclusions

- 2 FTE only provide support for many ATLAS services
  - This is an example of benefit driven by running a shared facility - Full / complete service are provided by using shared expertise & resources
  - The required base facility infrastructure is already provided at low cost due to the SDCC being shared

- ATLAS specific services provided:
  - Mattermost (shared service, with dedicated support channels provided)
  - Discourse (dedicated service)
  - Storage:
    - GPFS (dedicated)
    - Lustre (dedicated)
    - CVMFS (shared with dedicated repos)

Backup Slides

Discourse Server Usage Stats (cont.)