XSEDE Gateway attributes submission
An overview of new reporting method

Amit Chourasia, SDSC/UCSD

https://xsede-xdcdb-api.xsede.org/api/gateways

Deadline to make the switch: Nov 30, 2019

Slides: https://bit.ly/2kRPsaB

Agenda

Presentation (30 mins)

- Background

- Importance of gateway attributes

- Submission details

- Best practices

- Support information

Discussion (30 mins)

Background

Gateway attributes provide complimentary information about jobs that were run under a community account on XSEDE

- Report for whom the job was actually run (aka gateway user)

XSEDE mandates Gateways to report this information

SP = Service Provider i.e. site hosting the cluster e.g. psc, sdsc, tacc

Background: Attribute compilation

Service Providers report job info

- **Job ID – Slurm/PBS**
- User name - Community account
- Usage – Su's consumed
- Allocation
- etc

Gateways report job attributes

- **Job ID – Slurm/PBS**
- Gateway name
- Gateway user name
- Job submit time

XDCDB
Combines info received from SP & Gateway attributes based on **Job ID**

Slides: https://bit.ly/2kRPsaB

Importance of Gateway attributes

For Gateways

- Better visibility of SU burn rate
- View, report and share usage data via XDMoD interface with Sponsor

For each Service Provider (SP) aka site hosting the cluster

- Number of Gateways users using a system
- Usage by individual gateway users
- Better understanding of how Gateways are using the cluster
- Report this information to sponsor

For XSEDE

- Same as for SPs
- Better handle on how various clusters are being used to meet community needs
- Visibility into the breadth of users using XSEDE

Slides: https://bit.ly/2kRPsaB

XSEDE user Trend

XSEDE User Census
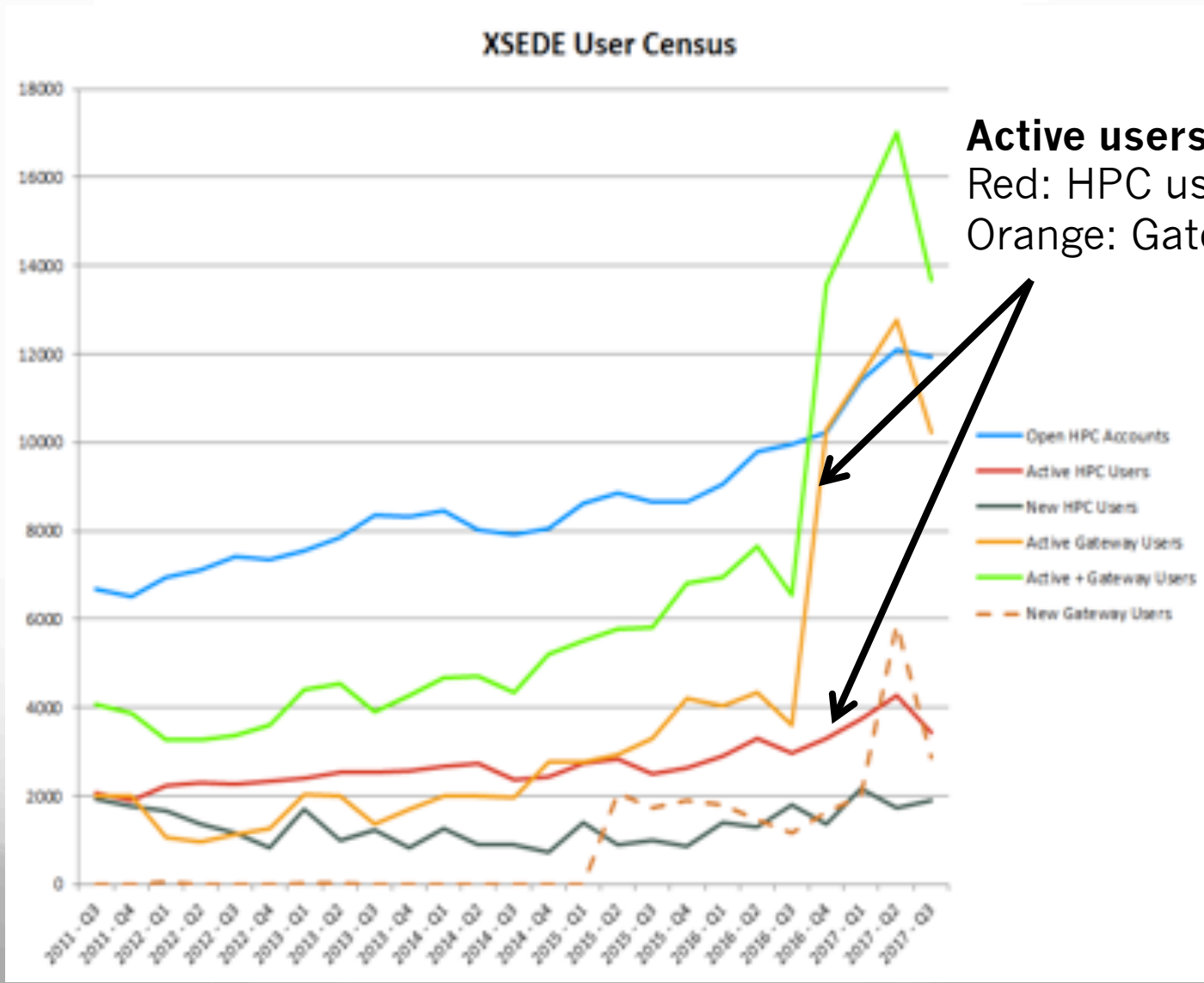
Active users
Red: HPC users
Orange: Gateway users

Open HPC Accounts
Active HPC Users
New HPC Users
Active Gateway Users
Active + Gateway Users
New Gateway Users

Deficiencies in existing method

Gateway Operators

- Submission methods vary between SPs, and Gateway operators must know this

- Can only submit attributes from login nodes only

- Issues with data validation and consistency checks

- No feedback to Gateway operator on submission

- Cannot check prior submission (submit a ticket)

XSEDE Service Providers

Must act as brokers and enable submission via custom scripts that they
- Install
- Manage
- Document
- No ability to audit who submitted the attributes

Other legacy issues

Slides: https://bit.ly/2kRPsaB

New submission method
(Gateway operator friendly)

Goal: Cater well to the needs of the gateway operators

- Self service

- Same submission method for any XSEDE site

- Single documentation spot

- Semantic arguments for submission

- Origin agnostic submission i.e. submit from anywhere, not just XSEDE login nodes.

- Tool agnostic submission

Slides: https://bit.ly/2kRPsaB

Gateway operator benefits

- Consistent & portable way to report across all XSEDE resources.

- Consistent documentation.

- New ability to report software(s) used.

- View and share reporting of above data via XDMoD interface with Sponsor.

Submission recipe

One time setup

Generate APIKey for your gateway

Recurring

For each job submit gateway attributes with your APIKey

One time setup

Generate APIKey for your gateway as noted at

https://xsede-xdcdb-api.xsede.org/api/gateways

- Preference order of XSEDE account to generate APIKey
  1. Community user account associated with the gateway
  2. Alternative account used to administer multiple gateways

- Choose a sound name for your Gateway.
  - Once name is set it cannot be changed.

- Save your APIKey to a file immediately
  - This can be only retrieved once
  - Reset existing to get a new one

List of attributes

1. **XSEDE resource name**

2. **Job ID**

3. **Submit time**

4. **Gateway user name**

5. **Software (optional, but recommended)**

6. debug (optional, useful for testing)

7. Usage (optional, for future use with JetStream)

8. VMID (optional, for future use with JetStream)

Slides: https://bit.ly/2kRPsaB

List resource names

Get a list of valid resource names as follows

$ curl -XPOST \

    --data @$HOME/.xsede-gateway-attributes-apikey \

    --data-urlencode "xsederesourcename=unknown" \

https://xsede-xdcdb-api.xsede.org/gateway/v2/job_attributes

acf.nics.xsede

bridges-gpu-ai.psc.xsede

...

Required attributes

xsederesourcename: MUST match to available list on the server

jobid: MUST be the ID assigned by the XSEDE

Service Provider's job scheduler

submittime: Approximate time when the job was submitted

to the cluster

gatewayuser: Unique identifier for your actual gateway user.

i.e the gateway user for whom the job was run

e.g. username, other unique ID or email if permitted

This is NOT Gateway's XSEDE community account

MUST NOT switch identifier method at a later time

Slides: https://bit.ly/2kRPsaB

Optional attributes

software: Multiple attributes can be submitted as comma separated list. MUST be in specified format.

    Key:Value

    Generic_name:Name_with_version

    MUST be official names as used by the software

    Examples

    MATLAB:MATLAB R2019a

    NumPy:NumPy v1.17

debug: Useful for initial testing. Submissions including this attribute are ephemeral (not stored).

usage: Relevant for JetStream only (future)

vmid: Relevant for JetStream only (future)

Slides: https://bit.ly/2kRPsaB

Test submissions with debug flag

```
$ curl -XPOST --data @$HOME/.xsede-gateway-attributes-apikey  \
  --data-urlencode "gatewayuser=John Doe"  \
  --data-urlencode "xsederesourcename=comet-gpu.sdsc.xsede"  \
  --data-urlencode "jobid=bogus"  \
  --data-urlencode "submittime=2019-01-31 23:45 PST" \
  --data-urlencode "debug=x"  \
  https://xsede-xdcdb-api.xsede.org/gateway/v2/job_attributes
```

debug parameter can have any value

Sample response

Job: bogus
Job summited: 2019-01-31 23:45:00-0800
Resource Name: comet-gpu.sdsc.xsede
Gateway Name: Tesseract
Gateway User: John Doe
Debug specified. Attributes not staged

Submission with debug are not stored

Real submission example

```
$ curl -XPOST --data @$HOME/.xsede-gateway-attributes-apikey \
  --data-urlencode "gatewayuser=John Doe" \
  --data-urlencode "xsederesourcename=comet-gpu.sdsc.xsede" \
  --data-urlencode "jobid=12345" \
  --data-urlencode "submittime=2019-01-31 23:45 PST" \
  --data-urlencode "software=MATLAB:MATLAB R2019a, NumPy:NumPy v1.17" \
  https://xsede-xdcdb-api.xsede.org/gateway/v2/job_attributes
```

Sample response

Job: 12345
Job summited: 2019-01-31 23:45:00-0800
Resource Name: comet-gpu.sdsc.xsede
Gateway Name: Tesseract
Gateway User: John Doe
Software: MATLAB:MATLAB R2019a, NumPy:NumPy v1.17

Review submitted attributes (anytime)

**$ curl -XPOST **
  **--data @$HOME/.xsede-gateway-attributes-apikey **
  https://xsede-xdcdb-api.xsede.org/gateway/v2/jobs

This url is different than one for submitting attributes

Returns last 10 submitted Gateway Attributes

The url above is different from one used for submission

Review submitted attributes
(after 24 hours)

$ xdusage -j -ja -s 2019-05-31 -e 2019-06-01 -p TG-STA110011S

. . .

job id=4939827 resource=comet.sdsc.xsede

submit=2019-05-31@14:32:28 start=2019-05-31@17:47:28

end=2015-05-31@17:57:37 nodecount=2 processors=32 queue=normal charge=24.89

job-attr id= 4939827 name=gateway-user value=John Doe

Supplied by you

This will only work when SPs send their piece of the information.
SPs typically report on a nightly basis.

Best practices
(Things to do)

1. APIKey MUST be handled securely

2. Store APIKey in a file and restrict its access to minimal permissions

3. If you operate more than one Gateway. Generate APIKey for each gateway, then report attributes for each of them separately

4. Incorporate mechanism to retry failed attribute submission

5. Report software attributes and stick to prescribed format

6. Use debug attribute for testing

Best practices
(Things not to do)

1. DO NOT generate multiple APIKey for the same Gateway

2. DO NOT use personal or non-community account to generate APIKey for a Gateway

3. APIKey MUST NOT be visible in command line based submissions
    curl -XPOST --data-urlencode "apikey=example-123456abcd" ← This is insecure

    See correct usage in the documentation

4. gatewayuser attribute MUST NOT switch identifier scheme that was selected initially. e.g. If username was chosen as a unique identifier initially, it should not be changed to something else later (e.g. email)

Requesting help

Submit a ticket to help@xsede.org

Include the following in your message

- "Gateway Attribute Submission"

- Server URL: https://xsede-xdcdb-api.xsede.org

Support until deadline Nov 30, 2019 deadline.

Reach out to me via email: amit _ AT _ sdsc.edu

Include same message as above

FAQs

- JetStream is not supported

- There is no bulk submission method

- Resubmitting duplicate entries will overwrite previous ones

- Stray entries that do not match with official XSEDE Service Providers JobID's are purged after few weeks.

Tasks ahead

Switch your Gateway to new submission method by Nov 30, 2019

Reference

Amit Chourasia, Scott Sakai, Michael Shapiro, and Steven Gallo. 2019. Improving Science Gateways usage reporting for XSEDE. In Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (learning) (PEARC '19). ACM, New York, NY, USA, Article 109, 4 pages. DOI: https://doi.org/10.1145/3332186.3338096

Slides: https://bit.ly/2kRPsaB

Acknowledgements

- Nancy Wilkins-Diehr (project initiator), Marlon Pierce, Robert Sinkovits

- (Alphabetic) Christopher Irving, David Hart, Eric Coulter, Eroma Abheysinghe, Ester Soriano, Lee Leming, Mahidhar Tatineni, Maytal Dahan, Nicole Wolter, Sudhakar Pamidighantam, Victor Hazelwood

- XCI Team: JP Navarro, Shava Smallen, Choonhan Yoon, Jim Basney and other reviewers

Slides: https://bit.ly/2kRPsaB