Course information is gathered from the [XSEDE Training Course Google spreadsheet](https://docs.google.com/spreadsheets/d/1pBvJKRBug9ivIAutF62JkJVhR23rYlg0yb-b_QXfxOo/edit?ts=5d81281b#gid=0) and the XSEDE Portal database. Information is written to tables in the XSEDE Warehouse database as “Type=Training”.

There are 2 types of records that will be created in the warehouse. The first is generic records for a training class that includes generic information about the class. The second is specific sessions of the class including where the session will be held and when it is being held. Multiple sessions of a training class can be held concurrently at different locations.

Duplicate entries from the sources are combined into one warehouse entry with the newest entries being used. Missing information from the entry will be gathered from the older, duplicate entries, where they exist. Entries in the spreadsheet will take precedence over entries from the Portal database.

**Transferring data from the spreadsheet**

The spreadsheet does not contain information about individual sessions of a training class. It has no start/end times or locations for individual sessions of a class. These records will be stored in the warehouse as generic training class records. In many cases the spreadsheet data will be duplicated in the portal DB data.

For each row in the spreadsheet, a record is created in the glue2.resource\_v3\_resourcev3local table:

* A GlobalID value is generated with the row number of the spreadsheet entry. It is stored as ID. urn:ogf:glue2:info.xsede.org:resource:google:XSEDE\_Training\_Course:{row}
* The current time is stored as CreationTime.
* Validity is an interval set to 14 days.
* Affiliation is set to xsede.org.
* The row number in the spreadsheet is stored as LocalID.
* LocalType is set to “Training”.
* The spreadsheet’s url value is stored as LocalURL.
* CatalogMetaURL
* A JSON entry is generated from all of the information in each row in the spreadsheet. It is stored as EntityJSON.

A record is also created in the glue2.resource\_v3\_resourcev3 table:

* The ID, Affiliation, and LocalID used above are stored in the values of the same name.
* Type is set to “Training”
* QualityLevel is set to “Production”.
* The spreadsheet’s “Title” value is stored as Name.
* If the spreadsheet’s “Delivery” value is “IN-PERSON” or “MULTICAST”, the “ResourceGroup” field is set to “Live Events”. Otherwise, it is set to “Streamed Events”.
* The spreadsheet’s “Title” and “Length” values in the spreadsheet are combined to create the “Short Description” entry in the database.
* If the provider already exists in the resource\_v3\_resourcev3 table its ID will be stored as the ProviderID for the record. Otherwise a new record will be added to both the resource\_v3\_resourcev3 table and to the resource\_v3\_resourcev3local table and the newly generated ID will be stored as the ProviderID.
  + In the resource\_v3\_resourcev3local table:
    - A GlobalID value is generated (using the row number for the entry in the spreadsheet) for the new provider and stored as ID. urn:ogf:glue2:info.xsede.org:resource:google:provider:{row}
    - The current time is stored as CreationTime.
    - Validity is an interval set to 14 days.
    - Affiliation is set to xsede.org.
    - The row number in the spreadsheet is stored as LocalID.
    - LocalType is set to “provider”.
    - LocalURL and CatalogMetaURL are set to NULL.
    - The Center and Logo spreadsheet values are stored in EntityJSON.
  + In the resource\_ve\_resourcev3 table:
    - The ID, Affiliation, and LocalID used above are stored in the values of the same name.
    - QualityLevel is set to “Production”.
    - The spreadsheet’s Center value is stored as the Name.
    - ResourceGroup is set to “Organizations”.
    - Type is set to “Provider”.
    - ShortDescription, Description, Topics, Keywords, and ProviderID are set to NULL.
    - Audience is set to the same value as Affiliation.
    - StartDateTime and EndDateTime are set to NULL.
* The “Description” value is stored as Description.
* The “Category” value is stored in the “Topics” field.
* The “Level” value will be added as a “Keyword” for the database entry.
* The “Subcategory” value is added to the “Keywords” field.
* Audience is set to the same value as Affiliation.
* StartDateTime and EndDateTime are NULL since the spreadsheet doesn’t contain training class sessions.

# **Transferring data from the portal database**

The portal database stores course information in the portal.training\_class table. There is another table (portal.training\_class\_session) that contains information about specific sessions of a class that are taught at a particular location at a particular time. It looks like the portal.training\_class table was used to store session information up until 2013-07-22. The portal.training\_class\_session table duplicates the information for those early rows and contains all of the information for sessions from that time up to the present so these early rows will be ignored.

For each record in the portal.training\_class table, a record is created in the glue2.resource\_v3\_resourcev3local table:

* An ID entry is generated 'urn:ogf:glue2:info.xsede.org:resource:portal:training\_class:{id}' where {id} is the id of the entry in Portal DB.
* The current time is stored as CreationTime.
* Validity is an interval set to 14 days.
* The site\_url value is stored as Affiliation.
* The id value is stored as LocalID.
* LocalType is set to “Training”.
* LocalURL
* CatalogMetaURL
* A JSON entry is generated from all of the information in each record. It is stored as EntityJSON.

A record is also created in the glue2.resource\_v3\_resourcev3 table:

* The ID, Affiliation, and LocalID used above are stored in the values of the same name.
* Type is set to “Training”
* QualityLevel is set to “Production”.
* The training\_name value is stored as Name.
* The “training\_type” field in the Portal DB is only used in records with a start date which are being ignored because they are duplicated in the training\_class\_session table. So ResourceGroup will be set to NULL.
* ShortDescription, Topics, and Keywords will all be set to NULL.
* If the location\_name already exists in the resource\_v3\_resourcev3 table as a provider, its ID will be stored as the ProviderID for the record. Otherwise it will be added to both the resource\_v3\_resourcev3 table and to the resource\_v3\_resourcev3local table and the newly generated ID will be stored as the ProviderID.
  + In the resource\_v3\_resourcev3local table:
    - A GlobalID value is generated (using the current record’s local id) for the new provider and stored as ID. urn:ogf:glue2:info.xsede.org:resource:portal:provider:{id}
    - The current time is stored as CreationTime.
    - Validity is an interval set to 14 days.
    - Affiliation is set to xsede.org.
    - The record’s id is stored as LocalID.
    - LocalType is set to “provider”.
    - LocalURL and CatalogMetaURL are set to NULL.
    - The site\_name, site\_url, contact\_name, contact\_phone, contact\_email, location\_name, location\_address\_1, location\_address\_2, location\_city, location\_state, location\_zip, location\_phone, and location\_url are stored in EntityJSON.
  + In the resource\_ve\_resourcev3 table:
    - The ID, Affiliation, and LocalID used above are stored in the values of the same name.
    - QualityLevel is set to “Production”.
    - The site\_name value is stored as the Name.
    - ResourceGroup is set to “Organizations”.
    - Type is set to “Provider”.
    - ShortDescription, Description, Topics, Keywords, and ProviderID are set to NULL.
    - Audience is set to the same value as Affiliation.
    - StartDateTime and EndDateTime are set to NULL.
* The training\_summary field is stored into the Description.
* Audience is set to the same value as Affiliation.

portal.training\_class\_session records contain a training\_class\_id that points to a record in the portal.training\_class table. At this point an entry has already been created for that training\_class record. For each record in the portal.training\_class\_session table, the record for its corresponding training\_class record is copied from the glue2.resource\_v3\_resourcev3local table and modified:

* An ID entry is generated 'urn:ogf:glue2:info.xsede.org:resource:portal:training\_class\_session:{id}' where {id} is the id of the entry in Portal DB.
* The current time is stored as CreationTime.
* Validity is an interval set to 14 days.
* The site\_url value is stored as Affiliation.
* The id value is stored as LocalID.
* LocalType is set to “Training”.
* LocalURL
* CatalogMetaURL
* All of the information for this record is added to the data from its training\_class record in EntityJSON.

A record is also created in the glue2.resource\_v3\_resourcev3 table. Information will be copied from its associated training\_class record. The following values will be set:

* The ID, Affiliation, and LocalID used above are stored in the values of the same name.
* The “training\_type” field in the Portal DB is only used in records with a start date which are being ignored because they are duplicated in the training\_class\_session table. So ResourceGroup will be set to NULL.
* ShortDescription will list the location information from the training\_class\_session table.
* The start\_date value will be stored as StartDateTime and the end\_date value will be stored as EndDateTime.

A record is also created in the glue2.resource\_v3\_resourcev3relation table:

* An ID entry is generated.
* training\_class\_id is stored as FirstResourceID.
* id is stored as SecondResourceID.
* RelationType is set to “Live Session”.

# **XSEDE Training Course spreadsheet columns**

* Center – (CAC, ESNET, IU, NCSA, NICS, PRACE, PSC, PURDUE, SDSC, TACC, UMICH)
* Logo – URL to center’s logo (NCSA has old logo)
* Title – Course title
* Vetted – (Yes, No)
* Exercise
* Upcoming – (Yes, No)
* Specific – (Yes, No)
* Level – (Advanced, All, Beginner, Intermediate)
* Prerequisites – “None” or description of prerequisite knowledge
* Length – Hours, days, or weeks
* Category - Applications, Architecture Accelerators and Coprocessors, ARCHITECTURE ACCELERATORS AND COPROCESSORS, Big Data File Movement and Workflows, BIG DATA FILE MOVEMENT AND WORKFLOWS, Comprehensive Courses, COMPREHENSIVE COURSES, Course Collection, Debuggers Profilers and Optimization Tools, Domain Specific Courses, Introductory Courses, Job Submission and Management, Linux /Unix Basics and Environment, Parallel and High Throughput Programming, PARALLEL AND HIGH THROUGHPUT PROGRAMMING, Programming and Libraries, Science gateways, Scientific workflows, Visualization
* Subcategory - Bioinformatics, C, Chemistry, Computing Environment, Data analysis, Data movement and storage, Debugging, Engineering, FORTRAN, Gateways, GPGPUs, Hadoop, Hybrid OpenMP/MPI, Intel Xeon Phi, I/O, Job Submission and Management, Life Sciences, Linux /Unix Basics, Machine specific training, Matlab, MPI, Multiple, New User, New User/Allocations, Numerical Libraries, OpenMP, Optimization for CPUs, Parallel Computing Basics, Parallel Computing: MOOC, Physics, Profiling and optimization, Programming and Libraries, Python, R, Science gateways, Scientific workflows, Security, Use Cases, Visualization, Workflows
* Delivery – (IN-PERSON, MULTICAST, Online, Training Portal, Webinar, YouTube)
* Assessment – (Yes, No)
* Description – Text description of course
* url – Link to course material, some links are broken

# 

# **Warehouse V3 tables**

## **Table "glue2.resource\_v3\_resourcev3"**

List of all resources. All generic training class and specific training class session records get added here.

* ID – (urn:)
* Affiliation - (uiuc.edu, xsede.org)
* LocalID = (# for uiuc, matches ID for xsede)
* QualityLevel - (Preliminary, Testing, Production, Unsupported, Pre-production, Decommissioned)
* Name - Resource Name
* ResourceGroup - (Data Resources, Guides, Live Events, Computing, Tools and Services, Organizations, Streamed Events, Software)
* Type - (Training, Online Service, Backup and Storage, Web Publishing and Communications, Research Computing, Executable Software, Networking and Security, Packaged Software, Programming, Event, Consulting and Support, Data, Provider, Software)
* ShortDescription - Short Description
* ProviderID - (urn:)
* Description - Description
* Topics - csv list of topics
* Keywords - csv list of keywords
* Audience - (uiuc.edu, xsede.org)
* StartDateTime - DateTime or NULL
* EndDateTime - DateTime or NULL

## **Table "glue2.resource\_v3\_resourcev3catalog"**

* ID – (urn:)
* Created\_at
* Updated\_at
* Name
* Affiliation
* ShortDescription
* CatalogMetaURL
* CatalogUserURL
* CatalogAPIURL
* CatlaogSchemaURL

## **Table "glue2.resource\_v3\_resourcev3local"**

One to one match to glue2.resource\_ve\_resourcev3 table?

* ID – (urn:)
* CreationTime - Time record was written
* Validity - (Planned, Pre-production, Decommissioned, Production)
* Affiliation
* LocalID
* LocalType - (curated\_guide, Support Contact, Executable Software, Packaged Software, resource, Resource Site Information, Network Service, provider, Science Gateway, Resource Information)
* LocalURL
* CatalogMetaURL
* EntityJSON

## **Table "glue2.resource\_v3\_resourcev3relation"**

* ID – (urn:glue2:GlobalResource:#.uiuc.edu)
* FirstResourceID - (urn:)
* SecondResourceID - (urn:)
* RelationType - (Hosted On, Guide Resource, Supported By, Gateway Integrated, Provided By, Resource Association, Operated By)

## 

# **XSEDE Portal Database**

## **Table "portal.training\_class"**

* Id -
* training\_name - Course title
* training\_url – Link to course material, some links are broken
* training\_summary - Kind of random
* site\_name - Site that hosts the class?
* site\_url - URL of the site
* contact\_name - mostly unused
* contact\_phone -
* contact\_email -
* fee - mostly unused
* fee\_details - comments about fees, mostly unused
* fee\_payment\_url - mostly unused
* training\_type - (ONLINE\_TRAINING, WEBCAST, IN-PERSON) mostly unused
* location\_name - similar to site\_name
* location\_address\_1 -
* location\_address\_2 -
* location\_city -
* location\_state -
* location\_zip -
* location\_phone -
* location\_url -
* start\_date - dates from 2012, 2013. Mostly unused, superseded by training\_class\_session?
* end\_date -
* timezone\_class - (America/Los\_Angeles, America/Denver, America/New\_York, America/Chicago) mostly unused
* registration\_open\_date – dates from 2012, 2013
* registration\_close\_date -
* timezone\_registration -
* registration\_limit – mostly unused
* published - (t,f)
* private - (t,f)
* password - plaintext password, mostly unused
* notification\_email -
* confirmation\_message -
* confirmation\_replyto -
* owner\_id - numeric ID, mostly unused
* created -
* no\_registration - (t,f)
* no\_registration\_type – (ANNOUNCEMENT, OPEN\_EVENT) mostly unused
* feedback\_request\_sent – (t,f)
* waitlist - (t,f)
* owner - XSEDE login ID of the owner of the entry

## **Table "portal.training\_class\_category"**

* id -
* category\_name -
* category\_weight –

**Table "portal.training\_class\_manager"**

* id -
* training\_class\_id -
* person\_id -
* assigned\_by\_person\_id -
* ts -
* username -

## **Table "portal.training\_class\_session"**

* id -
* training\_class\_id -
* contact\_name -
* contact\_phone -
* contact\_email -
* fee -
* fee\_details -
* fee\_payment\_url -
* training\_type -
* location\_name -
* location\_address\_1 -
* location\_address\_2 -
* location\_city -
* location\_state -
* location\_zip -
* location\_phone -
* location\_url -
* start\_date -
* end\_date -
* timezone\_class -
* registration\_open\_date -
* registration\_close\_date -
* timezone\_registration -
* registration\_limit -
* notification\_email -
* confirmation\_message -
* confirmation\_replyto -
* created -
* feedback\_request\_sent -

## **Table "portal.training\_class\_trainer"**

* id - (#)
* person\_id - (#)
* assigned\_by\_person\_id - (#)
* username - (login name)
* ts - (time)
* session\_id - (#)

## **Table "portal.training\_session\_to\_category"**

* category\_id - (#)
* session\_id - (#)
* session\_weight - (0)