Course/xU
Analytics Reports

As of 10 January 2014

About edX data: access and formats

Most edX data is stored in three general places:

- Student data in SQL database: registration info, answers to exercises, certificate info, etc.
- Course data in MongoDB database: courseware, tabs info, Studio settings, etc.
- Forum data in MongoDB database: threads and posts

This document discusses student data. To access it, you will need a Github account with an ssh key, and to send DevOps a copy of your public key and request for read-only access (Jira ticket). Olga can help you with this.

Then, you'll need to create your "query" or SQL script. Below is an example that gets student gender, year of birth, level of education, and date of enrollment for UTAustinX courses. Change the details as needed and save as NAMEofQUERY.sql:

```
select aup.user_id,
       ifnull(aup.gender,"") as gender,
       ifnull(aup.year_of_birth,"") as year_of_birth,
       ifnull(aup.level_of_education,"") as level_of_education,
       course_id,
       date(created)
from (
    select * from student_courseenrollment
    where course_id like "UTAustinX/UT._.01x/%"
    ) as sce
left outer join auth_userprofile as aup
on sce.user_id=aup.user_id
;
```

% means that anything could come after this, so long as the first part matches

_ indicates wildcard character

Note: Download and use Sublime Text 2 or TextWrangler for writing code, it makes things easier.

Next, you'll have to execute the query. Open up a terminal window, navigate to the directory containing your query file (`cd Documents/Analytics` or something like that) and use the following syntax to run the query and save the output file in the same location:

mysql --host=prod-analytics.rds.edx.org --user=read_only -p wwc < NAMEofQUERY.sql > NAMEofOUTFILE_DATE.txt

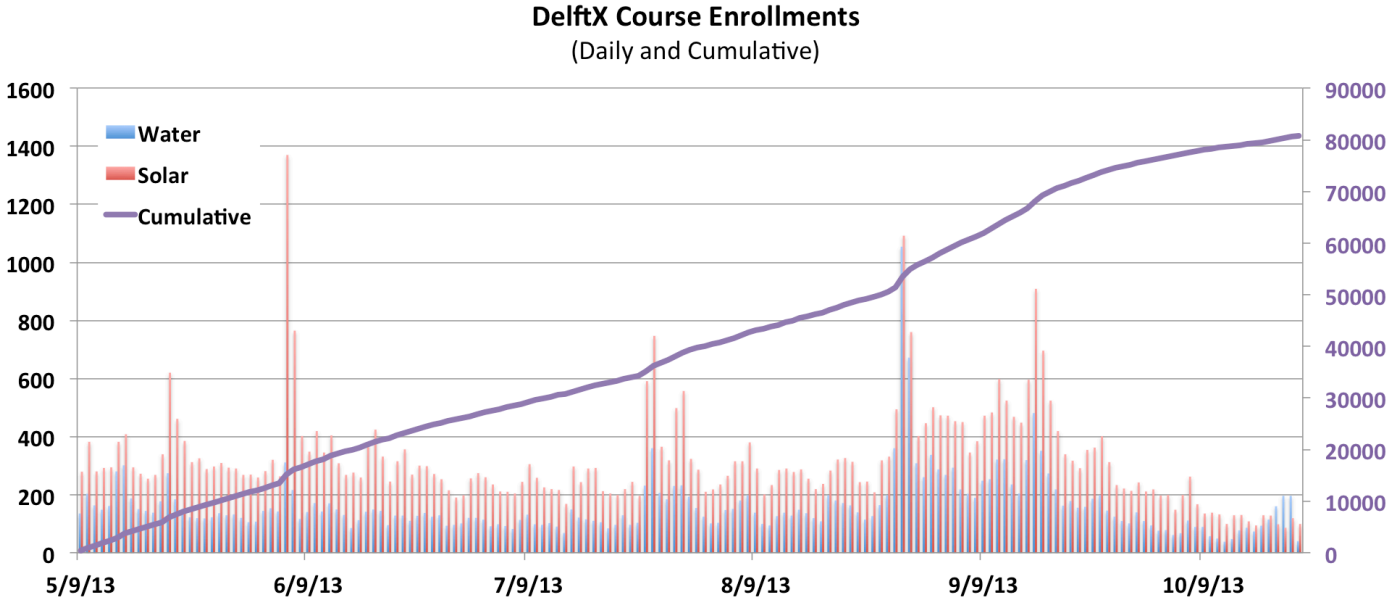
Need a place to start? Try a tutorial on UNIX commands to learn navigation, creating/deleting directories (technical term for folders), and common tasks like copying or moving files via the command line.

Confused about the format of the SQL database and its tables, like auth_userprofile? See docs.edx.org or the (probably outdated) cheat sheet at this end of this doc.

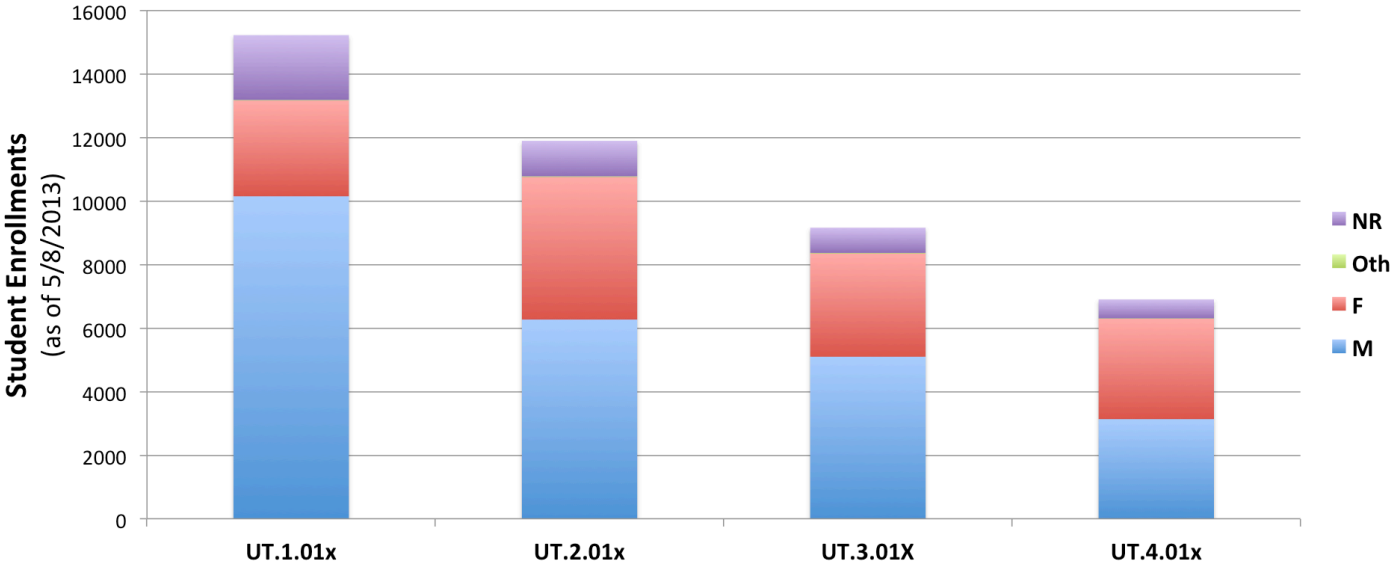
Before launch: enrollments & demographics

Between registration opening and the course launch, universities are mostly interested in how many students are signing up for their courses, when there are enrollment spikes (and what caused them, like an edX student newsletter), and who these students are demographically.

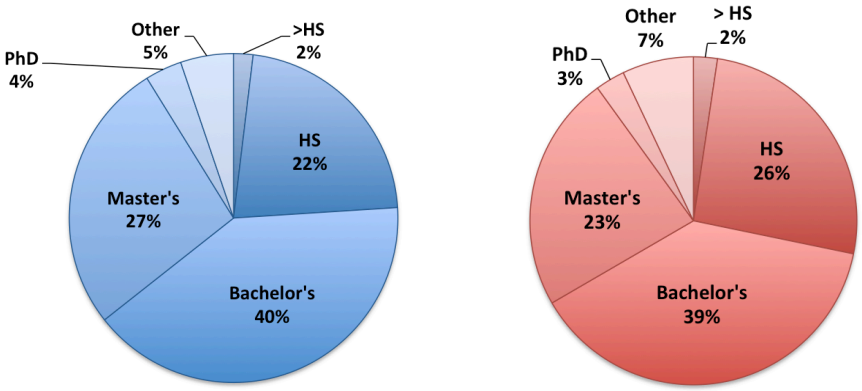
1.1 Enrollment timelines: can be per course or include all current courses for a university.



1.2 Enrollment summaries: gender ratios are also of great interest to partners, so include them if you like.

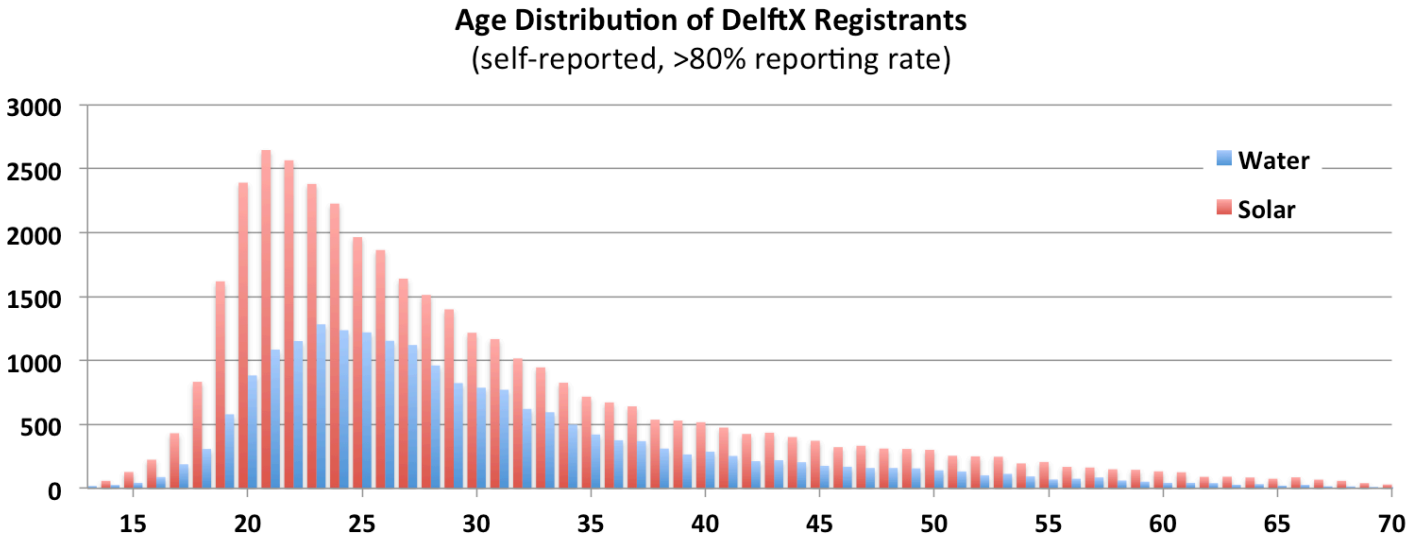


1.3 Level of education: most of our students already have postsecondary degrees, but this can vary.



DelftX Water (blue) and Solar (red) level of education distributions.

1.4 Age distribution: again, our users are generally not your typical undergraduates.



In order to get this data:

Create your "query" or SQL script. Change the details as needed and save as NAMEofQUERY.sql:

```
select aup.user_id,
       ifnull(aup.gender,"") as gender,
       ifnull(aup.year_of_birth,"") as year_of_birth,
       ifnull(aup.level_of_education,"") as level_of_education,
       course_id,
       date(created)
from (
    select * from student_courseenrollment
    where course_id like "UTAustinX/UT._.01x/%"
    ) as sce
left outer join auth_userprofile as aup
on sce.user_id=aup.user_id
;
```

Open a terminal, change directory to your desired file location, and run the following line of code:
mysql --host=prod-analytics.rds.edx.org --user=read_only -p wwc < NAMEofQUERY.sql > NAMEofOUTFILE_DATE.txt

NOTE: These queries might take up to 3 hours to run depending on the size of the course. Make sure your internet connection is secure and reliable!

After launch: student engagement

After the course launches and students start interacting with the courseware, the most common questions from course teams and university staff are: how many students are watching the videos, which videos are they watching (and for how long), and are they doing the assessments?

We can't help with the first two, although a document showing them how to go into their own YouTube CMS accounts and obtain the information themselves would be helpful.

We can provide very basic weekly summaries of student engagement with the assessments. It's not great, but it's enough information to let the course teams feel like they are no longer "flying blind."

Step 1: Obtain up-to-date course json maps & convert to csv

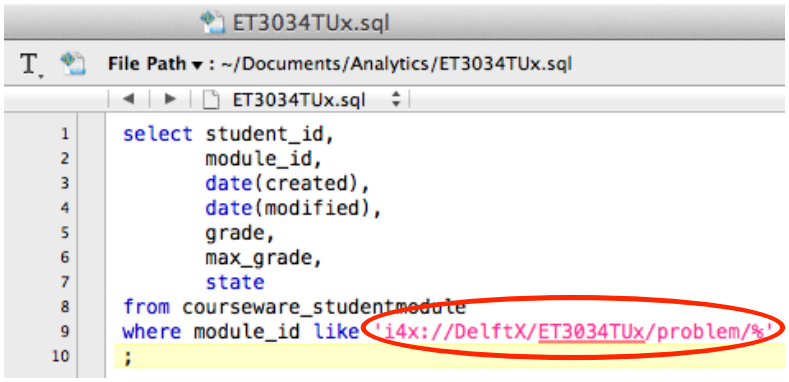
Make a list all of the courses (UnivX/Course101x/1T2014) that have requested reports, and note which ones are on Edge or edX.org. Put this list into a Jira ticket: Studio Chore, assigned to Chris Dodge. He will get these to you within 2-3 business days.

Once you have them, run the following command in a terminal to create a simple course content map:

```
$ python json2csvC.py MYCOURSEID.json > MYCOURSEID.csv
```

Step 2: Query the database

Now for the actual data! For each course, you'll need to create the following MYCOURSEID.sql file:



Then, for each MYCOURSEID.sql, run the following command in a terminal (all one line):

```
$ mysql --host=prod-analytics.rds.edx.org --user=read_only -p wwc <
MYCOURSEID.sql > MYCOURSEID_DATE.txt
```

NOTE: You're going to be running these regularly, so include the date in the output filename! Also, these queries might take up to 3 hours to run depending on the size of the course. Make sure your internet connection is secure and reliable!

Step 3: Tie it all together

Run the following command in a terminal, making sure that PscriptMod.py and MYCOURSEID_DATE.txt are in the same folder:

```
$ python PscriptMod.py MYCOURSEID_DATE.txt MYCOURSEID.csv
```

This will take the course map and the data and combine them into MYCOURSEID_DATEreadable.csv

The engagement summary for a course lists, in sequential order, all of the problems in a course and the number of students who earned credit for each part of the problem. It can be easily opened in Excel (comma delimited).

This summary does NOT reflect things like point values or, for multipart problems, which parts were answered correctly. For optimal information, course teams should create a new, separate component for each part of a problem.

For example, let's look at Question 1.3.2 in this course. This problem had 2 parts:
444 students earned zero credit,
627 students succeeded with one part but not the other,
9777 students earned full credit for completing both parts, and
27220 students looked at the learning sequence but did not attempt to answer.

Looked at the learning sequence or homework, but did not attempt to answer any questions

display_name	module_id	0	1	2	3	4	5	6	7	9	9999
About EdX	i4x://DelftX/ET3034TUx/problem/b3d30df864ca41ffa0170e790f01a783										27220
Demographics & Learning preference	i4x://DelftX/ET3034TUx/problem/ae408726ac4a43abbd661a9bd5df0c20										27220
Education and background	i4x://DelftX/ET3034TUx/problem/d68b6d0719594b368f8301636f1403fe										27220
You and this MOOC	i4x://DelftX/ET3034TUx/problem/c4840764159f452c814f99e6b2b7b99c										27220
Communication and collaboration	i4x://DelftX/ET3034TUx/problem/7e13dc5f7691487d8de158f65845a484										27220
Aspirations	i4x://DelftX/ET3034TUx/problem/8418dbd432a741c99de79c6c09d7ac88										27220
Stay updated about the research	i4x://DelftX/ET3034TUx/problem/51d31904ab594137ae4ee762948dd492										27220
QUESTION 1.2.1	i4x://DelftX/ET3034TUx/problem/8a353992676e45f1bde78c03ab3791ab	448	12286	1	*						15401
QUESTION 1.2.2	i4x://DelftX/ET3034TUx/problem/0c6efae23d624b07b32f9837f3d16c6c	894	11564								15678
QUESTION 1.3.1	i4x://DelftX/ET3034TUx/problem/81ede2a654bd4b1392a81c9c21281067	443	10541								15286
QUESTION 1.3.2	i4x://DelftX/ET3034TUx/problem/79e672094c8e407db4c042c1052758c3	444	627	9777							15422
QUESTION 1.3.3	i4x://DelftX/ET3034TUx/problem/30c3e9751be842bd9f65356c21023463	551	10057								15662
QUESTION 1.5.1	i4x://DelftX/ET3034TUx/problem/c9bc0ff23e4940349186f2242810f8b0	2270	866	4743							16533
QUESTION 1.5.2	i4x://DelftX/ET3034TUx/problem/22fff45893e24e3ea5ebd9c05fd498db	1901	145	5242							17124
QUESTION 1.6.1	i4x://DelftX/ET3034TUx/problem/b493b7973c6d436db037ba60c5ae65c0	1982	590	3824							17759
QUESTION 1.6.2	i4x://DelftX/ET3034TUx/problem/38b231210da04fd49b35c6cbf0c8a01a	844	433	477	3597						18804
1.1 Costs of PV Technology Generations	i4x://DelftX/ET3034TUx/problem/4b6d3ff228d440cf8585c768313c776d	2870	1046	2062	2777						16437
1.2 Renewables in Global Electricity Supply Mix	i4x://DelftX/ET3034TUx/problem/7c373beb47a84e9a8710c589040aabcb	2330	407	5272							17183
1.3 Power Spectral Density and Photon Flux	i4x://DelftX/ET3034TUx/problem/dc6cd763c50947629bc4fb96b6b5c59d	1898	2368	2612							18314
1.4 PV Potential Around the World	i4x://DelftX/ET3034TUx/problem/96fd71e4edb8429e8080b93043d59a34	679	931	1294	1927	2350					17951
1.5 PV System for the Smith Family	i4x://DelftX/ET3034TUx/problem/ac6905494f4441cd82070b90408781fc	1927	533	853	388	857	186	1848			18600
Question 2.1.1	i4x://DelftX/ET3034TUx/problem/614612af05c64e49bbf03ca2724696d0	757	4926								12370
Question 2.1.2	i4x://DelftX/ET3034TUx/problem/054152f2087d4416ab9864b45dc0fb34	622	5008								12423
Question 2.2.1	i4x://DelftX/ET3034TUx/problem/92991f18193f4c35b22dc60d186843da	266	5091								12236

* This is an artifact: this question had 2 parts during early development, but was published with only 1 part.

```
mysql> describe student_courseenrollment;
+-----------+--------------+------+-----+---------+----------------+
| Field     | Type         | Null | Key | Default | Extra          |
+-----------+--------------+------+-----+---------+----------------+
| id        | int(11)      | NO   | PRI | NULL    | auto_increment |
| user_id   | int(11)      | NO   | MUL | NULL    |                |
| course_id | varchar(255) | NO   | MUL | NULL    |                |
| created   | datetime     | YES  | MUL | NULL    |                |
+-----------+--------------+------+-----+---------+----------------+
```

```
mysql> describe certificates_generatedcertificate;
+---------------+--------------+------+-----+---------+----------------+
| Field         | Type         | Null | Key | Default | Extra          |
+---------------+--------------+------+-----+---------+----------------+
| id            | int(11)      | NO   | PRI | NULL    | auto_increment |
| user_id       | int(11)      | NO   | MUL | NULL    |                |
| download_url  | varchar(128) | NO   |     | NULL    |                |
| grade         | varchar(5)   | NO   |     | NULL    |                |
| course_id     | varchar(255) | NO   | MUL | NULL    |                |
| key           | varchar(32)  | NO   |     | NULL    |                |
| distinction   | tinyint(1)   | NO   |     | NULL    |                |
| status        | varchar(32)  | NO   |     | NULL    |                |
| verify_uuid   | varchar(32)  | NO   |     | NULL    |                |
| download_uuid | varchar(32)  | NO   |     | NULL    |                |
| name          | varchar(255) | NO   |     | NULL    |                |
| created_date  | datetime     | NO   |     | NULL    |                |
| modified_date | datetime     | NO   |     | NULL    |                |
| error_reason  | varchar(512) | NO   |     | NULL    |                |
+---------------+--------------+------+-----+---------+----------------+
```

```
mysql> describe auth_userprofile;
+--------------------+--------------+------+-----+---------+----------------+
| Field              | Type         | Null | Key | Default | Extra          |
+--------------------+--------------+------+-----+---------+----------------+
| id                 | int(11)      | NO   | PRI | NULL    | auto_increment |
| user_id            | int(11)      | NO   | UNI | NULL    |                |
| name               | varchar(255) | NO   | MUL | NULL    |                |
| language           | varchar(255) | NO   | MUL | NULL    |                |
| location           | varchar(255) | NO   | MUL | NULL    |                |
| meta               | longtext     | NO   |     | NULL    |                |
| courseware         | varchar(255) | NO   |     | NULL    |                |
| gender             | varchar(6)   | YES  | MUL | NULL    |                |
| mailing_address    | longtext     | YES  |     | NULL    |                |
| year_of_birth      | int(11)      | YES  | MUL | NULL    |                |
| level_of_education | varchar(6)   | YES  | MUL | NULL    |                |
| goals              | longtext     | YES  |     | NULL    |                |
+--------------------+--------------+------+-----+---------+----------------+
```

```
mysql> describe auth_user;
+------------------------------+--------------+------+-----+---------+----------------+
| Field                        | Type         | Null | Key | Default | Extra          |
+------------------------------+--------------+------+-----+---------+----------------+
| id                           | int(11)      | NO   | PRI | NULL    | auto_increment |
| username                     | varchar(30)  | NO   | UNI | NULL    |                |
| first_name                   | varchar(30)  | NO   |     | NULL    |                |
| last_name                    | varchar(30)  | NO   |     | NULL    |                |
| email                        | varchar(75)  | NO   | UNI | NULL    |                |
| password                     | varchar(128) | NO   |     | NULL    |                |
| is_staff                     | tinyint(1)   | NO   |     | NULL    |                |
| is_active                    | tinyint(1)   | NO   |     | NULL    |                |
| is_superuser                 | tinyint(1)   | NO   |     | NULL    |                |
| last_login                   | datetime     | NO   |     | NULL    |                |
| date_joined                  | datetime     | NO   |     | NULL    |                |
| status                       | varchar(2)   | NO   |     | NULL    |                |
| email_key                    | varchar(32)  | YES  |     | NULL    |                |
| avatar_type                  | varchar(1)   | NO   |     | NULL    |                |
| country                      | varchar(2)   | NO   |     | NULL    |                |
| show_country                 | tinyint(1)   | NO   |     | NULL    |                |
| date_of_birth                | date         | YES  |     | NULL    |                |
| interesting_tags             | longtext     | NO   |     | NULL    |                |
| ignored_tags                 | longtext     | NO   |     | NULL    |                |
| email_tag_filter_strategy    | smallint(6)  | NO   |     | NULL    |                |
| display_tag_filter_strategy  | smallint(6)  | NO   |     | NULL    |                |
| consecutive_days_visit_count | int(11)      | NO   |     | NULL    |                |
+------------------------------+--------------+------+-----+---------+----------------+
```

```
mysql> describe courseware_studentmodule;
+-------------+--------------+------+-----+---------+----------------+
| Field       | Type         | Null | Key | Default | Extra          |
+-------------+--------------+------+-----+---------+----------------+
| id          | int(11)      | NO   | PRI | NULL    | auto_increment |
| module_type | varchar(32)  | NO   | MUL | problem |                |
| module_id   | varchar(255) | NO   | MUL | NULL    |                |
| student_id  | int(11)      | NO   | MUL | NULL    |                |
| state       | longtext     | YES  |     | NULL    |                |
| grade       | double       | YES  | MUL | NULL    |                |
| created     | datetime     | NO   | MUL | NULL    |                |
| modified    | datetime     | NO   | MUL | NULL    |                |
| max_grade   | double       | YES  |     | NULL    |                |
| done        | varchar(8)   | NO   | MUL | NULL    |                |
| course_id   | varchar(255) | NO   | MUL | NULL    |                |
+-------------+--------------+------+-----+---------+----------------+
```