# Install the current official release

This note describes installation of tranSMART version 16.3, released on 6 August 2018 (installed on Ubuntu 16.04)

Discussion on upgrading to 16.3 from an earlier version can be found here.

## Overview

These instructions are based on the assumption that you are installing tranSMART version 16.3 on Ubuntu 14.04. They draw heavily on previous versions of the release notes, see the credits section of this document for details.

The instructions come in two flavors (1) instructions for an all-in-one script and (2) more detailed step-by-step instructions. The step-by-step instructions are useful when you need to see what is going on and possibly correct problems. They also will serve as a helpful guide for installation on other platforms.

As an overview, the install process implements these five steps. For the step-by-step instructions, follow each step in order, reading each step completely. Expect the process to take several hours (depending on the speed of your internet connection, the power of your machine, and the speed of your machine's memory and disk). You can also see these steps in the comments of the all-in-one script.

1. You will need the set up the necessary support tools (some basic unix command-line tools, PostgreSQL, Grails, Tomcat, R, RServe, and SOLR).
2. You will need to set the database (PostgreSQL is used in these instructions - tranSMART also runs on Oracle), and load the database with the core control data for transmart.
3. You will need to download and copy the release war files to tomcat.
4. You will need to start all the parts of the system
5. And you will need to load some number of datasets.

In addition to the browser-based web-application, there is also a set of R-Based modules that connect to the application's REST interface; these can be used to connect to the database, search for and select data, and to perform analyses.

There is also a Sanity Checklist for a fully functional system, to make sure that everything was correctly installed and it running properly. This may help you pinpoint any problems.

## Getting Help

These instructions attempt to be complete, however, unexpected problems can occur. If you are stuck, there are several sources for help, which can be found on the Getting Support page.

When email or posting a question or a request for help, please describe the situation with as much detail as possible: What were you trying to do? What actions were you taking or had you taken before the error occurred? What was the last command, choice, or action that you entered? What was the error message or last output on the terminal? Remember the person you are in communication with can always envision what you are seeing. Help them to do so.

If you know anyone in the community, especially someone who has recently installed tranSMART, contact them with your question. Often they will have seen the same problem.

Finally, if you learn anything that should be added to these instructions page feed it foward: please leave a comment (at the end)!

# Hardware and OS Requirements

The requirements vary based on your specific needs. Since these instructions are aimed at producing a demonstration system for review and evaluation, we are assuming very modest requirements, for example, no more than 100 subjects, a couple of studies with genomic data, and fewer than 10 users; in this case, you can combine the app server with the database server into a single VM or a standalone server with at least the following specs:

- CPU: Dual core
- Memory: 4 to 8GB (8 preferred; will work on as little as 4GB)
- Storage: 50 GB (for more studies increase the size of the disk storage)
- OS: linux (these instructions assume: Ubuntu 14.04).

## How to setup a VM on your desktop for evaluation

1. Install VirtualBox (Or VMWare)
2. Create a new VM with at least the above specifications on memory and disk space.
3. Download the ISO file for the linux operation system you want (these instructions assume: Ubuntu 14.04)
4. When booting the VM select the ISO as the disk to boot from.
5. Walk through the instructions of the ISO installation program.

# All-in-one Install Script, plus Data-Loading

The "all-in-one" instructions require the minimum of human interactions. In these instructions, there are three major steps. First a set of instructions guides you in setting up a **Scripts** folder on the home directory of **an account with sudo privileges**. In the second step, you run the install script that will install all of the support command line tools, create the required support directories, load the required toolsets for PostgreSQL, R and Rserve, load the required R packages, load and start SOLR and the SOLR web interface, and load the transmart war files into Tomcat 7. In addition, the second step sets up the transmart database. In the third step, you run a script that loads the demo datasets. Optionally, by editing a configuration file, you can select additional datasets to load.

Upon completion of the install you can (optionally) run a set of checking scripts that will check to make sure everything is installed correctly and running properly.

In an additional supplementary process, you can use the full PGP keys to verify the downloaded install artifacts. This process is described in this additional document. Therein you will see instructions for downloading tranSMART Foundation PGP keys and using them to verify the digital signature of the downloaded artifacts (that is, the zip files of, the install scripts, transmart-data, tranSMART-ETL, transmart.war, and gwava.war).

Make sure the machine or VM you are installing on is according to the Hardware and OS Requirements as specified above.

## Set up a transmart account

We assume that you have set up an account with sudo privileges to use in these instructions. For sake of illustration we will use an account named transmart. If you have not already set up such an account, you do so with the following commands (assuming, in this case, that you are logged in with the initial admin account, ubuntu).

**Setting up a sudo account, transmart**

```
# set the password for your new user 'transmart'
# use defaults for the other values if you wish
sudo adduser transmart

# add the user transmart to the sudo group
sudo adduser transmart sudo

# log out and log back in with your new user
```

## Set up the install script

To download the **Scripts** zip file and install its folder, run the following commands in the home directory of the account that you are going to use to run transmart (e.g in our case, the account named **transmart**).

**Download, install Scripts**

```
sudo apt-get install -y curl
sudo apt-get install -y unzip

# download the installation script
curl http://library.transmartfoundation.org/release/release16_3_0_artifacts/Scripts-release-16.3.zip -o Scripts-
release-16.3.zip
# unzip and rename the Scripts folder
unzip Scripts-release-16.3.zip
mv Scripts-release-16.3 Scripts
```

## Run the install script

Next, we run the install script. This script, if successful, installs tranSMART. It the script fails, read the last message, the error message, carefully; it will usually suggest a fix for the problem. Then, fix the problem and rerun the script. The script is robust with respect to being rerun; and will usually run to completion. To recover error messages, review the **install.log** file in the install user's home directory.

In the case that the error seems to have left things in an unrecoverable state, delete the folder ~/transmart (and its contents) and restart the install script as below. This will simply reload all the tranSMART artifacts, without having to reload the command-line tools.

Note: during this process, you will be prompted for the account's password, by the 'sudo' command, at several points along the way.

```
./Scripts/install-ubuntu/InstallTransmart.sh 2>&1 | tee ~/install.log
```

# Loading datasets

After completing the installation procedure above, you have an 'empty' instance of tranSMART. Now, you need to load data, either your own or some of the Curated Data Repository sets provided by the Foundation. In these instructions we will illustrate loading data using scripts that access the curated datasets. We will use the script in the l**oad datasets** code block, below, to load datasets into the tranSMART database. These instructions describe how to load a representative sample of curated datasets. However, they can be easily generalized; you can choose which datasets you wish loaded (the **select datasets** code block).

For complete details of the loadable datasets, and additional details on the loading process, see the instructions at the start of the Curated Data Repository wiki page.

These two files are in the **Scripts/install-ubuntu** directory are involved in this process:

- **datasetsList.txt** - the list of possible datasets to load, and
- **load_datasets.sh** - the script to load the datasets.

By default, a representative sample of the datasets are loaded. Specifically, these datasets are loaded, by default:
EtriksGSE2125, PfizerGSE22138, RanchoGSE11903, RanchoGSE4382, SanofiGSE38642 . The Curated Data Repository sets wiki page has details of the source and content of these datasets.

Optionally, you can edit the list of datasets to load. Using a text-file editor, like vim, as in the details instructions below, edit the file **datasetsList.txt**, to uncomment the lines indicating of the datasets that you wish to load and/or comment out the lines of the datasets already loaded or not to be loaded. Please note, it is generally not a good idea to try and load a dataset twice. For each dataset in that list, you can find a description on the  Curated Data Repository sets wiki page. The detailed command sequence, below, includes the loading of vim (a visual version of vi) and the editing of the list of datasets. If you are unsure about this editing process you can run the script without editing the file. Doing so will load the 5 representative datasets, above.

**select datasets (optional)**

```
# Note, you may use any text edit, to edit datasetList.txt; herein, we use vi
# this next line will runs the install that makes available the vi command used below
sudo apt-get install -y vim

# Note: in the 'cd' command, the assumption is that the Scripts directory
# is in the home directory of the current account
# if not, adjust accordingly
cd $HOME

vi ./Scripts/install-ubuntu/datasetList.txt
```

Once you are in the (**vi**) editor, you can use the following commands:

- **i** - to switch into "insert" mode (the tag **--INSERT--** will appear at the bottom of the screen)
- the arrow keys to move from line to line or in position within a line
- the 'delete' key to delete the character to the left of the cursor
- (when in insert mode) any character key to type that character (for example #, at the start of a line, to comment out that line)
- the ESC (escape) key to get out of insert mode
- ZZ  - (two upper-case z characters - when not in insert mode) to exit and save
- :q!  - (when not in insert mode) to exit immediately without saving

Edit the list of datasets by adding the # character at the start of a line to comment out those datasets that you do not which to load and deleting the # character at the front of the lines corresponding to the datasets that you which to load.

Once you have set up, in **datasetsList.txt**, the list of datasets you wish to load, then, run the script, **load_datasets.sh,** as indicated, here.

---

**load datasets**

```
# Note: in the 'cd' command, the assumption is that the Scripts directory
# is in the home directory of the current account
# if not, adjust accordingly
cd $HOME

./Scripts/install-ubuntu/load_datasets.sh
```

---

With this, the install is complete (with datasets loaded). Open a browser to http://localhost:8080/transmart/ where you should see the tranSMART Web Application's login page. On the Analyze tab, in the tree-interface on the left, you will see the list of datasets loaded. The default login is admin/admin, and you can change the password of that account at the Admin Page in the interface.

## Is the system running?

Finally, at any time, you can test to see if the install was successful, that is if all the "moving parts" were installed and are running, by using the following commands:

```
# Note: in the 'cd' command, the assumption is that the Scripts directory
# is in the home directory of the current account
# if not, adjust accordingly
cd $HOME

cd Scripts/install-ubuntu/checks/
./checkAll.sh 2>&1 2>&1 | tee ~/checks.log
```

Review the log file, **checks.log**, in your home directory. For details see the section, below, on using the web application, towards the end of this document. To restart the needed tool interfaces (after a reboot, for example) see the section, below on Running SOLR, Rserve, and Tomcat.

# Step-by-step Instructions

The rest of this document gives instructions for a step by step series of commands (which are also implemented by the all-in-one script) to better illustrate what is going on and give installers of transmart more opportunity to tune or modify the instructions to particular needs. In addition, seeing the detail of the install process may help you in debugging problems. Finally, it is instructive to go through the process of the install by yourself, in you have the time and inclination.

## Verifying tranSMART release artifacts

In these instructions, you will be download the tranSMART release artifacts from the tranSMART library. The current release is tranSMART release 16.3, and it is visible in the library at http://library.transmartfoundation.org/release/release16_3_0.html. As you can see from that page, the artifacts are presented as either zip or war files and each has a PGP signed "detached signature", and files containing the MD5 hash and a SHA1 hash. These can be used to verify the file being download. The process for verification is covered separately.

## Installing Support Tools

The support tools are, generally, unix command line tools. In the instructions that follow, we assume that your installing on the latest LTS version of Ubuntu (these instructions were developed and tested on Ubuntu 14.04) and are using the command line interface. Since the final product of these instructions is run in the browser, it is assumed that you have loaded (or are using) the desktop version of Ubuntu (or that of your target OS).

# Setup

Following tools and frameworks are required prior to installing and running TranSMART. These instructions use Ubuntu's apt-get command to install the supporting tools for the subsequent install of tranSMART. We assume that you are installing on a clean Ubuntu OS. To build Ubuntu OS in a VM, see the search for "installing Ubuntu on a VM".

Also, at least in this context, where you are setting up tranSMART for an introduction and exploration, it is a good idea to set up the initial (root) user as the user that will run the application. You can choose the name and password when you do the install, for example the user we use in these instructions is: name = transmart, password = transmart.

## Basics

In this case, for the Ubuntu OS, there is a predefined method for installing rsync, curl, tar, unzip, Java (JDK), php, and PostgreSQL; a make file in the transmart-data archive will load these required tools. However, we will install curl separately, as below, to load and extract **transmart-data;** so that we can run the script. The command line steps below will do the following:

- Create a folder, **transmart**, to contain the app, scripts, and data
- Within that folder, using curl, download https://github.com/tranSMART-Foundation/transmart-data/archive/release-16.3.zip.
- Expand that zip file into **~/transmart/transmart-data**
- Within that folder run the initial setup commands (these instructions are basically from the file ~/transmart/transmart-data/README.md)
- Run additional commands to install **ant** and **maven**

```
mkdir ~/transmart
cd ~/transmart
sudo apt-get install -y curl
curl http://library.transmartfoundation.org/release/release16_3_0_artifacts/transmart-data-release-16.3.zip -o
transmart-data-release-16.3.zip
unzip transmart-data-release-16.3.zip
mv transmart-data-release-16.3 transmart-data
cd transmart-data
sudo make -C env ubuntu_deps_root
make -C env ubuntu_deps_regular
sudo apt-get install -y ant
sudo apt-get install -y maven
```

NOTE: There is an error in the scripts above that installs the wrong version of groovy; this error will need to be fixed in the next release of the scripts. For now, this command-line, coupled with the install of groovy 2.4.5 in the next section is a work-around.

```
rm ~/transmart/transmart-data/env/groovy
```

In the following instructions, it will be necessary to edit configuration files. We are using the visually enhanced version of **vi** called **vim**.  Feel free to use another editing tool. If you wish to closely follow the editing instructions herein, download and install **vim**:

```
sudo apt-get install -y vim
```

For more detailed information on basic OS commands see this "expand" box.

The command-line artifacts loaded in this section are required for the scripts that follow. They are the following: postgresql make git rsync libcairo-dev php5-cli php5-json curl tar openjdk-7-jdk zip unzip

In addition, to compile R from source code you will need: gfortran g++ libreadline-dev libxt-dev libpango1.0-dev libprotoc-dev texlive-fonts-recommended tex-gyre liblz4-tool pv

On Mac OsX the native version of tar does not work, you need gnutar instead.

## Install Grails 2.x

This step will install Grails and Groovy - https://grails.org/download.html - also see http://grails.org/doc/latest/guide/gettingStarted.html

We will use the **SDKman** installer. In the following, set grails 2.3.11 as the default; set groovy 2.4.5 as the default. These can be changed with **sdk use** *<tool>* *<version>*, for example **sdk use grails 2.3.11** .
Note that SDKman works on most major OS types.

```
curl -s get.sdkman.io | bash -c "source $HOME/.sdkman/bin/sdkman-init.sh"
# during the following two commands you will be asked if you want to
# set the installed tool/version to be the default, you do
sdk install grails 2.3.11
sdk install groovy 2.4.5
```

## Install Tomcat

You will need a framework in which to run the transmart war files. Most commonly, for tranSMART exploratory installs, we use **Tomcat7** (version 8 should work as well, it has not been tested on version 9). For more general notes on installing Tomcat in other OS environments, visit the apache tomcat site.

```
sudo apt-get install -y tomcat7
```

The above install will also start tomcat, but we need to shut it down for now; we will start it later in this process.

```
sudo service tomcat7 stop
```

NOTE: There is a problem that occurs when the tranSMART webapp runs in tomcat7 (as it is, initially, installed), the application causes Tomcat to run out of Java Heap Space. The work-around, for now, is to edit the file that sets the default parameters for tomcat. Use an editor, with sudo, to edit the file /etc /default/tomcat7. Find the line that sets the JAVA_OPTS system variable and change it to read:

**JAVA_OPTS="-Djava.awt.headless=true -Xms512m -Xmx2g -XX:+UseConcMarkSweepGC"**

Specifically, using **vim**, enter the command to start vim editing the file and follow the VIM directives given below (**<R>** indicated the RETURN key)

```
sudo cp /etc/default/tomcat7 /etc/default/tomcat7-backup
sudo vim /etc/default/tomcat7
```

(In **vim**, you can type **:q!** at any time to quite without saving your changes.)

Once in **vim**,

- type the single character '**i**'  which puts you into input mode (the characters **- -  Insert - -** appear at the bottom left of the terminal screen).
- Then use the arrow keys to scroll to the line with **JAVA_OPTS** and edit that line to be the one above (by typing, use the delete and backspace keys, positioning with the arrow keys).
- Then type the single character <esc> (the Escape-key) to get out of input mode.
- Double-check your work.
- And then, type **ZZ** to save the file and quit.

## Install R and Rserve

To install R and Rserve type the following (this will take several minutes):

```
cd ~/transmart/transmart-data
cp vars.sample vars
. ./vars
make -C R install_packages
```

NOTE: There is a problem with the fact that R, as installed above, is not on the PATH System variable. The following three commands are a work-around to fix this problem. Eventually, it will have to be fixed in the scripts. These commands add the path for R to the PATH variable (for all processes); this is done by creating an (new) file Rpath.sh and adding it to the /etc/profile.d/ directory.

```
cd ~
echo "export PATH=${HOME}/transmart/transmart-data/R/root/bin:$PATH" > Rpath.sh
sudo mv Rpath.sh /etc/profile.d/
```

This change will take place when you next start the machine, login, or start a new terminal window or shell. To have them take effect immediately you must:

```
source /etc/profile.d/Rpath.sh
```

Check and make sure that you have the correct version of R on your PATH:

```
R --version
```

The version number should be **3.2.1 a**lthough it has been reported that other versions work. If errors occur in advanced workflows, check that the R version is working.

## Install SOLR

SOLR is a search engine that uses Lucene to build fast searches. In our case, it is built on top of Hibernate so that each search targets a pseudo-tables generated by a data-base query. It is used in multiple places in tranSMART. This process runs in java. So, it should be (mostly) independent of OS type.

In these instructions it is installed as a stand-alone web application. See the documentation for a full explanation of other ways in which it can be installed.

Install SOLR by using the transmart-data command line 'make' file:

```
cd ~/transmart/transmart-data
. ./vars
nohup make -C solr start &
```

This first downloads and installs solr, then starts it using the standard configuration for tranSMART in the solr directory. For evidence that it is finished loading, look for the line:

**INFO org.eclipse.jetty.server.AbstractConnector – Started SocketConnector@0.0.0.0:8983**

And once it is running, which takes a few minutes, kill it with **control-C**. You can ignore the resulting error message.

We will restart it, again, later in these instructions.

## The tranSMART Web Application

Set up a basic database that supports login and contains one study. Set up the configuration file. Set up the war files. Start Rserve, SOLR, and Tomcat.

### Setting up a new database and sample data

We use the make files in **transmart-data** to create a new database. These processes should run independent of OS type,

```
cd ~/transmart/transmart-data
. ./vars
make -j4 postgres
```

### Create the tranSMART Configuration Files

To create the configuration files for the tranSMART web application and the connection to the database:

```
cd ~/transmart/transmart-data
. ./vars
make -C config install
```

Unfortunately, these instructions set up the configuration files in the configuration directory for the user **transmart** in the directory **/home/transmart/.grails /transmartConfig**, but they need instead to be in the configuration directory of the user **tomcat7** which is in the directory **/usr/share/tomcat7/.grails /transmartConfig**. So, we copy them

```
sudo mkdir -p /usr/share/tomcat7/.grails/transmartConfig/
sudo cp /home/transmart/.grails/transmartConfig/*.groovy /usr/share/tomcat7/.grails/transmartConfig/
```

### Setting Up The Web Application

From the transmart-library web site, download the application's war files: transmart.v16.3.war and gwava.v16.3.war. We will copy then to a staging directory at ~/transmart/war-files . And then, install then in tomcat's webapps directory as transmart.war and gwava.war. Specifically,

```
cd ~/transmart
mkdir war-files
cd war-files
curl http://library.transmartfoundation.org/release/release16_3_0_artifacts/transmart.war --output transmart.war
curl http://library.transmartfoundation.org/release/release16_3_0_artifacts/gwava.war--output gwava.war
sudo cp *.war /var/lib/tomcat7/webapps/
```

### Loading Data into the Database

The description of loading data is covered on in two separate locations, see the Load Dataset section, above, in the all-in-one instructions, and/or the data loading instructions on the DataCuration page.

# Running SOLR, Rserve, and Tomcat

These instructions will need to be run (checked) each time you start up.

## SOLR

Since SOLR runs in Java, these instructions are easily translated into other OS types. Check to see that SOLR is running with:

```
ps aux | grep start.jar
```

If it is not found, start it with:

```
cd ~/transmart/transmart-data
. ./vars

nohup make -C solr start > ~/transmart/transmart-data/solr.log 2>&1 &
```

This last statement rebuilds all the indexes (should be done after each database load; and with SOLR running as above). You will also need to rebuild the index if you do any editing on the browse page in the tranSMART web application; browse page editing is not covered in these notes.

```
 make -C solr browse_full_import rwg_full_import sample_full_import
```

## Rserve

Check to see that Rserve is running with:

```
 ps aux | grep Rserve
```

If it is not found, start it with:

```
cd ~/transmart/transmart-data
sudo -u tomcat7 bash -c 'source vars; make -C R start_Rserve'
```

NOTE: Rserve must be running under the user **tomcat7**; as above.

## Tomcat

Check to see that Tomcat7 is running with:

```
 sudo service tomcat7 status
```

If is is running, restart it with:

```
sudo service tomcat7 restart
```

If is is not running, start it with:

```
sudo service tomcat7 start
```

# Using the Web Application

Browse to http://localhost:8080/transmart ; log in with username = admin, password = admin ; click on the Analyze tab.

You should see one study "changed".

Drag it to the selection box and click on "Summary Statistics": this will show a summary of statistics for that study.

In this case, transmart is up and running.

# Notes on PostgreSQL 9.3

The steps at "Basics", above, we installed PostgreSQL 9.3; those steps also created the database table-space directories and set the owner of those directories to **postgres**, as required by the data-loading make files used later. See this folder
`ls /var/lib/postgresql/tablespaces`

In addition, those install steps start PostgreSQL and set init files so that PostgreSQL will start on OS restart and after reboot.

This means that there is a **postgresql** service to start/stop and get status of PostgreSQL, specifically,

```
sudo service postgresql start
sudo service postgresql stop
sudo service postgresql status
```

# Using the R Interface (Optional)

Follow the instructions in the README.md file at this location:

https://github.com/tranSMART-Foundation/RInterface/tree/release-16.3

# FAQs

Start with the Sanity Checklist. If that is successful, you can work through the exercises provided in the user guide.

The security group settings should be determined based on the level of access to the data in the system that you want the users and the developers /administrators to have. Here are some suggestions:

**Users:**

- Open HTTP to the IP address range for the users or make it accessible to the world (i.e. 0.0.0.0/0)

**Developers/Administrators:**

- Open SSH to putty or ssh into the box
- Open 8080 for testing of the tomcat-installed application; or the port for HTTP access of your installation
- Open 8983 so you can access the Solr Admin
- Normally the DB ports are used only locally (on the Application server for a single host solution; or between App and DB servers for a multiple host solution); for development these may need to be opened - for SSH tunneling, for example.

You can use a band of IPs if you are unable to get an exact IP address.

# Credits

These install instructions rely on a number of contributions from multiple authors. The various make files in the **transmart-data** repository were originally written by Gustavo dos Santos Lopes (and teammates) at theHyve; later modified by him, Peter Rice (at Imperial College London) and Ruslan Forostianov (theHyve). In addition, instructions on earlier versions of this wiki, from which these instructions draw heavily, were written by Ruslan Forostianov, Janneke Schoots - van der Ploeg, Jan Kanis, Gustavo dos Santos Lopes, and Ward Weistra (from theHyve), with additional contributions from Zach Wright (University of Michigan), Jinlei Liu (ConvergeHealth by Deloitte), Dave John (while at ConvergeHEALTH by Deloitte), Vasudeva Mahavisno (while at University of Michigan), and Terry Weymouth (University of Michigan).