Published in final edited form as:

Curr Protoc Bioinformatics. ; 50: 9.10.1–9.10.10. doi:10.1002/0471250953.bi0910s50.

Installing a Local Copy of the Reactome Web Site and Knowledgebase

Sheldon J McKay and Joel Weiser

Ontario Institute for Cancer Research, Toronto, ON, Canada

Abstract

The Reactome project builds, maintains, and publishes a knowledgebase of biological pathways. The information in the knowledgebase is gathered from the experts in the field, peer reviewed, and edited by Reactome editorial staff and then published to the Reactome Web site, <http://www.reactome.org> (see *UNIT 8.7; Croft et al., 2013*). The Reactome software is open source and builds on top of other open-source or freely available software. Reactome data and code can be freely downloaded in its entirety and the Web site installed locally. This allows for more flexible interrogation of the data and also makes it possible to add one's own information to the knowledgebase.

INTRODUCTION

The Reactome project, described in *UNIT 8.7*, builds, maintains, and publishes a knowledgebase of biological pathways. The Reactome knowledgebase contains a curated collection of well-documented molecular reactions assembled into pathways ranging from intermediary metabolism through signal transduction to complex cellular events such as the cell cycle. The information in the knowledgebase is gathered from the experts in the field, then peer reviewed and edited by Reactome editorial staff. It is then published to the Reactome Web site. The Reactome Web site provides facilities to search and browse the knowledgebase contents as well as to export the data in various formats. Reactions and pathways can be exported in BioPAX and SBML formats, and in automatically created diagrams in several formats. The narrative in Reactome can be exported in RTF and PDF formats. While the Reactome Web site provides free access to the data, the data can also be downloaded in their entirety and the Reactome software installed locally. This allows for more flexible interrogation of the data, and also makes it possible to add one's own information to the knowledgebase. This unit describes setting up your own copy of Reactome knowledgebase together with the Web site software for accessing and viewing the data. Basic Protocol 1 describes automated installation of a local copy of the Reactome Web site. Basic Protocol 2 covers manual installation, and Basic Protocol 3 describes how to use a pre-loaded Amazon Web Services AMI for a cloud-based instance of Reactome.

Basic Protocol 1

Automated installation

If you are setting up the Reactome software and data on a newly configured server instance running either a Debian or Ubuntu Linux operating system, the shell script described in this

protocol can automate the download and installation of all of the necessary components and configure the various system components necessary to serve the Reactome web site. It relies on the Debian package manager to install system software. For non-Debian based Linux distributions or different server configurations, see Basic Protocol 2 for information on how to perform a manual installation.

Necessary Resources

Hardware: A computer or virtual machine running a Debian or Ubuntu Linux operating system, an internet connection, and at least 4 Gbyte free disk space. Processing power requirements depend on the planned use of the installation. 8 Gbyte or more of RAM is recommended. Administrative (root) or sudo access to the installation machine is required. If an Amazon EC2 instance is being used, the *m3.large* instance type is recommended.

Software: Minimally, wget and the apt-get package manager are required to begin the installation. These are normally available by default in Debian or Ubuntu Linux. Either the Debian 6 (or later) or Ubuntu 12.04 (or later) Linux distributions are recommended.

1. Use wget to download the required script from the reactome website.

```
wget http://reactome.org/download/current/install_reactome.sh
```

2. Run the script using the BASH interpreter.

The command below must either be run as root or using 'sudo'

```
bash install_reactome.sh
```

3. You will be prompted to enter a MySQL root password. Enter the password and hit the Enter key.

If there is an existing mysql installation, enter the root password for that installation. You can skip creating a mysql root password by hitting Enter without a password

.

```
IMPORTANT: If you have already set up MySQL, provide the
mysql 'root' user (administrator) password here.
Otherwise, the password you enter here will be
configured as the mysql root password.
The password can be left blank if desired
Enter the password for mysql user 'root' [default: none]:
```

4. Installation is now done. Enter the IP or web address of your Linux server into the address field of your web browser and your local installation of the Reactome web site (Figure 9.10.1) should load.

BASIC PROTOCOL 2

MANUAL INSTALLATION

The shell script described in Basic Protocol 1 automates the installation process using the steps described below. It assumes that you are using a newly configured server that is running either a Debian or Ubuntu Linux operating system. If you are performing the installation for a different Linux distribution, or on a server that has already been configured for other applications, it may be necessary to install the software manually in order to modify installation paths, database access credentials or Apache configuration. In any case, reading through this protocol will provide a better understanding of Reactome's components in case it is necessary to debug or modify the installation at a later date.

Necessary Resources

Hardware: A computer or virtual machine running a Debian or Ubuntu Linux operating system, an internet connection, and at least 4 Gbyte free disk space. Processing power requirements depend on the planned use of the installation. 8 Gbyte or more of RAM is recommended. Administrative (root) or sudo access to the installation machine is required. If an Amazon EC2 instance is being used, the *m3.large* instance type is recommended.

Software: Minimally, wget and the apt-get package manager are required to begin the installation. These are normally available by default in Debian or Ubuntu Linux. Either the Debian 6 (or later) or Ubuntu 12.04 (or later) Linux distributions are recommended.

Installing the reactome software

- 1 Create the path for the reactome web site.

The path below mirrors the one used on the reactome production servers. This and subsequent steps in this protocol are to be performed as the root user or using 'sudo'

```
mkdir -p /usr/local/reactomes/Reactome/production
cd /usr/local/reactomes/Reactome/production
```

- 2 Use wget to download the reactome web site software.

The reactome.tar.gz file is 700 Mbytes in size contains a compressed archive all of the website software components

```
wget http://reactome.org/download/current/reactome.tar.gz
```

- 3 Install the software.

```
tar xzf reactome.tar.gz
rm -f reactome.tar.gz
```

After the file reactome.tar.gz is decompressed, the directory tree will be as shown below. The AnalysisService and Solr directories contained serialized data used in the analysis and search components of the web site, respectively. The apache-tomcat directory contains the Java webapp components of the website and GKB contains the Perl and wordpress parts of the website. See Figure 9.10.2 for an architecture diagram

```
.
├── AnalysisService
├── apache-tomcat -> apache-tomcat-7.0.50
├── apache-tomcat-7.0.50
├── GKB
├── RESTful
└── Solr
```

For convenience, set up a symbolic link to the reactome software

```
cd /usr/local
ln -s reactomes/Reactome/production/GKB gkb
```

4 Install the configuration files.

Configuration files contain database and other access credentials and website configuration. By default, the access credentials are pre-configured to work without modification

```
cd /
tar xzf /usr/local/gkb/third_party_install/config.tar.gz
```

5 Set up temporary website folders.

```
cd /usr/local/gkb/website/html
mkdir img-fp
mkdir img-tmp
chown -R www-data img-*
```

Installing software dependencies

6 Use the apt-get package manager to install or update third-party software and associated dependencies.

The ‘\’ character shown below will be interpreted by the BASH shell to indicate that the command continues on the same line. During the installation process, you will be prompted to set a mysql root password if mysql is not already installed

```
.

apt-get clean
apt-get update
apt-get install \
  build-essential \
  perl \
  curl \
  mysql-server \
  apache2 \
  libexpat1 \
  libexpat1-dev \
  php5 \
  php5-mysql \
  libbio-perl-perl \
  libgd-gd2-perl \
  openjdk-7-jre-headless \
  openjdk-7-jdk
```

- 7 The CGI scripts expect Perl to be installed in /usr/local/bin. If necessary, use a symbolic link to create this path.

```
cd /usr/local/bin
ln -s /usr/bin/perl
```

- 8 Install the necessary remaining Perl modules and associated dependencies from CPAN (cpan.org) using the cpanminus application.

```
curl -L http://cpanmin.us | perl - --sudo App::cpanminus
cpanm -q \
HTTP::Tiny \
IO::String \
LWP::UserAgent \
MIME::Lite \
Net::OpenSSH \
XML::Simple \
Search::Tools \
Capture::Tiny \
WWW::SearchResult \
```

JSON \
 PDF::API2

Setting up the MySQL databases

- 9 Use wget to download the reactome databases.

The databases are compressed mysql dumps available from the reactome website

.

```
cd
wget \
http://reactome.org/download/current/databases/gk_current.sql.gz
wget \
http://reactome.org/download/current/databases/gk_wordpress.sql.gz
wget \
http://reactome.org/download/current/databases/gk_stable_ids.sql.gz
```

- 10 Create or re-initialize the databases.

If you have set a MySQL root password, use the -p flag in the commands below

.

```
mysql -uroot [-p] -e \
'DROP DATABASE IF EXISTS gk_current;
CREATE DATABASE gk_current;
DROP DATABASE IF EXISTS gk_stable_ids;
CREATE DATABASE gk_stable_ids;
DROP DATABASE IF EXISTS gk_wordpress;
CREATE DATABASE gk_wordpress;'
```

- 11 Load the databases.

These commands will take a while to run. Make sure you have at least 4 Gbyte of available space on the file system that contains the mysql databases (default: /var/lib/mysql). [-p] is only required if you have configured a root password for mysql

.

```
zcat gk_current.sql.gz | mysql -uroot [-p] gk_current
zcat gk_wordpress.sql.gz | mysql -uroot [-p] gk_wordpress
zcat gk_stable_ids.sql.gz | mysql -uroot [-p] gk_stable_ids
```

12 Set up database access permissions.

By default the reactome website components will use the database access credentials below. [-p] is only required if you have configured a root password for mysql

```
mysql -uroot [-p] -e "GRANT SELECT ON gk_stable_ids.* TO
'reactome_user'@'localhost' IDENTIFIED BY 'reactome_pass'"
mysql -uroot [-p] -e "GRANT SELECT ON gk_current.* TO
'reactome_user'@'localhost' IDENTIFIED BY 'reactome_pass'"
mysql -uroot [-p] -e "GRANT ALL ON gk_wordpress.* TO
'reactome_user'@'localhost' IDENTIFIED BY 'reactome_pass'"
```

Setting up the Apache web server**13** Make sure the apache2 web server has the required modules enabled.

```
cd /etc/apache2/mods-available
a2enmod \
mime \
include \
autoindex \
dir \
cgi \
alias \
proxy \
proxy_http \
rewrite
```

14 Install the reactome configuration file.

The setup shown here assumed that this is a new apache2 installation and no other websites are being served by this host. If this is not a new apache installation and other websites are running on this server, it may be necessary to adjust the virtual host configuration (see <http://httpd.apache.org/docs/2.2/vhosts/>)

```
cd /etc/apache2/sites-available
cp /usr/local/gkb/website/conf/http.conf reactome.conf
a2dissite default
a2ensite reactome.conf
```

- 15** If the apache2 software version is 2.4+, there is a syntax change in the configuration.

A simple test for apache version 2.4+ is the presence of the path below, which did not exist prior to version 2.4

```
ls /etc/apache2/conf-available
```

If necessary, use a text editor to modify /etc/apache2/sites-available/reactome.conf to uncomment the line below (it occurs twice in the file) by removing the '#'

```
#Require all granted
```

- 16** Restart the apache web server.

```
/etc/init.d/apache2 restart
```

Setting up the Apache Tomcat server

- 17** Create a tomcat7 user and group.

The tomcat7 user will own the files used by the java components of the website and control the tomcat server process once it has been started

```
groupadd tomcat7
useradd -g tomcat7 -s /sbin/nologin -d /opt/tomcat/temp tomcat7
```

- 18** Change permissions to allow the tomcat7 user to access files.

```
cd /usr/local/reactomes/Reactome/production
chown -R tomcat7:tomcat7 \
apache-tomcat-7.0.50\
Solr\
AnalysisService\
RESTful
```

- 19** Configure tomcat to start by default on system boot.


```
update-rc.d tomcat7 defaults
```

- 20 Start the tomcat server.

```
/etc/init.d/tomcat7 restart
```

- 21 Installation is done. Enter the IP or web address of your Linux server into the address field of your web browser and the reactome web site (Figure 9.10.1) should load.

BASIC PROTOCOL 3

USING THE PRE-CONFIGURED REACTOME AMAZON WEB SERVICES AMI

If you are more interested in using a Reactome instance rather than installation or customization, it is very convenient to launch a pre-configured virtual machine (an Amazon Machine Instance in this case). This protocol describes how to use Amazon Web Services to launch a cloud-based instance of Reactome.

Necessary Resources

Hardware: An *m3.large* Amazon EC2 instance type.

Software: A web browser and ssh client.

A pre-loaded, cloud-based instance of Reactome is available as an Amazon EC2 AMI. Visit <http://aws.amazon.com/ec2/> if you are new to Amazon EC2.

See <http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/launching-instance.html> for instructions on how to launch an instance of an amazon AMI.

1. Sign on to amazon AWS.
2. Go to the EC2 console (<https://console.aws.amazon.com/ec2/v2>)
3. Select the “N. Virginia”, Oregon, Ireland or Singapore availability zone using the pop down menu at the upper right of the screen. Reactome AMIs are available in each of these zones.
4. Click on the **Launch Instance** button.
5. Click on **Community AMIs**.
6. Search for the keyword **reactome**.
7. Click the **Select** button next to the desired reactome AMI.
8. On the left panel, click on **General purpose** to select an instance size. Choose **m3.large**.
9. Follow the remaining steps under **Continue...** button on the bottom right.

10. Select/create a security group that allows your connections to port 22 (ssh), 80 (apache2) and 8080 (apache tomcat).
11. Launch the instance.
12. The EC2 instance show up in the **Instances** panel of the console. Once it is running, select the instance to retrieve information about the instance including its public IP address.
13. The Reactome web site will be available by entering the IP address for your EC2 instance in a web browser.

COMMENTARY

Background Information

The concept of a pathway knowledgebase is not a novel one and there are numerous sources offering information under various access terms ranging from free-for-all to paying-subscriber only. However, the feature that distinguishes the Reactome project from many of its peers is that, in addition to freely accessible data, it also offers the possibility to download and replicate the whole knowledgebase and Web site. While the Reactome project attempts to provide easy access to various bits of information in various formats, having a local copy of the knowledgebase and API code gives the ultimate freedom and flexibility to extract whatever is necessary. While the Reactome project's own curation efforts concentrate mainly on human biology, the setup can be used to annotate biochemical processes of any cellular organism. Indeed, the Reactome project also produces orthology-based computational predictions of pathways in numerous other organisms. These can be used as a starting point for manual curation of pathways in other species. The Reactome Curator Tool, available from the Reactome download page at <http://www.reactome.org/download/>, is a stand-alone Java application that allows users to edit existing knowledgebase entries and to enter new information. The same Web page also offers access to the Reactome Author Tool, which provides a more graphical way to enter and edit the information and hides many of the intricacies of the Reactome data model. However, in order to write the information assembled in the Author Tool back to the knowledgebase, one has to use the Curator Tool. The Reactome project also makes available Perl and Java APIs for accessing the data in the knowledgebase. The Perl API comes as part of the Web site and code download, while the Java API is available as part of the Curator Tool installation. Although both of them are extensively used internally by the Reactome project, their documentation is limited; therefore, they should be approached only by individuals who are comfortable with writing software. Both the software developed as part of the Reactome project and the external software used by Reactome installation are open source and freely available. All website components are available on GitHub (github.com/reactome). An architectural diagram of the software is shown in Figure 9.10.1.

Critical Parameters and Troubleshooting

The instructions presented in this unit assume that the user has root privileges on the computer where the local copy of Reactome is being installed. These privileges are required for installation of software at system-wide locations, as well as for starting up the Web

servers. For the local installation of Reactome to work, both the Web and database servers have to be running. Perl has to be located at (or be symbolically linked from)/usr/local/bin/perl. The most useful resources for resolving issues with Reactome installation are the error messages that appear on the command line during installation. Also check the Web server error log file (/usr/local/gkb/website/logs/error.log) and the tomcat logs (/usr/local/reactomes/Reactome/production/apache-tomcat/logs) for other error messages. When requesting help, please describe the problem and paste all error messages into an email sent to help@reactome.org mailing list.

Acknowledgments

The Reactome project is supported by the National Human Genome Research Institute at the National Institutes of Health [P41 HG003751]; the Ontario Research (GL2) fund; the European Bioinformatics Institute; the European Commission (PSIMEx); Genome Canada; Google Summer of Code Program (2011–2013) and ORCID [RFP 2013-06-09].

Literature Cited

1. Croft D, Fabregat Mundo A, Haw R, Milacic M, Weiser J, Wu G, Caudy M, Garapati P, Gillespie M, Kamdar MR, Jassal B, Jupe S, Matthews L, May B, Palatnik S, Rothfels K, Shamovsky V, Song H, Williams M, Birney E, Hermjakob H, Stein L, D'Eustachio P. The Reactome pathway knowledgebase. *Nucleic Acids Res.* 42:D472–7. (Database issue). 10.1093/nar/gkt1102 [PubMed: 24243840]

REACTOME
A CURATED PATHWAY DATABASE

About Content Documentation Tools Community Download Contact

e.g. O95631, NTN1, signa Search

Browse Pathways **Analyze Data** **Reactome FI Network**

User Guide **Data Download** **Contact Us**

About Reactome

Reactome is a free, open-source, curated and peer reviewed pathway database. Our goal is to provide intuitive bioinformatics tools for the visualization, interpretation and analysis of pathway knowledge to support basic research, genome analysis, modeling, systems biology and education. The current version (v50) of Reactome was released on October 8, 2014.

Logos: OICR, NYU Langone MEDICAL CENTER, CSH Cold Spring Harbor Laboratory, EMBL-EBI

The development of Reactome is supported by a grant from the US National Institutes of Health (P41 HG003751), Ontario Research Fund, and the European Molecular Biology Laboratory.

Tweets

Current Version: Reactome hits 50

reactome @reactome 24 Oct

Go robots @TheTechMuseum
pic.twitter.com/maqfsWUcfR

Expand

Tweet to @reactome

About
About Reactome
News
Reactome Team
Scientific Advisory Board
Other Reactomes
License Agreement
Reactome Disclaimer

Content
Table of Contents
DOIs
Data Schema
Editorial Calendar
Statistics

Documentation
User Guide
Data Model
Orthology Prediction
Object/Relational Mapping
Wiki
Linking to Reactome
Referencing Reactome

Tools
Pathway Browser
Analyze Data
Species Comparison
Reactome FI Network
Advanced Search
Small Molecule Search
Analysis Service

Community
Reactome Outreach
Reactome Events
Reactome Training
Reactome Publications
Papers Citing Reactome
Resources Guide
Mailing List

Download **Contact**

f t You Tube

Figure 1.
The front page of the Reactome website.

Figure 2.

Architecture diagram of the Reactome software. Shaded boxes represent different server types. End users usually access Reactome knowledgebase via a web browser. Content is accessed or analyzed via various outward facing tools (square boxes). The pathway analysis tool uses performance-optimized serialized data files. The search tool is an implementation of Apache Solr, which uses serialized index files and also queries the main knowledgebase via a Java API. The pathway browser is a graphical tool that renders pathways diagrams using data obtained via the RESTful API, which uses a Java API to query the main knowledgebase. The Pathway browser is also a portal to the Analysis tool. The RESTful API is used internally to access the main knowledgebase and is also publically accessible. The Reactome web is served via a combination of the wordpress framework and Perl/CGI scripts. Wordpress uses its own database and the CGI scripts access the main database as well as the stable identifier database.