8.1 Back-End Access Via Insight

You may have existing reporting/visualization/statistical analysis tools you want to use to track, analyze, or report on your OpenClinica data. The most common method for doing this is using downloadable data extracts from the OpenClinica user interface. However several other options exist, including using OpenClinica's API, using the OpenClinica Insight reporting tools, using Insight's API, or by establishing a direct read-only connection to the Insight data warehouse. This guide provides an overview of accessing data through a direct read-only connection to the Insight data warehouse. For existing institutional toolchains, such as SAS, Qlik, Tableau, SSRS, Shiny, Jupyter, Stata, R, Python, Power BI, etc, the customer can establish a direct, live, readonly connection with the Insight database. This method can also be used to incorporate your OpenClinica data into other databases via Postgres Foreign Data Wrappers (FDWs, or linked tables). Since the Insight back-end is a PostgreSQL database, allowing connections is a matter of setting up a secure, read-only database connection.

Pros:

- Works where SSH does: Linux, Windows (PuTTY), etc.
- Use the tool of your choice, so long as it supports remote connections
- Live access to data, not snapshots as in data extracts (can be pro & con)

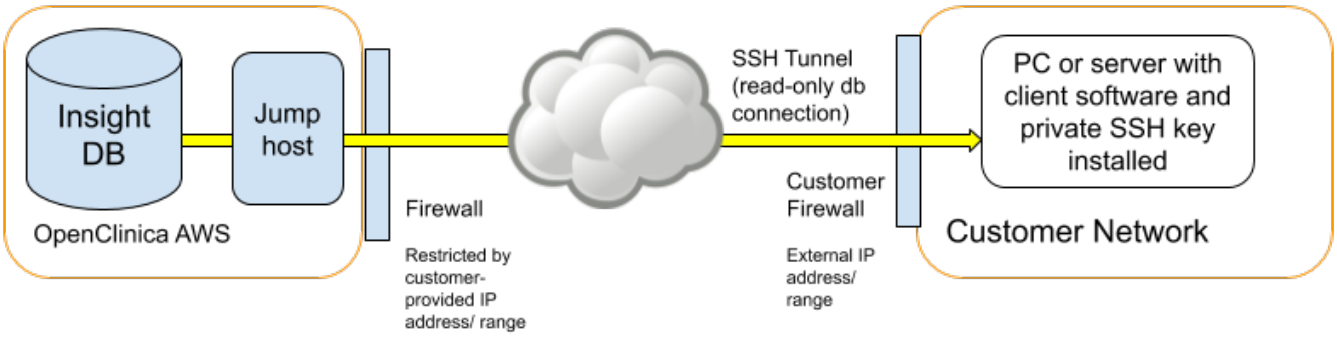
Cons:

- Mostly static data permissions (schema-level)
- User management is separate from the OpenClinica and Insight web interfaces
- OpenClinica team will support setup and maintenance of your connection, however, assistance with connectivity and use of your particular tool (such as SAS or Tableau) is best-effort. We are not able to support the wide variety of possible clients (SAS, Tableau, Power BI, etc) and cannot guarantee end-to-end connectivity, functionality, or level of service.

There are Three Main Parts of the Set Up:

1. Securing the connection itself
2. Authenticating the user that is connecting
3. Defining what the user is authorized to do

We do this with a secure, read-only, direct interface to the Insight Data Warehouse, via an SSH tunnel. It will require the involvement of your internal IT team



Technical Information

Connections to the Insight back-end (PostgreSQL) database are secured by using a SSH tunnel via a jump host. A "jump host" means that your Insight server is not directly reachable from the Internet, so instead connections are made via an intermediate server that is reachable from the Internet but from which connections are allowed to the Insight server. A "SSH tunnel" means that the nature of the connection to Insight will be via a SSH (secure shell) session between your computer (or server) and the jump host. This SSH session is then configured to forward database connection requests to a certain local port (such as 65432) to the Insight server's and database port (such as 5432). Database traffic forwarded in this way is secured. The exact configuration of back-end connections on your side depends on the intended setup. A SSH client will be required on your computer or server. On Linux, SSH is installed by default (ssh). On Windows, there is a SSH tool called "PuTTY" which has a user friendly interface. Alternatively on Windows, "Git for Windows" (Git is a change tracking tool) comes with a Windows version of Linux's SSH tool.

To Set Up Back-End Access Via Insight:

1. Consider the pros and cons and decide if a back-end connection is the right solution for you.
2. Submit a support ticket to the OpenClinica team to get the process started. Include:
    - The IP addresses from which you will connect. This could be your computer IP, or a server IP. This is used to restrict access to the jump host to only those nominated IP addresses, for additional security.
    - The user names and a "PEM" format SSH public key (see below) for the user(s) that will connect. If you are setting this up for a server then there could be just one "service account" type user, or there could be multiple individual users.
    - For each user, which study environment(s) / schemata that they should have access to. The Insight "databases" shown in the Metabase front-end are actually database schemata. The default is access to all schemata, but it is possible to assign read-only access to one schema or many.
    - The public key. These are files that end with ".pub" Please do not share the private key(s).
    - Steps to set up the SSH with your preferred tool.
3. Once the OpenClinica team has set up the connection, test it with your preferred tool.

The Following Links Describe How to Create SSH Keys:

- Linux/Mac: https://www.digitalocean.com/docs/droplets/how-to/add-ssh-keys/create-with-openssh/
- Windows: https://www.digitalocean.com/docs/droplets/how-to/add-ssh-keys/create-with-putty/

Please be sure to use a passphrase for your keys so that they cannot be used if they fall out of your possession. Once the SSH tunnel is in place, it will be possible to connect to the Insight back-end PostgreSQL server. If using ODBC, the "psqlODBC" driver will need to be installed on the connecting computer/server (https://odbc.postgresql.org/). This driver allows applications to talk to PostgreSQL in the ODBC protocol. Then, either a DSN or connection string can be set up in the desired application. We can provide examples of these with your settings.

Approved for publication by Kerry Tamm. Signed on 2020-12-01 9:17AM

Not valid unless obtained from the OpenClinica document management system on the day of use.