

CPCSSN Operating Procedures

SOP number:	CPCSSN-SOP-SRE_002
Version number:	1.00
Category	CPCSSN Secure Research Environment-User Guide
Effective Date:	23 November 2021

References

References
No References

Revision History

Version	Description of Change	Author	Date
NA	NA	NA	NA

Approvals

	Name, Surname, Title	Signature	Date
Authored by	Don White		22 nd March, 2021
Reviewed by			22 nd March, 2021
Approved by	Sabrina Wong		23 rd March, 2021

Status

Disclosure (Internal, External)	Internal, External
Confidentiality (Strictly, Confidential, Public)	Confidential
Status (Draft, Under review, Final)	Final

Table of Contents

Technical Requirements for SRE Access 2

Access – Logging into SRE 4

Connecting to the Database and Accessing / Importing data in your applications..... 10

Linking Tables in your Application 14

Purpose

Provide a step-by-step process to internal and external users to setup a virtual private network (VPN) to access the CPCSSN Secure Research Environment (SRE).

Scope

The document serves as a guide to internal and external users of the CPCSSN SRE.

Technical Requirements for SRE Access

The SRE is a self-contained secure Windows project workspace that is provisioned with all of the tools required to perform data analysis on data. That said there are a couple of components that a researcher will need to have installed on their local workstation to access the SRE. If your IT Department has policies or restrictions on installation of software or administrative privilege on your computers, it will be necessary to obtain their permission and assistance with the installation of OpenVPN and/or the Microsoft Remote Desktop for Mac.

As all processing is done on the SRE virtual desktop, the requirements for the user's computer are minimal. A Windows or Macintosh PC with the following free software is required:

OpenVPN client

The SRE employs a Virtual Private Network, or VPN, to create a secure, encrypted connection between the user's computer workstation and the SRE project workspace. Connecting to the VPN requires a VPN client be installed on the workstation. For Windows users this is the OpenVPN Client for Microsoft Windows. For Mac users this is the Tunnelblick client for OpenVPN.

- Windows - <https://openvpn.net/community-downloads/>
- Mac - <https://tunnelblick.net/downloads.html>

OpenVPN configuration file

The OpenVPN software requires a configuration file provided by the Queen's University Centre for Advanced Computing (CAC). This is sent to the end user as a ZIP archive containing the configuration file, as is a digital certificate called an SSH Key. Both of these are needed to identify the end user and to connect to the VPN.

Remote Desktop Client (RDP Connection)

The VPN software provides the network connection but a Remote Desktop Connection provides the user with a graphical user interface (Windows) to interact with the SRE. For Windows users, no additional software is required as the Remote Desktop Connection applet is included in all modern versions of Windows. For Mac users, the Remote Desktop Connection application can be downloaded from the iTunes store is required.

- Mac App Store - <https://apps.apple.com/app/microsoft-remote-desktop/id1295203466?mt=12>

SRE User Experience

The SRE project workspace should be familiar to users who have done data analysis in a Windows environment in the past. There are, however, a number of Windows features that have been disabled or restricted to enhance the security of the SRE

- Internet – there is no access to the Internet from the SRE workspace.
- Cut and Paste – users are able to cut and paste between applications running on the SRE – for example, between a text editor and an analysis or programming environment, such as SAS Studio or Jupyter Notebook. However, cut and paste between the SRE workspace and the user's remote workstation (and vice-versa) are disabled for security purposes.
- Copying of files – there is no facility for end users to copy files or data to or from the SRE. If users require data, programming source code, macros, or other working files copied into the SRE or need analysis results copied out of the SRE, the system administrator can enable and facilitate a process to do this.
- Software installation – users are not able to install their own software on the SRE. If an additional analysis or programming tool is needed, please contact the SRE System Administrator and / CPCSSN Operations Manager.

- Data Access – with the exception of flat file datasets which are stored on the SRE Windows Project Workspace, datasets are generally stored in a relational database (Microsoft SQL Server). Accessing data in the database requires a connection to the database.

Access – Logging into SRE

An approved researcher will need 3 sets of login credentials to access the SRE.

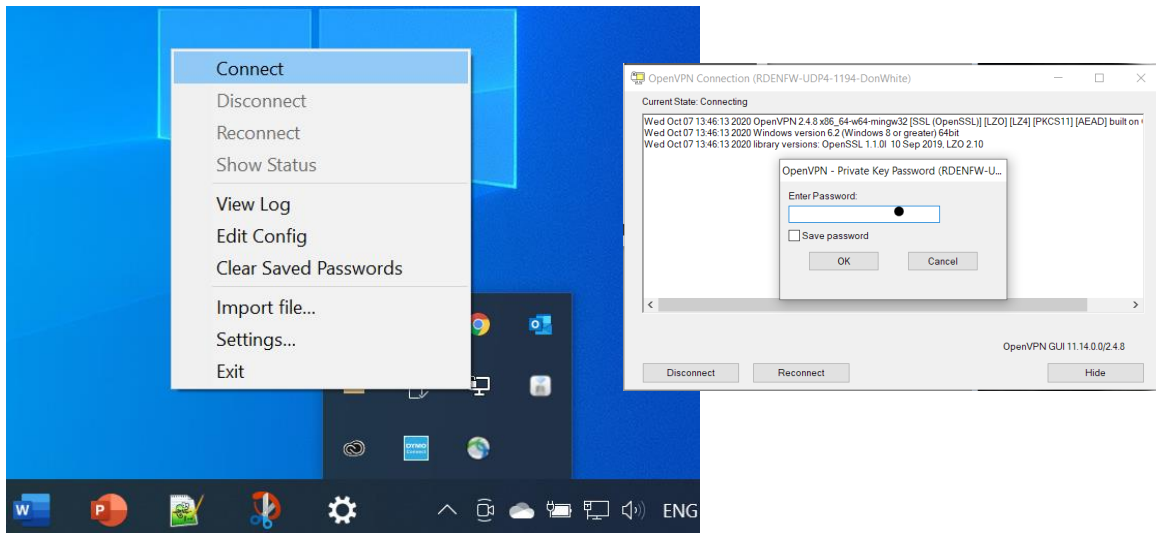
1. Queen’s CAC / CPCSSN SRE Virtual Private Network – this provides a secure, encrypted connection to the SRE network.
2. CPCSSN SRE Windows front end Virtual Machine
3. CPCSSN SRE Database user login and connection information. If your data are in a flat file (CSV, tab delimited, etc.) database credentials are not necessary.

Queen’s University / CPCSSN SRE Virtual Private Network

Once your request for access is approved, the Queen’s University Centre for Advanced Computing (CAC) will create a VPN Login Configuration (config) file and password for you. The config file will be emailed directly to you in a zip file and the password sent by text message.

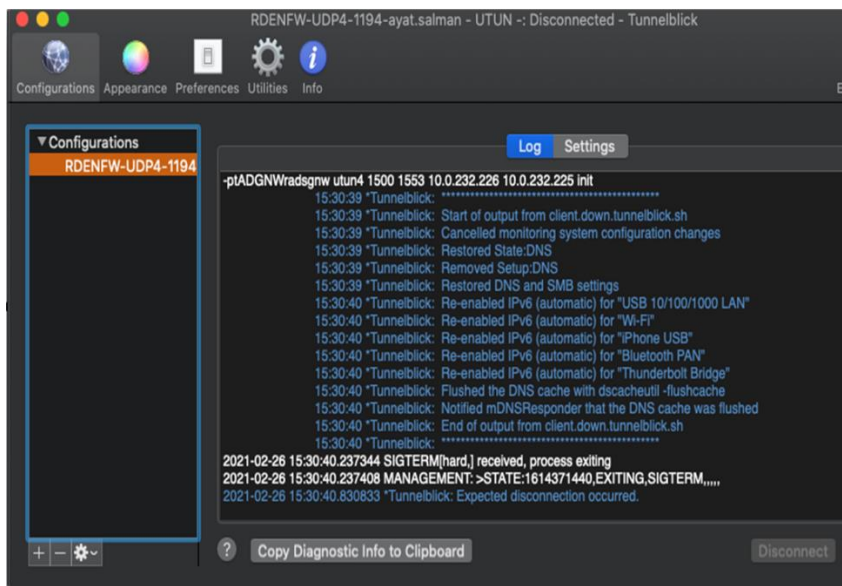
Windows Users Instructions

1. Download the OpenVPN Windows Client
<https://openvpn.net/index.php/open-source/downloads.html>
2. Install OpenVPN free community edition
3. Navigate to C:\Program Files\OpenVPN\Config and extract the contents of the ZIP archived configuration file provided by the CAC into this folder
4. Run OpenVPN Client as admin (right click OpenVPN client icon and select “Run as Administrator”)
5. As soon as OpenVPN icon appears in the bottom right corner of the screen (on the taskbar), right-click it and select Connect.
6. A connection dialog box will pop up and you will be prompted to enter your VPN password (provided by Queens CAC)



Mac Users Instructions

1. For macOS installation, we recommend using Tunnelblick client for OpenVPN. Please download the latest stable version from: <https://tunnelblick.net/downloads.html>
2. The following video provides step-by-step instructions for installation: <https://www.youtube.com/watch?v=FieduQ2hDPc>
3. Once TunnelBlick is installed, unzip the config file provided by the Queen's Centre for Advanced Computing (CAC). It contains a certificate that is password encrypted. The CAC will text you the password required to use the certificate.
4. Once the file is unzipped, drag the **.ovpn** file to the TunnelBlick icon located in the upper right-hand corner of your screen.
5. When installation is complete, press the "Connect" button. A secure connection to the VPN will be established.



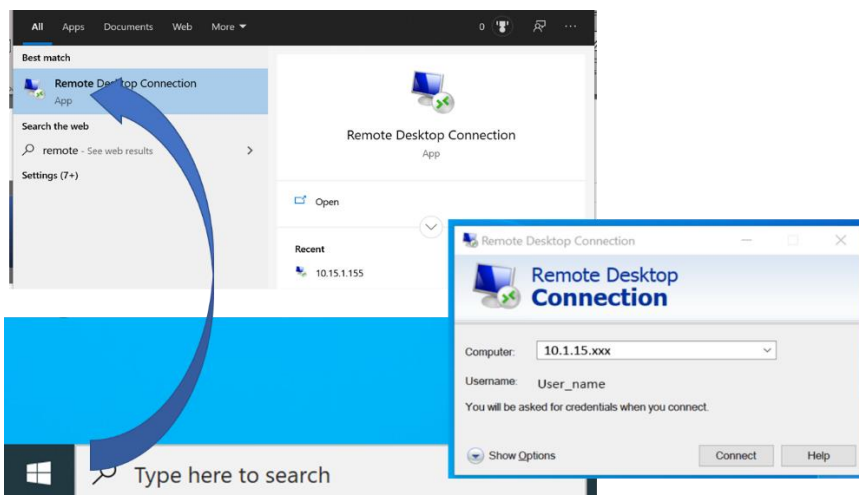
If you have followed these instructions, OpenVPN has created an encrypted “tunnel” connection between your computer and the SRE Network.

Connect to the Remote Server – Microsoft Remote Desktop Connection

OpenVPN has now provided the network connection. A remote Windows session is now required to access and interact with the SRE Windows project workspace. The Microsoft Remote Desktop Connection App is needed for this. Your account on the SRE Windows project workspace has both a username and password. The CPCSSN SRE System Administrator will email you the username and send your password by text message.

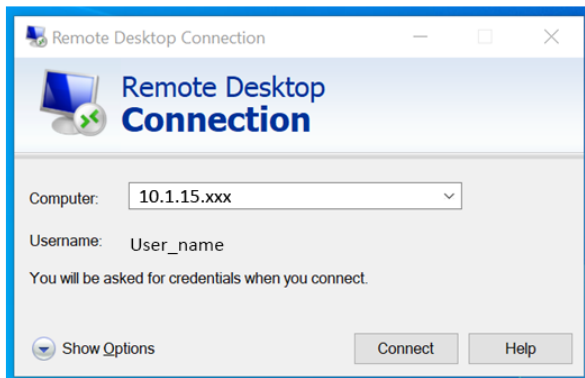
Windows Users Instructions

1. Click on the Windows Start Icon and start typing “Remote”
2. When the Remote Desktop Connection App comes up, click on it.
3. A connection dialog will be displayed.



4. Enter the Computer IP Address. This is unique to each SRE Project and is provided by the SRE Systems Administrator:
 - **10.15.1.xxx**
5. Enter your SRE username
 - If a different username is populated, select Show Options and then enter your SRE username provided by the SRE System Administrator

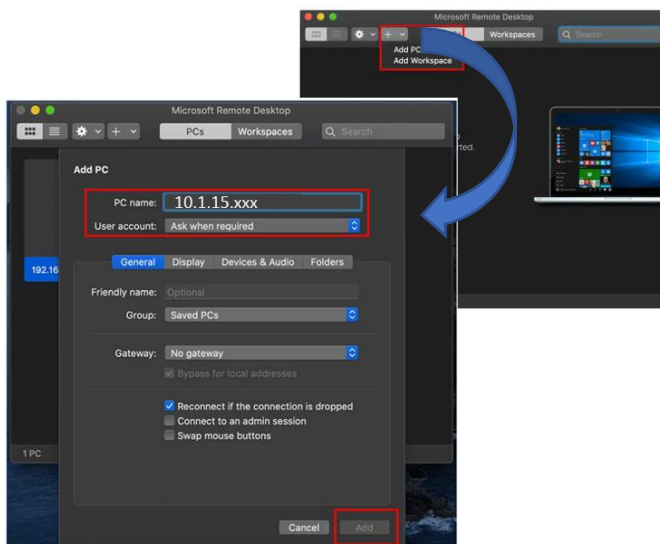
- You will be prompted for your password when you log in:



Note that the Computer IP Address is unique to each project. The username is provided by the SRE System Administrator.

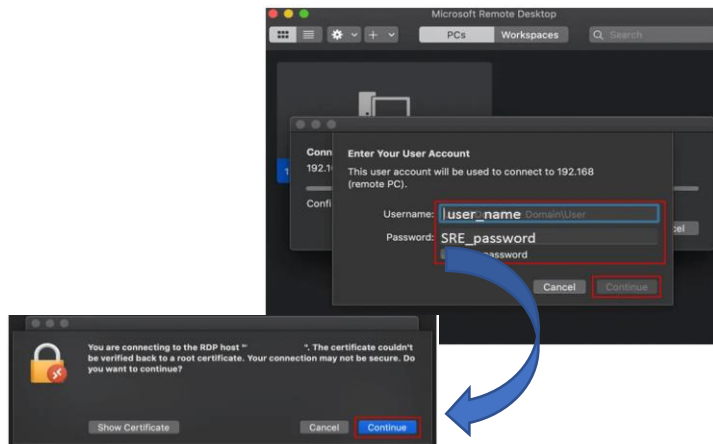
Mac Users Instructions

- Download the free Microsoft Remote Desktop for macOS from the Mac App Store.
- Once installed, open the application and click on the “+” button and add a new desktop or PC and fill out the Add PC form.
- PC Name: IP Address provided by the SRE Administrator.
 - **10.1.15.xxx**
- User account: SRE username provided by SRE Administrator.
- Provide a memorable name in the Friendly Name box. Click Add once the settings are confirmed.



- Your connection will appear, saved and ready to connect in the main Remote Desktop for Mac window. Double-click on the entry to begin your connection. You’ll be asked to provide the username and password for your Windows PC at this point. Provide them, then click Continue to connect.

7. If it's your first connection, you may need to accept a security warning. Click Continue to disregard the message and make the connection.



This will bring you to a Windows login page. Log in using your Windows credentials.



For convenience, your Windows desktop environment has been prepared with shortcuts to commonly used analytical and programming tools on the Desktop. Additionally, if your project needs to share data or programming files between multiple users, a shortcut to a shared folder has also been placed on the desktop.

Connecting to the Database and Accessing / Importing data in your applications

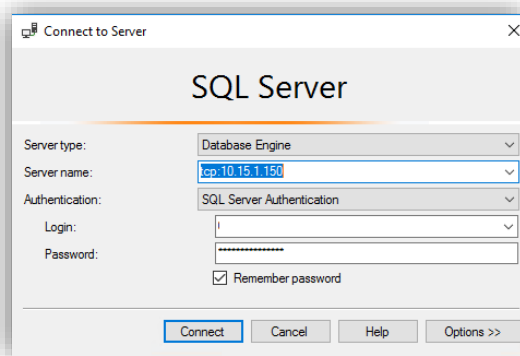
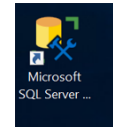
Your data are stored in an MS SQL / SQL Server database on the SRE. The CPCSSN SRE System Administrator has created a database user and password for you and configured an Open Database Connectivity (ODBC) connection in Windows for you.

The procedure to connect to the database and query or export results is specific to each application, but the Windows ODBC Connection provides a common interface for all programs.

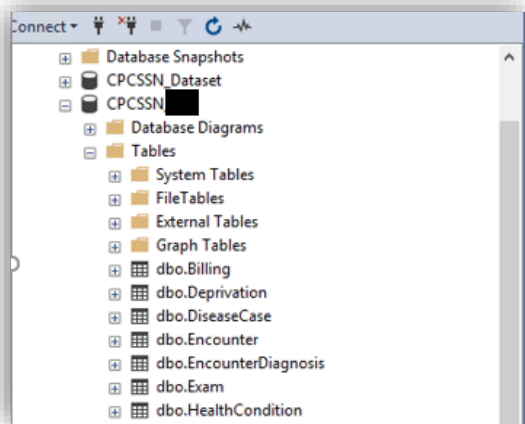
If your project's dataset is in a flat file (csv, tab delimited, etc.) format, a database connection is not required. Your dataset will be copied to your shared folder – accessible by a shortcut on the SRE Windows desktop.

Microsoft SQL Server Management Studio (SSMS)

1. Launch SSMS from the Windows start menu and open a new connection.
2. At the login page put in the server name **tcp:10.15.1.150**.



3. Then put in the database login and password that was provided to you.
4. Once you have successfully logged in, you should now be able to see a list of all the tables under your database, listed on the left hand side under 'Tables'.



Note: Your database will have a unique name and may contain a different subset of tables than shown.

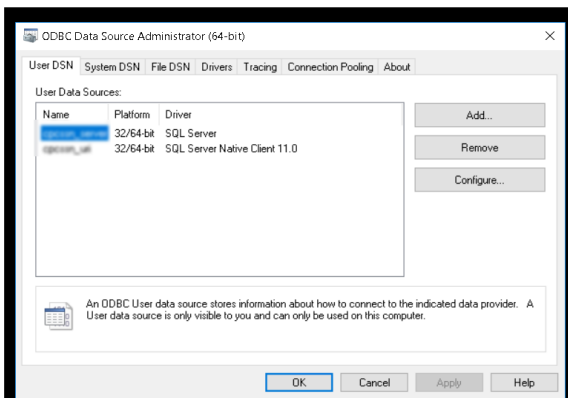
Open Database Connectivity (ODBC)

Open Database Connectivity (ODBC) is an Application Programming Interface (API) that provides a standardized set of libraries to enable applications to connect to and access data in a database. For most SRE users, ODBC is the easiest method to connect to the database.

The data is stored in SQL tables in a relational database or as provided. The best way to access the SQL tables in your statistics software (ex. SAS, R, SPSS, or Stata), while maintaining the integrity of the data, is to set up an ODBC connection. The SRE System Administrator has set up an ODBC connection for you, and the name of your ODBC data source will have been provided to you.

To Find name of pre-configured ODBC Connection:

1. From the Windows Start menu launch PDBC Data Sources (64 bit)



Your connection name will be shown next to the **SQL Server driver**.



SAS Studio

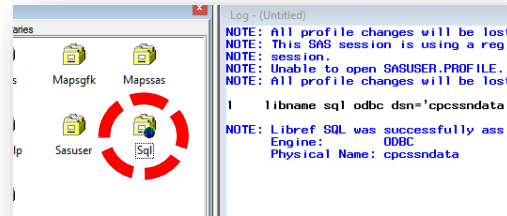
1. Open up SAS
2. In a new program window create a Library that points SAS to the SQL database (this code is also described in the SRE Documentation)

```
LIBNAME ODBC SQL DSN='cpcssn_uri' user=uuuuuu pw=ppppppp schema=dbo
```

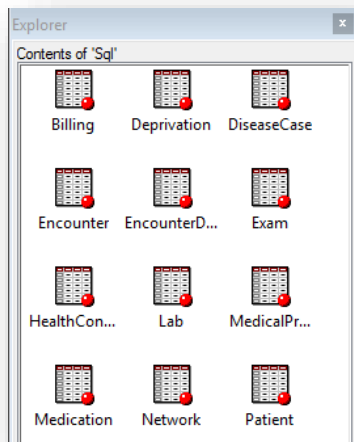
Where:

- xxxxx (as above) is the ODBC Connection name,
- uuuuuu is the database username, and
- ppppppp is the database password

3. Once you have created the new library it will be visible on the left-hand side under Libraries (it will have a little 'world' icon on it).

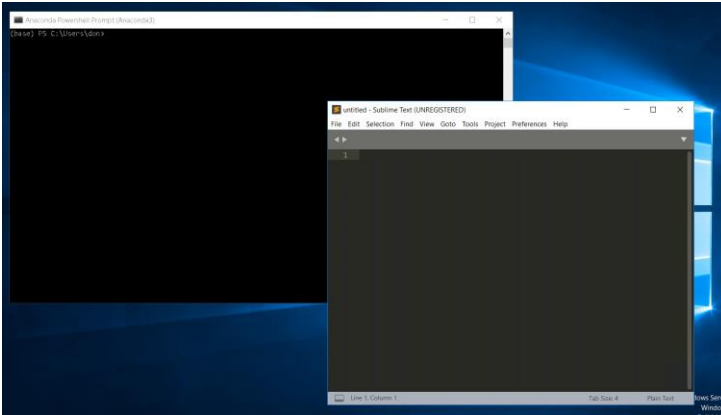


4. Opening up the Library you should now be able to see all your CPCSSN tables (note this image may have some tables that are not in your database).



Python

- **Programming Editor** – Sublime Text 3
- **Shell** – Anaconda Powershell Prompt (Anaconda 3)



- **Sample python connect script**

```
"""
=====
Sample Python connect to database
=====
"""
import pyodbc

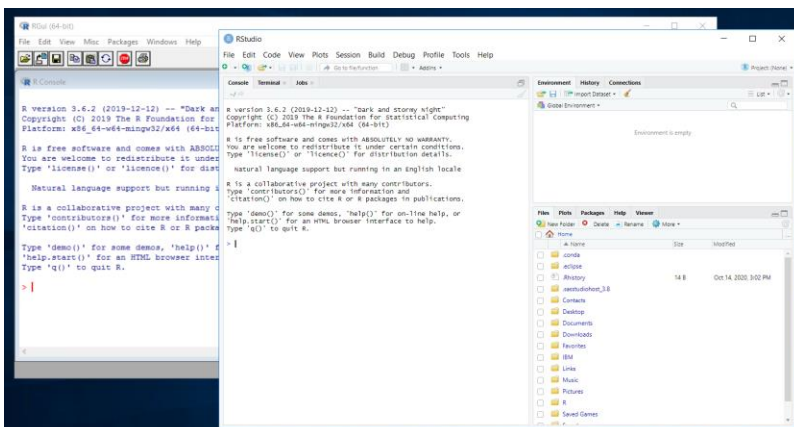
drv='{SQL Server Native Client 11.0}'
srv='10.15.1.150,1433'
dbase='your_database_name'
user='your_user_name'
pass='your_password'

conn=pyodbc.connect(driver=drv, server=srv, database=dbase, uid=user, pwd=pass)

curs=conn.cursor()
sql='SELECT TOP 10 * FROM Billing'
curs.execute(sql)
row=curs.fetchone()
while row:
    print(row)
    row=curs.fetchone()
conn.close()
```

R Studio

- **Programming Editor – RStudio**
- **Shell R x64 GUI**



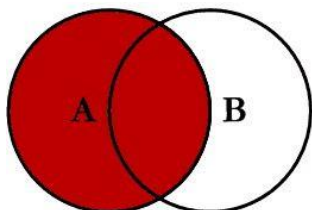
- **Sample R database connect script**

```
library(odbc)
con <- dbConnect(odbc(),
  Driver = "SQLServer",
  Server = "10.15.1.150",
  Database = "your_database",
  UID = "your_user_name",
  PWD = rstudioapi::askForPassword("Database password")
  Port = 1433)
```

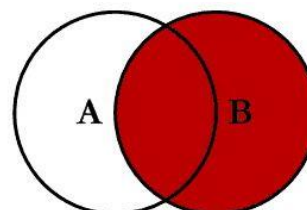
Linking Tables in your Application

Each analytical application has its own mechanism and language for linking data across tables. It is beyond the scope of this document to describe all of them. Most of these mechanisms will translate to a Structure Query Language (SQL) Join against the database tables. For reference, the different types of SQL joins are shown below.

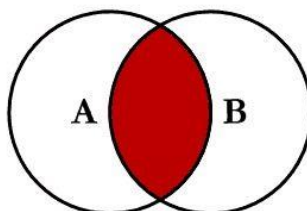
SQL JOINS



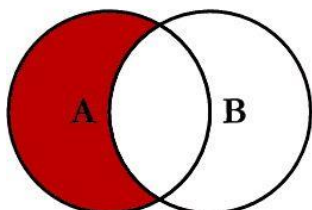
```
SELECT <select_list>
FROM TableA A
LEFT JOIN TableB B
ON A.Key = B.Key
```



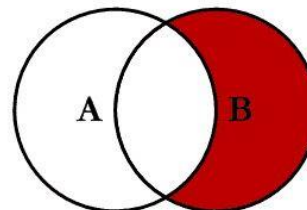
```
SELECT <select_list>
FROM TableA A
RIGHT JOIN TableB B
ON A.Key = B.Key
```



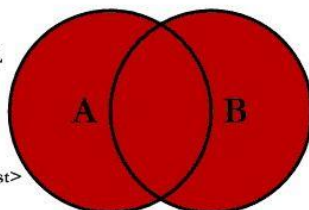
```
SELECT <select_list>
FROM TableA A
INNER JOIN TableB B
ON A.Key = B.Key
```



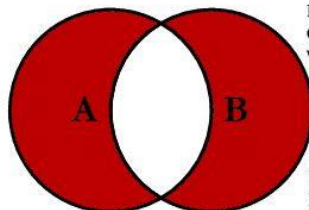
```
SELECT <select_list>
FROM TableA A
LEFT JOIN TableB B
ON A.Key = B.Key
WHERE B.Key IS NULL
```



```
SELECT <select_list>
FROM TableA A
RIGHT JOIN TableB B
ON A.Key = B.Key
WHERE A.Key IS NULL
```



```
SELECT <select_list>
FROM TableA A
FULL OUTER JOIN TableB B
ON A.Key = B.Key
```



```
SELECT <select_list>
FROM TableA A
FULL OUTER JOIN TableB B
ON A.Key = B.Key
WHERE A.Key IS NULL
OR B.Key IS NULL
```

© C.L. Moffatt, 2008