Secure Data Commons

Data Analyst User Guide

www.its.dot.gov/index.htm
Draft Report — September 24, 2018
FHWA-JPO-18-xxx

Produced by DTFH6116H00027
U.S. Department of Transportation

Notice

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for its contents or use thereof. The U.S. Government is not endorsing any manufacturers, products, or services cited herein and any trade name that may appear in the work has been included only because it is essential to the contents of the work.

Revision History

#	Author	Version	Revision Date	Revision Description
1	REAN Cloud	1.0	08/02/2018	Initial Draft
2	REAN Cloud	2.0	09/13/2018	Add Export Functionality

1. Report No.
FHWA-JPO-18-xxx

2. Government Accession No.

3. Recipient's Catalog No.

4. Title and Subtitle
Secure Data Commons Proof of Concept: Data User Guide

5. Report Date
September 24, 2018

6. Performing Organization Code

7. Author(s)
ICF, REAN Cloud

8. Performing Organization Report No.
Task 2 Report

9. Performing Organization Name And Address
ICF International, 1725 Eye St NW,
Washington DC, 20006
REAN Cloud, 2201 Cooperative Way #302
Herndon, VA 20171

10. Work Unit No. (TRAIS)

11. Contract or Grant No.
DTFH61-16-D-00052/Task 12

12. Sponsoring Agency Name and Address
U.S Department of Transportation 1200 New Jersey Ave, SE
Washington, DC 20590

13. Type of Report and Period Covered
Design Document, 09-14-2017 to 1/05/2018

14. Sponsoring Agency Code

15. Supplementary Notes
Ariel Gold (TOCOR), Harry Crump (COR), Bob Brown (CO)

16. Abstract

The SDC POC is an online cloud-based analytic sandbox that provides users access to data sets and programming environments to a number of transportation related data sets. The project performs the following development activities necessary to support the design, launch, and operation of the Secure Data Commons. The primary SDC concepts are:

- Enable scalable data storage, data analysis, and user access protocol via cloud based platforms (AWS S3, Azure Cloud Storage, etc.);
- Leverage cloud capabilities to share complex (high volume, velocity, and/or variety) transportation datasets with the transportation research community;
- Authorize user access through a data use agreement with revocable access terms to protect the sensitivity of the data;
- Provide users with pre-defined data analysis tools and encourage custom toolsets and open sharing amongst the user community;
- Ensure sensitive data is protected through implementation of DOT IT security standards; and
- Utilize agile development processes to develop increasing product functionality and leverage user feedback.

This document describes a guide for data users on the SDC platform.

17. Key Words
Secure Data, Cloud Services, AWS

18. Distribution Statement
This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classif. (of this report)
None

20. Security Classif. (of this page)
None

21. No. of Pages
44

22. Price
NA

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction and Document Overview

The Secure Data Commons (SDC) is a United States Department of Transportation (U.S DOT) sponsored cloud-based analytical sandbox designed to create wider access to sensitive transportation data sets, with the goal of advancing the state of the art of transportation research and state/local traffic management.

The SDC stores sensitive transportation data made available by participating data providers, and grants access to approved researchers to these datasets. The SDC also provides access to open-source tools, and allow researchers to collaborate and share code with other system users.

The SDC platform is a research environment that allows users to conduct analyses and do development and testing of new tools and software products. It is not intended to be an alternative to any local jurisdiction's traffic management center or local data repository. The existing SDC provides users with the following data, tools, and features:

- **Data**: The SDC is ingesting several datasets currently. Additional data sets will be added to the environment over time. Users can bring their own data into the environment to use along with the Waze data.
- **Tools**: The environment provides access to open source tools including Python, RStudio, Microsoft R, SQL Workbench, Power BI, and Jupyter Notebook. These tools are available on a virtual machine in the system enabling data analytics in the cloud.
- **Functionality**: Users can access and analyze data within the environment, save their work to a virtual machine, and publish processes and results to share with others

The SDC platform supports two major roles:

- **Data Providers** – These are entities that provide data hosted on the SDC platform. The data provider establishes the data protection needs and acceptable use terms for the data analysts.
- **Data Analysts** – These are entities that conduct analysis of the datasets hosted in the SDC system. Note that analysts can bring their own data and tools into the SDC system.

This document provides guidance for the **Data Analyst** role. A similar guide will be prepared for the data providers. The document is organized as follows:

- Initial Setup and Validation
- Workstation Access
- Sample Queries

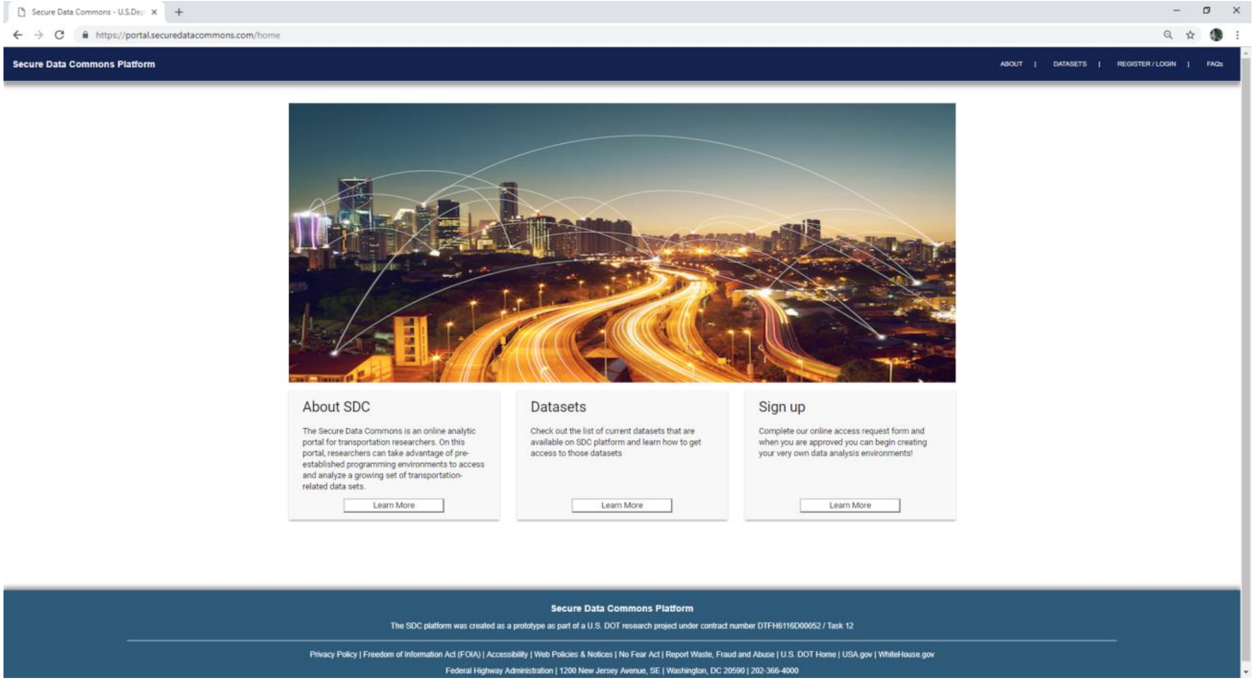
- Exporting Data
- Importing and Exporting Code
- Accessing External Data Sources within SDC
- Technical Support and Contact Information
- Frequently Asked Questions

Chapter 2. Initial Setup and Validation

This chapter provides guidance on the initial setup and validation of the user into the SDC system.

Accessing Portal

Users will access a SDC web portal by clicking on https://portal.securedatacommons.com.



Secure Data Commons - U.S.Dep
https://portal.securedatacommons.com/home
Secure Data Commons Platform
ABOUT | DATASETS | REGISTER / LOGIN | FAQs

About SDC

The Secure Data Commons is an online analytic portal for transportation researchers. On this portal, researchers can take advantage of pre-established programming environments to access and analyze a growing set of transportation-related data sets.

Learn More

Datasets

Check out the list of current datasets that are available on SDC platform and learn how to get access to those datasets

Learn More

Sign up

Complete our online access request form and when you are approved you can begin creating your very own data analysis environments!

Learn More

Secure Data Commons Platform

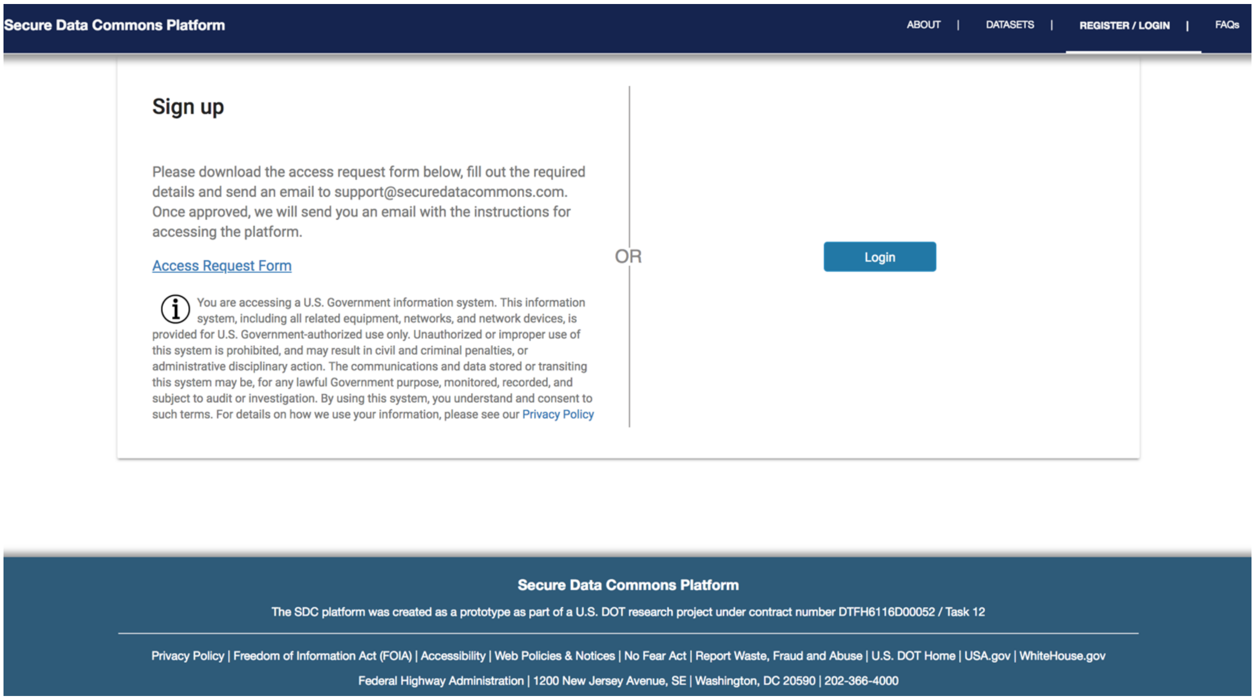
The SDC platform was created as a prototype as part of a U.S. DOT research project under contract number DTFH6116D00052 / Task 12

Privacy Policy | Freedom of Information Act (FOIA) | Accessibility | Web Policies & Notices | No Fear Act | Report Waste, Fraud and Abuse | U.S. DOT Home | USA.gov | WhiteHouse.gov

Federal Highway Administration | 1200 New Jersey Avenue, SE | Washington, DC 20590 | 202-366-4000

Figure 1. Screenshot of SDC web portal (Source: REAN Cloud).

Users can click on the “REGISTER/LOGIN” button on the right and perform the following actions on the login page:


Secure Data Commons Platform
ABOUT | DATASETS | REGISTER / LOGIN | FAQs
OR
Login

Sign up

Please download the access request form below, fill out the required details and send an email to support@securedatacommons.com. Once approved, we will send you an email with the instructions for accessing the platform.

Access Request Form

ⓘ You are accessing a U.S. Government information system. This information system, including all related equipment, networks, and network devices, is provided for U.S. Government-authorized use only. Unauthorized or improper use of this system is prohibited, and may result in civil and criminal penalties, or administrative disciplinary action. The communications and data stored or transiting this system may be, for any lawful Government purpose, monitored, recorded, and subject to audit or investigation. By using this system, you understand and consent to such terms. For details on how we use your information, please see our Privacy Policy

Secure Data Commons Platform

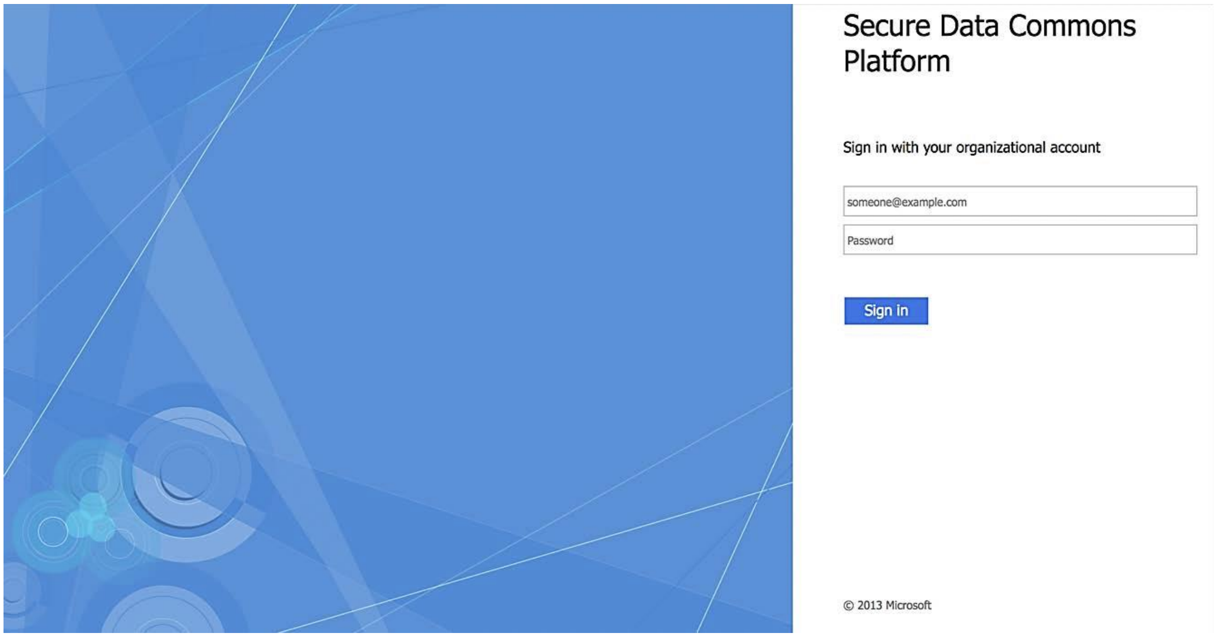
The SDC platform was created as a prototype as part of a U.S. DOT research project under contract number DTFH6116D00052 / Task 12

Privacy Policy | Freedom of Information Act (FOIA) | Accessibility | Web Policies & Notices | No Fear Act | Report Waste, Fraud and Abuse | U.S. DOT Home | USA.gov | WhiteHouse.gov

Federal Highway Administration | 1200 New Jersey Avenue, SE | Washington, DC 20590 | 202-366-4000

Figure 2. Screenshot of the REGISTER/LOGIN page (Source: REAN Cloud).

Clicking on login will redirect users to the Secure Data Commons platform login page



Secure Data Commons Platform

Sign in with your organizational account

someone@example.com

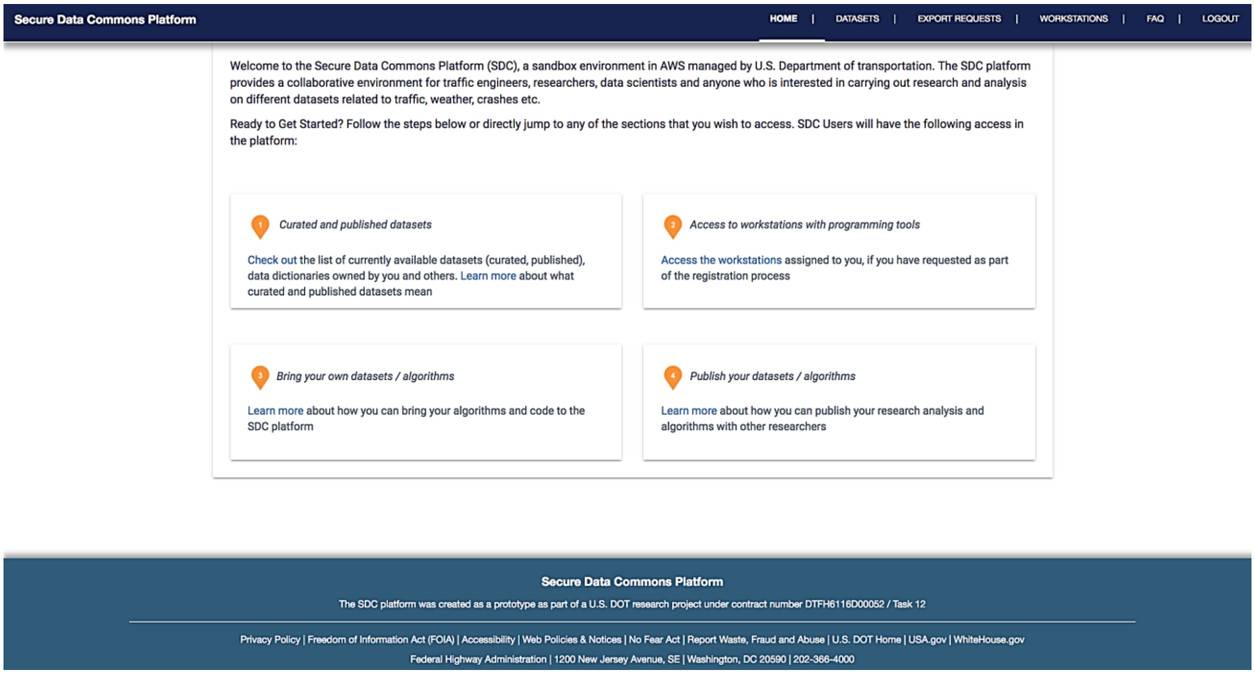
Password

© 2013 Microsoft

Figure 3. Screenshot of the SDC login page (Source: REAN Cloud).

If the user is accessing the portal for the first time, they will be prompted to change the password after entering the credentials provided in the welcome email.

Upon successfully logging in the user is redirected to the landing page, which provides an overview of Secure Data Commons and the different actions the user can perform from the web portal.


Secure Data Commons Platform
HOME | DATASETS | EXPORT REQUESTS | WORKSTATIONS | FAQ | LOGOUT
1
2
3
4

Welcome to the Secure Data Commons Platform (SDC), a sandbox environment in AWS managed by U.S. Department of transportation. The SDC platform provides a collaborative environment for traffic engineers, researchers, data scientists and anyone who is interested in carrying out research and analysis on different datasets related to traffic, weather, crashes etc.

Ready to Get Started? Follow the steps below or directly jump to any of the sections that you wish to access. SDC Users will have the following access in the platform:

Curated and published datasets

Check out the list of currently available datasets (curated, published), data dictionaries owned by you and others. Learn more about what curated and published datasets mean

Access to workstations with programming tools

Access the workstations assigned to you, if you have requested as part of the registration process

Bring your own datasets / algorithms

Learn more about how you can bring your algorithms and code to the SDC platform

Publish your datasets / algorithms

Learn more about how you can publish your research analysis and algorithms with other researchers

Secure Data Commons Platform

The SDC platform was created as a prototype as part of a U.S. DOT research project under contract number DTFH6116D00052 / Task 12

Privacy Policy | Freedom of Information Act (FOIA) | Accessibility | Web Policies & Notices | No Fear Act | Report Waste, Fraud and Abuse | U.S. DOT Home | USA.gov | WhiteHouse.gov

Federal Highway Administration | 1200 New Jersey Avenue, SE | Washington, DC 20590 | 202-366-4000

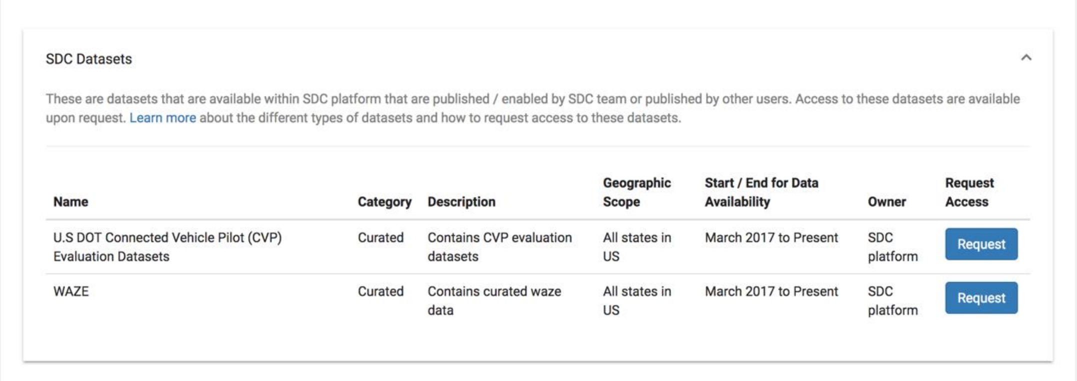
Figure 4. Screenshot of the SDC landing page (Source: REAN CLOUD)

Request Access to Datasets

The available datasets within SDC platform are published / enabled by SDC team or published by other users. Access to these datasets are available upon request.

- Once you are logged in go to tab ‘DATASETS’ in menu tab

Click on SDC Datasets. You will be able to see all available datasets in the SDC platform.



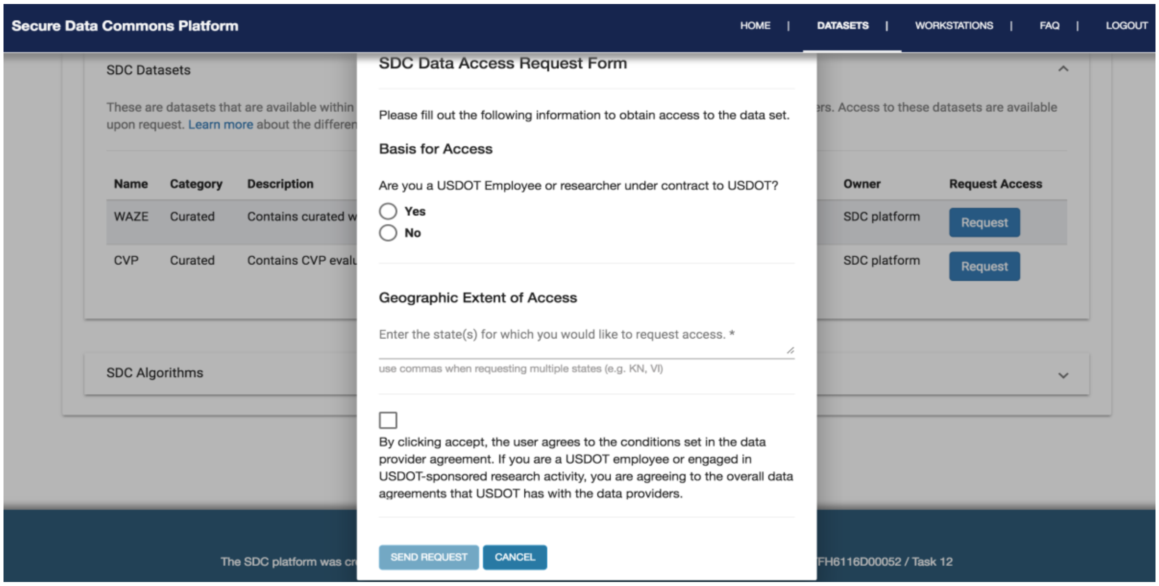
SDC Datasets

These are datasets that are available within SDC platform that are published / enabled by SDC team or published by other users. Access to these datasets are available upon request. Learn more about the different types of datasets and how to request access to these datasets.

Name	Category	Description	Geographic Scope	Start / End for Data Availability	Owner	Request Access
U.S DOT Connected Vehicle Pilot (CVP) Evaluation Datasets	Curated	Contains CVP evaluation datasets	All states in US	March 2017 to Present	SDC platform	Request
WAZE	Curated	Contains curated waze data	All states in US	March 2017 to Present	SDC platform	Request

Figure 5. Screenshot of SDC Datasets (Source: REAN CLOUD).

To access these datasets, click on the ‘Request’ button and complete the attached form. When completed, hit the ‘Send Request’ button.



SDC Data Access Request Form

Please fill out the following information to obtain access to the data set.

Basis for Access

Are you a USDOT Employee or researcher under contract to USDOT?

○ Yes
○ No

Geographic Extent of Access

Enter the state(s) for which you would like to request access. *

use commas when requesting multiple states (e.g. KN, VI)

□

By clicking accept, the user agrees to the conditions set in the data provider agreement. If you are a USDOT employee or engaged in USDOT-sponsored research activity, you are agreeing to the overall data agreements that USDOT has with the data providers.

SEND REQUEST CANCEL

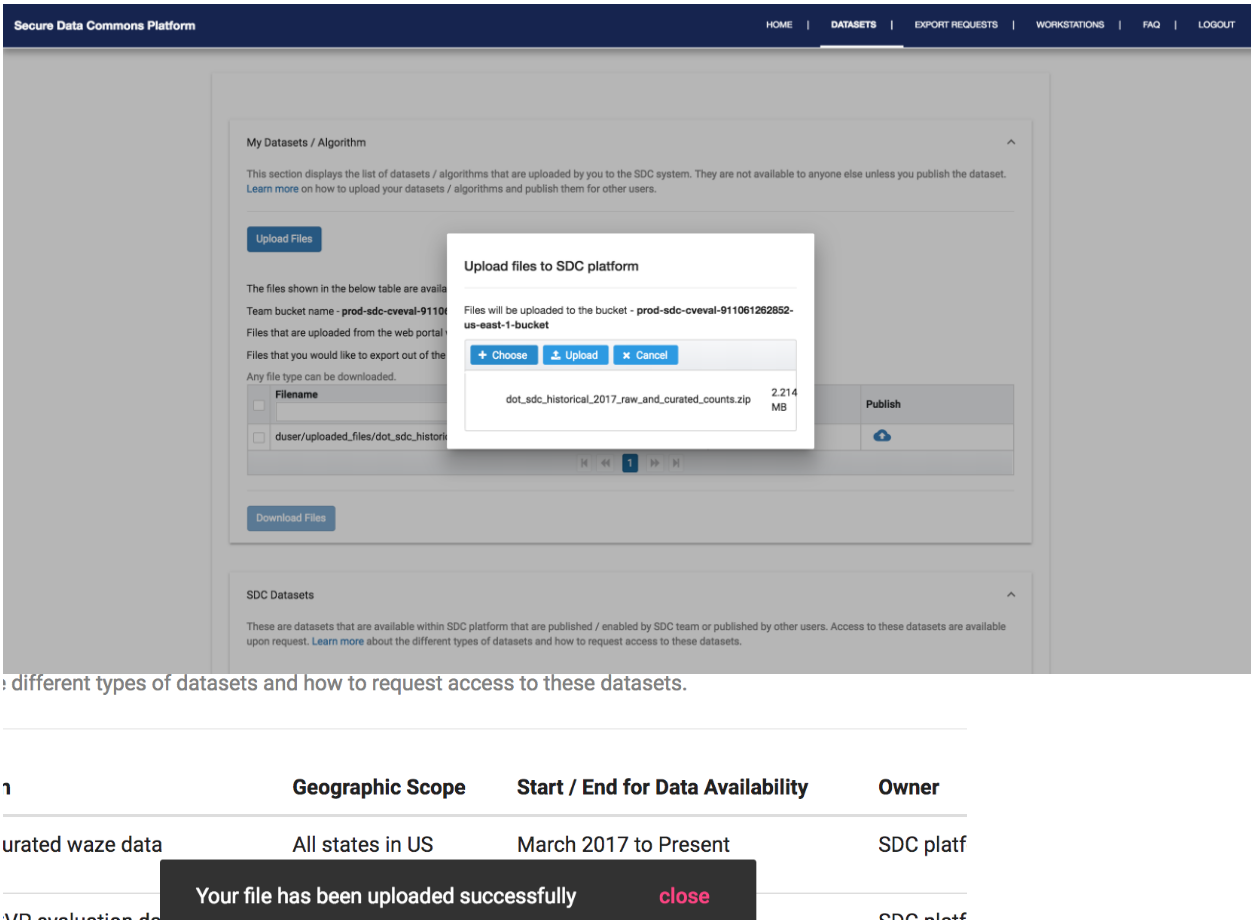
Figure 6. Screenshot of SDC Data Access Request Form (Source: REAN CLOUD)

The request will be sent to the support team and access to the requested dataset will be given upon validation of the information in the form.

Upload User Data to S3 Bucket Through Portal

Users can upload their own data to their assigned team/individual buckets through the portal.

1. Click on the Datasets button on the home page.
2. Click on the Upload Files button under My Datasets / Algorithm.
3. A pop up window will be opened allowing one to choose file/files and upload it to the assigned bucket. (Assigned bucket name will be displayed on the upload pop-up window)
4. A “success” message will be displayed upon a successful upload.


Secure Data Commons Platform
HOME | DATASETS | EXPORT REQUESTS | WORKSTATIONS | FAQ | LOGOUT
Upload Files
+ Choose
Upload
× Cancel
Filename
Publish
1
Download Files
close

My Datasets / Algorithm

This section displays the list of datasets / algorithms that are uploaded by you to the SDC system. They are not available to anyone else unless you publish the dataset. Learn more on how to upload your datasets / algorithms and publish them for other users.

The files shown in the below table are availa

Team bucket name - **prod-sdc-cveval-9110**

Files that are uploaded from the web portal

Files that you would like to export out of the

Any file type can be downloaded.

duser/uploaded_files/dot_sdc_histori

Upload files to SDC platform

Files will be uploaded to the bucket - **prod-sdc-cveval-911061262852-us-east-1-bucket**

dot_sdc_historical_2017_raw_and_curated_counts.zip 2.214 MB

SDC Datasets

These are datasets that are available within SDC platform that are published / enabled by SDC team or published by other users. Access to these datasets are available upon request. Learn more about the different types of datasets and how to request access to these datasets.

: different types of datasets and how to request access to these datasets.

ı	Geographic Scope	Start / End for Data Availability	Owner
urated waze data	All states in US	March 2017 to Present	SDC platf

Your file has been uploaded successfully

Figure 7. Screenshot of: a) Uploading files to SDC platform, and b) Success message upon upload (Source: REAN CLOUD).

Download User Data from S3 Bucket Through Portal

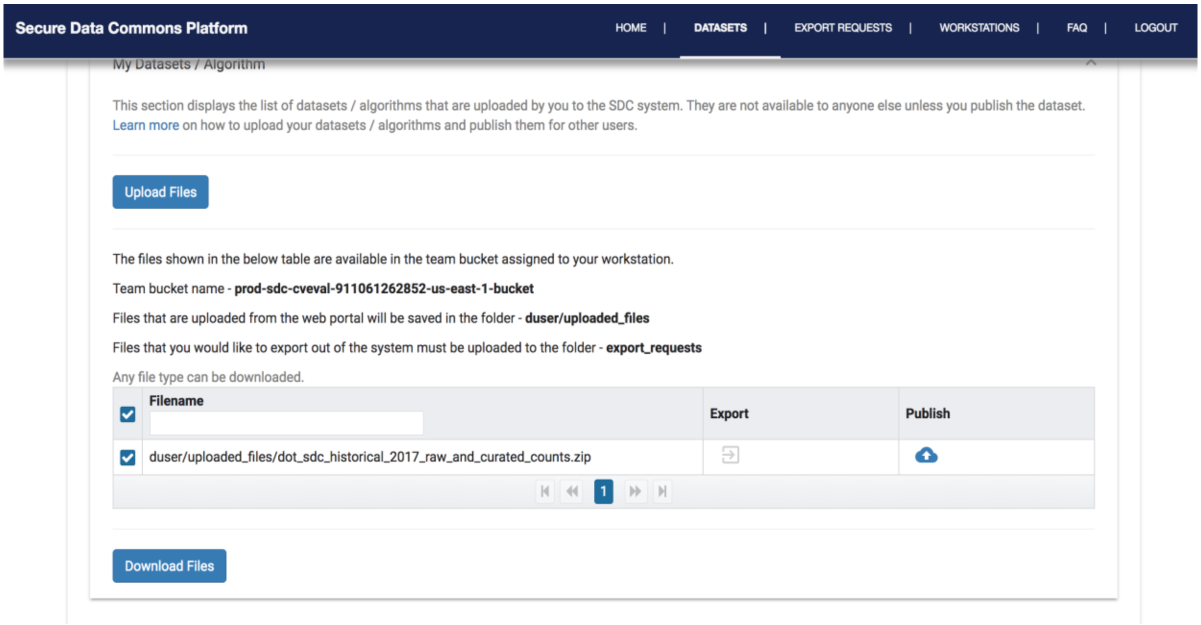
Users can download their data from their assigned team/individual buckets through the portal.

1. Click on the Datasets button on the home page.
2. All the available files under **username/uploaded_files** in the assigned bucket will be displayed along with the assigned bucket name under My Datasets / Algorithm.
3. Select the files that you want to download and click on “Download Files” button.
4. Click on the Download Files button under My Datasets / Algorithm.

Notes:

- Not all the files are downloaded directly. Files like .txt, .png, .pdf will be opened in tab from where we can download. All other files like .csv, .zip etc.., can be downloaded directly.
- Files are downloaded individually.

- The Filename box allows searches partial file names. This can be used to download all the contents of a sub-folder in an S3 bucket, by searching for the sub-folder name, and then clicking the box next to Filename to select all objects.


Secure Data Commons Platform
HOME | DATASETS | EXPORT REQUESTS | WORKSTATIONS | FAQ | LOGOUT
My Datasets / Algorithm

This section displays the list of datasets / algorithms that are uploaded by you to the SDC system. They are not available to anyone else unless you publish the dataset. Learn more on how to upload your datasets / algorithms and publish them for other users.

The files shown in the below table are available in the team bucket assigned to your workstation.

Team bucket name - **prod-sdc-cveval-911061262852-us-east-1-bucket**

Files that are uploaded from the web portal will be saved in the folder - **duser/uploaded_files**

Files that you would like to export out of the system must be uploaded to the folder - **export_requests**

Any file type can be downloaded.

☑	Filename	Export	Publish
☑	duser/uploaded_files/dot_sdc_historical_2017_raw_and_curated_counts.zip	⇥	⇧

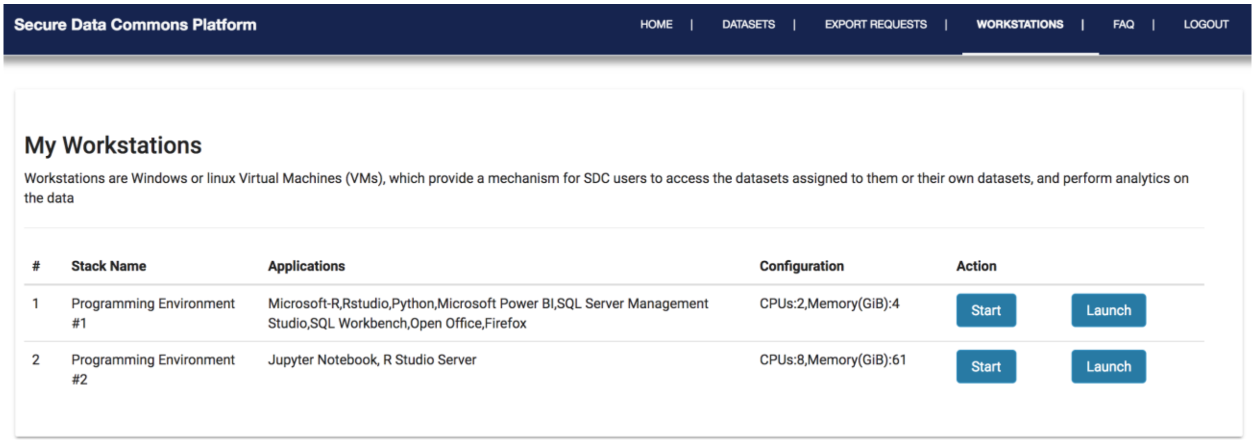
Figure 8. Screenshot of Download page (Source: REAN CLOUD).

Chapter 3. Accessing Workstations

Users are assigned cloud-based workstations to perform analysis on the datasets. This section provides a description of how to launch and use these workstations.

Launch Workstations

Users can see the assigned workstations by clicking on the workstations tab on the top right corner of the page.


Secure Data Commons Platform
HOME | DATASETS | EXPORT REQUESTS | WORKSTATIONS | FAQ | LOGOUT

My Workstations

Workstations are Windows or linux Virtual Machines (VMs), which provide a mechanism for SDC users to access the datasets assigned to them or their own datasets, and perform analytics on the data

#	Stack Name	Applications	Configuration	Action	
1	Programming Environment #1	Microsoft-R,Rstudio,Python,Microsoft Power BI,SQL Server Management Studio,SQL Workbench,Open Office,Firefox	CPUs:2,Memory(GiB):4	Start	Launch
2	Programming Environment #2	Jupyter Notebook, R Studio Server	CPUs:8,Memory(GiB):61	Start	Launch

Figure 9. Screenshot of Workstations (Source: REAN CLOUD)

By default, all the workstations are in the stopped state. Click on the **Start** button to start the workstation. The workstation should become available within five minutes; you may not see any change immediately. A message will appear when the workstation has been successfully started.

Now click the Launch button. This will provide access to your workstation within your browser. The workstation may take a few minutes to initialize. When complete, a login screen will appear.

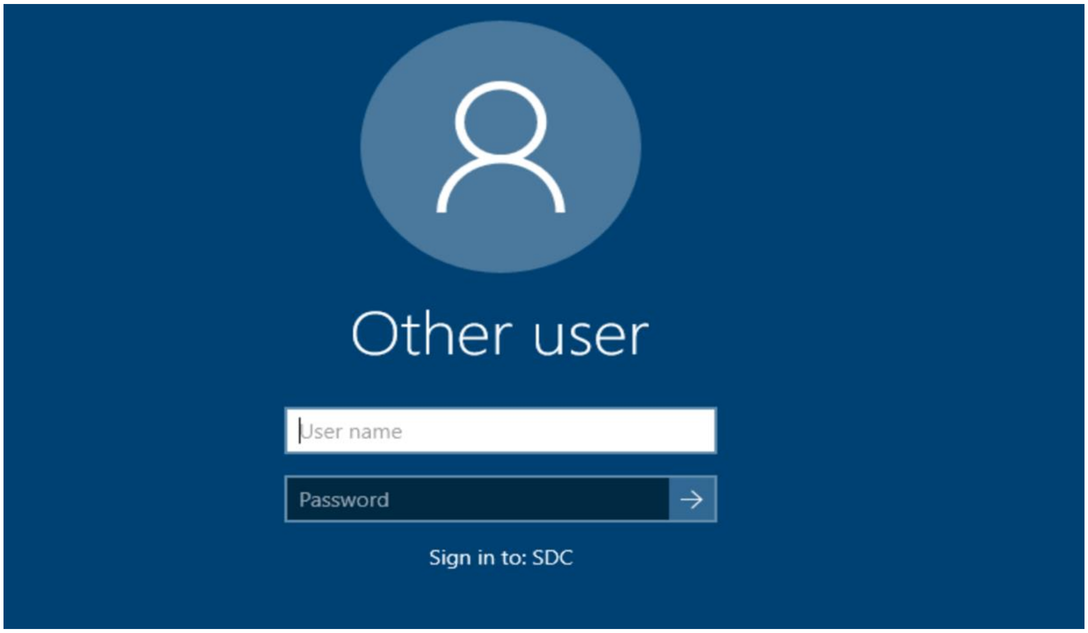


Figure 10. Screenshot of SDC's Login page (Source: REAN CLOUD)

Reenter your assigned username and password to access.

Software Validation

By default, users will have the following installed on their workstations:

- Java
- Python
- R, RStudio
- SQL Workbench
- Power BI
- AWS CLI
- Adobe
- Libre Office
- Visual Studio
- PuTTY
- Firefox

To test Python connectivity to the data warehouse, open the IDLE python editor and execute:

```
from impala.dbapi import connect
conn = connect(
  host='172.18.1.20',
  port=10000,
  auth_mechanism='PLAIN',
  user='<your_username>' ,password='<your_password>')
cursor = conn.cursor()
```

```
    cursor.execute('SHOW TABLES')
    print cursor.fetchall()
```

This should result in an array of tables displayed to the user.

Connecting to the Data Warehouse Using SQL Workbench

The following sections illustrate how the user can connect to the data stores available to the SDC.

Connecting to Waze data in Redshift

Launch SQL Workbench by double clicking the SQL Workbench shortcut on the desktop.

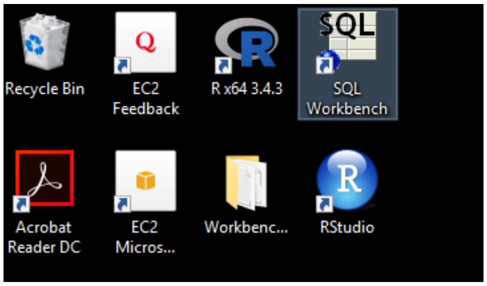
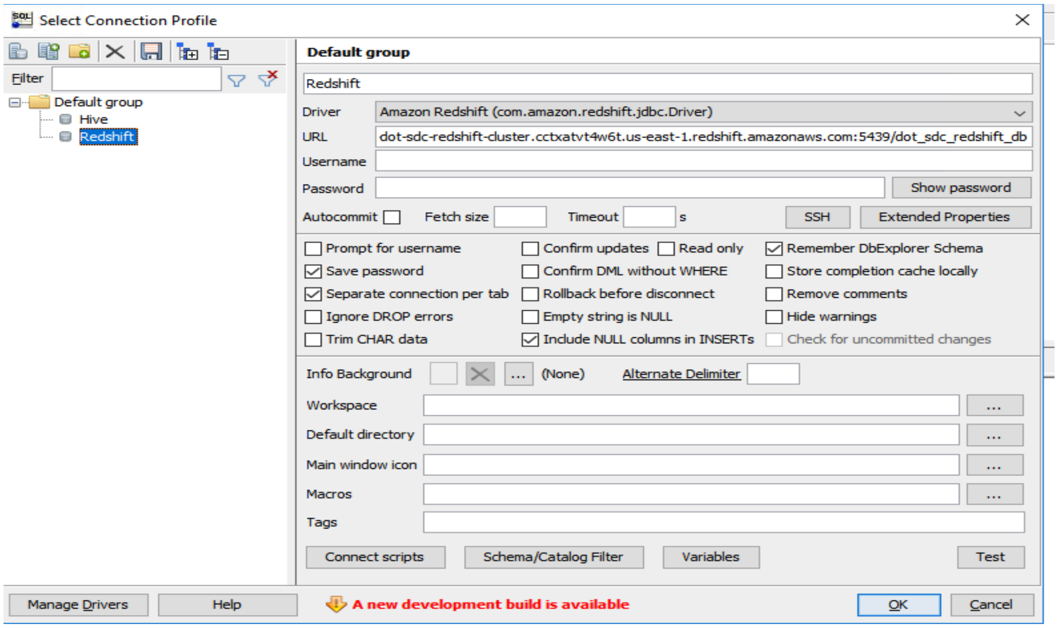


Figure 11. Screenshot of SQL Desktop Icon (Source: REAN CLOUD)

Create a Redshift connection profile to connect to Waze data

- Create new connection profile by selecting left top corner icon on the “Select Connection Profile”
- Select “Amazon Redshift Driver” under driver dropdown.
- Update the URL section with the Redshift URL provided in the cheat sheet provided later in the document[1].
- Provide your username and password received in the welcome email.
- Click on the Test button at the bottom to the connection. A dialog box will pop up confirming a successful or failed connection.

[1]



Redshift

Driver: Amazon Redshift (com.amazon.redshift.jdbc.Driver)

URL: dot-sdc-redshift-cluster.cctxatvt4w6t.us-east-1.redshift.amazonaws.com:5439/dot_sdc_redshift_db

Username:

Password:

Autocommit □ Fetch size Timeout s

□ Prompt for username
☑ Save password
☑ Separate connection per tab
□ Ignore DROP errors
□ Trim CHAR data
□ Confirm updates □ Read only
□ Confirm DML without WHERE
□ Rollback before disconnect
□ Empty string is NULL
☑ Include NULL columns in INSERTs
☑ Remember DbExplorer Schema
□ Store completion cache locally
□ Remove comments
□ Hide warnings
□ Check for uncommitted changes

Info Background ... (None) Alternate Delimiter

Workspace ...

Default directory ...

Main window icon ...

Macros ...

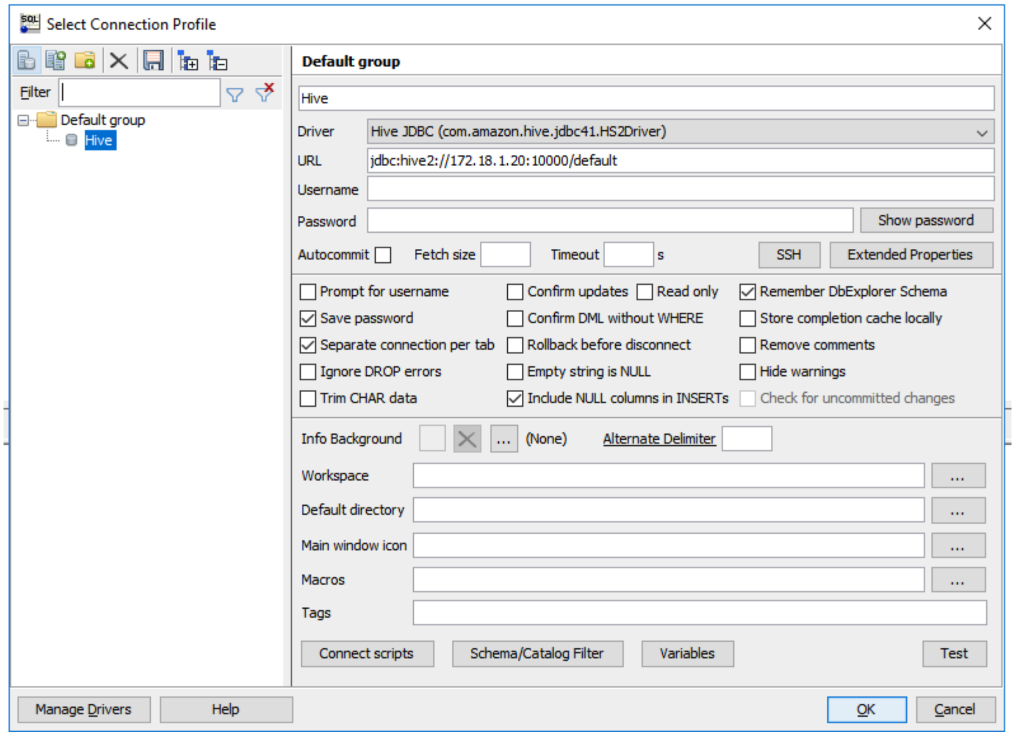
Tags

Figure 12. Screenshot of Redshift's connection profile creation page (Source: REAN CLOUD)

Connecting to CVP data in Hive Metastore

Launch SQL Workbench by double clicking the SQL Workbench shortcut on the desktop.

- Create a hive connection profile
  - Create a new connection profile by selecting the left top corner icon on the “Select Connection Profile”
  - Select “Hive JDBC” under driver dropdown.
  - Update URL section with the Hive URL provided in the cheat sheet.
  - Provide your username and password received in the welcome email.
  - Click on the Test button at the bottom to validate your connection. A pop up dialog will appear confirming a successful or failed connection.



Default group

Hive

Driver: Hive JDBC (com.amazon.hive.jdbc41.HS2Driver)

URL: jdbc:hive2://172.18.1.20:10000/default

Username:

Password:

Autocommit □ Fetch size Timeout s

□ Prompt for username
☑ Save password
☑ Separate connection per tab
□ Ignore DROP errors
□ Trim CHAR data
□ Confirm updates
□ Read only
□ Confirm DML without WHERE
□ Rollback before disconnect
□ Empty string is NULL
☑ Include NULL columns in INSERTs
☑ Remember DbExplorer Schema
□ Store completion cache locally
□ Remove comments
□ Hide warnings
□ Check for uncommitted changes

Info Background (None) Alternate Delimiter

Workspace

Default directory

Main window icon

Macros

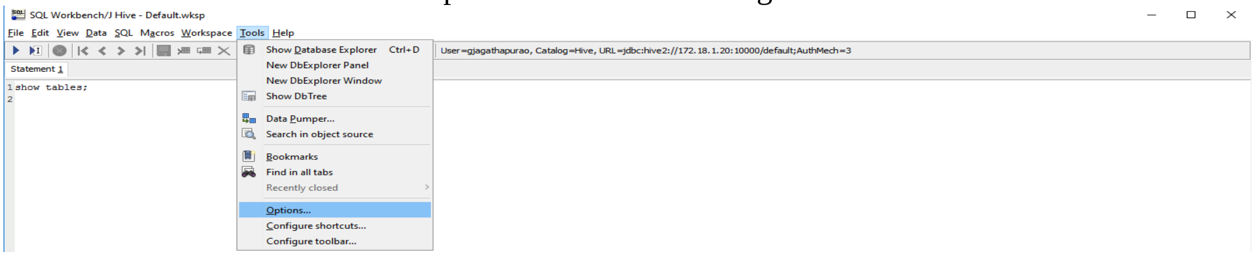
Tags

Figure 13. Screenshot of connecting to CVP data in Hive Metastore (Source: REAN CLOUD)

Update Data Formatting Settings in SQL Workbench

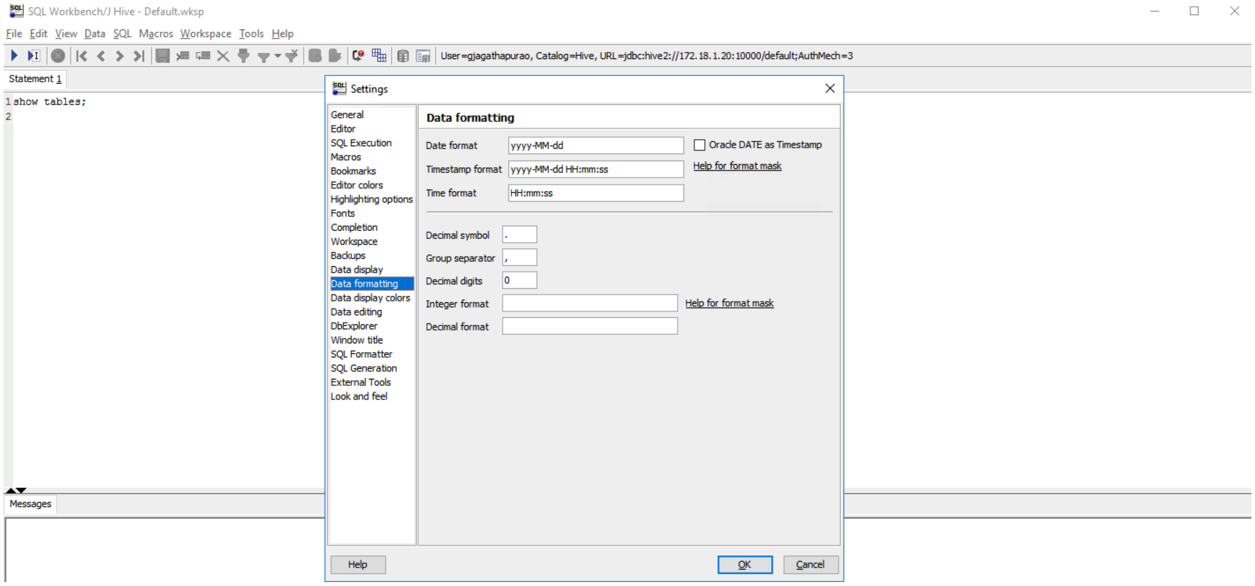
Credentials Once the connection has been established, navigate to Tools | Options | Data formatting and update the Decimal digits value to 0.

Tools → Options → Data formatting



User=gjagathapurao, Catalog=Hive, URL=jdbc:hive2://172.18.1.20:10000/default;AuthMech=3

```
show tables;
```



User=gjagathapurao, Catalog=Hive, URL=jdbc:hive2://172.18.1.20:10000/default;AuthMech=3

```
show tables;
```

Data formatting

Date format: yyyy-MM-dd

□ Oracle DATE as Timestamp

Timestamp format: yyyy-MM-dd HH:mm:ss

Time format: HH:mm:ss

Decimal symbol: .

Group separator: ,

Decimal digits: 0

Integer format:

Decimal format:

Figure 14. Screenshot of updating the data formatting settings in SQL Workbench (Source: REAN CLOUD)

Connecting to Redshift from Linux Environments

Credentials to access the Waze Redshift database are communicated from the SDC Administrator (support@securedatacommons.com) by a secure email service Virtu. Users being granted access will receive an email in the normal email client (such as Outlook) with an “Unlock Message” link. Clicking on this will link to secure.virtu.com/secure-reader, where the credentials will be displayed as an email in the web browser. A verification step may be required, where another email will be sent to Outlook, to confirm the user’s email address.

- In R, it is possible to connect to Redshift using multiple packages. The RPostgreSQL package provides a simple method. This package requires the PostgreSQL library to be installed at the system level; if it is not installed, it would be necessary to install as root in the terminal:

```
$ sudo yum install postgresql-devel
```

- In R, you may need to `install.packages(“RPostgreSQL”, dep=T)`
- Connect to Redshift using code like the following:

```
library(RPostgres)
# Specify username and password manually, once:
if(Sys.getenv("sdc_waze_username")==""){
  cat("Please enter SDC Waze username and password
manually, in the console, the first time accessing the
Redshift database, using: \n Sys.setenv('sdc_waze_username'
= <see email from SDC Administrator>) \n
Sys.setenv('sdc_waze_password' = <see email from SDC
Administrator>)")
}

redshift_host <- [redacted]"
redshift_port <- "5439"
redshift_user <- Sys.getenv("sdc_waze_username")
redshift_password <- Sys.getenv("sdc_waze_password")
redshift_db <- "dot_sdc_redshift_db"

#drv <- dbDriver("PostgreSQL")
conn <- dbConnect(
  RPostgres::Postgres(),
  host=redshift_host,
  port=redshift_port,
  user=redshift_user,
  password=redshift_password,
  dbname=redshift_db)
```

- The database can then be queried using the `dbGetQuery()` function.

Accessing Jupyter Notebook and RStudio Server

Linux users can access their Jupyter Notebook and RStudio Server using the Firefox web browser through windows workstation using below URLs.

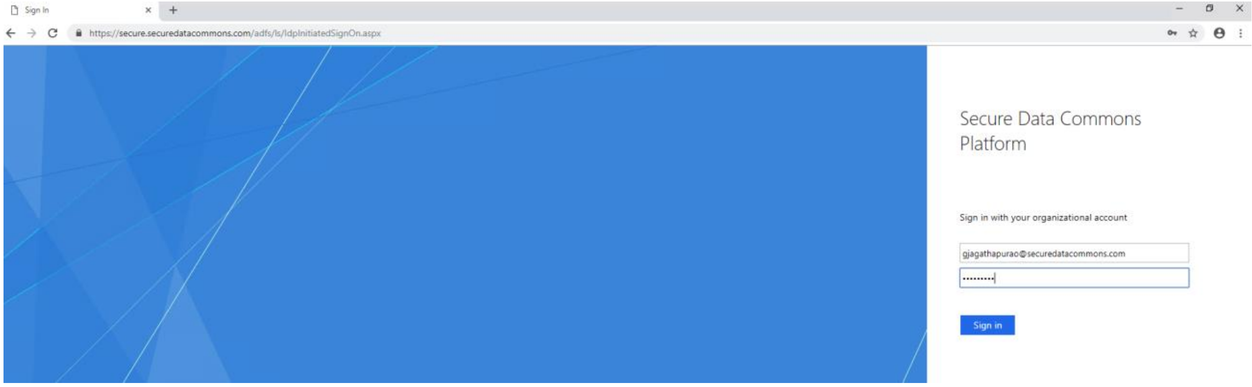
- RStudio – http://<username>-workspace.securedatacommons.internal:8787
- Jupyter Notebook – http://<username>-workspace.securedatacommons.internal:8888

Windows users can click on the “RStudio” shortcut icon present on the desktop to open RStudio console.

Accessing S3 Buckets from Workstations

Users can access their s3 buckets by logging into SDC AWS console. Each team/user is assigned with appropriate roles which will grant them access to only their team buckets and read only access to data lake bucket.

1. Login to SDC AWS account using the below link and using your SDC credentials from your workstation.
   https://secure.securedatacommons.com/adfs/ls/IdpInitiatedSignOn.aspx



Sign In

https://secure.securedatacommons.com/adfs/ls/IdpInitiatedSignOn.aspx

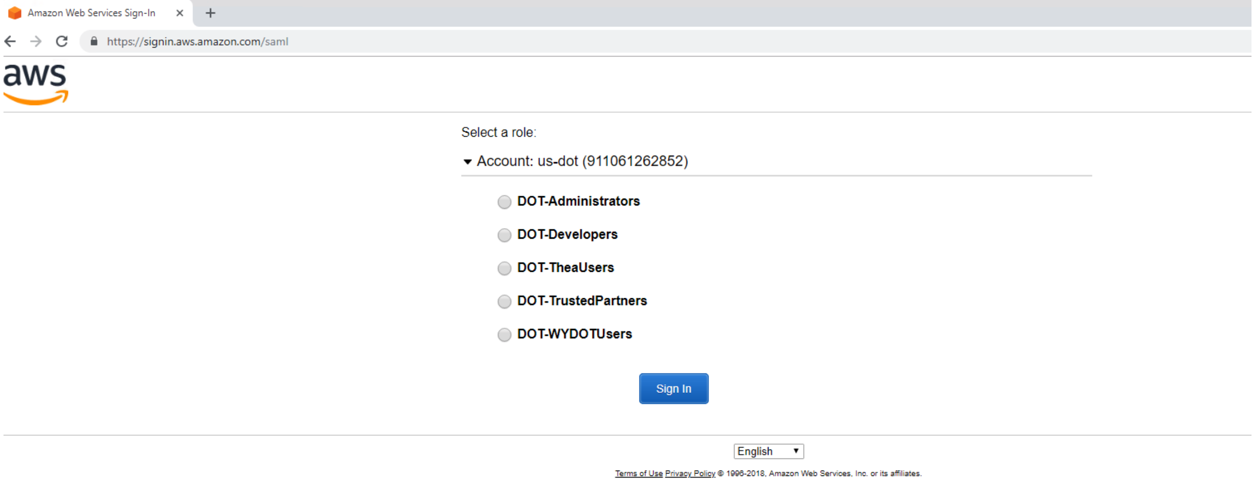
Secure Data Commons Platform

Sign in with your organizational account

gjagathapurao@securedatacommons.com

•••••••••

2. Upon successful login, you will be prompted to assume the role assigned to you based on your access levels.



Amazon Web Services Sign-In

https://signin.aws.amazon.com/saml

Select a role:

Account: us-dot (911061262852)

○ DOT-Administrators

○ DOT-Developers

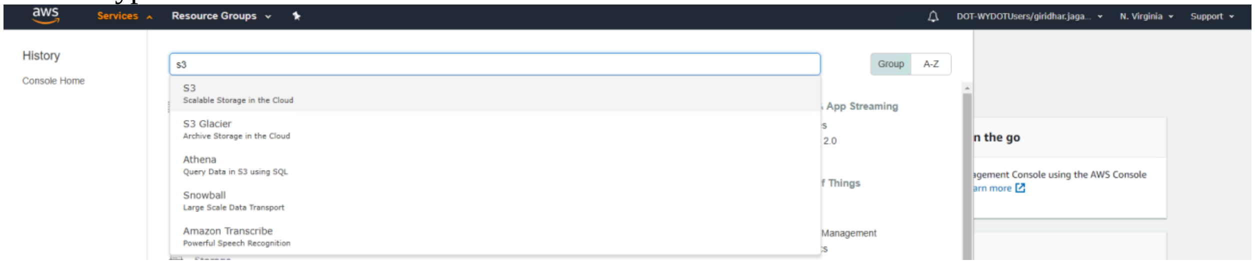
○ DOT-TheaUsers

○ DOT-TrustedPartners

○ DOT-WYDOTUsers

Terms of Use Privacy Policy © 1996-2018, Amazon Web Services, Inc. or its affiliates.

3. In this example, lets select DOT-WYDOTUsers as the trusted role and click signin. But doing so, you will have only access to wydot team bucket and read only access to datalake bucket.
4. To access s3 buckets, click on **Services** drop down on the top left corner of the console and type **S3** in the search bar.



s3

S3
Scalable Storage in the Cloud

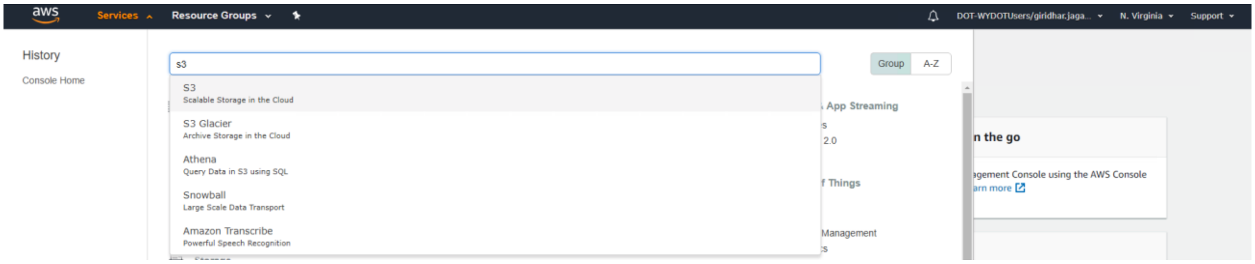
S3 Glacier
Archive Storage in the Cloud

Athena
Query Data in S3 using SQL

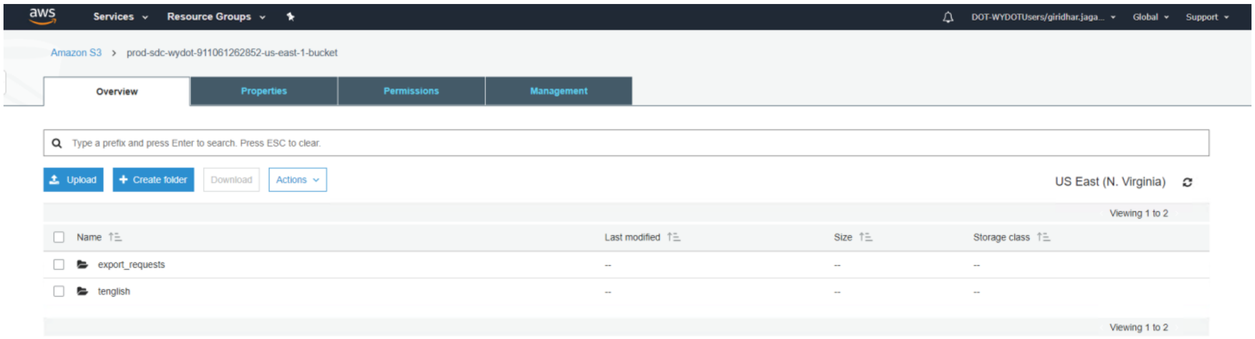
Snowball
Large Scale Data Transport

Amazon Transcribe
Powerful Speech Recognition

5. Upon selecting S3 service, you will be taken to S3 console as shown below.


aws
Services
Resource Groups
DOT-WYDOTUsers/giridhar.jaga...
N. Virginia
Support
History
Console Home
s3
Group
A-Z
S3
Scalable Storage in the Cloud
S3 Glacier
Archive Storage in the Cloud
Athena
Query Data in S3 using SQL
Snowball
Large Scale Data Transport
Amazon Transcribe
Powerful Speech Recognition
App Streaming
2.0
Things
Management
n the go
agement Console using the AWS Console
arn more

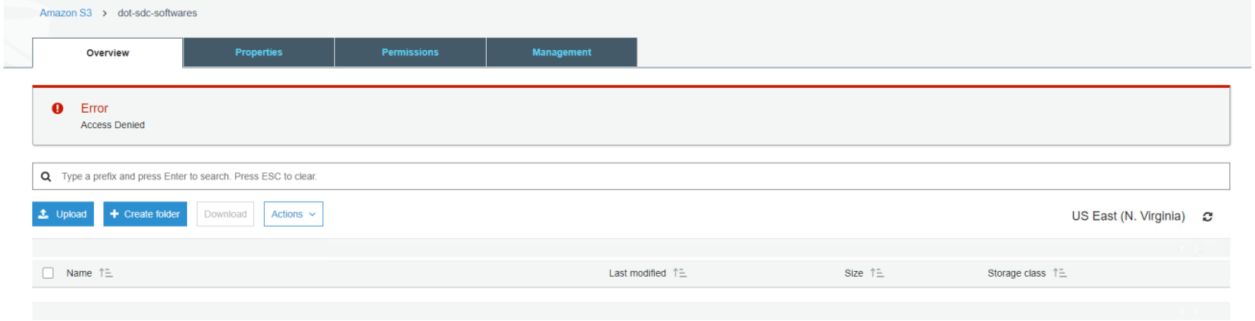
6. Now users would be able to access the assigned buckets by searching the bucket name in the search bar.


aws
Services
Resource Groups
DOT-WYDOTUsers/giridhar.jaga...
Global
Support
Amazon S3 › prod-sdc-wydot-911061262852-us-east-1-bucket
Overview
Properties
Permissions
Management
Type a prefix and press Enter to search. Press ESC to clear.
Upload
Create folder
Download
Actions
US East (N. Virginia)
Viewing 1 to 2

Name	Last modified	Size	Storage class
export_requests	--	--	--
tenglish	--	--	--

Viewing 1 to 2

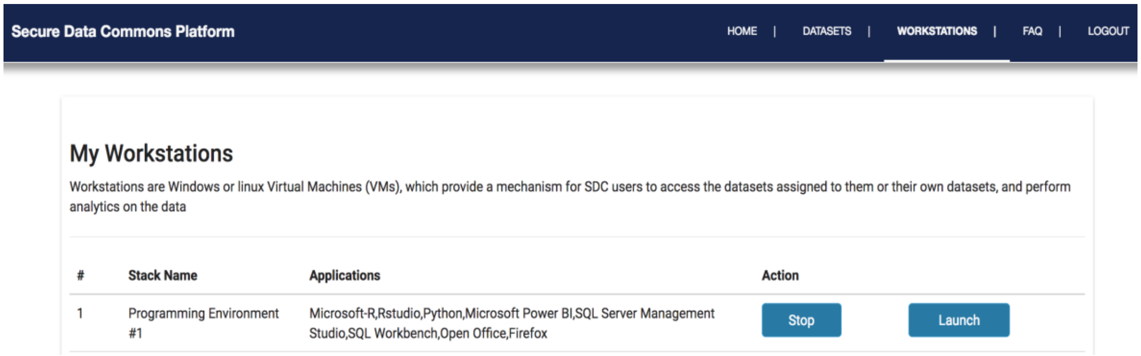
7. Users wouldn’t be able to access other buckets other than their team and data lake buckets.



Error
Access Denied

Stop Workstations

Users can see the assigned workstations by clicking on the workstations tab on the top right corner of the page. By default, all the workstations are scheduled to stop every day at 11 PM EST. Users can stop the workstations manually by clicking on the Stop button as shown below. A message will appear when the instance is successfully stopped.


Secure Data Commons Platform
HOME | DATASETS | WORKSTATIONS | FAQ | LOGOUT

My Workstations

Workstations are Windows or linux Virtual Machines (VMs), which provide a mechanism for SDC users to access the datasets assigned to them or their own datasets, and perform analytics on the data

#	Stack Name	Applications	Action	
1	Programming Environment #1	Microsoft-R,Rstudio,Python,Microsoft Power BI,SQL Server Management Studio,SQL Workbench,Open Office,Firefox	Stop	Launch

Figure 15. Screenshot of "My Workstations" page (Source: REAN CLOUD)

Chapter 4. Sample Queries for SDC Data Sets

The following section provides some sample queries for anchor data sets hosted by the SDC. Two datasets are currently available in the SDC system:

- Data provided by USDOT-sponsored research on Connected Vehicles (Connected Vehicle Pilots)
- Data provided by Waze on traffic jams and alerts

Sample queries are provided for each of the two data sets.

Connected Vehicle (CV) Data

Overview

The CV Pilot Data Warehouse is built on top of a Hadoop's HDFS cluster and utilizes Hive with a HiveQL querying language as a front-end querying package. Information about the HiveQL language and the language manual can be found online:

- Wikipedia info on Apache Hive: https://en.wikipedia.org/wiki/Apache_Hive
- Language manual: https://cwiki.apache.org/confluence/display/Hive/LanguageManual

WYDOT BSM Relational Tables

WYDOT BSM data currently exists within the SDC in two formats: "version 4" and "version 5". Currently incoming data conforms to the "version 5" format. Older data conforms to the legacy "version 4" format. For analysis, data from both versions have been combined into a set of relational tables. These are the tables you should be querying. They are:

```
wydot_bsm_core
wydot_bsm_partii
wydot_bsm_partii_crumbdata
```

Figure 16 illustrates their relationship:

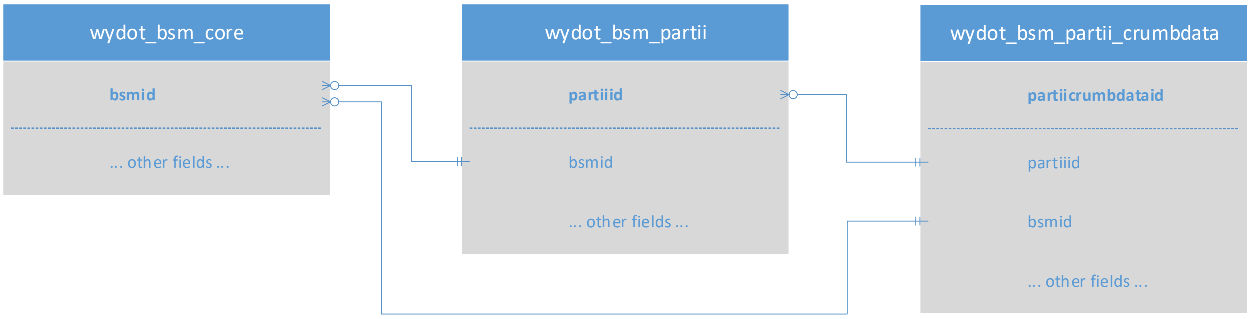


Figure 16. Relationship between dataset tables

Contents may be queried using SQL workbench provided that it's configured to connect to the data warehouse. Some examples:

```
select * from wydot_bsm_core limit 5;
select * from wydot_bsm_partii limit 5;
select * from wydot_bsm_partii_crumbdata limit 5;

select * from wydot_bsm_core join wydot_bsm_partii where
wydot_bsm_core.bsmid = wydot_bsm_partii.bsmid limit 5;

select * from wydot_bsm_partii join
wydot_bsm_partii_crumbdata where wydot_bsm_partii.partiiid
= wydot_bsm_partii_crumbdata.partiiid limit 5;

select * from wydot_bsm_core join wydot_bsm_partii join
wydot_bsm_partii_crumbdata where wydot_bsm_core.bsmid =
wydot_bsm_partii.bsmid and wydot_bsm_partii.partiiid =
wydot_bsm_partii_crumbdata.partiiid limit 10;
```

Wyoming DOT BSM and TIM Metadata RECORDGENERATEDAT vs ODERECEIVEDAT

Wyoming DOT (WYDOT) BSM and TIM messages include a metadata section, appended by the Operational Data Environment (ODE) component.[2] Among others, there are two timestamps of interest are presented in this section:

recordGeneratedAt: Closest time to which the record was created by a Vehicle.
odeReceivedAt: Time ODE received the data in UTC format. This time is the closest to which the CV PEP system received the record.

- Based on real-life conditions, odeReceivedAt may be days or even weeks after the recordGeneratedAt timestamp.

- The following queries result in distribution of recordGeneratedAt by odeRecevedAt times:
- WYDOT BSM messages:

  ```
  select SUBSTR(metadatarecordgeneratedat, 0, 10) as
  RECORDGENERATEDAT
  , SUBSTR(metadataodereceivedat, 0, 10) as ODERECEIVEDAT
  , count(SUBSTR(metadataodereceivedat, 0, 10)) as CNT
  from wydot_bsm_core
  group by SUBSTR(metadataodereceivedat, 0, 10)
  , SUBSTR(metadatarecordgeneratedat, 0, 10)
  order by RECORDGENERATEDAT limit 10000;
  ```

- WYDOT TIM messages:

  ```
  select SUBSTR(metadatarecordgeneratedat, 0, 10) as
  RECORDGENERATEDAT
  , SUBSTR(metadataodereceivedat, 0, 10) as ODERECEIVEDAT
  , count(SUBSTR(metadataodereceivedat, 0, 10)) as CNT
  from wydot_tim
  group by SUBSTR(metadataodereceivedat, 0, 10)
  , SUBSTR(metadatarecordgeneratedat, 0, 10)
  order by RECORDGENERATEDAT limit 10000;
  ```

WYDOT BSM: Getting Latitude and Longitude

```
select coredatalatitude
, coredatalongitude
from wydot_bsm_core limit 1;
```

WYDOT Speed Data

There are two WYDOT speed data sets in the CV PEP Data Warehouse: wydot_speed_unprocessed and wydot_speed_processed. The following sample query displays average vehicle speed distribution by lane and it will work against either of the tables:

```
select lane, avg(speedmph) as speed_average
from wydot_speed_unprocessed
group by lane
order by lane;
```

Geospatial Queries

- The Data Warehouse has geospatial querying capabilities. Functions such as ST_Point, ST_Polygon, ST_Contains and others can be used in queries. For the full list of supported functions see: https://github.com/Esri/spatial-framework-for-hadoop/wiki/UDF-Documentation
- As an example, here is a sample query to retrieve a count of messages generated by vehicles between latitudes 40 and 41 and longitudes between -106 and -105.

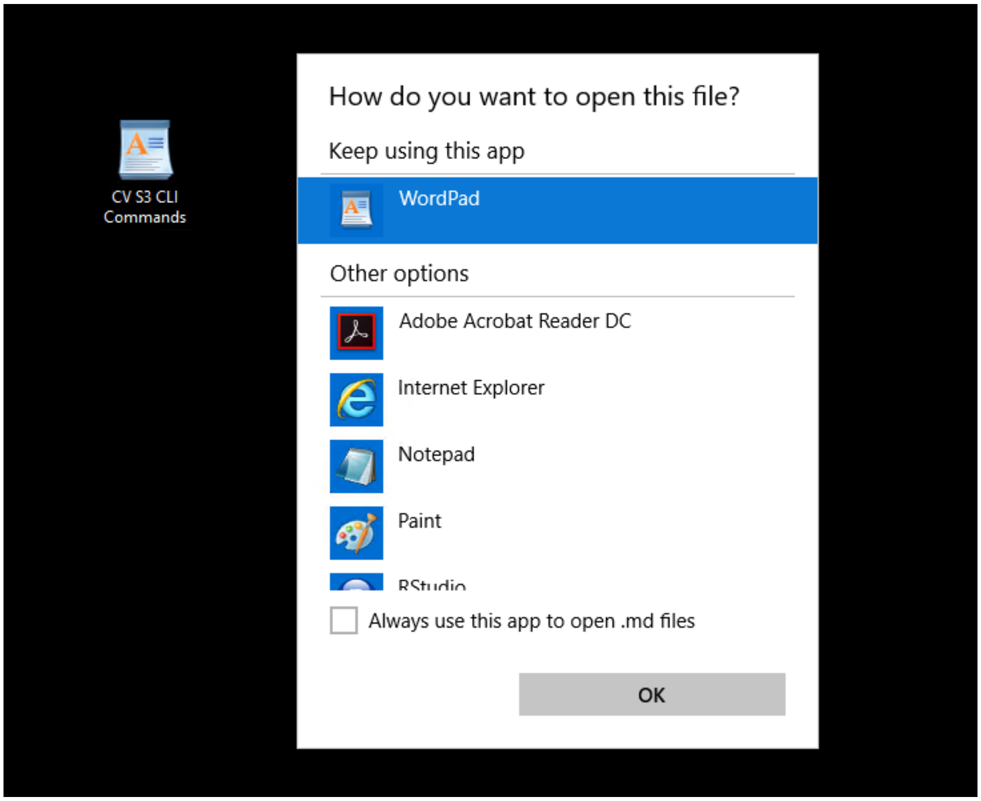
  ```
  select count(*) from wydot_bsm_core where
  ```

```
ST_Contains(ST_Polygon(40, -105, 41, -105, 41, -106, 40, -
106),
ST_Point(coredatalatitude, coredatalongitude));
```

Cheat Sheet

The cheat sheet is available on the desktop of all workstations with sets of useful commands. The cheat sheet is called “S3 CLI Commands.md or CV S3 CLI Commands.md” and can be opened using Wordpad.

1. Right Click on the File
2. Choose Open with… and select Wordpad.



How do you want to open this file?

Keep using this app

WordPad

Other options

Adobe Acrobat Reader DC

Internet Explorer

Notepad

Paint

RStudio

□ Always use this app to open .md files

Figure 17. Screenshot of Cheat Sheet and how to open it (Source: REAN CLOUD)

Waze Data

Overview

The SDC platform ingests and curates Waze data for all 50 states of United States of America using the Waze API. Users have access only to Waze data pertaining to the specific geographic region specified in their data access policies.

Waze Alert Data

An alert is a User Generated Incident (UGI) reported by a Waze user or group of users, as defined by Waze. In SQL Workbench, the following query can be used to test access to the *alert* table.

```
select distinct alert_uuid from alert where
alert_type='<alert_type>' and pub_utc_timestamp between
<start_time_stamp> and <end_time_stamp> and state
=<state_name>;

select distinct alert_uuid from alert where
alert_type='ACCIDENT' and pub_utc_timestamp between '2018-
01-01 04:59:43.00' and '2018-01-30 04:59:43.00' and state
='CA';
```

Waze Jam Data

Waze provides information of traffic jams and events that affect road conditions either from wazers or external sources. A traffic jam maybe associated with an alert.

```
select top 10 * from jam;
select top 10 * from jam_point_sequence;
```

Waze Irregularity Data

Irregularities are similar to Jams, where Waze derives these events based on unusual traffic patterns. These could also be a result of an alert or a jam.

```
select top 10 * from irregularity;
select top 10 * from irregularity_point_sequence;
select top 10 * from irregularity_alert;
select top 10 * from irregularity_jam;
```

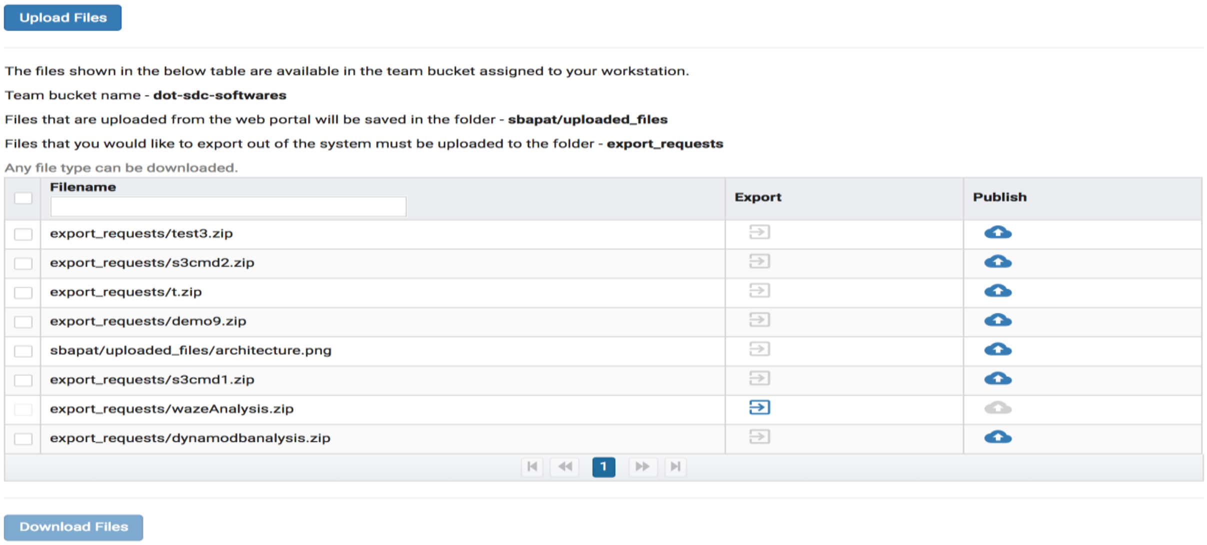
Chapter 6. Exporting Data from the SDC

Data Analysts should be able to export the data of the system, based on the compliance and data usage policies set forth by a Data Provider. There are two different types of analysts:

1. **General Analyst:** This type of analyst must provide justification of each data product to the data provider of why it can be exported out of the SDC system. The intent is to ensure that the data provider has oversight of the exported data.
   This type of analyst can also request for the trusted status to the data provider while filling the approval form.
2. **Trusted Analyst:** This type of analyst already has a trusted status which is provided to them by the data providers. The intent is to reduce the effort for exporting the data products of the analysis out of the SDC system. A trusted user has a preexisting and approved relationship with the data provider.

Once the Analyst completes creating derived datasets, either working on the SDC datasets or combining with other data sets that they import into the system, the analysts can export the derived datasets or share the datasets with other team members. Following are the steps the Data Analyst need to follow to export the data of their analysis form the SDC system to support their research:

1. Each data analyst is part of a team bucket which is displayed in the Datasets section. When ready to export, the Data Analysts select the file (or files) that they want to export out of the SDC system and places them in a separate staging folder (i.e. **export_requests)** in their team bucket. An analyst can request for exporting a file in this folder by clicking on the export symbol against the file he wants out of the SDC system.
   Note: Derived datasets placing under staging folder (i.e. **export_requests**) should be in compressed format. SDC currently supports .zip and .tar formats.



The files shown in the below table are available in the team bucket assigned to your workstation.

Team bucket name - **dot-sdc-softwares**

Files that are uploaded from the web portal will be saved in the folder - **sbapat/uploaded_files**

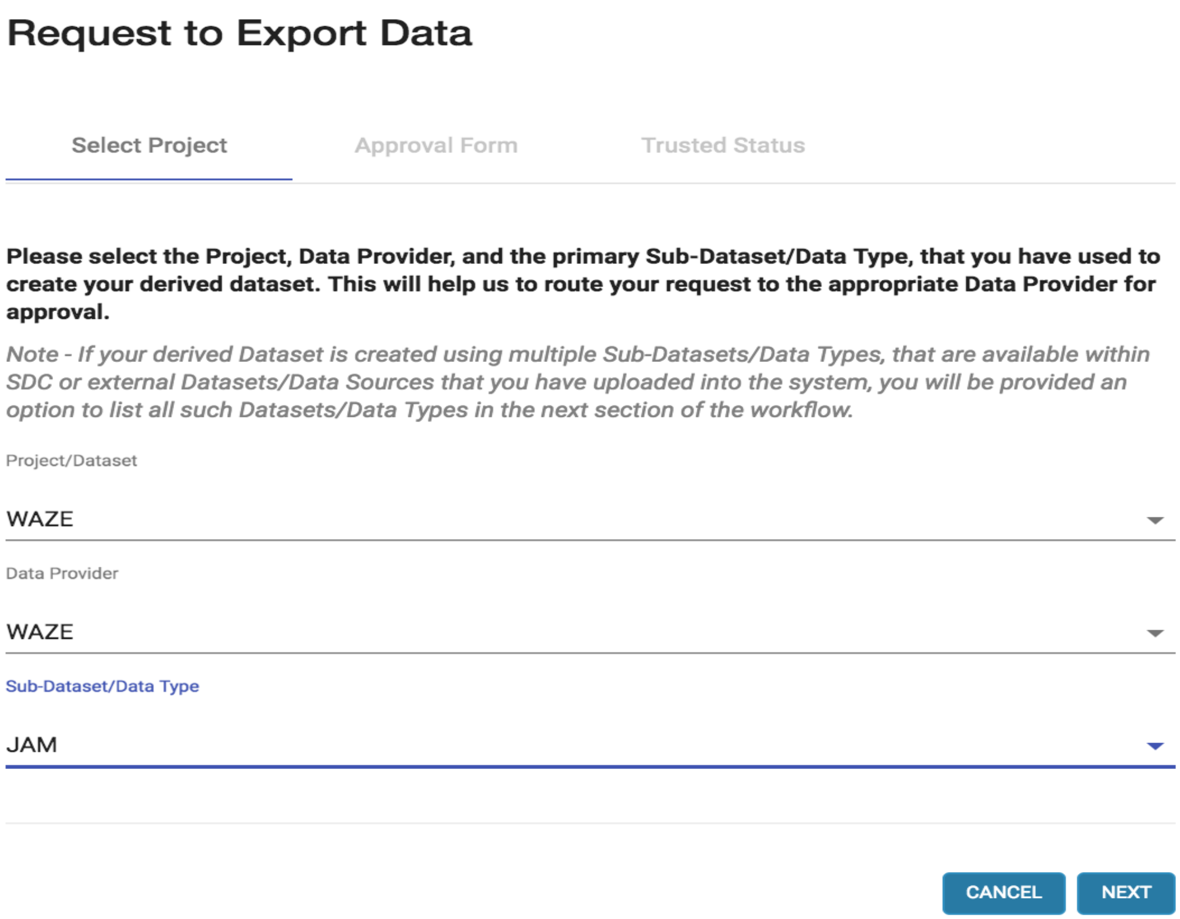
Files that you would like to export out of the system must be uploaded to the folder - **export_requests**

Any file type can be downloaded.

□	Filename	Export	Publish
□	export_requests/test3.zip	⇥	☁
□	export_requests/s3cmd2.zip	⇥	☁
□	export_requests/t.zip	⇥	☁
□	export_requests/demo9.zip	⇥	☁
□	sbapat/uploaded_files/architecture.png	⇥	☁
□	export_requests/s3cmd1.zip	⇥	☁
□	export_requests/wazeAnalysis.zip	⇥	☁
□	export_requests/dynamodbanalysis.zip	⇥	☁

Figure 18. Screenshot of the Datasets section (Source: REAN CLOUD)

2. Once the export button is clicked, a dialog box for requesting the export data will pop up. The analyst need to provide the details of the Project, Data Provider and Data Type that he has used to create his own dataset and click on the NEXT button.



Request to Export Data

Please select the Project, Data Provider, and the primary Sub-Dataset/Data Type, that you have used to create your derived dataset. This will help us to route your request to the appropriate Data Provider for approval.

Note - If your derived Dataset is created using multiple Sub-Datasets/Data Types, that are available within SDC or external Datasets/Data Sources that you have uploaded into the system, you will be provided an option to list all such Datasets/Data Types in the next section of the workflow.

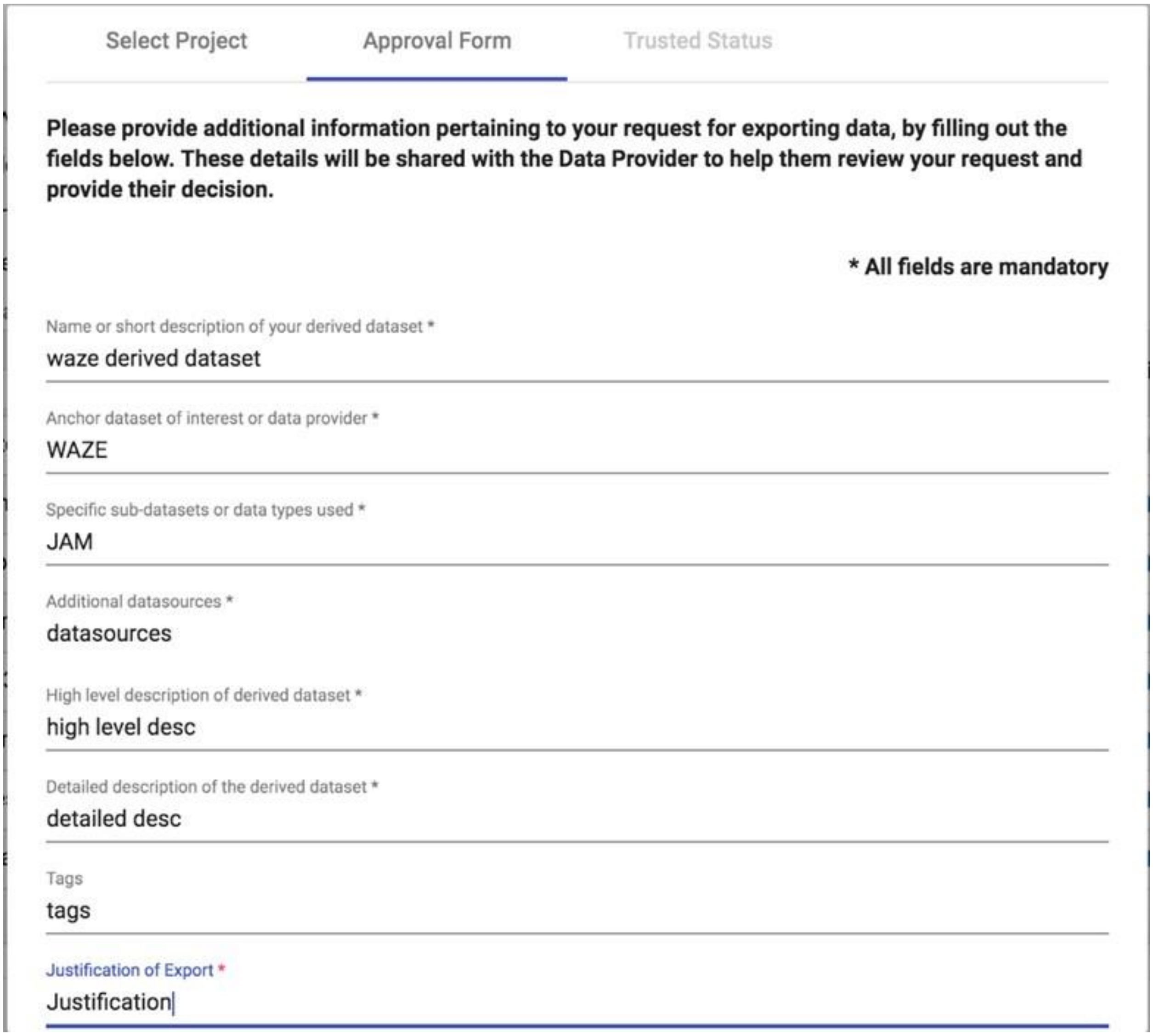
Project/Dataset: WAZE

Data Provider: WAZE

Sub-Dataset/Data Type: JAM

Figure 19. Screenshot of dialog box for requesting the export data (Source: REAN CLOUD)

3. The additional information regarding the request for exporting the data has to be filled in the approval form below. These details are shared with the data provider which helps them to accept or reject the request made by the data analyst.



Please provide additional information pertaining to your request for exporting data, by filling out the fields below. These details will be shared with the Data Provider to help them review your request and provide their decision.

*** All fields are mandatory**

Name or short description of your derived dataset *
waze derived dataset

Anchor dataset of interest or data provider *
WAZE

Specific sub-datasets or data types used *
JAM

Additional datasources *
datasources

High level description of derived dataset *
high level desc

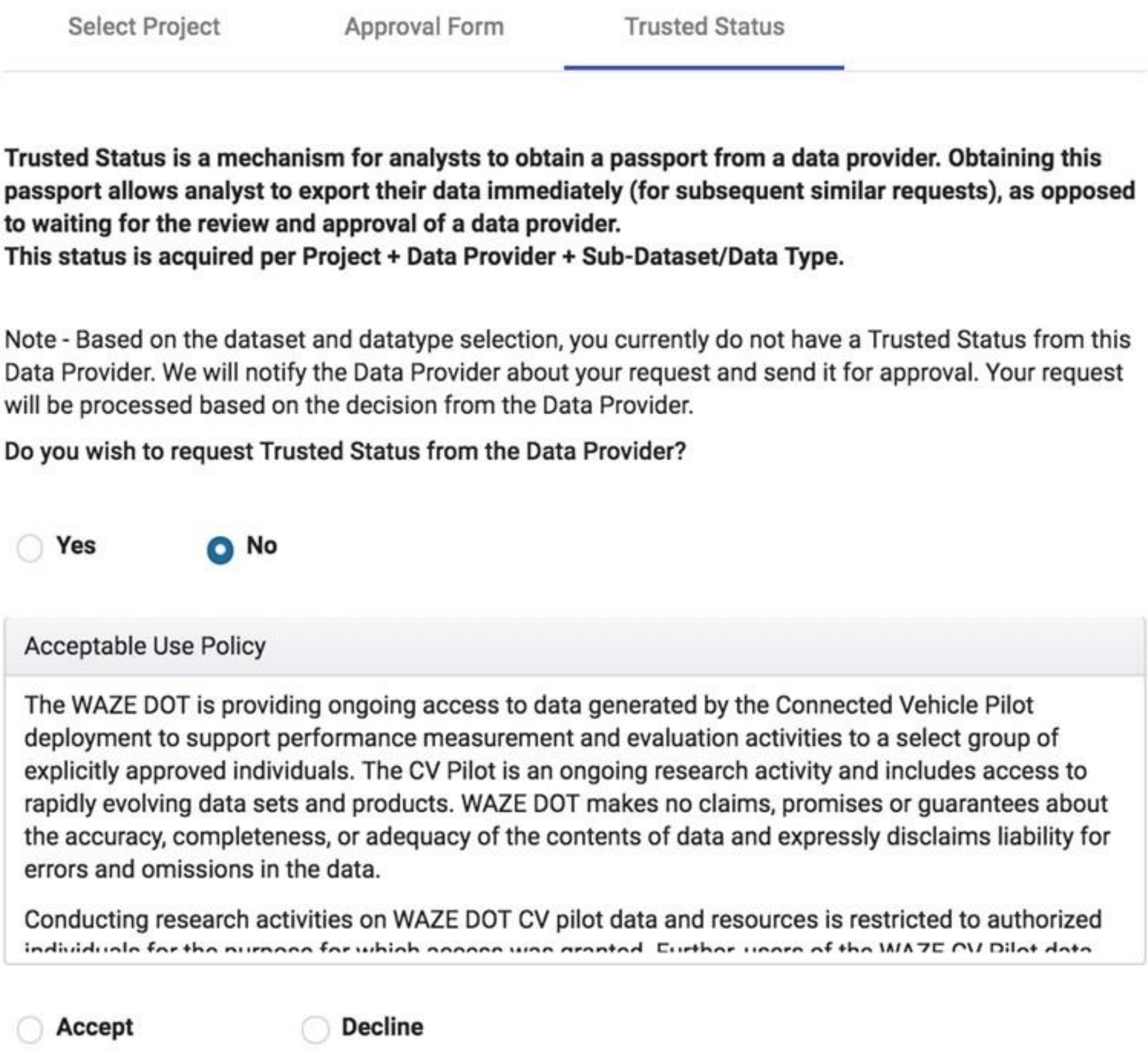
Detailed description of the derived dataset *
detailed desc

Tags
tags

Justification of Export *
Justification

Figure 20. Screenshot of Approval Form (Source: REAN CLOUD)

4. If the user is not a trusted user, he would be prompted with the option for requesting the trusted status from the data provider. This will allow the analyst to export the data immediately, as opposed to waiting for the review and approval from the data provider. The user has to accept the Acceptable Usage policy for the request to go through to the data provider.



Trusted Status is a mechanism for analysts to obtain a passport from a data provider. Obtaining this passport allows analyst to export their data immediately (for subsequent similar requests), as opposed to waiting for the review and approval of a data provider.
This status is acquired per Project + Data Provider + Sub-Dataset/Data Type.

Note - Based on the dataset and datatype selection, you currently do not have a Trusted Status from this Data Provider. We will notify the Data Provider about your request and send it for approval. Your request will be processed based on the decision from the Data Provider.

Do you wish to request Trusted Status from the Data Provider?

○ **Yes** ● **No**

Acceptable Use Policy

The WAZE DOT is providing ongoing access to data generated by the Connected Vehicle Pilot deployment to support performance measurement and evaluation activities to a select group of explicitly approved individuals. The CV Pilot is an ongoing research activity and includes access to rapidly evolving data sets and products. WAZE DOT makes no claims, promises or guarantees about the accuracy, completeness, or adequacy of the contents of data and expressly disclaims liability for errors and omissions in the data.

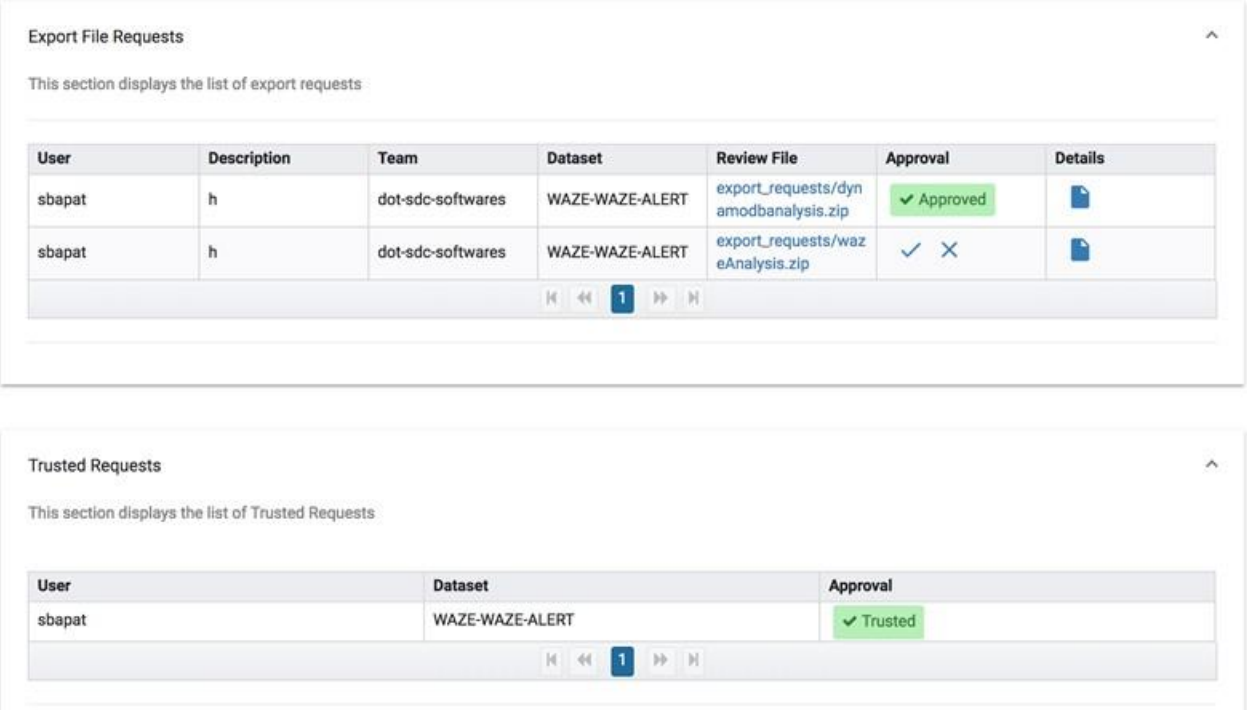
Conducting research activities on WAZE DOT CV pilot data and resources is restricted to authorized individuals for the purpose for which access was granted. Further, users of the WAZE CV Pilot data

○ **Accept** ○ **Decline**

Figure 21. Screenshot for requesting trusted status from the data provider (Source: REAN CLOUD)

5. Upon successful submission, the request will be sent to appropriate Data Providers. Data providers are responsible for accepting or rejecting the export requests for the data analysts so that they are able to get the data products out of the SDC system.
6. Once the Data Providers approves the request, data analysts will be able to download the dataset out of SDC through portal.

Data providers can see the requests in the EXPORT REQUESTS tab of the SDC web portal.



Export File Requests

This section displays the list of export requests

User	Description	Team	Dataset	Review File	Approval	Details
sbapat	h	dot-sdc-softwares	WAZE-WAZE-ALERT	export_requests/dynamodbanalysis.zip	✓ Approved	🗎
sbapat	h	dot-sdc-softwares	WAZE-WAZE-ALERT	export_requests/wazeAnalysis.zip	✓ ✕	🗎

Trusted Requests

This section displays the list of Trusted Requests

User	Dataset	Approval
sbapat	WAZE-WAZE-ALERT	✓ Trusted

Figure 22. Screenshot of Export Request tab (Source: REAN CLOUD)

There are two sections for the requests made by the data analysts:

1. Export File Requests: These requests correspond to the approval form submitted by the data analyst for exporting any data file out of the SDC system.
   The data provider can accept the request by clicking on the right mark and reject the request by clicking on the wrong mark as shown above.
   To get all the details regarding the request, the provider can click on the file symbol under the details column for each request.
   To review the file for which the request has been made, the data provider can click on the file link to download the file and review it before giving the access to the analyst.

2. Trusted Requests: These requests correspond to the request for getting the trusted status. The data provider can accept the request by clicking on the right mark and reject the request by clicking on the wrong mark.

Actions of Review:

1. Notify: The request of the data analyst for the trusted status or the export request is accepted automatically. The data provider is notified with an email for the request that is accepted.

2. NotifyReview: The request of the data analyst is sent to the data provider for approval by sending him a notification over an email. The data provider has to accept or reject the request from the SDC web portal under the section named "EXPORT REQUESTS"

Chapter 7. Technical Documentation and Contact Information

The following sections provide technical resources for SDC users.

Architecture Diagram

The figure below illustrates the SDC Architecture.

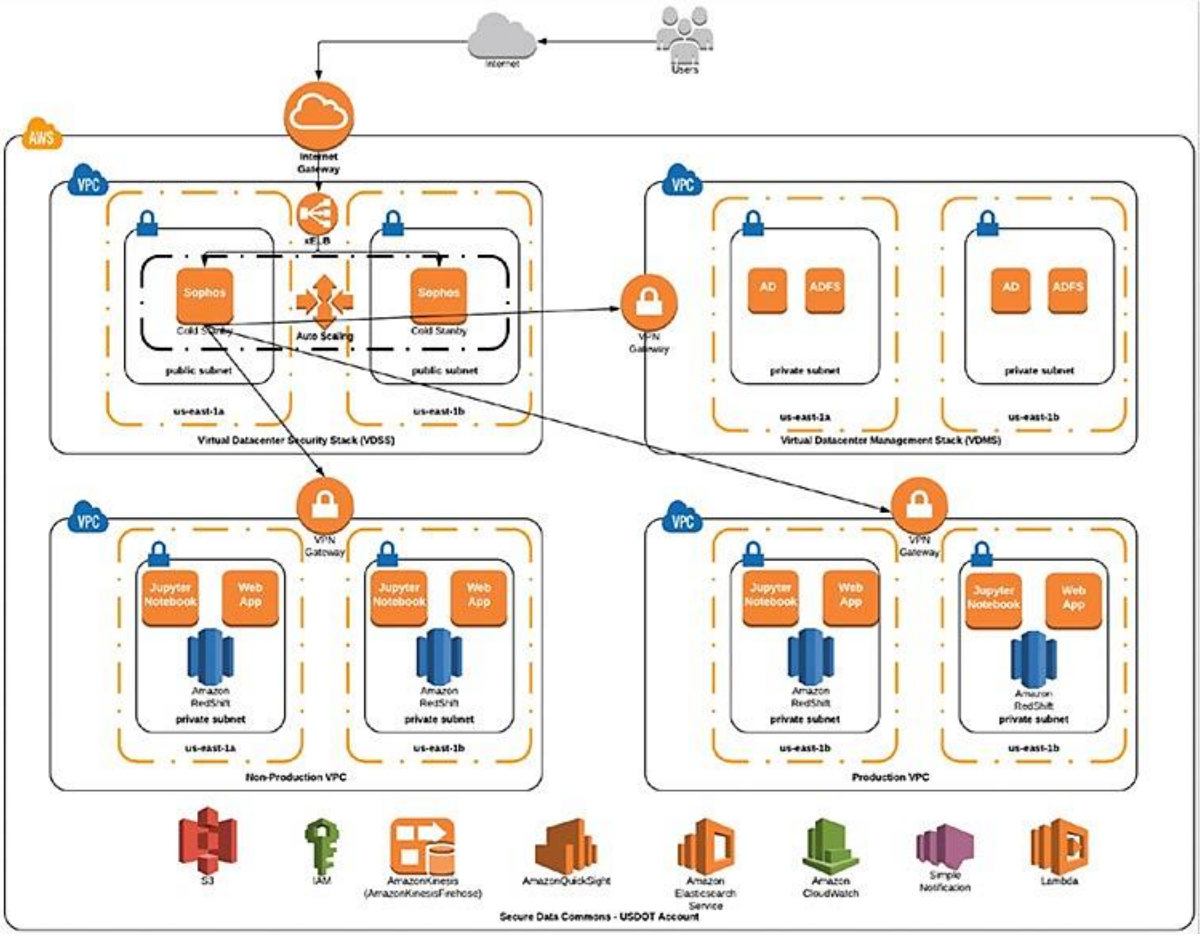


Figure 23. SDC Architecture Diagram (Source: REAN CLOUD)

Workstation Details

Table 1. Default Workstation details.

Work Station Type	Type
Linux Workstation	t2.medium
Windows Workstation	t2.medium

Note: Workstation size and type can be increased upon user request.

Tools and Versions

Table 2. List of tools used and their versions.

Tool Name	Version	Work Station
Java	1.8.0_162	Linux, Windows
Python	2.7.14	Linux, Windows
SQLWorkbench/J	Build 123	Windows
R	3.4.3	Linux, Windows
RStudio	1.1.423.0	Linux, Windows
Libre Office	5.3.6.1	Windows
Visual Studio	1.20.1.0	Windows
AWS CLI	1.14.46	Windows
7Zip	18.01	Windows
PuTTY	0.70	Windows
Firefox	59.0.2.0	Windows

Contact Information

SDC support team can be reached out at support@securedatacommons.com

Useful Links

- S3: Simple Storage Service, place to store data.
- Jupyter: An interactive, browser-based programming environment, mostly used for Python scripts but can also run R or other languages and can weave formatted text in with code and results of code into one ‘notebook’ file.
- Redshift: A database system, which can be queried with SQL.

AWS S3 CLI Commands

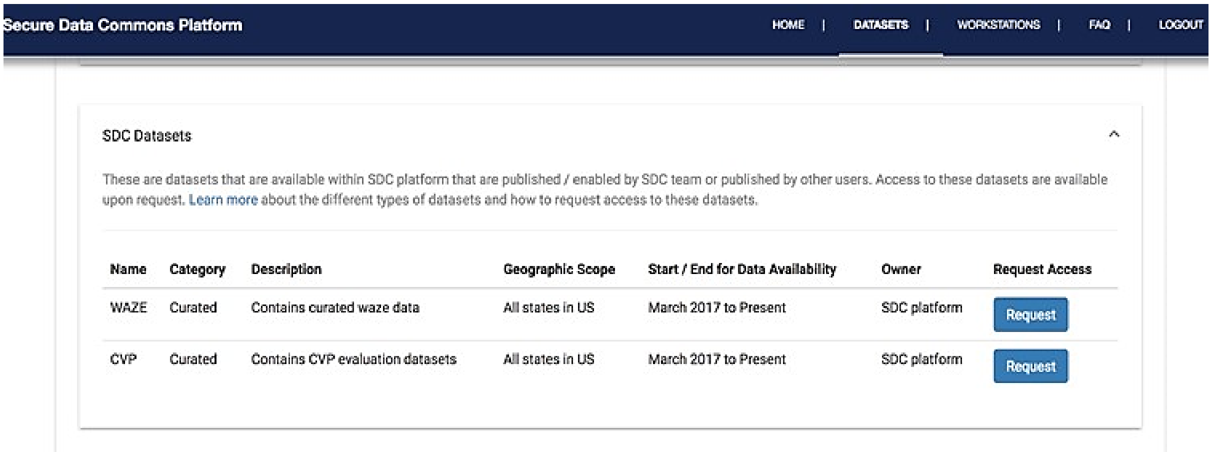
Following are the list of helpful commands to work with S3 from the terminal. In these commands, ‘local’ refers to the EC2 instance being run inside SDC:

- **List objects in a bucket -** If there are any files at the bucket level then the below command will return the list of files. If there are only folders/prefixes under the bucket, then it will return the top level folder/prefix names of that bucket
  aws s3 ls s3://<bucketName>
- **List objects under a folder/prefix -** The below command will list all the objects/files under that folder or prefix.
  aws s3 ls s3://<bucketName>/<prefix>/
- **Copy a local file to an s3 bucket -** This command will copy the file at the root of the bucket.
  aws s3 cp nohup.out s3://<bucketName>/
- **Copy a local file to a specific folder/prefix in an s3 bucket -** To create a new output folder and copy to that destination, simply add the desired path to the end, following bucket/.
  aws s3 cp <filename> s3://<bucketName>/<prefix>/
- **Get a file from S3 to the ec2 instance -** This command will help you get a file from S3 to the EC2 instance:
  aws s3 cp s3://<bucketName>/<prefix>/<fileName> .

Chapter 8. Frequently Asked Questions

How can I get access to the SDC Datasets?

These are datasets that are available within SDC platform that are published / enabled by SDC team or published by other users. Access to these datasets are available upon request.
Once you are logged in go to tab ‘Dataset’ in menu tab.


Secure Data Commons Platform
HOME | DATASETS | WORKSTATIONS | FAQ | LOGOUT

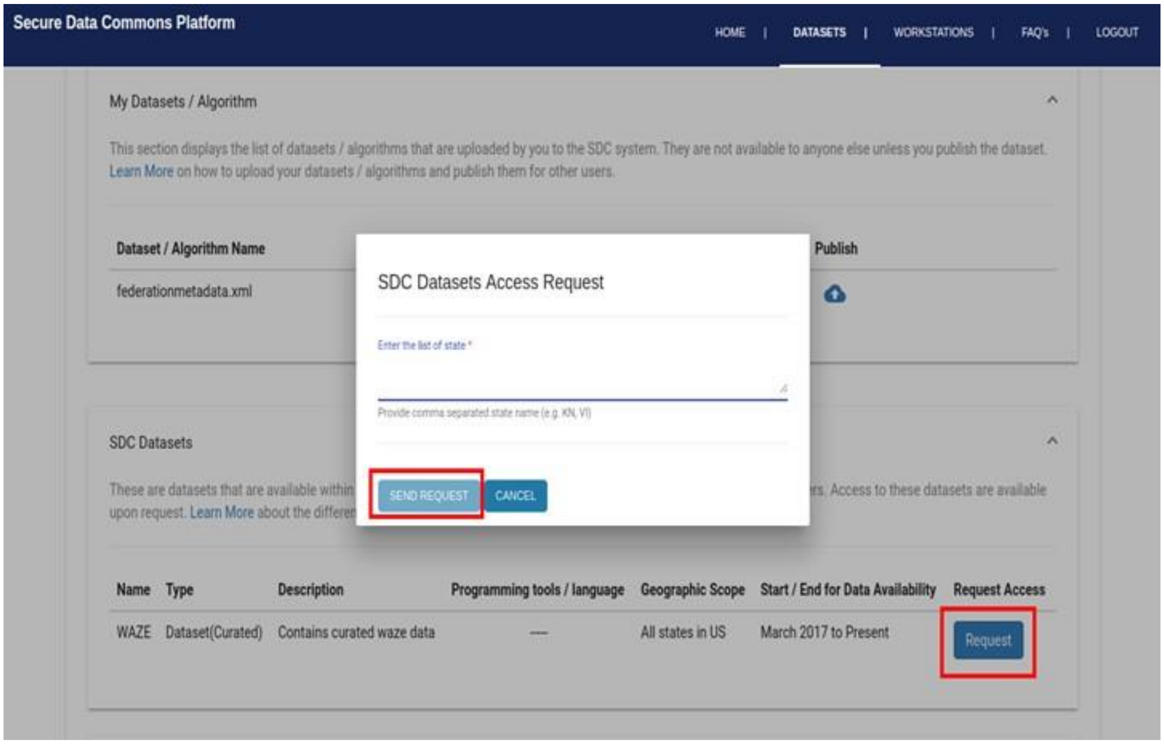
SDC Datasets

These are datasets that are available within SDC platform that are published / enabled by SDC team or published by other users. Access to these datasets are available upon request. Learn more about the different types of datasets and how to request access to these datasets.

Name	Category	Description	Geographic Scope	Start / End for Data Availability	Owner	Request Access
WAZE	Curated	Contains curated waze data	All states in US	March 2017 to Present	SDC platform	Request
CVP	Curated	Contains CVP evaluation datasets	All states in US	March 2017 to Present	SDC platform	Request

Figure 24. Screenshot of SDC Dataset tab (Source: REAN CLOUD)

After clicking on the ‘SDC Dataset’ Heading, you can see all the available datasets. To access these datasets, you need to click on ‘Request’ button. A form will pop up, fill that form and click on ‘Send Request’ button.


Secure Data Commons Platform
HOME | DATASETS | WORKSTATIONS | FAQ's | LOGOUT
SEND REQUEST
CANCEL
Request

My Datasets / Algorithm

This section displays the list of datasets / algorithms that are uploaded by you to the SDC system. They are not available to anyone else unless you publish the dataset. Learn More on how to upload your datasets / algorithms and publish them for other users.

Dataset / Algorithm Name	Publish
federationmetadata.xml	

SDC Datasets Access Request

Enter the list of state *

Provide comma separated state name (e.g. KN, VI)

SDC Datasets

These are datasets that are available within rs. Access to these datasets are available upon request. Learn More about the differen

Name	Type	Description	Programming tools / language	Geographic Scope	Start / End for Data Availability	Request Access
WAZE	Dataset(Curated)	Contains curated waze data	—	All states in US	March 2017 to Present	Request

Figure 25. Screenshot of SDC Dataset Access Request form (Source: REAN CLOUD)

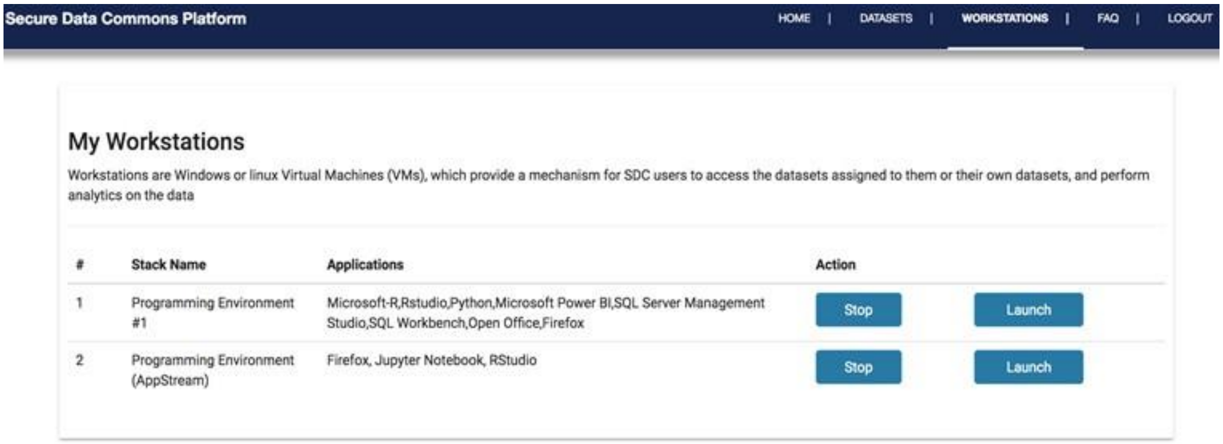
The request will be send to support team and access to the requested dataset will be given.

How will I understand what a particular dataset consists of?

Click on Name of Dataset, you can see README of that particular dataset below it.

How can I launch a workstation?

Click on 'Workstation' menu tab and click on 'Launch' button of any workstation which you want to access


Secure Data Commons Platform
HOME | DATASETS | WORKSTATIONS | FAQ | LOGOUT

My Workstations

Workstations are Windows or linux Virtual Machines (VMs), which provide a mechanism for SDC users to access the datasets assigned to them or their own datasets, and perform analytics on the data

#	Stack Name	Applications	Action	
1	Programming Environment #1	Microsoft-R,Rstudio,Python,Microsoft Power BI,SQL Server Management Studio,SQL Workbench,Open Office,Firefox	Stop	Launch
2	Programming Environment (AppStream)	Firefox, Jupyter Notebook, RStudio	Stop	Launch

Figure 26. Screenshot of Workstations tab (Source: REAN CLOUD).

For Programming Environment #1, you will be prompted with username and password to login to windows workstation.

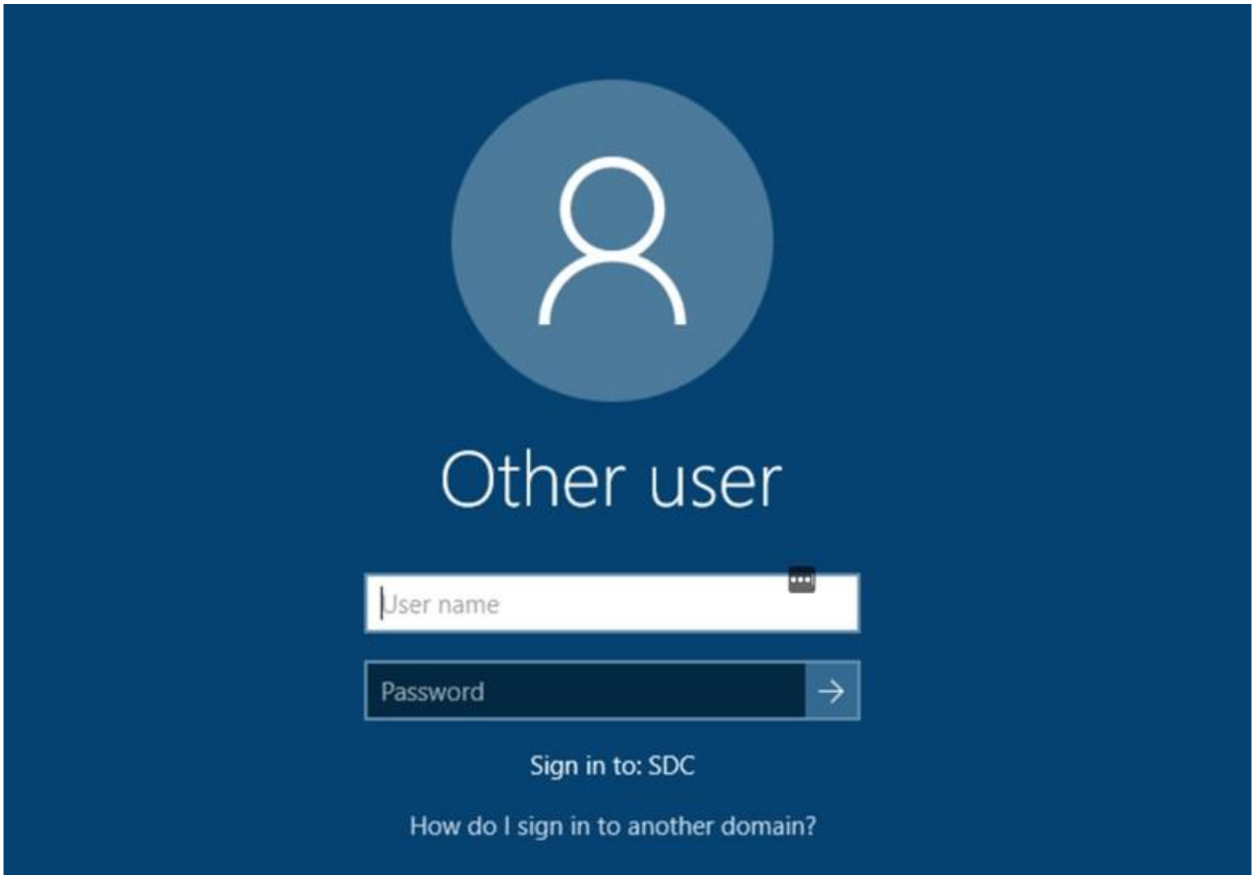


Figure 27. Screenshot of SDC login page for Programming Environment #1 (Source: REAN CLOUD)

Where do I store my data?

You can store your data in your team/individual bucket. Please refer to Upload User Data to S3 Bucket Through Portal

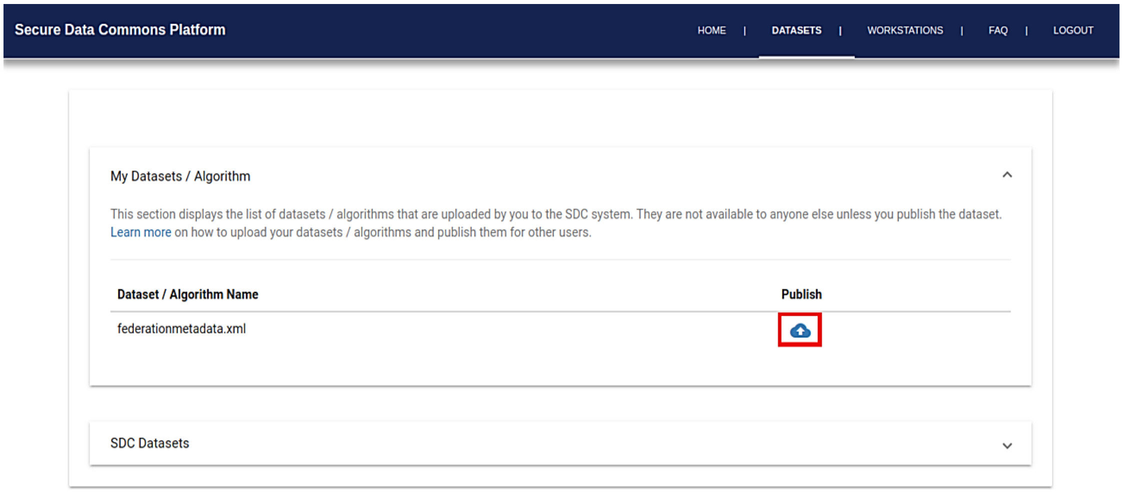
How can I bring my own datasets/algorithm to my workstation?

Please refer to Upload User Data to S3 Bucket Through Portal to bring your own datasets/algorithm to workstation.

How can I publish my dataset/algorithm?

Follow the below steps to publish your datasets / algorithms and share with other SDC Users. Navigate to Datasets page

1. Click on the upload icon under publish for the dataset/algorithm you wish to publish



Secure Data Commons Platform

My Datasets / Algorithm

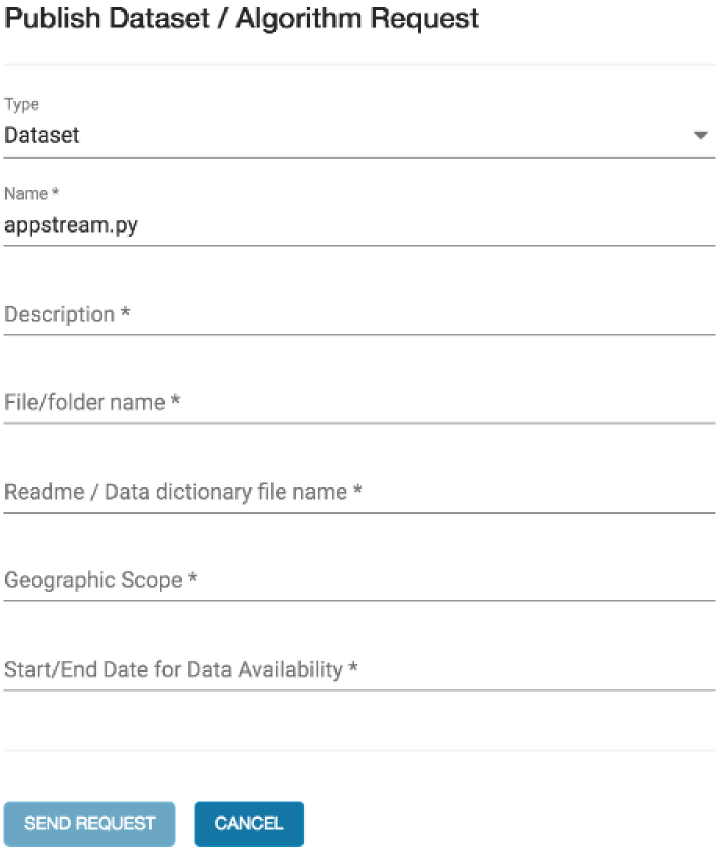
This section displays the list of datasets / algorithms that are uploaded by you to the SDC system. They are not available to anyone else unless you publish the dataset. Learn more on how to upload your datasets / algorithms and publish them for other users.

Dataset / Algorithm Name	Publish
federationmetadata.xml	

SDC Datasets

Figure 28. Screenshot of "upload" button within the Dataset tab (Source: REAN CLOUD)

2. In the pop-up window that shows up, you will have two options either a Dataset or Algorithm
   a. Select the dropdown value as Dataset and fill the required values



Publish Dataset / Algorithm Request

Type
Dataset

Name *
appstream.py

Description *

File/folder name *

Readme / Data dictionary file name *

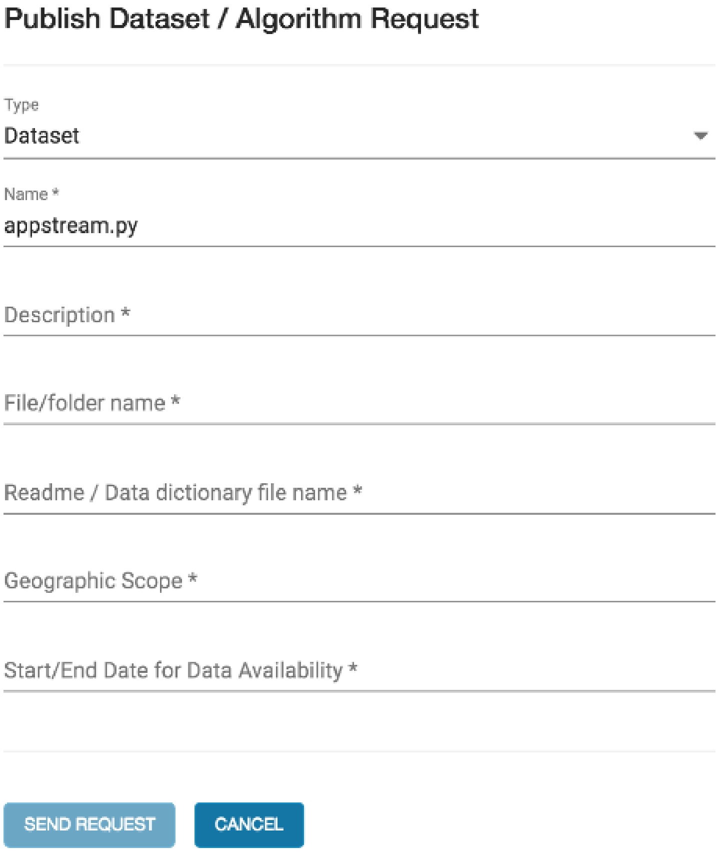
Geographic Scope *

Start/End Date for Data Availability *

Figure 29. Screenshot of Publish Dataset form (Source: REAN CLOUD)

i. Name - Name of the dataset, which you wish to call it. Users will see your dataset with this name under SDC Datasets section
ii. Description - Provide a short description so users can get an idea about your dataset
iii. File/folder name - Name of the file or folder where your dataset resides in your S3 Bucket. We need this information, so the support team can publish this dataset and make it available to other users
iv. Readme / Data dictionary file name - This file should provide detailed instructions about your dataset, how it was created or any relevant information that helps user to understand and use the dataset. Save this file in your home folder relative to the dataset file/folder name
v. Geographic scope - Indicate the geographic scope for your dataset whether it belongs to a specific state, region, country etc
vi. Start/End Date for data availability - Provide the start and end dates of the data that belongs in your dataset. For example, your dataset may contain data from March 2017 to August 2017

b. Select the dropdown value as Algorithm



Publish Dataset / Algorithm Request

Type
Dataset

Name *
appstream.py

Description *

File/folder name *

Readme / Data dictionary file name *

Geographic Scope *

Start/End Date for Data Availability *

Figure 30. Screenshot of Algorithm Request form (Source: REAN CLOUD)

i. Name - Enter the name for your algorithm. Users will see your algorithm with this name under SDC Datasets section
ii. Description - Provide a short description about your algorithm

iii. File/Folder name - Name of the file or folder where your algorithm resides in your S3 bucket. We need this information, so SDC support team can publish this algorithm and make it available to other users

iv. Readme / Data dictionary file name - This file should provide detailed instructions about your algorithm, how it was created or any relevant information that helps user to understand and use the algorithm. Save this file in your home folder relative to the algorithm file/folder name

v. Programming Tools/language - Provide the details of programming tools and/or languages that were used to create this algorithm, so users can leverage the same to run your program.

U.S. Department of Transportation
ITS Joint Program Office-HOIT
1200 New Jersey Avenue, SE
Washington, DC 20590

Toll-Free "Help Line" 866-367-7487
www.its.dot.gov

FHWA-JPO-18-XXX

U.S. Department of Transportation