Synapse Demo - File Uploads

Aryton Tediarjo

2021-03-17

This R markdown is used for showcasing basic interaction with Synapse using RStudio. It will cover on how to utilize our Synapse R client to do basic uploading/downloading files.

Import Libraries

```r
library(data.table)
library(tidyverse)
library(synapser)
library(synapserutils)
synLogin()
```

```
## Welcome, Aryton Tediarjo!

## NULL
```

Global Variables

```r
SYNAPSE_PARENT_ID <- "syn25257185" # where you will be saving all your files
FILEPATH <- file.path("data", "sample_1.tsv")
```

Part 1: Uploading to Synapse (1 File)

Step 1. Read Arbitrary Data

```r
data <- fread(FILEPATH) %>%
    tibble::as_tibble(.)
data
```

```
## # A tibble: 9 x 4
##   user        createdOn        x       y
##   <chr>       <date>       <dbl>   <dbl>
## 1 XiXkQAlsYH  2020-03-27  -0.802   0.508
## 2 UvLxhtBdz8  2020-04-04   0.378  -0.446
## 3 JFY3TTToMXA 2020-06-06  -0.664  -0.364
## 4 csHOlfCMMg  2020-08-08  -0.739  -0.137
```

```
## 5 KrmpbUmsCJ 2020-12-18  0.458  1.84
## 6 jane doe    2021-03-17  4      0
## 7 jane doe    2021-03-17  4      0
## 8 jane doe    2021-03-17  4      0
## 9 jane doe    2021-03-17  4      0
```

Step 2. Create File Entity Locally

```
file_entity <- synapser::File(FILEPATH, parentId = SYNAPSE_PARENT_ID)
file_entity
```

```
## File: sample_1.tsv (-)
##    md5=None
##    fileSize=None
##    contentType=None
##    externalURL=None
##    cacheDir=data
##    files=['sample_1.tsv']
##    path=data/sample_1.tsv
##    synapseStore=True
## properties:
##    concreteType=org.sagebionetworks.repo.model.FileEntity
##    name=sample_1.tsv
##    parentId=syn25257185
## annotations:
```

Step 3. Store to Synapse

```
synStore(file_entity)
```

```
## File: sample_1.tsv (syn25257188)
##    md5=77e434831fb7f969e5e03b0b2d78958e
##    fileSize=409
##    contentType=text/tab-separated-values
##    externalURL=None
##    cacheDir=data
##    files=['sample_1.tsv']
##    path=data/sample_1.tsv
##    synapseStore=True
## properties:
##    concreteType=org.sagebionetworks.repo.model.FileEntity
##    createdBy=3394042
##    createdOn=2021-03-17T18:43:49.599Z
##    dataFileHandleId=73914428
##    etag=09371171-6a52-4bee-a441-cecb5a03164b
##    id=syn25257188
##    isLatestVersion=True
##    modifiedBy=3394042
##    modifiedOn=2021-03-17T19:58:45.492Z
```

```
##   name=sample_1.tsv
##   parentId=syn25257185
##   versionLabel=37
##   versionNumber=37
## annotations:
```

Part 2: Files Versioning, Provenance, Annotations

File Entity Components:

Versioning: Let's say you want to add more data into the previously stored dataframe, Synapse will provide you the capability to version your files everytime you update it so that you can always access your previous data.

Provenance: You can also track the sources of how the data is generated

Annotations: Key-Value pair attached to a data

a. Versioning

```r
data_new <- data %>%
    add_row(tibble_row(
        user = "jane doe",
        createdOn = lubridate::as_date(lubridate::now()),
        x = 4, y = 0))
data_new %>%
    write.table(., FILEPATH, sep = "\t",
                row.names = FALSE,
                quote = FALSE)
```

b. Files Provenance

```r
USED_SYNAPSE_ID <- "syn23555388"
USED_NB_GIT_URL <- "https://github.com/arytontediarjo/synapse-demo/blob/main/file_upload.Rmd"

activity = synapser::Activity(
    "generating demo data",
    used = c(USED_SYNAPSE_ID),
    executed = c(USED_NB_GIT_URL)
)

store_to_synapse <- synStore(new_file_entity, activity = activity)
```

c. Annotations

```r
# try retrieving annotation
synGetAnnotations(store_to_synapse$properties$id)
```

```
## named list()
```

```r
#' set annotations
synSetAnnotations(
    store_to_synapse$properties$id,
    list("annotatedBy" = "Aryton"))
```

```
## $annotatedBy
## $annotatedBy[[1]]
## [1] "Aryton"
```

Part 3: Multi-File Uploads

This part will cover how to do bulk upload and bulk download to Synapse. It will use 'synapserutils' which is used for some utility use-cases built on top of 'synapser' Documentation can be found here: https://docs.synapse.org/articles/uploading_in_bulk.html

Bulk Upload

We will use a "manifest.tsv" (with pre-defined format requirement set in the documentation) file to bulk upload all your files.

```r
bulk_upload_data <- synapserutils::syncToSynapse("manifest.tsv")
```

```
## Validation and upload of: manifest.tsvValidating columns of manifest.....OKValidating that all paths
```

Bulk Download

We can also sync what has been downloaded using this command below

```r
bulk_download_data <- synapserutils::syncFromSynapse("syn25257395")
```