Group Project 2: Financial Data Modeling and Database Implementation

Part II: Database Creation

Making sure ipython-sql and psycopg2 libraries are installed so that we can use PostgreSQL.

In [1]:
```
!pip freeze | grep -E 'ipython-sql|psycopg2'
```

```
ipython-sql==0.4.1
psycopg2==2.9.5
psycopg2-binary==2.9.5
```

Dropping DB named SECDB (using PostgreSQL's dropdp command) if it already exists to ensure that this notebook can be run repeatedly.

In [2]:
```
!dropdb -U student SECDB
```

Using PostgreSQL's createdb command to create a database named SECDB

In [3]:
```
!createdb -U student SECDB
```

We want to utilize the query tools that SQL provides. To do so, we are installing the SQL extension to be able to run SQL in this notebook.

In [4]:
```
%load_ext sql
```

Connecting the SECDB database that we just created using the student user.

In [5]:
```
%sql postgresql://student@/SECDB
```

Creating Tables

Creating SUB Table

Use DROP TABLE command before creating the table SUB. This is good practice in case table SUB already exists and allows us to run the notebook repeatedly.

Use CREATE TABLE command to create table SUB that contains each entity and specifies the data type of each entity (and whether or not the column can take null values).

The SUB table does not contain any foreign keys. It does however contain a **primary key**, ADSH, which is referenced by other tables. ADSH uniquely identifies each row of SUB.

In [6]:
```
%%sql
DROP TABLE IF EXISTS SUB Cascade;
```

```
CREATE TABLE SUB (
    ADSH         VARCHAR(20) NOT NULL,
    CIK          NUMERIC(10,0) NOT NULL,
    NAME         VARCHAR(150) NOT NULL,
    SIC          NUMERIC(4,0) NULL,
    COUNTRYBA    VARCHAR(2) NOT NULL,
    STPRBA       VARCHAR(2) NULL,
    CITYBA       VARCHAR(30) NOT NULL,
    ZIPBA        VARCHAR(10) NULL,
    BAS1         VARCHAR(40) NULL,
    BAS2         VARCHAR(40) NULL,
    BAPH         VARCHAR(20) NULL,
    COUNTRYMA    VARCHAR(2) NULL,
    STPRMA       VARCHAR(2) NULL,
    CITYMA       VARCHAR(30) NULL,
    ZIPMA        VARCHAR(10) NULL,
    MAS1         VARCHAR(40) NULL,
    MAS2         VARCHAR(40) NULL,
    COUNTRYINC   VARCHAR(2) NOT NULL,
    STPRINC      VARCHAR(2) NULL,
    EIN          NUMERIC(10,0) NULL,
    FORMER       VARCHAR(150) NULL,
    CHANGED      VARCHAR(8) NULL,
    AFS          VARCHAR(5) NULL,
    WKSI         BOOLEAN NOT NULL,
    FYE          VARCHAR(4) NOT NULL,
    FORM         VARCHAR(10) NOT NULL,
    PERIOD       DATE NOT NULL,
    FY           NUMERIC(4,0) NOT NULL,
    FP           VARCHAR(2) NOT NULL,
    FILED        DATE NOT NULL,
    ACCEPTED     TIMESTAMP NOT NULL,
    PREVRPT      BOOLEAN NOT NULL,
    DETAIL       BOOLEAN NOT NULL,
    INSTANCE     VARCHAR(32) NOT NULL,
    NCIKS        NUMERIC(4,0) NOT NULL,
    ACIKS        VARCHAR(120) NULL,
    PUBFLOATUSD    NUMERIC(8,0) NULL,
    FLOATDATE    DATE NULL,
    FLOATAXIS    VARCHAR(255) NULL,
    FLOATMEMS    NUMERIC(1,0) NULL,

    PRIMARY KEY (ADSH)
);
```

```
 * postgresql://student@/SECDB
Done.
Done.
```

Out[6]: []

Adding a comment to table SUB

In [7]:
```
%%sql
COMMENT ON TABLE SUB IS
'SUB table: Submission data set includes one record for each XBRL submissior
```

```
 * postgresql://student@/SECDB
Done.
```

Out[7]: []

Adding comments for each column of SUB

In [8]:
```
%%sql
COMMENT ON COLUMN SUB.adsh IS 'adsh: is an accession number, a 20-character
COMMENT ON COLUMN SUB.cik IS 'cik: is Central Index Key (CIK) assigned by th
COMMENT ON COLUMN SUB.name IS 'name: name of the registrant (legal entity) a
COMMENT ON COLUMN SUB.sic IS 'sic: Standard Industrial Classification (SIC)
COMMENT ON COLUMN SUB.countryba IS 'countryba: ISO 3166-1 country code of th
COMMENT ON COLUMN SUB.stprba IS 'stprba: State or province of the registrant
COMMENT ON COLUMN SUB.cityba IS 'cityba: City of the registrant''s business
COMMENT ON COLUMN SUB.zipba IS 'zipba: Zip code of the registrant''s busines
COMMENT ON COLUMN SUB.bas1 IS 'bas1: First line of the street of the registr
COMMENT ON COLUMN SUB.bas2 IS 'bas2: Second line of the street of the regist
COMMENT ON COLUMN SUB.baph IS 'baph: Phone number of the registrant''s busir
COMMENT ON COLUMN SUB.countryma IS 'countryma: ISO 3166-1 country code of th
COMMENT ON COLUMN SUB.stprma IS 'stprma: State or province of the registrant
COMMENT ON COLUMN SUB.cityma IS 'cityma: City of the registrant''s mailing a
COMMENT ON COLUMN SUB.zipma IS 'zipma: Zip code of the registrant''s mailing
COMMENT ON COLUMN SUB.mas1 IS 'mas1: First line of the street of the registr
COMMENT ON COLUMN SUB.mas2 IS 'mas2: Second line of the street of the regist
COMMENT ON COLUMN SUB.countryinc IS 'countryinc: Country of incorporation fo
COMMENT ON COLUMN SUB.stprinc IS 'stprinc:  State or province of incorporati
COMMENT ON COLUMN SUB.ein IS 'ein:  Employee Identification Number (EIN) ass
COMMENT ON COLUMN SUB.former IS 'former: Most recent former name of the regi
COMMENT ON COLUMN SUB.changed IS 'changed: Date of change from the former na
COMMENT ON COLUMN SUB.afs IS 'afs: Filer status with the Commission at the t
COMMENT ON COLUMN SUB.wksi IS 'wksi: Well Known Seasoned Issuer (WKSI) statu
COMMENT ON COLUMN SUB.fye IS 'fye: Fiscal Year End Date (format: MMDD). fye
COMMENT ON COLUMN SUB.form IS 'form: The submission type of the registrant''
COMMENT ON COLUMN SUB.period IS 'period: Balance Sheet Date (format: YYYY-MM
COMMENT ON COLUMN SUB.fy IS 'fy: Fiscal Year Focus (the year of the financia
COMMENT ON COLUMN SUB.fp IS 'fp: Fiscal Period Focus (indicating the reporti
COMMENT ON COLUMN SUB.filed IS 'filed: The date of the registrant''s filing
COMMENT ON COLUMN SUB.accepted IS 'accepted: The acceptance date and time of
COMMENT ON COLUMN SUB.prevrpt IS 'prevrpt: TRUE indicates that the submissio
COMMENT ON COLUMN SUB.detail IS 'detail: TRUE indicates that the XBRL submis
COMMENT ON COLUMN SUB.instance IS 'instance: Name of the submitted XBRL Inst
COMMENT ON COLUMN SUB.nciks IS 'nciks: Number of Central Index Keys (CIK) of
COMMENT ON COLUMN SUB.aciks IS 'aciks: Additional CIKs of co-registrants inc
COMMENT ON COLUMN SUB.pubfloatusd IS 'pubfloatusd: Public float, in USD, if
COMMENT ON COLUMN SUB.floatdate IS 'floatdate: Date on which the public floa
COMMENT ON COLUMN SUB.floataxis IS 'floataxis: If the public float value was
COMMENT ON COLUMN SUB.floatmems IS 'floatmems: Public float, in USD, if prov
```

```
 * postgresql://student@/SECDB
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
```

Out[8]: []

Retrieve table comment using pg_class

In [9]:

```
%%sql
SELECT obj_description(oid, 'pg_class')
FROM pg_class
WHERE upper(relname) = 'SUB';
```

```
 * postgresql://student@/SECDB
1 rows affected.
```

Out[9]:

obj_description
SUB table: Submission data set includes one record for each XBRL submission. The set includes fields of information pertinent to the submission and the filing entity

Retrieve column comments using information_schema.columns and pg_class system views

In [10]:
```
%%sql
SELECT cols.column_name,
pg_catalog.col_description(c.oid, cols.ordinal_position::int)
FROM pg_catalog.pg_class c, information_schema.columns cols
 WHERE cols.table_name = c.relname  and upper(relname) = 'SUB';
```

```
 * postgresql://student@/SECDB
40 rows affected.
```

Out[10]:

column_name	col_description
adsh	adsh: is an accession number, a 20-character string formed from the 18-digit number assigned by the Commission to each EDGAR submission. Where the 20, is the max-size in length that the column can have and the value is NOT NULL because all the columns contain valid entries. adsh is classified as CHAR because the value is in fixed length
cik	cik: is Central Index Key (CIK) assigned by the Commission to each registrant. cik is classified as NUMERIC (10,0) because it makes it explicit that only integers (whole numbers) are stored in this column and precision (10) is the total number of digits the number can have. The max-size length is 10 and the value is NOT NULL because all the columns contain valid entries
name	name: name of the registrant (legal entity) as recorded in EDGAR as of the filing date. name is classified as VARCHAR because it is not fixed in length but its max-size in length is 150. The value is NOT NULL because all the columns contain valid entries
sic	sic: Standard Industrial Classification (SIC) code assigned by the Commission, indicating the registrant's type of business. sic is classified as NUMERIC (4,0) because it makes it explicit that only integers (whole numbers) are stored in this column and precision (4) is the total number of digits the number can have. The max-size length is 4. The value is NULL because allowing the SIC column to be NULL provides flexibility in situations where the SIC code is not available, applicable, or required
countryba	countryba: ISO 3166-1 country code of the registrant's business address. countryba is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC. The max-size length is 2 and the value is NOT NULL because all the columns contain valid entries
stprba	stprba: State or province of the registrant's business address (if COUNTRYBA is US or CA). stprba is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC. The max-size length is 2 and the value is NULL because there may be cases where a registrant in the US or CA does not provide a specific state or province, rendering the field optional or incomplete
cityba	cityba: City of the registrant's business address. cityba is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size length is 30 and the value is NOT NULL because all the columns contain valid entries
zipba	zipba: Zip code of the registrant's business address. zipba is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size length is 10 and the value is NULL because even though the cityba is NOT NULL, meaning a city must always be provided, the zipba could still be NULL if the zip code was missing, incomplete, or incorrectly formatted during data entry
bas1	bas1: First line of the street of the registrant's business address. bas1 is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size length is 40. The value is NULL because some of the first line of the registrant's business address may not be provided or left incomplete during data entry
bas2	bas2: Second line of the street of the registrant's business address. bas2 is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size length is 40. The value is NULL because some of the second line of the registrant's business address may not be provided or left incomplete during data entry
baph	baph: Phone number of the registrant's business address. baph is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size length is 20. The value is NULL because some of the phone number of the registrant's business address may not be provided or left incomplete during data entry
countryma	countryma: ISO 3166-1 country code of the registrant's mailing address. countryma is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-

column_name	col_description
	size length is 2. The value is NULL because some of ISO 3166-1 country code of the registrant's mailing address may not be provided or left incomplete during data entry
stprma	stprma: State or province of the registrant's mailing address (if COUNTRYMA is US or CA). stprma is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size length is 2. The value is NULL because some of state or province of the registrant's mailing address may not be provided or left incomplete during data entry
cityma	cityma: City of the registrant's mailing address. cityma is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size length is 30. The value is NULL because some of city of the registrant's mailing address may not be provided or left incomplete during data entry
zipma	zipma: Zip code of the registrant's mailing addres. zipma is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size length is 10. The value is NULL because some of zipcode of the registrant's mailing address may not be provided or left incomplete during data entry
mas1	mas1: First line of the street of the registrant's mailing address. mas1 is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size length is 40. The value is NULL because some of the first line of the street of the registrant's mailing address may not be provided or left incomplete during data entry
mas2	mas2: Second line of the street of the registrant's mailing address. mas2 is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size length is 40. The value is NULL because some of the second line of the street of the registrant's mailing address may not be provided or left incomplete during data entry
countryinc	countryinc: Country of incorporation for the registrant (ISO 3166-1 ALPHA 2). countryinc is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC. The max-size length is 3 and the value is NOT NULL because all the columns contain valid entries
stprinc	stprinc: State or province of incorporation for the registrant (if COUNTRYINC is US or CA). stprinc is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size length is 2. The value is NULL because some of the state or province of incorporation for the registrant may not be provided or left incomplete during data entry
ein	ein: Employee Identification Number (EIN) assigned by the IRS to business entities operating in the US. ein is classified as NUMERIC (10,0) because it makes it explicit that only integers (whole numbers) are stored in this column and precision (10) is the total number of digits the number can have. The max-size length is 10. The value is NULL because some of the EIN may not be provided or left incomplete during data entry
former	former: Most recent former name of the registrant, if any. former is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size length is 150. The value is NULL because some of the most recent former name of the registrant may not be provided or left incomplete during data entry. There can also be cases where there isn't any
changed	changed: Date of change from the former name, if any. changed is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size length is 8. The value is NULL because some of date may not have been provided or left incomplete during data entry. There can also be cases where there isn't any

column_name	col_description
afs	afs: Filer status with the Commission at the time of submission (e.g., LAF, ACC, SRA, NON, SML). afs is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size length is 5. The value is NULL because some of filer status may not have been provided or left incomplete during data entry
wksi	wksi: Well Known Seasoned Issuer (WKSI) status; TRUE if the issuer meets specific Commission requirements. afs is classified as BOOLEAN because it indicates a binary state, typically representing whether a business entity qualifies as a Well-Known Seasoned Issuer or not, which can only be true or false (1 or 0). The max-size length is 1. The value is NOT NULL because all the columns contain valid entries
fye	fye: Fiscal Year End Date (format: MMDD). fye is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size length is 4. The value is NOT NULL because all the columns contain valid entries
form	form: The submission type of the registrant's filing (e.g., 10-K, 10-Q). form is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size length is 10. The value is NOT NULL because all the columns contain valid entries
period	period: Balance Sheet Date (format: YYYY-MM-DD). period is classified as DATE because it stores the value as (yymmdd) in the column. The max-size length is 8. The value is NOT NULL because all the columns contain valid entries
fy	fy: Fiscal Year Focus (the year of the financial report). fy is classified as NUMERIC because it stores the value as (yyyy) in the column. The max-size length is 4. The value is NOT NULL because all the columns contain valid entries
fp	fp: Fiscal Period Focus (indicating the reporting period, e.g., FY, Q1, Q2). fp is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size length is 2. The value is NOT NULL because all the columns contain valid entries
filed	filed: The date of the registrant's filing with the Commission. filed is classified as DATE because it stores the value as (yymmdd) in the column. The max-size length is 8. The value is NOT NULL because all the columns contain valid entries
accepted	accepted: The acceptance date and time of the registrant's filing; filings accepted after 5:30pm EST are considered filed the next business day. accepted is classified as DATETIME because it is stored like this: (yyyy-mm-dd hh:mm:ss) in the column. The maz-size length is 19. The value is NOT NULL because all the columns contain valid entries
prevrpt	prevrpt: TRUE indicates that the submission information was subsequently amended prior to the end cutoff date of the data set. prevrpt is classified as boolean because it represents a binary condition, where the value indicates whether the submission information has been amended (1 for true) or not amended (0 for false). The max-size length is 1. The value is NOT NULL because all the columns contain valid entries
detail	detail: TRUE indicates that the XBRL submission contains quantitative disclosures within the footnotes and schedules at the required detail level. detail is classified as BOOLEAN because it indicates whether the XBRL submission includes quantitative disclosures at the required detail level, with 1 representing true (it does) and 0 representing false (it does not). The max-size length is 1. The value is NOT NULL because all the columns contain valid entries
instance	instance: Name of the submitted XBRL Instance Document (EX-101.INS) type data file, often beginning with the company. instance is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size length is 32. The value is NOT NULL because all the columns contain valid entries
nciks	nciks: Number of Central Index Keys (CIK) of registrants (i.e., business units) included in the consolidating entity's submitted filing. nciks is classified as NUMERIC (4,0) because it makes it explicit that only integers (whole numbers) are stored in this column and precision (4) is the total number of digits the number

column_name	col_description
	can have. The max-size length is 4. The value is NOT NULL because all the columns contain valid entries
aciks	aciks: Additional CIKs of co-registrants included in a consolidating entity's EDGAR submission, separated by spaces. aciks is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size length is 120. The value is NULL because if there are no other coregistrants (i.e., nciks = 1), the value of aciks is NULl
pubfloatusd	pubfloatusd: Public float, in USD, if provided in this submission. pubfloatusd is classified as NUMERIC (14,0) because it makes it explicit that only integers (whole numbers) are stored in this column and precision (14) is the total number of digits the number can have. The max-size length is 8. The pubfloatusd field may be null because the public float value was not disclosed or provided in the submission
floatdate	floatdate: Date on which the public float was measured by the filer. floatdate is classified as DATE because it stores the value as (yymmdd) in the column. The max-size length is 8. The value is NULL because The floatdate field may be null because the filer did not specify the date on which the public float was measured
floataxis	floataxis: If the public float value was computed by summing across several tagged values, this indicates the nature of the summation. floataxis is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size length is 255.The floataxis field may be null because some values may not be provided or left incomplete during data entry
floatmems	floatmems: Public float, in USD, if provided in this submission. floatmems is classified as NUMERIC (1,0) because it makes it explicit that only integers (whole numbers) are stored in this column and precision (1) is the total number of digits the number can have. The max-size length is 1. The field may be null because the some values may not be provided or left incomplete during data entry

Creating DIM Table

Use DROP TABLE command before creating the table DIM.

Use CREATE TABLE command to create table DIM that contains each entity and specifies the data type of each entity (and whether or not the column can take null values).

DIM also does not contain any foreign key but does contain the **primary key**, DIMHASH, which is used by the NUM table to reference DIM. DIMHASH uniquely identifies each row of DIM.

In [11]:
```
%%sql
DROP TABLE IF EXISTS DIM Cascade;

CREATE TABLE DIM (
    DIMHASH VARCHAR(32) NOT NULL,
    SEGMENTS VARCHAR(1024) NOT NULL,
    SEGT BOOLEAN NOT NULL,

    PRIMARY KEY (DIMHASH)
);
```

```
 * postgresql://student@/SECDB
Done.
Done.
```

Out[11]: []

Adding comment to table DIM

In [12]:
```
%%sql
COMMENT ON TABLE DIM IS
'DIM table: The DIM set contains all of the combinations of XBRL axis and me
```

```
 * postgresql://student@/SECDB
Done.
```

Out[12]: []

Adding comments to each column of DIM

In [13]:
```
%%sql
COMMENT ON COLUMN DIM.dimhash IS 'dimhash: MD5 hash of the segments field te
COMMENT ON COLUMN DIM.segments IS 'segments: Concatenation of tag names repr
COMMENT ON COLUMN DIM.segt IS 'segt: TRUE if the segments field would have b
```

```
 * postgresql://student@/SECDB
Done.
Done.
Done.
```

Out[13]: []

Retrieve table comment using pg_class

In [14]:
```
%%sql
SELECT obj_description(oid, 'pg_class')
FROM pg_class
WHERE upper(relname) = 'DIM';
```

```
 * postgresql://student@/SECDB
1 rows affected.
```

Out[14]:

obj_description
DIM table: The DIM set contains all of the combinations of XBRL axis and member used to tag any submission

Retrieve column comments using information_schema.columns and pg_class system views

In [15]:
```
%%sql
SELECT cols.column_name,
pg_catalog.col_description(c.oid, cols.ordinal_position::int)
FROM pg_catalog.pg_class c, information_schema.columns cols
 WHERE cols.table_name = c.relname  and upper(relname) = 'DIM';
```

```
 * postgresql://student@/SECDB
3 rows affected.
```

Out[15]:

column_name	col_description
dimhash	dimhash: MD5 hash of the segments field text, used as a unique identifier for dimension information. dimh is classified as VARCHAR because it represents a string of characters generated by the MD5 hash function, which is expressed in hexadecimal format. The maz-size in length is 32. The value is NOT NULL because all the columns contain valid entries
segments	segments: Concatenation of tag names representing the axis and members appearing in the XBRL segments. segments is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The maz-size in length is 1024. The value is NOT NULL because all the columns contain valid entries
segt	segt: TRUE if the segments field would have been longer than 1024 characters had it not been truncated, otherwise FALSE. segt is classified as BOOLEAN because it indicates whether the segments field was truncated due to exceeding 1024 characters, with TRUE representing that it was truncated and FALSE indicating it was not. The maz-size in length is 1. The value is NOT NULL because all the columns contain valid entries

Creating TAG Table

Use DROP TABLE command before creating the table TAG.

Use CREATE TABLE command to create table TAG that contains each entity and specifies the data type of each entity (and whether or not the column can take null values).

TAG also does not contain any foreign key but does contain a **primary key** which consists of a combination of TAG and VERSION which uniquely identifies each row in TAG. The entity TAG is referenced by both NUM and PRE through keys TAG and VERSION.

In [16]:
```
%%sql
DROP TABLE IF EXISTS TAG Cascade;

CREATE TABLE TAG (
    tag VARCHAR(256) NOT NULL,
    version VARCHAR(20) NOT NULL,
    custom BOOLEAN NOT NULL,
    abstract BOOLEAN NOT NULL,
    datatype VARCHAR(20) NULL,
    iord VARCHAR(1) NOT NULL,
    crdr VARCHAR(1) NULL,
    tlabel VARCHAR(512) NULL,
    doc VARCHAR(2048) NULL,

    PRIMARY KEY (tag, version)
);
```

```
 * postgresql://student@/SECDB
Done.
Done.
```

Out[16]:
```
[]
```

Adding comment to table TAG

In [17]:
```
%%sql
COMMENT ON TABLE TAG IS
```

```
'TAG table: The TAG data set contains all standard taxonomy tags, not just t
```

```
 * postgresql://student@/SECDB
Done.
```

Out[17]: []

Adding comments to each column of TAG

In [18]:
```
%%sql
COMMENT ON COLUMN TAG.tag IS 'tag: The unique identifier (name) for a tag ir
COMMENT ON COLUMN TAG.version IS 'version: Identifier for the taxonomy versi
COMMENT ON COLUMN TAG.custom IS 'custom: Indicates if the tag is custom (TRU
COMMENT ON COLUMN TAG.abstract IS 'abstract: Indicates if the tag is abstrac
COMMENT ON COLUMN TAG.datatype IS 'datatype: Data type of the tag (e.g., mor
COMMENT ON COLUMN TAG.iord IS 'iord: Indicates if the value is a point in ti
COMMENT ON COLUMN TAG.crdr IS 'crdr: The natural accounting balance of the t
COMMENT ON COLUMN TAG.tlabel IS 'tlabel: The label text provided by the taxo
COMMENT ON COLUMN TAG.doc IS 'doc: Detailed definition for the tag, truncate
```

```
 * postgresql://student@/SECDB
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
```

Out[18]: []

Retrieve table comment using pg_class

In [19]:
```
%%sql
SELECT obj_description(oid, 'pg_class')
FROM pg_class
WHERE upper(relname) = 'TAG';
```

```
 * postgresql://student@/SECDB
1 rows affected.
```

Out[19]:

obj_description
TAG table: The TAG data set contains all standard taxonomy tags, not just those appearing in submissions to date, and also includes all custom taxonomy tags defined in the submissions

Retrieve column comments using information_schema.columns and pg_class system views

In [20]:
```
%%sql
SELECT cols.column_name,
pg_catalog.col_description(c.oid, cols.ordinal_position::int)
FROM pg_catalog.pg_class c, information_schema.columns cols
 WHERE cols.table_name = c.relname  and upper(relname) = 'TAG';
```

```
 * postgresql://student@/SECDB
9 rows affected.
```

Out[20]:

column_name	col_description
tag	tag: The unique identifier (name) for a tag in a specific taxonomy release. tag is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size in length is 256. The value is NOT NULL because all the columns contain valid entries
version	version: Identifier for the taxonomy version; if a standard tag, it indicates the taxonomy of origin. version is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size in length is 20. The value is NOT NULL because all the columns contain valid entries
custom	custom: Indicates if the tag is custom (TRUE) or standard (FALSE). The custom field is classified as boolean because it indicates whether a tag is custom (1 for true) or standard (0 for false). The max-size in length is 1. The value is NOT NULL because all the columns contain valid entries
abstract	abstract: Indicates if the tag is abstract (TRUE) and not used to represent a numeric fact. The abstract field is classified as boolean because it signifies whether the tag is not used to represent a numeric fact, with 1 indicating true (it is abstract) and 0 indicating false (it is not abstract). The max-size in length is 1. The value is NOT NULL because all the columns contain valid entries
datatype	datatype: Data type of the tag (e.g., monetary); NULL if the tag is abstract. datatype is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size in length is 20. The value is NULL because if abstract=1, then NULL
iord	iord: Indicates if the value is a point in time ("I") or a duration ("D"); NULL if the tag is abstract. iord is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size in length is 1. The value is NULL because if abstract=1, then NULL
crdr	crdr: The natural accounting balance of the tag ("C" for credit or "D" for debit); NULL if not defined. crdr is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size in length is 1. The value is NULL because if it's not defined then NULL
tlabel	tlabel: The label text provided by the taxonomy for standard tags, or the text provided by the filer for custom tags. tlabel is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size in length is 512. The value is NULL because if a tag which had neither would have been a NULL value in the colunmn
doc	doc: Detailed definition for the tag, truncated to 2048 characters; may be from the taxonomy or provided by the filer. doc is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size in length is 2048. The value is NULL because some some tags have neither in which case the fielde is NULL in the column

Creating NUM Table

Use DROP TABLE command before creating the table NUM.

Use CREATE TABLE command to create table NUM that contains each entity and specifies the data type of each entity (and whether or not the column can take null values).

The **primary key** of NUM consists of a combination of ADSH, TAG, VERSION, DDATE, QTRS, UOM, DIMH, and IPRX which uniquely identifies each row in NUM.

NUM contains **foreign keys** ADSH, TAG, VERSION, and DIMHASH. ADSH on it's own references rows in SUB, the combination of TAG and VERSION references rows in the

TAG table, and DIMH references the attribute DIMHASH in the DIM table.

In [21]:
```
%%sql
DROP TABLE IF EXISTS NUM Cascade;

CREATE TABLE NUM (
  ADSH    VARCHAR(20) NOT NULL,
  TAG     VARCHAR(256) NOT NULL,
  VERSION VARCHAR(20) NOT NULL,
  DDATE   DATE NOT NULL,
  QTRS    NUMERIC(8,0) NOT NULL,
  UOM     VARCHAR(20) NOT NULL,
  DIMH    VARCHAR(32) NOT NULL,
  IPRX    NUMERIC(2,0) NOT NULL,
  VALUE   NUMERIC(16,0) NULL,
  FOOTNOTE VARCHAR (512) NULL,
  FOOTLEN NUMERIC(4,0) NOT NULL,
  DIMN    NUMERIC(1,0) NOT NULL,
  COREG   VARCHAR(256) NULL,
  DURP    NUMERIC NOT NULL,
  DATP    NUMERIC(4) NOT NULL,
  DCML    NUMERIC(2,0) NOT NULL,

  PRIMARY KEY (ADSH, TAG, VERSION, DDATE, QTRS, UOM, DIMH, IPRX),

  FOREIGN KEY (ADSH) REFERENCES SUB(ADSH),
  FOREIGN KEY (TAG, VERSION) REFERENCES TAG(TAG, VERSION),
  FOREIGN KEY (DIMH) REFERENCES DIM(DIMHASH)

);
```

```
 * postgresql://student@/SECDB
Done.
Done.
```

Out[21]: []

Adding comment to table NUM

In [22]:
```
%%sql
COMMENT ON TABLE NUM IS
'NUM table: The NUM data set contains numeric data, one row per data point i
```

```
 * postgresql://student@/SECDB
Done.
```

Out[22]: []

Adding comments to each column of NUM

In [23]:
```
%%sql
COMMENT ON COLUMN NUM.adsh IS 'adsh: is an accession number, a 20-character
COMMENT ON COLUMN NUM.tag IS 'tag: The unique identifier (name) for a tag ir
COMMENT ON COLUMN NUM.version IS 'version: For a standard tag, it''s an ider
COMMENT ON COLUMN NUM.ddate IS 'ddate: The end date for the data value, rour
COMMENT ON COLUMN NUM.qtrs IS 'qtrs: The count of the number of quarters rep
COMMENT ON COLUMN NUM.uom IS 'uom: The unit of measure for the value (e.g.,
COMMENT ON COLUMN NUM.dimh IS 'dimh: The 32-byte hexadecimal key for the dim
COMMENT ON COLUMN NUM.iprx IS 'iprx: A positive integer to distinguish diffe
COMMENT ON COLUMN NUM.value IS 'value: The actual numeric value reported in
COMMENT ON COLUMN NUM.footnote IS 'footnote: The plain text of any superscri
```

```
COMMENT ON COLUMN NUM.footlen IS 'footlen: Number of bytes in the plain text
COMMENT ON COLUMN NUM.dimn IS 'dimn: Small integer representing the number o
COMMENT ON COLUMN NUM.coreg IS 'coreg: If specified, indicates a specific co
COMMENT ON COLUMN NUM.durp IS 'durp: The difference between the reported fac
COMMENT ON COLUMN NUM.datp IS 'datp: The difference between the reported fac
COMMENT ON COLUMN NUM.dcml IS 'dcml: The value of the fact "decimals" attrib
```

```
 * postgresql://student@/SECDB
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
```

Out[23]: []

Retrieve table comment using pg_class

In [24]:
```
%%sql
SELECT obj_description(oid, 'pg_class')
FROM pg_class
WHERE upper(relname) = 'NUM';
```

```
 * postgresql://student@/SECDB
1 rows affected.
```

Out[24]:

obj_description
NUM table: The NUM data set contains numeric data, one row per data point in the financial statements

Retrieve column comments using information_schema.columns and pg_class system views

In [25]:
```
%%sql
SELECT cols.column_name,
pg_catalog.col_description(c.oid, cols.ordinal_position::int)
FROM pg_catalog.pg_class c, information_schema.columns cols
 WHERE cols.table_name = c.relname  and upper(relname) = 'NUM';
```

```
 * postgresql://student@/SECDB
16 rows affected.
```

Out[25]:

column_name	col_description
adsh	adsh: is an accession number, a 20-character string formed from the 18-digit number assigned by the Commission to each EDGAR submission. Where the 20, is the max-size in length that the column can have and the value is NOT NULL because all the columns contain valid entries. adsh is classified as CHAR because the value is in fixed length
tag	tag: The unique identifier (name) for a tag in a specific taxonomy release. tag is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size in length is 256. The value is NOT NULL because all the columns contain valid entries
version	version: For a standard tag, it's an identifier for the taxonomy;otherwise the accession number where the tag was defined. version is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size in length is 20. The value is NOT NULL because all the columns contain valid entries
ddate	ddate: The end date for the data value, rounded to the nearest month end (format: YYYY-MM-DD). ddate is classified as DATE because it stores the value as (yymmdd) in the column. The maz-size length is 8. The value is NOT NULL because all the columns contain valid entries
qtrs	qtrs: The count of the number of quarters represented by the data value, rounded to the nearest whole number. qtrs is classified as NUMERIC (8,0) because it makes it explicit that only integers (whole numbers) are stored in this column and precision (8) is the total number of digits the number can have. The max-size length is 8. The value is NOT NULL because all the columns contain valid entries
uom	uom: The unit of measure for the value (e.g., USD, shares). uom is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size in length is 20. The value is NOT NULL because all the columns contain valid entries
dimh	dimh: The 32-byte hexadecimal key for the dimensional information in the DIM data set. dimh is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size in length is 32. The value is NOT NULL because all the columns contain valid entries
iprx	iprx: A positive integer to distinguish different reported facts that otherwise would have the same primary key. iprx is classified as NUMERIC (2,0) because it makes it explicit that only integers (whole numbers) are stored in this column and precision (2) is the total number of digits the number can have. The max-size length is 8. The value is NOT NULL because all the columns contain valid entries
value	value: The actual numeric value reported in the financial statements, rounded to four decimal places. value is classified as NUMERIC (16,0) because it makes it explicit that only integers (whole numbers) are stored in this column and precision (16) is the total number of digits the number can have. The max-size length is 16. The value is NULL because there may be missing data points
footnote	footnote: The plain text of any superscripted footnotes on the value, if any, as shown on the statement page, truncated to 512 characters. footnote is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size in length is 512. The footnote field may be null because there were no superscripted footnotes associated with the value, indicating that no relevant text was provided in the submission
footlen	footlen: Number of bytes in the plain text of the footnote prior to truncation, indicating footnote length. footlen is classified as NUMERIC (4,0) because it makes it explicit that only integers (whole numbers) are stored in this column and precision (4) is the total number of digits the number can have. The max-size length is 4. The value is NOT NULL because all the columns contain valid entries
dimn	dimn: Small integer representing the number of dimensions associated with the reported value. dimn is classified as NUMERIC (1,0) because it makes it explicit that only integers (whole numbers) are stored in this column and precision (1) is

column_name	col_description
	the total number of digits the number can have. The max-size length is 4. The value is NOT NULL because all the columns contain valid entries
coreg	coreg: If specified, indicates a specific co-registrant, the parent company, or other entity; NULL indicates the consolidated entity. coreg is classified as VARCHAR because the table in the Financial Statement specifies its ALPHANUMERIC and there is no fixed length in the column. The max-size in length is 256. The value may be NULL because it was not specified in the submission, indicating that the entity is treated as a consolidated entity without a specific co-registrant or related party identified
durp	durp: The difference between the reported fact duration and the quarter duration, expressed as a fraction of 1. durp is classified as NUMERIC since it is a fraction of 1 and contains several digits after the decimal. Since these digits vary, the parameters for precision and scale are not specified. The max-size in length is 4. The value is NOT NULL because all the columns contain valid entries
datp	datp: The difference between the reported fact date and the month-end rounded date, expressed as a fraction of 1. datp is classified as NUMERIC (4) because it is a fraction of 1 with max-size in length of 4. It is not specified how many digits can be after the decimal, so we are not limitting the range by including a value for scale. The value is NOT NULL because all the columns contain valid entries
dcml	dcml: The value of the fact "decimals" attribute, with INF represented by 32767. dcml is classified as NUMERIC (2,0) because it makes it explicit that only integers (whole numbers) are stored in this column and precision (2) is the total number of digits the number can have. The max-size length is 2. The value is NOT NULL because all the columns contain valid entries

Creating PRE Table

Use DROP TABLE command before creating the table PRE.

Use CREATE TABLE command to create table PRE that contains each entity and specifies the data type of each entity (and whether or not the column can take null values).

The **primary key** of PRE consists of a combination of ADSH, REPORT, AND LINE which uniquely identifies each row in PRE.

NUM contains **foreign keys** ADSH, TAG, and VERSION. ADSH on it's own references rows in SUB, the combination of TAG and VERSION references rows in the TAG table, and the combination of ADSH, TAG, and VERSION refereces rows in the NUM table in accordance with the notes on the data. We do not reference the combination of ADSH, TAG, and VERSION as a foreign key below when building the table for PRE because it is redundant given ADSH and the combination of TAG and PRE are already being used as foreign keys. Additionally, when including ADSH, TAG, and VERSION as a foreign key referencing NUM, we see the error message that this key combination is not unique in the NUM table and therefore cannot be used.

In [26]:
```
%%sql
DROP TABLE IF EXISTS PRE Cascade;

CREATE TABLE PRE (
    ADSH VARCHAR(20) NOT NULL,
    REPORT NUMERIC(6,0) NOT NULL,
    LINE NUMERIC(6,0) NOT NULL,
    STMT VARCHAR(2) NOT NULL,
```

```
    INPTH BOOLEAN NOT NULL,
    TAG VARCHAR(256) NOT NULL,
    VERSION VARCHAR(20) NOT NULL,
    PROLE VARCHAR(50) NOT NULL,
    PLABEL VARCHAR(512) NOT NULL,
    NEGATING BOOLEAN NOT NULL,

    PRIMARY KEY (ADSH, REPORT, LINE),

    FOREIGN KEY (ADSH) REFERENCES SUB(ADSH),
    FOREIGN KEY (TAG, VERSION) REFERENCES TAG(TAG, VERSION)

);
```

```
 * postgresql://student@/SECDB
Done.
Done.
```

Out[26]: []

In [27]:
```
%%sql
COMMENT ON TABLE PRE IS
'PRE table: Presentation data set; this provides information about how the t
```

```
 * postgresql://student@/SECDB
Done.
```

Out[27]: []

In [28]:
```
%%sql
COMMENT ON COLUMN PRE.adsh IS 'adsh: is an accession number, a 20-character
COMMENT ON COLUMN PRE.report IS 'report: represents the report grouping, whi
COMMENT ON COLUMN PRE.line IS 'line: represents the tags presentation line o
COMMENT ON COLUMN PRE.stmt IS ' stmt: the financial statement location to wh
COMMENT ON COLUMN PRE.inpth IS'inpth: 1 indicates whether the value was pres
COMMENT ON COLUMN PRE.tag IS 'tag: the tag chosen by the filer for this line
COMMENT ON COLUMN PRE.version IS 'version: the taxonomy identifier if the ta
COMMENT ON COLUMN PRE.prole IS 'prole: the XBRL link "role" of the preferred
COMMENT ON COLUMN PRE.plabel IS 'plabel: the text presented on the line item
COMMENT ON COLUMN PRE.negating IS 'negating: flag to indicate whether the pr
```

```
 * postgresql://student@/SECDB
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
Done.
```

Out[28]: []

Retrieve table comment using pg_class

In [29]:
```
%%sql
SELECT obj_description(oid, 'pg_class')
FROM pg_class
WHERE upper(relname) = 'PRE';
```

```
 * postgresql://student@/SECDB
1 rows affected.
```

Out[29]:

obj_description
PRE table: Presentation data set; this provides information about how the tags and numbers were presented in the primary financial statements

Retrieve column comments using information_schema.columns and pg_class system views

In [30]:

```
%%sql
SELECT cols.column_name,
pg_catalog.col_description(c.oid, cols.ordinal_position::int)
FROM pg_catalog.pg_class c, information_schema.columns cols
 WHERE cols.table_name = c.relname  and upper(relname) = 'PRE';
```

```
 * postgresql://student@/SECDB
10 rows affected.
```

Out[30]:

column_name	col_description
adsh	adsh: is an accession number, a 20-character string formed from the 18-digit number assigned by the Commission to each EDGAR submission. Where the 20, is the max-size in length that the column can have and the value is NOT NULL because all the columns contain valid entries. adsh is classified as VARCHAR because the value is not fixed length
report	report: represents the report grouping, which is a numeric value referring to the "R file" as computed by the renderer. Where we have classified report as a numeric (6,0) because it contains whole numbers with max-length of 6 and no digits after the decimal. It is NOT NULL because all the columns contain valid entries
line	line: represents the tags presentation line order for a given report, allowing for the correct ordering of line items. Line is classified as anumeric (6,0) because it contains whole numbers with max-length of 6 and no digits after the decimal. It is NOT NULL because all the columns contain valid entries
stmt	stmt: the financial statement location to which the value of the "report" field pertains (e.g., CP for Cover Page, BS for Balance Sheet, IS = Income Statement, CF = Cash Flow, EQ = Equity, UN = Unclassifiable Statement). stmt is classified as a VARCHAR because the values are in fixed length. As seen in the table, the abbreviated words are shown in two letters only, "CP","BS".. hence why the max-size in length is also 2. The value is NOT NULL because all the columns contain valid entries
inpth	inpth: 1 indicates whether the value was presented "parenthetically" instead of in fields within the financial statements. inpth is classified as a BOOLEAN field type because it represents a binary condition—either the value is presented parenthetically (1 or TRUE), or it is not (0 or FALSE). The value is NOT NULL because all the columns contain valid entries
tag	tag: the tag chosen by the filer for this line item, which corresponds to a specific taxonomy tag used in XBRL. tag is classified as VARCHAR because it is not fixed in length but its max-size in length is 256. The value is NOT NULL because all the columns contain valid entries
version	version: the taxonomy identifier if the tag is a standard tag; otherwise, it is the accession number where the tag was defined. version is classified as VARCHAR because it is not fixed in length but its max-size in length is 20. The value is NOT NULL because all the columns contain valid entries
prole	prole: the XBRL link "role" of the preferred label, indicating the role of the tag in the financial statement. prole is classified as VARCHAR because it is not fixed in legth but its max-size in length is 50. The value is NOT NULL because all the columns contain valid entries
plabel	plabel: the text presented on the line item, also known as a "preferred" label, which describes the financial data. plabel is classified as VARCHAR because it is not fixed in legth but its max-size in length is 512. The value is NOT NULL because all the columns contain valid entries
negating	negating: flag to indicate whether the prole is treated as negating by the renderer, affecting how the value is interpreted. This is classified as a BOOLEAN because it represents a binary state—either the profile is treated as negating by the renderer (TRUE or 1), or it is not (FALSE or 0). The value is NOT NULL because all the columns contain valid entries

Uploading Financial Data

Unzipping data for each month

Each of the monthly financial data files were downloaded and the zip files were uploaded to the directory '/home/ubuntu/notebooks'.

In [31]:
```
!pwd
```

```
/home/ubuntu/notebooks
```

The code below is run to unzip each of monthly files.

In [33]:
```
!unzip 2024_01_notes.zip -d Jan2024
```

```
Archive:  2024_01_notes.zip
  inflating: Jan2024/sub.tsv
  inflating: Jan2024/tag.tsv
  inflating: Jan2024/dim.tsv
  inflating: Jan2024/ren.tsv
  inflating: Jan2024/cal.tsv
  inflating: Jan2024/pre.tsv
  inflating: Jan2024/num.tsv
  inflating: Jan2024/txt.tsv
  inflating: Jan2024/readme.htm
  inflating: Jan2024/notes-metadata.json
```

In [34]:
```
!unzip 2024_02_notes.zip -d Feb2024
```

```
Archive:  2024_02_notes.zip
  inflating: Feb2024/sub.tsv
  inflating: Feb2024/tag.tsv
  inflating: Feb2024/dim.tsv
  inflating: Feb2024/ren.tsv
  inflating: Feb2024/cal.tsv
  inflating: Feb2024/pre.tsv
  inflating: Feb2024/num.tsv
  inflating: Feb2024/txt.tsv
  inflating: Feb2024/readme.htm
  inflating: Feb2024/notes-metadata.json
```

In [35]:
```
!unzip 2024_03_notes.zip -d Mar2024
```

```
Archive:  2024_03_notes.zip
  inflating: Mar2024/sub.tsv
  inflating: Mar2024/tag.tsv
  inflating: Mar2024/dim.tsv
  inflating: Mar2024/ren.tsv
  inflating: Mar2024/cal.tsv
  inflating: Mar2024/pre.tsv
  inflating: Mar2024/num.tsv
  inflating: Mar2024/txt.tsv
  inflating: Mar2024/readme.htm
  inflating: Mar2024/notes-metadata.json
```

In [38]:
```
!unzip 2024_04_notes.zip -d Apr2024
```

```
Archive:  2024_04_notes.zip
  inflating: Apr2024/sub.tsv
  inflating: Apr2024/tag.tsv
  inflating: Apr2024/dim.tsv
  inflating: Apr2024/ren.tsv
  inflating: Apr2024/cal.tsv
  inflating: Apr2024/pre.tsv
  inflating: Apr2024/num.tsv
  inflating: Apr2024/txt.tsv
  inflating: Apr2024/readme.htm
  inflating: Apr2024/notes-metadata.json
```

Converting all tsv files into csv files for all entities

In [39]:
```
!csvformat -t ./Jan2024/sub.tsv > jan_sub.csv
!csvformat -t ./Feb2024/sub.tsv > feb_sub.csv
!csvformat -t ./Mar2024/sub.tsv > mar_sub.csv
!csvformat -t ./Apr2024/sub.tsv > apr_sub.csv
```

In [40]:
```
!csvformat -t ./Jan2024/tag.tsv > jan_tag.csv
!csvformat -t ./Feb2024/tag.tsv > feb_tag.csv
!csvformat -t ./Mar2024/tag.tsv > mar_tag.csv
!csvformat -t ./Apr2024/tag.tsv > apr_tag.csv
```

In [41]:
```
!csvformat -t ./Jan2024/dim.tsv > jan_dim.csv
!csvformat -t ./Feb2024/dim.tsv > feb_dim.csv
!csvformat -t ./Mar2024/dim.tsv > mar_dim.csv
!csvformat -t ./Apr2024/dim.tsv > apr_dim.csv
```

In [42]:
```
!csvformat -t ./Jan2024/pre.tsv > jan_pre.csv
!csvformat -t ./Feb2024/pre.tsv > feb_pre.csv
!csvformat -t ./Mar2024/pre.tsv > mar_pre.csv
!csvformat -t ./Apr2024/pre.tsv > apr_pre.csv
```

In [43]:
```
!csvformat -t ./Jan2024/num.tsv > jan_num.csv
!csvformat -t ./Feb2024/num.tsv > feb_num.csv
!csvformat -t ./Mar2024/num.tsv > mar_num.csv
!csvformat -t ./Apr2024/num.tsv > apr_num.csv
```

Stacking files for each month into one csv file for each entity

!csvstack jan_sub.csv feb_sub.csv mar_sub.csv apr_sub.csv > final_sub.csv

!csvstack jan_tag.csv feb_tag.csv mar_tag.csv apr_tag.csv > final_tag.csv

!csvstack jan_dim.csv feb_dim.csv mar_dim.csv apr_dim.csv > final_dim.csv

!csvstack jan_pre.csv feb_pre.csv mar_pre.csv apr_pre.csv > final_pre.csv

!csvstack jan_num.csv feb_num.csv mar_num.csv apr_num.csv > final_num.csv

IMPORTANT

We worked on the data for all four months for each entity. In loading in the data to the tables we were able to do it for all four months up until NUM but since NUM data is very large we could execute the last code on NUM on our instances even after multiple attempts. Hence, we decided to work with two months data. Again, we faced issues with loading it in for NUM and had to come down to working on just one month Jan. We leave some of the code as markdowns to show our attempt to work on full data.

Stacking CSV for two months (JAN and FEB 2024)

!csvstack jan_sub.csv feb_sub.csv > final_sub.csv

!csvstack jan_tag.csv feb_tag.csv > final_tag.csv

!csvstack jan_dim.csv feb_dim.csv > final_dim.csv

!csvstack jan_pre.csv feb_pre.csv > final_pre.csv

!csvstack jan_num.csv feb_num.csv > final_num.csv

Loading data into the tables created for each attribute

Loading SUB dataset into the SUB table

In [43]:
```
!head -n 1000 jan_sub.csv | csvstat
```

```
/home/ubuntu/.local/lib/python3.8/site-packages/agate/table/from_csv.py:74:
RuntimeWarning: Error sniffing CSV dialect: Could not determine delimiter
```

1. "adsh"

Type of data:	Text
Contains null values:	False
Unique values:	999
Longest value:	20 characters
Most common values:	0000081061-24-000003 (1x) 0000090168-23-000081 (1x) 0000090168-23-000083 (1x) 0000093410-24-000002 (1x) 0000107140-24-000002 (1x)

2. "cik"

Type of data:	Number
Contains null values:	False
Unique values:	864
Smallest value:	2186
Largest value:	2006191
Sum:	1360083967
Mean:	1361445.412
Median:	1498710
StDev:	506268.034
Most common values:	216877 (6x) 1325878 (5x) 1029744 (4x) 1564708 (4x) 1832168 (3x)

3. "name"

Type of data:	Text
Contains null values:	False
Unique values:	864
Longest value:	57 characters
Most common values:	PISMO COAST VILLAGE INC (6x) FEDERAL HOME LOAN BANK OF TOPEKA (5x) SONIC FOUNDRY INC (4x) NEWS CORP (4x) LONGBOARD PHARMACEUTICALS, INC. (3x)

4. "sic"

Type of data:	Number
Contains null values:	True (excluded from calculations)
Unique values:	206
Smallest value:	100
Largest value:	8742
Sum:	4429547
Mean:	4667.594
Median:	3845
StDev:	1990.612
Most common values:	2834 (140x) None (50x) 2836 (47x) 6770 (42x) 6798 (41x)

5. "countryba"

```
      Type of data:          Text
      Contains null values:  True (excluded from calculations)
      Unique values:         20
      Longest value:         2 characters
      Most common values:    US (933x)
                             CA (14x)
                             GB (8x)
                             CN (8x)
                             KY (7x)
```

6. "stprba"

```
      Type of data:          Text
      Contains null values:  True (excluded from calculations)
      Unique values:         54
      Longest value:         2 characters
      Most common values:    CA (148x)
                             NY (112x)
                             MA (99x)
                             TX (75x)
                             FL (68x)
```

7. "cityba"

```
      Type of data:          Text
      Contains null values:  True (excluded from calculations)
      Unique values:         405
      Longest value:         30 characters
      Most common values:    NEW YORK (85x)
                             BOSTON (36x)
                             HOUSTON (26x)
                             CHICAGO (21x)
                             AUSTIN (15x)
```

8. "zipba"

```
      Type of data:          Text
      Contains null values:  True (excluded from calculations)
      Unique values:         630
      Longest value:         10 characters
      Most common values:    10036 (14x)
                             02116 (13x)
                             10022 (11x)
                             10019 (10x)
                             02110 (9x)
```

9. "bas1"

```
      Type of data:          Text
      Contains null values:  True (excluded from calculations)
      Unique values:         838
      Longest value:         40 characters
      Most common values:    165 S DOLLIVER ST (6x)
                             200 CLARENDON STREET (5x)
                             500 SW WANAMAKER ROAD (5x)
                             222 W. WASHINGTON AVENUE (4x)
                             1211 AVENUE OF THE AMERICAS (4x)
```

10. "bas2"

```
       Type of data:          Text
       Contains null values:  True (excluded from calculations)
       Unique values:         256
       Longest value:         40 characters
       Most common values:    None (546x)
                              SUITE 300 (20x)
                              SUITE 200 (19x)
                              SUITE 100 (15x)
                              SUITE 400 (12x)
```

11. "baph"

```
       Type of data:          Text
       Contains null values:  True (excluded from calculations)
       Unique values:         849
       Longest value:         19 characters
       Most common values:    8057735649 (6x)
                              785 233 0507 (5x)
                              6084431600 (4x)
                              212-416-3400 (4x)
                              619-592-9775 (3x)
```

12. "countryma"

```
       Type of data:          Text
       Contains null values:  True (excluded from calculations)
       Unique values:         19
       Longest value:         2 characters
       Most common values:    US (935x)
                              CA (15x)
                              GB (8x)
                              CN (8x)
                              KY (7x)
```

13. "stprma"

```
       Type of data:          Text
       Contains null values:  True (excluded from calculations)
       Unique values:         53
       Longest value:         2 characters
       Most common values:    CA (150x)
                              NY (114x)
                              MA (97x)
                              TX (75x)
                              FL (69x)
```

14. "cityma"

```
       Type of data:          Text
       Contains null values:  True (excluded from calculations)
       Unique values:         407
       Longest value:         30 characters
       Most common values:    NEW YORK (86x)
                              BOSTON (33x)
                              HOUSTON (26x)
                              CHICAGO (21x)
                              AUSTIN (15x)
```

15. "zipma"

```
       Type of data:          Text
```

```
       Contains null values:  True (excluded from calculations)
       Unique values:         630
       Longest value:         10 characters
       Most common values:    10036 (14x)
                              10022 (11x)
                              02116 (10x)
                              10019 (9x)
                              02110 (9x)
```

16. "mas1"

```
       Type of data:          Text
       Contains null values:  True (excluded from calculations)
       Unique values:         839
       Longest value:         40 characters
       Most common values:    165 S DOLLIVER ST (6x)
                              500 SW WANAMAKER ROAD (5x)
                              222 W. WASHINGTON AVENUE (4x)
                              1211 AVENUE OF THE AMERICAS (4x)
                              590 MADISON AVENUE (4x)
```

17. "mas2"

```
       Type of data:          Text
       Contains null values:  True (excluded from calculations)
       Unique values:         260
       Longest value:         40 characters
       Most common values:    None (541x)
                              SUITE 200 (20x)
                              SUITE 300 (20x)
                              SUITE 100 (15x)
                              SUITE 400 (12x)
```

18. "countryinc"

```
       Type of data:          Text
       Contains null values:  True (excluded from calculations)
       Unique values:         13
       Longest value:         2 characters
       Most common values:    US (811x)
                              None (102x)
                              KY (42x)
                              CA (20x)
                              GB (7x)
```

19. "stprinc"

```
       Type of data:          Text
       Contains null values:  True (excluded from calculations)
       Unique values:         43
       Longest value:         2 characters
       Most common values:    DE (514x)
                              None (187x)
                              NV (84x)
                              MD (56x)
                              FL (19x)
```

20. "ein"

```
       Type of data:          Number
       Contains null values:  False
```

```
       Unique values:         662
       Smallest value:        0
       Largest value:         990375676
       Sum:                   446053117934
       Mean:                  446499617.552
       Median:                452552528
       StDev:                 348901443.161
       Most common values:    0 (226x)
                              952990441 (6x)
                              480561319 (5x)
                              391783372 (4x)
                              845009619 (3x)
```

21. "former"

```
       Type of data:          Text
       Contains null values:  True (excluded from calculations)
       Unique values:         520
       Longest value:         60 characters
       Most common values:    None (407x)
                              NEWS CORP (4x)
                              ADITX THERAPEUTICS, INC. (3x)
                              HOMELAND RESOURCES LTD. (3x)
                              SPHERE 3D CORP (3x)
```

22. "changed"

```
       Type of data:          Number
       Contains null values:  True (excluded from calculations)
       Unique values:         472
       Smallest value:        19690520
       Largest value:         20230928
       Sum:                   11896422208
       Mean:                  20095307.784
       Median:                20120858.5
       StDev:                 108831.456
       Most common values:    None (407x)
                              19920703 (29x)
                              19930328 (5x)
                              20210331 (4x)
                              20130628 (4x)
```

23. "afs"

```
       Type of data:          Text
       Contains null values:  True (excluded from calculations)
       Unique values:         4
       Longest value:         5 characters
       Most common values:    4-NON (587x)
                              1-LAF (274x)
                              2-ACC (109x)
                              None (29x)
```

24. "wksi"

```
       Type of data:          Boolean
       Contains null values:  False
       Unique values:         2
       Most common values:    False (998x)
                              True (1x)
```

25. "fye"

Type of data:	Number
Contains null values:	True (excluded from calculations)
Unique values:	14
Smallest value:	131
Largest value:	1231
Sum:	1095169
Mean:	1109.594
Median:	1231
StDev:	263.083
Most common values:	1231 (775x)
	930 (56x)
	630 (37x)
	331 (32x)
	831 (23x)

26. "form"

Type of data:	Text
Contains null values:	False
Unique values:	23
Longest value:	7 characters
Most common values:	8-K (888x)
	8-K/A (24x)
	10-Q (24x)
	DEF 14A (10x)
	N-CSR (10x)

27. "period"

Type of data:	Number
Contains null values:	True (excluded from calculations)
Unique values:	12
Smallest value:	20051231
Largest value:	20240131
Sum:	20049020159
Mean:	20231100.06
Median:	20231231
StDev:	5791.136
Most common values:	20231231 (890x)
	20231130 (31x)
	20230930 (27x)
	20231031 (26x)
	20240131 (9x)

28. "fy"

Type of data:	Number
Contains null values:	True (excluded from calculations)
Unique values:	4
Smallest value:	2022
Largest value:	2024
Sum:	84988
Mean:	2023.524
Median:	2024
StDev:	0.552
Most common values:	None (957x)
	2024 (23x)
	2023 (18x)
	2022 (1x)

29. "fp"

Type of data:	Text
Contains null values:	True (excluded from calculations)
Unique values:	5
Longest value:	2 characters
Most common values:	None (957x) FY (11x) Q2 (11x) Q3 (11x) Q1 (9x)

30. "filed"

Type of data:	Number
Contains null values:	False
Unique values:	4
Smallest value:	20240102
Largest value:	20240105
Sum:	20219863487
Mean:	20240103.591
Median:	20240104
StDev:	1.124
Most common values:	20240105 (280x) 20240104 (260x) 20240102 (230x) 20240103 (229x)

31. "accepted"

Type of data:	DateTime
Contains null values:	False
Unique values:	631
Smallest value:	2023-12-29 17:44:00
Largest value:	2024-01-05 17:30:00
Most common values:	2024-01-05 16:06:00 (11x) 2024-01-05 16:30:00 (10x) 2024-01-05 17:01:00 (10x) 2024-01-03 16:31:00 (8x) 2024-01-04 16:16:00 (8x)

32. "prevrpt"

Type of data:	Boolean
Contains null values:	False
Unique values:	2
Most common values:	False (940x) True (59x)

33. "detail"

Type of data:	Boolean
Contains null values:	False
Unique values:	2
Most common values:	False (948x) True (51x)

34. "instance"

Type of data:	Text

```
       Contains null values:  False
       Unique values:         881
       Longest value:         32 characters
       Most common values:    form8-k_htm.xml (89x)
                              form8-ka_htm.xml (7x)
                              form8k_htm.xml (4x)
                              f8k_010424_htm.xml (4x)
                              pcv-20231213_htm.xml (3x)
```

35. "nciks"

```
       Type of data:          Number
       Contains null values:  False
       Unique values:         2
       Smallest value:        1
       Largest value:         2
       Sum:                   1012
       Mean:                  1.013
       Median:                1
       StDev:                 0.113
       Most common values:    1 (986x)
                              2 (13x)
```

36. "aciks"

```
       Type of data:          Number
       Contains null values:  True (excluded from calculations)
       Unique values:         14
       Smallest value:        57183
       Largest value:         1959472
       Sum:                   12379906
       Mean:                  952300.462
       Median:                1045610
       StDev:                 638299.3
       Most common values:    None (986x)
                              1004036 (1x)
                              1959472 (1x)
                              81033 (1x)
                              57183 (1x)
```

37. "pubfloatusd"

```
       Type of data:          Number
       Contains null values:  True (excluded from calculations)
       Unique values:         11
       Smallest value:        0
       Largest value:         144946000
       Sum:                   280152338
       Mean:                  28015233.8
       Median:                4462000.5
       StDev:                 50475574.688
       Most common values:    None (989x)
                              13037758 (1x)
                              0 (1x)
                              96011 (1x)
                              6424001 (1x)
```

38. "floatdate"

```
       Type of data:          Number
       Contains null values:  True (excluded from calculations)
```

```
        Unique values:          6
        Smallest value:         20220630
        Largest value:          20231231
        Sum:                    202294408
        Mean:                   20229440.8
        Median:                 20230331
        StDev:                  3109.924
        Most common values:     None (989x)
                                20230331 (6x)
                                20231231 (1x)
                                20220630 (1x)
                                20230131 (1x)

 39. "floataxis"

        Type of data:           Boolean
        Contains null values:   True (excluded from calculations)
        Unique values:          1
        Most common values:     None (999x)

 40. "floatmems"

        Type of data:           Boolean
        Contains null values:   True (excluded from calculations)
        Unique values:          2
        Most common values:     None (989x)
                                True (10x)

Row count: 999
```

In [32]:
```
%%sql
COPY SUB FROM '/home/ubuntu/notebooks/jan_sub.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
(psycopg2.errors.NotNullViolation) null value in column "fy" violates not-n
ull constraint
DETAIL:  Failing row contains (0000081061-24-000003, 81061, PUBLIX SUPER MA
RKETS INC, 5411, US, FL, LAKELAND, 33811, 3300 PUBLIX CORPORATE PARKWAY, nu
ll, 863-688-1188, US, FL, LAKELAND, 33811, 3300 PUBLIX CORPORATE PARKWAY, n
ull, US, FL, 590324412, null, null, 4-NON, f, 1231, 8-K, 2023-12-31, null,
null, 2024-01-02, 2024-01-02 14:19:00, f, f, ck0000081061-20240101_htm.xml,
1, null, null, null, null, null).
CONTEXT:  COPY sub, line 2: "0000081061-24-000003,81061,PUBLIX SUPER MARKET
S INC,5411,US,FL,LAKELAND,33811,3300 PUBLIX CORPORATE ..."

[SQL: COPY SUB FROM '/home/ubuntu/notebooks/jan_sub.csv'
CSV
HEADER;]
(Background on this error at: https://sqlalche.me/e/14/gkpj)
```

Explain error:

The error shows up for the "fy" or "Fiscal Year Focus" column. We looked through the dataset and found that there are more than one missing values in the "fy" column. When loading the data on the database SUB table it looks like it is considering the missing cells as NULL when trying to upload the data. This violates the constraint of "Not NULL" on the "fy" attribute and hence returns the error.

Next Step:

We would try to update the constraint from "NOT NULL" to "NULL" and upload the data. The reason we take that approach is because when studying the data we found the "fy" field gives the fiscal year in focus. There already exists fields for fiscal year end date "fye" which gives the months and the balance sheet date "period" which gives the year month and date. It seems like making the field constraint null may not effect our analysis for the dataset. These values cannot be dropped since there are more missing fields in this column than filled in fields.

Updating the constraint on the attribute "fy" in SUB

In [33]:
```
%%sql

ALTER TABLE sub
ALTER COLUMN fy DROP NOT NULL;
```

```
 * postgresql://student@/SECDB
Done.
```

Out[33]:
```
[]
```

Load data again

In [34]:
```
%%sql
COPY SUB FROM '/home/ubuntu/notebooks/jan_sub.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
(psycopg2.errors.NotNullViolation) null value in column "fp" violates not-n
ull constraint
DETAIL:  Failing row contains (0000081061-24-000003, 81061, PUBLIX SUPER MA
RKETS INC, 5411, US, FL, LAKELAND, 33811, 3300 PUBLIX CORPORATE PARKWAY, nu
ll, 863-688-1188, US, FL, LAKELAND, 33811, 3300 PUBLIX CORPORATE PARKWAY, n
ull, US, FL, 590324412, null, null, 4-NON, f, 1231, 8-K, 2023-12-31, null,
null, 2024-01-02, 2024-01-02 14:19:00, f, f, ck0000081061-20240101_htm.xml,
1, null, null, null, null, null).
CONTEXT:  COPY sub, line 2: "0000081061-24-000003,81061,PUBLIX SUPER MARKET
S INC,5411,US,FL,LAKELAND,33811,3300 PUBLIX CORPORATE ..."

[SQL: COPY SUB FROM '/home/ubuntu/notebooks/jan_sub.csv'
CSV
HEADER;]
(Background on this error at: https://sqlalche.me/e/14/gkpj)
```

Error:

The "fp" column returns a similar error. We looked through the dataset and found most of the cells in this column of the dataset have missing values. The data is related to financial period in focus within fiscal year.

Updating the constraint on the attribute "fp" in SUB

In [35]:
```
%%sql
```

```
ALTER TABLE sub
ALTER COLUMN fp DROP NOT NULL;
```

```
 * postgresql://student@/SECDB
Done.
```

Out[35]: []

Try to load data again

In [36]:
```
%%sql
COPY SUB FROM '/home/ubuntu/notebooks/jan_sub.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
(psycopg2.errors.NotNullViolation) null value in column "countryinc" violates not-null constraint
DETAIL:  Failing row contains (0000898173-24-000003, 898173, O REILLY AUTOMOTIVE INC, 5531, US, MO, SPRINGFIELD, 65802, 233 S PATTERSON AVE, null, 417-829-5878, US, MO, SPRINGFIELD, 65802, 233 S PATTERSON AVE, null, null, null, 274358837, null, null, 1-LAF, f, 1231, 8-K, 2023-12-31, null, null, 2024-01-02, 2024-01-02 16:32:00, f, f, orly-20240102x8k_htm.xml, 1, null, null, null, null, null).
CONTEXT:  COPY sub, line 11: "0000898173-24-000003,898173,O REILLY AUTOMOTIVE INC,5531,US,MO,SPRINGFIELD,65802,233 S PATTERSON AVE..."

[SQL: COPY SUB FROM '/home/ubuntu/notebooks/jan_sub.csv'
CSV
HEADER;]
(Background on this error at: https://sqlalche.me/e/14/gkpj)
```

Error:

The error in the countryinc column is because some rows in the column are empty. Since the constraint on the attribute is NOT NULL, while loading the data on the table its returning an error as the database system takes these missing values as NULL.

To avoid this problem we will try to remove the constraint on "countryinc" and try to load the data.

In [37]:
```
%%sql

ALTER TABLE sub
ALTER COLUMN countryinc DROP NOT NULL;
```

```
 * postgresql://student@/SECDB
Done.
```

Out[37]: []

Try loading the data again

In [38]:
```
%%sql
COPY SUB FROM '/home/ubuntu/notebooks/jan_sub.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
(psycopg2.errors.NotNullViolation) null value in column "fye" violates not-
null constraint
DETAIL:  Failing row contains (0000950170-24-000306, 1484612, OUTSET MEDICA
L, INC., 3845, US, CA, SAN JOSE, 95134, 3052 ORCHARD DRIVE, null, 669-231-8
200, US, CA, SAN JOSE, 95134, 3052 ORCHARD DRIVE, null, US, DE, 200514392,
HOME DIALYSIS PLUS, LTD., 20100219, 1-LAF, f, null, 8-K, 2023-12-31, null,
null, 2024-01-02, 2024-01-02 16:41:00, f, f, om-20240102_htm.xml, 1, null,
null, null, null, null).
CONTEXT:  COPY sub, line 31: "0000950170-24-000306,1484612,"OUTSET MEDICAL,
INC.",3845,US,CA,SAN JOSE,95134,3052 ORCHARD DRIVE,,66..."

[SQL: COPY SUB FROM '/home/ubuntu/notebooks/jan_sub.csv'
CSV
HEADER;]
(Background on this error at: https://sqlalche.me/e/14/gkpj)
```

Error:

The error in the "fye" (Fiscal Year End Date) column is because the constraint is set to NOT NULL and there are a few missing rows in the column. While loading the data on the table, the database system treats the missing values as NULL and hence returns an error.

We will try to fix this by dropping the constraint on this column.

In [39]:
```
%%sql

ALTER TABLE sub
ALTER COLUMN fye DROP NOT NULL;
```

```
 * postgresql://student@/SECDB
Done.
```

Out[39]: []

Try loading the data again

In [40]:
```
%%sql
COPY SUB FROM '/home/ubuntu/notebooks/jan_sub.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
(psycopg2.errors.NotNullViolation) null value in column "period" violates n
ot-null constraint
DETAIL:  Failing row contains (0001193125-24-000495, 1949594, FIDELITY MULT
I-STRATEGY CREDIT FUND, null, US, MA, BOSTON, 02110, 245 SUMMER STREET, nul
l, 617-563-7000, US, MA, BOSTON, 02110, 245 SUMMER STREET, null, US, DE, 0,
null, null, null, f, 0630, 424B3, null, null, null, 2024-01-02, 2024-01-02
12:23:00, f, f, d59653d424b3_htm.xml, 1, null, null, null, null, null).
CONTEXT:  COPY sub, line 102: "0001193125-24-000495,1949594,FIDELITY MULTI-
STRATEGY CREDIT FUND,,US,MA,BOSTON,02110,245 SUMMER STRE..."

[SQL: COPY SUB FROM '/home/ubuntu/notebooks/jan_sub.csv'
CSV
HEADER;]
(Background on this error at: https://sqlalche.me/e/14/gkpj)
```

Error:

The error in the "period" (Balance Sheet Date) column is because the constraint is set to NOT NULL and there are a few missing rows in the column. While loading the data on the table, the database system treats the missing values as NULL and hence returns an error.

Next step:

We will try to fix this by dropping the constraint on this column. Since there are very few missing fields, we will try to impute these values once the data is successfully loaded.

In [41]:
```
%%sql

ALTER TABLE sub
ALTER COLUMN period DROP NOT NULL;
```

```
 * postgresql://student@/SECDB
Done.
```

Out[41]: []

Try loading the data again

In [42]:
```
%%sql
COPY SUB FROM '/home/ubuntu/notebooks/jan_sub.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
(psycopg2.errors.NotNullViolation) null value in column "countryba" violate
s not-null constraint
DETAIL:  Failing row contains (0001213900-24-001269, 1376231, VPR BRANDS, L
P., 7372, null, null, null, null, null, null, 305-537-6607, US, FL, SUNRIS
E, 33323, 1141 SAWGRASS CORPORATE PARKWAY, null, null, null, 900191208, SOL
EIL CAPITAL L.P., 20100401, 4-NON, f, 1231, 8-K, 2023-12-31, null, null, 20
24-01-05, 2024-01-04 20:34:00, f, f, ea191186-8k_vprbrands_htm.xml, 1, nul
l, null, null, null, null).
CONTEXT:  COPY sub, line 884: "0001213900-24-001269,1376231,"VPR BRANDS, L
P.",7372,,,,,,,305-537-6607,US,FL,SUNRISE,33323,1141 SAWG..."

[SQL: COPY SUB FROM '/home/ubuntu/notebooks/jan_sub.csv'
CSV
HEADER;]
(Background on this error at: https://sqlalche.me/e/14/gkpj)
```

Error:

The error in the "countryba" (country of the registrant's business address) column is because the constraint is set to NOT NULL and there are a few missing rows in the column. While loading the data on the table, the database system treats the missing values as NULL and hence returns an error.

Next Step:

We will try to fix this by dropping the constraint on this column. Since there are very few missing fields, we will try to drop these rows once the data is successfully loaded.

In [43]:
```
%%sql
```

```
ALTER TABLE sub
ALTER COLUMN countryba DROP NOT NULL;
```

```
 * postgresql://student@/SECDB
Done.
```

Out[43]: []

Try loading in the data again

In [44]:
```
%%sql
COPY SUB FROM '/home/ubuntu/notebooks/jan_sub.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
(psycopg2.errors.NotNullViolation) null value in column "cityba" violates n
ot-null constraint
DETAIL:  Failing row contains (0001213900-24-001269, 1376231, VPR BRANDS, L
P., 7372, null, null, null, null, null, null, 305-537-6607, US, FL, SUNRIS
E, 33323, 1141 SAWGRASS CORPORATE PARKWAY, null, null, null, 900191208, SOL
EIL CAPITAL L.P., 20100401, 4-NON, f, 1231, 8-K, 2023-12-31, null, null, 20
24-01-05, 2024-01-04 20:34:00, f, f, ea191186-8k_vprbrands_htm.xml, 1, nul
l, null, null, null, null).
CONTEXT:  COPY sub, line 884: "0001213900-24-001269,1376231,"VPR BRANDS, L
P.",7372,,,,,,,305-537-6607,US,FL,SUNRISE,33323,1141 SAWG..."

[SQL: COPY SUB FROM '/home/ubuntu/notebooks/jan_sub.csv'
CSV
HEADER;]
(Background on this error at: https://sqlalche.me/e/14/gkpj)
```

Error:

The error in the "cityba" (city of the registrant's business address) column is because the constraint is set to NOT NULL and there are a few missing rows in the column. While loading the data on the table, the database system treats the missing values as NULL and hence returns an error.

Next Step:

We will try to fix this by dropping the constraint on this column. Since there are very few missing fields, we will try to drop these rows once the data is successfully loaded.¶

In [45]:
```
%%sql

ALTER TABLE sub
ALTER COLUMN cityba DROP NOT NULL;
```

```
 * postgresql://student@/SECDB
Done.
```

Out[45]: []

Try loading in the data again

In [46]:
```
%%sql
COPY SUB FROM '/home/ubuntu/notebooks/jan_sub.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
(psycopg2.errors.NumericValueOutOfRange) numeric field overflow
DETAIL:  A field with precision 8, scale 0 must round to an absolute value
less than 10^8.
CONTEXT:  COPY sub, line 985, column pubfloatusd: "144946000.00"

[SQL: COPY SUB FROM '/home/ubuntu/notebooks/jan_sub.csv'
CSV
HEADER;]
(Background on this error at: https://sqlalche.me/e/14/9h9h)
```

Error:

The error above is on the "pubfloatusd" column because the column as float values whereas the constraint assigned in the table only alllows integers with upto 8 digits).

Next Step:

The solution is to first increase the max size of the field as the maximum value is upto 13 digits and allow for digits to go upto 2 decimal places. We increase this if an error shows up with a field size exceeding that.

In [47]:
```
%%sql

ALTER TABLE sub
ALTER COLUMN pubfloatusd TYPE NUMERIC(12,2);
```

```
 * postgresql://student@/SECDB
Done.
```

Out[47]: []

Try loading in the data again

In [48]:
```
%%sql
COPY SUB FROM '/home/ubuntu/notebooks/jan_sub.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
(psycopg2.errors.NumericValueOutOfRange) numeric field overflow
DETAIL:  A field with precision 12, scale 2 must round to an absolute value
less than 10^10.
CONTEXT:  COPY sub, line 2742, column pubfloatusd: "165530000000.00"

[SQL: COPY SUB FROM '/home/ubuntu/notebooks/jan_sub.csv'
CSV
HEADER;]
(Background on this error at: https://sqlalche.me/e/14/9h9h)
```

In [49]:
```
%%sql

ALTER TABLE sub
ALTER COLUMN pubfloatusd TYPE NUMERIC(15,2);
```

```
 * postgresql://student@/SECDB
Done.
```

Out[49]: []

Try loading in the data again

In [50]:
```
%%sql
COPY SUB FROM '/home/ubuntu/notebooks/jan_sub.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
5904 rows affected.
```

Out[50]: []

In [51]:
```
!csvstat --count jan_sub.csv
```

```
5904
```

Data successfully loaded:

The rows from the Jan SUB csv file matches the number of rows loaded on the SUB table in the database which confirms that our data was successfully loaded.

Explain error:

It looks the highest value in digits this column takes is upto 20 digits. So we update the constraint to cater to such values.

In [18]:
```
%%sql

ALTER TABLE sub
ALTER COLUMN pubfloatusd TYPE NUMERIC(20,2);
```

```
 * postgresql://student@/SECDB
Done.
```

Out[18]: []

In [20]:
```
%%sql
COPY SUB FROM '/home/ubuntu/notebooks/Assignment2/final_sub.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
(psycopg2.errors.NumericValueOutOfRange) numeric field overflow
DETAIL:  A field with precision 1, scale 0 must round to an absolute value
less than 10^1.
CONTEXT:  COPY sub, line 14169, column floatmems: "16"

[SQL: COPY SUB FROM '/home/ubuntu/notebooks/Assignment2/final_sub.csv'
CSV
HEADER;]
(Background on this error at: https://sqlalche.me/e/14/9h9h)
```

Explain error:

The floatmems column (number of terms in the summation if the public float was computed) returns a constraint violation error as the max size is set to 1 whereas the summation can have more than 1 digit. We will try solving this by increasing the max-size of the field to 2.

In [19]:
```
%%sql
```

```
ALTER TABLE sub
ALTER COLUMN floatmems TYPE NUMERIC(2,0);
```

```
 * postgresql://student@/SECDB
Done.
```

Out[19]: []

In [20]:
```
%%sql
COPY SUB FROM '/home/ubuntu/notebooks/Assignment2/final_sub.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
15304 rows affected.
```

Out[20]: []

In [3]:
```
!csvstat --count final_sub.csv
```

```
15304
```

The rows from the final SUB csv file matches the number of rows loaded on the SUB table in the database which confirms that our data was successfully loaded.

Loading TAG dataset into the TAG table

In [64]:
```
!head -n 1000 jan_tag.csv | csvstat
```

1. "tag"

Type of data:	Text
Contains null values:	False
Unique values:	766
Longest value:	116 characters
Most common values:	AccountsPayableRelatedPartyCurrent (7x)
	AccountingPoliciesLineItems (6x)
	AccountingPoliciesTable (6x)
	A0.875NotesDue2025Member (5x)
	AccountingForWarrantsPolicyTextBlock (5x)

2. "version"

Type of data:	Text
Contains null values:	False
Unique values:	261
Longest value:	20 characters
Most common values:	0001035443-24-000072 (74x)
	us-gaap/2023 (70x)
	0000732712-24-000003 (29x)
	0000732712-24-000007 (29x)
	0001193125-24-013874 (29x)

3. "custom"

Type of data:	Boolean
Contains null values:	False
Unique values:	2
Most common values:	True (900x)
	False (99x)

4. "abstract"

Type of data:	Boolean
Contains null values:	False
Unique values:	2
Most common values:	True (782x)
	False (217x)

5. "datatype"

Type of data:	Text
Contains null values:	True (excluded from calculations)
Unique values:	11
Longest value:	14 characters
Most common values:	member (696x)
	monetary (149x)
	None (86x)
	textBlock (29x)
	percent (15x)

6. "iord"

Type of data:	Text
Contains null values:	True (excluded from calculations)
Unique values:	3
Longest value:	1 characters
Most common values:	None (782x)
	I (120x)
	D (97x)

```
  7. "crdr"

        Type of data:          Text
        Contains null values:  True (excluded from calculations)
        Unique values:         3
        Longest value:         1 characters
        Most common values:    None (850x)
                               C (93x)
                               D (56x)

  8. "tlabel"

        Type of data:          Text
        Contains null values:  False
        Unique values:         789
        Longest value:         137 characters
        Most common values:    Accounting Policies [Line Items] (6x)
                               Accounting Policies [Table] (6x)
                               0.875% Notes Due 2025 [Member] (5x)
                               Accounting Policies, by Policy (Policies) [L
ine Items] (5x)
                               Accounting Policies, by Policy (Policies) [T
able] (5x)

  9. "doc"

        Type of data:          Text
        Contains null values:  True (excluded from calculations)
        Unique values:         646
        Longest value:         608 characters
        Most common values:    None (196x)
                               A written promise to pay a note to a third p
arty. (8x)
                               Line items represent financial concepts incl
uded in a table. These concepts are used to disclose reportable information
associated with domain members defined in one or many axes to the table. (5
x)
                               0.875% Notes Due 2025 [Member] (4x)
                               2014 Equity Incentive Plan [Member] (4x)


Row count: 999
```

In [52]:
```
%%sql
COPY TAG FROM '/home/ubuntu/notebooks/jan_tag.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
(psycopg2.errors.NotNullViolation) null value in column "iord" violates not
-null constraint
DETAIL:  Failing row contains (A0.0001EUR1850Million20202028NovartisFinance
S.A.LuxembourgLuxemb..., 0001370368-24-000004, t, t, member, null, null, 0.
000% 1 EUR 1 850 million 2020/2028 Novartis Finance S.A., Luxe..., null).
CONTEXT:  COPY tag, line 2: "A0.0001EUR1850Million20202028NovartisFinanceS.
A.LuxembourgLuxembourgIssuedAt99.354Member,0001370368-..."

[SQL: COPY TAG FROM '/home/ubuntu/notebooks/jan_tag.csv'
CSV
HEADER;]
(Background on this error at: https://sqlalche.me/e/14/gkpj)
```

Error:

IORD is indicated as not being able to take null values in the notes to the financial statement with the exception that it can be null IF ABSTRACT=1.

Because of this stipulation, we have updated the constraint to allow for IORD to take null values.

In [54]:
```
%%sql

ALTER TABLE tag
ALTER COLUMN iord DROP NOT NULL;
```

```
 * postgresql://student@/SECDB
Done.
```

Out[54]: []

Try load in the data again

In [55]:
```
%%sql
COPY TAG FROM '/home/ubuntu/notebooks/jan_tag.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
84276 rows affected.
```

Out[55]: []

In [56]:
```
!wc -l /home/ubuntu/notebooks/jan_tag.csv
```

```
84277 /home/ubuntu/notebooks/jan_tag.csv
```

Data successfully loaded:

84277 rows in jan_tag.csv *including* the header for a total of 84276 rows of data. In table TAG, we see that there are also 84276 rows, validating that the counts of the csv match the counts in the database.

The error message above is occurring because some key combinations of version and tag occurs more than once across ALL months in the final_tag.csv. While the primary key is unique in each month, we find duplicates upon stacking. To verify that the data in each column of these rows remain consistent across each occurrence of these combinations, we ran the below code. From this we can see that these rows are perfect duplicates of each other. In this case, we see that this version tag combo occurs exactly 2 times, one for each month of data.

In [14]:
```
!grep 'AcceleratedShareRepurchasesAdjustmentToRecordedAmount', 'us-gaap/2023
```

```
grep: us-gaap/2023: No such file or directory
/home/ubuntu/notebooks/final_tag.csv:AcceleratedShareRepurchasesAdjustmentT
oRecordedAmount,us-gaap/2023,0,0,monetary,D,D,"Accelerated Share Repurchase
s, Adjustment to Recorded Amount","The adjustment needed to reconcile previ
ously recorded amounts to the actual aggregate amount paid, whether in cash
or other consideration, to acquire all of the shares purchased under an Acc
elerated Share Repurchase arrangement."
/home/ubuntu/notebooks/final_tag.csv:AcceleratedShareRepurchasesAdjustmentT
oRecordedAmount,us-gaap/2023,0,0,monetary,D,D,"Accelerated Share Repurchase
s, Adjustment to Recorded Amount","The adjustment needed to reconcile previ
ously recorded amounts to the actual aggregate amount paid, whether in cash
or other consideration, to acquire all of the shares purchased under an Acc
elerated Share Repurchase arrangement."
```

Creating a table, TAG1 that does not include primary key constraint. This allows us to upload the data without recieving the error about unique contraints.

In [22]:
```
%%sql
DROP TABLE IF EXISTS TAG1 Cascade;

CREATE TABLE TAG1 (
    TAG VARCHAR(256) NOT NULL,
    VERSION VARCHAR(20) NOT NULL,
    CUSTOM BOOLEAN NOT NULL,
    ABSTRACT BOOLEAN NOT NULL,
    DATATYPE VARCHAR(20) NULL,
    IORD VARCHAR(1) NULL,
    CRDR VARCHAR(1) NULL,
    TLABEL VARCHAR(512) NULL,
    DOC VARCHAR(2048) NULL
);
```

```
 * postgresql://student@/SECDB
Done.
Done.
```

Out[22]:
```
[]
```

Verifying the number of rows in final_tag.csv

In [51]:
```
!wc -l /home/ubuntu/notebooks/Assignment2/final_tag.csv
```

```
619123 /home/ubuntu/notebooks/Assignment2/final_tag.csv
```

In [23]:
```
%%sql
COPY TAG1 FROM '/home/ubuntu/notebooks/Assignment2/final_tag.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
619122 rows affected.
```

Out[23]:
```
[]
```

619123 rows in final_tag.csv *including* the header for a total of 619122 rows of data. In table TAG1, we see that there are also 619122 rows, validating that the counts of the csv match the counts in the database prior to dropping duplicate rows from the database.

We can use the DISTINCT feature to select only unique rows from TAG1 to be added to table TAG. We can then go ahead and alter the new TAG table to include the primary key combination of tag and version.

In [24]:
```
%%sql
DROP TABLE IF EXISTS TAG Cascade;
CREATE TABLE TAG AS
SELECT DISTINCT tag,version,custom,abstract,datatype,iord,crdr,tlabel,doc
FROM TAG1;
```

```
 * postgresql://student@/SECDB
Done.
608029 rows affected.
```

Out[24]: []

In [25]:
```
%%sql
ALTER TABLE TAG ADD PRIMARY KEY (TAG, VERSION)
```

```
 * postgresql://student@/SECDB
Done.
```

Out[25]: []

In [26]:
```
%%sql
DROP TABLE IF EXISTS TAG1 Cascade
```

```
 * postgresql://student@/SECDB
Done.
```

Out[26]: []

Queries to determine if only NULL values for IORD exists when ABSTRACT=1.

In [8]:
```
%%sql
SELECT COUNT(ABSTRACT)
FROM TAG
WHERE IORD is NULL;
```

```
 * postgresql://student@/SECDB
1 rows affected.
```

Out[8]:

count
274627

In [10]:
```
%%sql
SELECT COUNT(ABSTRACT)
FROM TAG
WHERE IORD is NULL
AND ABSTRACT is TRUE;
```

```
 * postgresql://student@/SECDB
1 rows affected.
```

Out[10]:

count
274627

Loading DIM dataset into the DIM table

In [57]:
```
%%sql
COPY DIM FROM '/home/ubuntu/notebooks/jan_dim.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
(psycopg2.errors.StringDataRightTruncation) value too long for type charact
er varying(32)
CONTEXT:  COPY dim, line 2, column dimhash: "0xef558142c09cb75426d7735bd81e
cdde"

[SQL: COPY DIM FROM '/home/ubuntu/notebooks/jan_dim.csv'
CSV
HEADER;]
(Background on this error at: https://sqlalche.me/e/14/9h9h)
```

Error:

The initial varchar(32) constraint was too short for some of the incoming data. There are values that are more than 32 characters.

Next Step:

By increasing the length to 34 characters, we ensure that longer dimhash values can be stored without causing errors during data loading. This command alters the dimhash column to change its data type to varchar(34) in order to accommodate longer dimhash values.

In [58]:
```
%%sql

ALTER TABLE dim
ALTER COLUMN dimhash TYPE varchar(34);
```

```
 * postgresql://student@/SECDB
Done.
```

Out[58]:
```
[]
```

Try loading data again

In [126…
```
%%sql
COPY DIM FROM '/home/ubuntu/notebooks/jan_dim.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
(psycopg2.errors.NotNullViolation) null value in column "segments" violates
not-null constraint
DETAIL:  Failing row contains (0x00000000, null, f).
CONTEXT:  COPY dim, line 20896: "0x00000000,,0"

[SQL: COPY DIM FROM '/home/ubuntu/notebooks/jan_dim.csv'
CSV
HEADER;]
(Background on this error at: https://sqlalche.me/e/14/gkpj)
```

Error:

When running the COPY DIM FROM command, we encountered an error due to NULL values in the segments column, which has a NOT NULL constraint.

Next Step:

We had to alter the table to allow NULL values and successfully load the data.

In [59]:
```
%%sql

ALTER TABLE DIM
ALTER COLUMN segments DROP NOT NULL;
```

```
 * postgresql://student@/SECDB
Done.
```

Out[59]: []

Try load the data again

In [60]:
```
%%sql

COPY DIM FROM '/home/ubuntu/notebooks/jan_dim.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
43201 rows affected.
```

Out[60]: []

In [61]:
```
!csvstat --count jan_dim.csv
```

```
43201
```

Data successfully loaded:

The rows from the Jan DIM csv file matches the number of rows loaded on the DIM table in the database which confirms that our data was successfully loaded.

Before copying the dim data into the dim table, we need to create a unique dim.csv file, as we encountered the error psycopg2.errors.UniqueViolation. This error occurs because the dimhash column has a uniqueness constraint, which ensures that no two rows in the dim table can have the same value for dimhash. Since we are loading the data for two months, the primary key might have the duplication for dimensions dataset. To resolve this, we used the command below to create another csv with unique dim values.

In [74]:
```
!{ head -n 1 final_dim.csv && tail -n +2 final_dim.csv | sort | uniq; } > ur
```

In [29]:
```
%%sql
COPY DIM FROM '/home/ubuntu/notebooks/Assignment2/unique_dim.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
460390 rows affected.
```

Out[29]: []

In [11]:
```
!csvstat --count unique_dim.csv
```

```
460390
```

Loading NUM dataset into the NUM table

In [62]:
```
%%sql

COPY NUM FROM '/home/ubuntu/notebooks/jan_num.csv'
```

```
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
(psycopg2.errors.StringDataRightTruncation) value too long for type charact
er varying(32)
CONTEXT:  COPY num, line 4, column dimh: "0xad66b56c81571b0e9d4882f564c0617
4"

[SQL: COPY NUM FROM '/home/ubuntu/notebooks/jan_num.csv'
CSV
HEADER;]
(Background on this error at: https://sqlalche.me/e/14/9h9h)
```

Error

The error we encountered whows the value is too long for type character varying(32), indicating that the data being inserted into the dimh column exceeds the maximum length defined for that column, which is 32 characters.

Next Step

Since the initial varchar(32) constraint was too short for some of the incoming data, so by increasing the length to 34 characters, we want to ensure that longer dimh values can be stored without causing errors during data loading. After running the command below, the dimh column will be able to store strings up to 34 characters long.

In [63]:
```
%%sql

ALTER TABLE NUM
ALTER COLUMN dimh TYPE varchar(34);
```

```
 * postgresql://student@/SECDB
Done.
```

Out[63]:
```
[]
```

Load data again

In [64]:
```
%%sql

COPY NUM FROM '/home/ubuntu/notebooks/jan_num.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
(psycopg2.errors.NumericValueOutOfRange) numeric field overflow
DETAIL:  A field with precision 2, scale 0 must round to an absolute value
less than 10^2.
CONTEXT:  COPY num, line 283, column dcml: "-32768"

[SQL: COPY NUM FROM '/home/ubuntu/notebooks/jan_num.csv'
CSV
HEADER;]
(Background on this error at: https://sqlalche.me/e/14/9h9h)
```

Error

The error message refers to a column dcml defined as NUMERIC(2,0). The precision (2) allows a maximum of 2 digits. The scale (0) indicates no decimal places, meaning only

whole numbers are permitted. The data includes a value -32768, which has five digits and is negative. Since the column only allows values between -99 and 99 (absolute values within two digits), -32768 exceeds this limit, causing the overflow.

Next Step

To accommodate larger numbers, we increase the column's precision to NUMERIC(5,0). This would allow values with up to five digits, which would include -32768.

In [65]:
```
%%sql

ALTER TABLE num
ALTER COLUMN dcml TYPE NUMERIC(5,0);
```

```
 * postgresql://student@/SECDB
Done.
```

Out[65]: []

Load data again

In [66]:
```
%%sql

COPY NUM FROM '/home/ubuntu/notebooks/jan_num.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
598843 rows affected.
```

Out[66]: []

In [67]:
```
!wc -l /home/ubuntu/notebooks/jan_num.csv
```

```
598844 /home/ubuntu/notebooks/jan_num.csv
```

Data successfully loaded

598844 rows in jan_num.csv *including* the header for a total of 598843 rows of data. In table NUM, we see that there are also 598843 rows, validating that the counts of the csv match the counts in the database.

In [82]:
```
%%sql

ALTER TABLE NUM
ALTER COLUMN FOOTLEN TYPE NUMERIC(5,0);
```

```
 * postgresql://student@/SECDB
Done.
```

Out[82]: []

In [83]:
```
%%sql

COPY NUM FROM '/home/ubuntu/notebooks/Assignment2/final_num.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
(psycopg2.errors.UniqueViolation) duplicate key value violates unique const
raint "num_pkey"
DETAIL:  Key (adsh, tag, version, ddate, qtrs, uom, dimh, iprx)=(0000002488
-24-000012, AccountsPayableCurrent, us-gaap/2023, 2022-12-31, 0, USD, 0x000
00000, 0) already exists.
CONTEXT:  COPY num, line 2

[SQL: COPY NUM FROM '/home/ubuntu/notebooks/Assignment2/jan_num.csv'
CSV
HEADER;]
(Background on this error at: https://sqlalche.me/e/14/gkpj)
```

In [33]:
```
%%sql

ALTER TABLE NUM
ALTER COLUMN VALUE TYPE NUMERIC(20,4);
```

```
 * postgresql://student@/SECDB
Done.
```

Out[33]: []

In [9]:
```
%%sql

COPY NUM FROM '/home/ubuntu/notebooks/Assignment2/final_num.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
(psycopg2.errors.NumericValueOutOfRange) numeric field overflow
DETAIL:  A field with precision 20, scale 4 must round to an absolute value
less than 10^16.
CONTEXT:  COPY num, line 734836, column value: "75000000000000000.0000"

[SQL: COPY NUM FROM '/home/ubuntu/notebooks/final_num.csv'
CSV
HEADER;]
(Background on this error at: https://sqlalche.me/e/14/9h9h)
```

In [34]:
```
%%sql

ALTER TABLE NUM
ALTER COLUMN VALUE TYPE NUMERIC(22);
```

```
 * postgresql://student@/SECDB
Done.
```

Out[34]: []

In [ ]:
```
%%sql

COPY NUM FROM '/home/ubuntu/notebooks/Assignment2/final_num.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
```

In [34]:
```
!wc -l /home/ubuntu/notebooks/Assignment2/jan_num.csv
```

```
598844 /home/ubuntu/notebooks/Assignment2/jan_num.csv
```

IF THE COMMAND ABOVE RUNS SUCCESSFULLY THEN PROCEED TO LOADING PRE

Loading PRE dataset into the PRE table

In [68]:
```
%%sql

COPY PRE FROM '/home/ubuntu/notebooks/jan_pre.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
(psycopg2.errors.NotNullViolation) null value in column "stmt" violates not
-null constraint
DETAIL:  Failing row contains (0001828852-24-000024, 1, 22, null, f, Amendm
entFlag, dei/2022, terseLabel, Amendment Flag, f).
CONTEXT:  COPY pre, line 2: "0001828852-24-000024,1,22,,0,AmendmentFlag,de
i/2022,terseLabel,Amendment Flag,0"

[SQL: COPY PRE FROM '/home/ubuntu/notebooks/jan_pre.csv'
CSV
HEADER;]
(Background on this error at: https://sqlalche.me/e/14/gkpj)
```

Error:

Since stmt does have NULL values, it violates the not-null constraint. One approach to resolve this would be to fill in all the missing values based on the financial statement each row belongs to. However, we do not have access to the original XBRL submission data or documentation that specifies this information, so we cannot go forward with this approach.

Next Step

As per the documentation, the acceptable values for the column are: CP = Cover Page, BS = Balance Sheet, IS = Income Statement, CF = Cash Flow, EQ = Equity, CI = Comprehensive Income, and UN = Unclassifiable Statement. Since we're unsure which value applies to each row, we can use UN (Unclassifiable Statement) as a temporary default for rows where the specific financial statement is unknown. We would first remove the NOT NULL constraint and then once the data is loaded we will update the table.

In [69]:
```
%%sql
ALTER TABLE PRE
ALTER COLUMN stmt DROP NOT NULL;
```

```
 * postgresql://student@/SECDB
Done.
```

Out[69]:
```
[]
```

Load the data again

In [70]:
```
%%sql

COPY PRE FROM '/home/ubuntu/notebooks/jan_pre.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
(psycopg2.errors.NotNullViolation) null value in column "plabel" violates n
ot-null constraint
DETAIL:  Failing row contains (0001185185-24-000066, 2, 31, BS, f, Stockhol
dersEquityBeforeTreasuryStock, us-gaap/2023, totalLabel, null, f).
CONTEXT:  COPY pre, line 14418: "0001185185-24-000066,2,31,BS,0,Stockholder
sEquityBeforeTreasuryStock,us-gaap/2023,totalLabel,,0"

[SQL: COPY PRE FROM '/home/ubuntu/notebooks/jan_pre.csv'
CSV
HEADER;]
(Background on this error at: https://sqlalche.me/e/14/gkpj)
```

Error:

The "plabel" field provides descriptive labels for financial statement tags, but some values are missing, which violates the "not-null" rule.

Next Step

Normally, we'd try to find and fill in these missing values, but we don't have enough information from the documentation or the sponsor. However, since "plabel" is a string and mainly adds extra detail to each tag in the PRE table, it doesn't seem crucial for calculations, especially because the "tag" field itself is not null. This means that even if "plabel" is missing, we can still get the necessary information from "tag." Given that there are only 270 missing values in a large dataset, we decided to leave them as is by dropping the NULL constraint on it to load the data.

In [71]:
```
%%sql
ALTER TABLE PRE
ALTER COLUMN plabel DROP NOT NULL;
```

```
 * postgresql://student@/SECDB
Done.
```

Out[71]:
```
[]
```

We created an ALTER TABLE statement for both the stmt and plabel attributes to change their constraints from NOT NULL to NULL. Although the financial statement specifies that these fields may not be null, our data analysis reveals actual null values in these columns. To address this discrepancy and avoid error messages, it's necessary to update the constraints accordingly.

In [72]:
```
%%sql

COPY PRE FROM '/home/ubuntu/notebooks/jan_pre.csv'
CSV
HEADER;
```

```
 * postgresql://student@/SECDB
511413 rows affected.
```

Out[72]:
```
[]
```

In [73]:
```
!wc -l /home/ubuntu/notebooks/jan_pre.csv
```

```
511414 /home/ubuntu/notebooks/jan_pre.csv
```

Data sucessfully loaded:

511414 rows in jan_pre.csv *including* the header for a total of 511413 rows of data. In table PRE, we see that there are also 511413 rows, validating that the counts of the csv match the counts in the database.

Imputing values for stmt

To resolve the ERROR: column 'stmt' contains NULL values, which violates the NOT NULL constraint, we first have to figure out how many NULL values are there. This (SELECT) command checks how many rows in the PRE table have NULL values in the stmt column. To do this, it selects the count of version, a column that has NO NULL values, where stmt is NULL. By counting version instead of stmt, we ensure each row is accounted for accurately, as using stmt would result in an output of "0 rows affected" due to the presence of NULL values. This approach gives us a clear count of rows in which stmt is missing data. After we run the code, we see that there are **"413897"** rows that are NULL in stmt.

In [74]:
```
%%sql
SELECT COUNT(version)
FROM PRE
WHERE stmt IS NULL;
```

```
 * postgresql://student@/SECDB
1 rows affected.
```

Out[74]:

count
413897

Secondly, since we determined that the best approach is to enter "UN" for the missing NULL values in the stmt column, we need to implement this. The UPDATE and SET command will update the stmt column and replace any NULL values with "UN." After running the code, we get the same result: **"413897"**, indicating that our command was successful and "UN" has been entered for all the missing values.

In [75]:
```
%%sql
UPDATE PRE
SET stmt = 'UN'
WHERE stmt IS NULL;
```

```
 * postgresql://student@/SECDB
413897 rows affected.
```

Out[75]:
```
[]
```

To address our *second error, ERROR: column 'plabel' contains NULL values, which violates the NOT NULL constraint, we need to determine how many NULL values are present in that column. We will use the same method as before, referencing the version column in our SELECT command because the plabel column contains NULL values, which would result in "0 rows affected" if we used it directly. By replacing stmt with plabel in the command, we obtain an output of 87 count*. Since this is a small number, we have decided to leave it as is, as it has little significance, as previously explained.

In [76]:
```
%%sql
SELECT COUNT(version)
```

```
FROM PRE
WHERE plabel IS NULL;
```

```
 * postgresql://student@/SECDB
1 rows affected.
```

Out[76]:

count
87

Part III: Data Analysis

Queries:

Q1

This question asks for information about submissions from the database. We need to formulate a query that selects specific columns to display: "Submission ID": The unique identifier for each submission. "Company Name": The name associated with company that made the submission. "Submission Date": The date when the submission was filed. "Mailing Address": The complete mailing address derived from concatenating the "MAS1", "CITYMA", "ZIPMA", and "COUNTRYMA" columns. The query should filter the results to include submissions that fall into two specific Standard Industrial Classification (SIC) code ranges: 8000 to 8093: This range corresponds to Health Services. 3821 to 3873: This range corresponds to Medical Equipment and Supplies.

In [153…

```
%%sql
SELECT ADSH as "Submission ID",
    NAME as "Name",
    FILED as "Submission Date",
    MAS1 || ', ' || CITYMA || ' ' || ZIPMA || ', ' || COUNTRYMA as "Mailing
FROM SUB
WHERE SIC BETWEEN 8000 AND 8093
    OR SIC BETWEEN 3821 AND 3873
```

```
 * postgresql://student@/SECDB
463 rows affected.
```

Out[153]:

Submission ID	Name	Submission Date	Mailing Address
0000920148-24-000002	LABORATORY CORP OF AMERICA HOLDINGS	2024-01-02	358 S MAIN ST, BURLINGTON 27215, US
0000950170-24-000014	UNIVERSAL HEALTH SERVICES INC	2024-01-02	367 S GULPH ROAD, KING OF PRUSSIA 19406, US
0000950170-24-000306	OUTSET MEDICAL, INC.	2024-01-02	3052 ORCHARD DRIVE, SAN JOSE 95134, US
0001104659-24-000146	ORTHOPEDIATRICS CORP	2024-01-02	2850 FRONTIER DRIVE, WARSAW 46582, US
0001104659-24-000386	PRECIPIO, INC.	2024-01-02	4 SCIENCE PARK, NEW HAVEN 06511, US
0001193125-24-000620	AKUMIN INC.	2024-01-02	8300 W SUNRISE BLVD, PLANTATION 33322, US
0001193125-24-000684	TRANSMEDICS GROUP, INC.	2024-01-02	200 MINUTEMAN ROAD, ANDOVER 01810, US
0001213900-24-000317	BLUEJAY DIAGNOSTICS, INC.	2024-01-02	360 MASSACHUSETTS AVENUE, SUITE 203, ACTON 01720, US
0001437749-24-000113	PULSE BIOSCIENCES, INC.	2024-01-02	3957 POINT EDEN WAY, HAYWARD 94545, US
0001493152-23-046641	MANGOCEUTICALS, INC.	2024-01-02	15110 DALLAS PKWY, SUITE 600, DALLAS 75248, US
0001493152-24-000035	DAXOR CORP	2024-01-02	350 5TH AVENUE, NEW YORK 10118, US
0001493152-24-000071	GLUCOTRACK, INC.	2024-01-02	301 RT 17 NORTH, RUTHERFORD 07070, US
0001493152-24-000083	AGAPE ATP CORP	2024-01-02	1705-1708 LEVEL 17, TOWER 2, FABER TOWER, KUALA LUMPUR 58100, MY
0001493152-24-000109	MICROBOT MEDICAL INC.	2024-01-02	25 RECREATION PARK DR SUITE 108, HINGHAM 02043, US
0001493152-24-000132	BONE BIOLOGICS CORP	2024-01-02	2 BURLINGTON WOODS DRIVE,, BURLINGTON 01803, US
0001683168-24-000004	MEDICALE CORP.	2024-01-02	9314 FOREST HILL BLVD, WELLINGTON 33411, US
0000950170-24-000544	ALPHATEC HOLDINGS, INC.	2024-01-03	5818 EL CAMINO REAL, CARLSBAD 92008, US
0001079973-24-000004	NEMAURA MEDICAL INC.	2024-01-03	57 WEST 57TH STREET, MANHATTAN 10019, US
0001079973-24-000007	PRO DEX INC	2024-01-03	2361 MCGAW AVENUE, IRVINE 92614, US
0001079973-24-000008	NEMAURA MEDICAL INC.	2024-01-03	57 WEST 57TH STREET, MANHATTAN 10019, US
0001104659-24-000868	INVIVO THERAPEUTICS HOLDINGS CORP.	2024-01-03	ONE KENDALL SQUARE, CAMBRIDGE 02139, US
0001104659-24-000906	ASSURE HOLDINGS CORP.	2024-01-03	7887 EAST BELLEVIEW AVENUE, GREENWOOD VILLAGE 80111, US

Submission ID	Name	Submission Date	Mailing Address
0001104659-24-000910	ASSURE HOLDINGS CORP.	2024-01-03	7887 EAST BELLEVIEW AVENUE, GREENWOOD VILLAGE 80111, US
0001140361-24-000342	STRATA SKIN SCIENCES, INC.	2024-01-03	5 WALNUT GROVE DRIVE, HORSHAM 19044, US
0001140361-24-000483	SANUWAVE HEALTH, INC.	2024-01-03	3360 MARTIN FARM RD, SUWANEE 30024, US
0001193125-24-000976	ENOVIS CORP	2024-01-03	2711 CENTERVILLE ROAD, WILMINGTON 19808, US
0001193125-24-001477	COGNEX CORP	2024-01-03	ONE VISION DRIVE, NATICK 01760, US
0001437749-24-000290	PERSPECTIVE THERAPEUTICS, INC.	2024-01-03	2401 ELLIOTT AVENUE, SEATTLE 98121, US
0001477932-24-000016	HAWKEYE SYSTEMS, INC.	2024-01-03	6605 ABERCORN, SAVANNAH 31405, US
0001493152-24-000338	INVO BIOSCIENCE, INC.	2024-01-03	5582 BROADCAST COURT, SARASOTA 34240, US
0001493152-24-000339	ENDONOVO THERAPEUTICS, INC.	2024-01-03	6320 CANOGA AVENUE, WOODLAND HILLS 91367, US
0000805928-24-000005	AXOGEN, INC.	2024-01-04	13631 PROGRESS BLVD., ALACHUA 32615, US
0000950157-24-000024	ILLUMINA, INC.	2024-01-04	5200 ILLUMINA WAY, SAN DIEGO 92122, US
0001037646-24-000002	METTLER TOLEDO INTERNATIONAL INC/	2024-01-04	1900 POLARIS PARKWAY, COLUMBUS 43240, US
0001079973-24-000018	PRO DEX INC	2024-01-04	2361 MCGAW AVENUE, IRVINE 92614, US
0001104659-24-001401	SENSEONICS HOLDINGS, INC.	2024-01-04	20451 SENECA MEADOWS PARKWAY, GERMANTOWN 20876, US
0001116463-24-000003	ORASURE TECHNOLOGIES INC	2024-01-04	220 E FIRST ST, BETHLEHEM 18015, US
0001171843-24-000073	PREDICTIVE ONCOLOGY INC.	2024-01-04	91 43RD STREET, PITTSBURGH 15201, US
0001213900-24-001160	NOCTURNE ACQUISITION CORP	2024-01-04	3 GERMAY DRIVE, WILMINGTON 19804, US
0001323885-24-000004	ATRICURE, INC.	2024-01-04	7555 INNOVATION WAY, MASON 45040, US
0001437749-24-000610	ASENSUS SURGICAL, INC.	2024-01-04	1 TW ALEXANDER DRIVE, DURHAM 27703, US
0001477932-24-000045	MIDWEST ENERGY EMISSIONS CORP.	2024-01-04	1810 JESTER DRIVE, CORSICANA 75109, US
0001493152-24-000514	SINTX TECHNOLOGIES, INC.	2024-01-04	1885 WEST 2100 STREET, SALT LAKE CITY 84119, US
0001493152-24-000517	FG GROUP HOLDINGS INC.	2024-01-04	4201 CONGRESS STREET, CHARLOTTE 28209, US
0001493152-24-001075	THERAPEUTIC SOLUTIONS INTERNATIONAL, INC.	2024-01-04	701 WILD ROSE LANE, ELK CITY 83525, US

Submission ID	Name	Submission Date	Mailing Address
0001493152-24-001103	INTELLIGENT BIO SOLUTIONS INC.	2024-01-04	142 WEST 57TH ST FL11, NEW YORK 10019, US
0001606498-24-000002	AVANOS MEDICAL, INC.	2024-01-04	5405 WINDWARD PARKWAY, ALPHARETTA 30004, US
0001645113-24-000002	NOVOCURE LTD	2024-01-04	NO. 4 THE FORUM, ST. HELIER JE2 4UF, JE
0001654954-24-000194	CATHETER PRECISION, INC.	2024-01-04	1670 HIGHWAY 160 WEST, FORT MILL 29708, US
0001683168-24-000066	ACLARION, INC.	2024-01-04	8181 ARISTA PLACE, BROOMFIELD 80021, US
0001683168-24-000077	CARDIFF LEXINGTON CORP	2024-01-04	3753 HOWARD HUGHES PARKWAY, LAS VEGAS 89169, US
0000805928-24-000011	AXOGEN, INC.	2024-01-05	13631 PROGRESS BLVD., ALACHUA 32615, US
0000950170-24-002303	INOGEN INC	2024-01-05	859 WARD DRIVE, GOLETA 93111, US
0000950170-24-002304	PERSONALIS, INC.	2024-01-05	6600 DUMBARTON CIRCLE, FREMONT 94555, US
0000950170-24-002495	CARMELL CORP	2024-01-05	1177 AVENUE OF THE AMERICAS, NEW YORK 10036, US
0001090872-24-000003	AGILENT TECHNOLOGIES, INC.	2024-01-05	5301 STEVENS CREEK BLVD, MS 1A-LC, SANTA CLARA 95052-8059, US
0001104659-24-001640	STANDARD BIOTOOLS INC.	2024-01-05	2 TOWER PLACE, SOUTH SAN FRANCISCO 94080, US
0001104659-24-001643	GREENBROOK TMS INC.	2024-01-05	890 YONGE STREET, 7TH FLOOR, TORONTO M4W 3P4, CA
0001104659-24-001818	TRISALUS LIFE SCIENCES, INC.	2024-01-05	6272 WEST 91ST AVENUE, WESTMINSTER 80031, US
0001104659-24-001822	CLEARSIGN TECHNOLOGIES CORP	2024-01-05	8023 E. 63RD PLACE, SUITE 101, TULSA 74133, US
0001140361-24-000830	ANGIODYNAMICS INC	2024-01-05	14 PLAZA DRIVE, LATHAM 12110, US
0001171843-24-000099	HARVARD BIOSCIENCE INC	2024-01-05	84 OCTOBER HILL ROAD, HOLLISTON 01746, US
0001185185-24-000044	PACIFIC HEALTH CARE ORGANIZATION INC	2024-01-05	1201 DOVE STREET, NEWPORT BEACH 92260, US
0001193125-24-002878	ZIMMER BIOMET HOLDINGS, INC.	2024-01-05	345 EAST MAIN STREET, WARSAW 46580, US
0001193125-24-003341	T2 BIOSYSTEMS, INC.	2024-01-05	101 HARTWELL AVENUE, LEXINGTON 02421, US
0001193125-24-003356	CERUS CORP	2024-01-05	1220 CONCORD AVENUE, CONCORD 94520, US
0001193125-24-003380	SOLENO THERAPEUTICS INC	2024-01-05	203 REDWOOD SHORES PARKWAY, REDWOOD CITY 94065, US

Submission ID	Name	Submission Date	Mailing Address
0001193125-24-003383	AKILI, INC.	2024-01-05	71 COMMERCIAL STREET, BOSTON 02109, US
0001213900-24-001552	SPECTRAL AI, INC.	2024-01-05	2515 MCKINNEY AVE #1000, DALLAS 75201, US
0001437749-24-000752	PERSPECTIVE THERAPEUTICS, INC.	2024-01-05	2401 ELLIOTT AVENUE, SEATTLE 98121, US
0001477932-24-000065	BIOCORRX INC.	2024-01-05	2390 EAST ORANGEWOOD AVENUE, ANAHEIM 92806, US
0001493152-24-001145	OPTEX SYSTEMS HOLDINGS INC	2024-01-05	1420 PRESIDENTIAL DRIVE, RICHARDSON 75081, US
0001493152-24-001155	STRONG GLOBAL ENTERTAINMENT, INC.	2024-01-05	5960 FAIRVIEW ROAD, SUITE 275, CHARLOTTE 28210, US
0001493152-24-001392	SYNTEC OPTICS HOLDINGS, INC.	2024-01-05	1111 LINCOLN ROAD, MIAMI BEACH 33139, US
0001558370-24-000087	HURCO COMPANIES INC	2024-01-05	ONE TECHNOLOGY WAY, INDIANAPOLIS 46268, US
0001558370-24-000115	HURCO COMPANIES INC	2024-01-05	ONE TECHNOLOGY WAY, INDIANAPOLIS 46268, US
0001628280-24-000516	AGILON HEALTH, INC.	2024-01-05	6210 E HWY 290, AUSTIN 78723, US
0001628280-24-000518	AGILON HEALTH, INC.	2024-01-05	6210 E HWY 290, AUSTIN 78723, US
0001628280-24-000678	MIRION TECHNOLOGIES, INC.	2024-01-05	1218 MENLO DRIVE, ATLANTA 30318, US
0001628945-24-000005	CUE HEALTH INC.	2024-01-05	4980 CARROLL CANYON ROAD, SAN DIEGO 92121, US
0001683168-24-000097	ODYSSEY HEALTH, INC.	2024-01-05	2300 WEST SAHARA AVENUE, LAS VEGAS 89102, US
0001683168-24-000099	RADNET, INC.	2024-01-05	1510 COTNER AVE, LOS ANGELES 90025, US
0001829126-24-000068	INNOVATIVE EYEWEAR INC	2024-01-05	66 W FLAGLER ST. SUITE 900, MIAMI 33130, US
0000313616-24-000004	DANAHER CORP /DE/	2024-01-08	2200 PENNSYLVANIA AVE. N.W., WASHINGTON 20037-1701, US
0000719135-24-000003	APYX MEDICAL CORP	2024-01-08	5115 ULMERTON ROAD, CLEARWATER 33760, US
0000805928-24-000013	AXOGEN, INC.	2024-01-08	13631 PROGRESS BLVD., ALACHUA 32615, US
0000856982-24-000004	MERIT MEDICAL SYSTEMS INC	2024-01-08	1600 WEST MERIT PARKWAY, SOUTH JORDAN 84095, US
0000885725-24-000003	BOSTON SCIENTIFIC CORP	2024-01-08	300 BOSTON SCIENTIFIC WAY, MARLBOROUGH 01752-1234, US
0000885725-24-000007	BOSTON SCIENTIFIC CORP	2024-01-08	300 BOSTON SCIENTIFIC WAY, MARLBOROUGH 01752-1234, US
0000950170-24-002635	SINGULAR GENOMICS SYSTEMS, INC.	2024-01-08	3010 SCIENCE PARK ROAD, SAN DIEGO 92121, US

Submission ID	Name	Submission Date	Mailing Address
0000950170-24-002636	RXSIGHT, INC.	2024-01-08	100 COLUMBIA STREET, ALISO VIEJO 92656, US
0000950170-24-002661	LYRA THERAPEUTICS, INC.	2024-01-08	480 ARSENAL WAY, WATERTOWN 02472, US
0000950170-24-002703	ALLURION TECHNOLOGIES, INC.	2024-01-08	11 HURON DR STE 200, NATICK 01760, US
0000950170-24-002719	TREACE MEDICAL CONCEPTS, INC.	2024-01-08	100 PALMETTO PARK PLACE, PONTE VEDRA 32081, US
0000950170-24-002726	CERUS CORP	2024-01-08	1220 CONCORD AVENUE, CONCORD 94520, US
0000950170-24-002731	OUTSET MEDICAL, INC.	2024-01-08	3052 ORCHARD DRIVE, SAN JOSE 95134, US
0000950170-24-002736	ALPHATEC HOLDINGS, INC.	2024-01-08	5818 EL CAMINO REAL, CARLSBAD 92008, US
0000950170-24-002738	STAAR SURGICAL CO	2024-01-08	1911 WALKER AVE, MONROVIA 91016, US
0000950170-24-002823	BRUKER CORP	2024-01-08	40 MANNING RD, BILLERICA 01821, US
0000950170-24-002924	ORTHOFIX MEDICAL INC.	2024-01-08	3451 PLANO PARKWAY, LEWISVILLE 75056, US
0001093557-24-000003	DEXCOM INC	2024-01-08	6340 SEQUENCE DRIVE, SAN DIEGO 92121, US
0001104659-24-001980	AKOYA BIOSCIENCES, INC.	2024-01-08	100 CAMPUS DRIVE, MARLBOROUGH 01762, US
0001104659-24-001982	ACCELERATE DIAGNOSTICS, INC	2024-01-08	3950 S. COUNTRY CLUB ROAD #470, TUCSON 85714, US
0001104659-24-002071	ORTHOPEDIATRICS CORP	2024-01-08	2850 FRONTIER DRIVE, WARSAW 46582, US
0001104659-24-002089	CVRX, INC.	2024-01-08	9201 WEST BROADWAY AVENUE, MINNEAPOLIS 55445, US
0001104659-24-002092	STANDARD BIOTOOLS INC.	2024-01-08	2 TOWER PLACE, SOUTH SAN FRANCISCO 94080, US
0001124140-24-000003	EXACT SCIENCES CORP	2024-01-08	5505 ENDEAVOR LANE, MADISON 53719, US
0001140361-24-001200	BIONANO GENOMICS, INC.	2024-01-08	9540 TOWNE CENTRE DRIVE, SAN DIEGO 92121, US
0001140361-24-001215	NUWELLIS, INC.	2024-01-08	12988 VALLEY VIEW ROAD, EDEN PRAIRIE 55344, US
0001140361-24-001339	STRATA SKIN SCIENCES, INC.	2024-01-08	5 WALNUT GROVE DRIVE, HORSHAM 19044, US
0001157523-24-000028	HOLOGIC INC	2024-01-08	250 CAMPUS DRIVE, MARLBOROUGH 01752, US
0001193125-24-003816	DELCATH SYSTEMS, INC.	2024-01-08	1633 BROADWAY, NEW YORK 10019, US
0001193125-24-003861	NEUROPACE INC	2024-01-08	455 N. BERNARDO AVENUE, MOUNTAIN VIEW 94043, US

Submission ID	Name	Submission Date	Mailing Address
0001193125-24-003922	ATRICURE, INC.	2024-01-08	7555 INNOVATION WAY, MASON 45040, US
0001193125-24-003936	VIRIDIAN THERAPEUTICS, INC.DE	2024-01-08	221 CRESCENT STREET, WALTHAM 02453, US
0001193125-24-003949	VERACYTE, INC.	2024-01-08	6000 SHORELINE COURT, SUITE 300, SOUTH SAN FRANCISCO 94080, US
0001193125-24-003966	NEURONETICS, INC.	2024-01-08	3222 PHOENIXVILLE PIKE, MALVERN 19355, US
0001193125-24-003983	ENOVIS CORP	2024-01-08	2711 CENTERVILLE ROAD, WILMINGTON 19808, US
0001193125-24-003996	PARAGON 28, INC.	2024-01-08	14445 GRASSLANDS DRIVE, ENGLEWOOD 80112, US
0001193125-24-004446	GUARDANT HEALTH, INC.	2024-01-08	3100 HANOVER STREET, PALO ALTO 94304, US
0001193125-24-004495	BETTER THERAPEUTICS, INC.	2024-01-08	548 MARKET ST. #49404, SAN FRANCISCO 94101, US
0001213900-24-001761	SS INNOVATIONS INTERNATIONAL, INC.	2024-01-08	1600 SE 15TH STREET, FORT LAUDERDALE 33316, US
0001274737-24-000006	EXAGEN INC.	2024-01-08	1261 LIBERTY WAY, VISTA 92081, US
0001285550-24-000016	CLEARPOINT NEURO, INC.	2024-01-08	120 S. SIERRA AVENUE, SOLANA BEACH 92075, US
0001299130-24-000005	PACIFIC BIOSCIENCES OF CALIFORNIA, INC.	2024-01-08	1305 O'BRIEN DRIVE, MENLO PARK 94025, US
0001388658-24-000006	IRHYTHM TECHNOLOGIES, INC.	2024-01-08	699 8TH STREET, San Francisco 94103, US
0001425450-24-000002	ORTHOPEDIATRICS CORP	2024-01-08	2850 FRONTIER DRIVE, WARSAW 46582, US
0001437749-24-000811	ASENSUS SURGICAL, INC.	2024-01-08	1 TW ALEXANDER DRIVE, DURHAM 27703, US
0001437749-24-000813	CRAWFORD UNITED CORP	2024-01-08	10514 DUPONT AVE, CLEVELAND 44108, US
0001437749-24-000821	MESA LABORATORIES INC /CO/	2024-01-08	12100 W 6TH AVE, LAKEWOOD 80228, US
0001447362-24-000006	CASTLE BIOSCIENCES INC	2024-01-08	505 S FRIENDSWOOD DRIVE, FRIENDSWOOD 77546, US
0001459839-24-000003	SI-BONE, INC.	2024-01-08	471 EL CAMINO REAL, SUITE 101, SANTA CLARA 95050, US
0001493152-24-001677	BIOSIG TECHNOLOGIES, INC.	2024-01-08	55 GREENS FARMS ROAD, WESTPORT 06880, US
0001493152-24-001690	SIGYN THERAPEUTICS, INC.	2024-01-08	9190 W OLYMPIC BLVD # 263, BEVERLY HILLS 90212, US
0001493152-24-001710	PAVMED INC.	2024-01-08	360 MADISON AVENUE, NEW YORK 10017, US
0001493152-24-001712	AURORA TECHNOLOGY ACQUISITION CORP.	2024-01-08	4 EMBARCADERO CENTER SUITE 1449, SAN FRANCISCO 94105, US

Submission ID	Name	Submission Date	Mailing Address
0001493152-24-001719	CLEARDAY, INC.	2024-01-08	15511 W. STATE HWY 71, AUSTIN 78738, US
0001603756-24-000017	AXONICS, INC.	2024-01-08	26 TECHNOLOGY DRIVE, IRVINE 92618, US
0001603756-24-000023	AXONICS, INC.	2024-01-08	26 TECHNOLOGY DRIVE, IRVINE 92618, US
0001609550-24-000004	INSPIRE MEDICAL SYSTEMS, INC.	2024-01-08	5500 WAYZATA BLVD, GOLDEN VALLEY 55416, US
0001628280-24-000776	BIOLIFE SOLUTIONS INC	2024-01-08	3303 MONTE VILLA PARKWAY, BOTHELL 98021, US
0001628280-24-000857	ANGIODYNAMICS INC	2024-01-08	14 PLAZA DRIVE, LATHAM 12110, US
0001638833-24-000006	SURGERY PARTNERS, INC.	2024-01-08	310 SEVEN SPRINGS WAY, BRENTWOOD 37027, US
0001639691-24-000006	LIVANOVA PLC	2024-01-08	20 EASTBOURNE TERRACE, LONDON W2 6LG, GB
0001645113-24-000003	NOVOCURE LTD	2024-01-08	NO. 4 THE FORUM, ST. HELIER JE2 4UF, JE
0001759655-24-000004	PRIVIA HEALTH GROUP, INC.	2024-01-08	950 N. GLEBE RD., SUITE 700, ARLINGTON 22203, US
0001770787-24-000004	10X GENOMICS, INC.	2024-01-08	6230 STONERIDGE MALL ROAD, PLEASANTON 94588, US
0001818331-24-000003	GENEDX HOLDINGS CORP.	2024-01-08	333 LUDLOW STREET, STAMFORD 06902, US
0001965040-24-000002	FORTREA HOLDINGS INC.	2024-01-08	8 MOORE DRIVE, DURHAM 27709, US
0000818479-24-000004	DENTSPLY SIRONA INC.	2024-01-09	13320 BALLANTYNE CORPORATE PLACE, CHARLOTTE 28277-3607, US
0000885978-24-000002	U S PHYSICAL THERAPY INC /NV	2024-01-09	1300 WEST SAM HOUSTON PARKWAY, HOUSTON 77043, US
0000950170-24-003196	NEVRO CORP	2024-01-09	1800 BRIDGE PARKWAY, REDWOOD CITY 94065, US
0000950170-24-003420	ORTHOFIX MEDICAL INC.	2024-01-09	3451 PLANO PARKWAY, LEWISVILLE 75056, US
0001035267-24-000006	INTUITIVE SURGICAL INC	2024-01-09	1020 KIFER ROAD, SUNNYVALE 94086, US
0001079973-24-000043	OPGEN INC	2024-01-09	9717 KEY WEST AVENUE, ROCKVILLE 20850, US
0001104659-24-002451	NATERA, INC.	2024-01-09	13011 MCCALLEN PASS, AUSTIN 78753, US
0001110803-24-000002	ILLUMINA, INC.	2024-01-09	5200 ILLUMINA WAY, SAN DIEGO 92122, US
0001110803-24-000004	ILLUMINA, INC.	2024-01-09	5200 ILLUMINA WAY, SAN DIEGO 92122, US
0001140361-24-001564	NUWELLIS, INC.	2024-01-09	12988 VALLEY VIEW ROAD, EDEN PRAIRIE 55344, US

Submission ID	Name	Submission Date	Mailing Address
0001157523-24-000035	REVVITY, INC.	2024-01-09	940 WINTER STREET, WALTHAM 02451, US
0001193125-24-004805	CYTEK BIOSCIENCES, INC.	2024-01-09	47215 LAKEVIEW BOULEVARD, FREMONT 94538, US
0001193125-24-004832	WATERS CORP /DE/	2024-01-09	34 MAPLE STREET, MILFORD 01757, US
0001193125-24-004834	CAREDX, INC.	2024-01-09	8000 MARINA BLVD, BRISBANE 94005, US
0001213900-24-001983	NORTHVIEW ACQUISITION CORP	2024-01-09	207 WEST 25TH ST, 9TH FLOOR, NEW YORK 10001, US
0001213900-24-002254	NOCTURNE ACQUISITION CORP	2024-01-09	3 GERMAY DRIVE, WILMINGTON 19804, US
0001213900-24-002281	SPECTAIRE HOLDINGS INC.	2024-01-09	3109 W 50TH ST., #207, MINNEAPOLIS 55410, US
0001213900-24-002307	FRESH2 GROUP LTD	2024-01-09	650 FIFTH AVENUE, NEW YORK 10019-6108, US
0001332349-24-000004	BROOKDALE SENIOR LIVING INC.	2024-01-09	111 WESTWOOD PLACE, BRENTWOOD 37027, US
0001397702-24-000002	SILK ROAD MEDICAL INC	2024-01-09	1213 INNSBRUCK DR., SUNNYVALE 94089-2918, US
0001437749-24-000939	PULSE BIOSCIENCES, INC.	2024-01-09	3957 POINT EDEN WAY, HAYWARD 94545, US
0001531048-24-000003	INARI MEDICAL, INC.	2024-01-09	6001 OAK CANYON, SUITE 100, IRVINE 92618, US
0001558370-24-000141	GLAUKOS CORP	2024-01-09	229 AVENIDA FABRICANTE, SAN CLEMENTE 92672, US
0001558370-24-000152	MOVING IMAGE TECHNOLOGIES INC.	2024-01-09	17760 NEWHOPE STREET, FOUNTAIN VALLEY 92708, US
0001628280-24-000890	CUTERA INC	2024-01-09	3240 BAYSHORE BOULEVARD, BRISBANE 94005, US
0001645113-24-000004	NOVOCURE LTD	2024-01-09	NO. 4 THE FORUM, ST. HELIER JE2 4UF, JE
0001654954-24-000372	AEHR TEST SYSTEMS	2024-01-09	400 KATO TERRACE, FREMONT 94539, US
0001688757-24-000005	ESTABLISHMENT LABS HOLDINGS INC.	2024-01-09	PO BOX 3140, COMMERCE HOUSE, TORTOLA VG1110, VG
0001753926-24-000076	OCEANTECH ACQUISITIONS I CORP.	2024-01-09	515 MADISON AVE., SUITE 8133, NEW YORK 10022, US
0001822479-24-000003	SOTERA HEALTH CO	2024-01-09	9100 SOUTH HILLS BLVD, SUITE 300, BROADVIEW HEIGHTS 44147, US
0000937556-24-000004	MASIMO CORP	2024-01-10	52 DISCOVERY, IRVINE 92618, US
0000950170-24-003448	TREACE MEDICAL CONCEPTS, INC.	2024-01-10	100 PALMETTO PARK PLACE, PONTE VEDRA 32081, US
0000950170-24-003512	EYEPOINT PHARMACEUTICALS, INC.	2024-01-10	480 PLEASANT STREET, WATERTOWN 02472, US

Submission ID	Name	Submission Date	Mailing Address
0000950170-24-003514	RENALYTIX PLC	2024-01-10	FINSGATE, LONDON EC1V 9EE, GB
0000950170-24-003621	AVITA MEDICAL, INC.	2024-01-10	28159 AVENUE STANFORD, VALENCIA 91355, US
0001037646-24-000004	METTLER TOLEDO INTERNATIONAL INC/	2024-01-10	1900 POLARIS PARKWAY, COLUMBUS 43240, US
0001104659-24-002962	AKOYA BIOSCIENCES, INC.	2024-01-10	100 CAMPUS DRIVE, MARLBOROUGH 01762, US
0001140361-24-001728	SANUWAVE HEALTH, INC.	2024-01-10	3360 MARTIN FARM RD, SUWANEE 30024, US
0001171843-24-000199	INTEGER HOLDINGS CORP	2024-01-10	5830 GRANITE PARKWAY., SUITE 1150, PLANO 75024, US
0001185185-24-000060	MITESCO, INC.	2024-01-10	505 BEACHLAND BLVD., SUITE 1377, VERO BEACH 32963, US
0001213900-24-002712	TENON MEDICAL, INC.	2024-01-10	104 COOPER CT., LOS GATOS 95032, US
0001380106-24-000009	RAPID MICRO BIOSYSTEMS, INC.	2024-01-10	1001 PAWTUCKET BLVD., LOWELL 01854, US
0001437749-24-001024	THERMOGENESIS HOLDINGS, INC.	2024-01-10	2711 CITRUS ROAD, RANCHO CORDOVA 95742, US
0001437749-24-001079	THERMOGENESIS HOLDINGS, INC.	2024-01-10	2711 CITRUS ROAD, RANCHO CORDOVA 95742, US
0001466258-24-000007	TRANE TECHNOLOGIES PLC	2024-01-10	C/O TRANE TECHNOLOGIES, DAVIDSON 28036, US
0001493152-24-001892	BONE BIOLOGICS CORP	2024-01-10	2 BURLINGTON WOODS DRIVE,, BURLINGTON 01803, US
0001493152-24-001940	INVO BIOSCIENCE, INC.	2024-01-10	5582 BROADCAST COURT, SARASOTA 34240, US
0001493152-24-001962	INVO BIOSCIENCE, INC.	2024-01-10	5582 BROADCAST COURT, SARASOTA 34240, US
0001558370-24-000182	HURCO COMPANIES INC	2024-01-10	ONE TECHNOLOGY WAY, INDIANAPOLIS 46268, US
0001562762-24-000006	GLOBUS MEDICAL INC	2024-01-10	2560 GENERAL ARMISTEAD AVENUE, AUDUBON 19403, US
0001628280-24-000993	P3 HEALTH PARTNERS INC.	2024-01-10	2045 W GRAND AVE STE B, CHICAGO 60612-1577, US
0001654954-24-000428	AEHR TEST SYSTEMS	2024-01-10	400 KATO TERRACE, FREMONT 94539, US
0001753926-24-000087	VIVANI MEDICAL, INC.	2024-01-10	1350 S. LOOP ROAD, ALAMEDA 94502, US
0001822359-24-000005	DOCGO INC.	2024-01-10	35TH STREET WEST 35TH, NEW YORK 10001, US
0000929638-24-000084	SEASTAR MEDICAL HOLDING CORP	2024-01-11	3513 BRIGHTON BLVD, DENVER 80216, US
0000950170-24-003948	CAREMAX, INC.	2024-01-11	1000 NW 57 COURT, SUITE 400, MIAMI 33126, US

Submission ID	Name	Submission Date	Mailing Address
0001140361-24-001849	ANGIODYNAMICS INC	2024-01-11	14 PLAZA DRIVE, LATHAM 12110, US
0001193125-24-006298	ALTITUDE ACQUISITION CORP.	2024-01-11	1240 WEST WESLEY ROAD, ATLANTA 30327, US
0001213900-24-002970	HEALTHLYNKED CORP	2024-01-11	1265 CREEKSIDE PARKWAY, NAPLES 34108, US
0001213900-24-003023	NUTRIBAND INC.	2024-01-11	121 S. ORANGE AVE., ORLANDO 32801, US
0001213900-24-003047	TENON MEDICAL, INC.	2024-01-11	104 COOPER CT., LOS GATOS 95032, US
0001437749-24-001197	PERSPECTIVE THERAPEUTICS, INC.	2024-01-11	2401 ELLIOTT AVENUE, SEATTLE 98121, US
0001437749-24-001204	ARCH THERAPEUTICS, INC.	2024-01-11	235 WALNUT STREET, SUITE 6, FRAMINGHAM 01702, US
0001493152-24-002010	BONE BIOLOGICS CORP	2024-01-11	2 BURLINGTON WOODS DRIVE,, BURLINGTON 01803, US
0001493152-24-002022	IMAC HOLDINGS, INC.	2024-01-11	1605 WESTGATE CIRCLE, BRENTWOOD 37027, US
0001493152-24-002048	SIGYN THERAPEUTICS, INC.	2024-01-11	9190 W OLYMPIC BLVD # 263, BEVERLY HILLS 90212, US
0001493152-24-002097	GLUCOTRACK, INC.	2024-01-11	301 RT 17 NORTH, RUTHERFORD 07070, US
0001558370-24-000191	QUANTERIX CORP	2024-01-11	900 MIDDLESEX TURNPIKE, BILLERICA 01821, US
0001628280-24-001153	ACUTUS MEDICAL, INC.	2024-01-11	2210 FARADAY AVE, CARLSBAD 92008, US
0001654954-24-000467	CATHETER PRECISION, INC.	2024-01-11	1670 HIGHWAY 160 WEST, FORT MILL 29708, US
0001683168-24-000184	FOCUS UNIVERSAL INC.	2024-01-11	2311 EAST LOCUST STREET, ONTARIO 91761, US
0000920148-24-000004	LABORATORY CORP OF AMERICA HOLDINGS	2024-01-12	358 S MAIN ST, BURLINGTON 27215, US
0000929638-24-000095	SEASTAR MEDICAL HOLDING CORP	2024-01-12	3513 BRIGHTON BLVD, DENVER 80216, US
0000950170-24-004263	ONTO INNOVATION INC.	2024-01-12	16 JONSPIN ROAD, WILMINGTON 01887, US
0001104659-24-003645	HURCO COMPANIES INC	2024-01-12	ONE TECHNOLOGY WAY, INDIANAPOLIS 46268, US
0001104659-24-003755	ASSURE HOLDINGS CORP.	2024-01-12	7887 EAST BELLEVIEW AVENUE, GREENWOOD VILLAGE 80111, US
0001157523-24-000058	HOLOGIC INC	2024-01-12	250 CAMPUS DRIVE, MARLBOROUGH 01752, US
0001193125-24-007127	TERADYNE, INC	2024-01-12	600 RIVERPARK DRIVE, NORTH READING 01864, US
0001193125-24-007527	ROCKWELL AUTOMATION, INC	2024-01-12	1201 SOUTH 2ND STREET, MILWAUKEE 53204, US

Submission ID	Name	Submission Date	Mailing Address
0001193125-24-007640	BIOVENTUS INC.	2024-01-12	4721 EMPEROR BOULEVARD, SUITE 100, DURHAM 27703, US
0001193125-24-007674	INOVIO PHARMACEUTICALS, INC.	2024-01-12	6769 MESA RIDGE RD., SAN DIEGO 92121, US
0001213900-24-003096	BLUERIVER ACQUISITION CORP.	2024-01-12	250 WEST NOTTINGHAM DRIVE, SAN ANTONIO 78209, US
0001437749-24-001286	ARCH THERAPEUTICS, INC.	2024-01-12	235 WALNUT STREET, SUITE 6, FRAMINGHAM 01702, US
0001437749-24-001289	MILESTONE SCIENTIFIC INC.	2024-01-12	220 SOUTH ORANGE AVENUE, LIVINGSTON 07039, US
0001493152-24-002201	MICROBOT MEDICAL INC.	2024-01-12	25 RECREATION PARK DR SUITE 108, HINGHAM 02043, US
0001493152-24-002206	AURORA TECHNOLOGY ACQUISITION CORP.	2024-01-12	4 EMBARCADERO CENTER SUITE 1449, SAN FRANCISCO 94105, US
0001493152-24-002233	BONE BIOLOGICS CORP	2024-01-12	2 BURLINGTON WOODS DRIVE,, BURLINGTON 01803, US
0001493152-24-002266	BIOSIG TECHNOLOGIES, INC.	2024-01-12	55 GREENS FARMS ROAD, WESTPORT 06880, US
0001501134-24-000003	INVITAE CORP	2024-01-12	1400 16TH STREET, SAN FRANCISCO 94103, US
0001654954-24-000525	AEHR TEST SYSTEMS	2024-01-12	400 KATO TERRACE, FREMONT 94539, US
0000066740-24-000002	3M CO	2024-01-16	3M CENTER, ST. PAUL 55144-1000, US
0000917491-24-000001	FARO TECHNOLOGIES INC	2024-01-16	FARO TECHNOLOGIES INC, LAKE MARY 32746, US
0000920148-24-000006	LABORATORY CORP OF AMERICA HOLDINGS	2024-01-16	358 S MAIN ST, BURLINGTON 27215, US
0000937556-24-000006	MASIMO CORP	2024-01-16	52 DISCOVERY, IRVINE 92618, US
0001037868-24-000002	AMETEK INC/	2024-01-16	1100 CASSATT ROAD, BERWYN 19312, US
0001062993-24-000849	DYNATRONICS CORP	2024-01-16	7030 PARK CENTER DR, SALT LAKE CITY 84121, US
0001062993-24-000888	SALONA GLOBAL MEDICAL DEVICE CORP	2024-01-16	6160 INNOVATION WAY, CARLSBAD 92009, US
0001140361-24-002200	NORTHVIEW ACQUISITION CORP	2024-01-16	207 WEST 25TH ST, 9TH FLOOR, NEW YORK 10001, US
0001193805-24-000047	ELECTROCORE, INC.	2024-01-16	200 FORGE WAY, ROCKAWAY 07866, US
0001213900-24-003737	SIENTRA, INC.	2024-01-16	3333 MICHAELSON DR, IRVINE 92612, US
0001437749-24-001387	INTEST CORP	2024-01-16	804 EAST GATE DR., SUITE 200, MT. LAUREL 08054, US
0001493152-24-002270	CLEARDAY, INC.	2024-01-16	15511 W. STATE HWY 71, AUSTIN 78738, US

Submission ID	Name	Submission Date	Mailing Address
0001493152-24-002447	NEXGEL, INC.	2024-01-16	2150 CABOT BLVD WEST,, LANGHORNE 19047, US
0001493152-24-002505	INVO BIOSCIENCE, INC.	2024-01-16	5582 BROADCAST COURT, SARASOTA 34240, US
0001683168-24-000307	FOCUS UNIVERSAL INC.	2024-01-16	2311 EAST LOCUST STREET, ONTARIO 91761, US
0001770787-24-000007	10X GENOMICS, INC.	2024-01-16	6230 STONERIDGE MALL ROAD, PLEASANTON 94588, US
0001822359-24-000007	DOCGO INC.	2024-01-16	35TH STREET WEST 35TH, NEW YORK 10001, US
0000950170-24-004758	CARMELL CORP	2024-01-17	1177 AVENUE OF THE AMERICAS, NEW YORK 10036, US
0000950170-24-004762	CARMELL CORP	2024-01-17	1177 AVENUE OF THE AMERICAS, NEW YORK 10036, US
0000950170-24-004775	CARMELL CORP	2024-01-17	1177 AVENUE OF THE AMERICAS, NEW YORK 10036, US
0000950170-24-004782	CARMELL CORP	2024-01-17	1177 AVENUE OF THE AMERICAS, NEW YORK 10036, US
0000950170-24-004783	CARMELL CORP	2024-01-17	1177 AVENUE OF THE AMERICAS, NEW YORK 10036, US
0000950170-24-004786	CARMELL CORP	2024-01-17	1177 AVENUE OF THE AMERICAS, NEW YORK 10036, US
0000950170-24-004791	CARMELL CORP	2024-01-17	1177 AVENUE OF THE AMERICAS, NEW YORK 10036, US
0000950170-24-004820	AVEANNA HEALTHCARE HOLDINGS, INC.	2024-01-17	400 INTERSTATE NORTH PARKWAY, ATLANTA 30339, US
0000950170-24-004821	RENALYTIX PLC	2024-01-17	FINSGATE, LONDON EC1V 9EE, GB
0000950170-24-005016	MARPAI, INC.	2024-01-17	615 CHANNELSIDE DRIVE, TAMPA 33602, US
0000950170-24-005068	ORTHOFIX MEDICAL INC.	2024-01-17	3451 PLANO PARKWAY, LEWISVILLE 75056, US
0001097149-24-000003	ALIGN TECHNOLOGY INC	2024-01-17	410 NORTH SCOTTSDALE ROAD, SUITE 1300, TEMPE 85288, US
0001104659-24-004548	CYTOSORBENTS CORP	2024-01-17	305 COLLEGE ROAD EAST, PRINCETON 08540, US
0001193125-24-008759	GIGCAPITAL5, INC.	2024-01-17	1731 EMBARCADERO ROAD SUITE 200, PALO ALTO 94303, US
0001193125-24-009264	MYOMO, INC.	2024-01-17	137 PORTLAND STREET, BOSTON 02114, US

Submission ID	Name	Submission Date	Mailing Address
0001213900-24-004149	ENVOY MEDICAL, INC.	2024-01-17	4875 WHITE BEAR PARKWAY, WHITE BEAR LAKE 55110, US
0001437749-24-001500	PERSPECTIVE THERAPEUTICS, INC.	2024-01-17	2401 ELLIOTT AVENUE, SEATTLE 98121, US
0001437749-24-001501	PERSPECTIVE THERAPEUTICS, INC.	2024-01-17	2401 ELLIOTT AVENUE, SEATTLE 98121, US
0001437749-24-001502	PERSPECTIVE THERAPEUTICS, INC.	2024-01-17	2401 ELLIOTT AVENUE, SEATTLE 98121, US
0001437749-24-001533	PERSPECTIVE THERAPEUTICS, INC.	2024-01-17	2401 ELLIOTT AVENUE, SEATTLE 98121, US
0001493152-24-002728	NANOVIBRONIX, INC.	2024-01-17	525 EXECUTIVE BLVD, ELMSFORD 10523, US
0001493152-24-002746	SYNTEC OPTICS HOLDINGS, INC.	2024-01-17	1111 LINCOLN ROAD, MIAMI BEACH 33139, US
0001493152-24-002790	THERAPEUTIC SOLUTIONS INTERNATIONAL, INC.	2024-01-17	701 WILD ROSE LANE, ELK CITY 83525, US
0001628280-24-001503	ACUTUS MEDICAL, INC.	2024-01-17	2210 FARADAY AVE, CARLSBAD 92008, US
0001654954-24-000690	LAKELAND INDUSTRIES INC	2024-01-17	1525 PERIMETER PARKWAY, SUITE 325, HUNTSVILLE 35806, US
0000950157-24-000061	ARTIVION, INC.	2024-01-18	1655 ROBERTS BOULEVARD N W, KENNESAW 30144, US
0000950170-24-005186	HEART TEST LABORATORIES, INC.	2024-01-18	550 RESERVE ST, SUITE 360, SOUTHLAKE 76092, US
0000950170-24-005385	MARPAI, INC.	2024-01-18	615 CHANNELSIDE DRIVE, TAMPA 33602, US
0001140361-24-002619	NUWELLIS, INC.	2024-01-18	12988 VALLEY VIEW ROAD, EDEN PRAIRIE 55344, US
0001140361-24-002644	U.S. NEUROSURGICAL HOLDINGS, INC.	2024-01-18	2400 RESEARCH BLVD, ROCKVILLE 20850, US
0001140361-24-002733	HOLOGIC INC	2024-01-18	250 CAMPUS DRIVE, MARLBOROUGH 01752, US
0001145197-24-000004	INSULET CORP	2024-01-18	100 NAGOG PARK, ACTON 01720, US
0001185185-24-000095	MITESCO, INC.	2024-01-18	505 BEACHLAND BLVD., SUITE 1377, VERO BEACH 32963, US
0001213900-24-004358	NOCTURNE ACQUISITION CORP	2024-01-18	3 GERMAY DRIVE, WILMINGTON 19804, US
0001213900-24-004490	NORTHVIEW ACQUISITION CORP	2024-01-18	207 WEST 25TH ST, 9TH FLOOR, NEW YORK 10001, US
0001213900-24-004491	SENSUS HEALTHCARE, INC.	2024-01-18	851 BROKEN SOUND PARKWAY NW, BOCA RATON 33487, US
0001213900-24-004577	ENVOY MEDICAL, INC.	2024-01-18	4875 WHITE BEAR PARKWAY, WHITE BEAR LAKE 55110, US
0001493152-24-002794	PONO CAPITAL TWO, INC.	2024-01-18	4348 WAIALAE AVE., #632, HONOLULU 96816, US

Submission ID	Name	Submission Date	Mailing Address
0001493152-24-002804	RETINALGENIX TECHNOLOGIES INC.	2024-01-18	1450 NORTH MCDOWELL BOULEVARD, PETALUMA 94954, US
0001493152-24-002827	MICROBOT MEDICAL INC.	2024-01-18	25 RECREATION PARK DR SUITE 108, HINGHAM 02043, US
0001493152-24-002828	AURORA TECHNOLOGY ACQUISITION CORP.	2024-01-18	4 EMBARCADERO CENTER SUITE 1449, SAN FRANCISCO 94105, US
0001493152-24-002848	SHARPS TECHNOLOGY INC.	2024-01-18	105 MAXESS ROAD, MELVILLE 11747, US
0001501134-24-000005	INVITAE CORP	2024-01-18	1400 16TH STREET, SAN FRANCISCO 94103, US
0001558370-24-000365	RESHAPE LIFESCIENCES INC.	2024-01-18	1001 CALLE AMANECER, SAN CLEMENTE 92673, US
0001645113-24-000005	NOVOCURE LTD	2024-01-18	NO. 4 THE FORUM, ST. HELIER JE2 4UF, JE
0001770141-24-000003	UPHEALTH, INC.	2024-01-18	14000 S. MILITARY TRAIL, DELRAY BEACH 33484, US
0000950170-24-005565	BRUKER CORP	2024-01-19	40 MANNING RD, BILLERICA 01821, US
0000950170-24-005804	SINGULAR GENOMICS SYSTEMS, INC.	2024-01-19	3010 SCIENCE PARK ROAD, SAN DIEGO 92121, US
0001104659-24-005246	STANDARD BIOTOOLS INC.	2024-01-19	2 TOWER PLACE, SOUTH SAN FRANCISCO 94080, US
0001140361-24-002904	VENUS CONCEPT INC.	2024-01-19	235 YORKLAND BLVD., TORONTO M2J 4Y8, CA
0001140361-24-002905	KEYSIGHT TECHNOLOGIES, INC.	2024-01-19	1400 FOUNTAINGROVE PARKWAY, SANTA ROSA 95403, US
0001193125-24-010779	BIOVENTUS INC.	2024-01-19	4721 EMPEROR BOULEVARD, SUITE 100, DURHAM 27703, US
0001193125-24-011225	ACADIA HEALTHCARE COMPANY, INC.	2024-01-19	6100 TOWER CIRCLE, FRANKLIN 37067, US
0001193125-24-011355	VIRIDIAN THERAPEUTICS, INC.DE	2024-01-19	221 CRESCENT STREET, WALTHAM 02453, US
0001213900-24-004669	MODULAR MEDICAL, INC.	2024-01-19	16772 WEST BERNARDO DRIVE, SAN DIEGO 92127, US
0001213900-24-004755	SPECTAIRE HOLDINGS INC.	2024-01-19	3109 W 50TH ST., #207, MINNEAPOLIS 55410, US
0001493152-24-003069	SHARPS TECHNOLOGY INC.	2024-01-19	105 MAXESS ROAD, MELVILLE 11747, US
0001753926-24-000119	OCEANTECH ACQUISITIONS I CORP.	2024-01-19	515 MADISON AVE., SUITE 8133, NEW YORK 10022, US
0000719135-24-000005	APYX MEDICAL CORP	2024-01-22	5115 ULMERTON ROAD, CLEARWATER 33760, US
0001104659-24-005384	NATERA, INC.	2024-01-22	13011 MCCALLEN PASS, AUSTIN 78753, US

Submission ID	Name	Submission Date	Mailing Address
0001104659-24-005618	GREENBROOK TMS INC.	2024-01-22	890 YONGE STREET, 7TH FLOOR, TORONTO M4W 3P4, CA
0001140361-24-002997	ANGIODYNAMICS INC	2024-01-22	14 PLAZA DRIVE, LATHAM 12110, US
0001193125-24-011707	MIMEDX GROUP, INC.	2024-01-22	1775 W OAK COMMONS COURT, NE, MARIETTA 30062, US
0001193125-24-012171	MKS INSTRUMENTS INC	2024-01-22	2 TECH DRIVE, ANDOVER 01810, US
0001213900-24-004961	NORTHVIEW ACQUISITION CORP	2024-01-22	207 WEST 25TH ST, 9TH FLOOR, NEW YORK 10001, US
0001213900-24-005356	ACHARI VENTURES HOLDINGS CORP. I	2024-01-22	60 WALNUT AVENUE, SUITE 400, CLARK 07066, US
0001437749-24-001852	PERSPECTIVE THERAPEUTICS, INC.	2024-01-22	2401 ELLIOTT AVENUE, SEATTLE 98121, US
0001493152-24-003170	THERAPEUTIC SOLUTIONS INTERNATIONAL, INC.	2024-01-22	701 WILD ROSE LANE, ELK CITY 83525, US
0001493152-24-003174	AGAPE ATP CORP	2024-01-22	1705-1708 LEVEL 17, TOWER 2, FABER TOWER, KUALA LUMPUR 58100, MY
0001493152-24-003188	MOTUS GI HOLDINGS, INC.	2024-01-22	1301 EAST BROWARD BOULEVARD, FT. LAUDERDALE 33301, US
0001493152-24-003322	MICROBOT MEDICAL INC.	2024-01-22	25 RECREATION PARK DR SUITE 108, HINGHAM 02043, US
0001493152-24-003331	BRUUSH ORAL CARE INC.	2024-01-22	128 WEST HASTINGS STREET, UNIT 210, VANCOUVER V6B 1G8, CA
0001501134-24-000007	INVITAE CORP	2024-01-22	1400 16TH STREET, SAN FRANCISCO 94103, US
0001554859-24-000003	SEMLER SCIENTIFIC, INC.	2024-01-22	2340-2348 WALSH AVENUE, SUITE 2344, SANTA CLARA 95051, US
0001654954-24-000793	SCIENTIFIC INDUSTRIES INC	2024-01-22	70 ORVILLE DR, BOHEMIA 11716, US
0001683168-24-000371	ACLARION, INC.	2024-01-22	8181 ARISTA PLACE, BROOMFIELD 80021, US
0000066740-24-000005	3M CO	2024-01-23	3M CENTER, ST. PAUL 55144-1000, US
0001035267-24-000009	INTUITIVE SURGICAL INC	2024-01-23	1020 KIFER ROAD, SUNNYVALE 94086, US
0001062993-24-001181	SALONA GLOBAL MEDICAL DEVICE CORP	2024-01-23	6160 INNOVATION WAY, CARLSBAD 92009, US
0001104659-24-005943	DIGITAL HEALTH ACQUISITION CORP.	2024-01-23	980 N FEDERAL HWY, BOCA RATON 33432, US
0001193125-24-012979	SIGHT SCIENCES, INC.	2024-01-23	4040 CAMPBELL AVE,, MENLO PARK 94025, US
0001213900-24-005420	ETAO INTERNATIONAL CO., LTD.	2024-01-23	C/O MAPLES CORPORATE SERVICES LIMITED, GRAND

Submission ID	Name	Submission Date	Mailing Address
			CAYMAN KY1-1104, KY
0001213900-24-005801	SPECTRAL AI, INC.	2024-01-23	2515 MCKINNEY AVE #1000, DALLAS 75201, US
0001493152-24-003338	MANGOCEUTICALS, INC.	2024-01-23	15110 DALLAS PKWY, SUITE 600, DALLAS 75248, US
0001493152-24-003385	SINTX TECHNOLOGIES, INC.	2024-01-23	1885 WEST 2100 STREET, SALT LAKE CITY 84119, US
0001493152-24-003491	STRONG GLOBAL ENTERTAINMENT, INC.	2024-01-23	5960 FAIRVIEW ROAD, SUITE 275, CHARLOTTE 28210, US
0001493152-24-003494	FG GROUP HOLDINGS INC.	2024-01-23	4201 CONGRESS STREET, CHARLOTTE 28209, US
0001683168-24-000377	ACLARION, INC.	2024-01-23	8181 ARISTA PLACE, BROOMFIELD 80021, US
0000929638-24-000189	SEASTAR MEDICAL HOLDING CORP	2024-01-24	3513 BRIGHTON BLVD, DENVER 80216, US
0000950170-24-006969	INOGEN INC	2024-01-24	859 WARD DRIVE, GOLETA 93111, US
0001140361-24-003589	VENUS CONCEPT INC.	2024-01-24	235 YORKLAND BLVD., TORONTO M2J 4Y8, CA
0001193125-24-014223	RESMED INC	2024-01-24	9001 SPECTRUM CENTER BLVD., SAN DIEGO 92123, US
0001437749-24-002001	SYPRIS SOLUTIONS INC	2024-01-24	101 BULLITT LN, LOUISVILLE 40222, US
0001493152-24-003550	OPTEX SYSTEMS HOLDINGS INC	2024-01-24	1420 PRESIDENTIAL DRIVE, RICHARDSON 75081, US
0001501134-24-000010	INVITAE CORP	2024-01-24	1400 16TH STREET, SAN FRANCISCO 94103, US
0001753926-24-000142	OCEANTECH ACQUISITIONS I CORP.	2024-01-24	515 MADISON AVE., SUITE 8133, NEW YORK 10022, US
0000319201-24-000003	KLA CORP	2024-01-25	ONE TECHNOLOGY DRIVE, MILPITAS 95035, US
0000707388-24-000002	STAR EQUITY HOLDINGS, INC.	2024-01-25	53 FOREST AVE, OLD GREENWICH 06870, US
0000943819-24-000003	RESMED INC	2024-01-25	9001 SPECTRUM CENTER BLVD., SAN DIEGO 92123, US
0000950170-24-007222	HEART TEST LABORATORIES, INC.	2024-01-25	550 RESERVE ST, SUITE 360, SOUTHLAKE 76092, US
0001104659-24-006362	ACCELERATE DIAGNOSTICS, INC	2024-01-25	3950 S. COUNTRY CLUB ROAD #470, TUCSON 85714, US
0001104659-24-006675	CVRX, INC.	2024-01-25	9201 WEST BROADWAY AVENUE, MINNEAPOLIS 55445, US
0001104659-24-006732	ASSURE HOLDINGS CORP.	2024-01-25	7887 EAST BELLEVIEW AVENUE, GREENWOOD VILLAGE 80111, US
0001104659-24-006749	TRISALUS LIFE SCIENCES, INC.	2024-01-25	6272 WEST 91ST AVENUE, WESTMINSTER 80031, US

Submission ID	Name	Submission Date	Mailing Address
0001140361-24-003890	SANUWAVE HEALTH, INC.	2024-01-25	3360 MARTIN FARM RD, SUWANEE 30024, US
0001193125-24-014842	INSULET CORP	2024-01-25	100 NAGOG PARK, ACTON 01720, US
0001193125-24-015380	INOVIO PHARMACEUTICALS, INC.	2024-01-25	6769 MESA RIDGE RD., SAN DIEGO 92121, US
0001193805-24-000095	ICAD INC	2024-01-25	98 SPIT BROOK ROAD, SUITE 100, NASHUA 03062, US
0001213900-24-006290	SS INNOVATIONS INTERNATIONAL, INC.	2024-01-25	1600 SE 15TH STREET, FORT LAUDERDALE 33316, US
0001213900-24-006496	NUTRIBAND INC.	2024-01-25	121 S. ORANGE AVE., ORLANDO 32801, US
0001493152-24-003755	PETVIVO HOLDINGS, INC.	2024-01-25	5251 EDINA INDUSTRIAL BLVD, EDINA 55439, US
0001493152-24-003785	MANGOCEUTICALS, INC.	2024-01-25	15110 DALLAS PKWY, SUITE 600, DALLAS 75248, US
0001493152-24-003809	PRESSURE BIOSCIENCES INC	2024-01-25	14 NORFOLK AVENUE, SOUTH EASTON 02375, US
0001628945-24-000009	CUE HEALTH INC.	2024-01-25	4980 CARROLL CANYON ROAD, SAN DIEGO 92121, US
0001654954-24-000932	CATHETER PRECISION, INC.	2024-01-25	1670 HIGHWAY 160 WEST, FORT MILL 29708, US
0000010795-24-000006	BECTON DICKINSON & CO	2024-01-26	ONE BECTON DR, FRANKLIN LAKE 07417, US
0000319201-24-000006	KLA CORP	2024-01-26	ONE TECHNOLOGY DRIVE, MILPITAS 95035, US
0000950170-24-007968	TIVIC HEALTH SYSTEMS, INC.	2024-01-26	25821 INDUSTRIAL BLVD., SUITE 100, HAYWARD 94545, US
0000950170-24-007975	TALIS BIOMEDICAL CORP	2024-01-26	1100 ISLAND DRIVE, REDWOOD CITY 94065, US
0001157523-24-000120	REVVITY, INC.	2024-01-26	940 WINTER STREET, WALTHAM 02451, US
0001157523-24-000122	BADGER METER INC	2024-01-26	4545 W BROWN DEER RD, MILWAUKEE 53223, US
0001171843-24-000431	PSYCHEMEDICS CORP	2024-01-26	125 NAGOG PARK, ACTON 01720, US
0001213900-24-006717	SS INNOVATIONS INTERNATIONAL, INC.	2024-01-26	1600 SE 15TH STREET, FORT LAUDERDALE 33316, US
0001213900-24-006969	LOGICMARK, INC.	2024-01-26	2801 DIODE LANE, LOUISVILLE 40299, US
0001437749-24-002304	AVINGER INC	2024-01-26	400 CHESAPEAKE DRIVE, REDWOOD CITY 94063, US
0001493152-24-003973	INTELLIGENT BIO SOLUTIONS INC.	2024-01-26	142 WEST 57TH ST FL11, NEW YORK 10019, US
0001609550-24-000008	INSPIRE MEDICAL SYSTEMS, INC.	2024-01-26	5500 WAYZATA BLVD, GOLDEN VALLEY 55416, US

Submission ID	Name	Submission Date	Mailing Address
0000066740-24-000007	3M CO	2024-01-29	3M CENTER, ST. PAUL 55144-1000, US
0000950170-24-008220	SUNLINK HEALTH SYSTEMS INC	2024-01-29	900 CIRCLE 75 PARKWAY, ATLANTA 30339, US
0000950170-24-008384	CAREMAX, INC.	2024-01-29	1000 NW 57TH COURT, SUITE 400, MIAMI 33126, US
0001104659-24-007544	HURCO COMPANIES INC	2024-01-29	ONE TECHNOLOGY WAY, INDIANAPOLIS 46268, US
0001104659-24-007734	GREENBROOK TMS INC.	2024-01-29	890 YONGE STREET, 7TH FLOOR, TORONTO M4W 3P4, CA
0001140361-24-004271	KEYSIGHT TECHNOLOGIES, INC.	2024-01-29	1400 FOUNTAINGROVE PARKWAY, SANTA ROSA 95403, US
0001193125-24-018104	TRANSMEDICS GROUP, INC.	2024-01-29	200 MINUTEMAN ROAD, ANDOVER 01810, US
0001193125-24-018115	TERADYNE, INC	2024-01-29	600 RIVERPARK DRIVE, NORTH READING 01864, US
0001193125-24-018126	TERADYNE, INC	2024-01-29	600 RIVERPARK DRIVE, NORTH READING 01864, US
0001193125-24-018193	GIGCAPITAL5, INC.	2024-01-29	1731 EMBARCADERO ROAD SUITE 200, PALO ALTO 94303, US
0001193125-24-018256	BIOLASE, INC	2024-01-29	27042 TOWNE CENTRE DRIVE, FOOTHILL RANCH 92610, US
0001213900-24-007142	SIENTRA, INC.	2024-01-29	3333 MICHAELSON DR, IRVINE 92612, US
0001213900-24-007204	VICARIOUS SURGICAL INC.	2024-01-29	78 FOURTH AVENUE, WALTHAM 02451, US
0001213900-24-007478	FRESH2 GROUP LTD	2024-01-29	650 FIFTH AVENUE, NEW YORK 10019-6108, US
0001213900-24-007479	FRESH2 GROUP LTD	2024-01-29	650 FIFTH AVENUE, NEW YORK 10019-6108, US
0001437749-24-002387	COHU INC	2024-01-29	12367 CROSTHWAITE CIRCLE, POWAY 92064-6817, US
0001493152-24-004179	BRUUSH ORAL CARE INC.	2024-01-29	128 WEST HASTINGS STREET, UNIT 210, VANCOUVER V6B 1G8, CA
0001642545-24-000016	SHOCKWAVE MEDICAL, INC.	2024-01-29	5403 BETSY ROSS DRIVE, SANTA CLARA 95054, US
0001683168-24-000519	ACLARION, INC.	2024-01-29	8181 ARISTA PLACE, BROOMFIELD 80021, US
0001683168-24-000525	WELSIS CORP.	2024-01-29	BULEVAR MIHAJLA PUPINA 115, BELGRADE 11070, RS
0001759655-24-000011	PRIVIA HEALTH GROUP, INC.	2024-01-29	950 N. GLEBE RD., SUITE 700, ARLINGTON 22203, US
0000310764-24-000005	STRYKER CORP	2024-01-30	2825 AIRVIEW BLVD, KALAMAZOO 49002, US

Submission ID	Name	Submission Date	Mailing Address
0000313616-24-000006	DANAHER CORP /DE/	2024-01-30	2200 PENNSYLVANIA AVE. N.W., WASHINGTON 20037-1701, US
0000864749-24-000008	TRIMBLE INC.	2024-01-30	10368 WESTMOOR DR, WESTMINSTER 80021, US
0000929638-24-000283	SEASTAR MEDICAL HOLDING CORP	2024-01-30	3513 BRIGHTON BLVD, DENVER 80216, US
0001096906-24-000127	UTAH MEDICAL PRODUCTS INC	2024-01-30	None
0001104659-24-008229	QUOIN PHARMACEUTICALS, LTD.	2024-01-30	23 HATA'AS STREET, KFAR SABA 44425, IL
0001104659-24-008231	QUANTERIX CORP	2024-01-30	900 MIDDLESEX TURNPIKE, BILLERICA 01821, US
0001104659-24-008352	ASSURE HOLDINGS CORP.	2024-01-30	7887 EAST BELLEVIEW AVENUE, GREENWOOD VILLAGE 80111, US
0001178913-24-000304	REWALK ROBOTICS LTD.	2024-01-30	3 HATNUFA ST., YOKNEAM ILIT 2069203, IL
0001185185-24-000124	MITESCO, INC.	2024-01-30	505 BEACHLAND BLVD., SUITE 1377, VERO BEACH 32963, US
0001193125-24-018748	HCA HEALTHCARE, INC.	2024-01-30	ONE PARK PLAZA, NASHVILLE 37203, US
0001193125-24-019500	SOLENO THERAPEUTICS INC	2024-01-30	203 REDWOOD SHORES PARKWAY, REDWOOD CITY 94065, US
0001206774-24-000062	TRANSCAT INC	2024-01-30	35 VANTAGE POINT DRIVE, ROCHESTER 14624, US
0001213900-24-007613	NOCTURNE ACQUISITION CORP	2024-01-30	3 GERMAY DRIVE, WILMINGTON 19804, US
0001437749-24-002508	TRANSCAT INC	2024-01-30	35 VANTAGE POINT DRIVE, ROCHESTER 14624, US
0001493152-24-004231	MICROBOT MEDICAL INC.	2024-01-30	25 RECREATION PARK DR SUITE 108, HINGHAM 02043, US
0001493152-24-004238	QHSLAB, INC.	2024-01-30	901 NORTHPOINT PARKWAY, WEST PALM BEACH 33407, US
0001493152-24-004245	BIOSIG TECHNOLOGIES, INC.	2024-01-30	55 GREENS FARMS ROAD, WESTPORT 06880, US
0001493152-24-004320	SIGYN THERAPEUTICS, INC.	2024-01-30	9190 W OLYMPIC BLVD # 263, BEVERLY HILLS 90212, US
0001493152-24-004345	LUCID DIAGNOSTICS INC.	2024-01-30	360 MADISON AVENUE, NEW YORK 10017, US
0001493152-24-004348	INTELLIGENT BIO SOLUTIONS INC.	2024-01-30	142 WEST 57TH ST FL11, NEW YORK 10019, US
0001493152-24-004349	BONE BIOLOGICS CORP	2024-01-30	2 BURLINGTON WOODS DRIVE,, BURLINGTON 01803, US
0001493152-24-004350	PAVMED INC.	2024-01-30	360 MADISON AVENUE, NEW YORK 10017, US
0001753926-24-000175	OCEANTECH ACQUISITIONS I CORP.	2024-01-30	515 MADISON AVE., SUITE 8133, NEW YORK 10022, US

Submission ID	Name	Submission Date	Mailing Address
0001753926-24-000178	SEP ACQUISITION CORP.	2024-01-30	3737 BUFFALO SPEEDWAY, HOUSTON 77098, US
0000097745-24-000003	THERMO FISHER SCIENTIFIC INC.	2024-01-31	168 THIRD AVENUE, WALTHAM 02451, US
0000882835-24-000002	ROPER TECHNOLOGIES INC	2024-01-31	6496 UNIVERSITY PARKWAY, SARASOTA 34240, US
0000885725-24-000011	BOSTON SCIENTIFIC CORP	2024-01-31	300 BOSTON SCIENTIFIC WAY, MARLBOROUGH 01752-1234, US
0000885725-24-000014	BOSTON SCIENTIFIC CORP	2024-01-31	300 BOSTON SCIENTIFIC WAY, MARLBOROUGH 01752-1234, US
0000950170-24-009126	TERADYNE, INC	2024-01-31	600 RIVERPARK DRIVE, NORTH READING 01864, US
0000950170-24-009160	HEART TEST LABORATORIES, INC.	2024-01-31	550 RESERVE ST, SUITE 360, SOUTHLAKE 76092, US
0000950170-24-009384	ACCURAY INC	2024-01-31	1240 DEMING WAY, MADISON 53717, US
0001024478-24-000006	ROCKWELL AUTOMATION, INC	2024-01-31	1201 SOUTH 2ND STREET, MILWAUKEE 53204, US
0001024478-24-000008	ROCKWELL AUTOMATION, INC	2024-01-31	1201 SOUTH 2ND STREET, MILWAUKEE 53204, US
0001035267-24-000021	INTUITIVE SURGICAL INC	2024-01-31	1020 KIFER ROAD, SUNNYVALE 94086, US
0001097149-24-000007	ALIGN TECHNOLOGY INC	2024-01-31	410 NORTH SCOTTSDALE ROAD, SUITE 1300, TEMPE 85288, US
0001104659-24-008555	CVRX, INC.	2024-01-31	9201 WEST BROADWAY AVENUE, MINNEAPOLIS 55445, US
0001140361-24-004664	NUWELLIS, INC.	2024-01-31	12988 VALLEY VIEW ROAD, EDEN PRAIRIE 55344, US
0001174947-24-000144	CONMED CORP	2024-01-31	11311 CONCEPT BOULEVARD, LARGO 33773, US
0001193125-24-020900	DELCATH SYSTEMS, INC.	2024-01-31	1633 BROADWAY, NEW YORK 10019, US
0001213900-24-008429	PIXIE DUST TECHNOLOGIES, INC.	2024-01-31	2-2-1 YAESU, CHUO-KU, TOKYO 104-0028, JP
0001213900-24-008465	TENON MEDICAL, INC.	2024-01-31	104 COOPER CT., LOS GATOS 95032, US
0001213900-24-008486	NOCTURNE ACQUISITION CORP	2024-01-31	3 GERMAY DRIVE, WILMINGTON 19804, US
0001437749-24-002702	TRANSCAT INC	2024-01-31	35 VANTAGE POINT DRIVE, ROCHESTER 14624, US
0001437749-24-002708	AMERICAN SHARED HOSPITAL SERVICES	2024-01-31	601 MONTGOMERY STREET, SAN FRANCISCO 94111, US
0001493152-24-004396	SINTX TECHNOLOGIES, INC.	2024-01-31	1885 WEST 2100 STREET, SALT LAKE CITY 84119, US

Submission ID	Name	Submission Date	Mailing Address
0001493152-24-004411	INSPIREMD, INC.	2024-01-31	4 MENORAT HAMAOR ST., TEL AVIV 6744832, IL
0001493152-24-004422	NEXGEL, INC.	2024-01-31	2150 CABOT BLVD WEST,, LANGHORNE 19047, US
0001493152-24-004424	NEURAXIS, INC	2024-01-31	11611 N. MERIDIAN ST, CARMEL 46032, US
0001493152-24-004491	BIOSIG TECHNOLOGIES, INC.	2024-01-31	55 GREENS FARMS ROAD, WESTPORT 06880, US
0001493152-24-004502	SIGYN THERAPEUTICS, INC.	2024-01-31	9190 W OLYMPIC BLVD # 263, BEVERLY HILLS 90212, US
0001531048-24-000009	INARI MEDICAL, INC.	2024-01-31	6001 OAK CANYON, SUITE 100, IRVINE 92618, US
0001651944-24-000005	DERMTECH, INC.	2024-01-31	12340 EL CAMINO REAL, SAN DIEGO 92130, US
0001659166-24-000023	FORTIVE CORP	2024-01-31	6920 SEAWAY BLVD., EVERETT 98203, US
0001683168-24-000570	AUTONOMIX MEDICAL, INC.	2024-01-31	21 WATERWAY AVENUE, SUITE 300, THE WOODLANDS 77380, US
0001683168-24-000576	CARDIFF LEXINGTON CORP	2024-01-31	3753 HOWARD HUGHES PARKWAY, LAS VEGAS 89169, US
0001683168-24-000585	ACLARION, INC.	2024-01-31	8181 ARISTA PLACE, BROOMFIELD 80021, US
0001688757-24-000011	ESTABLISHMENT LABS HOLDINGS INC.	2024-01-31	BUILDING B15 AND 25 COYOL FREE ZONE, ALAJUELA 20101, CR
0001822359-24-000009	DOCGO INC.	2024-01-31	35TH STREET WEST 35TH, NEW YORK 10001, US

Q2

This question asks for the following: Extract the month from the "FILED" date column. Counts the number of unique submissions (ADSH values) for each month. Group the results by month, creating a summary table that shows the total number of submissions per month.

Since we were only able to load in one month of data the query would just have the total number of submissions and not the average.

In [78]:
```
%%sql
SELECT EXTRACT(MONTH FROM FILED) as "Month",
    COUNT(ADSH) as "Total Number of Submissions"
FROM SUB
GROUP BY EXTRACT(MONTH FROM FILED);
```

```
 * postgresql://student@/SECDB
1 rows affected.
```

Out[78]:

Month	Total Number of Submissions
1.0	5904

Q3

This question asks to identify and rank the most common tags associated with companies in the Finance and Insurance sector (SIC codes 6021-6411). We filter for companies in the finance sector and ensures each has a relevant tag in the NUM table. By retrieving each tag's ID, label (name), and counting distinct companies (cik) linked to each tag, the query provides the number of unique companies associated with each tag. Grouping by tag ID and label, it then orders the results by the company count in descending order, highlighting the most frequently associated tags in this sector at the top of the list.

In [79]:

```
%%sql
SELECT
    TAG.tag as "Tag ID",
    TAG.tlabel as "Tag Label",
    COUNT(DISTINCT SUB.cik) AS "Total Companies"
FROM TAG, SUB, NUM
WHERE
    TAG.tag = NUM.tag
    AND SUB.adsh = NUM.adsh
    AND SUB.sic BETWEEN 6021 AND 6411
GROUP BY
    TAG.tag,
    TAG.tlabel
ORDER BY
    "Total Companies" DESC;
```

```
 * postgresql://student@/SECDB
2773 rows affected.
```

Out[79]:

Cas

Cas

CashCashEquivalentsRestrictedCashAndRestrictedCashE

IncomeLossFromContinuingOperat

CashCashEquivalentsRestrictedCashAndRestrictedCashE

AccumulatedDe

ShareBasedCompensationArrangementByShareBasedPayı

ShareBasedCompensationArrangementByShareBasedPaymentAwardEquityInstrumentsO

ShareBasedCompensationArrangementByShareBasedPaymentAwardEquityInstrumentsOther

ShareBasedCompensationArrangementByShareBasedPa

Effecti

ShareBasedCompensationArrangementByShareBasedPaymentAwardEquityInstrumentsOther

ShareBasedCompensationArrangementByShareBasedPa

ShareBasedCompensationArrangementByShareBasedPaymentAwardEquityInstrumentsO

ShareBasedCompensationArrangementByShareBasedPaymentAwardEquityInstrumentsOther

AdjustmentsToAdditionalPaidInCapitalSh

Effecti

ShareBasedCompensationArrangementByShareBasedPa

AdjustmentsToAdditionalPaidInCapitalSh

OtherComprehensiveIncomeUnrea

FinancingRec

EffectOfExchangeRateOnCas

ComprehensiveIncor

Financing

FinancingReceivableExcludingAccru

Finan

CashCashEquivalentsRestrictedCashAndRestrictedCas

FinancingReceivableExcludingAccr

FinancingReceivableExcludingAccru

FinancingF

FinancingReceivab

FinancingReceivableExcludingAccru

FinancingReceivableExcludingAccr

ShareBasedCompensationArrangementByShareBasedPaym

OtherComprehensiveIncomeLoss

Stockh

Stockh

S

FinancingRe

FairValueMeasurement

EmployeeServiceShareBasedCompensationNonvestedAwardsTotalCompensa

EmployeeServiceShareBasedCompensa

EmployeeServiceShareBasedCompensa

OtherComprehensiveIncom

FairValueMeasurementWithU

ShareBasedCompensationArrangementByShareBasedPaymentAwardE
FairValueMeasurementWithUnobservableInput
ShareBasedCompensationArrang
FairValueInvestmentsEntiti
OtherComprehensiveIncomeLossP

ShareBasedCompensationArrang

Inc

Ir

IncomeLossFromContinuingOperationsBeforeIncomeTaxe

Long

Fir

IncomeTaxReconcili

IncomeTaxReconcili

IncomeTaxRec

F

Ne

EffectOfExchangeRateOnCashCashEquivalentsRestrictedCashAndRestrictedCas

FinancingReceivableExcludingAccr

Fin

In

FinancingReceivableExcludingAccruedIntere

Incre

ReclassificationFrom

EffectiveIncomeTaxR

FinancingReceivableExcludingAccruedInter

FinancingReceivableExcludingAccruedIntere

FinancingReceivableExclu

FinancingReceivableExcludingAccruedIntere

FinancingReceivableExcludingAccruedInter

Effecti

FinancingReceivableModificat

De

Business

Incremental
AvailableF
AvailableF
AvailableForSaleS
AvailableFo
FinancingReceivableUr
AvailableForSaleS
AvailableFor
Availa
DebtSecuritiesAvaila

DebtSecuritiesAvailableForSaleContin

DebtSecuritiesAvaila

DebtSecuritiesAvailableForSaleContin

Availa

DebtSe

OtherComprehensiveIncomeForeignCurrency

OtherComprehensiveInc

AvailableForSaleSe

ShareBasedCompensationArrangementByShareE

ShareBasedCompensationArrangementByShareE

ShareBasedCompensationArrangementByShareBased

ShareBasedCompensationArrangementByShareBased

Other

ShareBasedCompensationArrangement

BusinessCombinationRe

Class

Class

SharebasedCompensationArrangementByS

FairValu

FairValueLia

St

St

StockIs

BusinessCombinationRecognizedIdent

Sto

FairValueMeasurementWithUr

S

S

FairValueMeasurementWithUnobserv

Antidilutive

Antidilutive

Derivati

FairV

C

FederalHomeLoanBank

Unrecogn

U

Unrecogniz

Unrecog

DisposalGroupIncludi

Disposa

D
Disposa
DisposalGr
DisposalGro
Disposal
DisposalGroupIn
Di
DisposalG
Dis
Dispo
Disposa
Disposa
Dispo
Dispos
DisposalGrou
DisposalGr
DisposalGroupIncludir

EffectiveIncomeT
EffectiveIncomeT
Effec
EffectiveIncomeTaxRateRecc
EffectiveIncomeTaxRateRecc
EffectiveIncomeTaxRateRec
Effec
EffectsOfCorporateF
Emp
EmployeeServiceShareBasedComp
EquityMethodI
EquityMethoc
EquityMethodInv
EquityMethodIr

E

EquityMet

EquityMethodInv

EquitySecuritiesWit

Exchang

ExpectedReclassificationFromAccumulat

ExposureTo

Ex

FairValueAssetsMeasuredOnRecurringBasisCha

FairValueGainsLossesOnFinancialAssetsReclassifiedOutOfFinancial

FairValueLi

FairValueLiabilitiesMeasuredOnRecurringBasisCha

FairValueMeasurementW

FairValueMeasurementWit

FairValueMeasurementWithUnobservableInputsReconciliationRecurrin

FairValueMeasurementW

FairValueMeasurementWithUnobservableI

FairValueMeasurement

FairValueMeasurementWithUnobservab

FairValueMeasurementWithUn

FairValueMeasurementWithUnobservab

FairValueMeasurementWithUnobservableInputsI

FairValueMeasurementWit

FairValueMeasurementWithUn

FairValueMeasurementW

FairValueMeasurementWithUno

FairValueMeasurementWithUnob

FairValueMeasurementWit

FairValueMeasurementwithUnobservableInputsF

FairValueMeasurementwithUnobs

FairValueMeasurementwithUnobser

FairValueNetDerivativeAssetLiabilityMeasuredOnRecurringBas

FairValueNetDerivativeA

FairValueOfOve

FairValueOfOv

FairValueOf(

FairValueOfOv

FairValue

FairValueOptio

FairValueOptior

FairValue

FairValueOptionLoansHeldAsAs

FederalHomeLoanB

Fi

FinancialAssetsSoldUnderAgreementsToRep

FinancialInst

Financia

F

Financi

F

FinancingReceivableAll

FinancingReceivableAllowanceFor

F

Fi

FinancingReceivab

FinancingReceivableAllowance

FinancingReceivab

Fina

FinancingReceivableBe

Finan

Financing

F

FinancingReceiva
FinancingReceivableExcluding
FinancingReceivableExcluding
FinancingRecei
FinancingReceivabl
FinancingReceivab
FinancingReceivable
Finan
Financing
FinancingRec
FinancingReceivableE
FinancingReceivableExcludingAccruedInteres
FinancingReceiv
Financ
F
FinancingReceivab
I
FinancingReceival
FinancingReceivableRatioOfGradeClassToG

FinancingReceivableRatioOfGradeClassTo
FinancingReceivableRatioOfGradeClassToG
FinancingReceivableRatio
FinancingReceivableRatioOfGradeClassToG
FinancingReceivableRatioOfGradeClassTo
F
FinancingReceivableRatioOfTotalPastDueLoansToTo
FinancingReceiv
Financi
Financing
Financing
FinancingReceivableRec
FinancingR
FinancingRece
FinancingReceiva
FinancingReceivab
F
FinancingReceivab
FinancingReceiva
F
FiniteLived
Fin

FiniteLivedIntangibleAssetsI

FiniteLivedIntangibleAsset

FiniteLi

Finite

FiniteLived

GainLossOnForeignCurr

GainLossOnInterest

GainLossRela

GainsLossesOnFinancialAssetsMe

GainsLossesOnFinancialAssetsMe

GainsLossesOı

GainsOrLossesReali

GuaranteeObli

HeldToMaturitySecuritiesContin

HeldToMatur

Held

HeldToMaturitySecuritiesContin

HeldToMaturitySecuritiesCo

HeldToMaturitySecuritiesCo

Impa
IncomeLossFromContinuin
IncomeLossFro
I
In
IncomeLo
Income
Income

IncomeTaxRelatingToE

IncomeTaxRelatingToFinancial

In

In

Increasel

IncreaseDecreasel

Increas

IncreaseDecrea

IncreaseDecreaseInProfitOrLo

IncreaseDec

I

IncreaseDecre

IncrementalCommonSharesAttributableTo

IncrementalComr

Inde

InterestA

Interes

InterestExpen

InvestmentB

InvestmentCompanyRe

InvestmentOwnec
Investme
Investm
Iss
Is
Issuan
Issuance
Iss

l

LetterOfCreditF

LiquidationB
Liquida
L
Lo
LongTermDebtAndCa
LongTermDebtAndCap

Mandat

Mandatc

Mar

Minimu

MinorityInt

NetI

NetIncomeLossIncl

NonInv

NoncashOrPartNoncas

NoncontrollingInterestConvers

NoncurrentAssetsOrDisposalG

NotionalExposureOfPortfolioOf

NumberOfCou

NumberOfNearTermVixFu

Numb

Num

Nu

Nu

OffBalanceSheetI

O

(

OtcD

Otc

OtcDerivati

O

Oth

OtherComprehensiveIncomeAvailablef

OtherComprehensiveIncomeDefinedBe

OtherComprehensiveIncomeForeignCurrencyTransactio

OtherComprehensiveIncomeForeignCurrencyT
C
OtherComprehensiveI
OtherComprehensiveInc
OtherComprehensiveIn
OtherComprehensiv
OtherComprehens
OtherComprehensiv
OtherComprehen
OtherComp
OtherComprehensiveIncomeLossDerivativeExclude
OtherComprehensiveIncomeLossFinalizatior
OtherComprehensiveIncomeLossFinancialLiabilityFairValueOpti
OtherComp
OtherComprehensive
OtherComprehensiveIncomeLossP
OtherComprehensiveIncomeLossPensionAndOtherPostretirementBer
OtherComprehensiveIncomeLossRe
OtherComprehensiveIncomeI
OtherComprehensiveIncomeLossReclassificationAdjustmentFromAOCIPensio
OtherComprehensiveIncomeNetOfTaxChangeInFairValu
OtherComp
OtherComprehensiveIncom

O

OtherTaxEffectsF

OwnershipPerce

Partic

Particip

Payment

PaymentsFc

PercentageForIssuedA

Perce

P

PercentageOfThePublicShareholdingTol

Pr

ProceedsFromIssuanceOfSharesUnderInc

ProceedsF

Proceeds

ProceedsFr

Proc

ProceedsFromSaleOfWhc

Proc
ProceedsFromSa
PropertyPlantAndEquipmentAndOperatingLease
PropertyPlantAndEquipmentLessorAssetUnderOperatingLeas
PropertyPlar
PropertyS

PurchaseA
Purch
Purch

Reductiontoaccr

ResearchAn
Restructu
R
RevenueNotFromCo
RevenueNo
Re
Revenu

Reverse

RsuW

SecuredBo

SecuritiesLoanedIncludingNotSubjectToMaste

Sec

SecuritiesPurcha

SecuritiesPurchasedUnderAgreementsToResellOffsetAgainstCollat

Secur

SecuritiesSol

Sec

Securities

ShareBasedCompensationArranger

ShareBasedCompensationArrangementByShareBase

ShareBasedCompensationArrangementByShareBasedPayme

ShareBasedCompensationArrangementByShareBasedPaymentAwardEquityInstrumentsOthe

ShareBasedCompensationArrangementByShareBasedPaymentAwa

ShareBasedCompensationArrangementByShareBasedPaymentA

ShareBasedCompensationArrangementByShareBasedP

ShareBasedCompensationArrangementByShareBasedP

ShareBasedCompensationArrangement

ShareBasedCompensationArrangementByShareB

ShareBasedCompensationArrangementByShareB

ShareBasedCompensationArrangementByShar

ShareBasedCompensationArrangementByShar

ShareBasedCompensationArrangementBy
ShareBasedCompensationArrangementBy
ShareBasedCompensationArrangementByShareBasedPaym
ShareBasedCompensationArrangementByShareBasedPaym
ShareBasedCompensationArrangementByShareBasedPaym
ShareBasedCompensationArrangementByShareBasedPaym
ShareBasedCompensationArrangementBy
ShareBasedCompensationArr
ShareBasedCompensationArra
ShareBasedCompensationArra
ShareBasedCompensationArrang
ShareBasedCompensationArrangementByShareBasedP
ShareBasedCompensationArrangementByShareI
ShareBasedCompensationArrangement
ShareBasedCompensationArrangementByS
ShareBasedCompensationArrangementsByShareBasedPayme
ShareBasedCompensationArrangementsByShareBasedPayment
ShareBasedCompensationArrangementsByShareBasedPayment
ShareBasedCompensationArrangementsByShareBasedPayı
ShareBasedCompensationArrangementsByShareBasedPayı
ShareBasedPaymentArrangementNonvest
ShareBasedPaymentArrangementNonveste
ShareBasedPaymentArrangementNonvest
S
ShareOfProfitLoss
SharebasedCompensationArrangeme

SharebasedCompensationArrangeme
SharebasedCompensationRequisiteServicePeriodRecognitionandStoc
S
SharesBasedCompensationArrangementByShareBasedPaymentAwardEquityInstrumer
St
StockIssu
StockIssue

StockIssue

StockIssu

StockRepurchaseI

Su

SubsidiaryOf

TaxEffectOfRevenuesExemptFromTaxationAndTaxEffec

TemporaryEquityCarry

TerminationOfConting

UnadvancedLineofCr

UnadvancedLongterm

Ur

UnrecognizedTaxBe

Unrecognize

UnrecognizedTaxB

UnrecognizedTaxBenefitsI

Ur

Va

VariableInterestEntityConsolida

WeightedAverageExercisePriceO

WeightedAverageExercisePrice

WeightedAverageExercisePrice

WeightedAverageExercisePriceOfS

AccumulatedDepreciationDepletionAndA

Accumulated

Accumulate

AccumulatedOtherCo

AccumulatedOtherComprehensiveIncomeLossOtherThanTempora

AccumulatedOtherComprehensiveIncomeLossOtherthanTemporaryIn

AcquisitionsTh

Acquis

AdditionsOtherThanTh

AdditionsOther

Adj

AdjustmentsF

AdjustmentsForDecreaseIncreaseInNoncurrentAssetsOrDisposalG

AdjustmentsForImpa

AdjustmentsForIncreaseDecreaseInDepositsFromBanksAccruedPayabl

Ad

AdjustmentsToAdditionalPaidI

AmountOfRest

AmountOfRestrictedNetAssetsForConsolidatedAndUnconsolidatedSubsidiari

AmountRecognizedInNetPe

AnnualTargetedLong

AnnualTargetedLongTermCompensationPerformanceMeasure

AnnualTargetedLongTermCompensationPerformanceMeasurementBenchmarkGrowthRateInTotalShar

AnnualTargetedLongTermCompensationPerformanceMeasurementBenchmarkGrowthRateInTotalShar

AnnualTargetedLongTermCompensationPerformanceMeasurementBenc

AntidilutiveSecuritiesExcludedFromComputationOfEarningsI

AssetsDisposedOfByMet

AssetsTh

AvailableFor

BusinessAc

BusinessAcquisitions

BusinessCombir
BusinessCombinationConsiderationTransf
BusinessCombinationContingentConsiderationA
BusinessCombinationContingentConsiderationArr
BusinessCombinationProFormaIr
BusinessCombinationProf
BusinessCombinationRecognizedIdentifiableAssetsAc
BusinessCombinationRecognizedIdent
BusinessCombinationRecognizedIdentifiableAssetsAcqui
BusinessCombinationRecognizedIdentifiableAssetsAcqui
BusinessCombinationRecognizedIdent
BusinessCombinationRecognizedIdent
BusinessCombinationRecognizedIder
BusinessCombinationRecognizedIdentifiableAssets
BusinessCombinationRecognizedIdentifiableAssetsAcquiredAndLiabilitiesAssumedBr
BusinessCombinationRecognizedIdent
BusinessCombinationRecognizedIdentifiableAssetsAcquiredAndLiabilitiesAssumedC
BusinessCombinationRecognizedIdent
BusinessCombinationRecognizedIdentifiableAssetsAcquiredAndL
BusinessCombinationRecognizedIdentifiable
BusinessCombinationRecognizedIdentifiableAssetsA
BusinessCombinationRecognizedIdentifia
BusinessCombinationRecognizedIdentif
BusinessCombinationRecognizedIdentifi
BusinessCombinationRecognizedIdentifiableAssetsAcqui
BusinessCombinationRecognizedIdentifiableAssetsAcquiredAndLiabilitie
BusinessCombinationRecognizedIdentifiableAssetsAcquiredAndLiabilitiesAssumedInvestmentsIr
BusinessCombinationRecognized
BusinessCombinationRecog

BusinessCombinationRecognizedIdentif
BusinessCombination
BusinessCombinationRecognizedIdentifiableAssetsAc
BusinessCombinationRecognizedIdentifiableAs
BusinessCombinationRecognizedIdentifiableAssetsAcquiredAndLiabilitie
BusinessCombinationRecognizedIdentifiableAssetsAc
BusinessCombinationRecogni:
BusinessCombinationStepAcquisitionEqu
BusinessCombina

Capi
CashAndCashEquivalentsIncreaseDecrea
CashAndSecuritiesS
CashAndSecur
CashCashEquivalentsRes
CashCashEquivalentsRestrictedCashAndRestrictedCashEquivalentsPeriodIncreaseDecreaseIncludingExc
CashFlo
CashFlowsUsedInObtainingControl
Ca

ChangeInUnreal
ChangeInUnrealizedGainLo
ChangeInUnrealizedGainLo
ClassOf\
Collat
CollateralizedFi
CollateralizedFinancingTransactio
CollateralizedFinancingTransa
CommonSharesAr
CommonSharesAndWarr

Com

ComprehensiveIncom

Com

ConcentrationOfRiskPercentageFor

Contra

Cry

Currer

DebtM

DebtSecuri

DebtSecuritiesAvailableForSaleAmo

DebtSecuritiesAvailableForSaleAmorti

DebtSecuritiesContin

DebtSecuritie

DebtSecuritiesContin

DebtSecuritie

DebtSecur

DebtSecuritiesHeldT

DebtSecuritiesHeldToMa

DecreaseIncreaseInCashAndSecuritiesSegregatedAndOnDepositForRegulator

DeemedDiv
DeferredCompensationArrar
Deferred

DeferredTaxAssets

DeferredTaxAssetsTaxDef

DeferredTaxAssetsTaxDeferredE

DeferredTaxAssetsTaxDeferredE

DeferredTaxAs

DeferredTaxA

DeferredTaxEx

D

DeferredTaxExpenseIncom

Deferre

DeferredTaxLiabil

Defe

Define

DefinedBenefit
DefinedBenefitPlanA
DefinedBenefitPlanAssumptionsUsedCalcu
DefinedBene
Define
De
DefinedCo
DefinedCo
DepositsFromBanksAccruedPayabl

Derivativ

Der

Deriv

DerivativeInstrumentsGainLossRecognizedInIncomeIne

Deri

DerivativeLiabilityFairValueHedging

DerivativeLiabilitySecuritiesSoldUnderAgre

DiscontinuedC
D
DiscontinuedOp
DiscontinuedOperationIncomeI
DisposalGroupIncludi
DisposalGroupIncludingDisc
DisposalGroupIncludingDiscor

Q4

The question asks to retrieve the top 5 companies with the highest total assets in 2024, displaying each company's name, sector, and the count of unique tags in their filings. The query we use joins the SUB, NUM, and TAG tables to access company details, asset values, and tags, then calculates total assets by summing values where the tag has "Assets" in it. We group it by company name and sector. The query then orders results by total assets in descending order and limits the output to the top 5 companies.

In [80]:
```
%%sql
SELECT
    SUB.name AS "Company Name",
    SUB.sic AS "Sector",
    COUNT(DISTINCT TAG.tag) AS "Unique Tags Used",
    SUM(CASE WHEN TAG.tag = 'Assets' THEN NUM.value ELSE 0 END) AS "Total As
FROM
    SUB
JOIN
    NUM ON SUB.adsh = NUM.adsh
JOIN
    TAG ON NUM.tag = TAG.tag AND NUM.version = TAG.version
GROUP BY
    SUB.name,
    SUB.sic
ORDER BY
    "Total Assets" DESC
LIMIT 5;
```

```
 * postgresql://student@/SECDB
5 rows affected.
```

Out[80]:

Company Name	Sector	Unique Tags Used	Total Assets
ECOPETROL S.A.	1311	384	924163282000000
AMERICA MOVIL SAB DE CV/	4813	246	9655132705000
GENERAL MOTORS CO	3711	473	2741695000000
MICROSOFT CORP	7372	249	882534000000
BOEING CO	3721	540	822336000000