# Setting up Joins

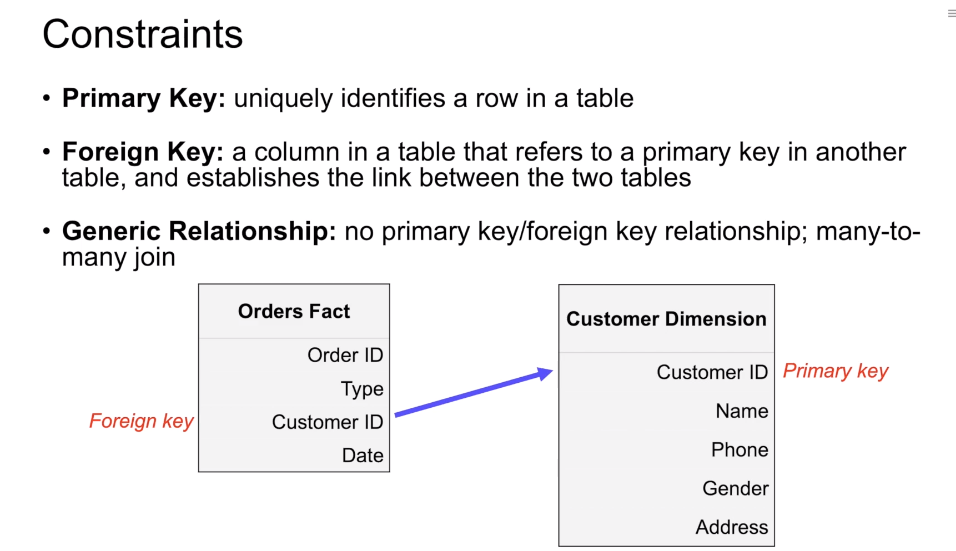
## Agenda

|  |
| --- |
| **Understand table joins** |
| **Set joins on the database level** |
| **Set joins in ThoughtSpot** |
| **Modify Worksheet join settings** |

**No joins :** in thoughtspot, other table will be greyed and disabled

**Worksheet: collection of tables (same as universe)**

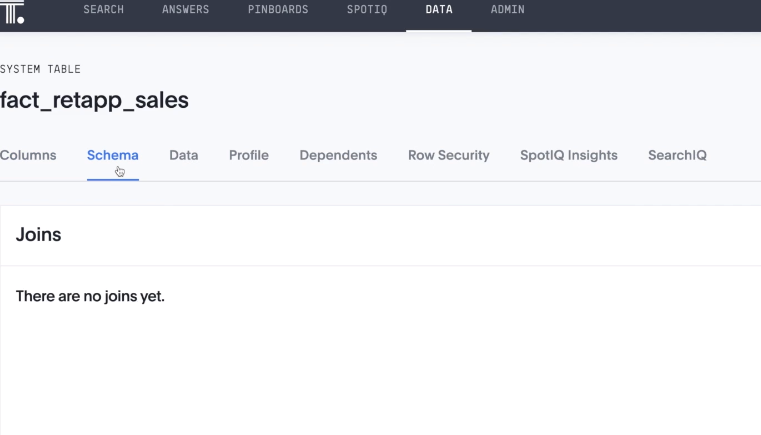
**Constraints: primary key and foreign key and generic relationship**

****

For many –many to resolve aggregate issue,use bridge tables

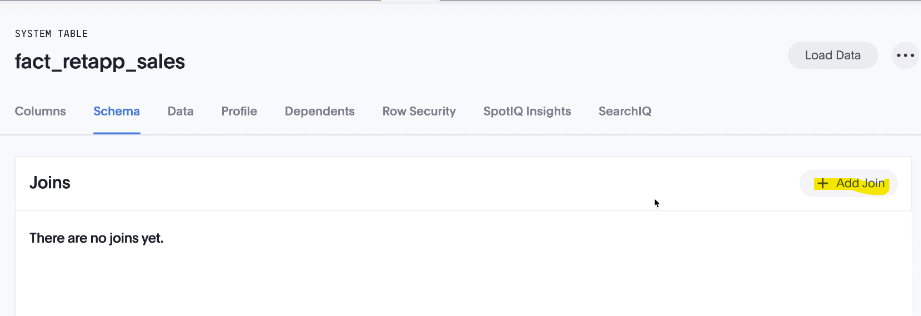
**Database join**

To see if join s are established in database level, goto data, table > schema and we can find joing there

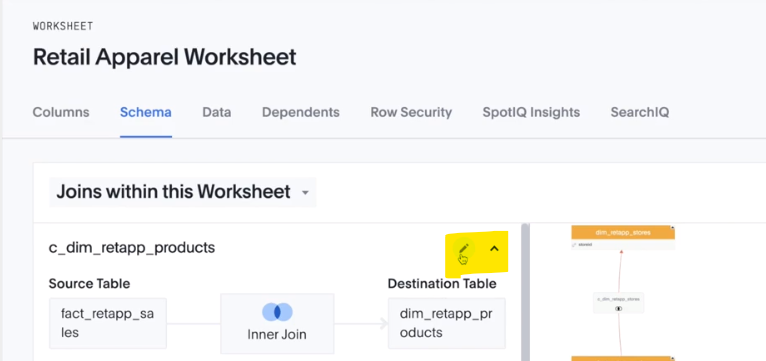


Access : to create join, can administer or can manage data(+can edit permission)

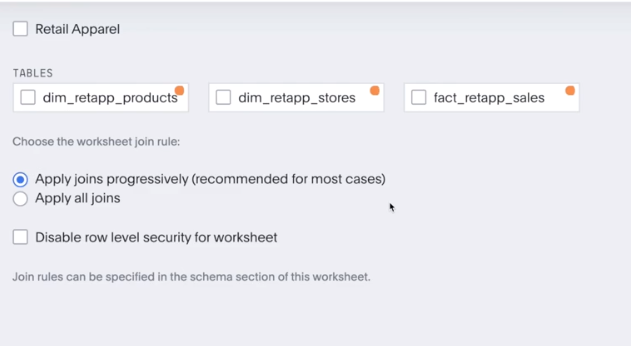
**Join:** always start from source(fact) table where there is many relations from the table



**To edit join:**



**Worksheet join rule:**

****

**[[1]](#footnote-1)**

# 2.Worksheet design

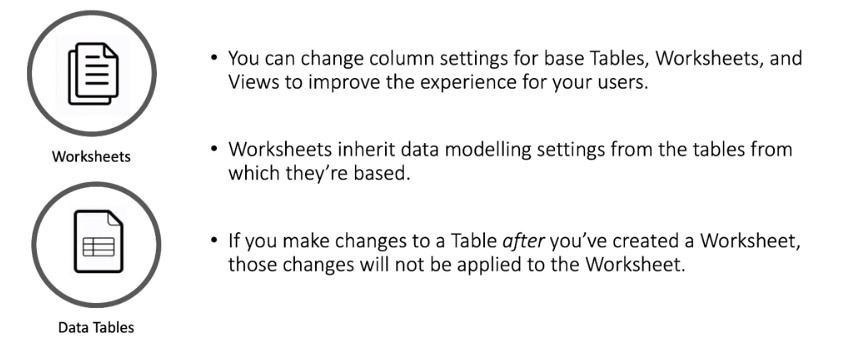
Worksheet: same as universe

Two ways to create worksheet

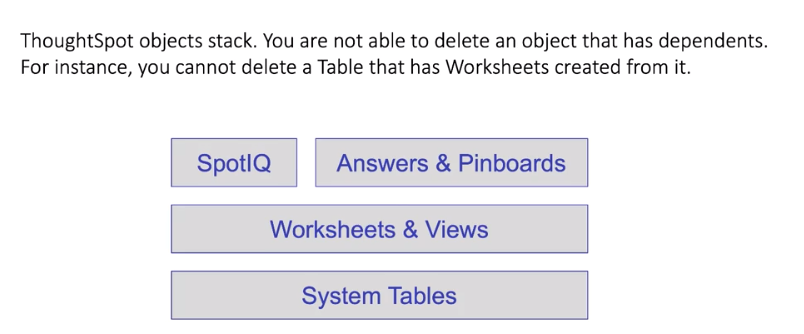
1. Import from TML
2. New worksheet from table

Steps : select table > then select column and then rename (there is sample records for each column)

## =>Modify Column settings

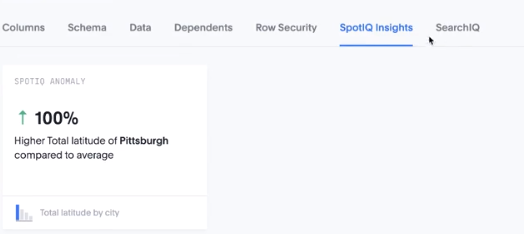


Hierarchy



\*Row security : cannot be set on worksheet

\*Spot IQ insights :

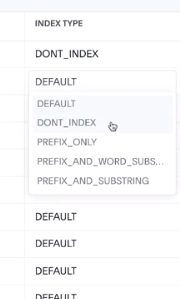


\***Search I**Q : thoughtspot NL search (plain English search) – to optimize this source for search IQ

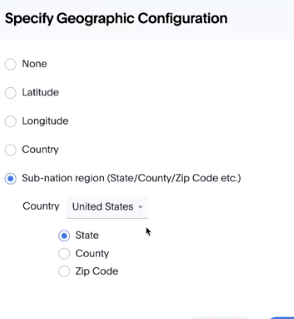
## Columns settings :ALL settings:

* data type cannot be changed
* aggregation is available for measures and also dimension (to see, change additive to yes and then proceed)
* Hidden (not be expose columns to users)
* Synonym : search terminology ;separated by commas
* Index type: any attribute , not date, prefix and substr is used when searching the field.

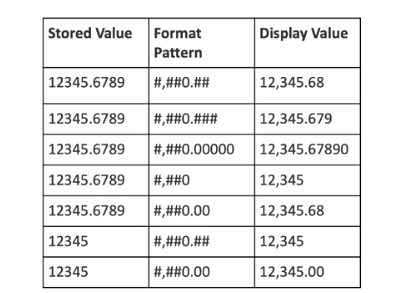
If long text, only prefix is used



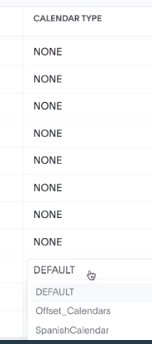
* Suggested to use “don’t unidex” for columns that have large text(like comments)
* Geo config : Location field (e.g state)



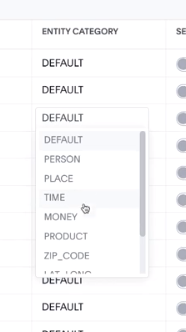
* Index priority : order in which search suggestion appear (by default it is UBR-user based ranking). If all is set to same value, it is UBR. If we set it, we are overwriting UBR
* Format pattern:



* Currency type: Currency has a separate field
  + Infer from browser
  + From a column (another column)
  + Specify ISO Code (commonly used to get currency based on country)
* Attribute Dimension : useful with Chasm trap (2 fact table related to shared dimension)
* When column in that dimension table should not be used to resolve a chasm trap. In which case, turn off the attribute dimension. Basically it says it looks like chasm trap and not oneby turning off.
* Calendar type : we can use custom calendar in this field.



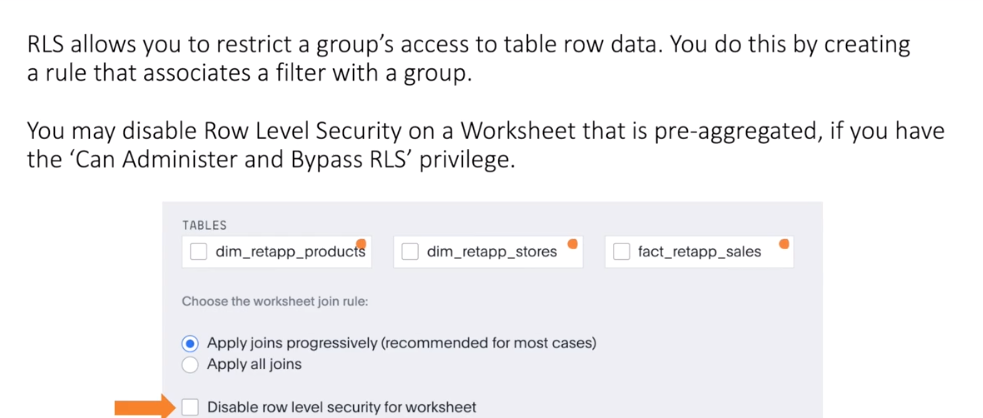
* To use custom calendar use TSL option. We can use formula to use the date that is setup on custom calendar.
* Entity category : search IQ (natural language search)



* Search IQ enabled: when turned on, only those column are available for search IQ

## Advanced: Row level Security

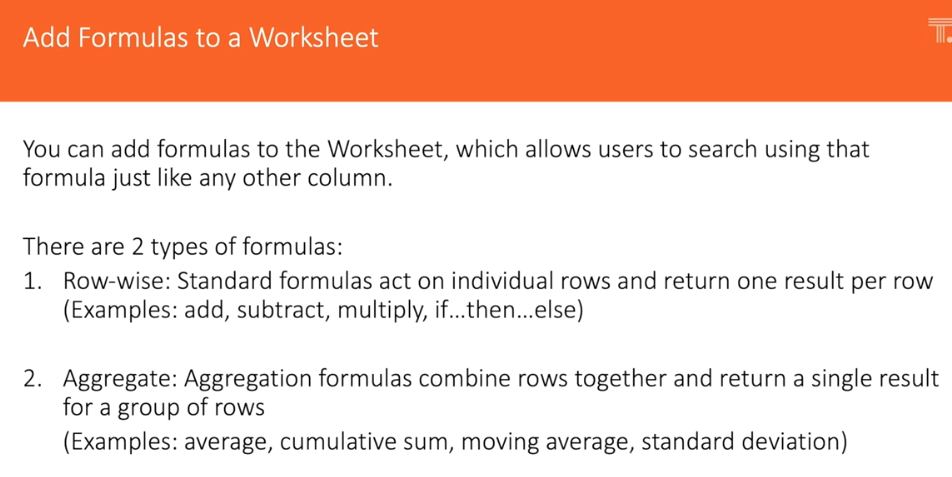
RLS : only on tables and worksheet inherit that.



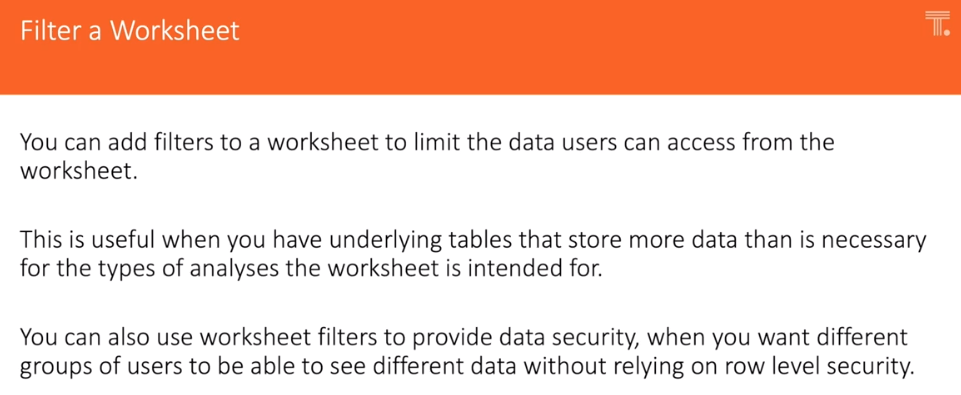
Why row level security disable option: if anyone if trying to look at overall sales, if row level security is defined, then users of that region can see only their region and the overall sales will be same as their region sakes

**Access : can administer and bypass RLS to see “disable row level security”**

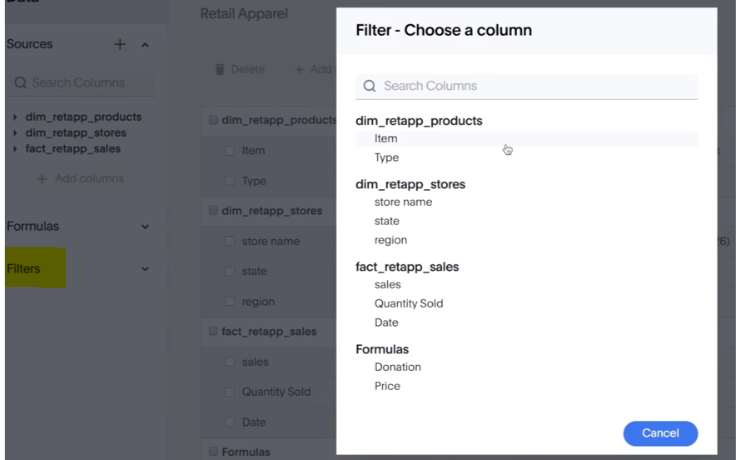
## Advanced: Formula



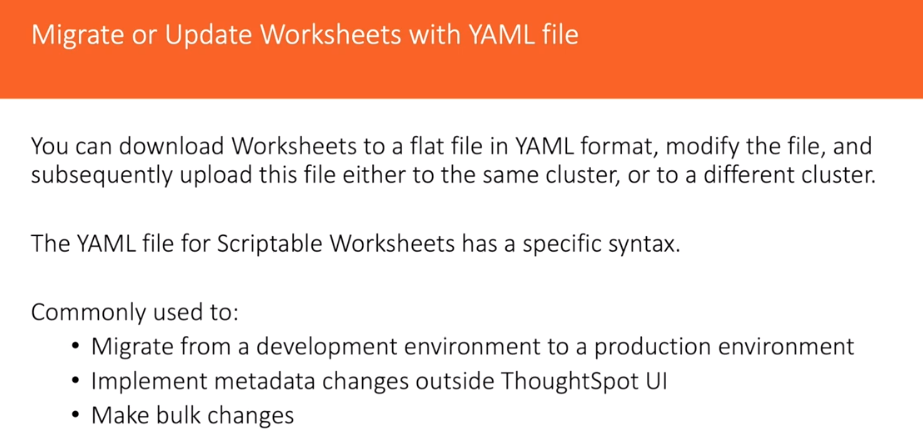
## Advanced: Add filter to worksheet

****

**How to?**

****

## Advanced:YAML Mgrate or update



**How to : goto worksheet > … > export worksheet**

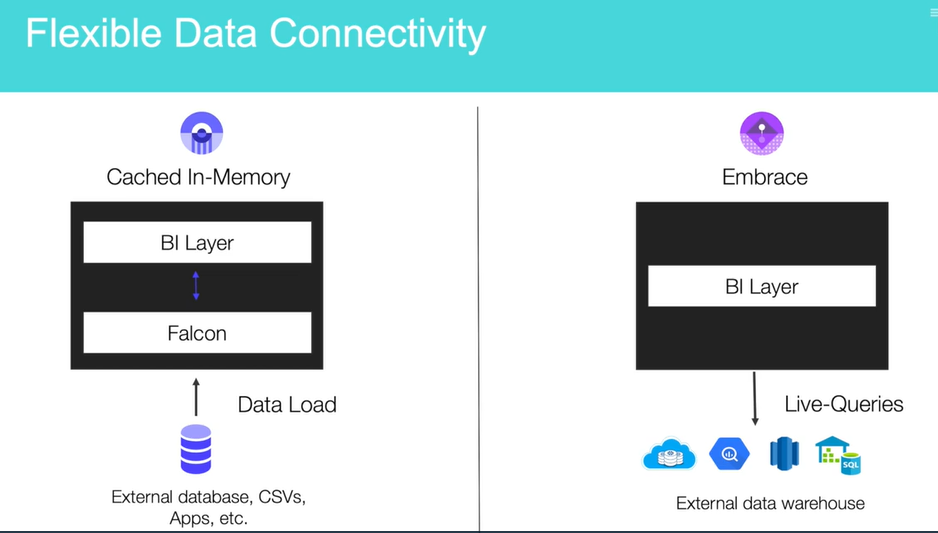
# 3.Loading Data with DataFlow

**Overview:**

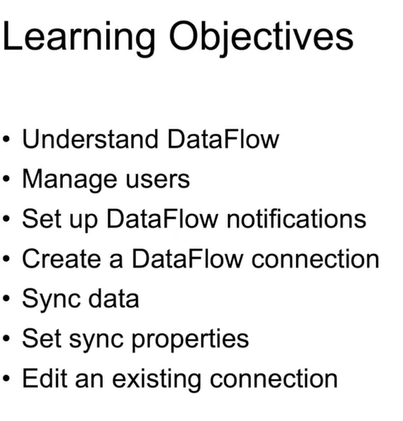
|  |
| --- |
| Understand DataFlow |
| Manage users |
| Set up DataFlow notifications |
| Create DataFlow connections & sync data:      Overview      MySQL Database      Salesforce      File System (S3 bucket) |
| Join tables created through DataFlow |
| Edit an existing connection |

**Thoughtspot :**

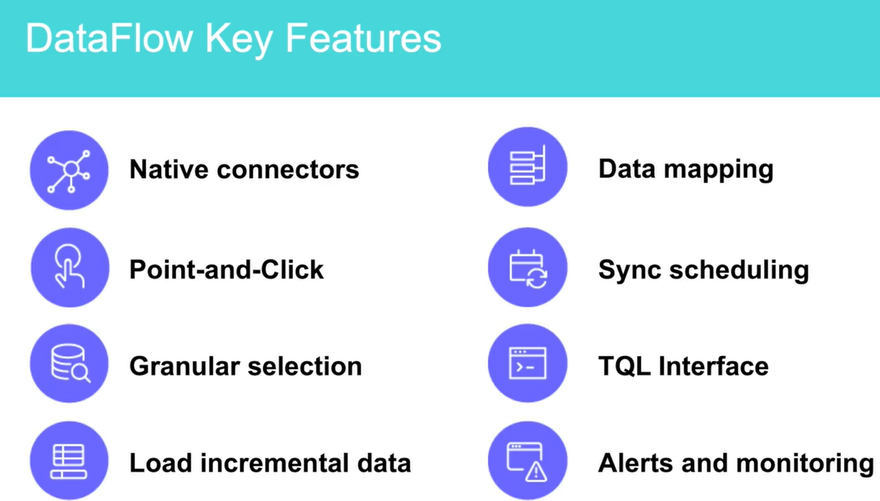
* **In memory database (Falcon)**
* **TS Load: Thoughtspot high speed bulk loader**
* **Dataflow : Graphical interface to load data**
* **Schedule to refresh data**

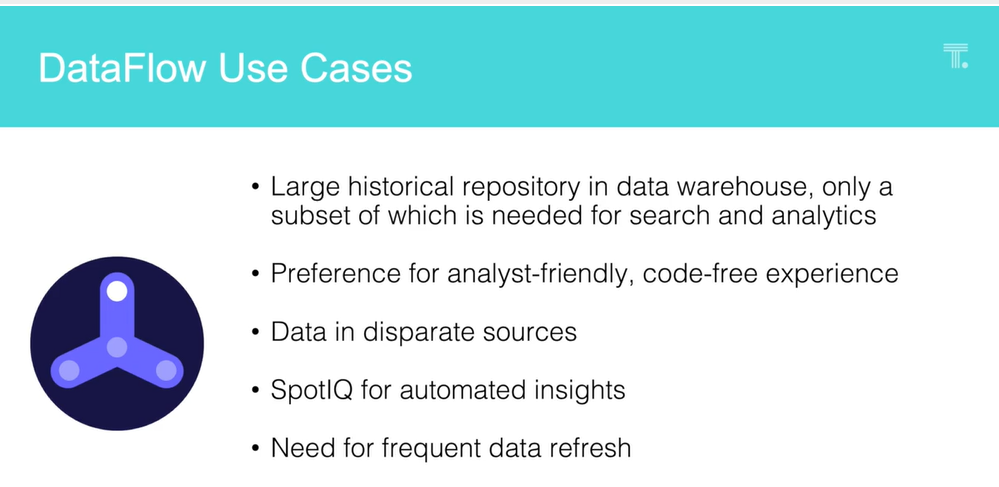
****

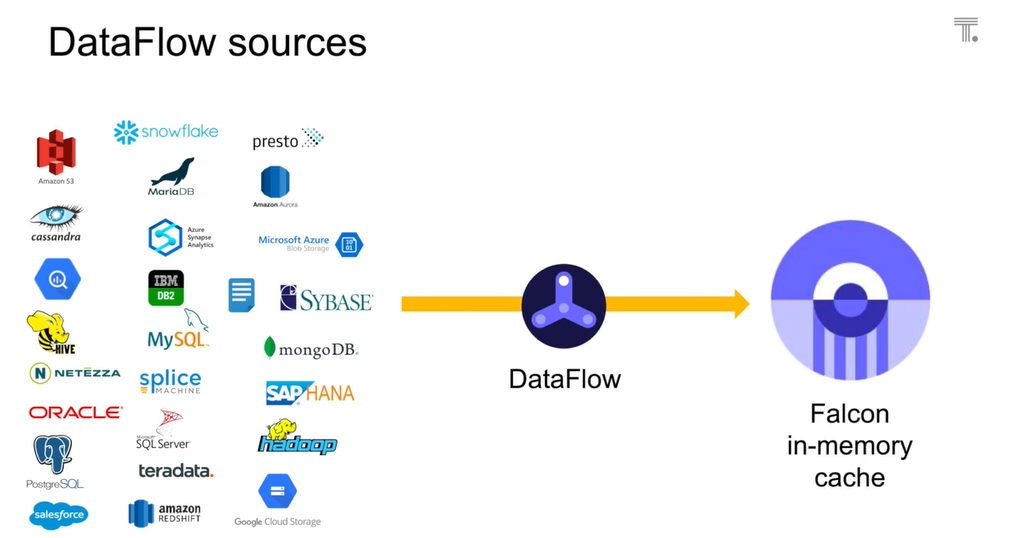
* **Data flow can be used to load data INOT FALCON only(to load data to snowflake or external database is not available**

****

* **Granular data selection to load only few columns of the table is available**

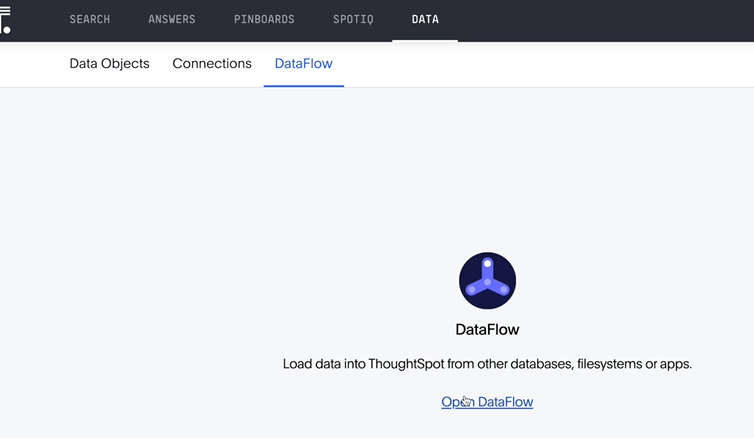
****

****

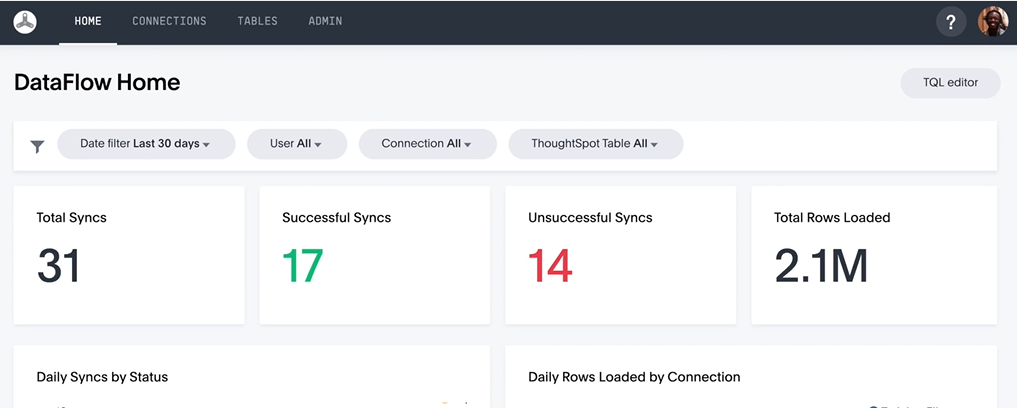
****

[[2]](#footnote-2)

## 3.a.Manage users

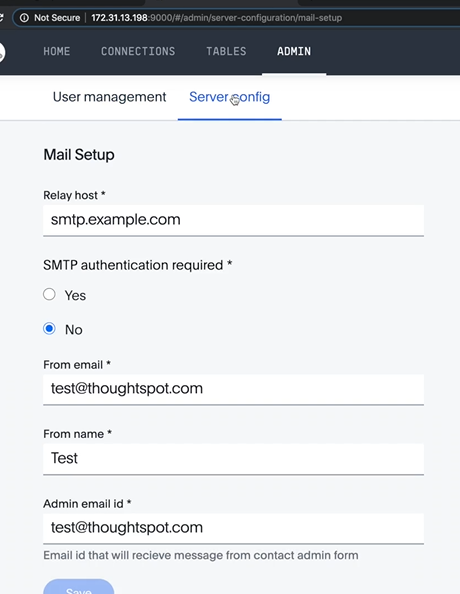


To add new user : Goto Admin tab and “Add users”

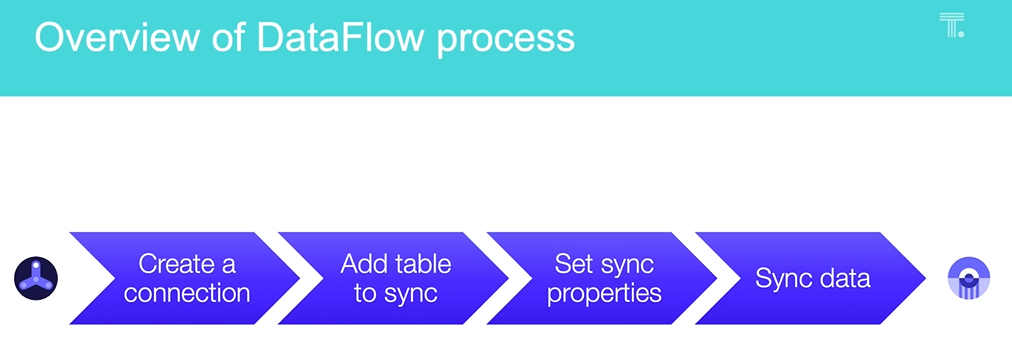


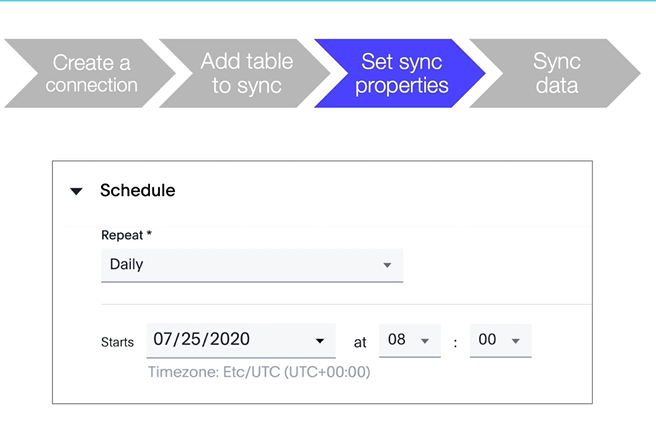
Note: If user don’t have “TS” account, they cannot associate to data flow account. They have to setup email id association here

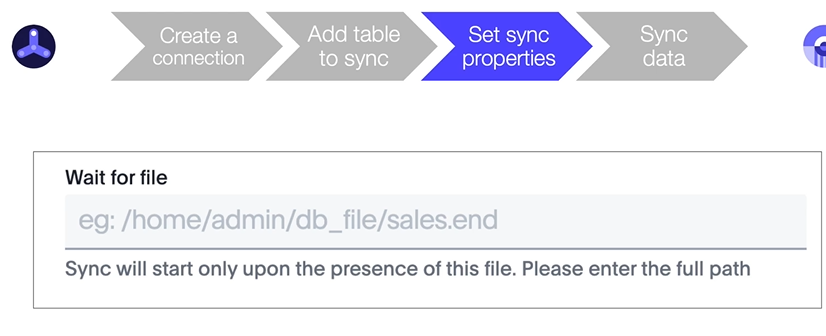
## 3.b.Data flow notification



## 3.c Steps



* Add one table at a time
* Select frequency of sync,data fields to sync and TS Load options
* Options to setup trigger file to initiate the load



* Joins will not carry over to data flow(only in worksheet)..

## 3.d.Demo of MYSQL database

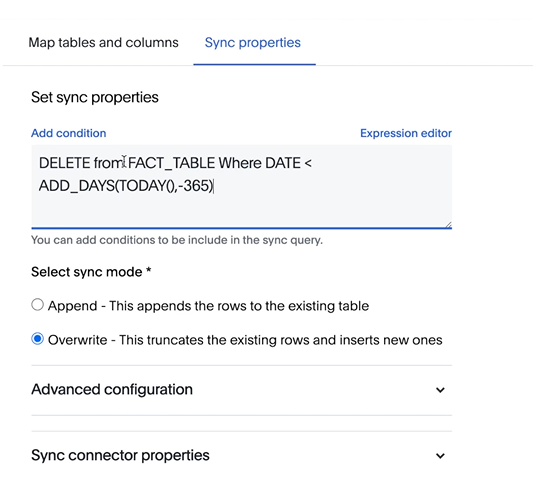
1) Data > Data Flow

2) Home : shows sync performance across org

3) shards are used when setting up data flows on large tables for performance and enables us to mentoned number of shards

If shard =1, the table is maintained in one node. If >1 , the table is broken into smaller tables and distributed across multiple nodes

1. To use a pre sql before loading



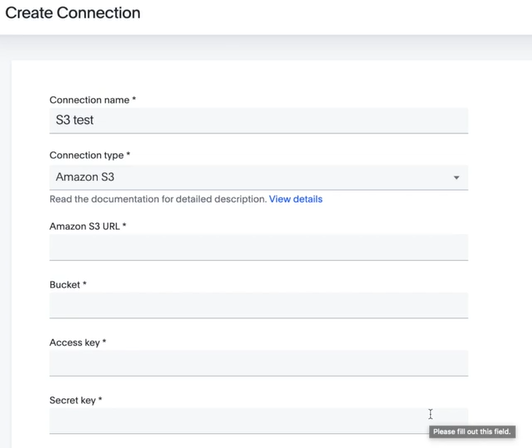
## 3.e : can be done for salesforce

-for salesforce connection , user must have API for salesforce SFDC.

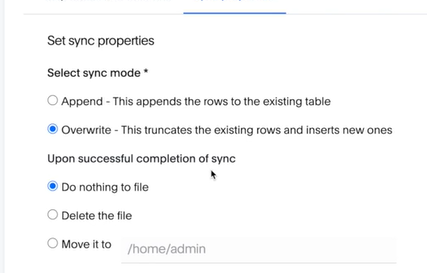
-password => password+securitytoken

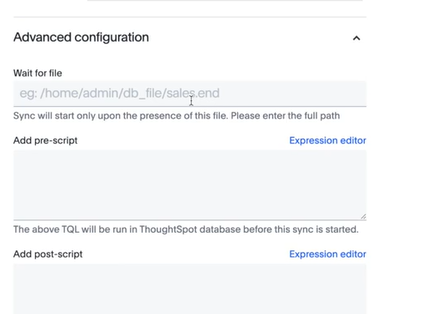
## 3.f : connection to S3

URL : Host name URL, bucket, access key and secret key

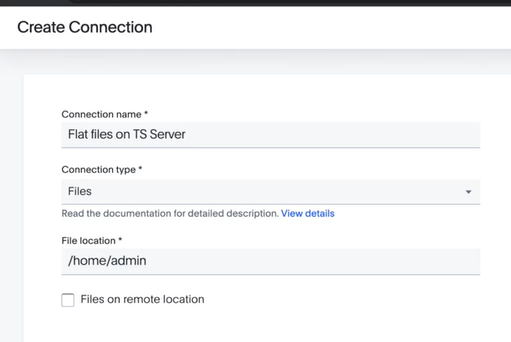


S3’ special setting: wait for trigger ; delete or retain file in source S3





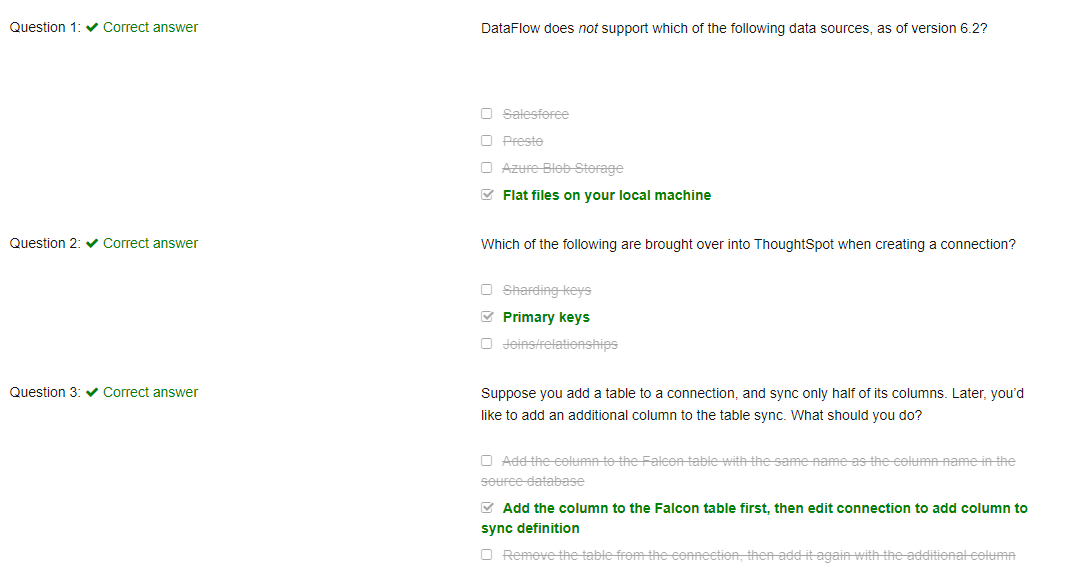
## 3.g Flat Files on Thoughtspot server

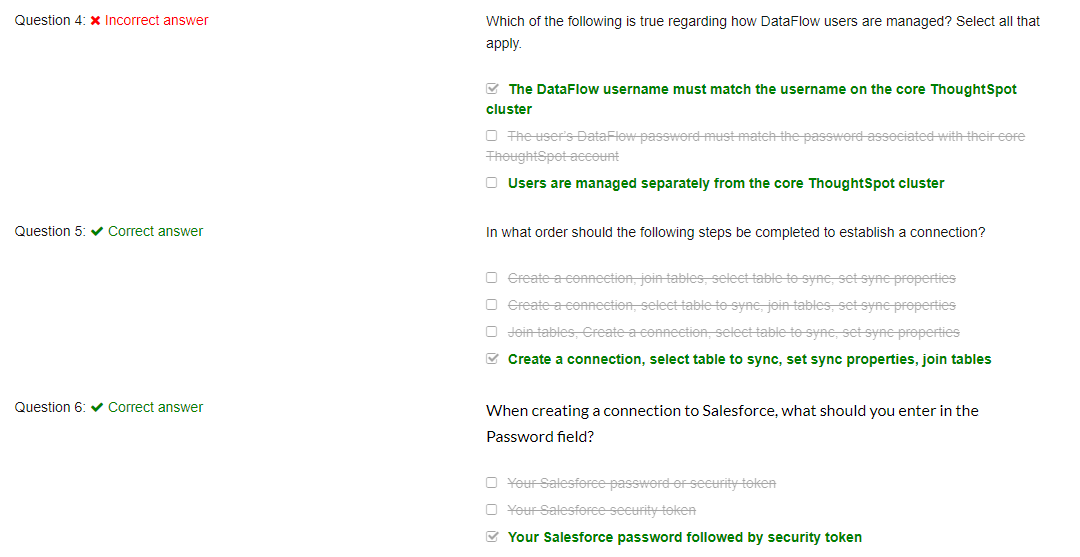


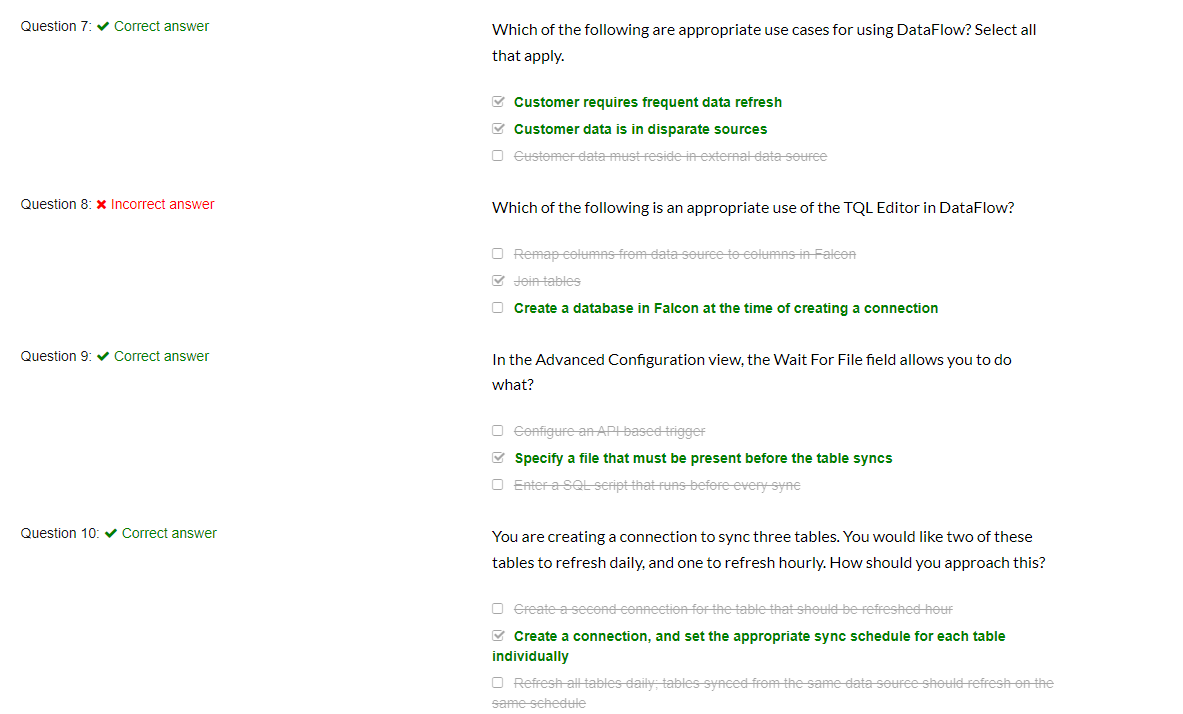
NOTE:

* To add new field to sync definition, first add this field to the underlying falcon table and then update the sync
* When a field is deleted from the source, until the sync definition is updated to remove this column, it will fail,.

## DATAFLOW QUIZ



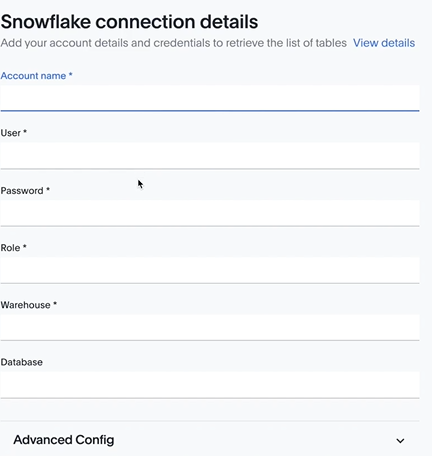


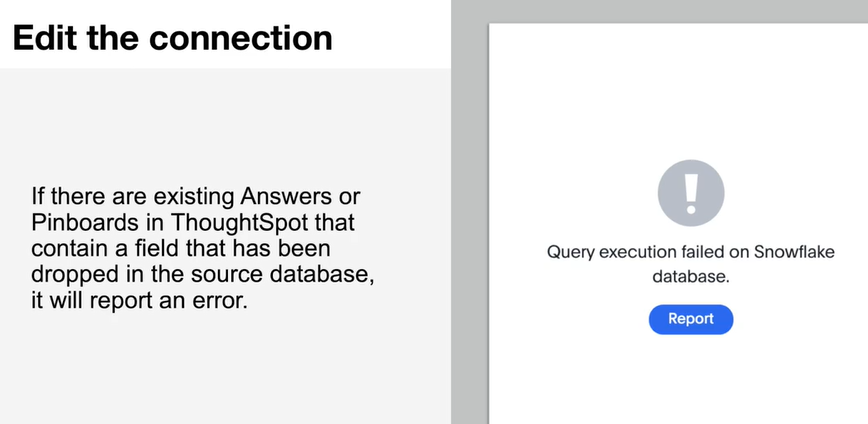


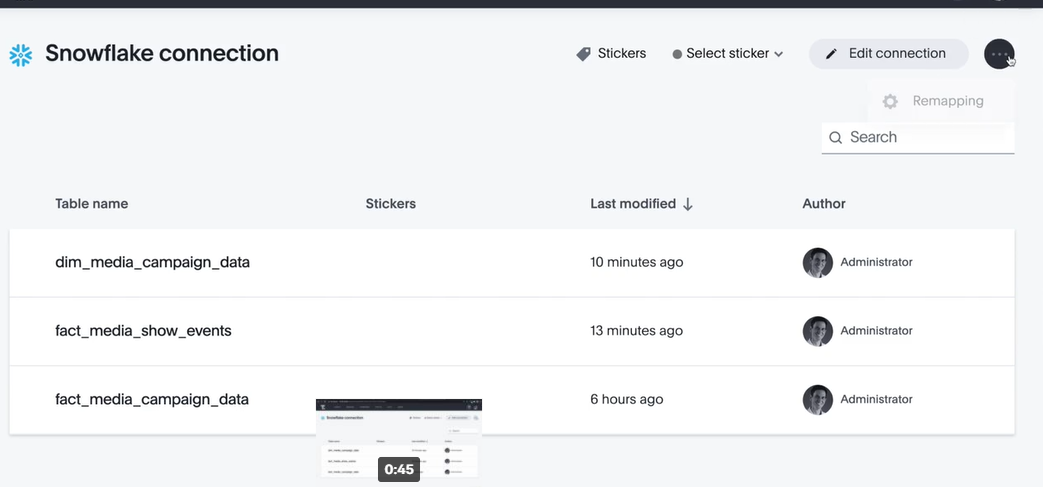
# **4.Connecting Data with Embrace(renamed to connection)**

|  |
| --- |
| **Embrace overview** |
| **Create a connection and Search data** |
| **Edit or remap a connection      Edit an existing connection      Remap a connection**  * Direct connection to sources. No need to move data          * Joins, pk,foreign key are imported to thoughtspot..it becomes a linked connectivity to the external data source * To access “variant” datatype in snowflake, create a view in source database and use this view in thoughtspot |

**=>NOTE: row level security must be set in TS even if source database has row level security**

****

****

****

# **5. Scriptability**

Worksheet, Ansers and punboards can be downloads as TSL(Thoughtspot Script Language) and updated in an editable format

\*Note : when an worksheet is downloaded, we can edit either formula name or formula expression. We cannot edit both and push it up the same time. (it needs two migration

\* TSL is useful when moving objects between environments, making bulk updates

\* Worksheets can be imported, updated and exported to overwrite the **same** worksheets

\* Answers/Pinboards can be imported, updated and exported and cannot overwrite the **same** worksheets(not beofore version 6.2)

\* When importing Answers, we will have option to import the dependant worksheet and so does pinboards

\* when importing updated worksheets to new environment, the underlying tables, connections needs to be setup the same in the new environment

# **6.Architecture**

=> Falcon is a in memory database that can be loaded using Data sync with On prem Thoughtspot

=> Thoughtspot Cloud donot have Falcon or Data Sync. It uses Embrace

=> Embrace/connection connect to cloud databases and use “Index” (instead of search asin falcon) as if its reading from in-memory to address performace

=? Best practives : when using tables with billions of rows, disable index

* Use more auto resume suspend to take advantage of snowflake virtual warehouse cache(virtual warehouse cache is clear when warehouse is suspended)

# **\*\*NOTES**

Access : to create join, can administer or can manage data(+can edit permission)

You cannot delete an object if there is a dependency on them

Row level security is only applied on table (not on worksheet)

**Join:** always start from source(fact) table where there is many relations from the table

We can use conditional join using TML

Salesforce native connector is present in Thoughtspot

Table sync in data flow can be done one table at a time

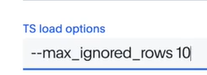
# Acronym:

TQL

TML

YAML

TS Load Option : used in data flow (refer TS load flag)



# vsBOBJ/Tableau:

Any changes to the table after creating the worksheet will not be applied (cons in TS)

Also, any new columns in table will not be automatically added

Can we have custom sql?

Similarities with BOBJ : just like downloading universe for environment move, we can download YAML file of worksheet from one environment to other

Yaml file is like editable format of worksheet with tables, fields, join info

# \*How to?

YAML : Yet Another Markup Language

Difference between TML (Thoughspot Modeling Language) download and YAML download of worksheet? Can TML also include answer,pinboards migration to a new environment?

Is RLS is setup after the worksheet is setup on a table, will it be reflected

**-can mv\_corp\_calendar use as customer calrndar type in setting up date attribute**

**UBR based on each user or each source?**

**-setup join in snowflake?**

**-create views : save search on top of search**

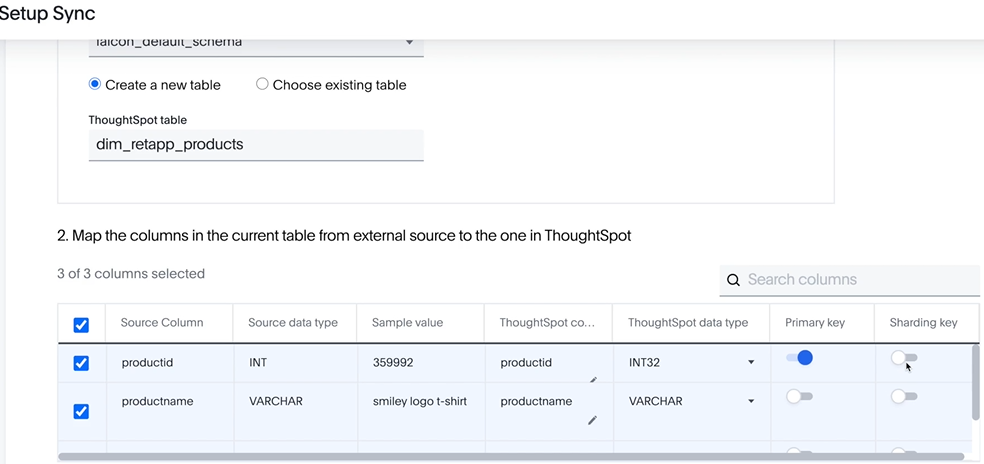
**-data tables:raw tables loaded by administrator**

**How to convert a vertical stacked bar to horizontal**

**- data tables vs. raw tables – what is the difference?**

**What is sharding key in data flow?**

**Can I do table sync in cloud?**

****

1. [↑](#footnote-ref-1)
2. [↑](#footnote-ref-2)