

Week 1 :- Fivetran course

- » Fivetran is platform which fully manages data movement and data ingestion & normalize data coming from various sources to storage destinations.
- » Principle → ELT → Extract, Load, Transform. → visualize (unseen).
- » Flow :-
 - Authenticate → fire & forget handles → normalize → create → load in schemas somewhere.
See queries & API
- » Sources can be :-
 - Application,
 - Databases → on-premise
→ Database
→ cloud
 - Events
 - source agents.
 - Files
- » Set-up →
 - ① Add connector
 - ② choose connector
 - ③ Authorize source by entering credentials
 - ④ Start initial sync
 - ⑤ Set-up time interval for incremental set-up.
- » Destination can be :- cloud data warehouses :- snowflake, bigquery.
Databases : MySQL, PostgreSQL, PostgreSQL.

» Connectors →

- ① Application connector : delivers relational schemas
- ② Database connector: all db's have diff logging methods
- ③ Report API connector : retrieve data from 3rd party sites.
ex: fbapis, instagram.
- ④ File connectors:- ① flat file c. ② Magic folder.
- ⑤ Event connectors:- Sync JSON from kafka etc as packed data

» Difference b/w Database, Warehouse and Data Lake

- ① Database → Systematic formatted and stored data may contain info. best for business or personal use.
- ② Data Warehouse → DW is basically advanced version of DB' where DB directly reports to DW on regular time intervals.
^(ETL)
- ③ Data Lake → Collection of DW and DB which can include all type of format.
Ex. Videos, audios, text
Ex. AWS, Azure, Bigquery.

Also defined as advancement of ETL.

** Destinations »

① Snowflake → Create warehouse and worksheet and run
(DBaaS)

SQL query present in firecracker dashboard and
authenticate it & save test it for connection.

» Create new worksheet.

» Run script (SQL) from firecracker dashboard

» Timezone → US, PST (-8)

» Beware of all Roles of users. Like admin, Secur etc.

» Port no ⇒ 443

② Redshift → ① Open redshift ^{search} dashboard (serverless)

② Created a namespace with default workgroup

③ Create cluster | redshift-cluster-1

dc2.large | 2 nodes | 320gb

④ Click on cluster, copy endpoint for
host string remove : he boardlets.

⑤ Go to properties → network & security grp chg:

⑥ Also give public access grant in some tab

⑦ Click on security group → click on id

⑧ Edit inbound rules →

⑨ custom TCP-5439 Port - link me to T2 the
who IP address CIDR.

* Bigquery →

- » Create new project and workspace
- » copy project id and specs.
- » Assign billing address or account to project
- » In IAM, grant access by add user + copy id from first.
- » Set up destination at fivetron platform.

Assign bigquery + Bearer ↴

* Databricks →

- » Create a new user in Databricks by signing up.
- » Give permission to user as admin
- » Create metastore in data section. with Mumbra IP.
- » Create S3L warehouse → Create S3L warehouse
Set-up with config mentioned in dashboard
- » check connection details copy →① Server hostname
② Port no.
③ HTTP path
④ Create Personal access token
- » Copy all these things into fivetron absord.
- » Run setup.

» Sources configured → ① Google Sheets

② Google Drive

③ Hubspot

④ Spotify

⑤ GitHub

» Transformations → Process of revising, computing, combining raw data into analytics.

↳ Reporting, ML & AI & Compliance

» Why transform not in BI tool → • Can't handle., • inefficient.

» Fivetran completely manages whole transformation by writing and maintaining Data models.

» User functions → We have no connector for that.

Solution which not supported → API, Events, file ingestion.

① Client writes script

② Client host script

③ Client tell fivetron where connection is

④ Rest → fivetron.



* Core elements of CF -> ① Authentication ③ Response
② Request ④ Cursoring.

>> Webhooks -> Webhook API is a RESTful API accessed by using either an HTTP client such as wget or curl.