



GROUP BY

Data Heroes;

How to execute the T in ELT



DAAN BAKBOORD, MANAGER DATA & ANALYTICS at PONG BV | JULY 2020

DAAN BAKBOORD - DAANALYTICS

DAAN BAKBOORD

DATA HEROES

MEMBER SINCE 2018

DAAN BAKBOORD

GROUP AFFILIATION
DaAnalytics

BASE OF OPERATIONS
Schiedam, Netherlands

JOINED SNOWFLAKE COMMUNITY
2018

SPECIALTIES

- Connecting BI Tooling to Data warehouses / Data Sources
- Modelling the Semantic Layer

IN THEIR OWN WORDS

My passion is collecting, processing, storing and presenting data. I help organisations to use this internal and external data in such a way that they can take action. My goal is simple. A customer must be able to make decisions based on facts and within the right context. DaAnalytics is my personal label. I provide Data & Analytics services and I have been active in Oracle Analytics since the mid-2000s. As of the end of 2017, Cloud Data Analytics has my specific attention. Initially focused on Snowflake. I am now Snowflake SnowPro Core certified. In addition, I focus on a number of other products from the Snowflake Eco-System.

DATA HEROES



<http://www.daanbakboord.com>



<http://www.daanalytics.nl/blog>



<http://www.linkedin.com/in/daanbakboord>



<http://www.twitter.com/daanbakboord>





How to Execute the T in ELT

Agenda

GOVERN DATA
LOADING DATA
TRANSFORMING DATA
CONCLUSION



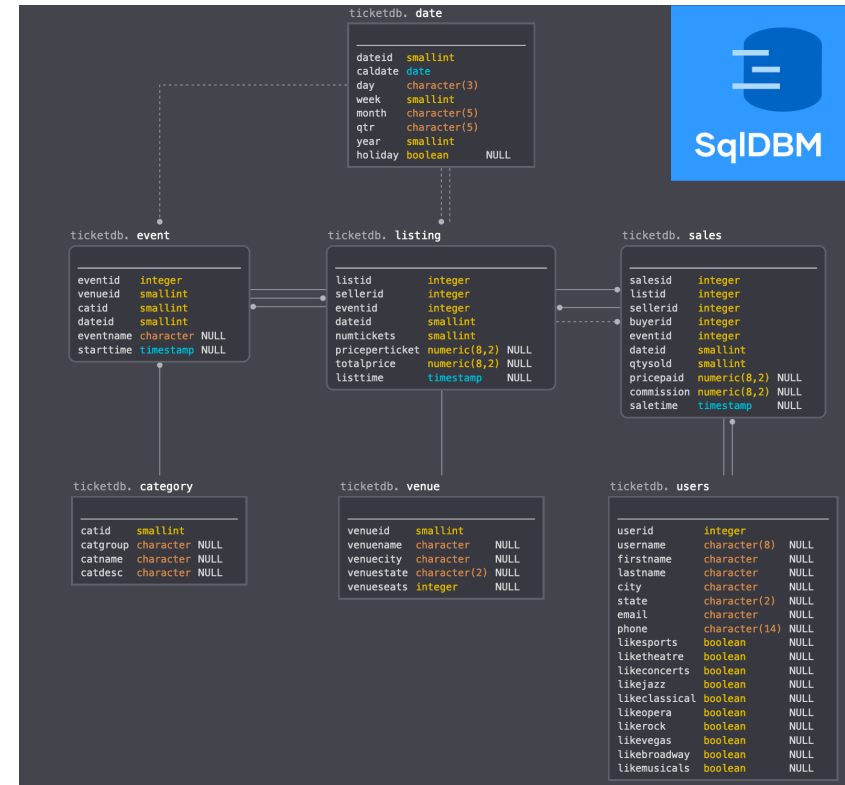
GOVERN DATA



SQLDBM – DATABASE MODELING

- Database Modelling
- Project Conversion
- DB Documentation
- Compare Revisions
- Forward Engineering
 - Snowflake Direct Db Connection
- Reverse Engineering
 - Bring your database

“Without writing a single line of code”



SQLDBM – PARTNERCONNECT

Snowflake Partner Connect

Get started with loading and analyzing your data in minutes. Automatically connect your Snowflake account with our partner applications available for a **free trial**.

Check back often as we will be adding new partners regularly.

“Try for yourself!”

The screenshot displays the Snowflake Partner Connect interface. At the top, there's a header for 'Snowflake Partner Connect' and a brief description of the service. Below this, a grid of partner logos is visible, including Fivetran, Stitch, Signo, Databricks, Snowplow, R7VERY, CHARTIO, and others. A modal dialog titled 'Connect to SqDBM' is open in the center. The dialog contains the following information:

Connect to SqDBM

SqDBM requires the following information to create your new trial account:

In order to configure the connection with Snowflake, the following objects will be created in your Snowflake account

Database :	PC_SQLDBM_DB
Warehouse :	PC_SQLDBM_WH (X-Small)
System User :	PC_SQLDBM_USER
System Password :	Autogenerated & Randomized
System Role :	PC_SQLDBM_ROLE

Role PUBLIC will be granted to the PC_SQLDBM_ROLE
Role PC_SQLDBM_ROLE will be granted to the SYSADMIN role

By clicking on **Connect** you are instructing Snowflake to create the above objects and provide all of the above information to SqDBM. SqDBM's processing of this information, and your use of SqDBM, are governed solely by SqDBM's [Terms of Use](#) and SqDBM's [Privacy Policy](#) and not your agreement with Snowflake.

Please contact [Snowflake Support](#) if you have questions about connecting with SqDBM.

Buttons: **Cancel** and **Connect**

The bottom right corner of the grid shows the SqDBM logo and a description: 'Cloud Data Modeling for Snowflake. Develop DB/DW without writing a single line of code.'



LOADING DATA



PARTNER ECO SYSTEM

Select Partner Type

ETL / ELT / DATA MOVEMENT



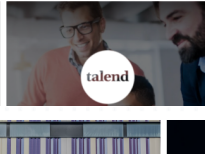
FIVETRAN



INFORMATICA



MATILLION



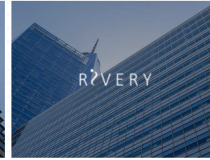
TALEND



MPARTICLE



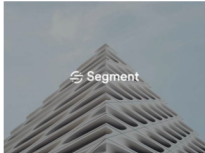
NEXT PATHWAY



RIVERY



SESAME SOFTWARE



SEGMENT



STITCH



ABINITIO



SYNSORT



TAMR



WANDISCO



XPLENTY



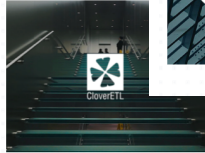
STRIIM



BRYTEFLOW



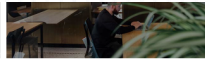
CDATA



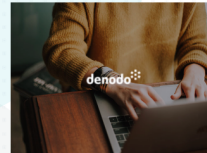
CLOVERETL



ZETARIS



COMPILERWORKS



DENODO



DIYOTTA



ETLEAP



HEVO DATA



HVR



IRI



LOOME



LYFTRON

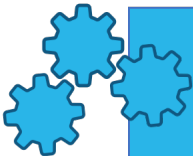


LOADING DATA

Structured & Semi-Structured



Data Load



- Snowflake
 - Manual
 - Snowpipe
 - PartnerConnect
- Fivetran
- Matillion
- Informatica

“SQL is THE Language to Query Data”



TRANSFORMING DATA



TRANSFORMING DATA

- Cleansing
 - formatting
 - validating
- Combining
 - joining
 - Integrating
- Aggregating
 - summarizing
 - cross-column calcs
- Separating
 - filtering
 - column-splitting



FIVETRAN TRANSFORMATIONS

Prepare your Data for Reporting

- Easy to use
- Data analysts' language of choice: SQL.
- Create Table
 - Name & Schema
- Schedule of choosing



Fivetran.

SQL Script *

```
1 /* ---- SAMPLE ----
2 CREATE OR REPLACE TABLE new_schema.new_table AS
3 SELECT column1, column2
4 FROM original_schema.original_table;
5 ----- */
```

Enter your transformation script above. Please ensure that the Fivetran user has correct permissions on your destination to be able to run the script successfully.

Schedule Type *

New Data

Run SQL when fresh data arrives from Fivetran

Time Schedule

Run SQL on a predictable schedule

SAVE & RUN

SAVE



DBT; THE T IN ELT

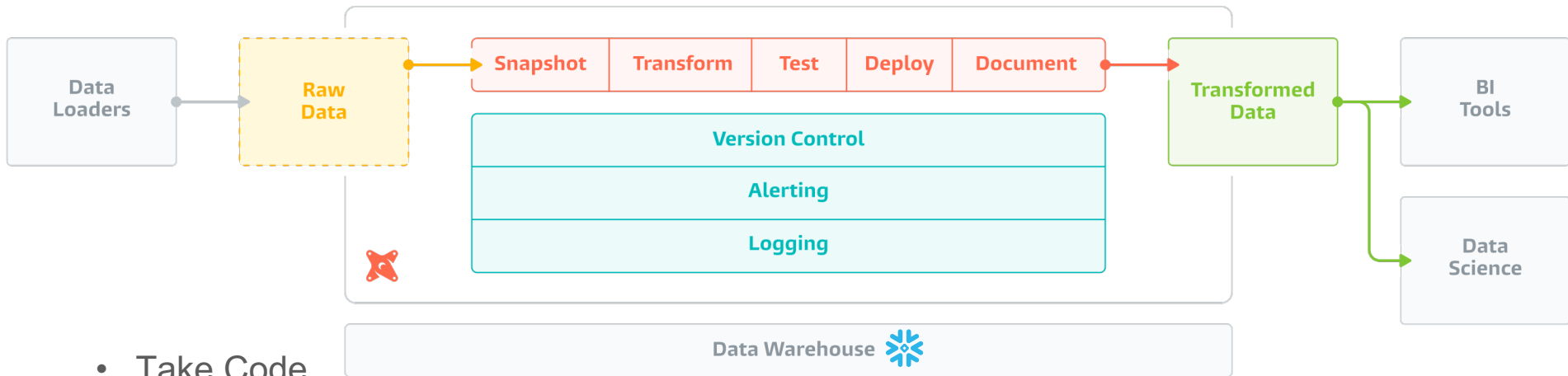
- Data already in DWH
- Logic in SQL
- Transformation in batch
- Transformations are idempotent



DBT is NOT an ETL/EL-T Tool!



DBT IN ACTION

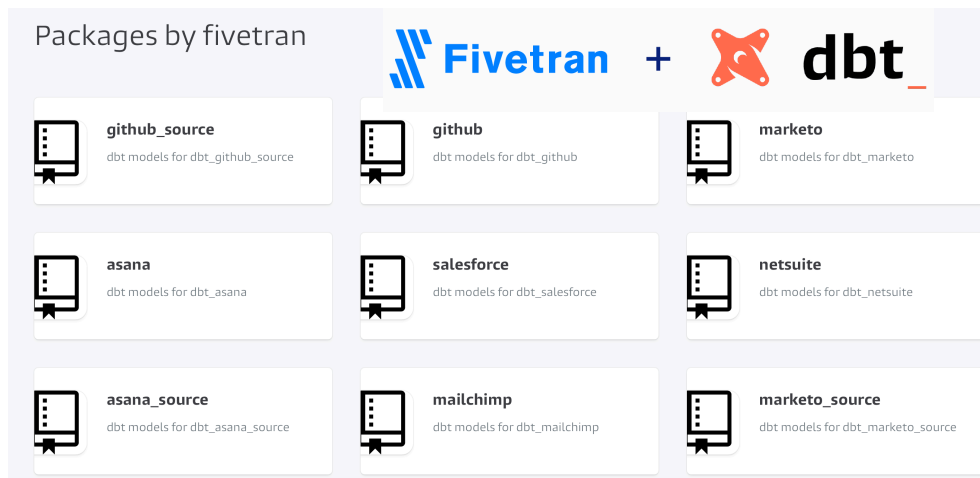


- Take Code
- Compile it to SQL
- Run against the Database



FIVETRAN & DBT

- Fivetran
 - (E) – Extract from Data Source
 - (L) – Load Data into Cloud DWH
- DBT
 - (T) Transform Data in Cloud DWH



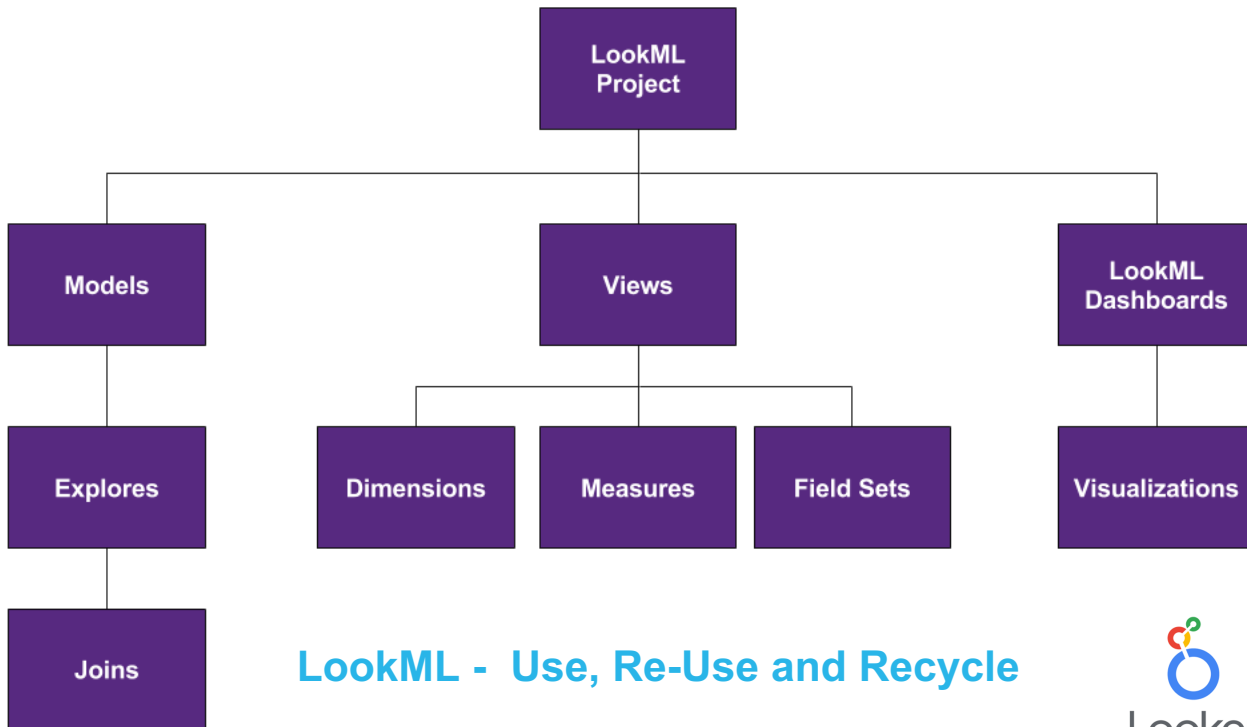
EL-T using the Best of both Worlds!



LOOKER'S APPROACH TO DATA MODELLING

02_users.view ▾

```
1 view: users {
2   sql_table_name: users ;;
3   ## Demographics ##
4
5   dimension: id {
6     primary_key: yes
7     type: number
8     sql: ${TABLE}.id ;;
9     tags: ["user_id"]
10  }
11
12  dimension: first_name {
13    hidden: yes
14    sql: INITCAP(${TABLE}.first_name) ;;
15  }
16
17  dimension: last_name {
18    hidden: yes
19    sql: INITCAP(${TABLE}.last_name) ;;
20  }
21
22  dimension: name {
23    sql: ${first_name} || ' ' || ${last_name} ;;
24  }
25
26  dimension: age {
27    type: number
28    sql: ${TABLE}.age ;;
29  }
30
31  dimension: age_tier {
32    type: tier
33    tiers: [0, 10, 20, 30, 40, 50, 60, 70]
34    style: integer
35    sql: ${age} ;;
36  }
37
38  dimension: gender {
39    sql: ${TABLE}.gender ;;
40  }
41
42  dimension: gender_short {
43    sql: LOWER(LEFT(${gender},1)) ;;
44  }
45 }
```



LookML - Use, Re-Use and Recycle



LOOKER DERIVED TABLE

- Native – LookML
- SQL-based
- Persisted
 - Stored outside Looker
 - Refreshed on schedule

order			
order_id	time	amount	customer_id
1	2015-01-01 12:00:00	100.00	1
2	2015-01-02 14:00:00	200.00	1
3	2015-01-03 08:00:00	150.00	2
...

Your original
"order" table

```
SELECT customer_id,  
       MIN(DATE(time)) AS first_order_date,  
       SUM(amount) AS lifetime_amount  
FROM   order  
GROUP BY customer_id
```

The SQL either created
by the NDT LookML or
directly in the "sql:"
parameter

customer_order_facts		
customer_id	first_order_date	lifetime_amount
1	2015-01-01	300.00
2	2015-01-03	150.00
...

The resulting
derived table



FIVETRAN & LOOKER

- Looker Blocks[®]
 - Library of LookML's
- Components of Business Logic
- Reusable & Customizable



Salesforce by Fivetran

The Salesforce Block by Fivetran provides analysis on sales opportunities and operations.



JIRA by Fivetran

JIRA is a project management software that enables development teams to coordinate, collaborate, and track progress...



Netsuite by Fivetran

The Fivetran block for NetSuite lets you easily visualize and analyze your Netsuite data.



Zendesk by Fivetran

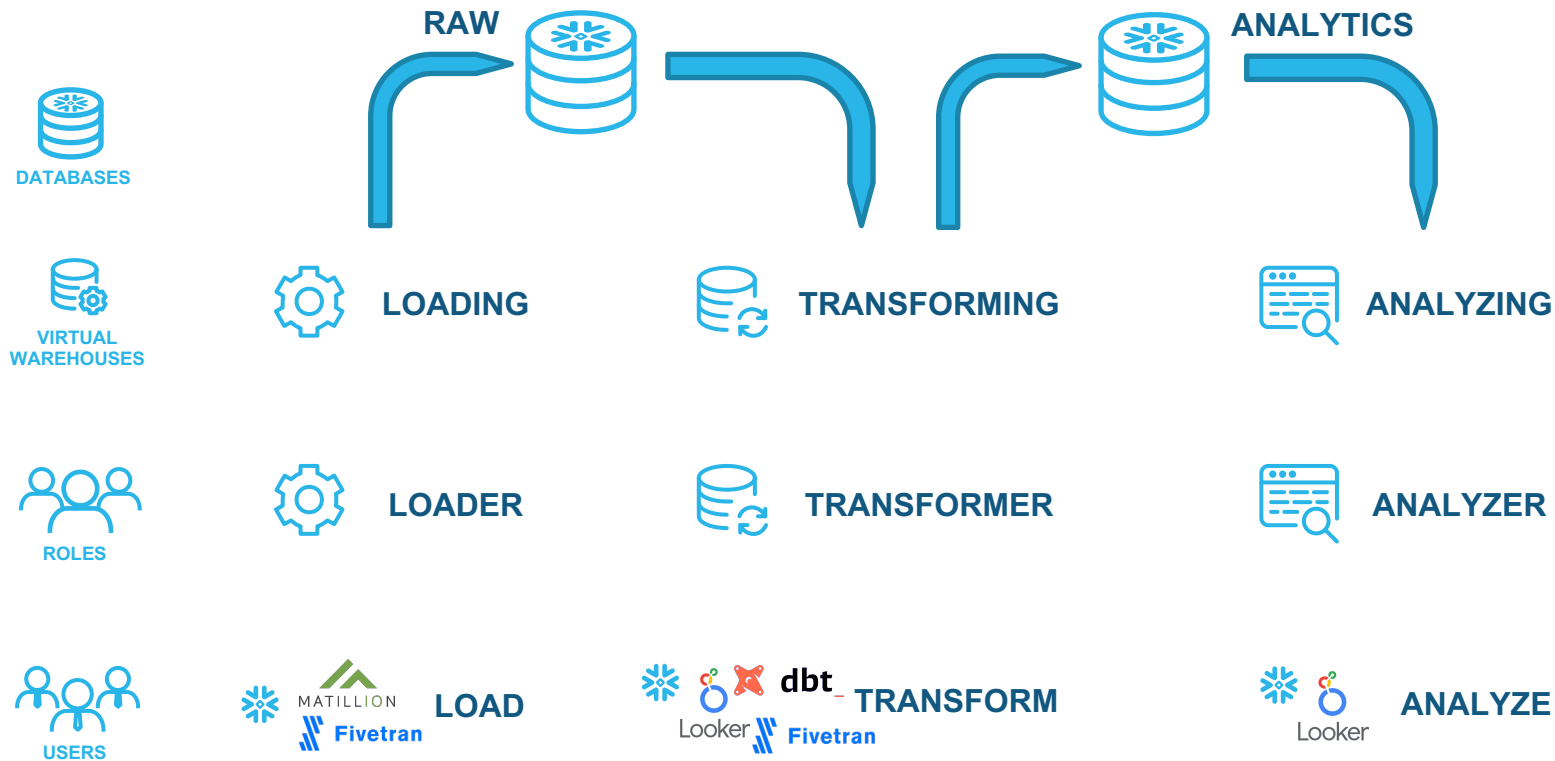
The Zendesk Block by Fivetran provides analysis on ticket information and customer support issues.



GETTING STARTED



SNOWFLAKE MODERN DATA ARCHITECTURE



CONCLUSION

- Ease of Use
 - Dwaas
- Flexibility
- Snowflake eco-system

“SQL is THE Language to Query Data”

- ANSI-SQL
- Fully ACID compliant

“Simply Load
and
Query Data”



GROUP BY

Data Heroes;

THANK YOU

