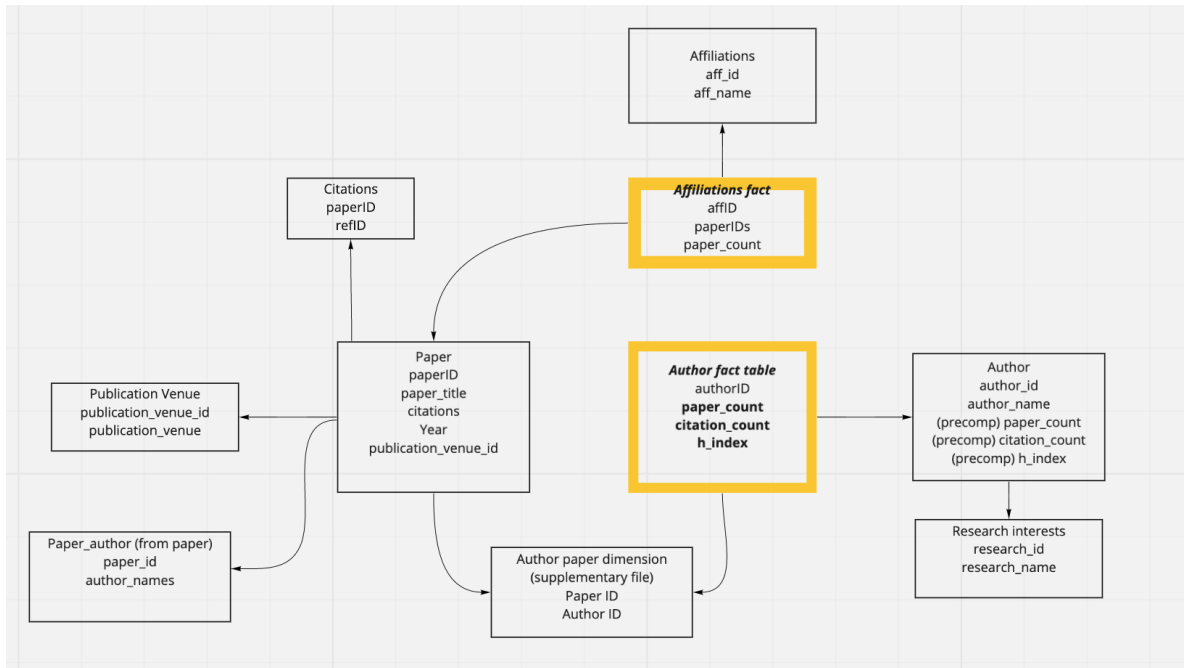


Task 1 - DILA Exercise - Query Results

Data warehouse schema:



Q1.2 Compute paper count per unique affiliation

affiliation	papers_count
nagoya univ, nago...	30
nam kwang enginee...	1
univ of technolog...	6
univ of texas at ...	41
purdue univ, west...	389
laboratoires de m...	3
padova univ, pad...	1
royal swedish aca...	1
dept of ee cs, u...	1
loughborough univ...	1
univ of quebec at...	1
shanghai jiaotong...	4
univ of aberdeen,...	5
department of ele...	20
beyond words publ...	1
western illinois ...	5
the bdm corporati...	1
the university of...	2
university of mia...	45
baylor college of...	11

only showing top 20 rows

Q1.1 Validate precomputed paper/ref counts and h indexes for the authors.

There are 1712364 unique author_ids for which the validation is being performed.

From our computations: (If you want to double check the numbers → look at the computations done on [line 100-103](#) (load_csv_into_schema.ipynb))

- citation_count → Approximately 7% of the authors pass the validation phase.
- paper_count → Approximately 99% of the authors pass the validation phase.
- h_index → Approximately 30% of the authors pass the validation phase.

author_id	validated_paper_count	validated_citation_count	validated_h_index	citation_count	h_index	paper_count
1	1	1	1	0	0	1
6	1	1	1	0	0	1
12	1	1	1	0	0	1
13	9	51	4	14	2	9
16	2	27	2	1	1	2
22	1	1	1	0	0	1
26	1	1	1	0	0	1
27	1	1	1	0	0	1
28	2	6	1	5	1	2
31	1	1	1	0	0	1

only showing top 10 rows