Score Avg by Month & Year

Year ● 2022

200

150

March

Month

Score Avg

Total Score by Month

February

March | 0.2K

0K   2K

Total Score

Month

Post	Score	Date ▲	Link
Introducing Kestra infinitely scalable open source orchestration and scheduling platform.	108	26-Feb-22	🔗
Why this subreddit dislikes the so-called Modern Data Stack?	126	27-Feb-22	🔗
questions to all DEs who hire other entry-level or Junior DEs for their Company or team.	68	28-Feb-22	🔗
Apache Airflow for Beginners Tutorial Series	213	28-Feb-22	🔗
Quarterly Salary Discussion	58	01-Mar-22	🔗
Wtf is a datalake?	178	01-Mar-22	🔗
Anyone noticing a trend where Data Engineer jobs are advertised but the role is really BI or Analyst work?	165	02-Mar-22	🔗
Every Freaking Time	219	03-Mar-22	🔗
Dbt Tests Vs Great Expectations	65	04-Mar-22	🔗

Date

2/26/2022

3/24/2022

○———○

Score

☐ 58
☐ 59
☐ 61
☐ 64
☐ 65
☐ 66
☐ 68
☐ 69
☐ 71

Top Comment

▲

A data lake is an intermediate step between a source and a data warehouse that is used to hold raw data.

Let's say I make a twitter bot for r/dataengineering called "Data Engineers Cursing": it runs once an hour downloads all comments in all threads of this subreddit and tweets any comment that contains some list of curse words: ['shit' 'fuck' 'damn'].

Let's say after running the script for 6 months I want to expand my list of curse words or I want to support other languages. If I just did source -> ETL -> Data Warehouse I could only make those changes going forward. If instead I had done source -> data lake -> ETL -> Data Warehouse I could retroactively re-process the data.
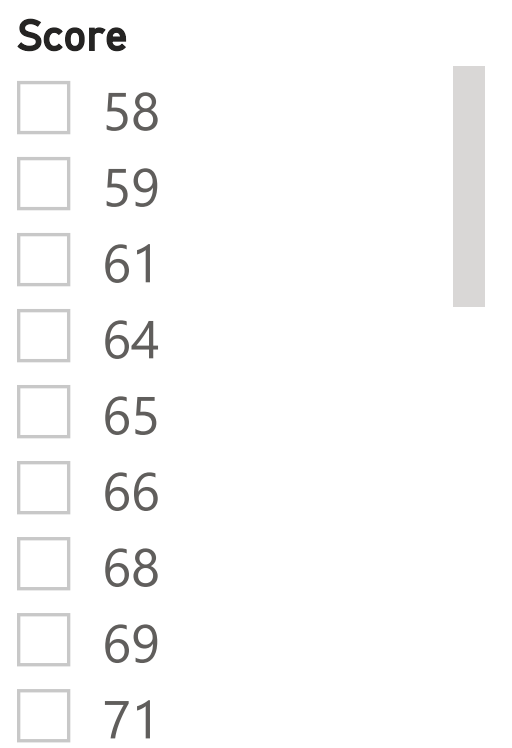
This need to retroactively re-process data and use data that you didn't think you needed is very common today.