Alternate Approach

Remove dependencies on 3rd party API

The Challenge: Accessing Public/External Data

Restrictive Limits

Even with no limits passed in on the API payload. The response is limited to 100 records

Complex Queries

Limited to queries that 3rd party has provided. No ability to handle complex queries

No Guarantees

There are no agreements between business and the vendor of this API. Meaning no notifications of API being down. Or if changes cause impacts to data flows.
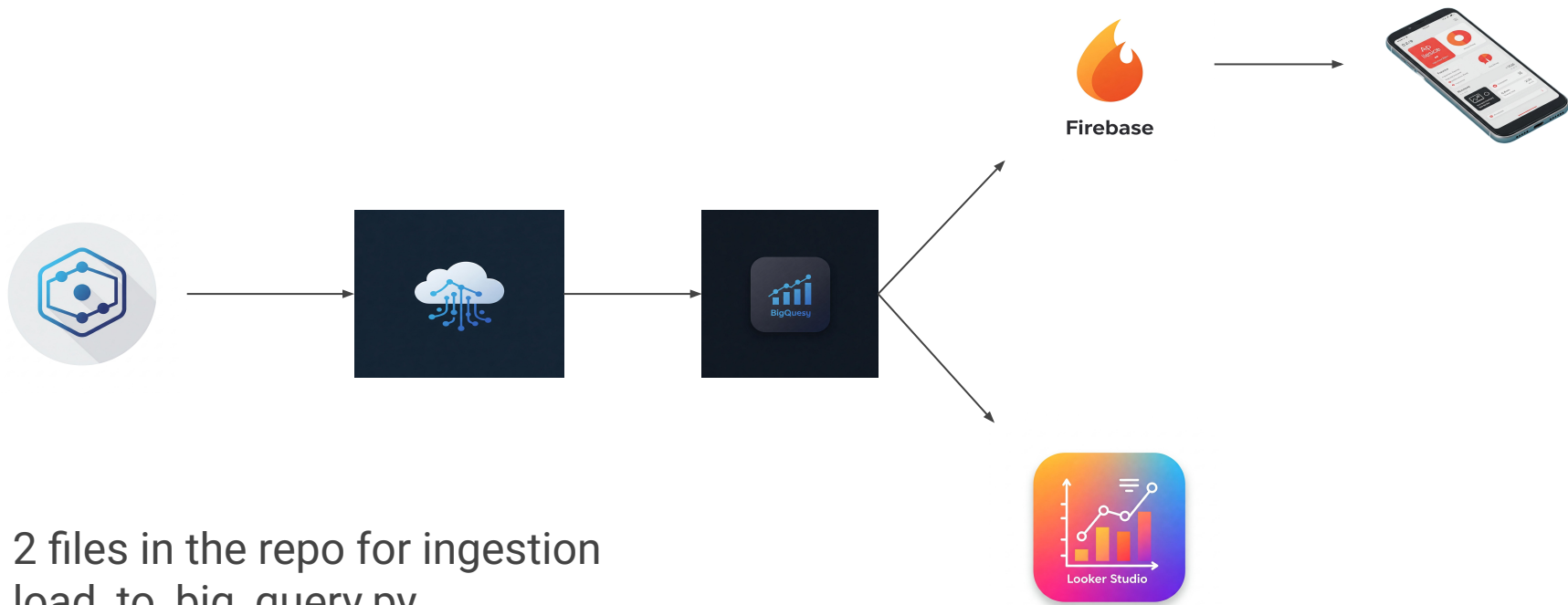
Solution

Instead of querying record-by-record, download the entire dataset as a CSV file

Many data portals (e.g., CKAN sites) offer direct download links.

Typically a simple HTTP GET request.

Firebase

Looker Studio

2 files in the repo for ingestion
load_to_big_query.py
load_to_databricks.py

business_names_australia

Overview | **Sample Data** | Details | Permissions | History | Lineage | Insights | Quality

Open in a dashboard | Create

Ask Genie Preview

	_R_E_G_I_S_T_E_R___N_A_M_E_	_B_N___N_A_M_E_	_B_N___S_T_A_T_
1	BUSINESS NAMES	TRUE POTENTIAL CARE	Registered
2	BUSINESS NAMES	True Potential Therapy	Registered
3	BUSINESS NAMES	True Power Engineering	Registered
4	BUSINESS NAMES	TRUE POWER SOLUTIONS	Registered
5	BUSINESS NAMES	TRUE PRESENCE	Registered
6	BUSINESS NAMES	TRUE RADIANCE BEAUTY	Registered
7	BUSINESS NAMES	True Rare	Registered
8	BUSINESS NAMES	TRUE RECOVERY	Registered
9	BUSINESS NAMES	TRUE REFLECTION DETAILING	Registered
10	BUSINESS NAMES	true relationships & reproductive health	Deregistered
11	BUSINESS NAMES	TRUE RESULT	Registered
12	BUSINESS NAMES	True Results Patient Management	Registered
13	BUSINESS NAMES	True Revolution Automotive	Registered
14	BUSINESS NAMES	TRUE ROOFING SA	Registered
15	BUSINESS NAMES	TRUE ROOTS	Registered
16	BUSINESS NAMES	TRUE SELF ADVOCACY	Registered
17	BUSINESS NAMES	True Self Space	Registered
18	BUSINESS NAMES	True Sense	Registered

Advantages of Data set ingestion

- Data is warehouse in our own infrastructure. Meaning ability to cluster and index to maximise performance.

- Creation of custom API for data access.

- Available for reporting purposes and to inrich internal datamodels.

- Create pipe line to firebase making data easily available to end appilcation and easily integrating with react JS.