Privacy and Security in Distributed Data Markets

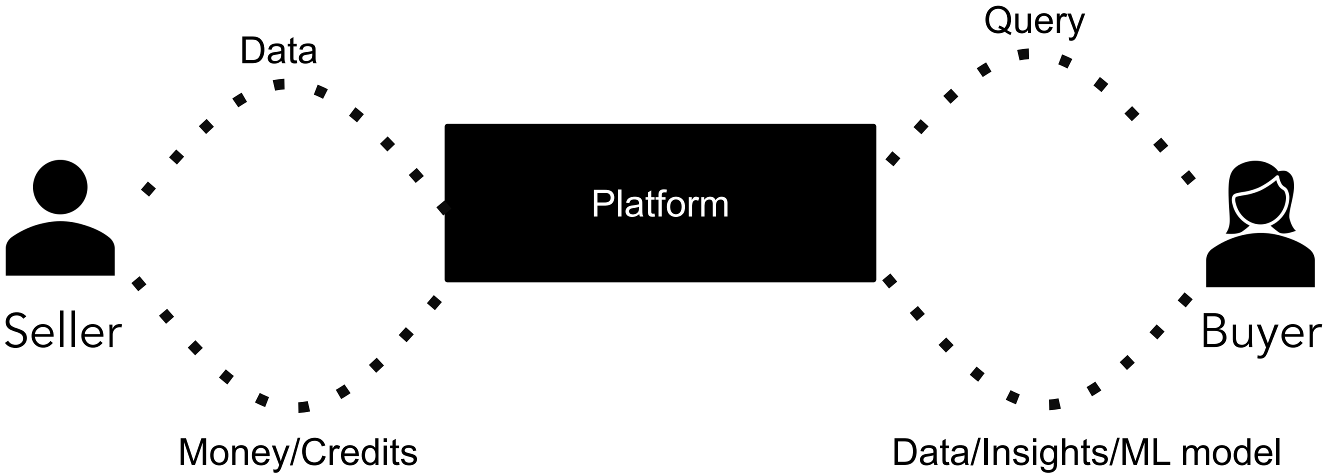
Daniel Alabi, Sainyam Galhotra, Shagufta Mehnaz, Zeyu Song, Eugene Wu

SIGMOD 2025 Tutorial

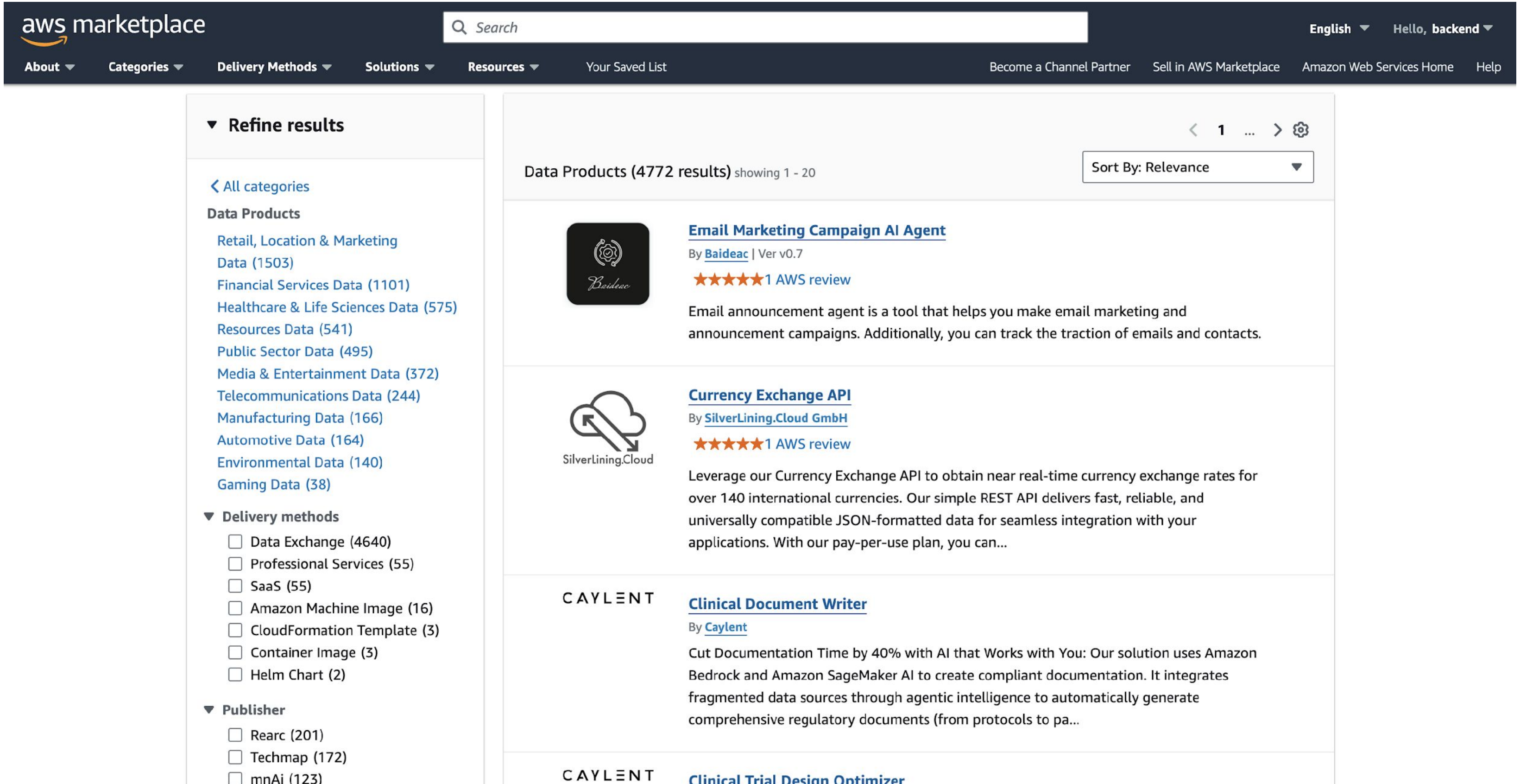
Overview of Data Markets

What is a data market?

A platform where data is bought, sold or exchanged (much like a traditional marketplace)



Many types of data markets



Many types of data market

- Keyword search over repositories
- Clean rooms
- Data labeling market
- Synthetic data market
- Curated alternative data

Keyword/NL Search

Keyword search over a table repository

- Index metadata or embeddings
- Return tables or links to tables

OpenData (data.gov), Academic (ICPSR, Dryad)

- Returns tables

Huggingface

- Returns models or training data

Google, Snowflake Marketplace, …

- Returns links

Enterprise Data Clean Rooms

Secure, privacy-preserving joins between orgs

- SQL aggregation over shared schemas
- Supports collaborative analytics (advertiser + publisher)

Example: Snowflake Clean Room

- Walmart w/ loyalty program & in-store purchases
- Discover w/ transaction & demographic data
- Can’t share raw customer data (PII)
- Join anonymized keys mediated by clean room
- Compute sales lift, cross-channel attribution

Others: AWS Cleanroom, BigQuery, InfoSum

Data Labeling Markets

Acquire labels for training models

- User provides task, data, instructions, and goal schema
- Workers complete tasks, checked with reviewers/algorithms
- Pay for high quality labels

Examples: Scale AI, Sama, Surge AI

- Waymo has millions of raw LiDAR frames
- Wants 3D bounding boxes, semantic segmentation
- Submits task definitions and raw data to Scale AI platform
- Labelers + AI-assisted workflows produce structured annotations
- Outputs used to train NN models

Synthetic Data Markets

Simulate real data without exposing real records

- User uploads data
- Train on secure platform
- Return synthetic data/model
- Used for testing, demos, edge cases, sharing

Examples: Gretel.ai, MostlyAI

- LendingClub has loan applications (income, SSN, credit)
- Can’t share or use raw data for model testing due to compliance
- Uploads sample data to generate synthetic dataset
- Uses output to train and validate credit risk models internally

Data Brokers

Curates data about sectors, companies, metrics, tickers, …

- Sources from web & vendors
- Reduce noise, integrate, clean, enforce schema, align w/ business concepts
- Sells datasets, subscriptions to data feeds, or faceted/keyword access

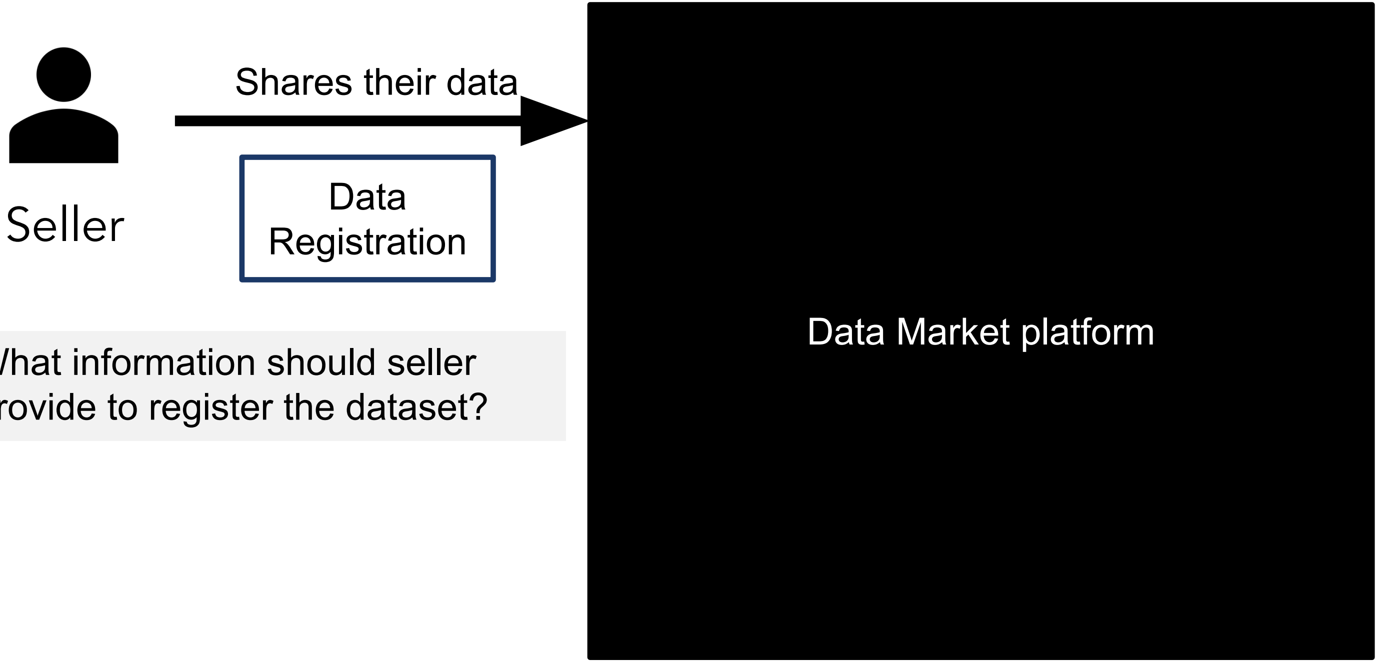
Example: Thinknum

- Crawls web: hiring pages, app store rankings, product pricing, retail inventory
- Differences data day-to-day
- Sells cleaned data feeds of changes e.g., Walmart + sales job postings

Others: Acxiom, Nielsen, Bloomberg, Morningstar, YipitData
https://oag.ca.gov/data-brokers

Category	Example	Query	Discovery	Incentive	Output
Clean Rooms	AWS Clean Rooms	SQL	Invite/catalog	Mutual value	Aggregated results
Labeling Markets	ScaleAI	Task	API	Payment	Labeled data
Alternative Data	Thinknum	Topic	Catalog/team	Subscription	Curated tables
Open Data Portals	NYC Open Data	Keyword	Tags/Portal	Public value	CSVs / APIs
Dataset Search	Google Dataset Search	NL (Keyword)	Metadata indexing	Visibility	External links
Model-as-Data	Hugging Face Datasets	Task	Benchmarks/Tags	Citation	Task-ready datasets
Academic Data	ICPSR	Structured	Metadata schema	Citation	Research tables
Synthetic Data	Gretel.ai	Schema	API	Privacy	Synthetic tabular data

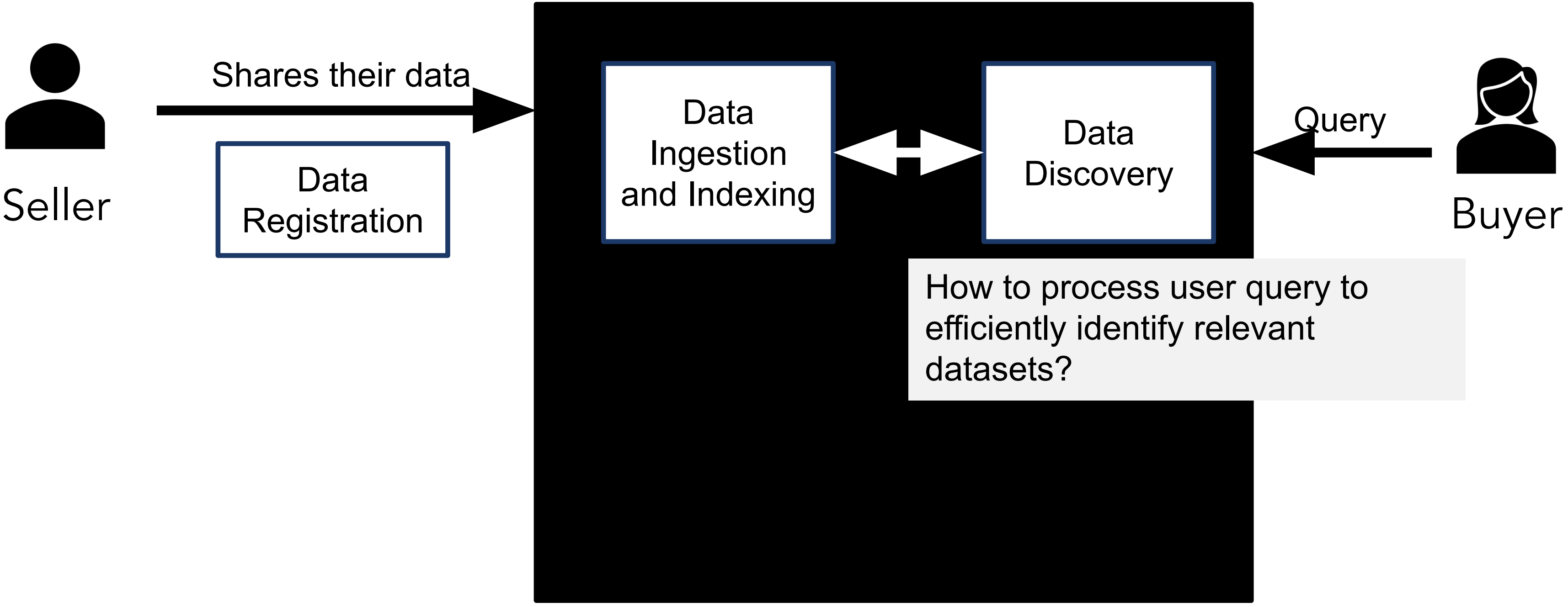
Key Components of a Data Market



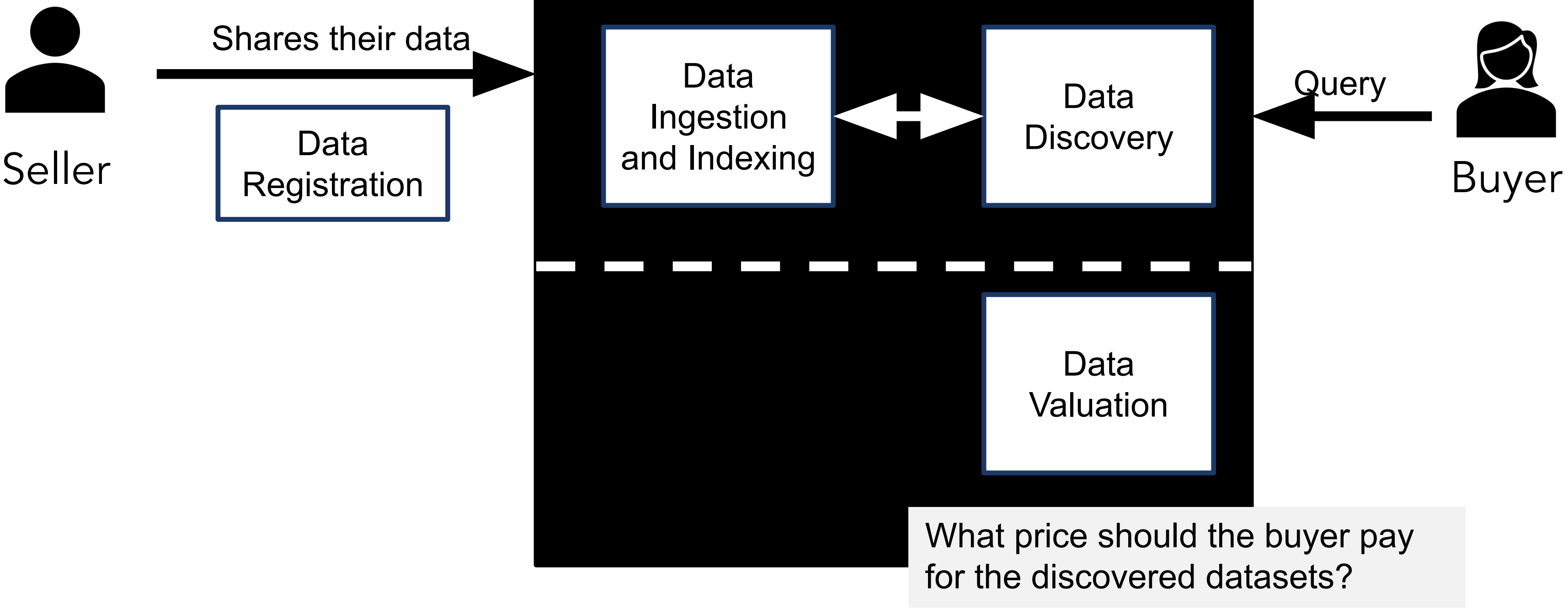
What information should seller provide to register the dataset?

Key Components of a Data Market

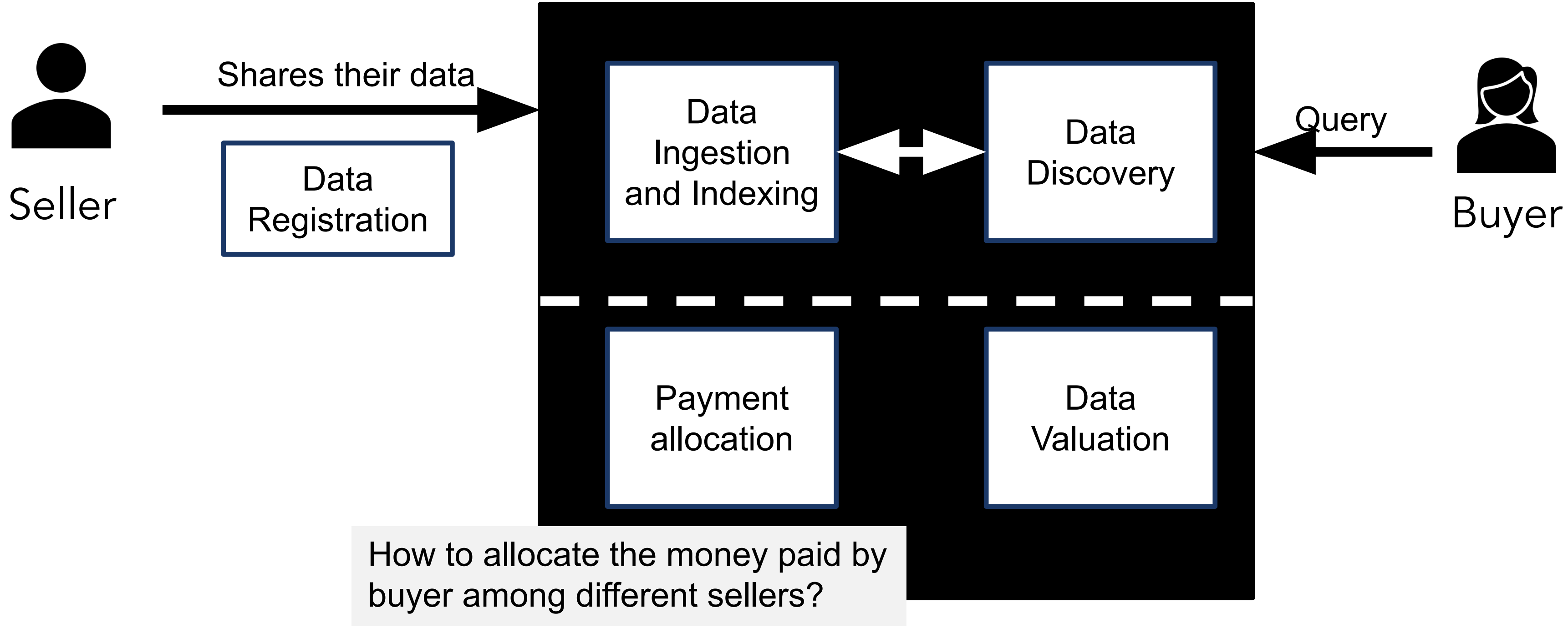
Key Components of a Data Market



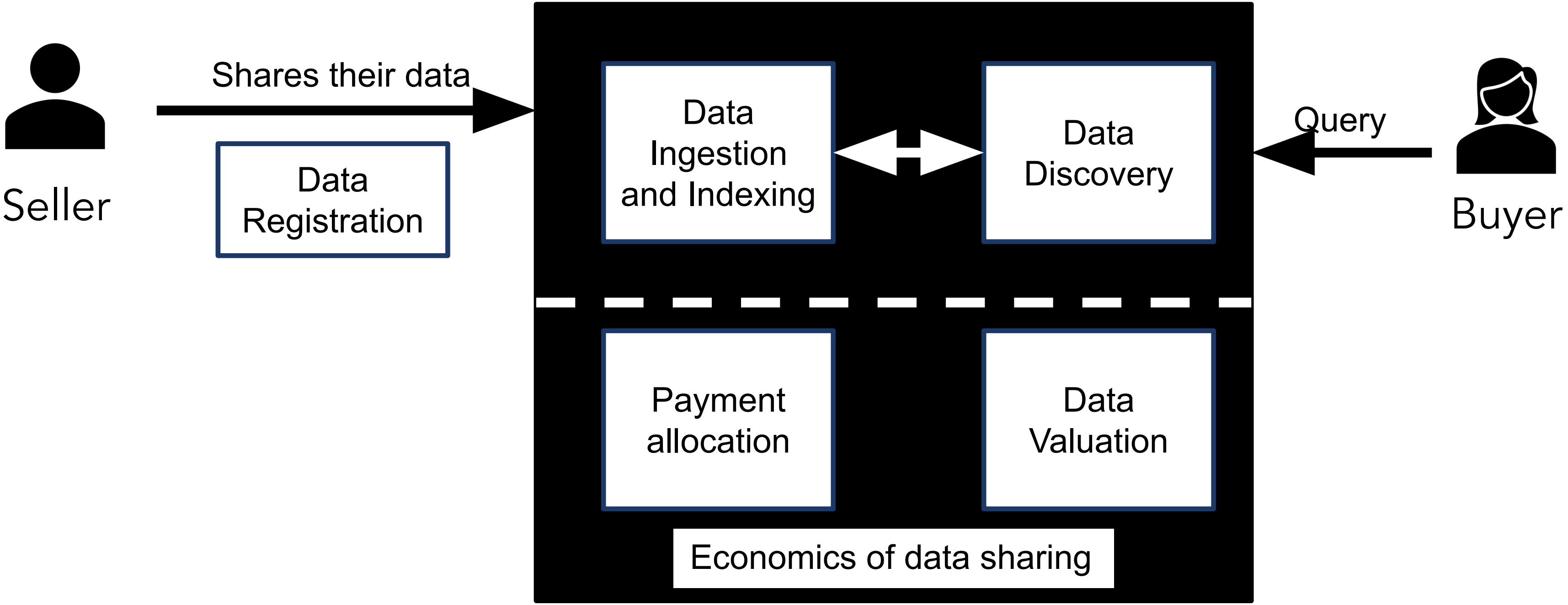
Key Components of a Data Market



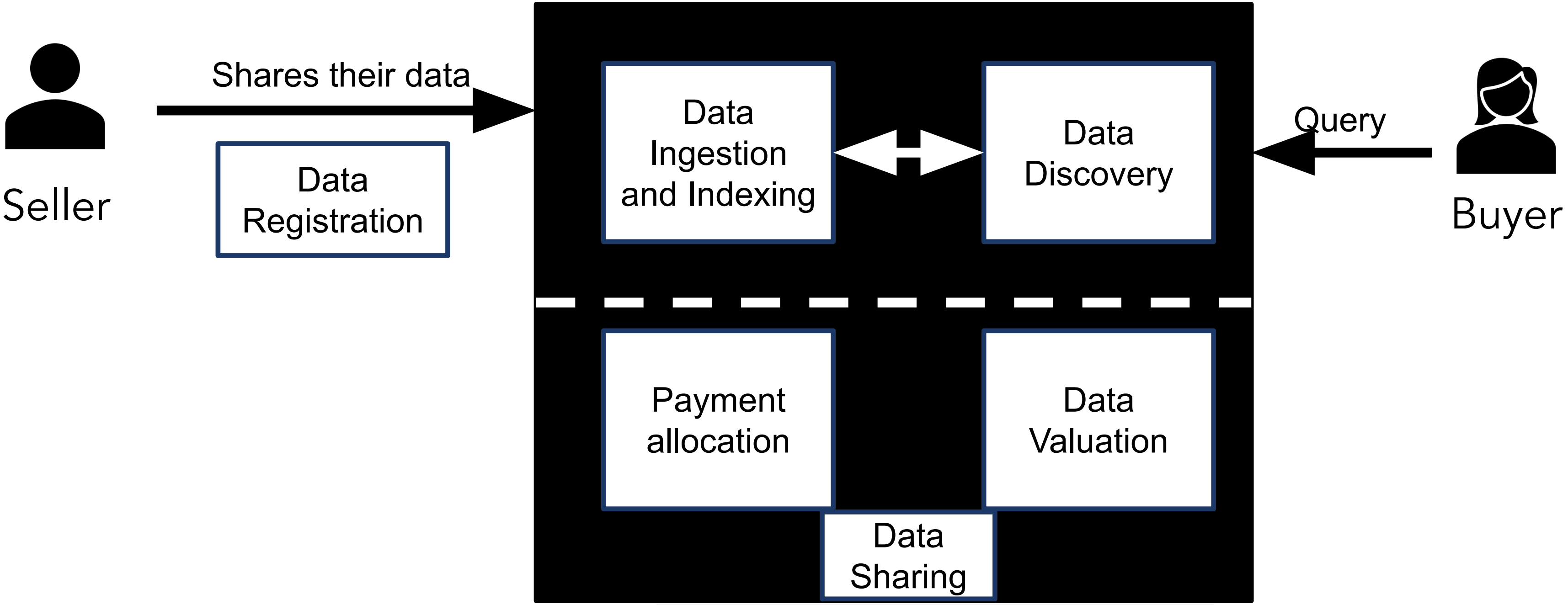
Key Components of a Data Market



Key Components of a Data Market

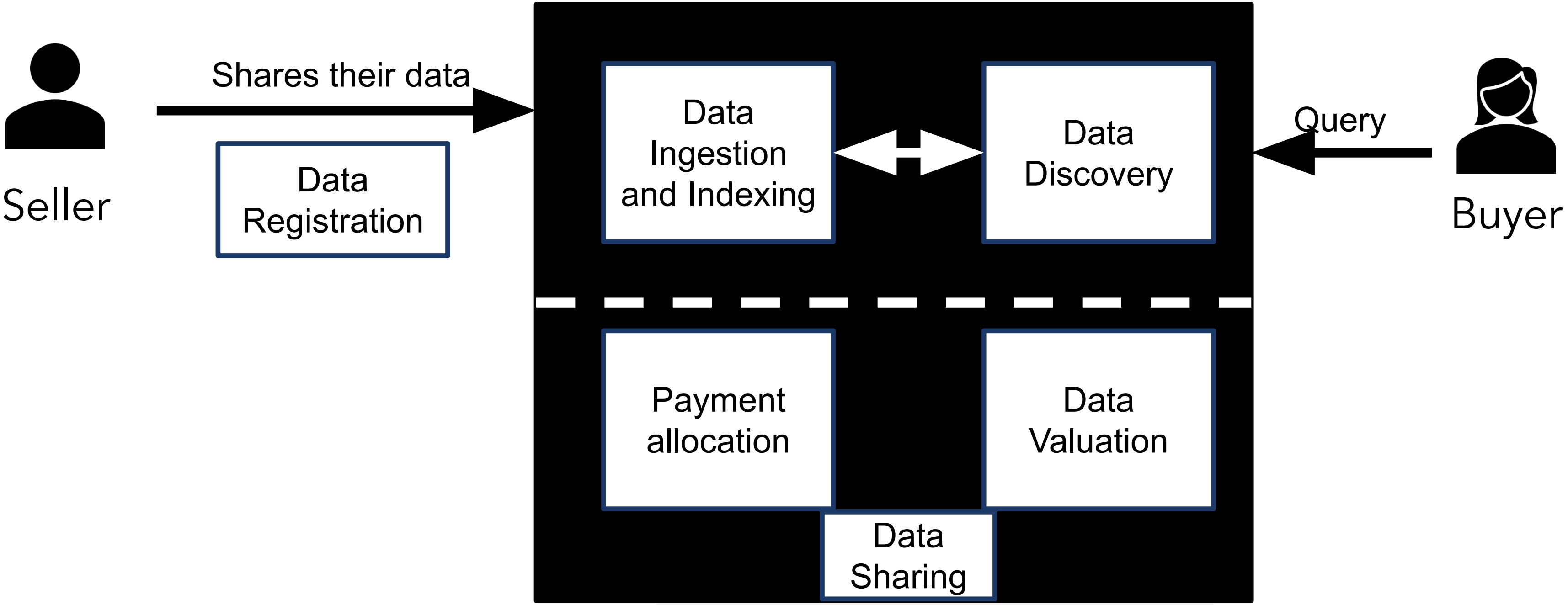


Key Components of a Data Market



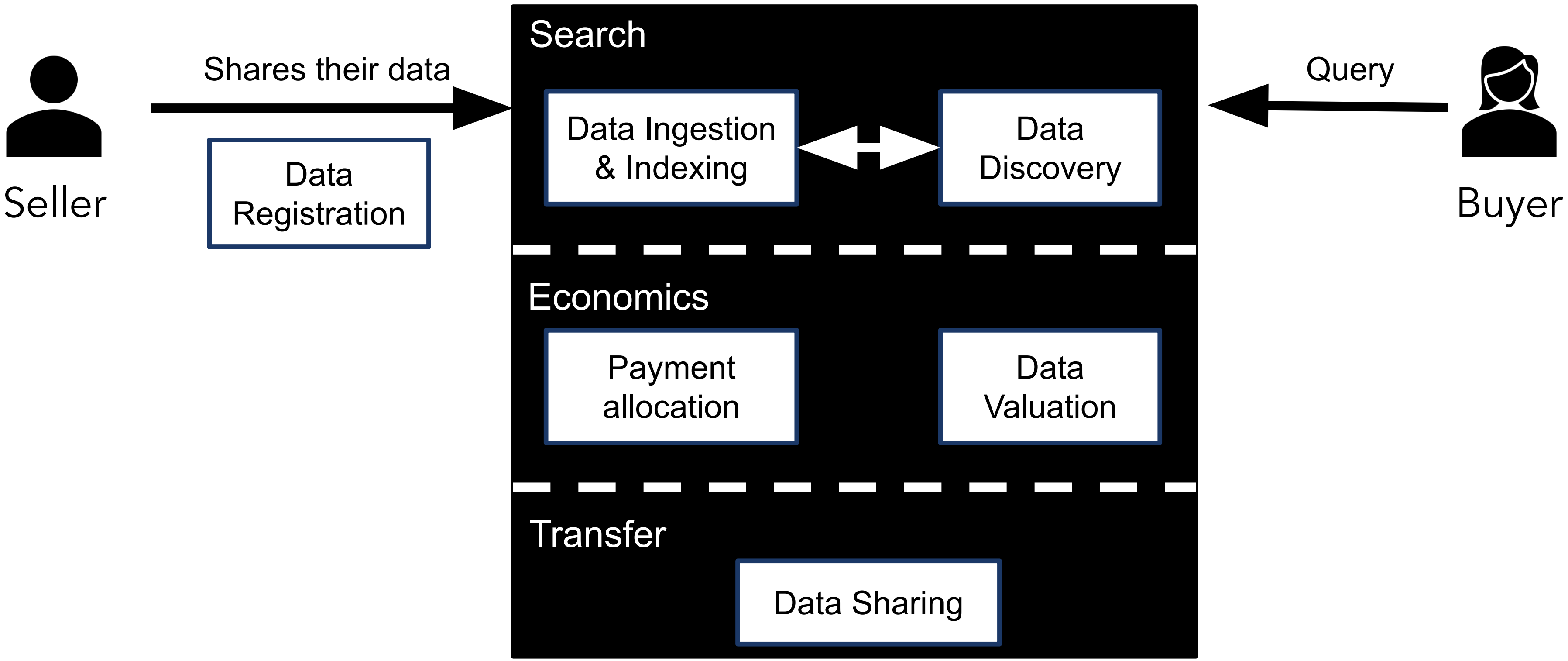
How to share data with the buyer?

Key Components of a Data Market



How to share data with the buyer?

Key Components of a Data Market



Summary: Challenges of a data market

- Data Registration and Discovery:
  - What information should a seller provide?
  - How to store these datasets?
  - How to efficiently discover datasets for a buyer?
- Data Sharing (or acquisition)
  - Arrows information paradox
  - What does the seller get? How is the final dataset shared?
- Data valuation:
  - How to price datasets?
- Payment allocation
  - How to allocate the money paid by the buyers amongst the sellers

Systems challenge

Economics challenge

Focus of this tutorial

How to ensure security and privacy?

- Protect buyers from malicious sellers
- Protect sellers from malicious buyers
- Prevent *unauthorized* users from accessing:
  - Seller private data
  - Buyer private data
  - Platform private data
- Prevent manipulation of data acquisition mechanisms:
  - Data discovery
  - Data valuation
  - Data negotiation
  - Data delivery

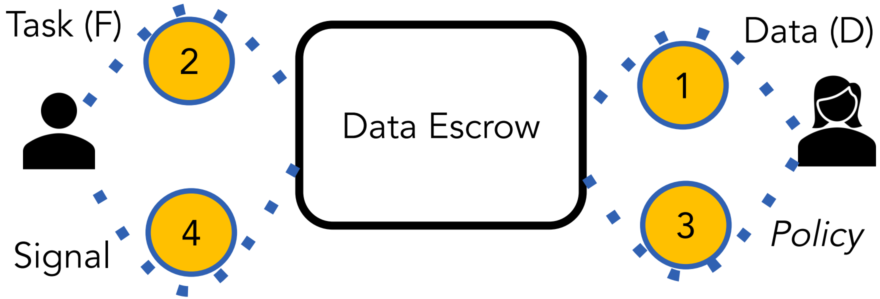
Tutorial Organization

- Part I: Data acquisition and search (Eugene and Sainyam)
- Part II: Privacy and Security Risks (Daniel)
- Part III: (Shagufta, Zeyu, and Eugene)
- Part IV: Regulatory Considerations (Daniel)
- Part V: Open Questions (Daniel, Eugene and Sainyam)

How to control what buyers can acquire?

Data Escrow [VLDB'22]

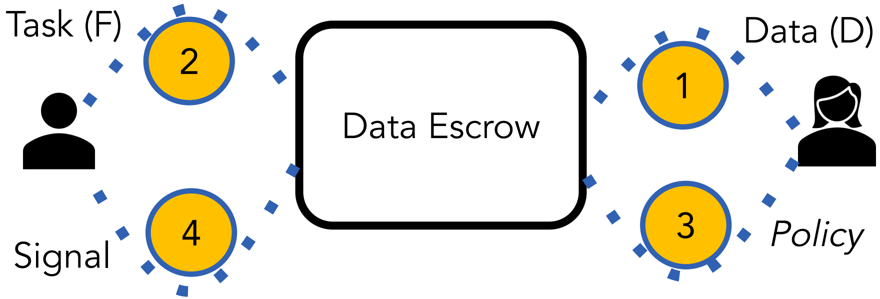
- A software system that controls dataflows
  - Sellers send their data; buyers send their tasks
  - Escrow runs buyers' tasks on seller's data



Slides borrowed from Raul Castro Fernandez, an author of this paper

Data Escrow [VLDB'22]

- A software system that controls dataflows
  - Sellers send their data; buyers send their tasks
  - Escrow runs buyers' tasks on seller's data

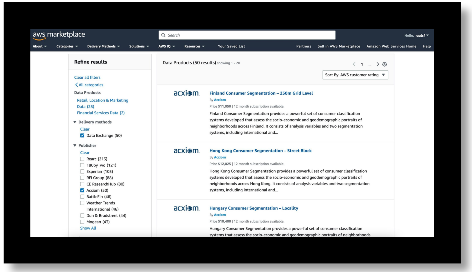


- <u>Guarantee</u>: no data* leaves the escrow without explicit permission, i.e., without an explicit *policy*

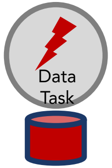
Using the Escrow to *Signal* Dataflow results

Data Markets

Seller



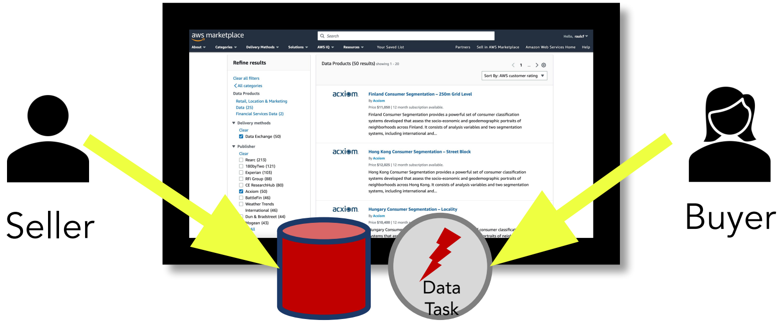
Buyer



With my data,
Accuracy: 0.63

Using the Escrow to *Signal* Dataflow results

Data Markets



With my data,
Accuracy: 0.63

Using the Escrow to *Signal* Dataflow results

With my data,
Accuracy: 0.63

With market data,
Accuracy: 0.94

Using the Escrow to *Signal* Dataflow results

Data Markets

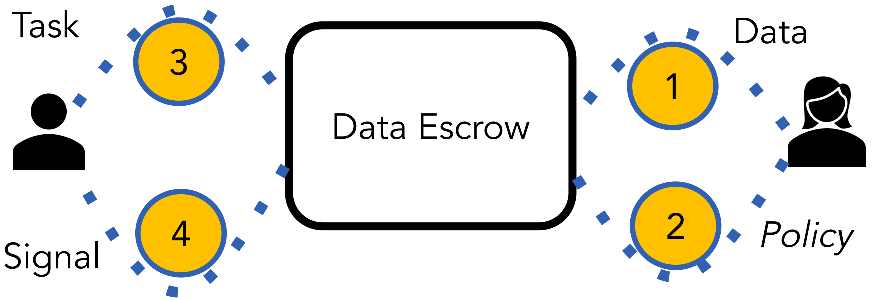
With my data,
Accuracy: 0.63

With market data,
Accuracy: 0.94

Market (escrow) signals the accuracy is 0.94. Buyer learns (0.94-0.63) is the benefit of buying

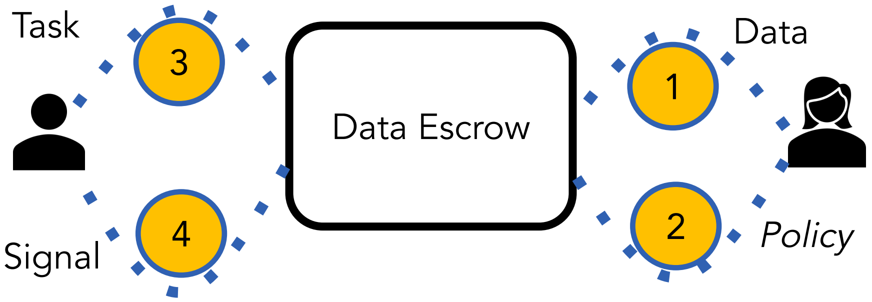
How do we delegate tasks, create signals,
i.e., how do we <u>control dataflows</u>?

Programmable Dataflows



Programmable Dataflows

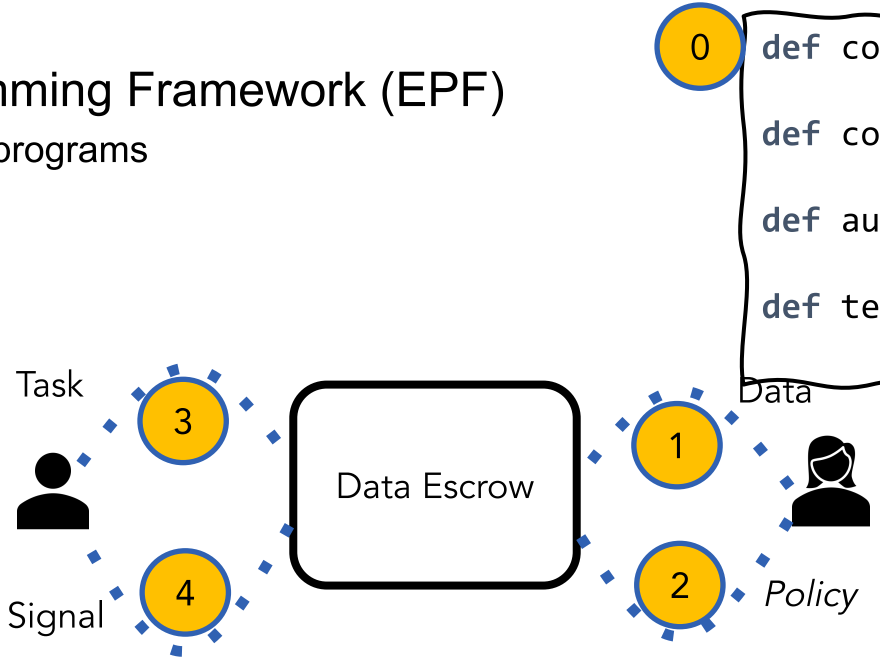
- Escrow Programming Framework (EPF)



Programmable Dataflows

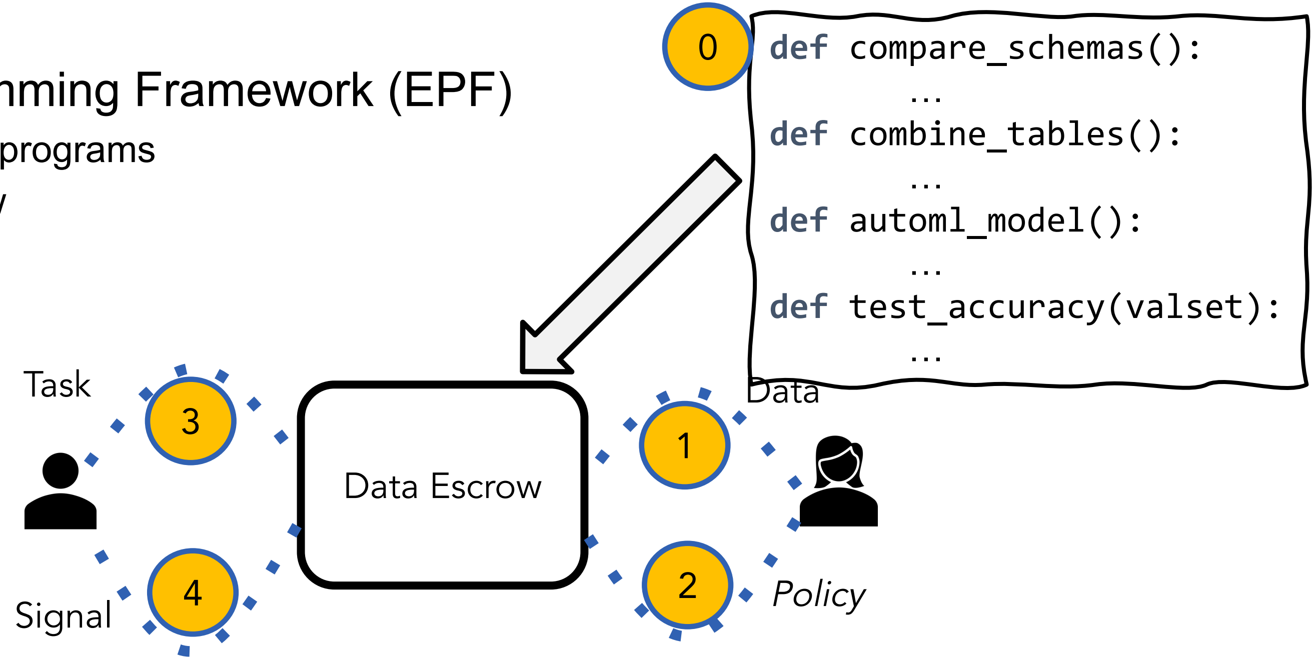
- Escrow Programming Framework (EPF)
  1. Developers write programs

```
def compare_schemas():
        ...
def combine_tables():
        ...
def automl_model():
        ...
def test_accuracy(valset):
        ...
```



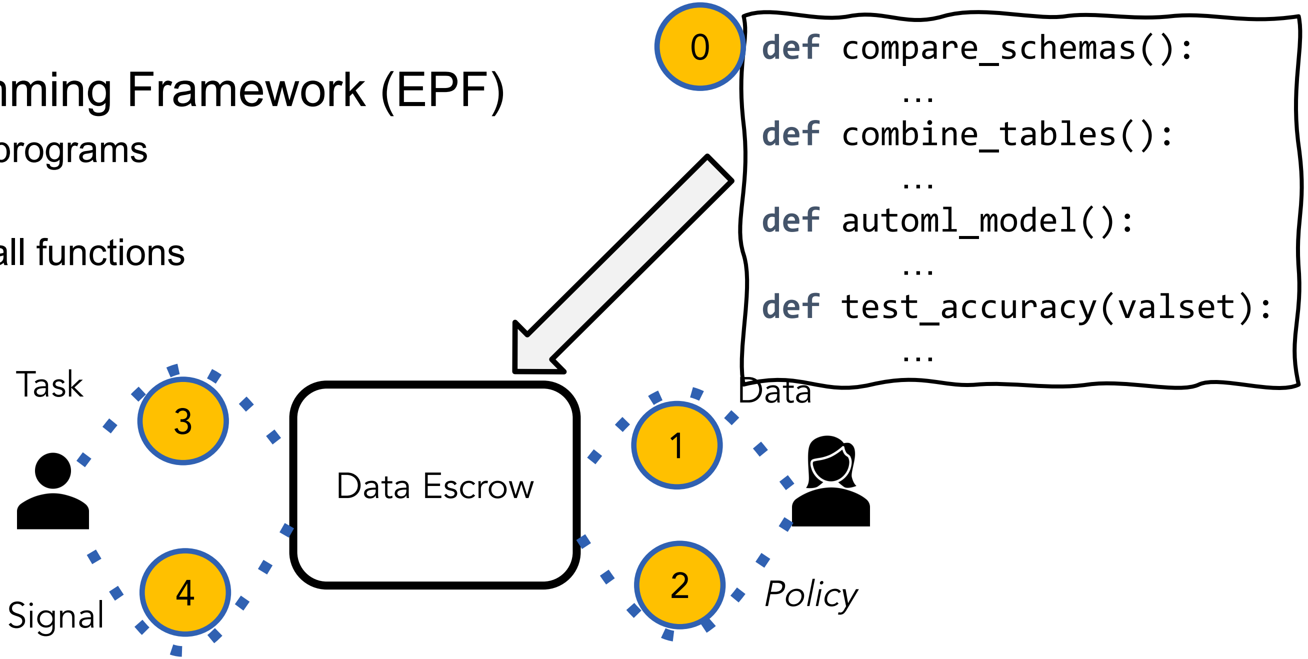
Programmable Dataflows

- Escrow Programming Framework (EPF)
  1. Developers write programs
  2. Deploy on escrow



Programmable Dataflows

- Escrow Programming Framework (EPF)
  1. Developers write programs
  2. Deploy on escrow
  3. Agents join and call functions

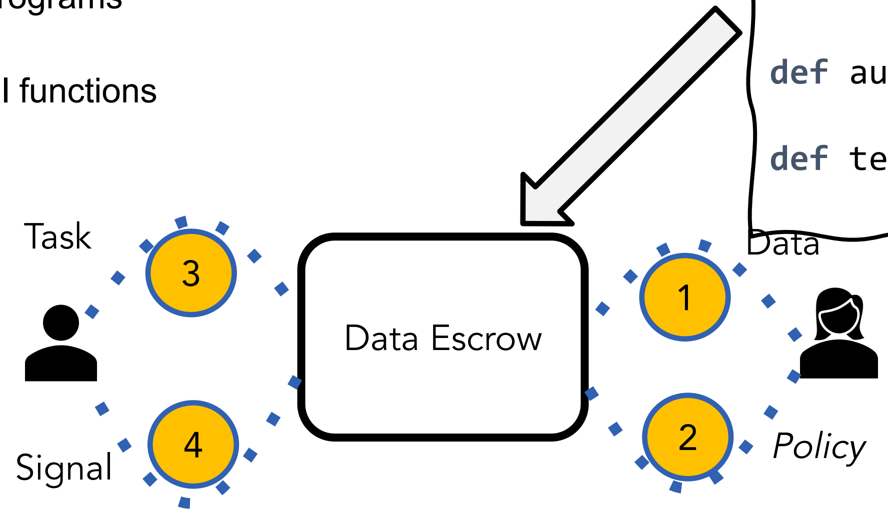


Programmable Dataflows

- Escrow Programming Framework (EPF)
  1. Developers write programs
  2. Deploy on escrow
  3. Agents join and call functions

0

```
def compare_schemas():
       ...
def combine_tables():
       ...
def automl_model():
       ...
def test_accuracy(valset):
       ...
```

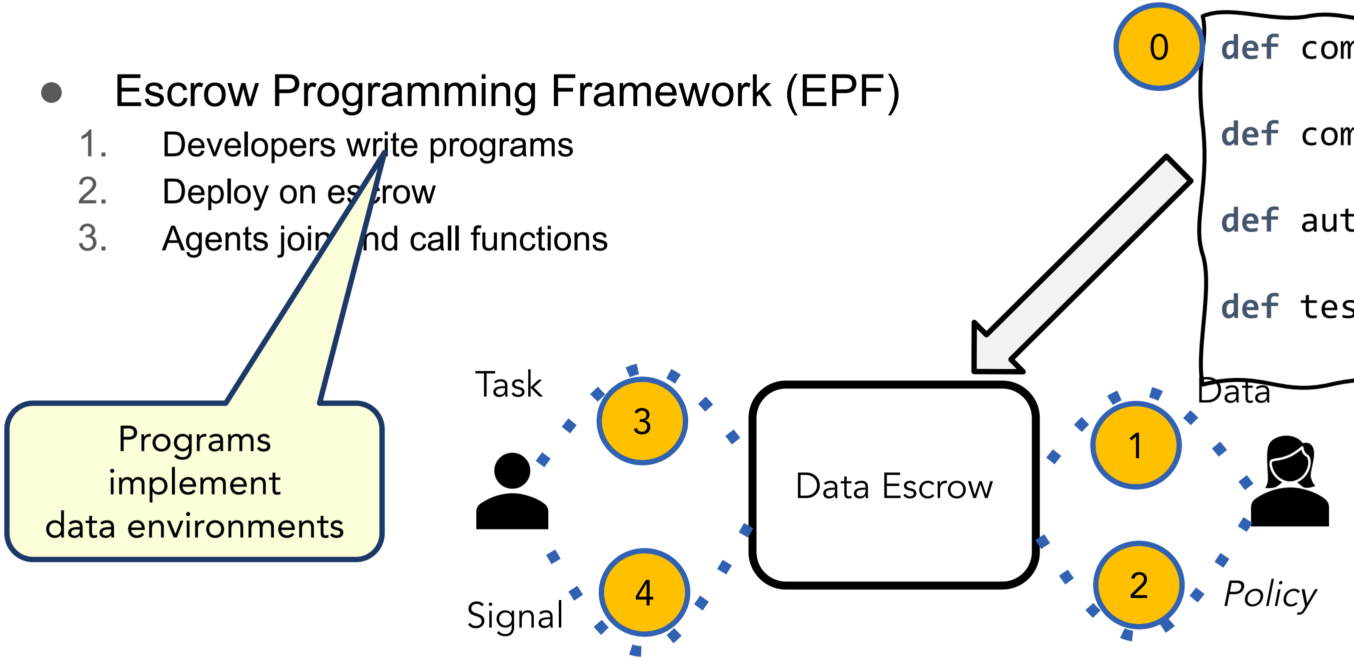


- Program implements communication and logic via *contracts*

Programmable Dataflows

- Escrow Programming Framework (EPF)
    1. Developers write programs
    2. Deploy on escrow
    3. Agents join and call functions

```
def compare_schemas():
    ...
def combine_tables():
    ...
def automl_model():
    ...
def test_accuracy(valset):
    ...
```

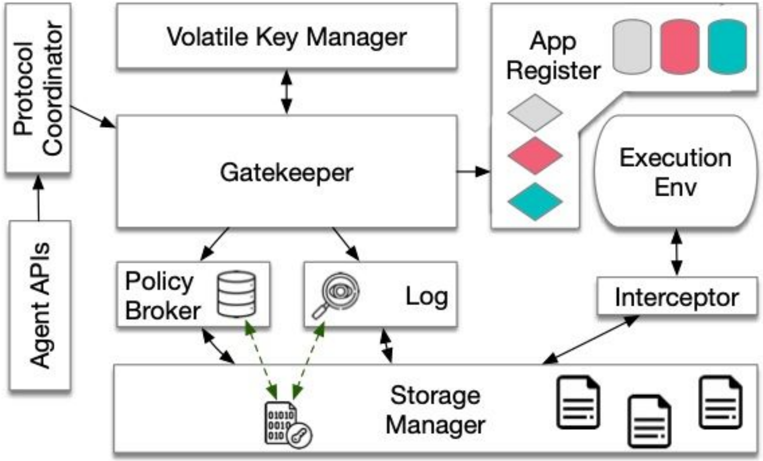


- Program implements communication and logic via *contracts*

Delegated, Auditable, Trustworthy

- What happens in the escrow, stays in the escrow
  - Except when it needs to be available to auditors and 3-party officers

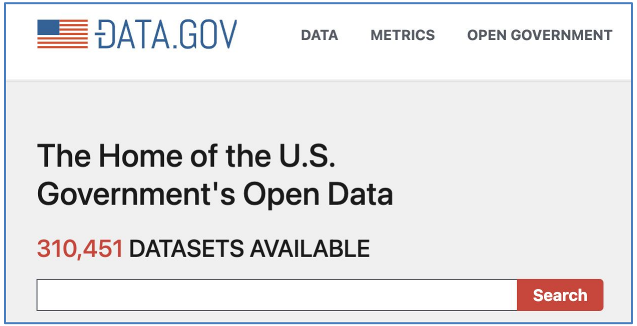
- Data is encrypted end-to-end
  - At rest *and* during computation
    - Use of secure hardware enclaves
  - Encrypted Write-Ahead Log (EWAL)
  - Cryptographic protocols for IO
    - Key exchange and recovery after failures…

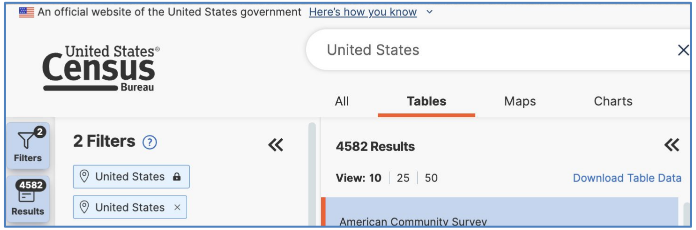


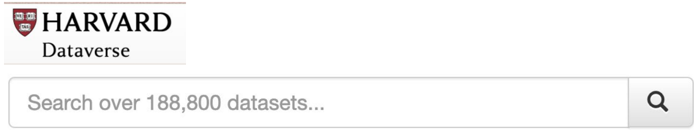
Data Search

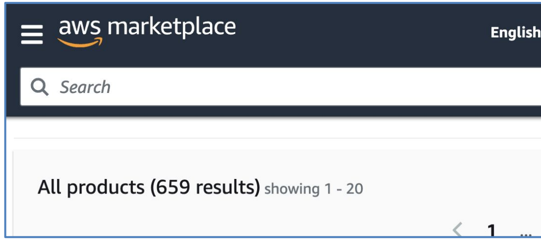
Unlimited Storage → Massive Data Repos

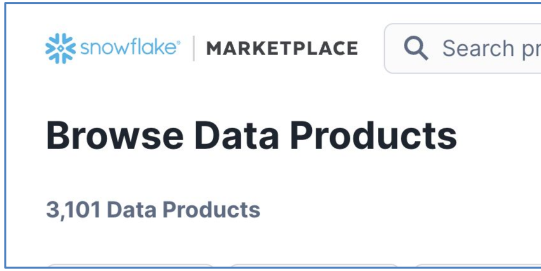
Gov Portals
Data Markets
Data Lakes
Web Tables
Data coalitions
…

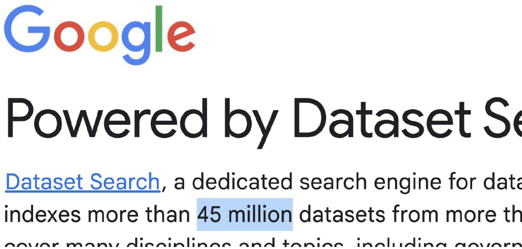












What Can We *Do* with 1M+ Tables?

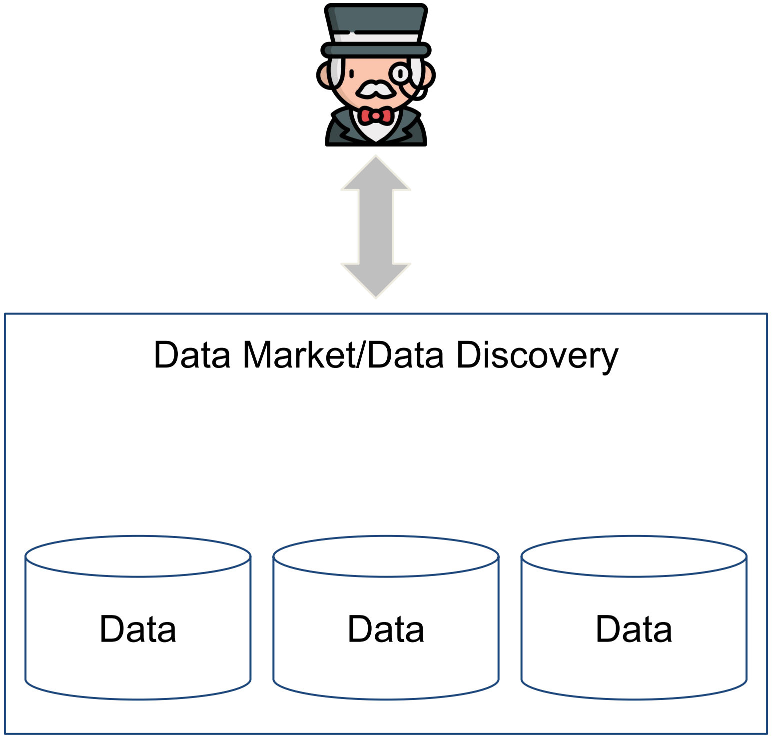
Gov Portals
Data Markets
Data Lakes
Web Tables
Data coalitions
…

?

Scientific phenomena
Economic theories
Investment hypotheses
Customer analysis
…

Step 1: Find Relevant Tabular Dataset

Centralized Data Search Systems

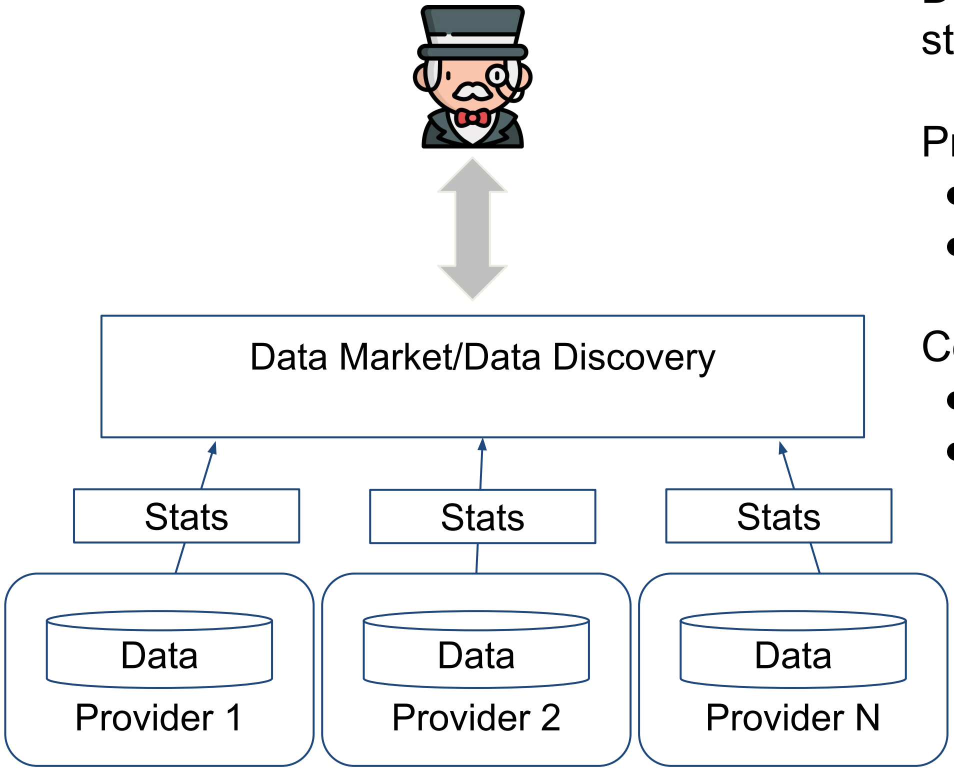


A single system stores & manages the datasets

Pros:
- Fits an organization’s data lake
- Easier access to raw data, experts, metadata
- Easier to tightly integrate with use cases

Cons
- Limited to a single organization

Decentralized Data Search Systems



Data is federated, and system has access to statistics rather than raw data

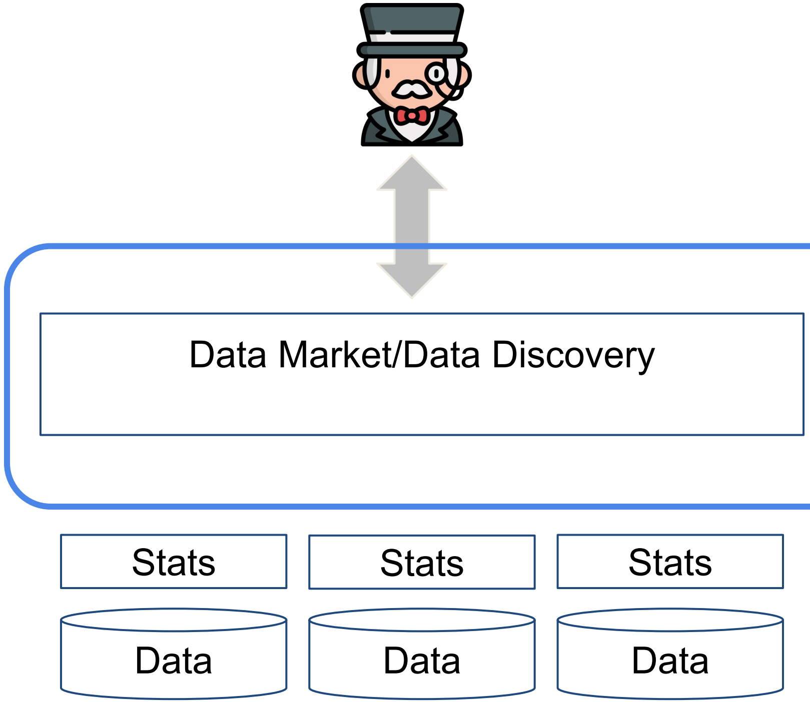
Pros

- Clear separation of privacy concerns
- More realistic for a public data market

Cons

- More difficult to provide utility
- Hard to manage multiple providers

Challenges in Data Search Systems



Query specification
Privacy protection

Query interface
Latency, Scalability
Privacy protection

Data acquisition
Data preparation
Privacy protection

3 Classes of Systems

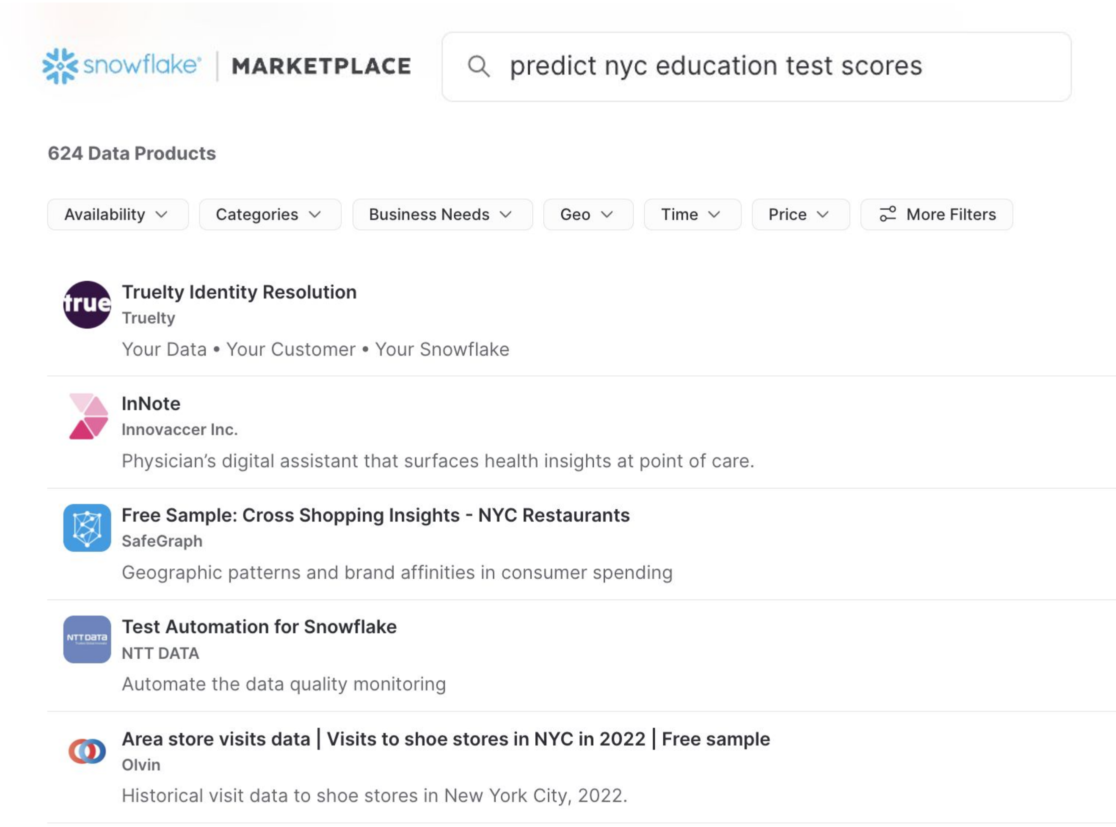
Keyword/Metadata Search

Data Discovery

Task-based Search

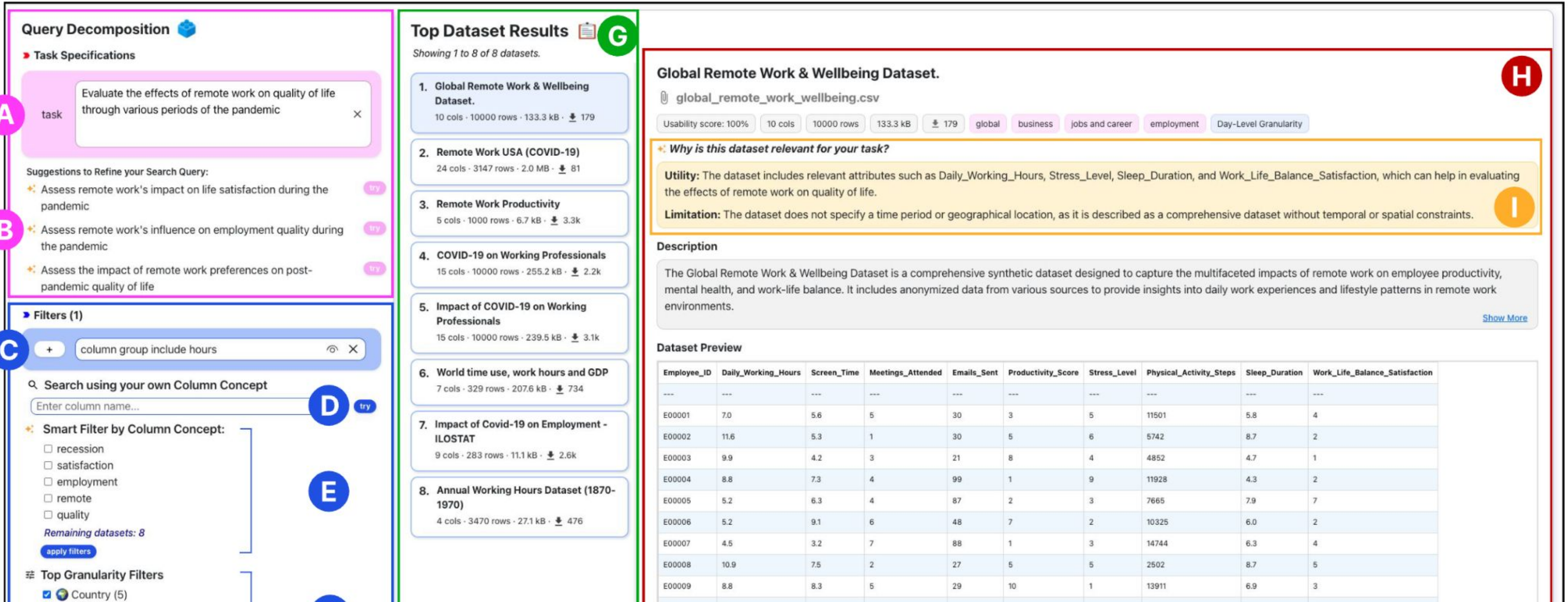
Keyword Search

narrative
snowflake
Datarade
data.gouv.fr
Google Dataset Search
DATA.GOV
NYC OpenData
City of London Open Data



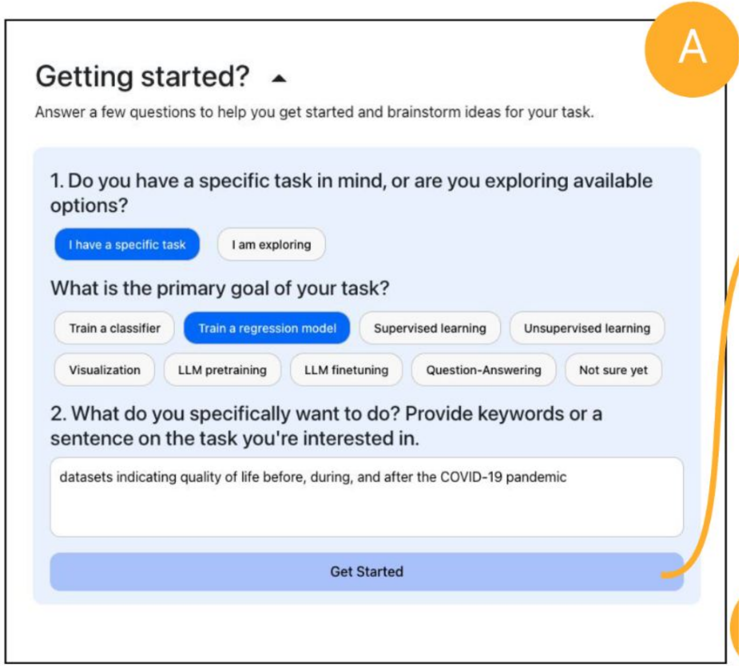
Keyword Search as Sensemaking

DataScout. Rachel Lin, Bhavya C., Wenjing L., Shreya S., Madelon H., Aditya P.



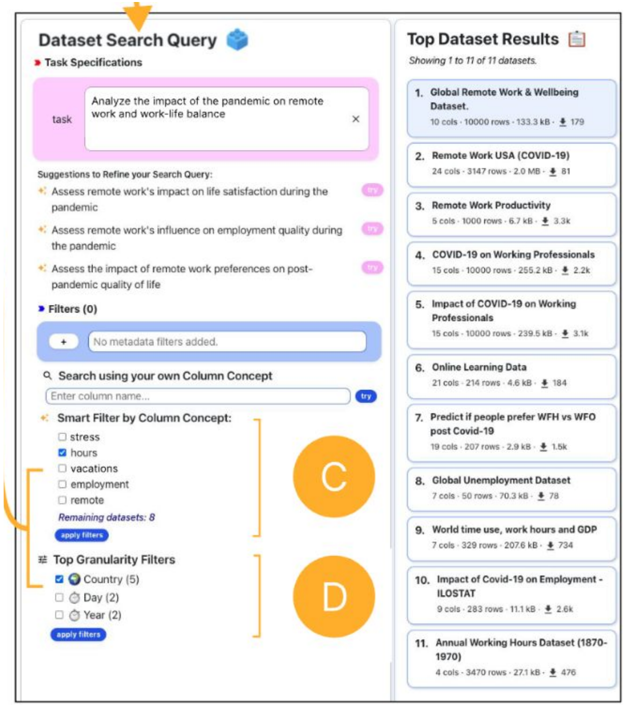
Keyword Search as Sensemaking

DataScout. Rachel Lin, Bhavya C., Wenjing L., Shreya S., Madelon H., Aditya P.



Keyword Search as Sensemaking

DataScout. Rachel Lin, Bhavya C., Wenjing L., Shreya S., Madelon H., Aditya P.



Keyword Search

Pros

- Fast, doesn’t need access to actual data
- Filters and ranks datasets
- Dominant data search approach today

Cons

- Users need to evaluate datasets against actual data task
- Users in the critical path of search

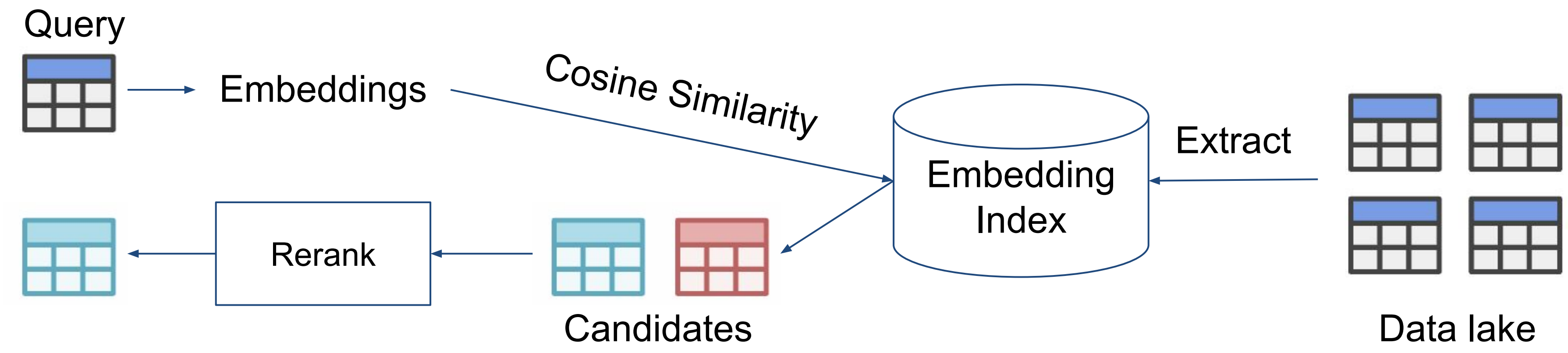
Data Discovery

Search by using a table or distribution as the query
Ling13,Zhu16,Nargesian18,Fernandez19,Rezig22,Santos21,Fan23,...

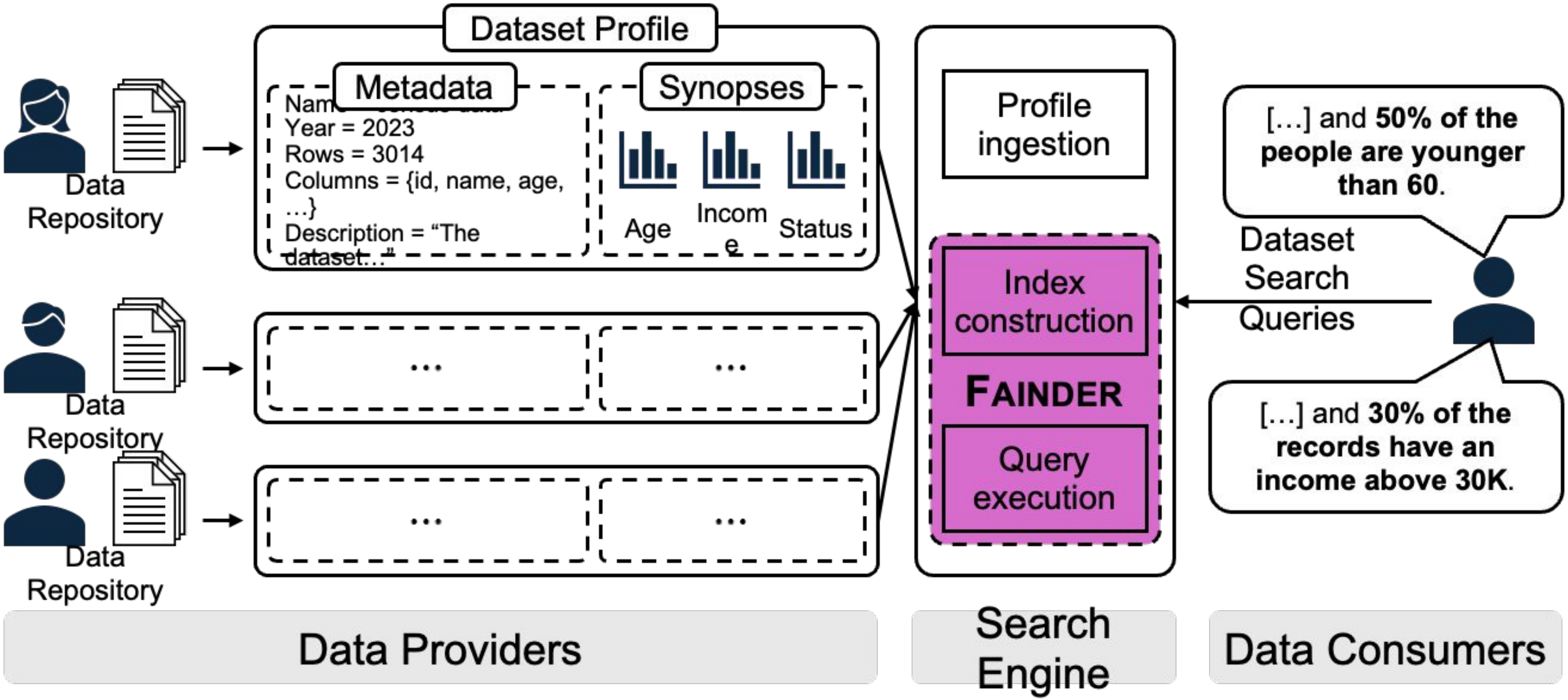
Rank based on

- Similarity,
- Joinability,
- Correlations,
- Unionability,
- Predicate satisfiability,
- …

Starmie: Table Union Search [Fan23]



Distribution-based Data Discovery [Behme24]



Data Discovery

Pros

- Results specific to the query table
- Scalable, leverages table representations

Cons

- Unless query is a retrieval task, users still need to evaluate datasets against actual data task
- Users in the critical path of search

Data Task as Search Query

Task $\mathrm{T(D)}\rightarrow\mathrm{goodness}$ is function of table D

Prediction ARDA, AUCTUS, Galhotra23

- $\mathrm{T(D)}$: train predictive model
- Given training dataset D, find augmentations that improve $\mathrm{T(D)}$

Causal Inference Suna Liu25, MetaM Galhotra23

- $\mathrm{T(D)}$: estimate Average Treatment Effect
- Given D with treatment and outcome, find likely confounders

Data Task as Search Query

Pros

- Ranks directly based on user's task
- Can incorporate cleaning, integration, transformation

Potential Cons

- Evaluating task can be slow
- Hard to quantify task quality

Two Examples of Task-Based Search

Based on

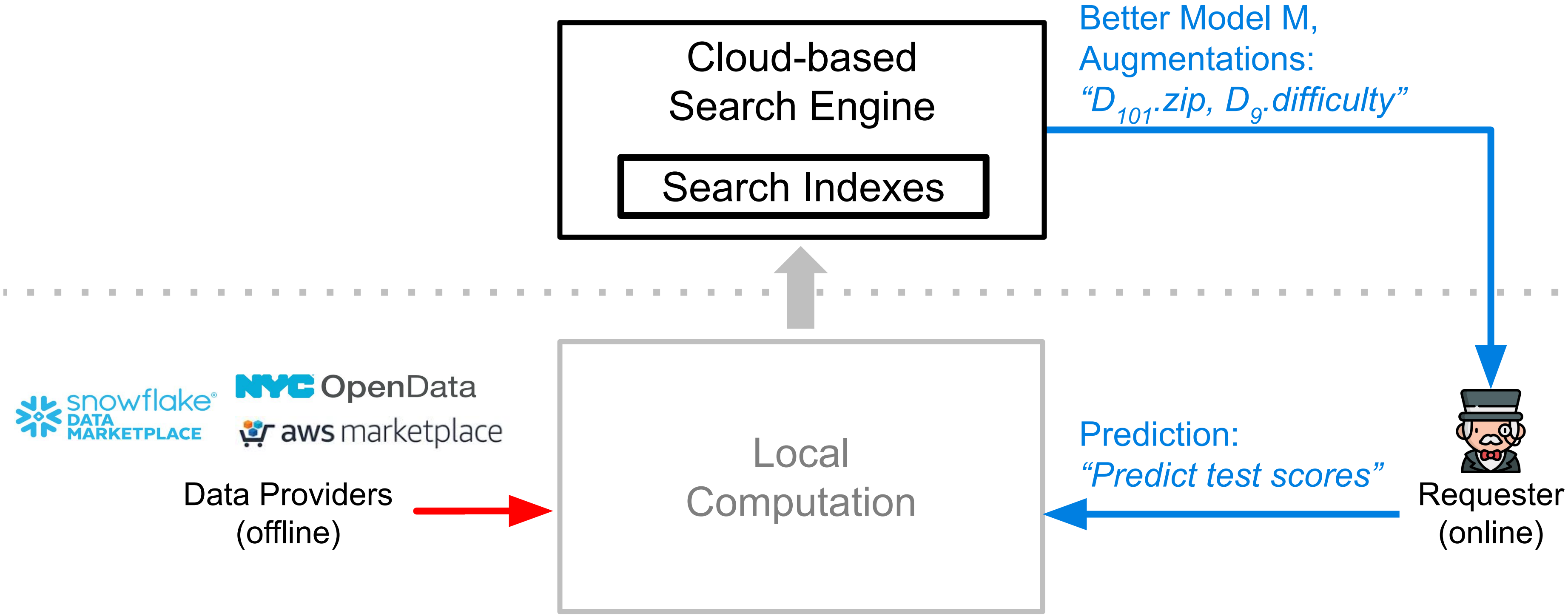
Kitana: A Data-as-a-Service Platform. Zach Huang23

The Fast and the Private: Task-based Dataset Search. Huang24

Saibot: A Differentially Private Data Search Platform. Huang23

Suna: Scalable Causal Confounder Discovery over Relational Data. Liu25

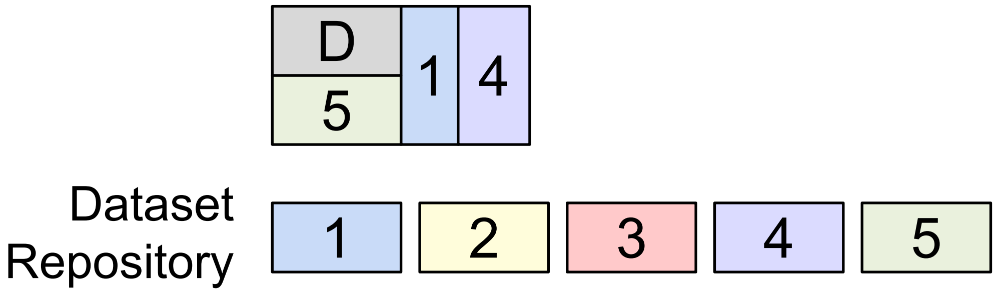
Data Task as Search Query



Prediction Task

Given training data D
greedily find augmentation plan A
that maximizes accuracy of model trained on A(D)

$$A(D) = D \cup 5 \bowtie 1 \bowtie 4$$



Basic Search Algorithm

```
D = initial training dataset
for A in all candidate augmentation plans
    eval(apply A to D)
return best A
```

Basic Search Algorithm

```
D = initial training dataset
for A in all candidate augmentation plans
    eval(apply A to D)
return best A
```

Basic Search Algorithm

```
D = initial training dataset
for A in all candidate augmentation plans
    eval(apply A to D)
return best A
```

Slow!

```
D = initial training dataset
for A in all candidate augmentation plans
    eval(apply A to D)
return best A
```

Combinatorial

Expensive!
Materialize A(D)
Retrain & Cross-validate

Reduce Search Space

- ARDA: join all relations + feature selection
- MetaM: cluster datasets and iteratively prune

Accelerate Eval()

- Auctus: find joinable correlations

Relies on access to raw data

Example Sysetm: Kitana

```
D = initial training dataset
for next augmentation α
    if eval(apply A to D) is best so far
        Keep α in A
return best A
```

Greedy Search

Expensive!
Materialize A(D)
Retrain & Cross-validate

Ideas

- Greedily find single best augmentation in each iteration
- Use sketches to accelerate & parallelize `eval()`

Sketches: Count(D⋈S)

Naïve join generates big intermediate relation

D

A	B
1	1
1	2
2	3

$\longrightarrow \bowtie \longleftarrow$

S

A	C
1	4
1	5
2	6

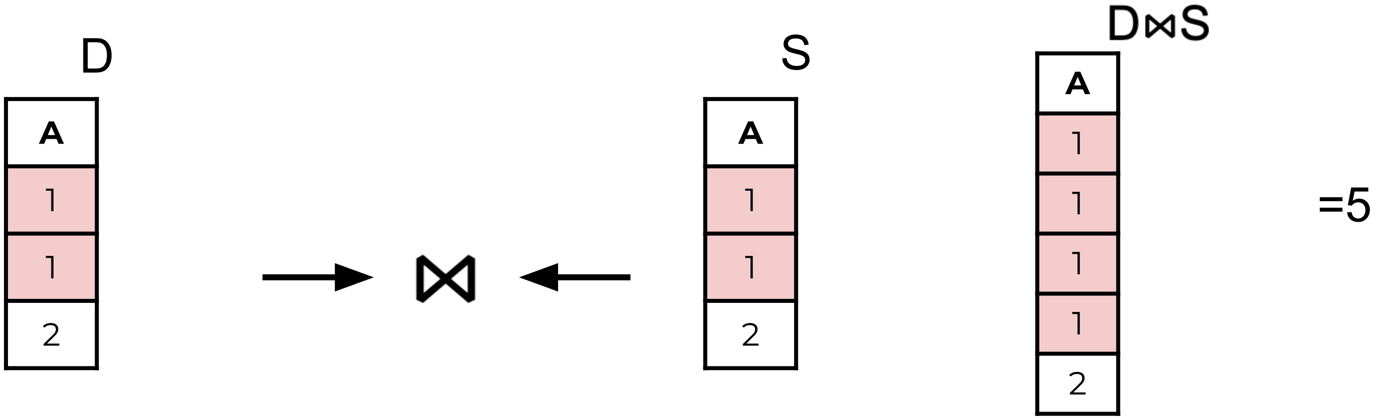
D⋈S

A	B	C
1	1	4
1	1	5
1	2	4
1	2	5
2	3	6

=5

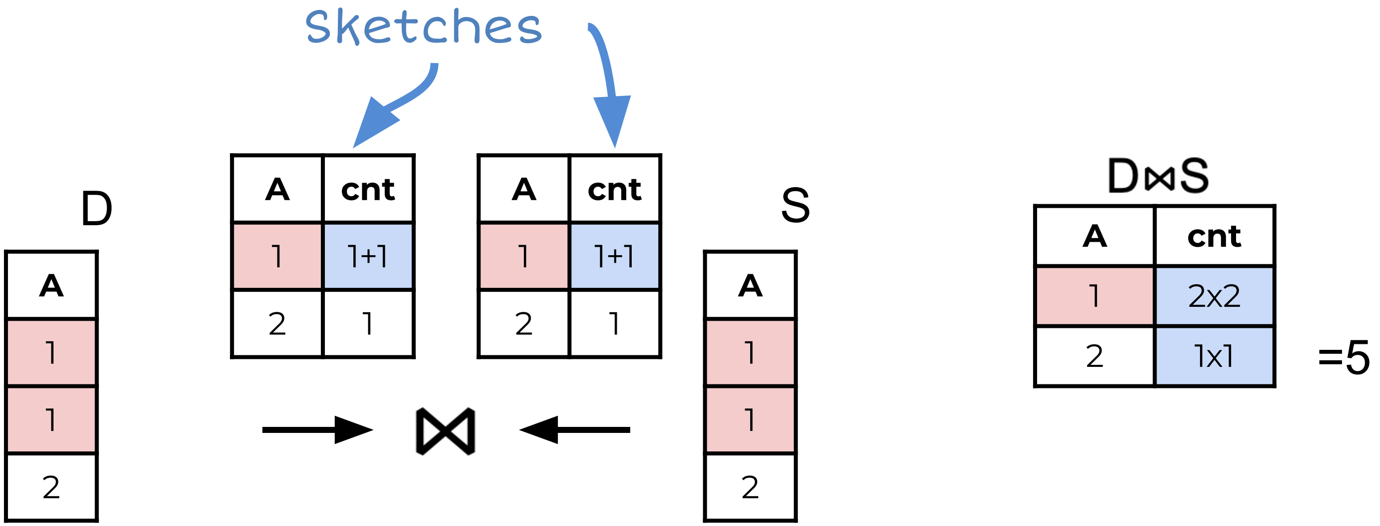
Sketches: Count(D⋈S)

Optimization: drop irrelevant columns



Sketches: Count(D⋈S)

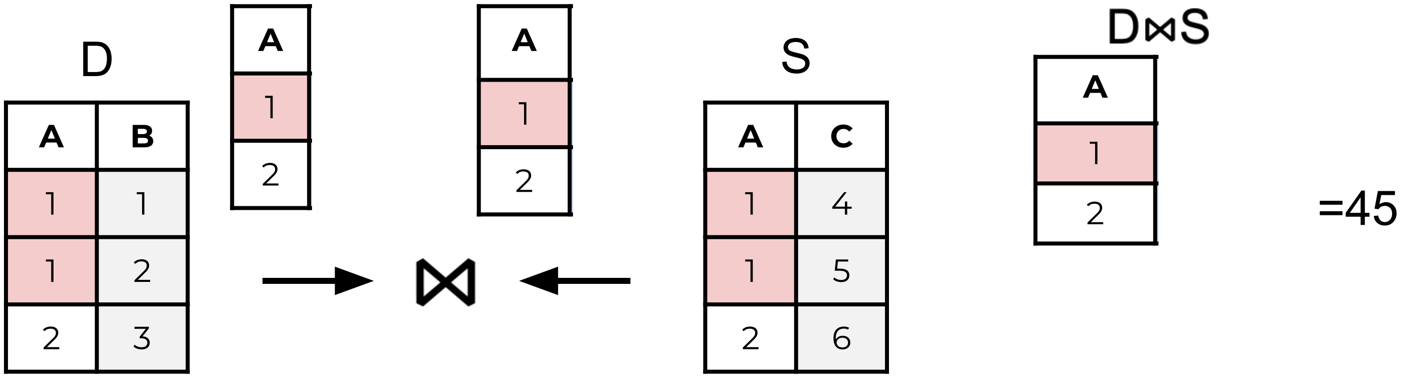
Optimization: sufficient statistics



Sketches: $\mathrm{Sum}_{b*c}(D \bowtie S)$

Optimization: sufficient statistics

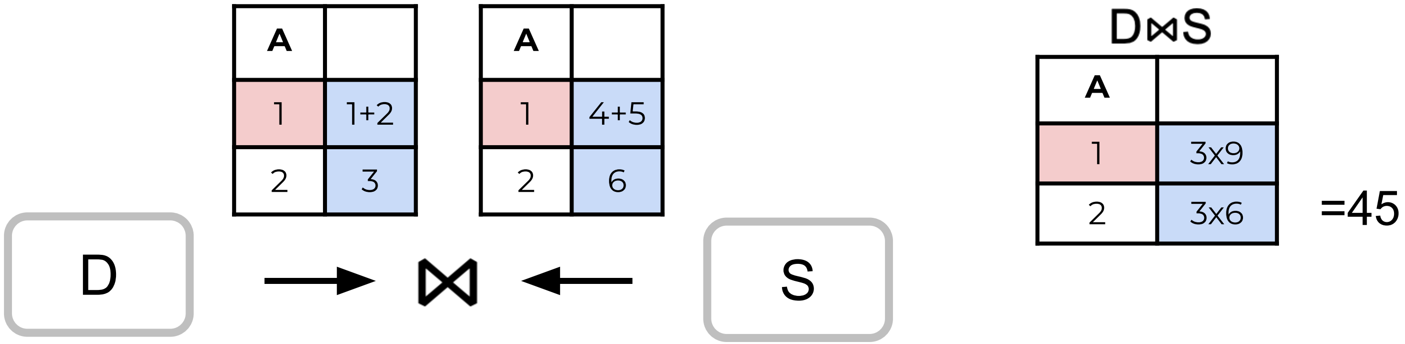
Sketches defined for common stats, ML models.



Sketches: trainAndEval(D⋈S)

Optimization: sufficient statistics

Sketches defined for common stats, ML models.

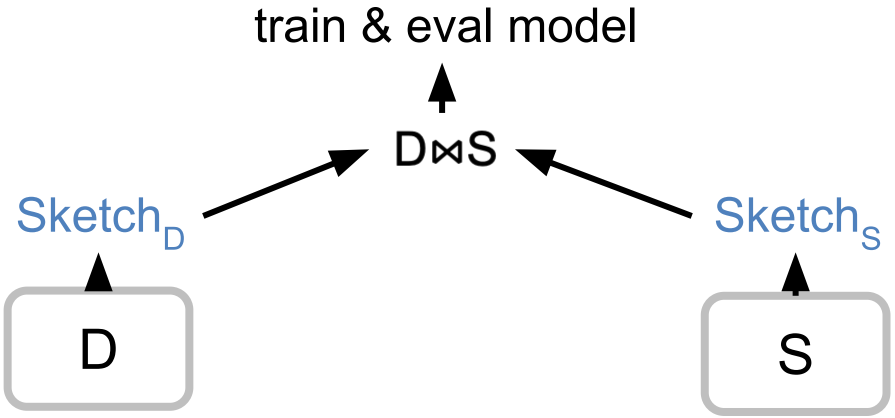


Sketches: train(D⋈S)

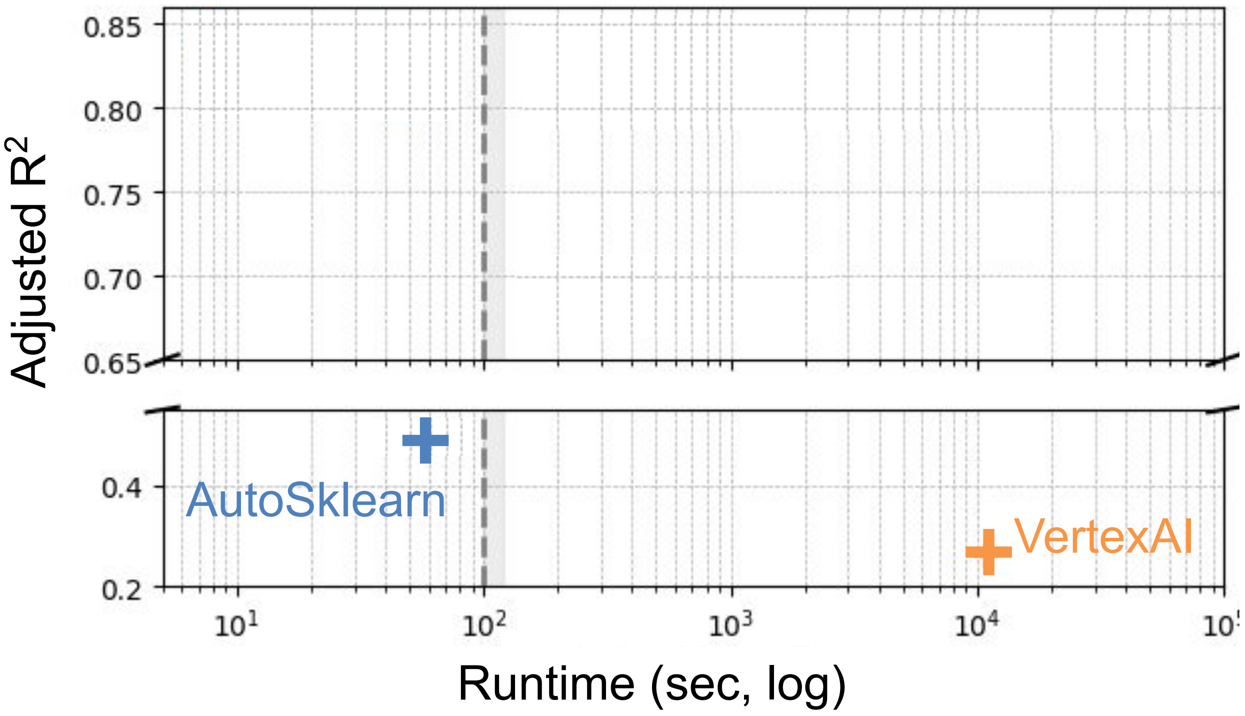
Optimization: sufficient statistics

Sketches defined for common stats, ML models.

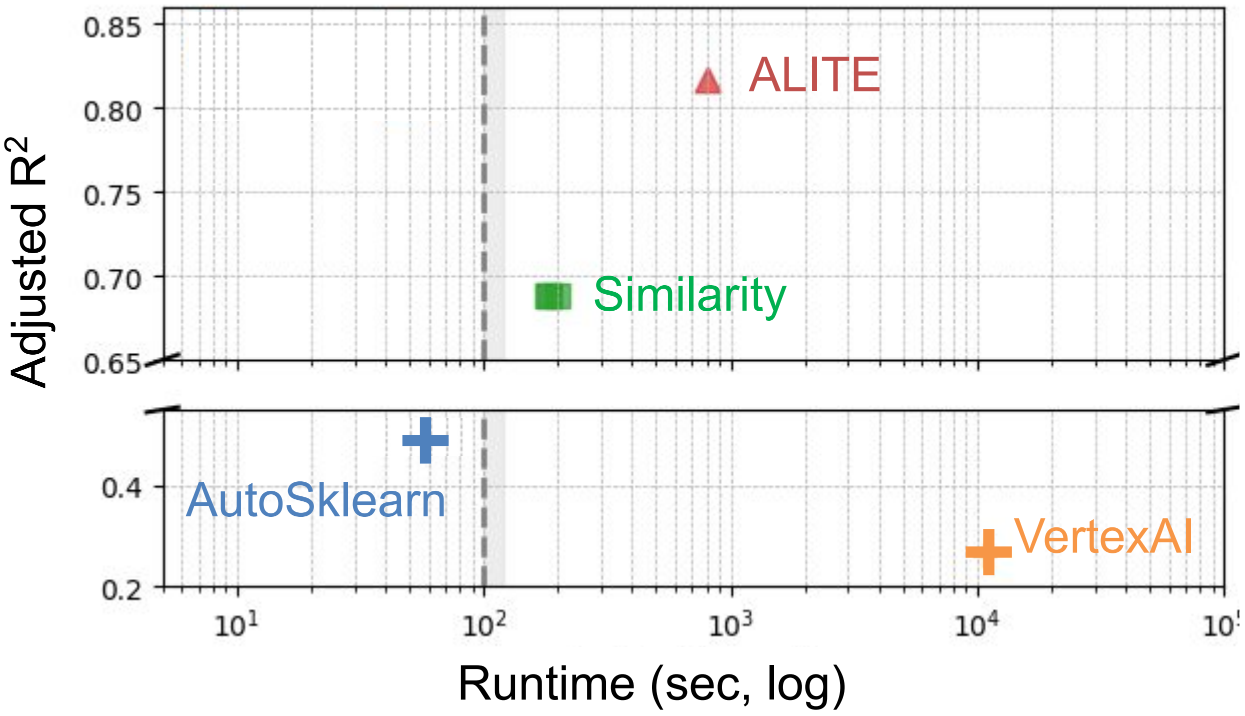
Linear Regression as a *proxy model* during search



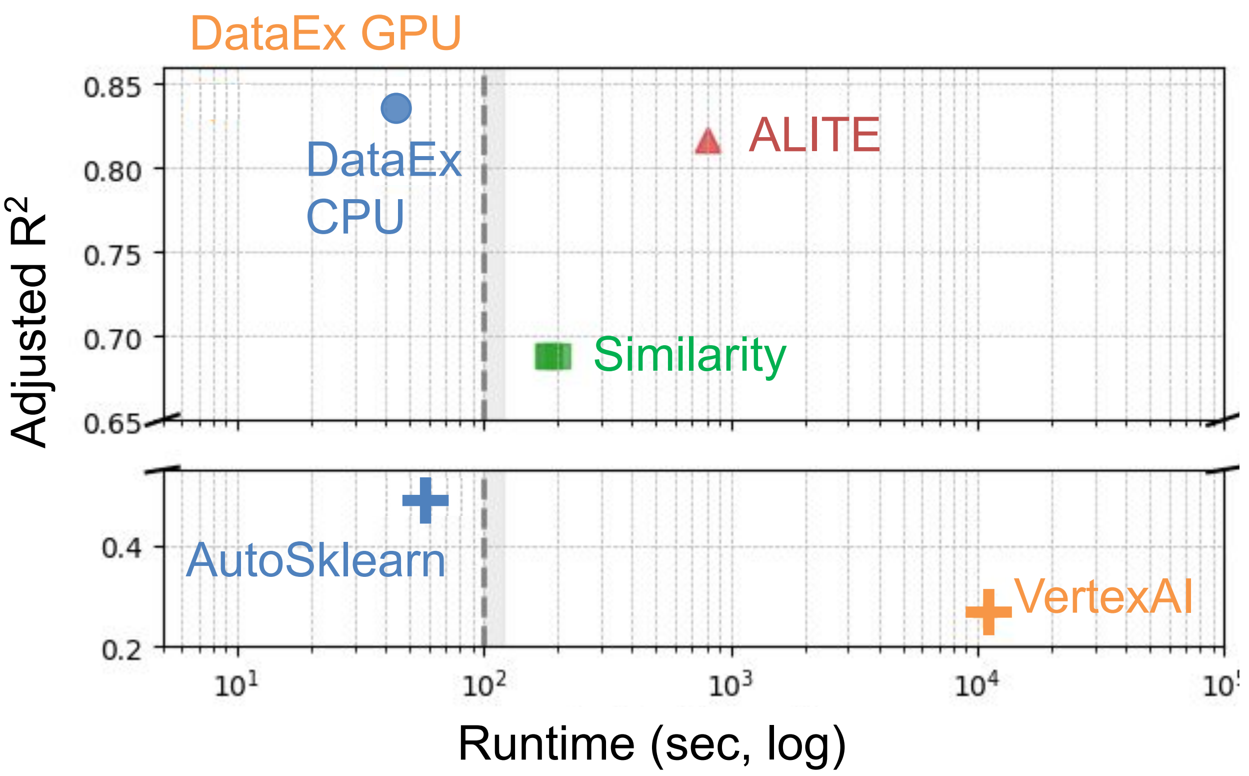
Evaluation on 8376 Kaggle Tables



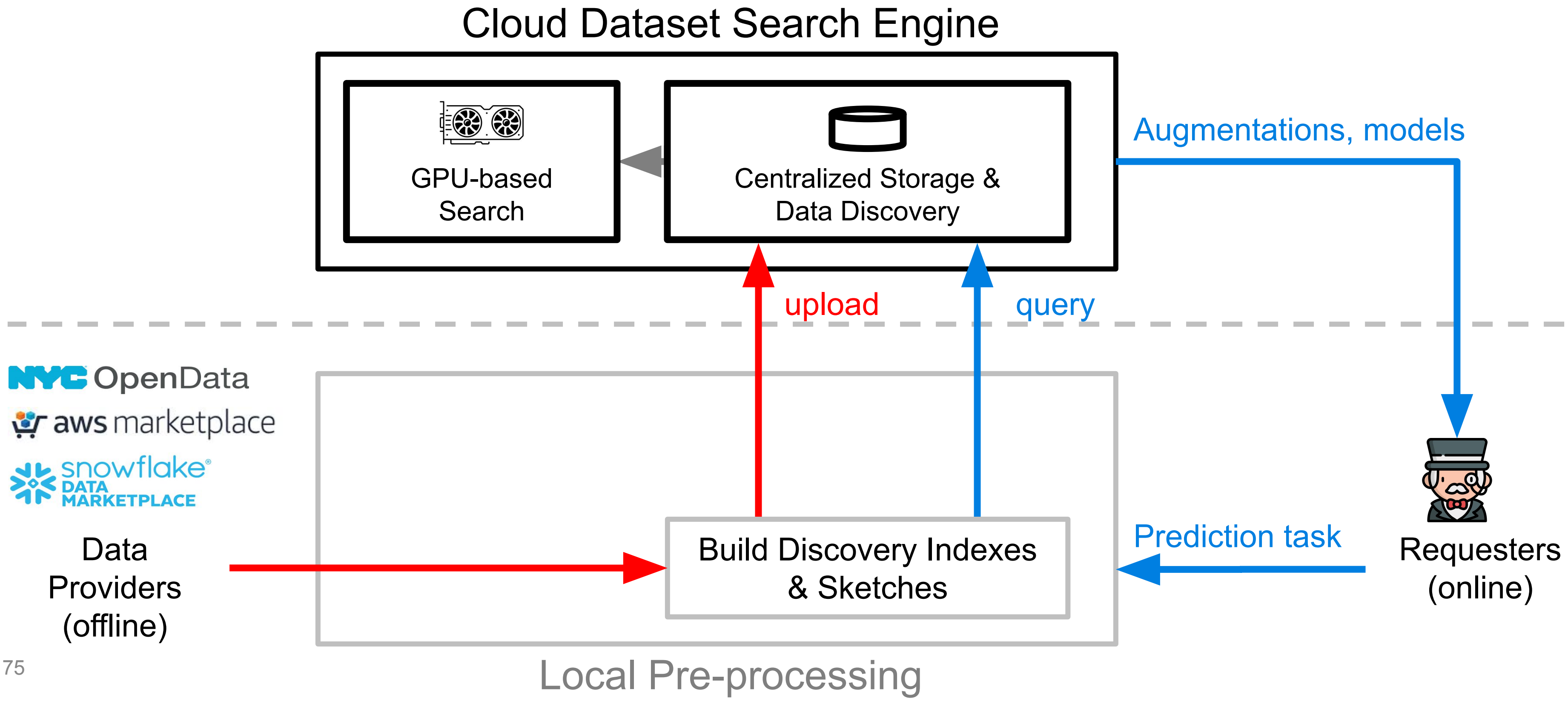
Evaluation on 8376 Kaggle Tables



Evaluation on 8376 Kaggle Tables

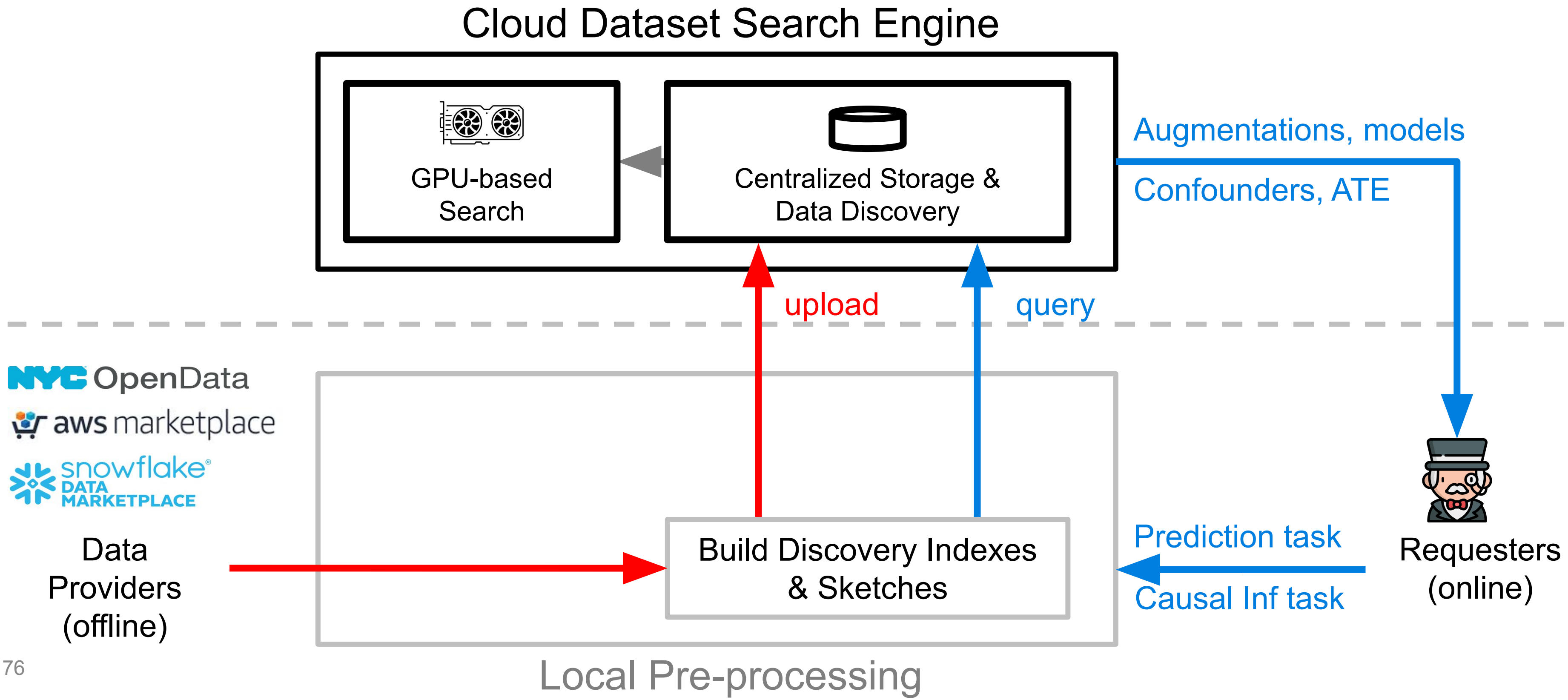


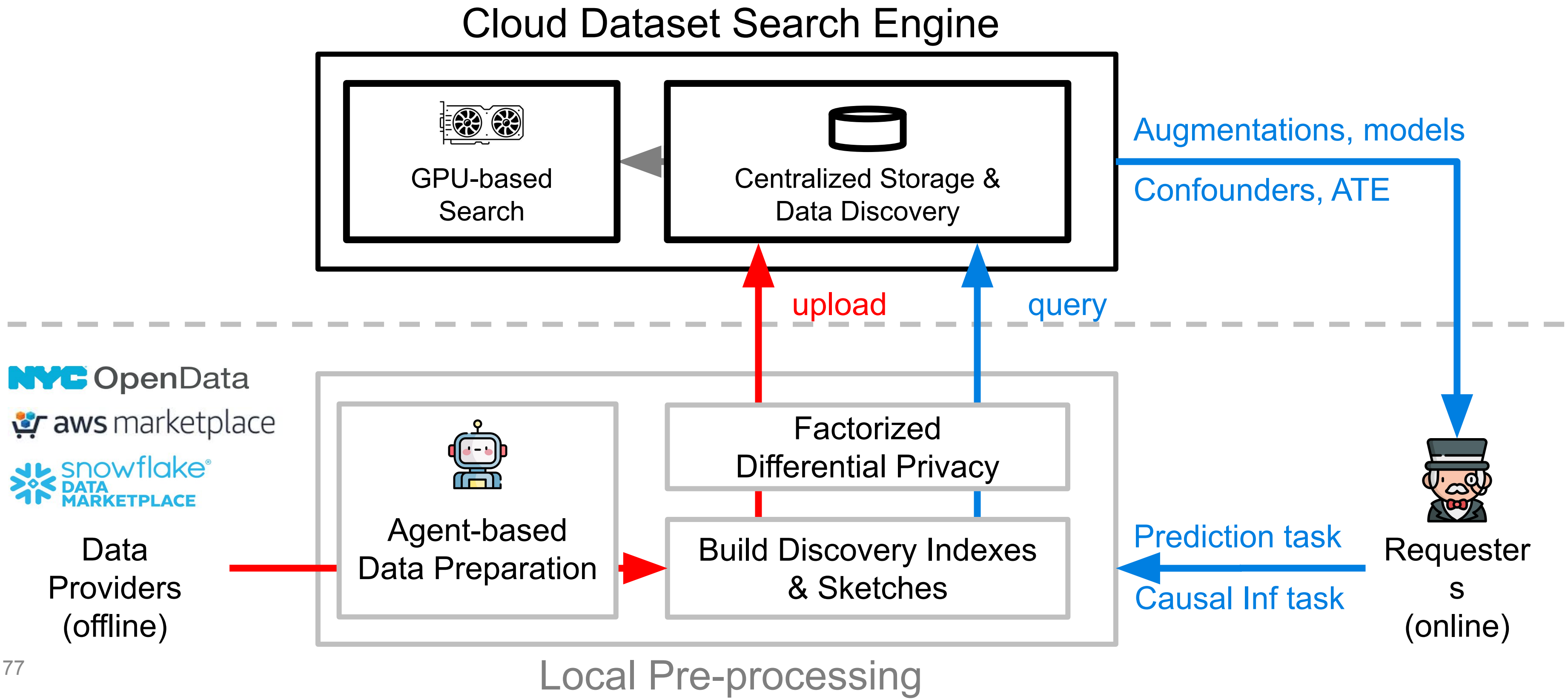
DataEx



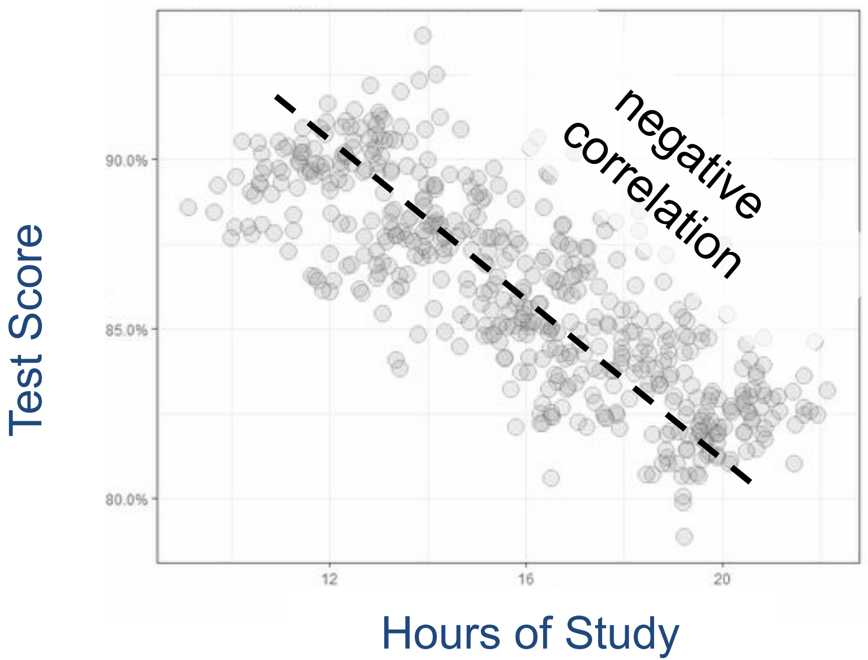
Cloud Dataset Search Engine
GPU-based Search
Centralized Storage & Data Discovery
Augmentations, models
upload
query
OpenData
aws marketplace
snowflake DATA MARKETPLACE
Data Providers (offline)
Build Discovery Indexes & Sketches
Prediction task
Requesters (online)
Local Pre-processing

DataEx

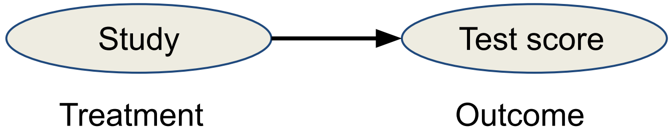

Cloud Dataset Search Engine
GPU-based Search
Centralized Storage & Data Discovery
Augmentations, models
Confounders, ATE
upload
query
OpenData
aws marketplace
snowflake
DATA MARKETPLACE
Data Providers (offline)
Build Discovery Indexes & Sketches
Prediction task
Causal Inf task
Requesters (online)
Local Pre-processing


Cloud Dataset Search Engine
GPU-based Search
Centralized Storage & Data Discovery
Augmentations, models
Confounders, ATE
upload
query
NYC OpenData
aws marketplace
snowflake DATA MARKETPLACE
Data Providers (offline)
Agent-based Data Preparation
Factorized Differential Privacy
Build Discovery Indexes & Sketches
Prediction task
Causal Inf task
Requesters (online)
Local Pre-processing

Confounders in Causal Analysis



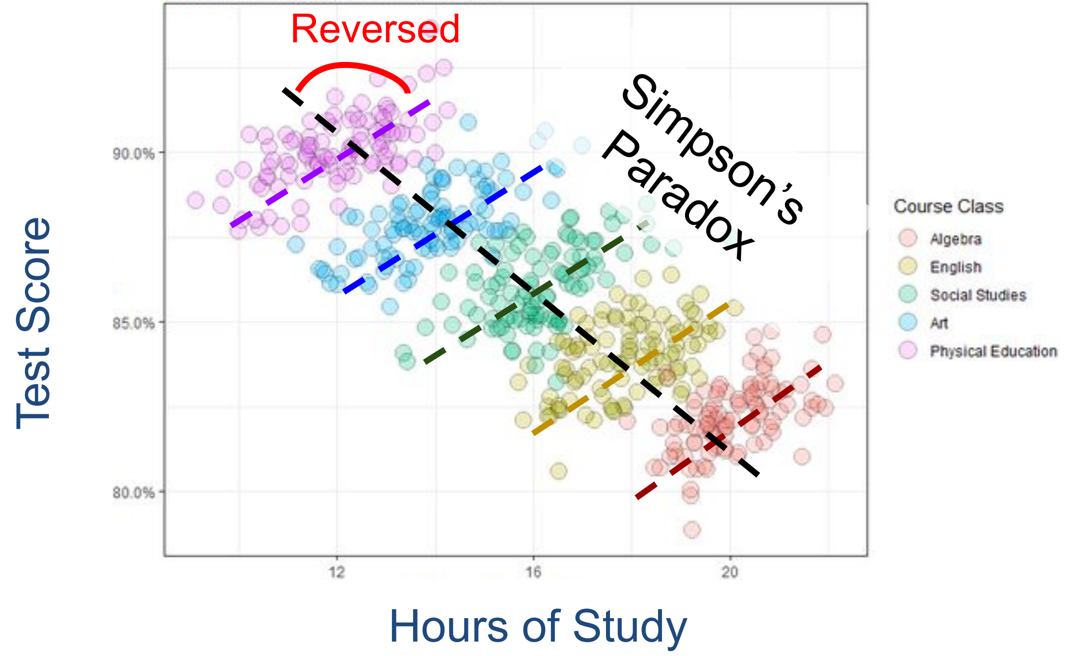
Causal Diagram



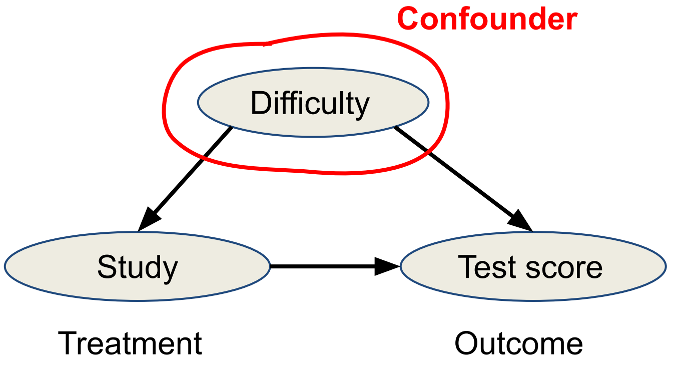
Studying causes poor grades?
Study → TestScores

https://math.stackexchange.com/questions/83756/more-examples-of-simpsons-paradox-barring-the-ones-on-wikipedia-titanic-and

Confounders in Causal Analysis



Causal Diagram



Studying causes poor grades?

Why does the trend reverse?

Study → TestScores

https://math.stackexchange.com/questions/83756/more-examples-of-simpsons-paradox-barring-the-ones-on-wikipedia-titanic-and

Confounders in Causal Analysis

User Query

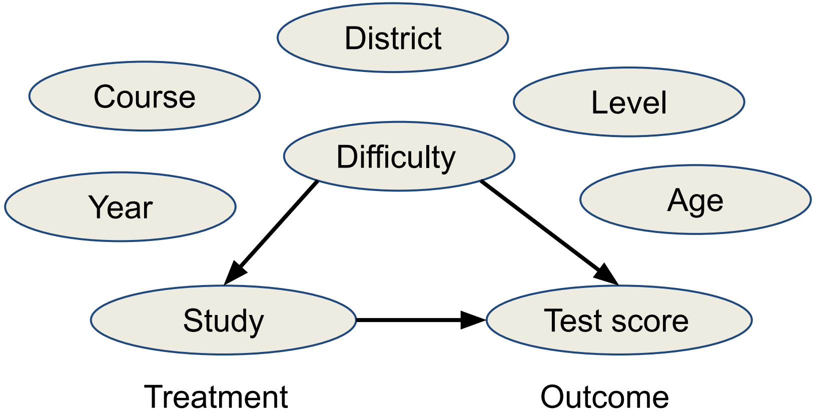
	Treatment	Outcome
ID	Study	Score
1	15 hr	75
2	10 hr	90
3	20 hr	85

Data Repository

ID	Course	Difficulty
1	CS101	3
2	CS102	1
3	CS103	4

ID	District
1	1
2	2
3	3

...



Confounders in Causal Analysis

User Query

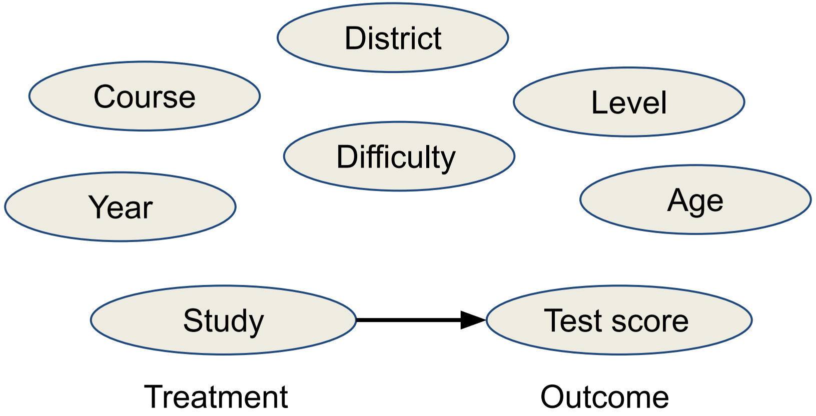
	Treatment	Outcome
ID	Study	Score
1	15 hr	75
2	10 hr	90
3	20 hr	85

Data Repository

ID	Course	Difficulty
1	CS101	3
2	CS102	1
3	CS103	4

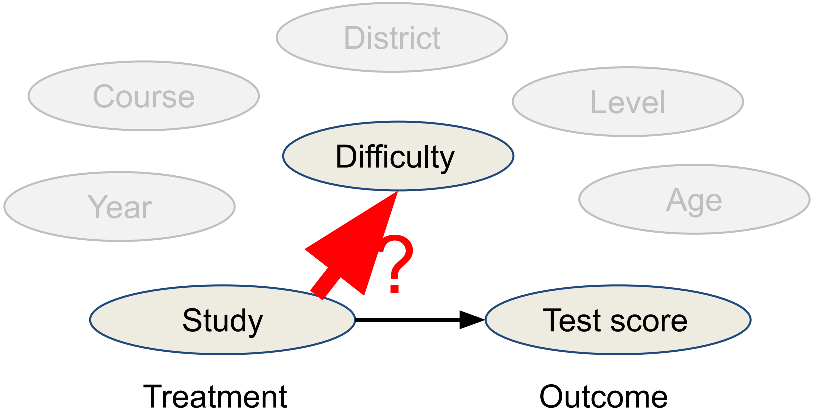
ID	District
1	1
2	2
3	3

...



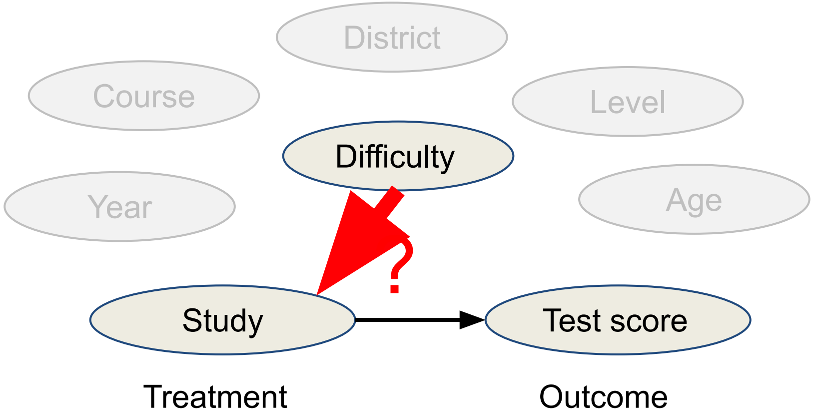
Confounders in Causal Analysis

Background: Bivariate Causal Discovery (BCD) estimates → or ← edges from data



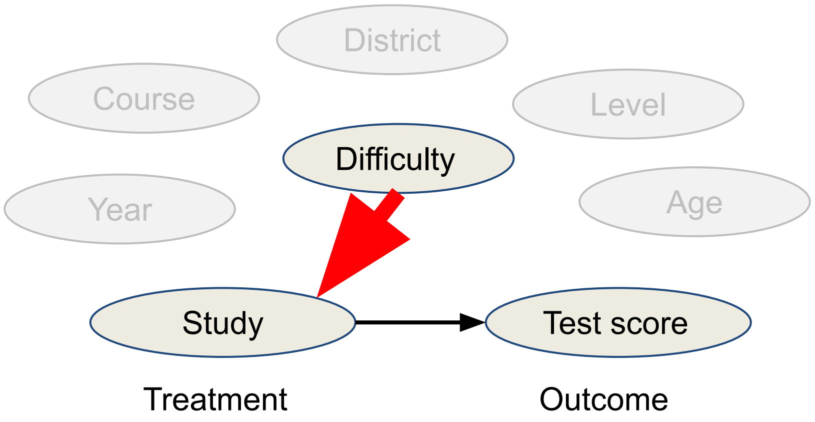
Confounders in Causal Analysis

Background: Bivariate Causal Discovery (BCD) estimates → or ← edges from data



Confounders in Causal Analysis

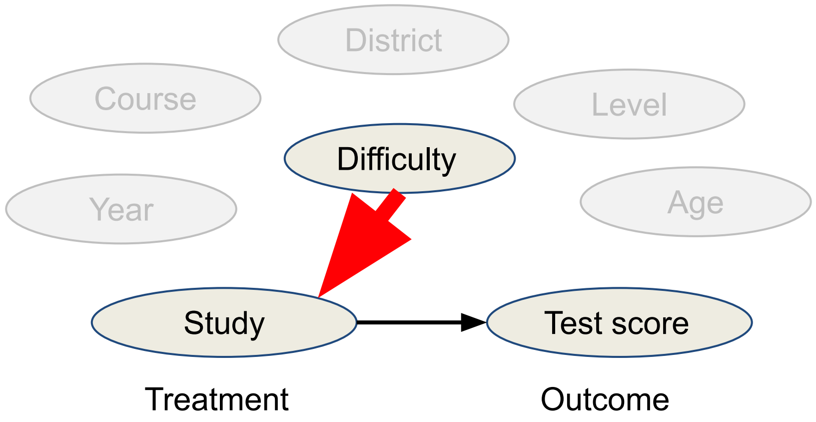
Background: Bivariate Causal Discovery (BCD) estimates → or ← edges from data



Confounders in Causal Analysis

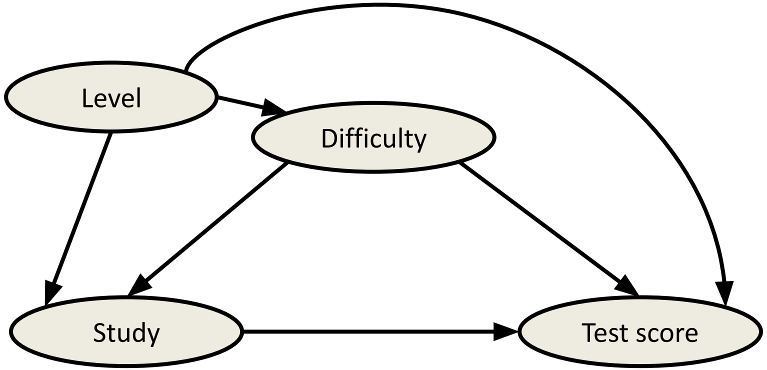
Background: Bivariate Causal Discovery (BCD) estimates → or ← edges from data

Proof: existence of confounder reduces to BCD estimating "*Ancestors* ↝ *Treatment*"



Discovering Confounders with BCD

Building adjustment set for Study → TestScore

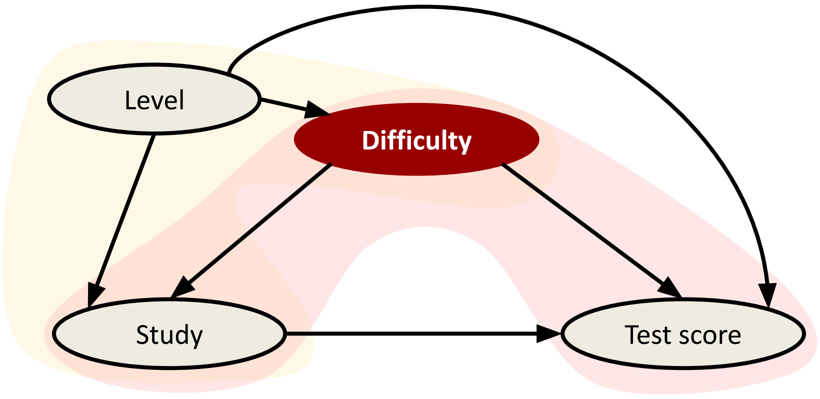


Key Observation:

treatment and outcome confounded: BCD will flag confounder → treatment.

Discovering Confounders with BCD

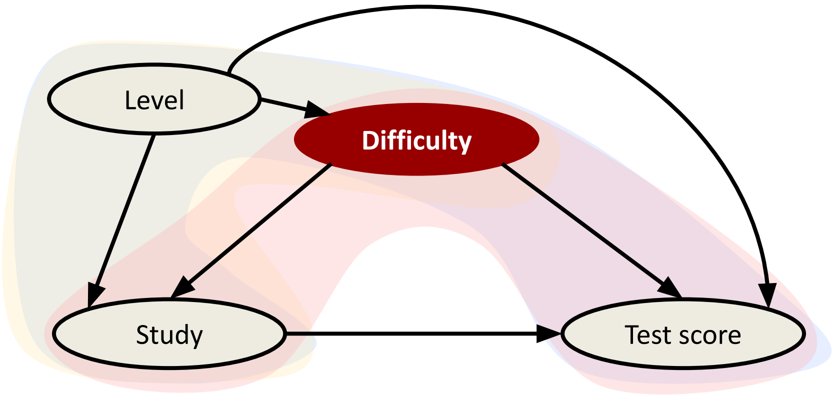
Building adjustment set for Study $\longrightarrow$ TestScore



Difficulty is a confounder, not flagged by BCD because confounded by **Level**

Discovering Confounders with BCD

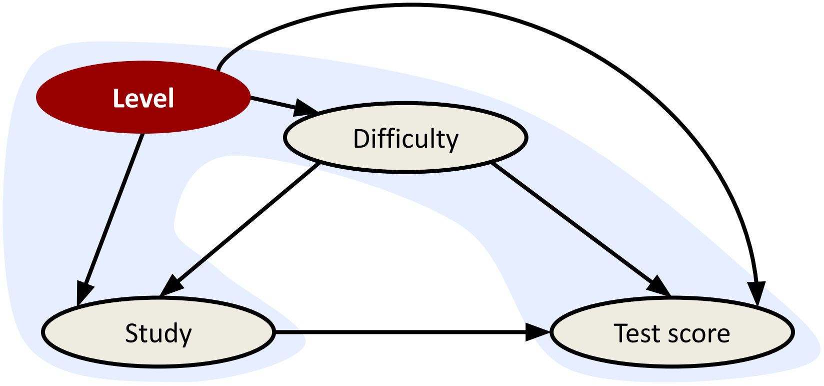
Building adjustment set for Study ⟶ TestScore



Difficulty is a confounder, not flagged by BCD because confounded by **Level**

Discovering Confounders with BCD

Building adjustment set for Study ⟶ TestScore



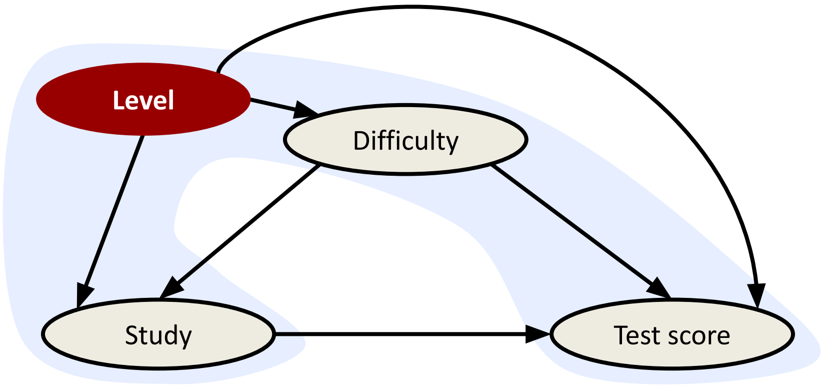
Level is also a confounder, flagged by BCD

Key Insight:

Level < Difficulty topologically – we prove a confounder always flagged by BCD

Discovering Confounders with BCD

Building adjustment set for Study ⟶ TestScore



Theorem 1: *If ∃ confounder between treatment and outcome, ∃ attribute **A** s.t*

- ***A*** → treatment and ∄ *confounder between **A** and treatment.* **Flagged by BCD**
- ***A*** *is a confounder between treatment and outcome.* **Selected heuristically**

Confounders in Causal Analysis

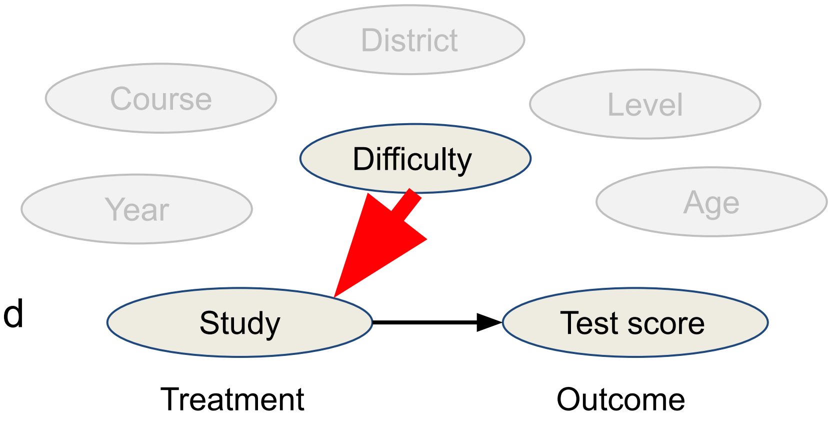
Background: Bivariate Causal Discovery (BCD) estimates → or ← edges from data

Proof: existence of confounder reduces to BCD estimating "*Ancestors* ↝ *Treatment*"

Algorithm: Use BCD to find superset of *Ancestors* and iteratively reduce until it is an admissible set.

System: develop novel sketches to accelerate BCD evaluation, scale using GPUs

- Level= β_1·Study + ϵ_1
- Estimate: MI(Study, Level-β_1·Study)
- Push mutual information through joins

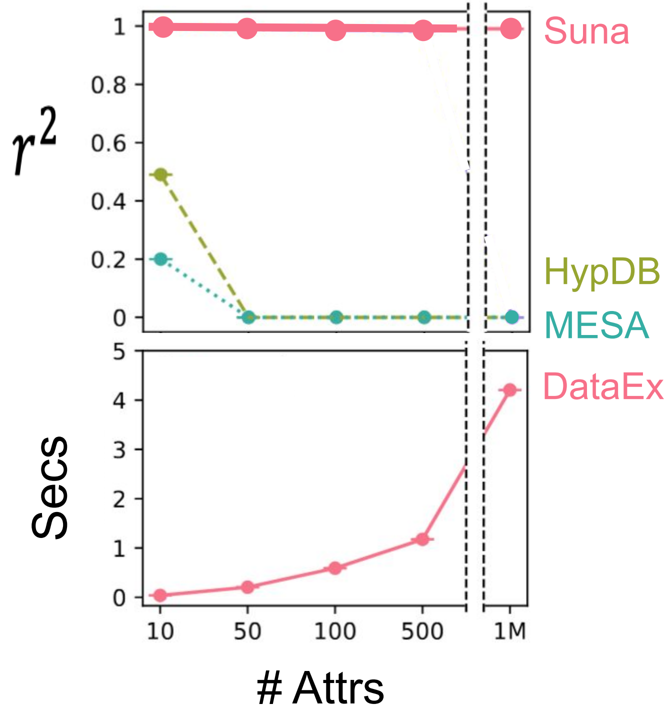


Experimental Results

Real Data: Reproduces Known Confounders

Dataset	Query	Suna
SO	What is the effect of education level on salary?	Cost of Living & Rent Index
ELA	What is the effect of each school's extra credit performance score on students' ELA score?	Enrollment % Poverty
Ratio	What is the effect of each school's pupil-to-teacher ratio on student's ELA score?	Level 4: % % Students with Disabilities Minimum Class Size
SAT	What is the effect of test takers numbers on SAT score?	# Safety Incidents Enrollment Total Regents #

Synthetic Data: Accurate & Fast



Summary of Task-based Search

Task-evaluation is bottleneck

- Identify hardware and parallelization-friendly sketches to accelerate task evaluation

Need algorithms to avoid combinatorial search

Arbitrary tasks can be supported, but are very difficult…

Metam: Task-agnostic search [Galhotra23]

Problem Setup

GIVEN:

An initial dataset D,

a collection of attributes Γ,

a task implementation t

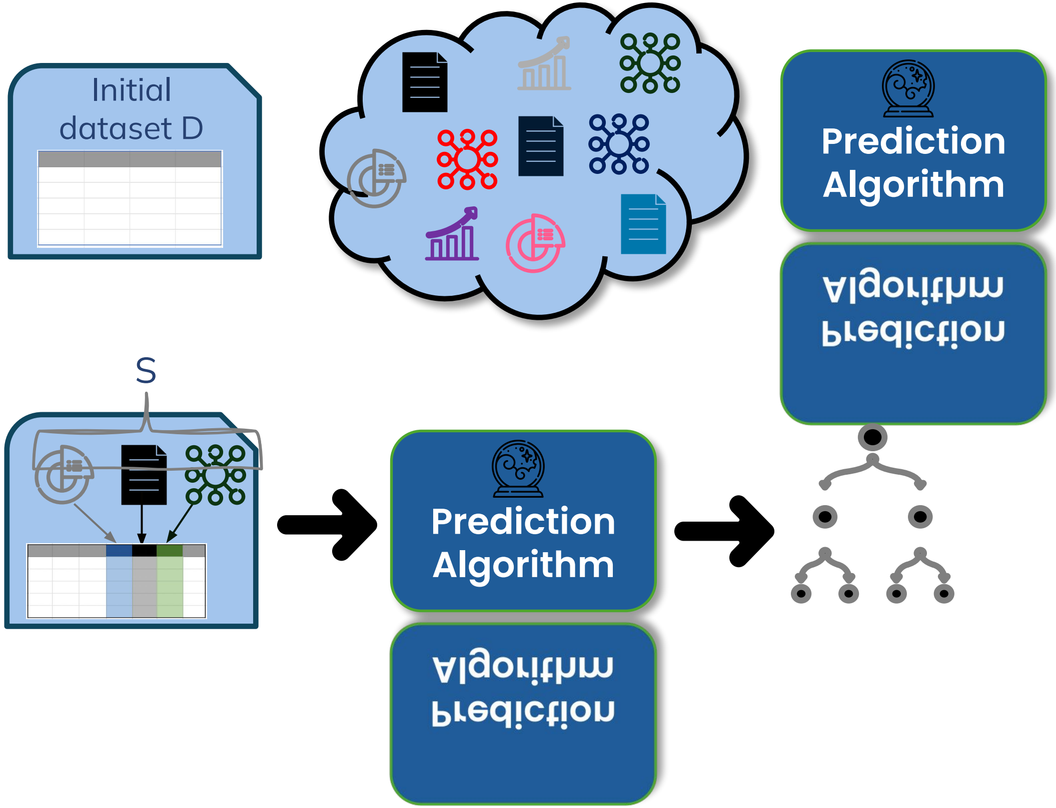
OBJECTIVE: max $\text{utility}_t\,(D \bowtie S)$

CONSTRAINT:

$S \subseteq \Gamma$

$|S| \leq k$ (a constant)

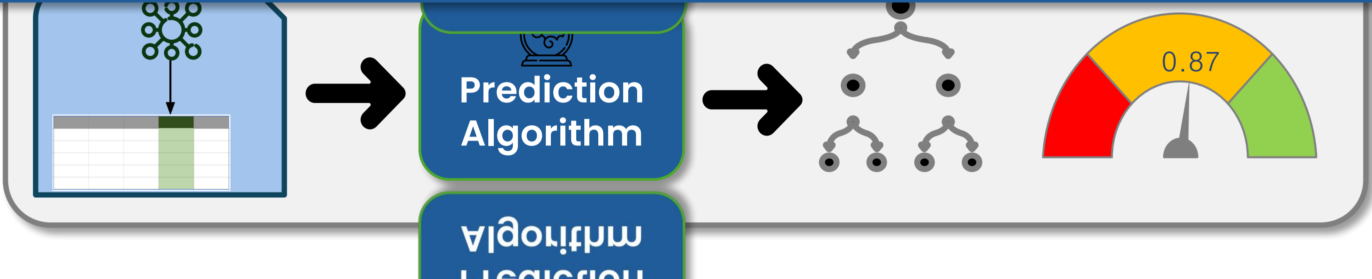
S is minimal



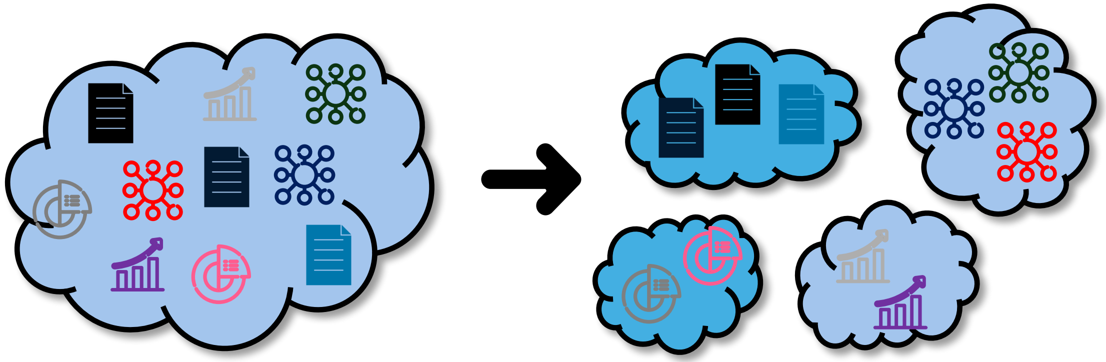
How to solve the problem?

⚠️ **Requires n^k queries! Infeasible when n is in the order of millions**

Can we prioritize subsets?



Clustering helps to diversify the search process



Similar datasets have similar utility!

Using Data Properties As Features To Cluster

Id	Address	....	Zip Code	Crime
1	153 JFK, NY		12543	Low
2	543 Albert Street, NY		?	?
3	432 MK road		14656	High
4	5432 Dud Dr		54637	Low
5	6732 Psycho Path		?	?
6	23 Main Street		?	?

Properties of the newly added attribute

Fraction of missing values: 0.4

Correlation (Crime, Area): 0.65

Approach 1: Diversify the Search Process

IDEAL SCENARIO: Probability of sampling an informative attribute from the cluster C

EXPLORE-EXPLOIT DILEMMA:

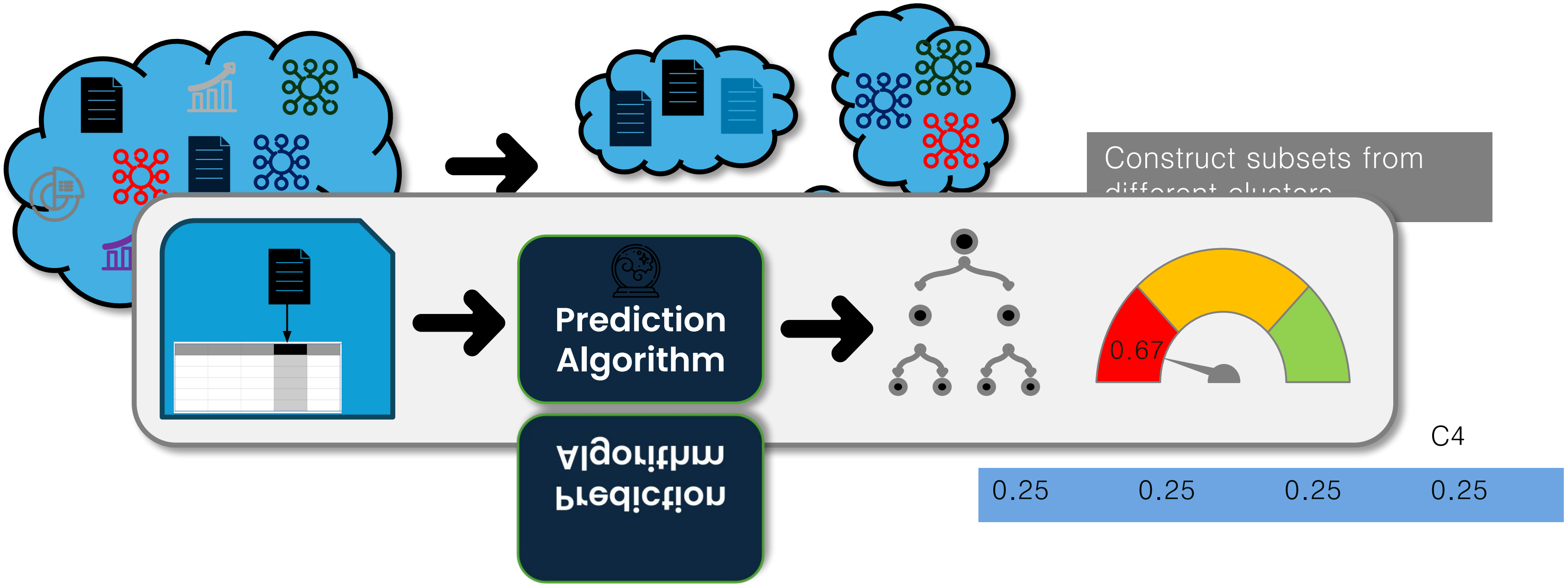
Should I sample more datasets from cluster C_i?

OR

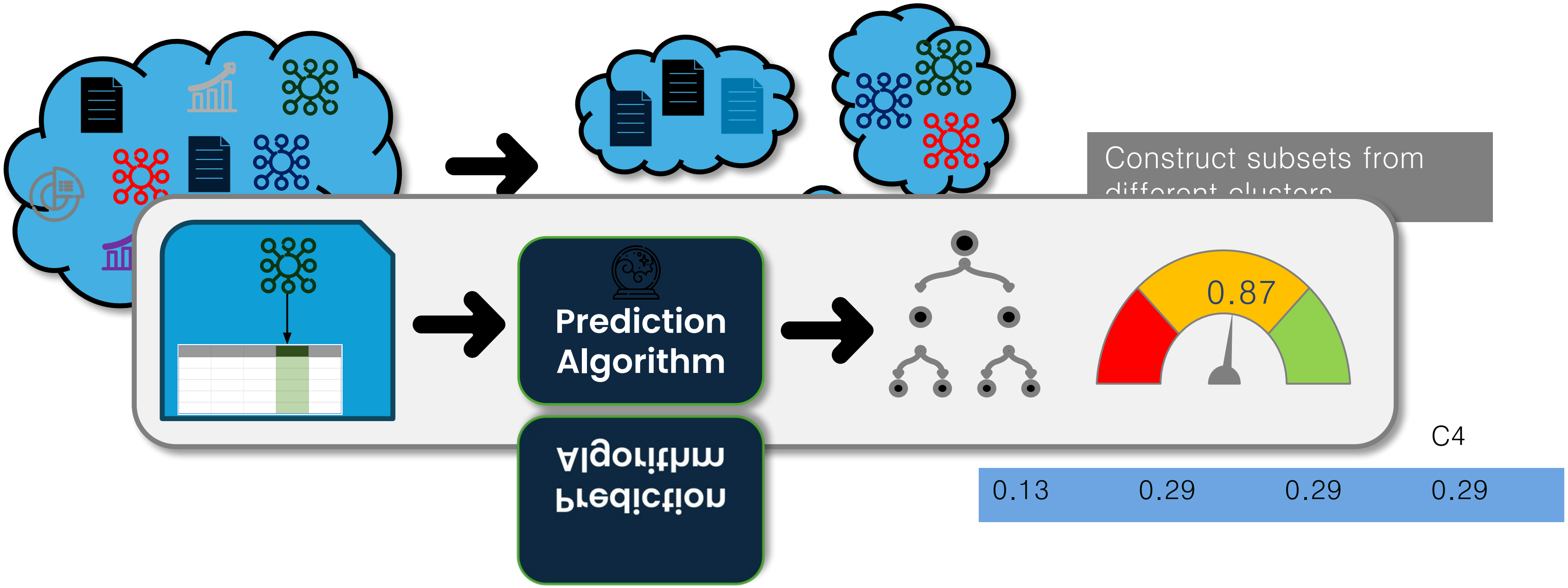
Should I explore different clusters?

SOLUTION: Bandit-based approach

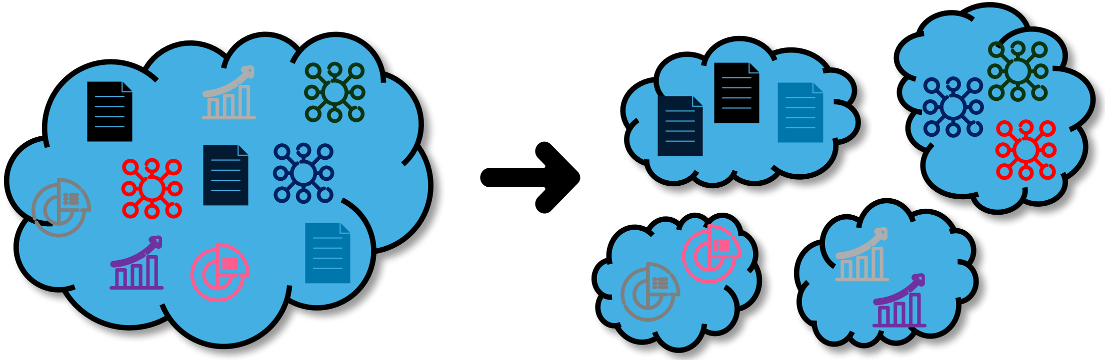
Approach 1: Diversify the Search Process



Approach 1: Diversify the Search Process



Approach 1: Diversify the Search Process



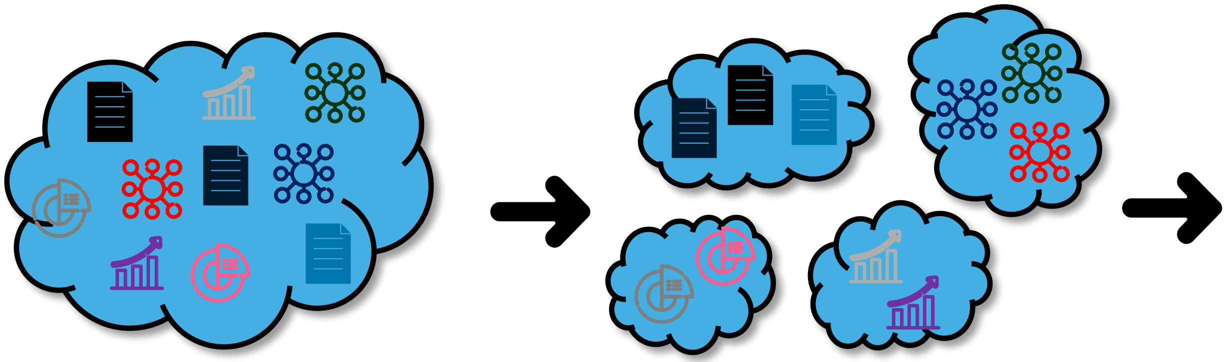
Construct subsets from different clusters

Number of queries: 100

C1	C2	C3	C4
0.02	0.84	0.10	0.04

Converged scores

Approach 1: Diversify the Search Process



Bandit-based Approach

Prediction Algorithm

Leverages data profiles to search efficiently

Complexity $O(|C|^k)$

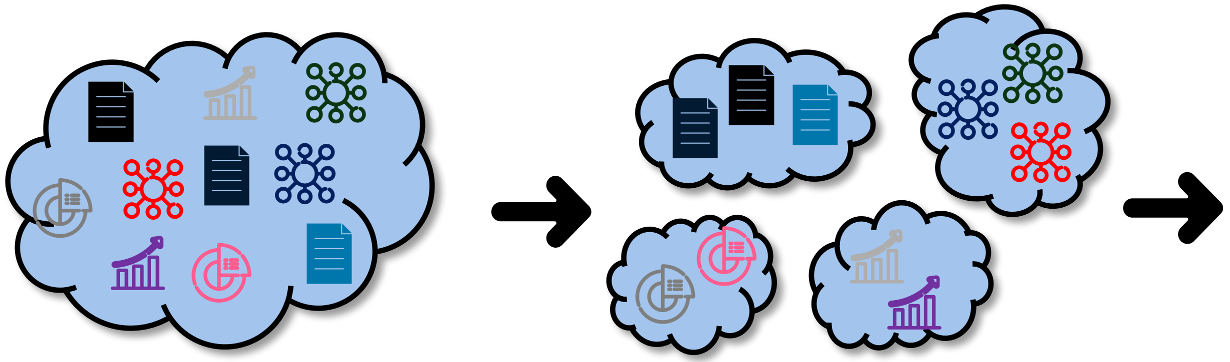
Approach 2: Leverage Monotonicity of Utility Metric

- **Monotonicity**: Easy to guarantee
  $u(D \cup T_2) \geq u(D \cup T_1)\ \forall\, T_1 \subseteq T_2$

- What if the utility is **submodular** too?
  - Diminishing returns property:
    $u(T_1 \cup \{X\}) - u(T_1) \geq u(T_2 \cup \{X\}) - u(T_2)\ \ \forall\, T_1 \subseteq T_2$

Solution: Greedily choose the best augmentation

Approach 2: Leverage Monotonicity of Utility Metric



Choose the best cluster

Choose the best cluster

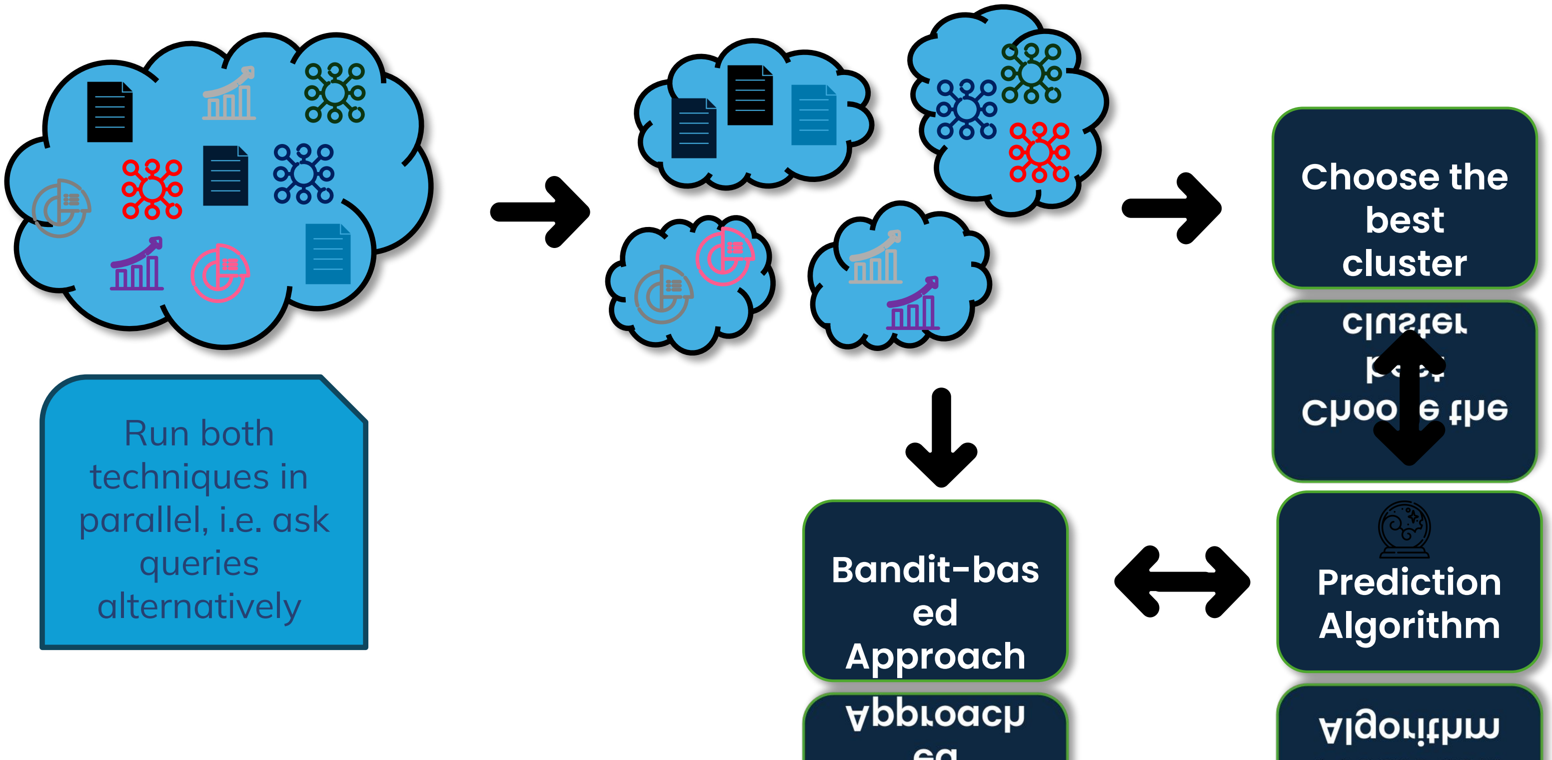
Prediction Algorithm

Algorithm

Highly efficient to identify useful augmentations

Can fail if utility is super modular (unlikely but possible)

Final Approach: Combining All Ideas



Privacy Challenges Are Everywhere!

Query specification
Privacy protection

Data Market/Data Discovery

Query interface
Latency, Scalability
Privacy protection

Stats	Stats	Stats
Data	Data	Data

Data acquisition
Data preparation
Privacy protection

Tutorial Organization

- Part I: Data acquisition and search (Eugene and Sainyam)
- Part II: (Daniel)
- Part III: (Shagufta, Zeyu, and Eugene)
- Part IV: Regulatory Considerations (Daniel)
- Part V: Open Questions (Daniel, Eugene and Sainyam)

BACKUP SLIDES

View Presentation

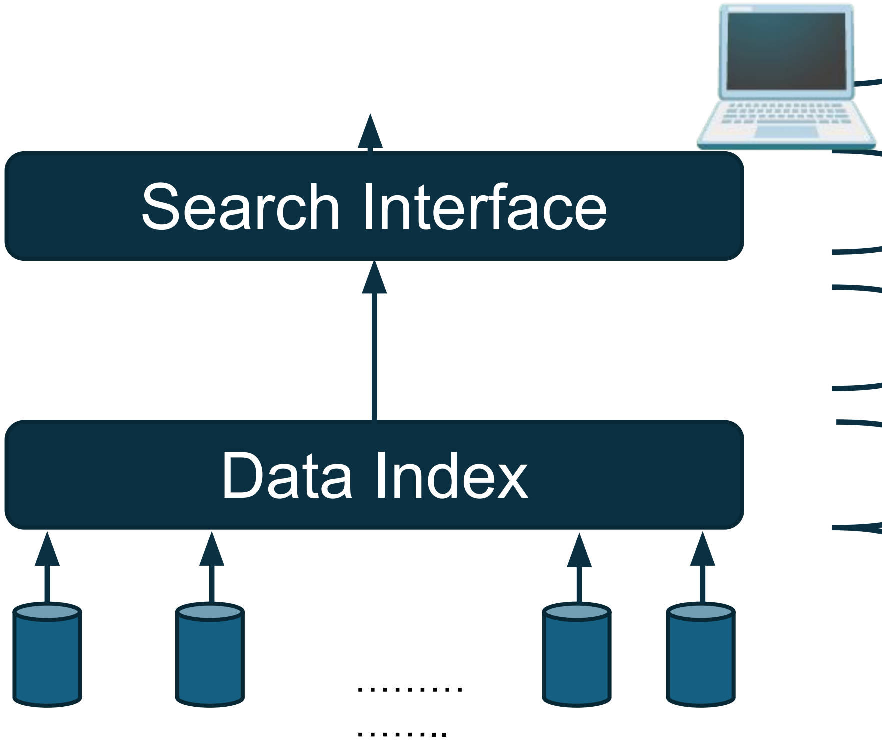
`Do you want to shortlist datasets containing the attribute:  long_name`

- ( ) Yes, my data must contain this attribute
- ( ) No, my data should not contain this attribute
- (•) Does not matter

Submit

Show Shortlisted d…

Data Discovery



How do we show datasets to the users?

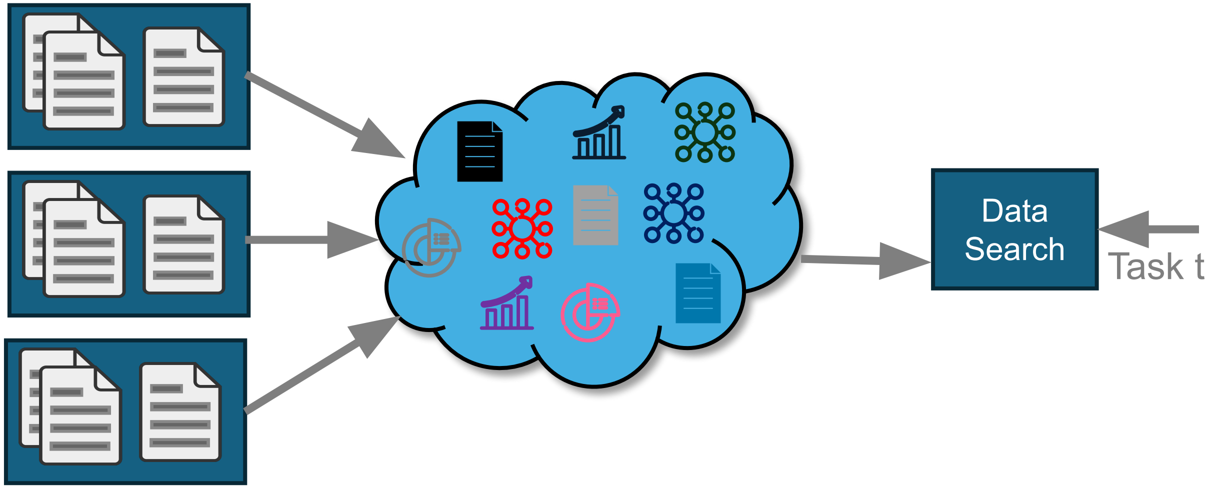
How can users express their queries?

How can we search efficiently?

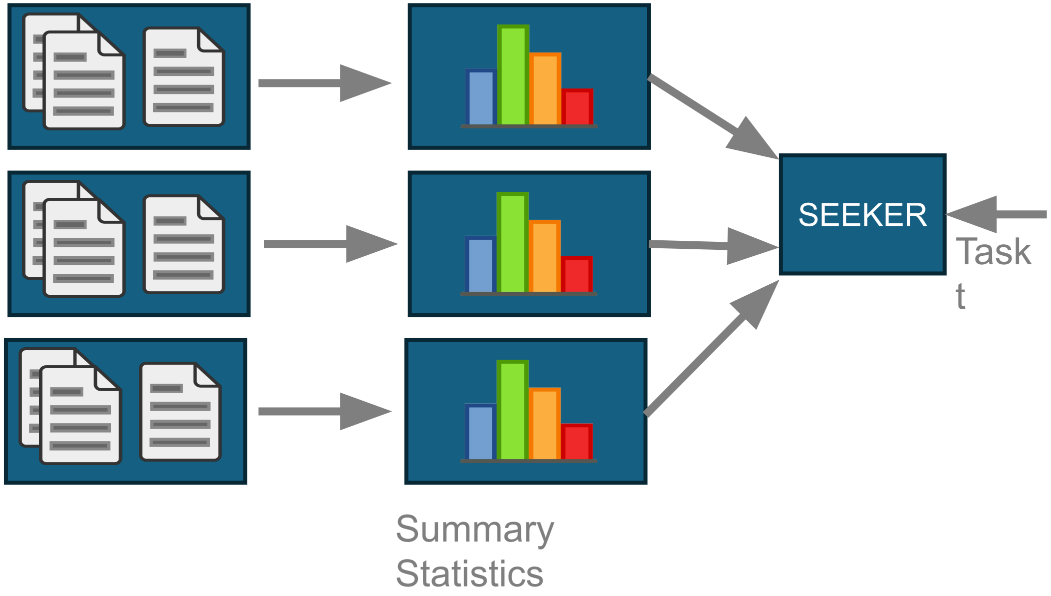
Can we precompute statistics to help with search?

How can we store these datasets to help with search?

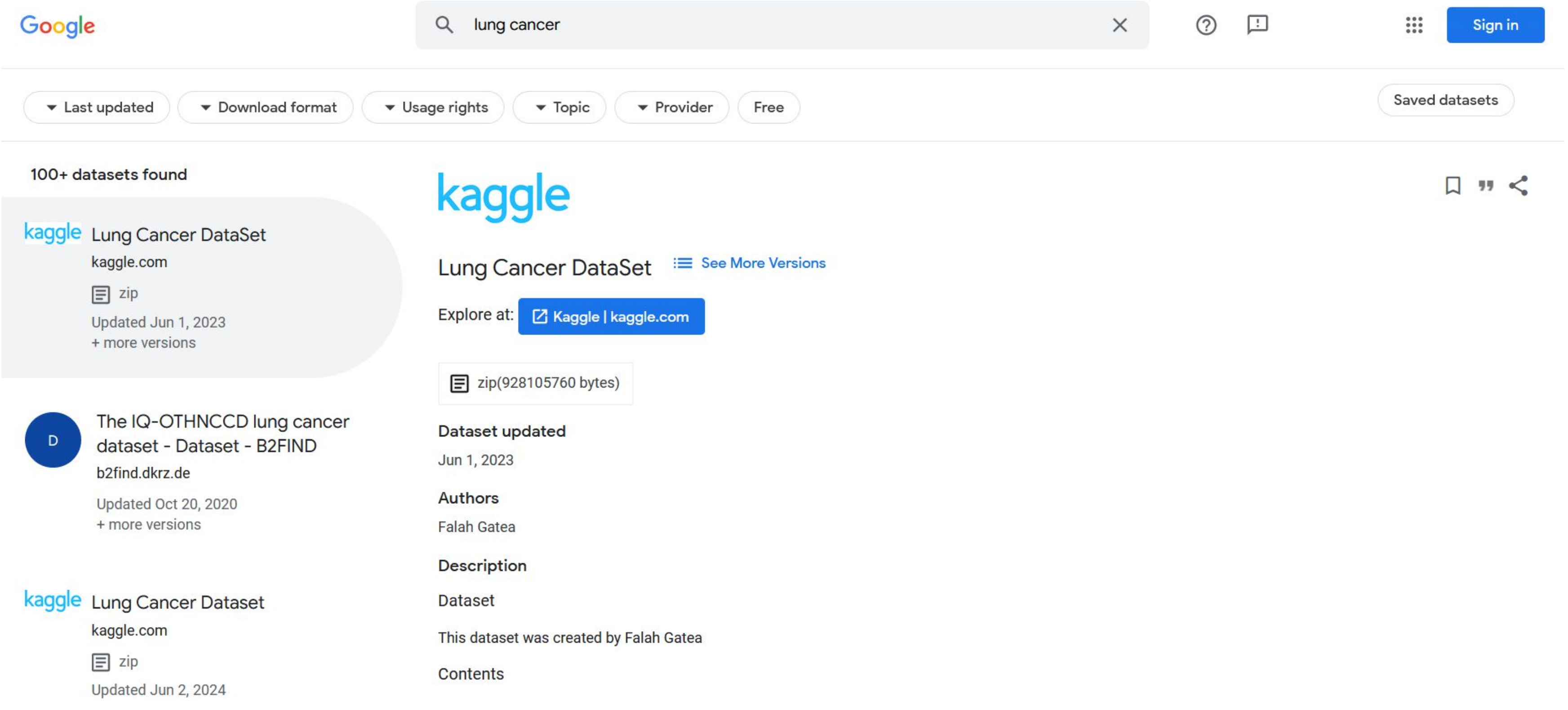
Centralized Data Discovery



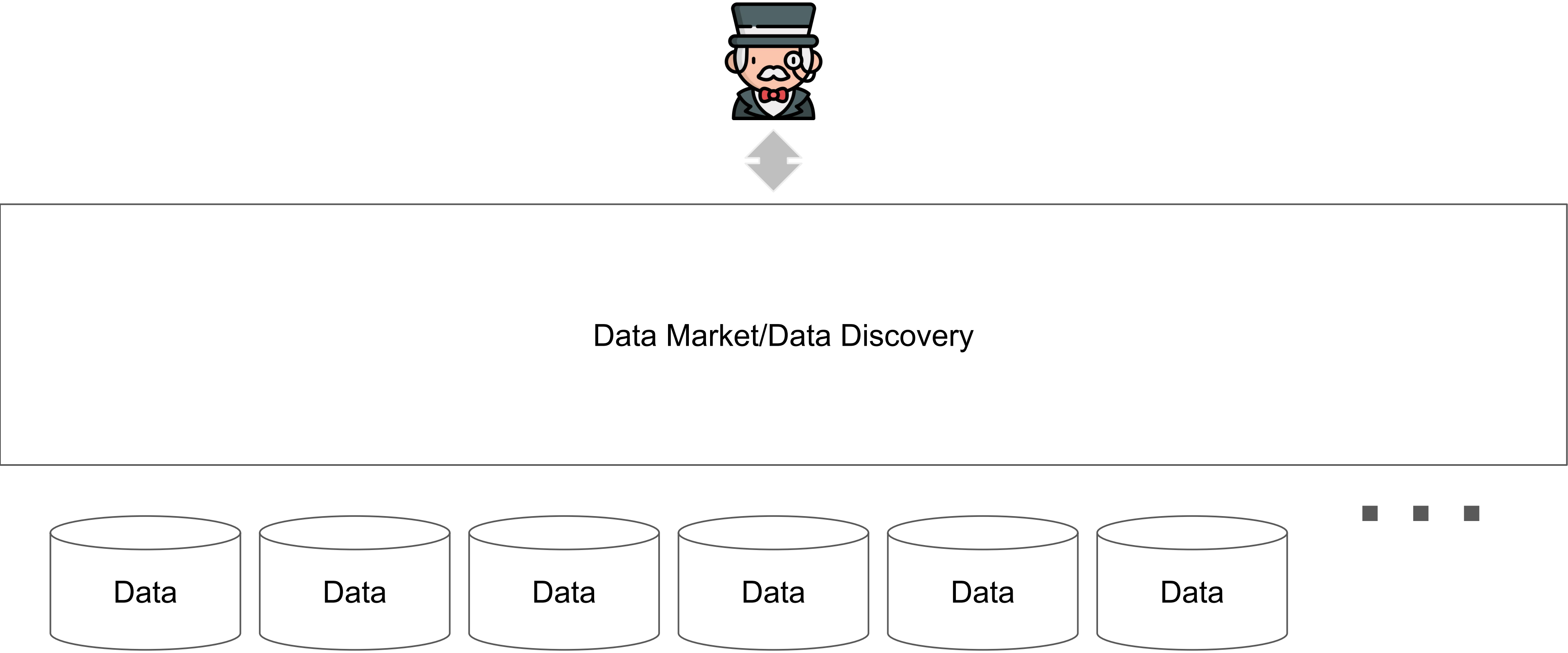
Federated Data Discovery



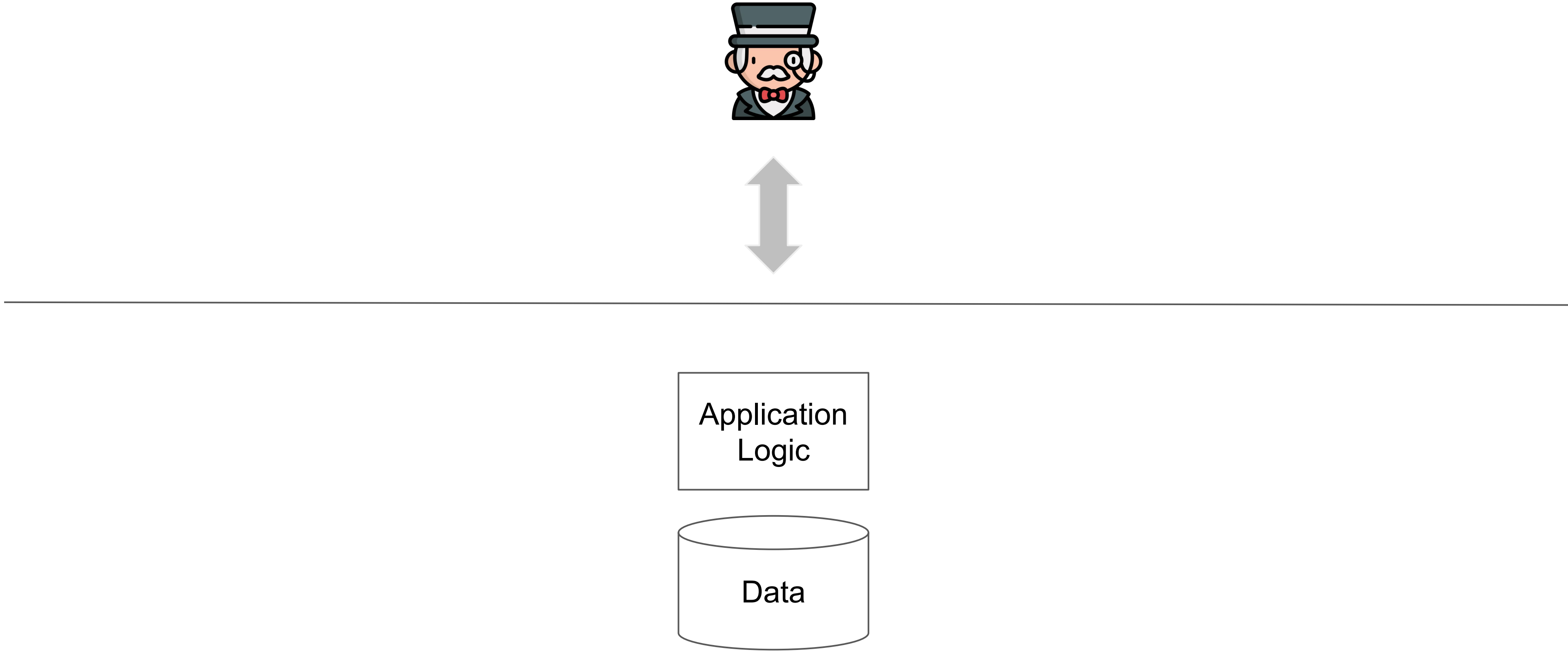
Distribution-Aware Dataset Search

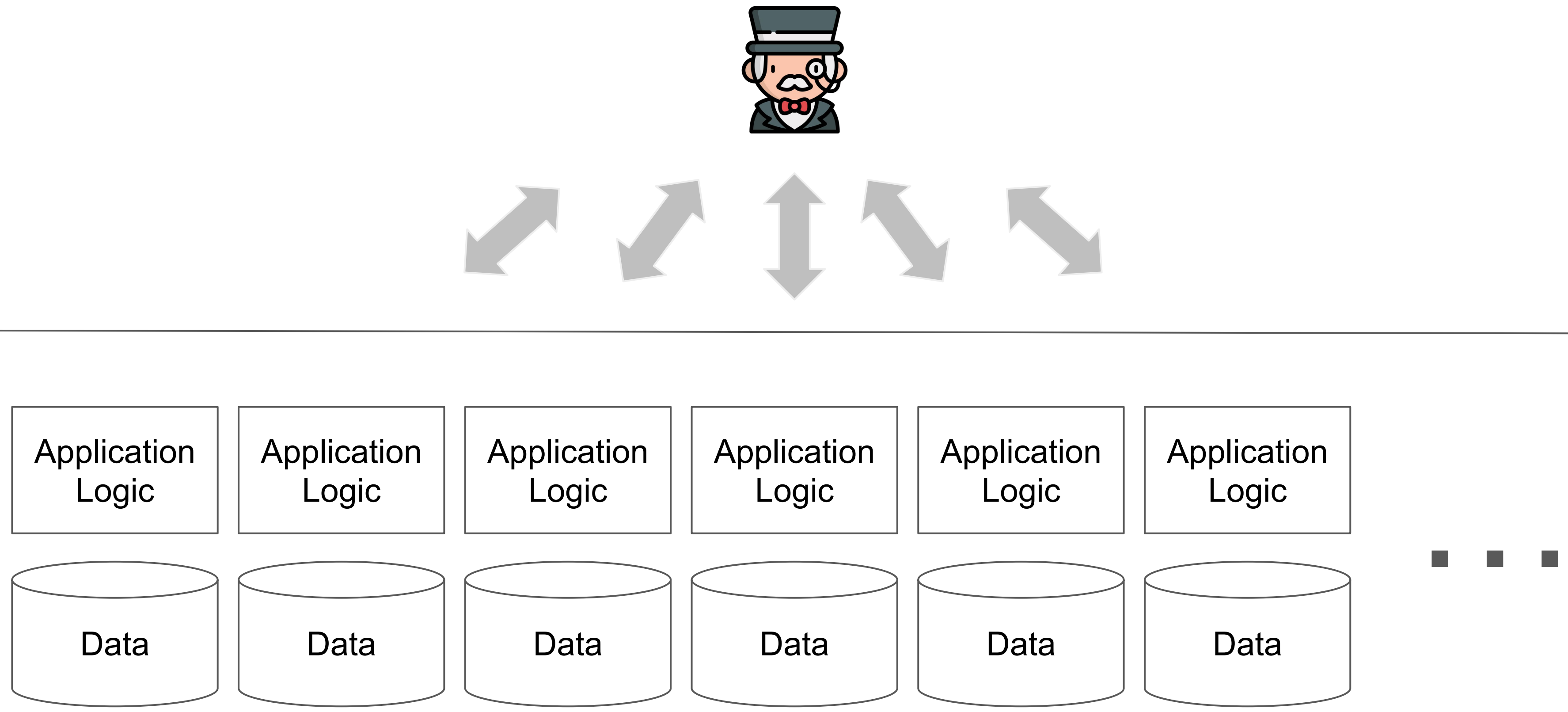


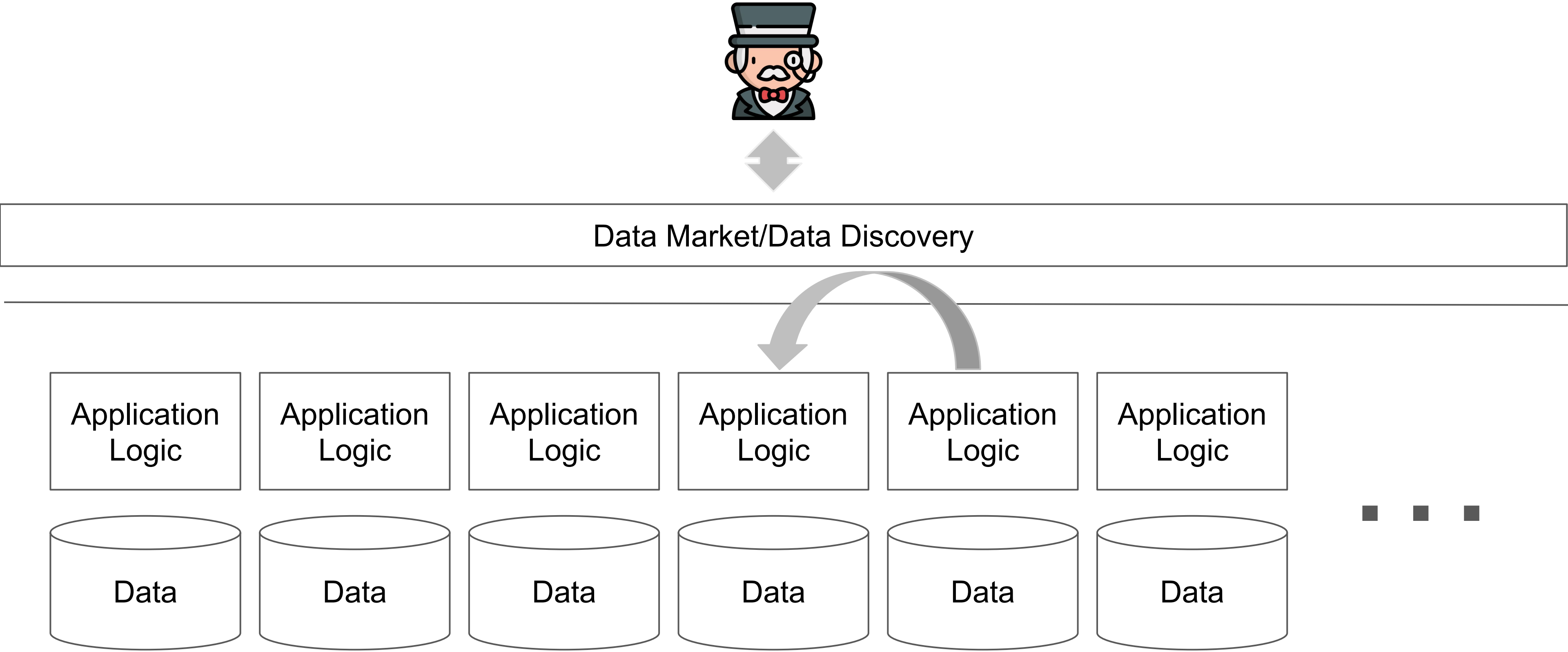
Speculating About the Future



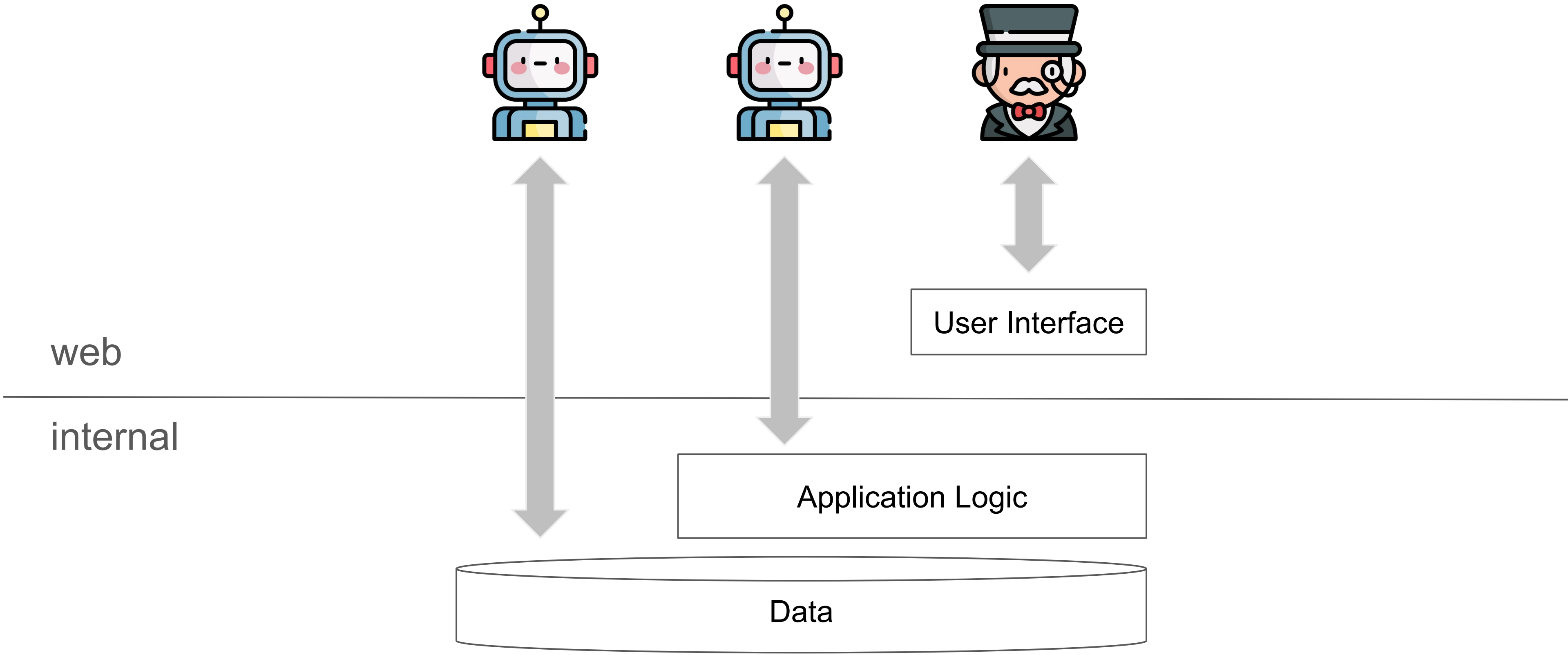
Data Market/Data Discovery
Data
Data
Data
Data
Data
Data


Application Logic
Data


Application Logic
Application Logic
Application Logic
Application Logic
Application Logic
Application Logic
Data
Data
Data
Data
Data
Data


Data Market/Data Discovery
Application Logic
Application Logic
Application Logic
Application Logic
Application Logic
Application Logic
Data
Data
Data
Data
Data
Data
...

Will Data Markets Be More Important In the Future?



Will Data Markets Be More Important In the Future?