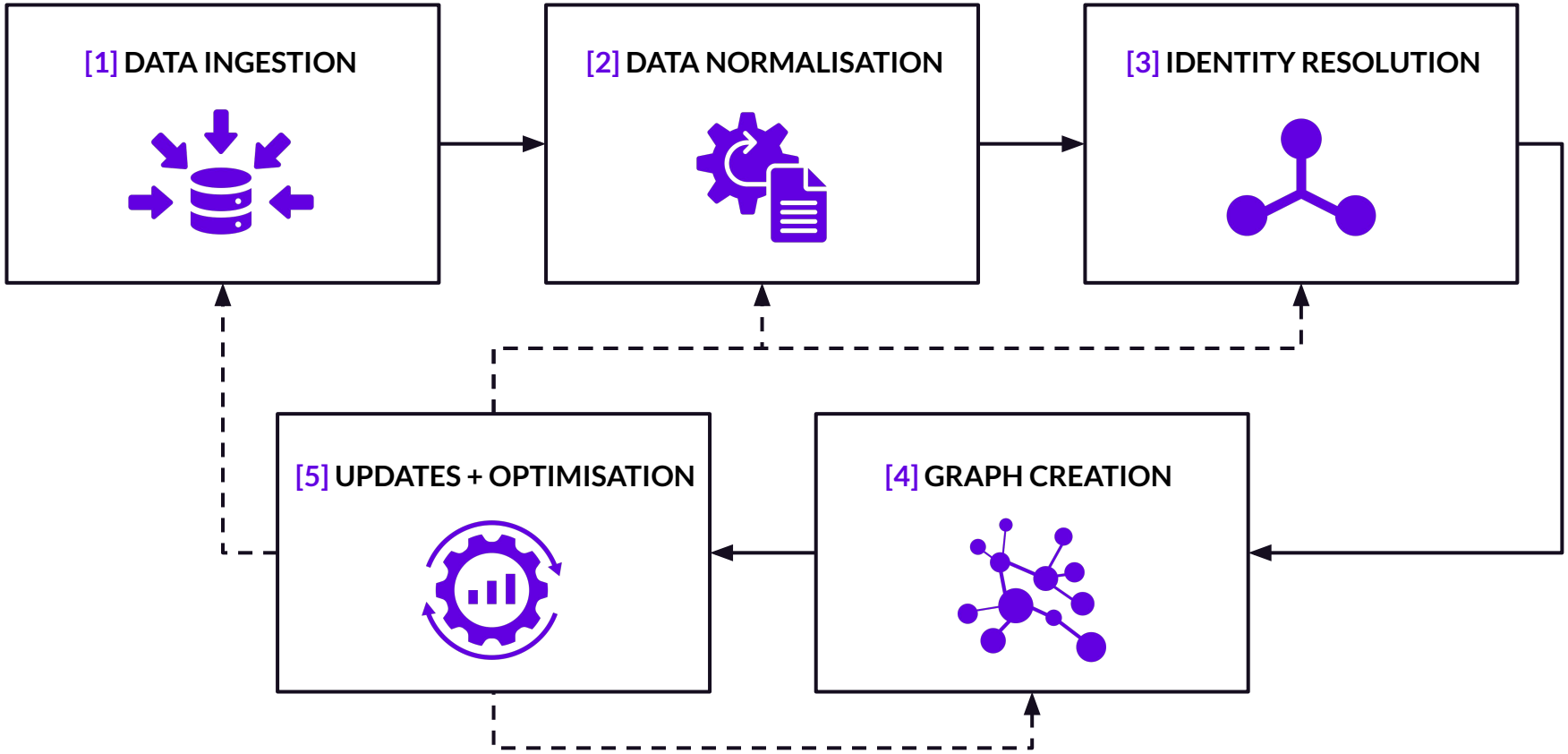
MADTECH MENTAL MODELS

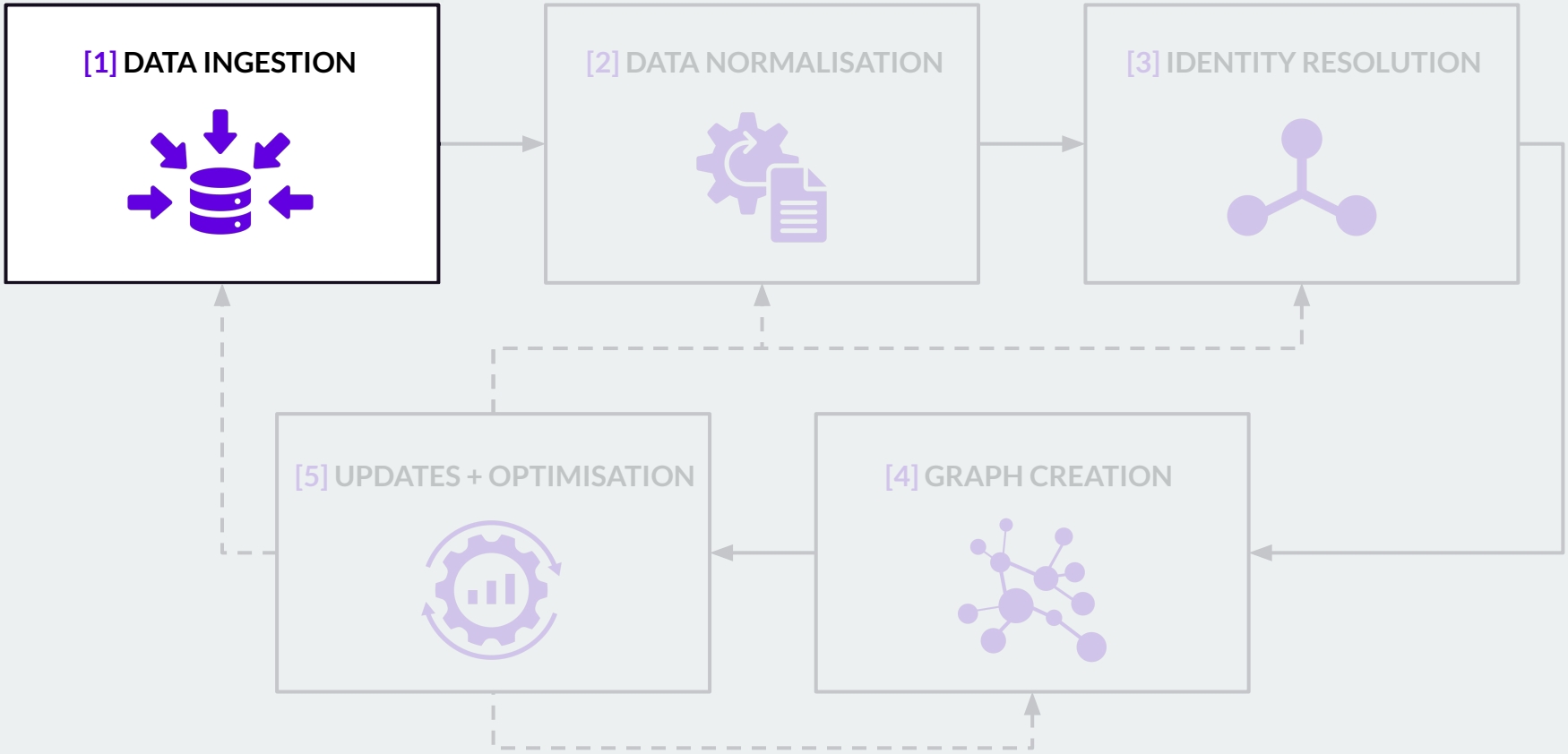
IDENTITY GRAPHS

An Identity Graph (or ID Graph) is a database that collects and links all data points related to a single user.

They’re used to create and store unified customer profiles (or single customer views) to enable and augment marketing use cases.

How Identity Graphs are created (*conceptually*)


[1] DATA INGESTION
[2] DATA NORMALISATION
[3] IDENTITY RESOLUTION
[5] UPDATES + OPTIMISATION
[4] GRAPH CREATION


[1] DATA INGESTION
[2] DATA NORMALISATION
[3] IDENTITY RESOLUTION
[4] GRAPH CREATION
[5] UPDATES + OPTIMISATION

DATA TYPES

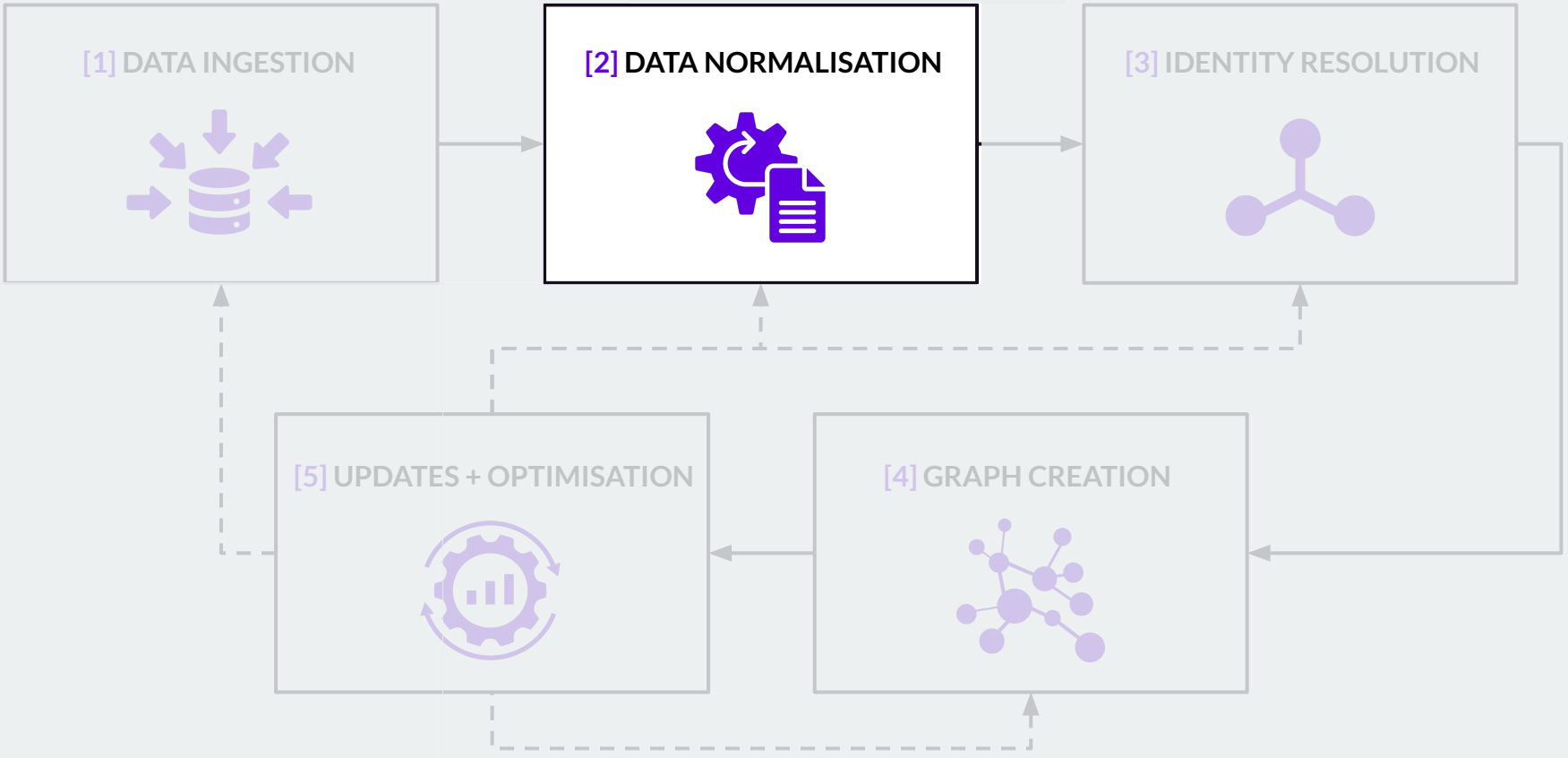
- PII
- Digital IDs
- Behavioral
- Interest
- Demographic

DATA SOURCES

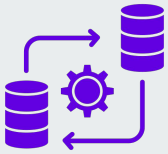
- 1P - Direct user interactions
- 2P - Direct partners
- 3P - Indirect partners

COLLECTION METHODS

- Direct collection
- Tracking/Tagging
- SDKs/APIs
- Data partnerships
- Bidstream


[1] DATA INGESTION
[2] DATA NORMALISATION
[3] IDENTITY RESOLUTION
[4] GRAPH CREATION
[5] UPDATES + OPTIMISATION

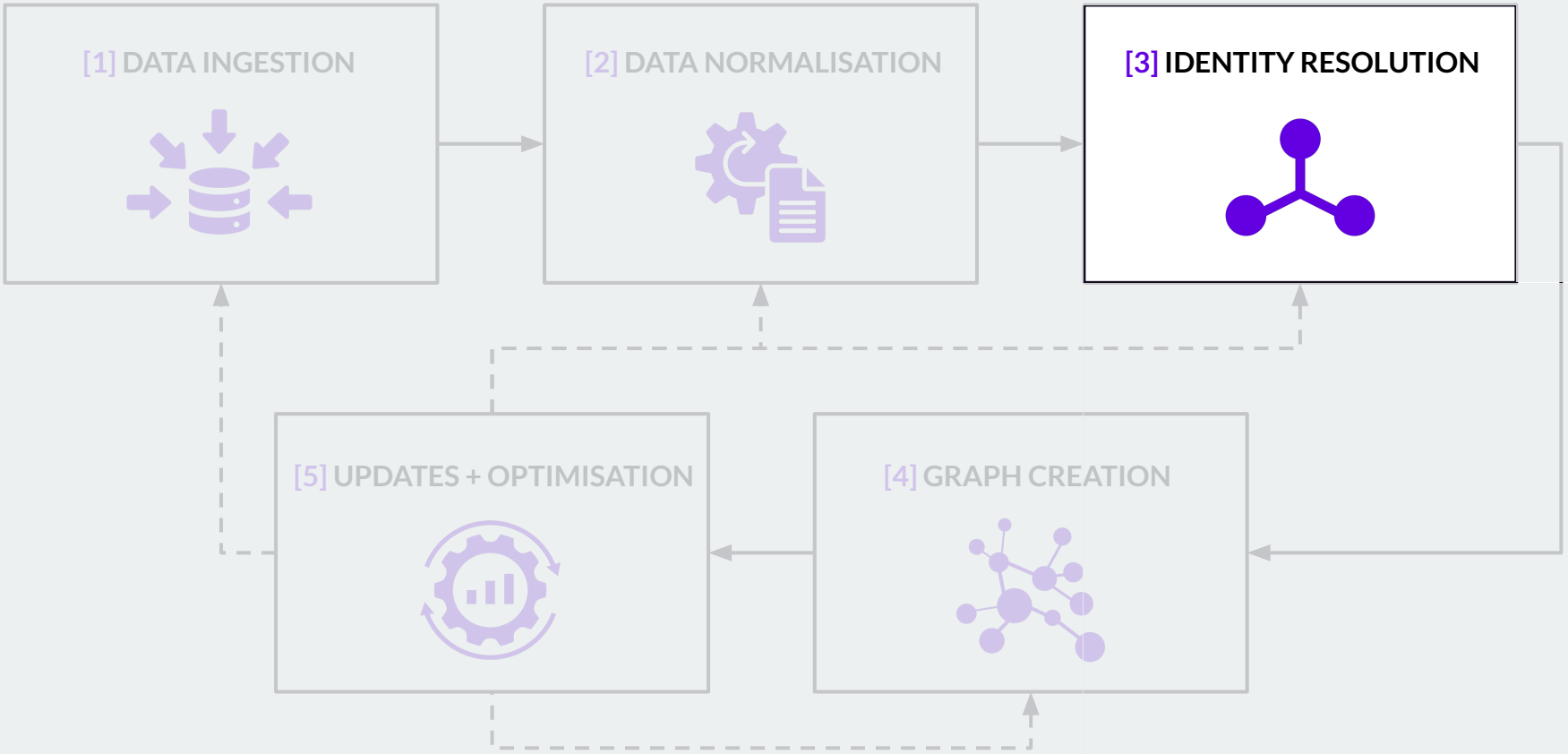
Ensuring integrity and consistency across different data types



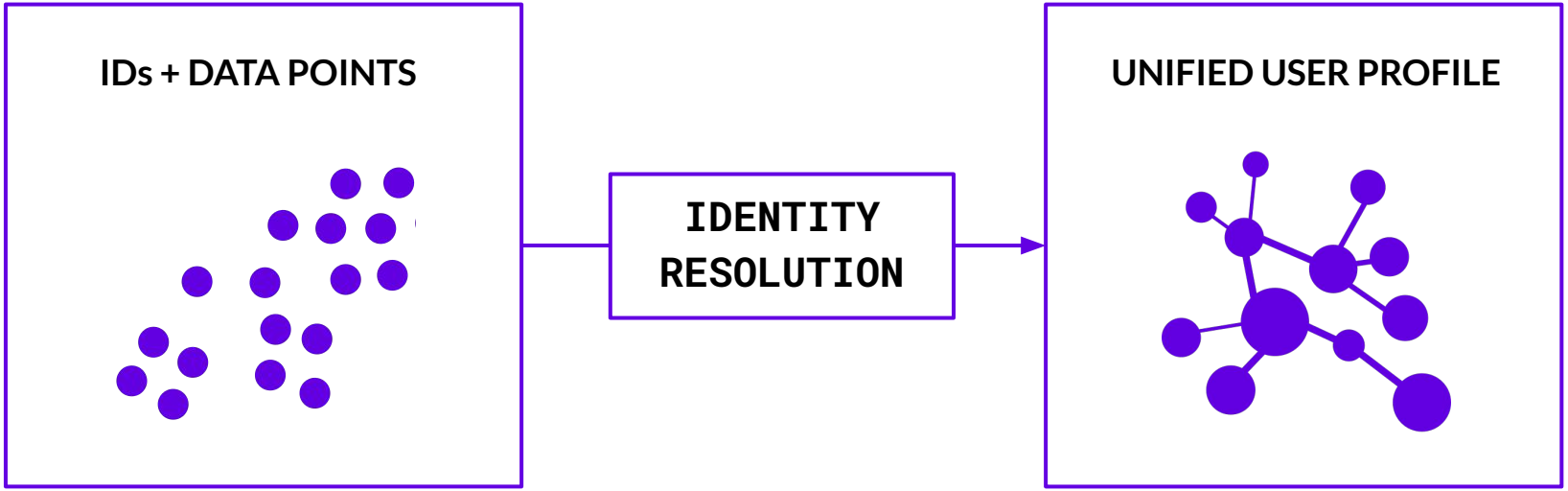
Convert data into common format

Clean data to remove discrepancies & duplicates

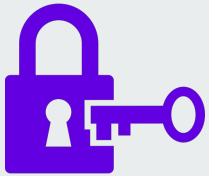
Standardise data points


[1] DATA INGESTION
[2] DATA NORMALISATION
[3] IDENTITY RESOLUTION
[4] GRAPH CREATION
[5] UPDATES + OPTIMISATION

Identity resolution is the process of matching data points from various sources to create a unified view of an individual.

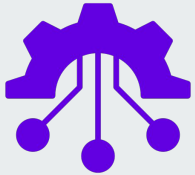

IDs + DATA POINTS
IDENTITY RESOLUTION
UNIFIED USER PROFILE

The process uses two matching methods…



DETERMINISTIC

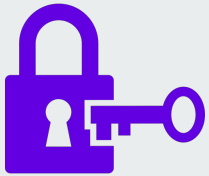
Directly link data points to an individual using unique, direct IDs (join keys)



PROBABILISTIC

Use models and algorithms to infer likelihood data points belong to same individual

...that work in tandem



DETERMINISTIC

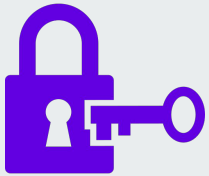
Directly link data points to an individual using unique, direct IDs (join keys)

PROBABILISTIC

Use models and algorithms to infer likelihood data points belong to same individual

Deployed together

...that work in tandem



DETERMINISTIC

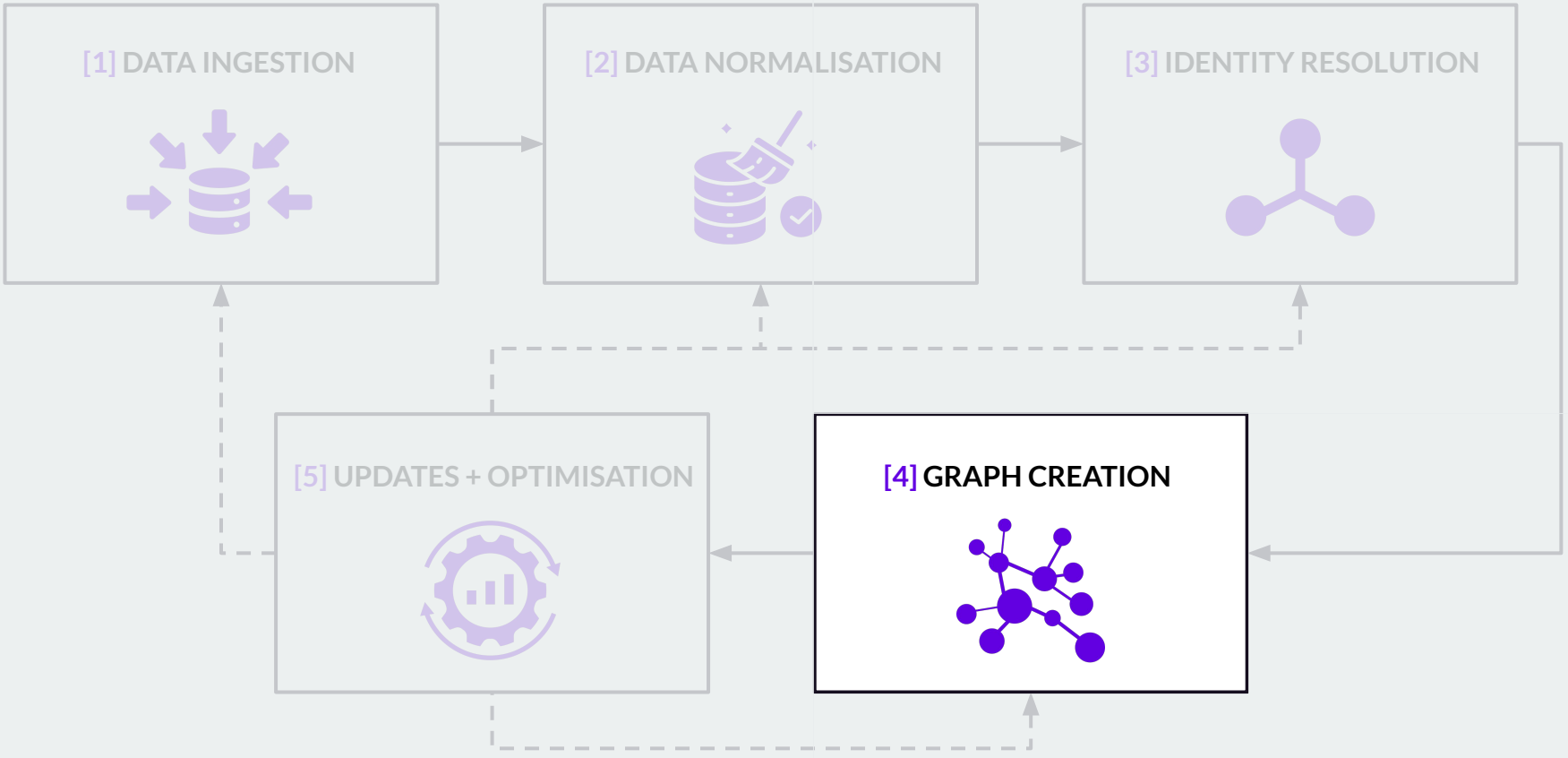
Directly link data points to an individual using unique, direct IDs (join keys)

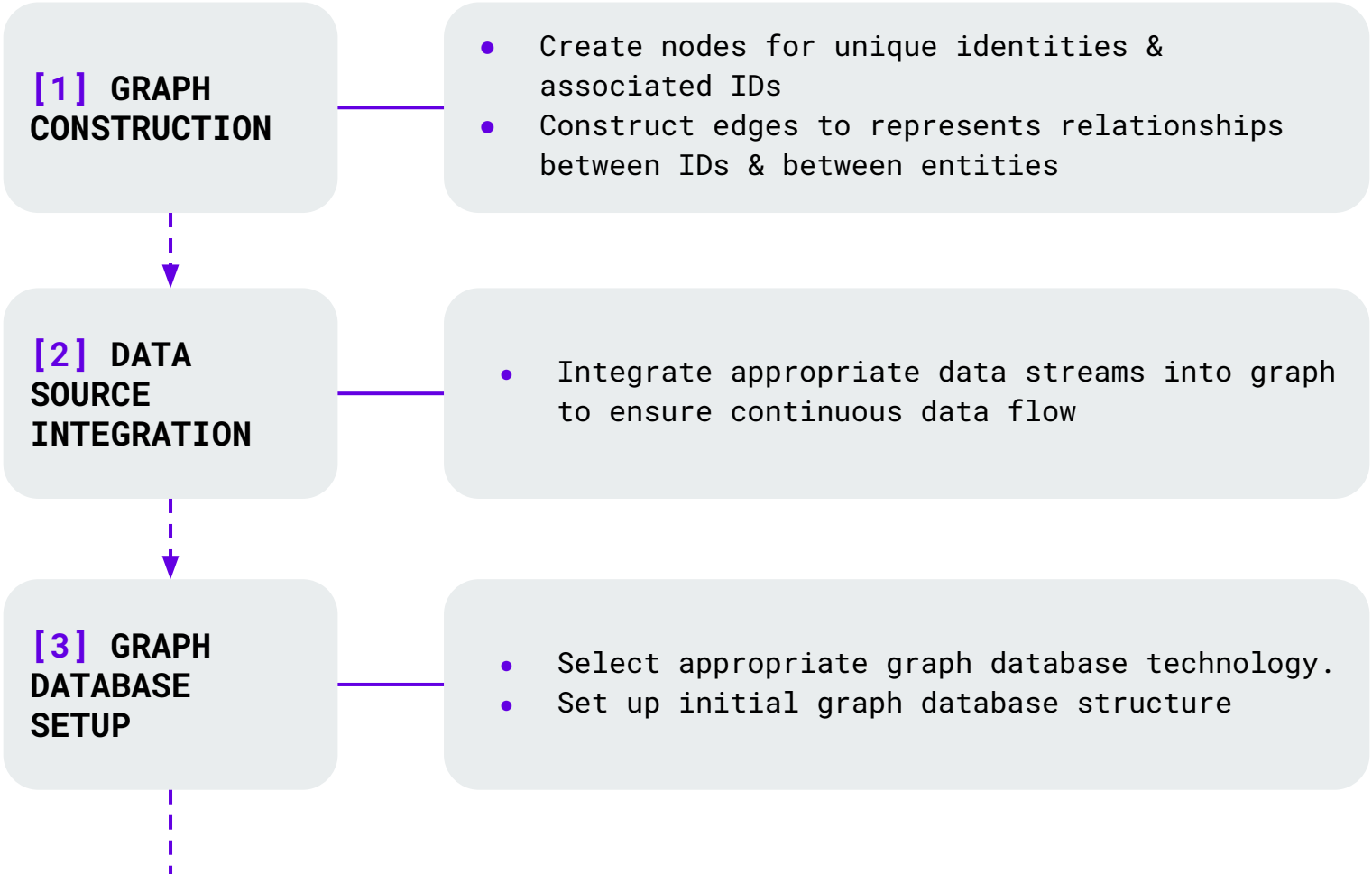
PROBABILISTIC

Use models and algorithms t
infer likelihood data po
belong to same indivi

CERTAIN METHODS ARE UNDER INCREASED SCRUTINY

Deployed together


[1] DATA INGESTION
[2] DATA NORMALISATION
[3] IDENTITY RESOLUTION
[4] GRAPH CREATION
[5] UPDATES + OPTIMISATION


[1] GRAPH CONSTRUCTION
Create nodes for unique identities & associated IDs
Construct edges to represents relationships between IDs & between entities
[2] DATA SOURCE INTEGRATION
Integrate appropriate data streams into graph to ensure continuous data flow
[3] GRAPH DATABASE SETUP
Select appropriate graph database technology.
Set up initial graph database structure

[4] TESTING AND QA

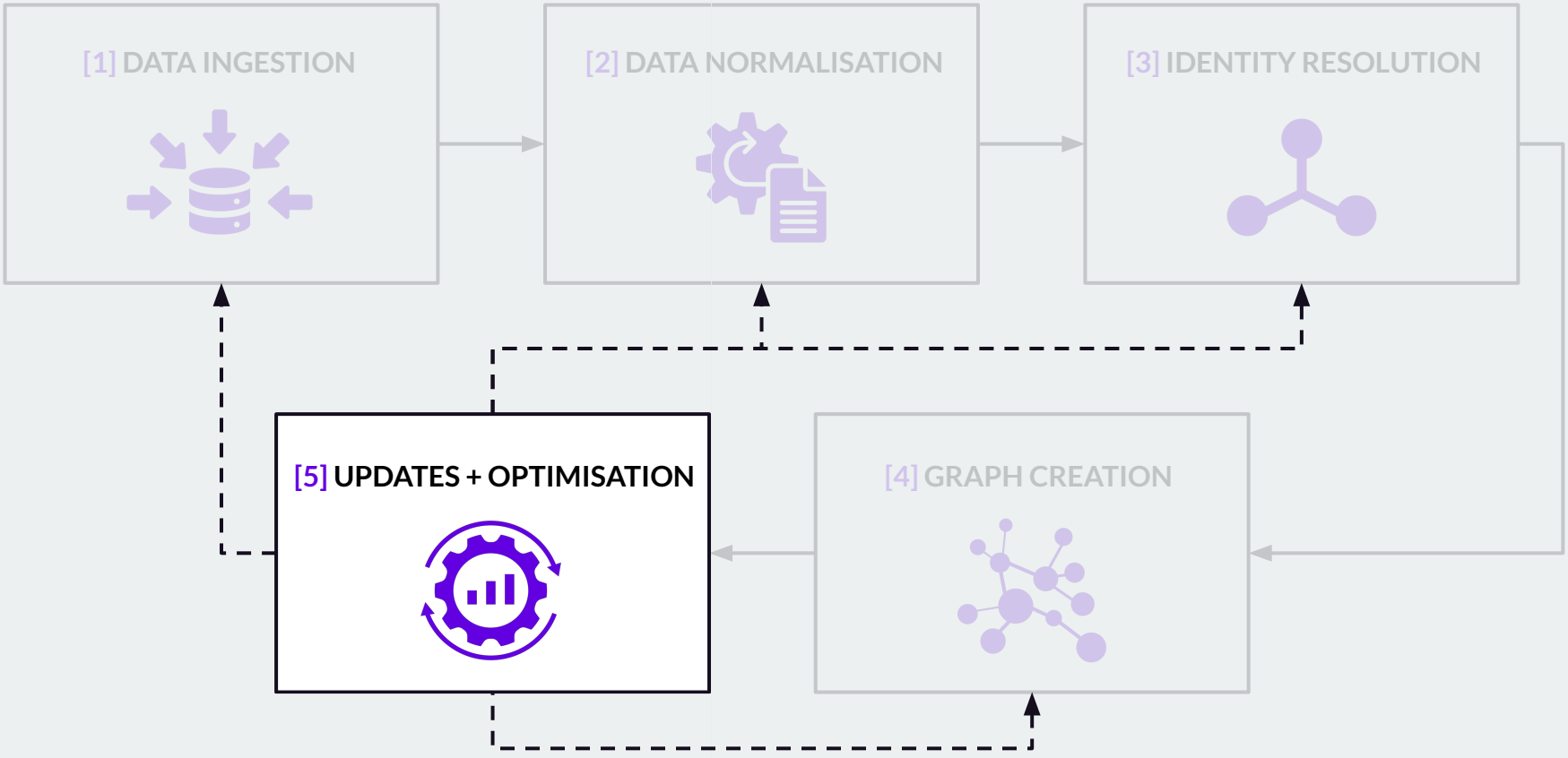
- Perform tests to ensure graph accurately reflect data relationships
- Validate graph integrity against known data sets

[5] PRIVACY AND SECURITY

- Establish data privacy protocols
- Implement security measures to protect data

[6] GRAPH ACTIVATION

- Integrate graph with MadTech platforms
- Make graph accessible for querying & insight generation


[1] DATA INGESTION
[2] DATA NORMALISATION
[3] IDENTITY RESOLUTION
[4] GRAPH CREATION
[5] UPDATES + OPTIMISATION

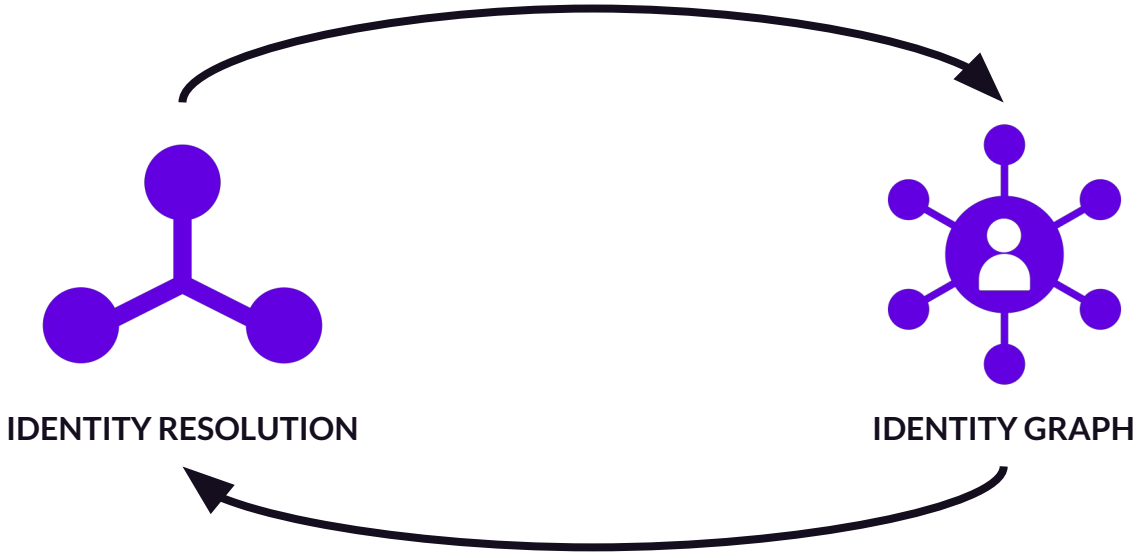
Regularly updating & optimizing the graph ensures it remains accurate & functional over time while also ensuring adherence to privacy & security standards. Here are some of the main steps that go into the process:

1. **(New) Data ingestion** - Gather new data from various sources while automating ingestion to facilitate real-time updates
2. **Data normalisation** - Check new data for veracity and clean to correct for discrepancies and ensure consistency
3. **Identity resolution updates** - Apply ID resolution to new data and update profiles with new identifiers and remove/update outdated data
4. **Database updates** - Add new nodes & edges for new identifiers and relationships while removing nodes & edges that are no longer relevant
5. **Model optimization** - Fine tune and improve probabilistic identity resolution algorithms based on new data patterns and insights
6. **Privacy compliance checks** - Regularly review and update data handling practices to ensure compliance with evolving regulations

How Identity Graphs enable marketing use cases

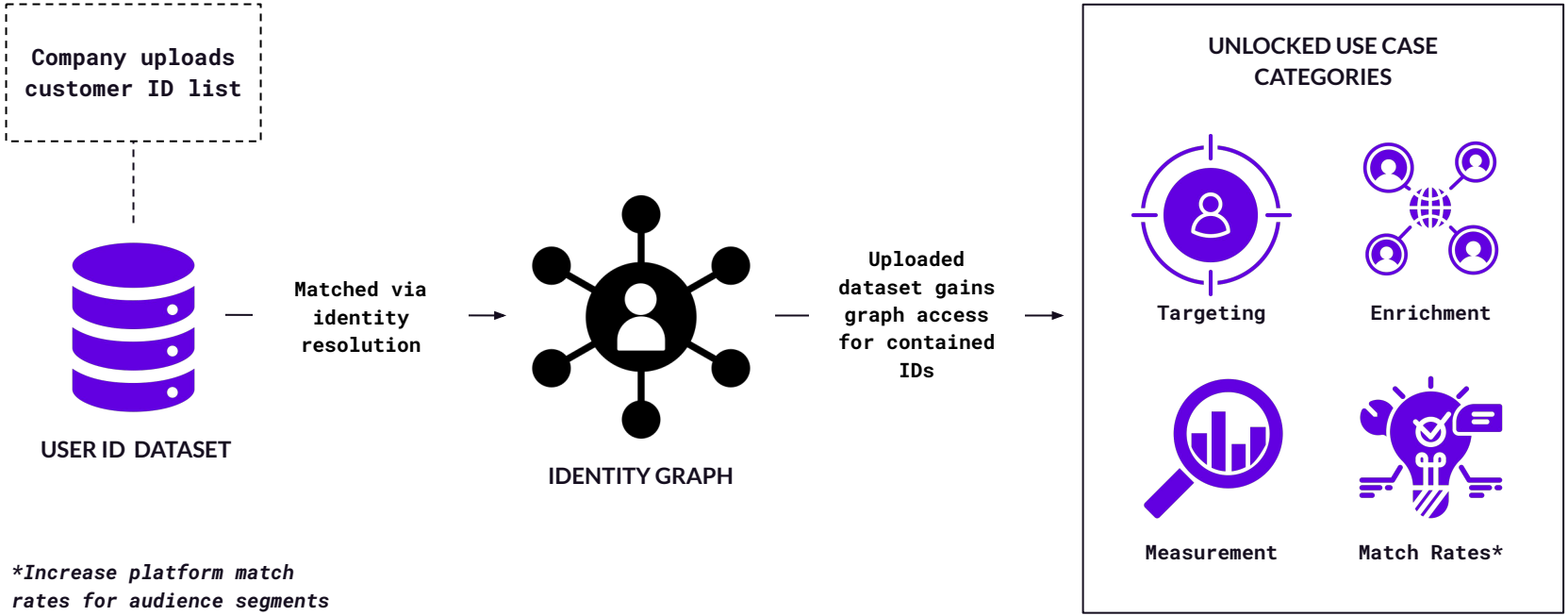
Identity Graphs facilitate marketing use cases primarily through identity resolution.

Identity Resolution to create Identity Graphs



Identity Graphs to facilitate Identity Resolution

Identity Graphs facilitate use cases by extending the utility of external datasets matched to the graph.



Increase platform match rates for audience segments

	How ID graph unlocks	Unlocked use cases (examples)
 **Targeting**	Matches uploaded IDs to existing resolved user profiles containing attached IDs across environments	• On-platform targeting • Cross-environment targeting & exclusion • Cross environment frequency capping
 **Enrichment**	Matches uploaded IDs to behavioral, interest, demographic attributes attached to existing resolved user profiles	• Audience insight generation • Audience discovery • Audience expansion/modeling • Audience segmentation

	How ID graph unlocks	Unlocked use cases (examples)
 **Measurement**	Matches uploaded IDs to existing resolved user profiles containing attached touchpoints (IDs and timestamps)	• Multi-touch attribution • Cross-channel attribution • Customer journey analysis • Lifetime value calculation
 **Match Rates**	Matches uploaded IDs to existing resolved user profiles containing attached IDs increasing match “surface area”	• Improve match rates for media activation platforms • Improve match rates for CDPs and Data Clean Rooms • Extend utility of Universal IDs

Where Identity Graphs show up (*kinda everywhere…*)

Identity Graphs are a core feature in all MadTech platforms and technologies with identity and audience based capabilities.

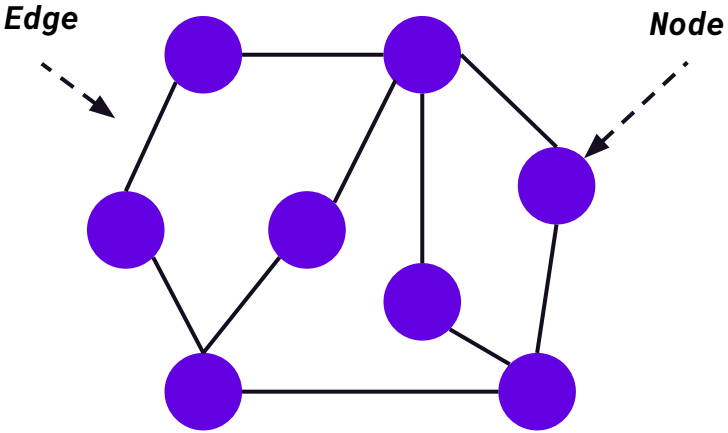
Platform/Tech	Enabled and/or Enhanced Functions (examples)
Media Buying Platforms	• Cross-device targeting, attribution, reporting • Enhanced user segmentation • Audience expansion/Lookalike modeling • Improved insight generation
SSPs/Ad Exchanges	• Cross-device campaign reporting • Improved buy-side/sell-side match rates • Deal/Inventory enrichment • Bid request enrichment
Web/App Analytics	• Cross-environment targeting, tracking, attribution • Enhanced user segmentation • Improved personalisation/UX optimisation • Advanced cohort analysis/behavioral insights • Privacy compliance/Consent management

Platform/Tech	Enabled and/or Enhanced Functions (examples)
Universal ID providers	• Single cross-ecosystem identifier • Identity resolution (improved match rates) • Cross-ecosystem targeting • Cross-ecosystem tracking/attribution
Data Clean Rooms	• Identity resolution (improved match rates) • Enhanced data collaboration utility (improves all use cases)
Customer Data Platforms	• Identity resolution (single customer view) • Enhanced customer insights • Cross-channel personalisation, tracking, attribution • Customer journey mapping • Privacy compliance/Consent management

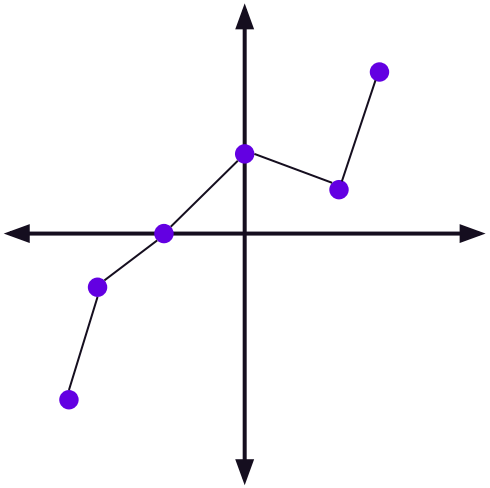
What the “Graph” in Identity Graph stands for

The “Graph” in Identity Graph refers to…

GRAPH THEORY

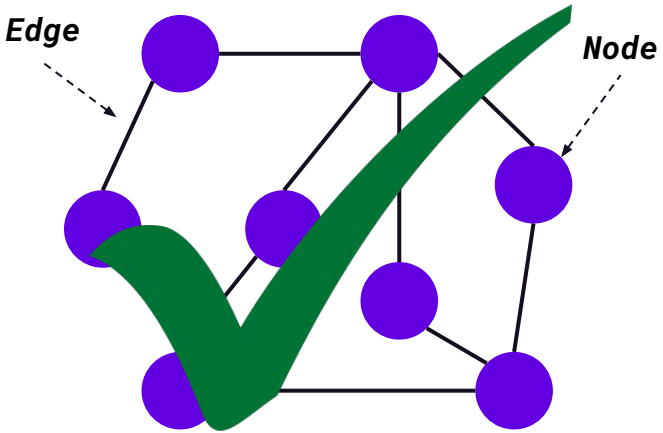


CARTESIAN PLANE

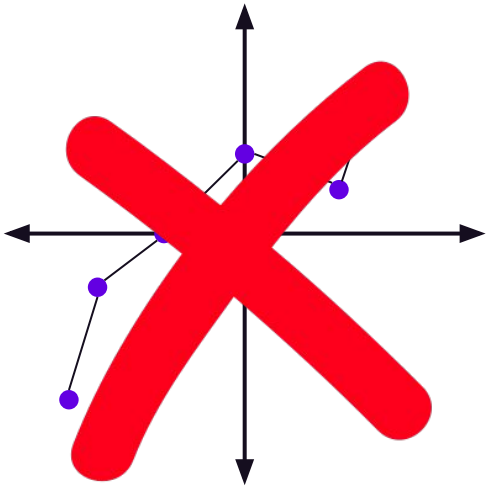


The “Graph” in Identity Graph refers to…

GRAPH THEORY



CARTESIAN PLANE



The role of Graph Databases

Graph databases are specialised databases designed to manage and query highly connected data in real-time.

Most Identity Graphs are built using graph databases.

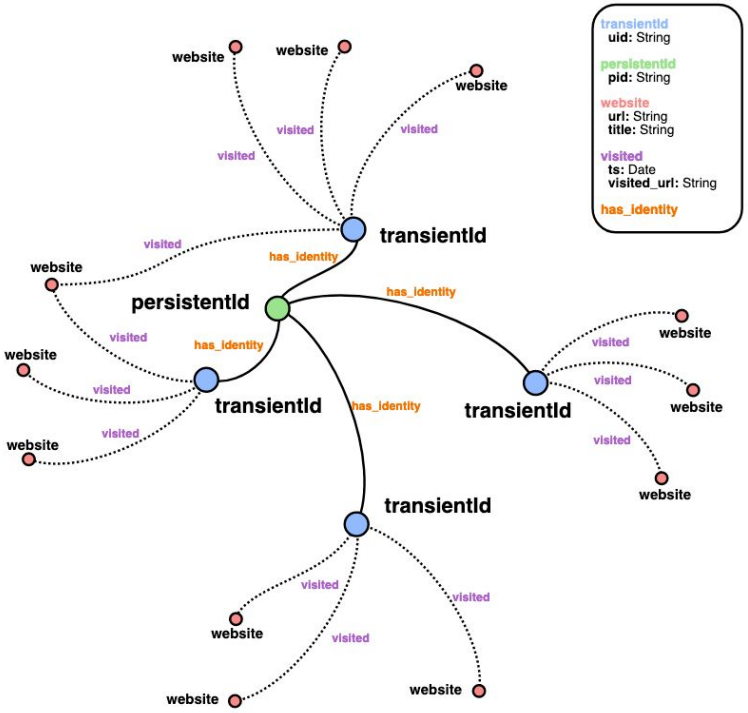
Prominent providers include:

neo4j

Dgraph

Amazon Neptune

ArangoDB



Identity Graph on Graph Database abstraction (AWS)

A simplified Identity Graph schema example

NODES (Entities)

1. USER NODE (Primary)

- **Properties:** UserID, FirstName, LastName, Email
- **Example:** UserID:"U123", FirstName:"Alice", LastName:"Smith", Email:"alice@example.com"

2. DEVICE NODE

- **Properties:** DeviceID, Type, OS
- **Example:** DeviceID:"D456", Type:"Mobile", OS:"iOS"

3. INTERACTION NODE

- **Properties:** InteractionID, Type, Timestamp, URL
- **Example:** InteractionID:"I789", Type:"Purchase", Timestamp:"2023-01-15T12:34:56Z", URL:"/product/42"

4. SESSION NODE

- **Properties:** SessionID, StartTime, IP
- **Example:** SessionID: "S101", StartTime: "2023-01-15T11:00:00Z", EndTime: "2023-01-15T12:00:00Z", IP: "198.51.100.1"

EDGES (Relationships)

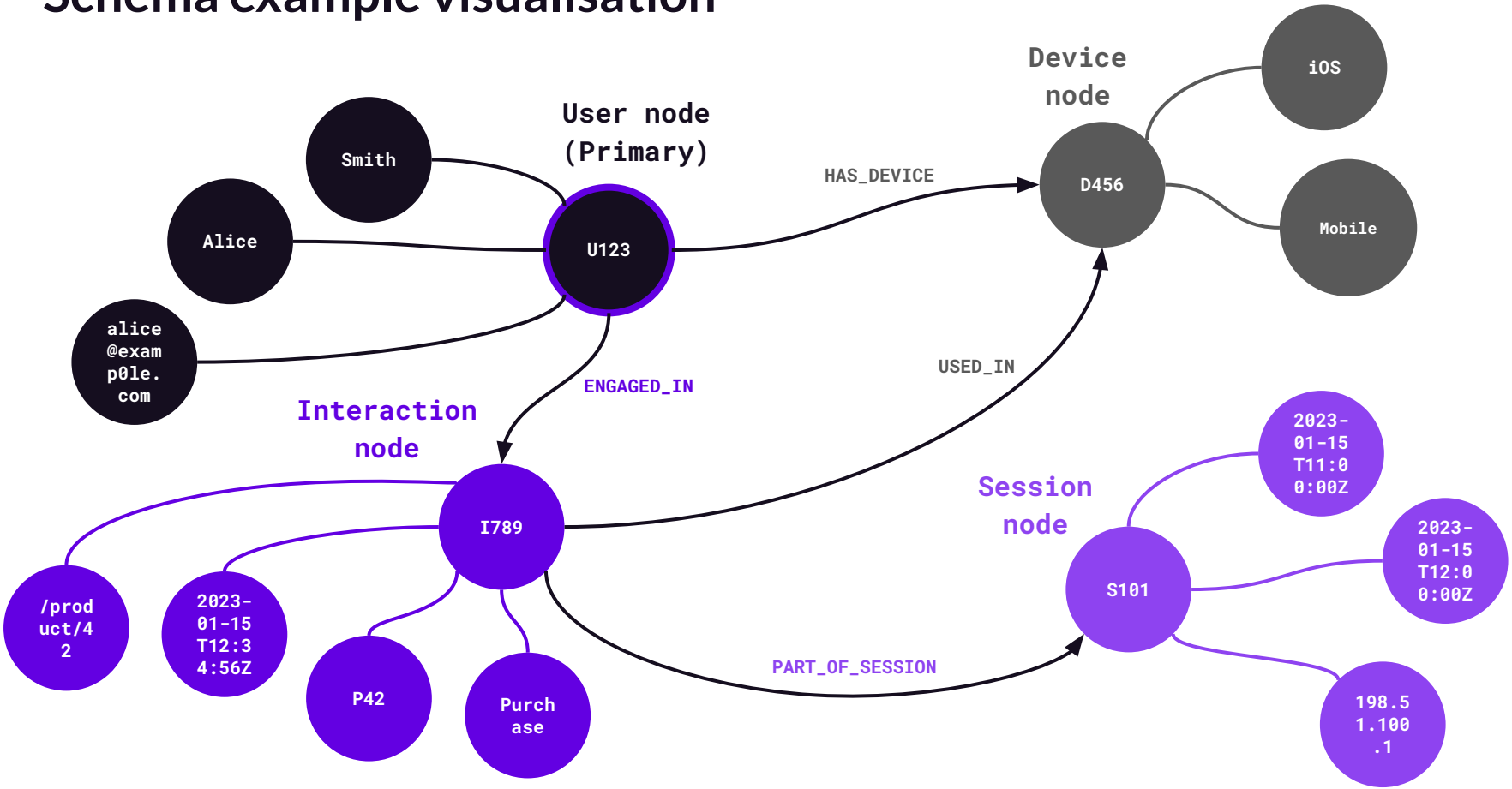
- **HAS_DEVICE**: Connects User to Device
- **ENGAGED_IN**: Connects User to Interaction
- **USED_IN**: Connects Device to Interaction, indicating which device was used for a particular interaction
- **PART_OF_SESSION**: Connects Interaction to Session, indicating which session includes a particular interaction

Schema example as a code snippet

```
(User:User {UserID: "U123", FirstName: "Alice", LastName: "Smith", Email: "alice@example.com"})
     [:HAS_DEVICE](Device:Device {DeviceID: "D456", Type: "Mobile", OS: "iOS"})
     [:ENGAGED_IN](Interaction:Interaction {InteractionID: "I789", Type: "Purchase", Timestamp:
     "2023-01-15T12:34:56Z", URL: "/product/42", ProductID: "P42"})
               [:USED_IN](Device:Device {DeviceID: "D456", Type: "Mobile", OS: "iOS"})
               [:PART_OF_SESSION](Session:Session {SessionID: "S101", StartTime:
               "2023-01-15T11:00:00Z", EndTime: "2023-01-15T12:00:00Z", IP: "198.51.100.1"})
```

- **Each User node is**
  - Connected to a Device node via a HAS_DEVICE relationship.
  - Connected to an Interaction node via an ENGAGED_IN relationship
- **Each Interaction node is**
  - Connected back to the Device node that it was USED_IN.
  - Connected to a Session node via a PART_OF_SESSION relationship

Schema example visualisation



While most ID Graphs are built on Graph Databases, an ID Graph is defined by its ability to link identifiers to individual users, not the underlying tech. Thus, there are exceptions where ID graphs are built using tech other than graph databases.

Thanks!

contact@skeletonkey.digital