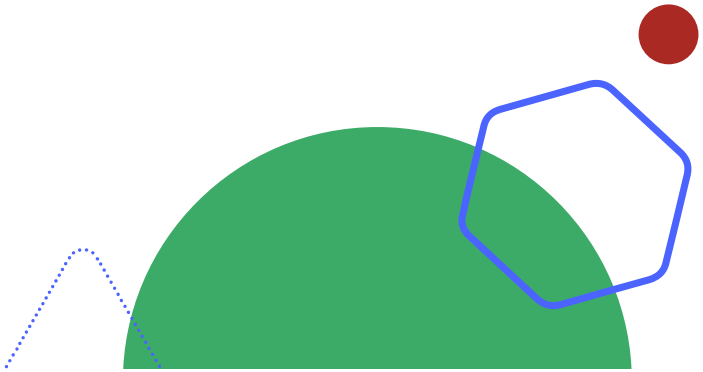
ADVANCING ANALYTICS

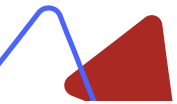
Data Watch

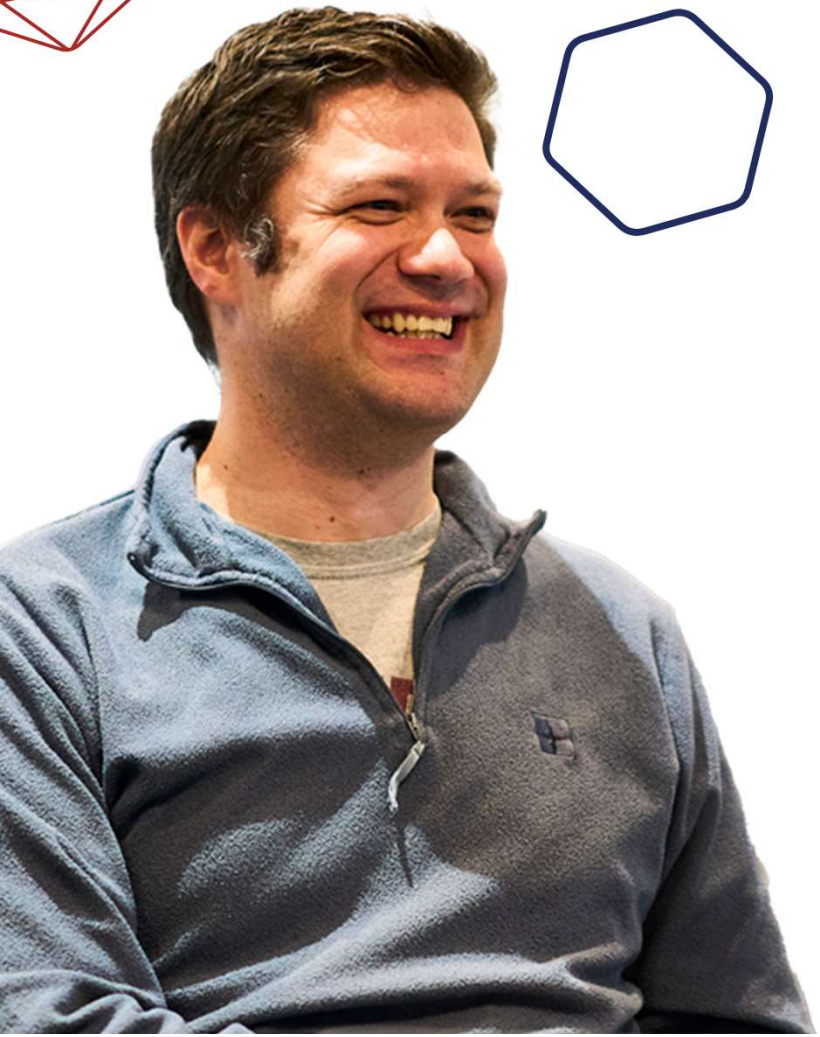
Seeing the Unseen in Data Platforms and Products through Observability



Agenda

- What is Observability
- Why observability is important
- How to observe (Demos)



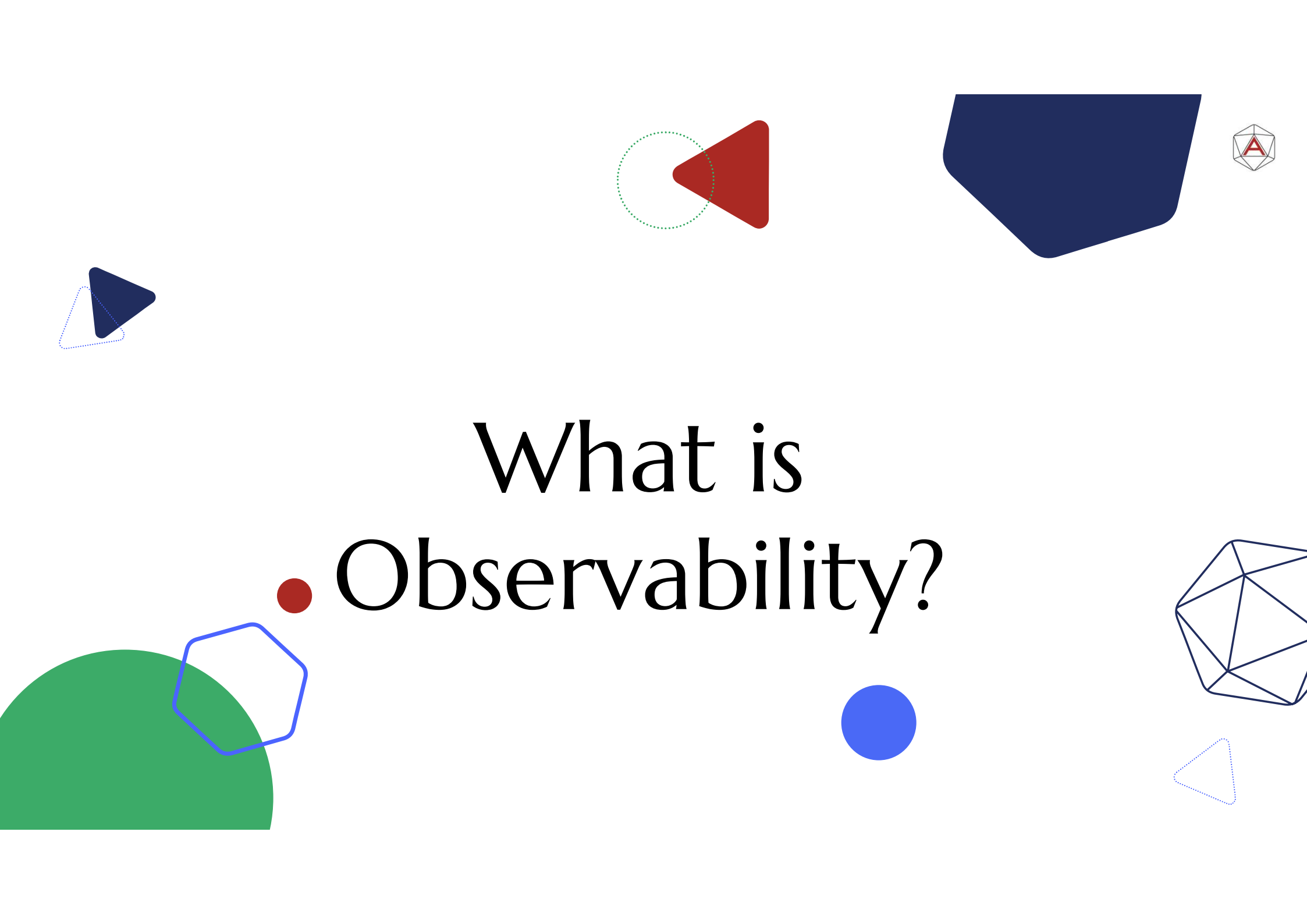


About Ust.

Head of Analytics at Advancing Analytics

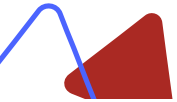
LinkedIn: Ust Oldfield
X (Twitter): @UstDoesTech
Email: ust@advancinganalytics.co.uk

databricks
PARTNER PROGRAM
SOLUTIONS ARCHITECT CHAMPION


What is
Observability?

It is the ability to infer the internal state of a system based on its external outputs.

https://en.wikipedia.org/wiki/Observability



What are we observing?

Metrics

Metadata

Lineage

Logs

Behaviour

Metrics

Freshness

Volume

Duplication

Accuracy

Metadata

~~Meta~~data about the data

Name: Product Orders

Sensitivity: Commercial Sensitive

Location: AdventureWorks

Schema:

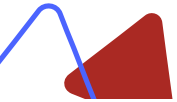
ID	INT
SKU	STRING
Location	STRING
Qty	INT
Updated	DATE
Created	DATE

Processing: Databricks

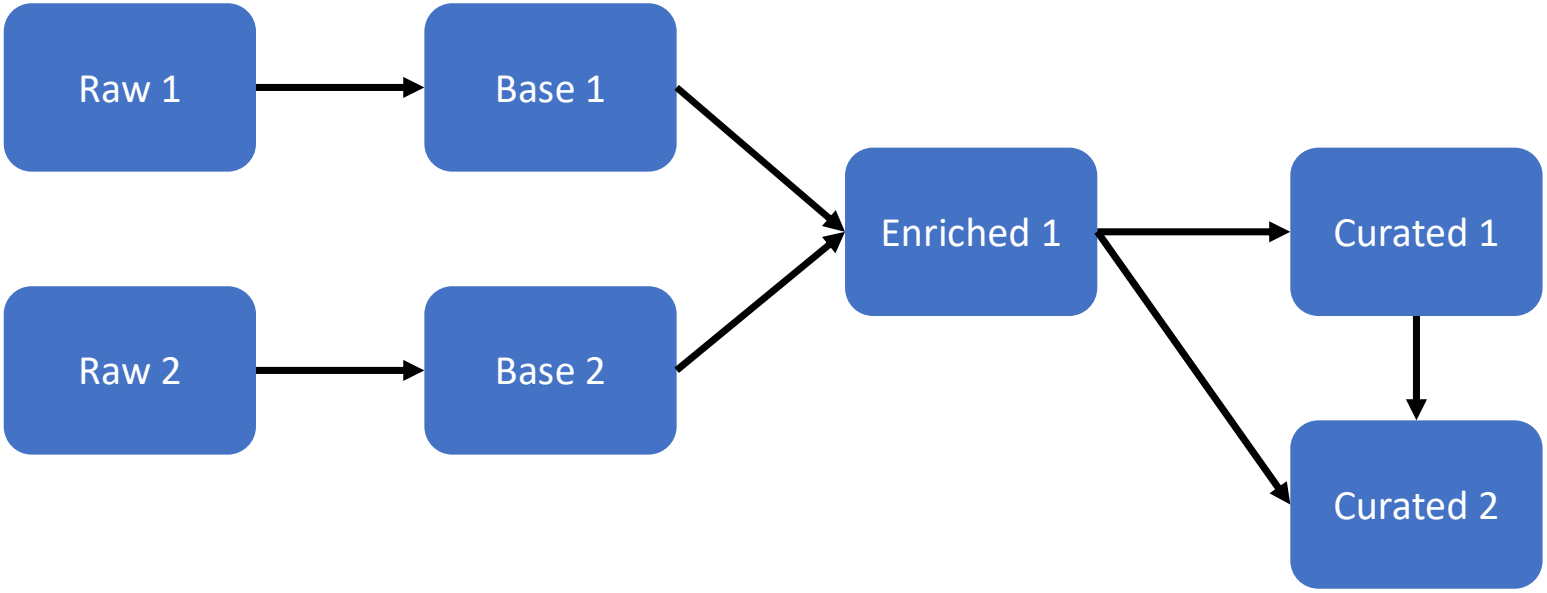
Owner: Sales

Format: Delta

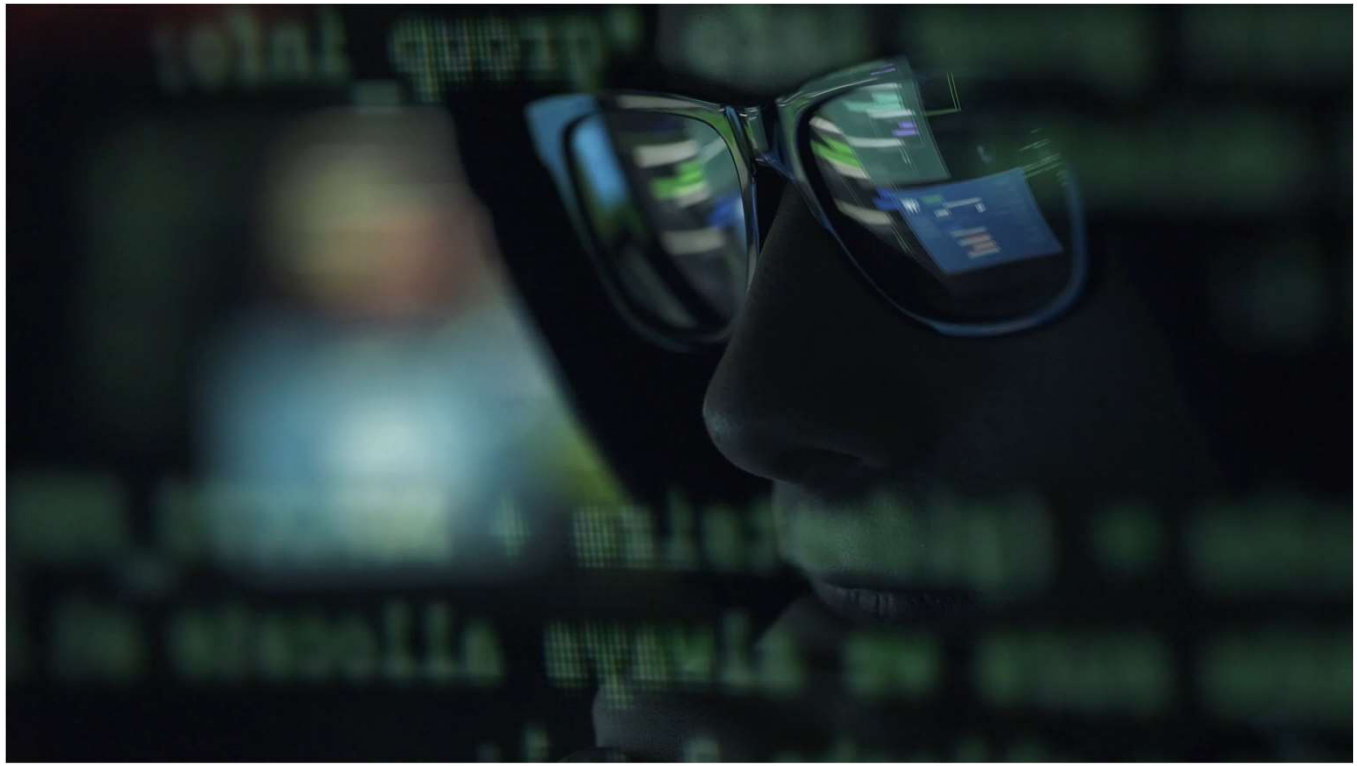
Consumers: Finance, Product



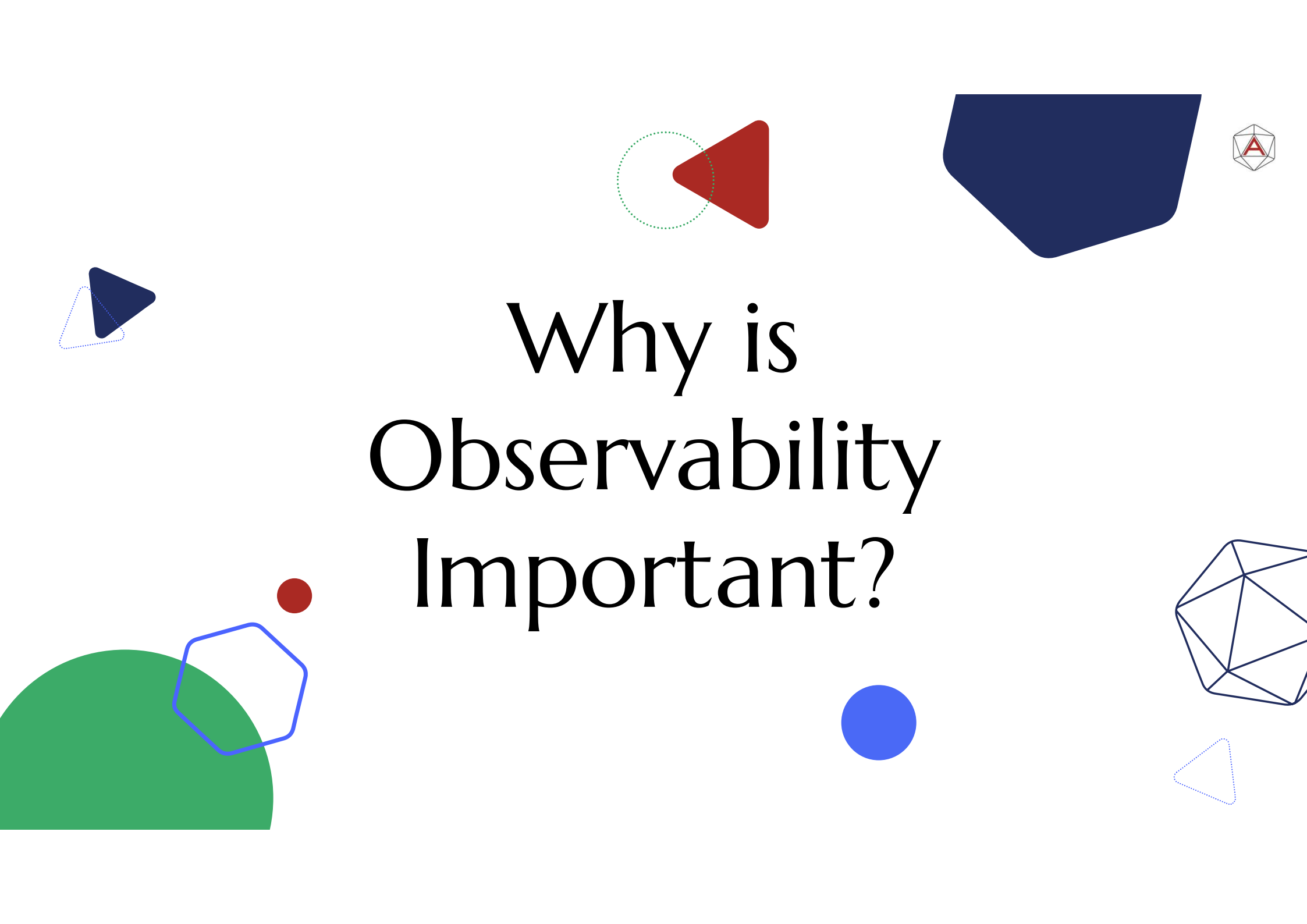
Lineage

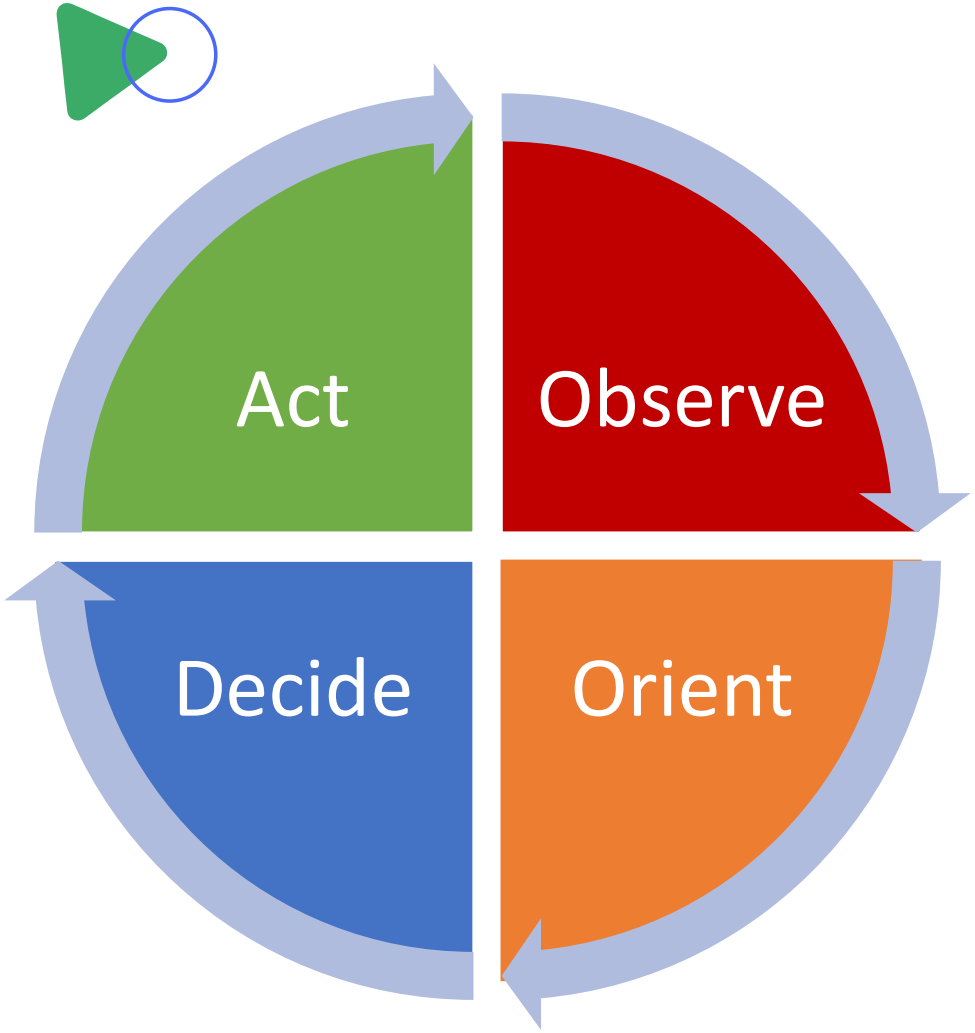


Logs



Behaviour


Why is
Observability
Important?

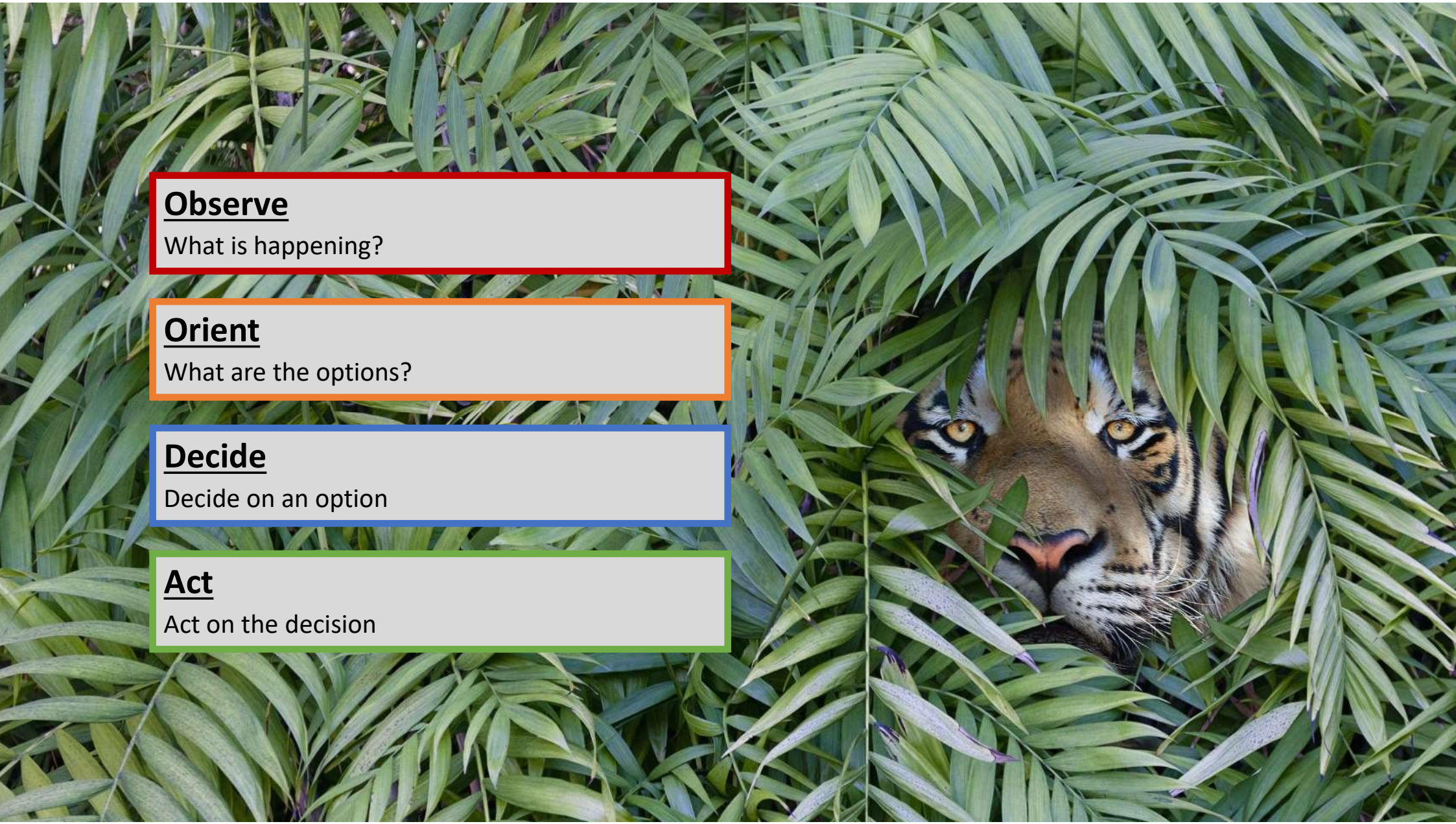


Observe
What is happening?

Orient
What are the options?

Decide
Decide on an option

Act
Act on the decision


Observe
What is happening?
Orient
What are the options?
Decide
Decide on an option
Act
Act on the decision

Build it and they will come...?

Success

Financially Viable

Users Like It

Users Use It

Behavioural Change

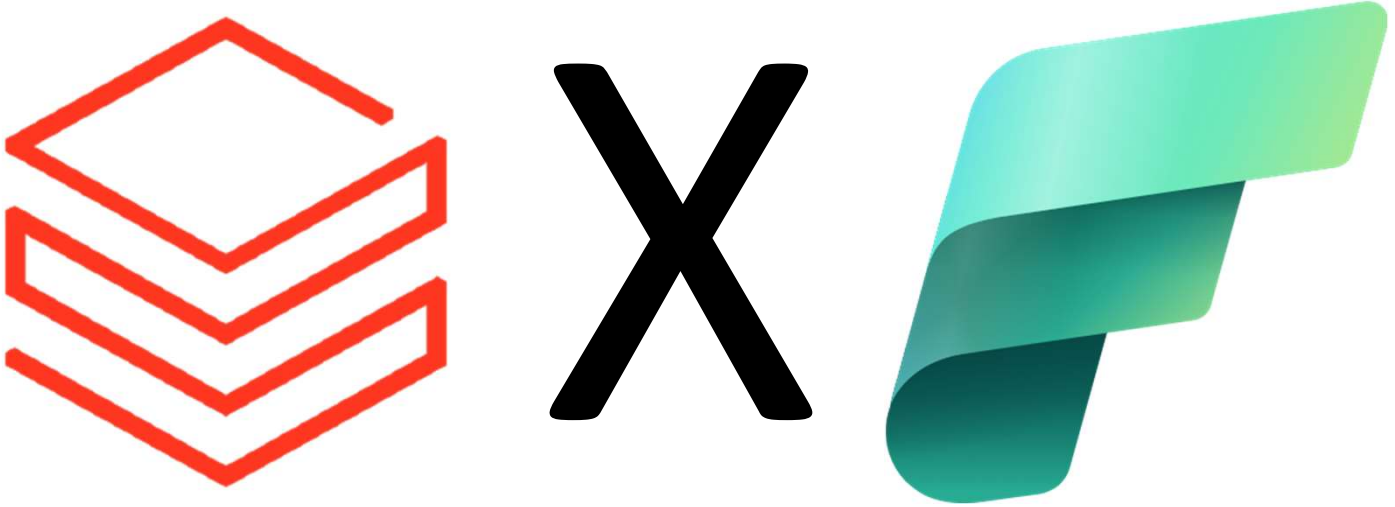
Data Strategy

Data Culture

Hawthorne Effect

How to implement Observability

Data Platforms



Platform Observability

Metrics

Metadata

Lineage

Logs

Behaviour

Metrics Questions

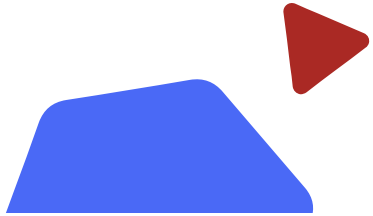
How is the platform's overall performance? Are there any bottlenecks?

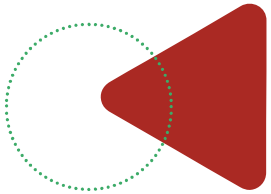
How utilised are the resources?

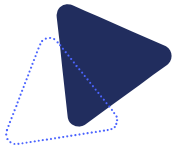
How frequently do data processing jobs fail or succeed?

How long do different jobs and processes take to complete?

How is the platform handling increased data loads and user demands?

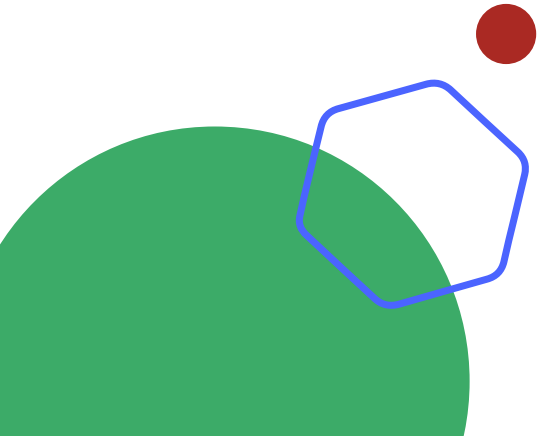






Data Platform Metrics

Demo Time

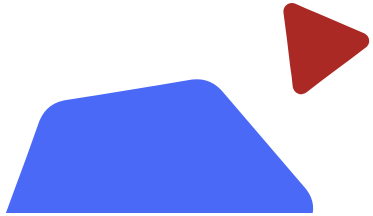


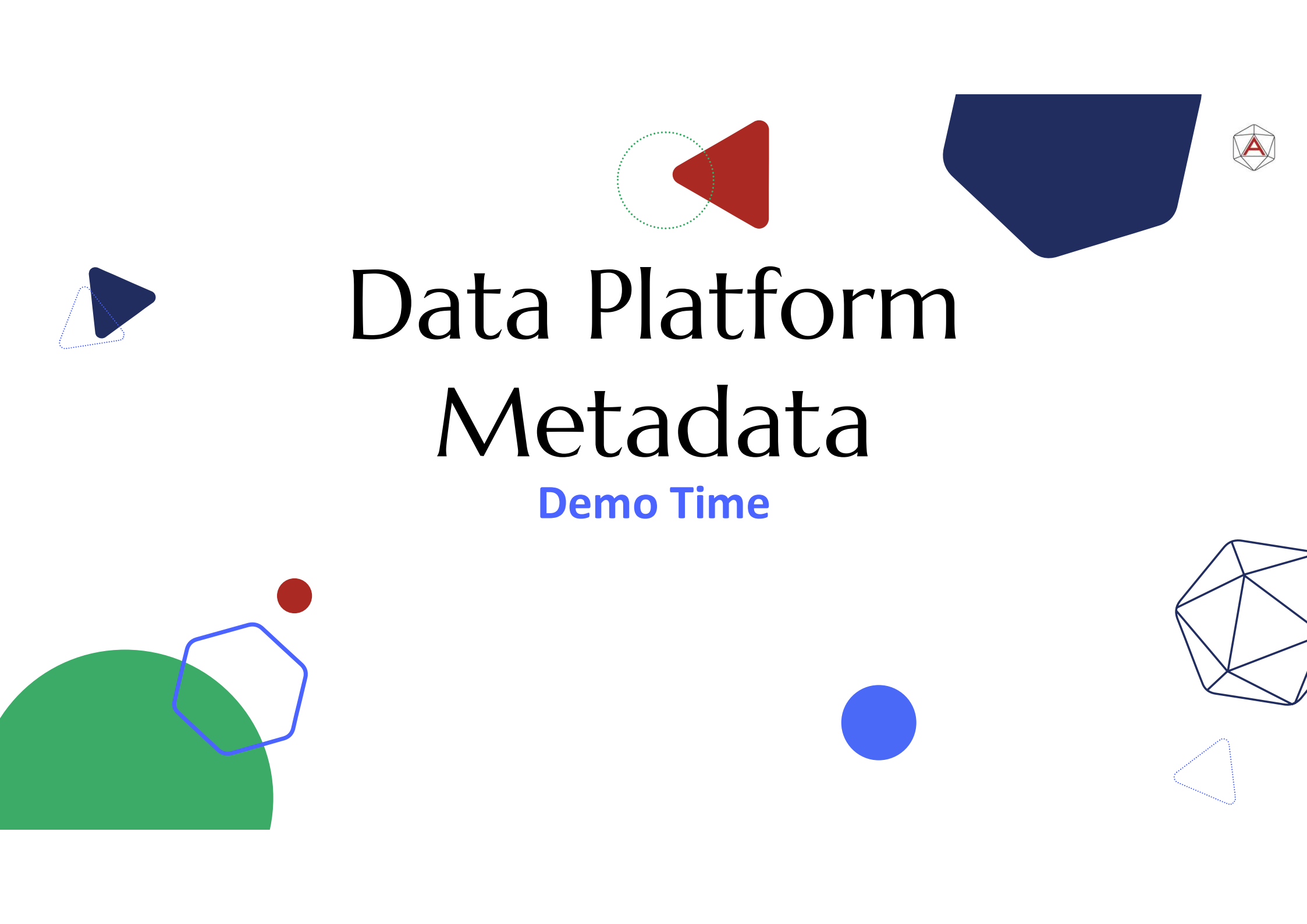
Metadata Questions

What are the current configurations of the various components within the data platform?

What versions of software, libraries, and dependencies are currently in use across the platform?

When were the last updates or patches applied to the system components?

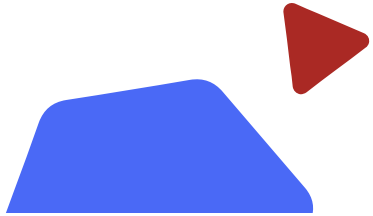


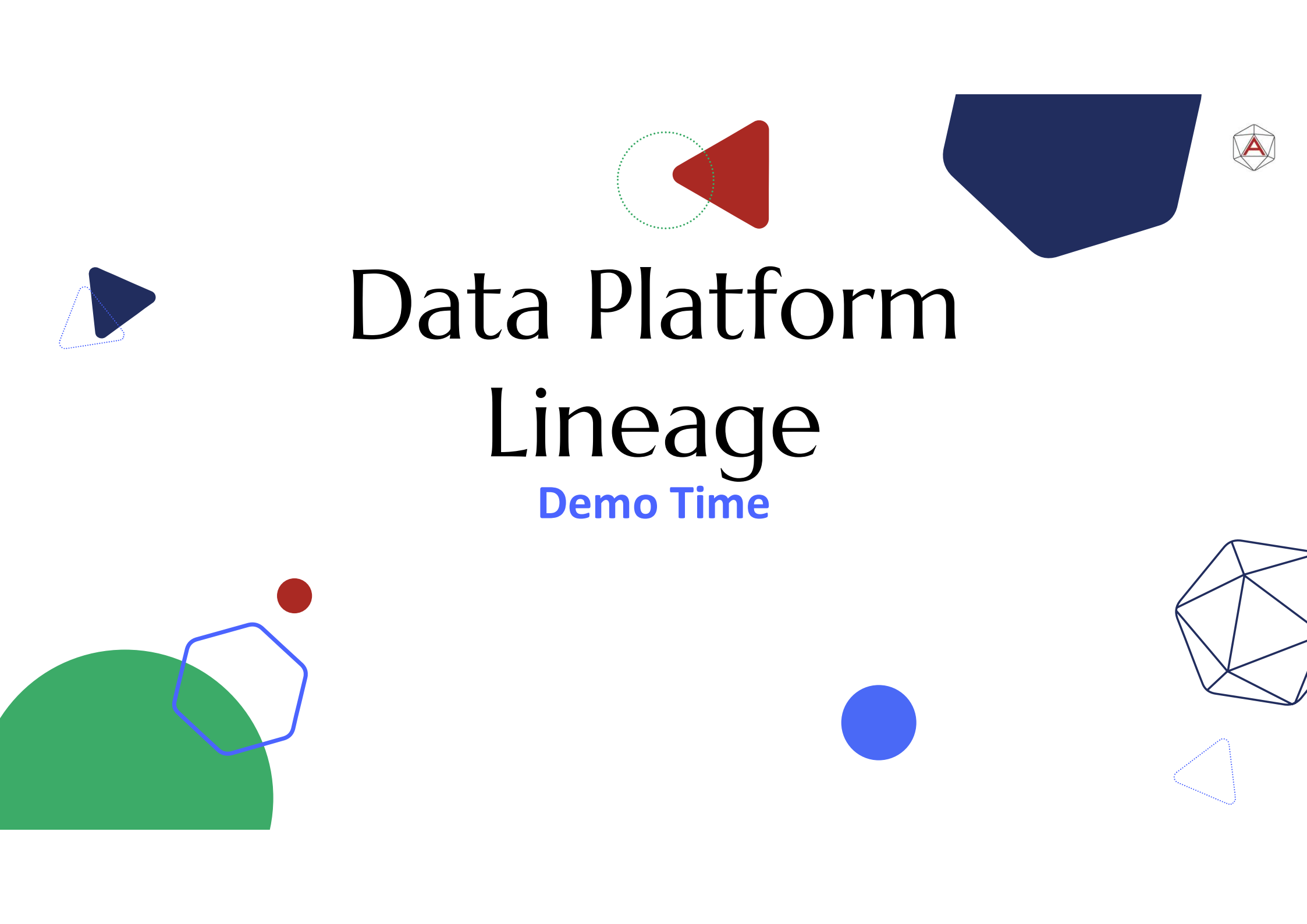

Data Platform
Metadata
Demo Time

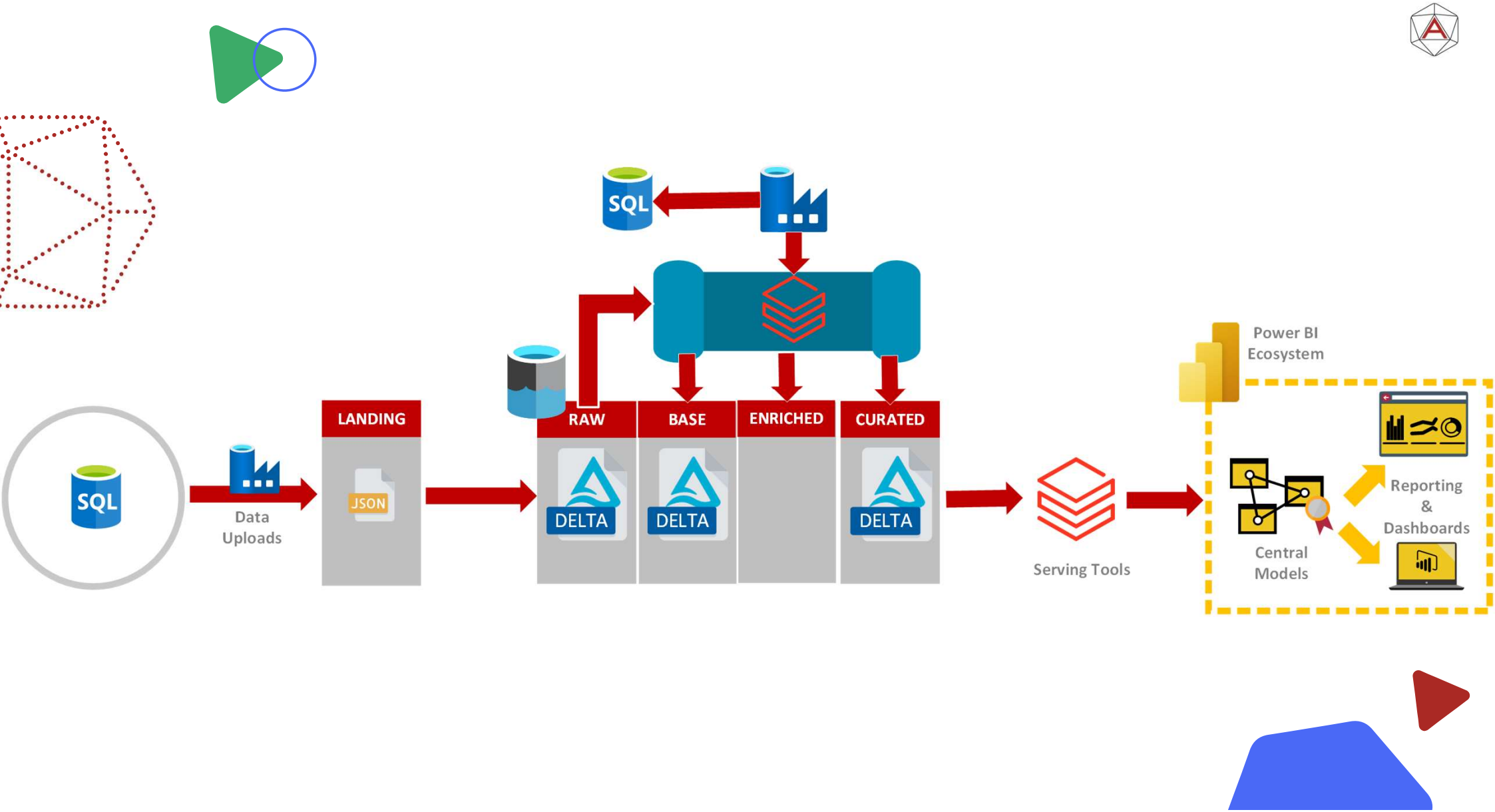
Lineage Questions

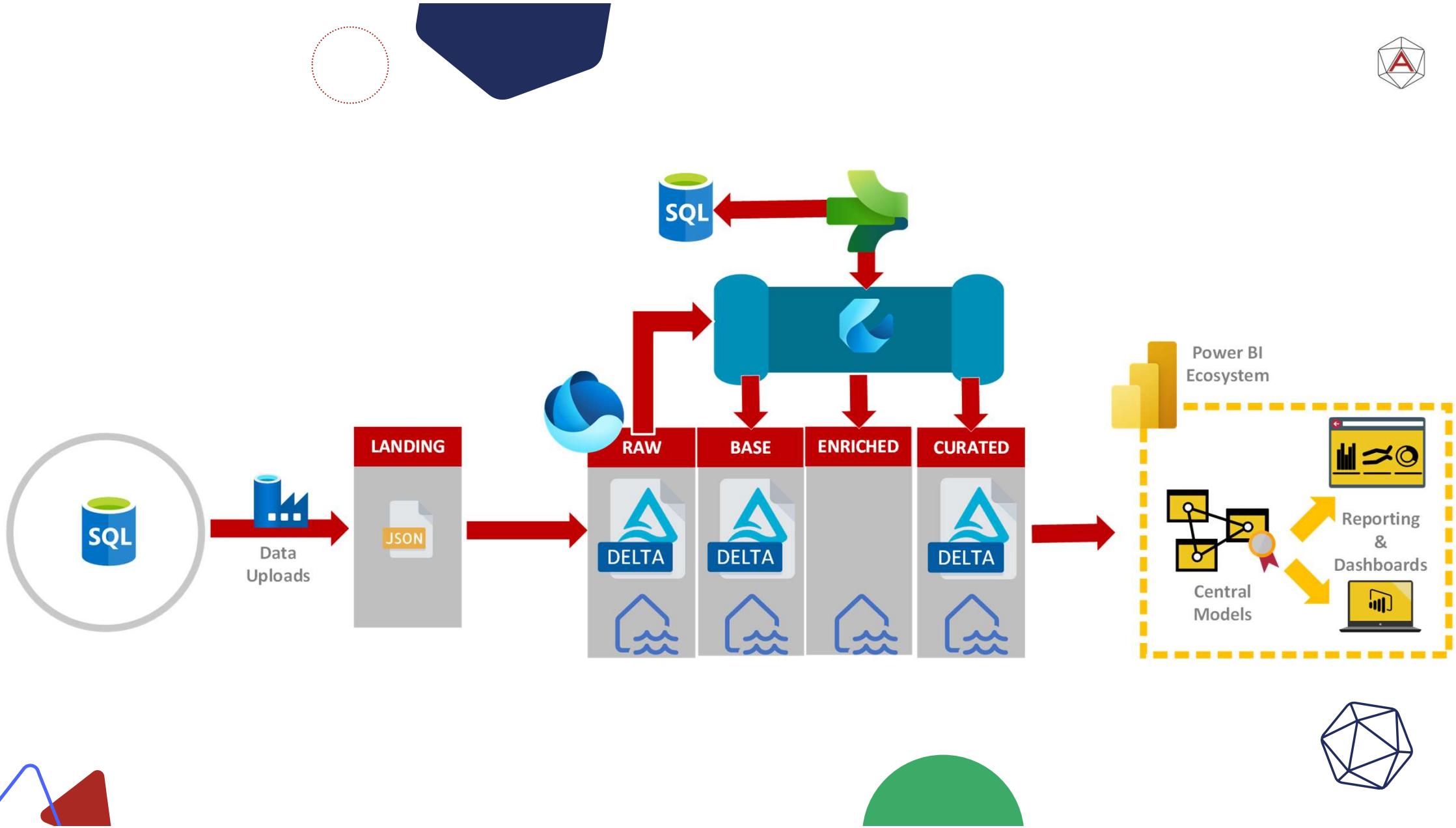
What are the dependencies between components?

What are the dependencies processing schedules?




Data Platform
Lineage
Demo Time

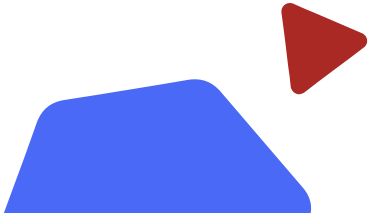

SQL
Data
Uploads
LANDING
JSON
SQL
RAW
BASE
ENRICHED
CURATED
DELTA
DELTA
DELTA
Serving Tools
Power BI
Ecosystem
Central
Models
Reporting
&
Dashboards

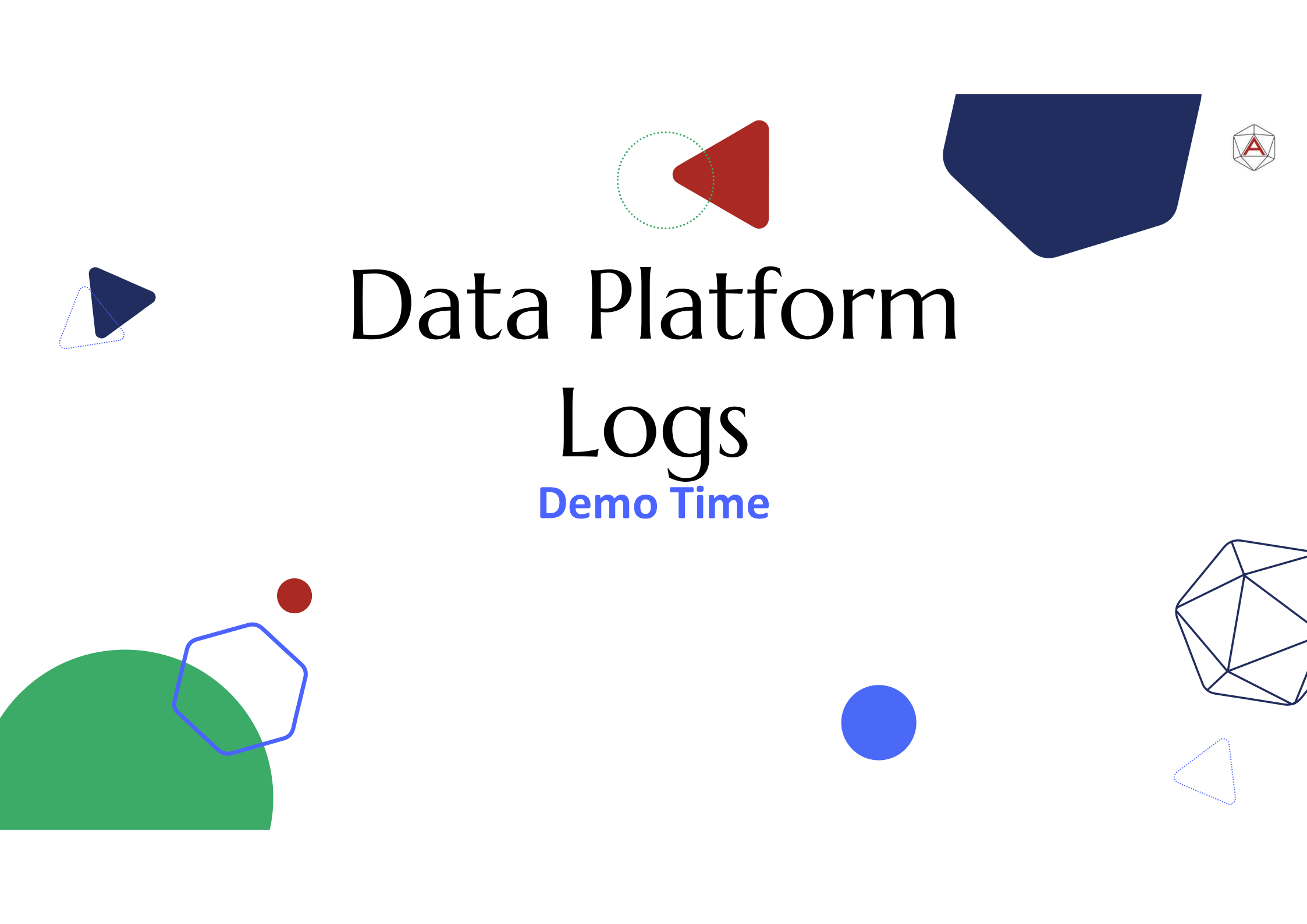

SQL
Data
Uploads
LANDING
JSON
SQL
RAW
BASE
ENRICHED
CURATED
DELTA
DELTA
DELTA
Power BI
Ecosystem
Central
Models
Reporting
&
Dashboards

Logs Questions

What events happened that impacted the performance of the platform?

Who or what is using the platform?




Data Platform
Logs
Demo Time

Platform Observability Impact

Optimised Performance

Increased Reliability

Scalability

Regulatory Compliance

Cost Efficiency

Data Products

Defining Data Products

A domain-bound set of data services, including the underlying infrastructure & code to produce/maintain

It Is:

- A Data Model
- A Reporting Table
- ML Model

It is NOT:

- A Dashboard
- An Application
- An Entire Data Warehouse / Lake

Product Observability

Metrics

Metadata

Lineage

Logs

Behaviour

Metrics Questions

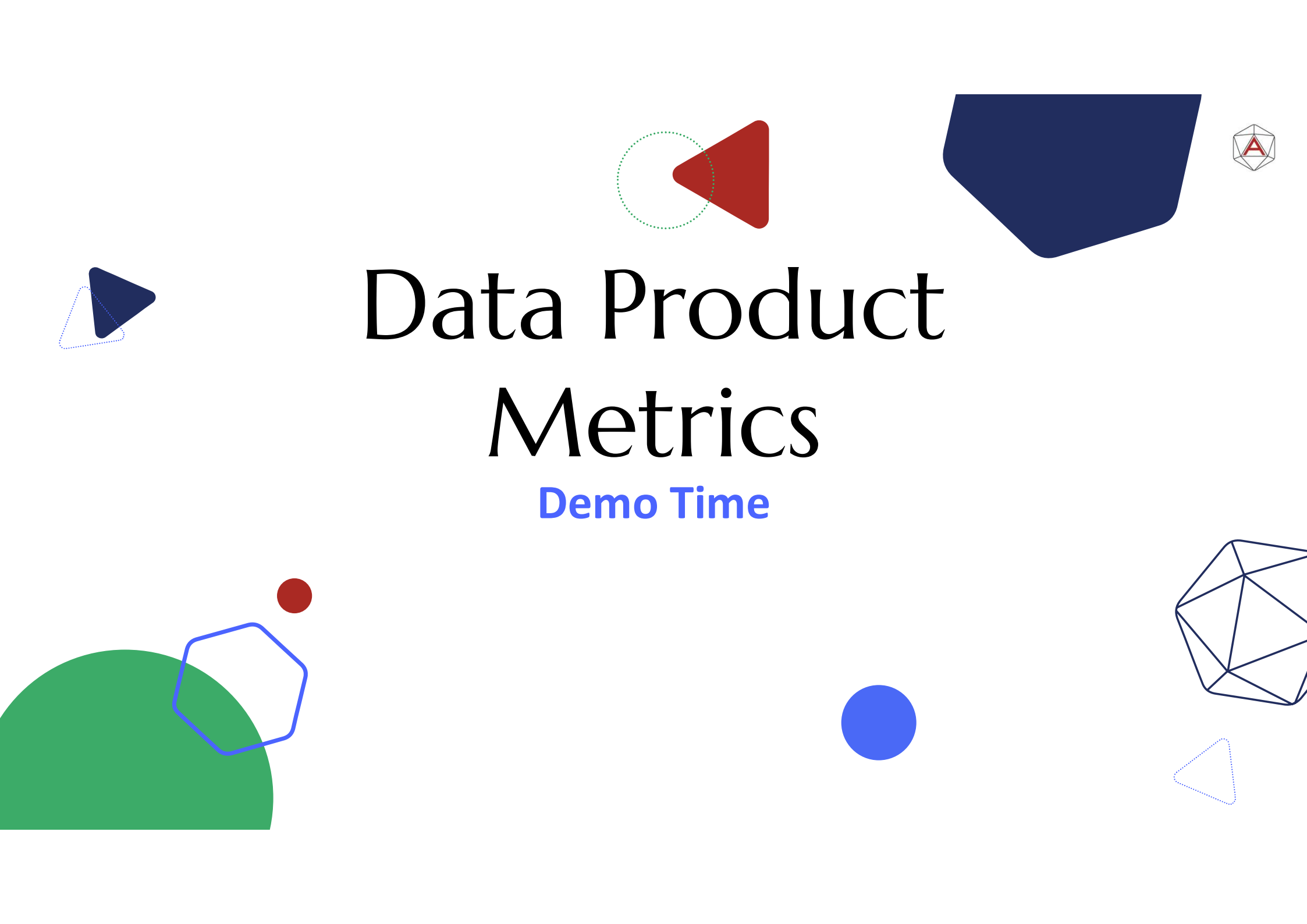
How current is the data available to users?

What's the Data Quality of the Data Product?

How accurate is the data product?

How quickly are queries being processed?

What is the response time for data access requests?


Data Product
Metrics
Demo Time

great expectations

Setup

- Stores
- Plugins
- Data Docs

Connect

- Data Sources

Expectations

- Expectations
- Profilers

Validate

- Checkpoints

Manges

- Metrics
- Validation Results
- Data Docs Contents

Metadata Questions

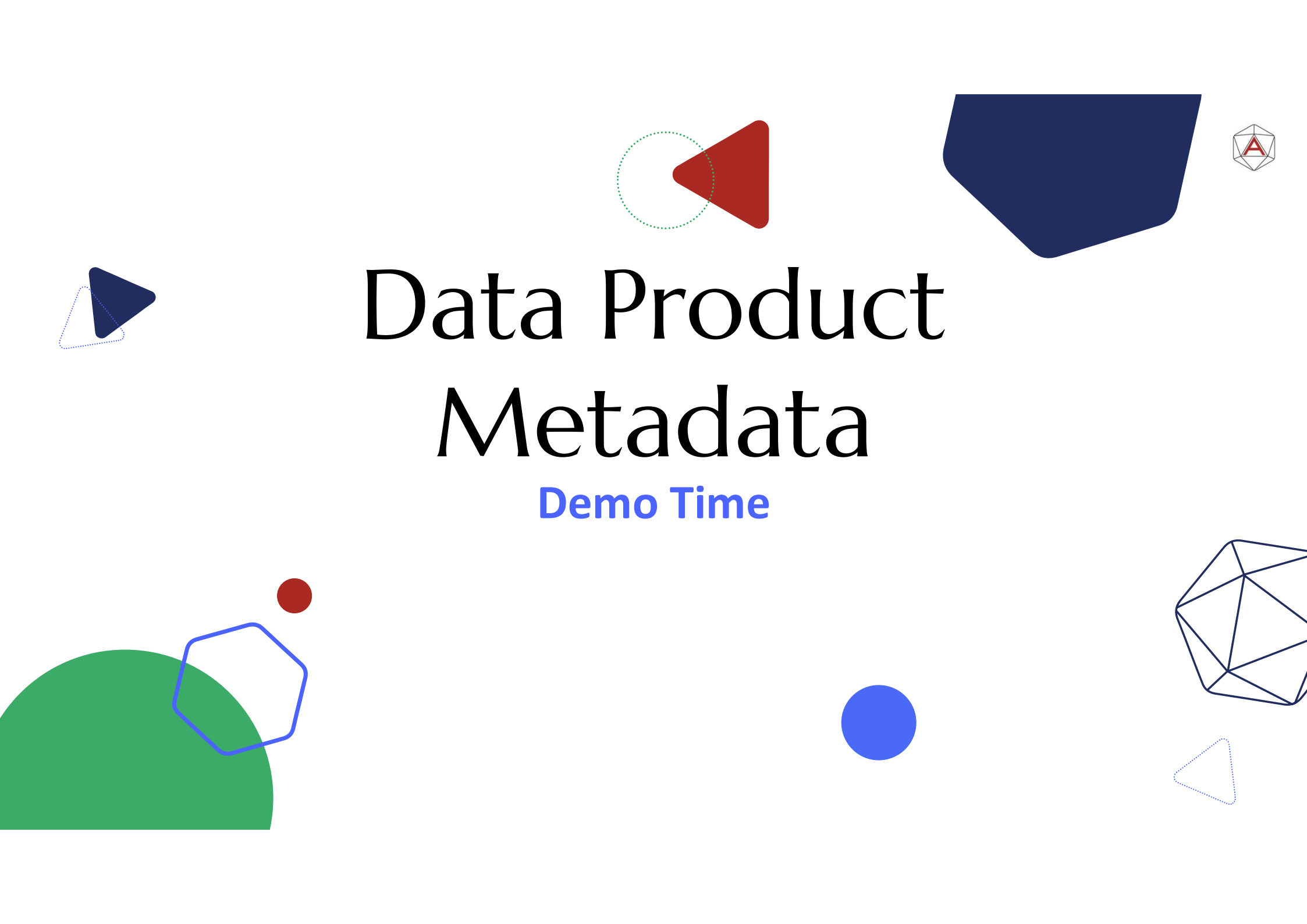
What is the classification for the Data Product?

Who can use or see the Data Product?

What are the DQ rules and validation checks for this data product?

What is the schema of the data product, including data types and constraints?

What tags and descriptions help in discovering and understanding this data product?


Data Product
Metadata
Demo Time

Lineage Questions

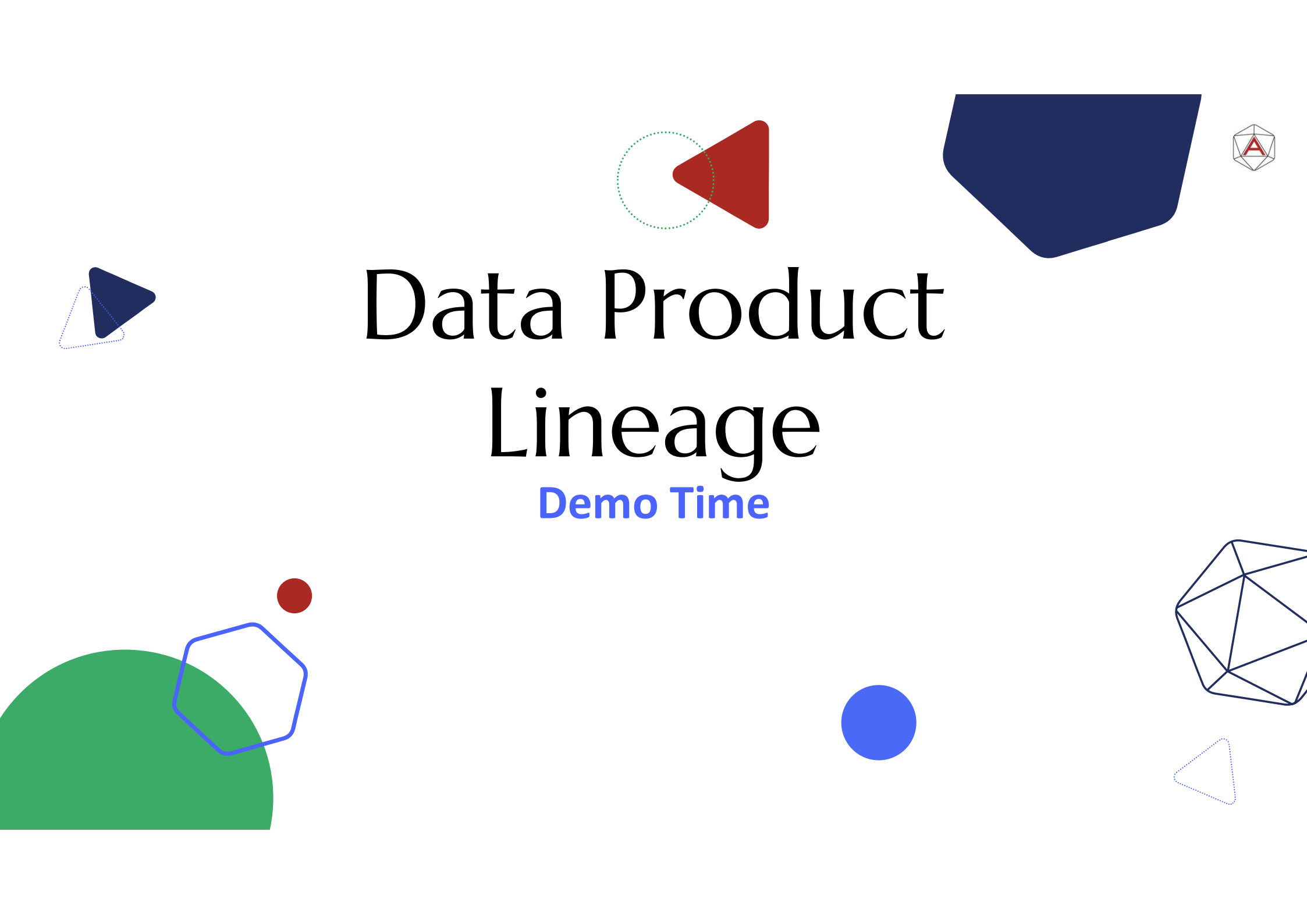
Where has the data come from for the Data Product?

What transformations has the data gone through?

How does a change in a data source or process affect downstream data products?

Is there a clear and complete audit trail of the data's lifecycle?

How has the data changed over time, and what are the different versions?


Data Product
Lineage
Demo Time

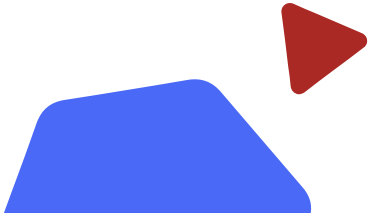
Logs Questions

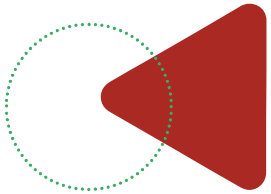
Who is using the Data Product?

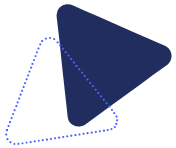
How are they using the Data Product?

How are access controls and permissions being enforced and monitored?

How are access controls and permissions being enforced and monitored?

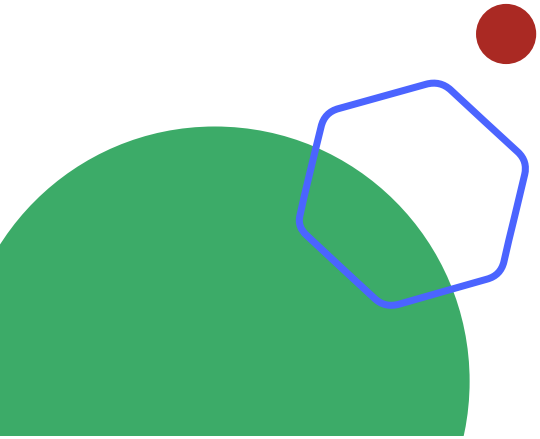






Data Product Logs

Demo Time



Data Product Logs Requirements

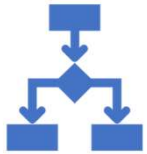
Entra App API Permissions

Resource	Permission	Type
Microsoft Graph	Application.Read.All	Application
Microsoft Graph	User.Read.All	Application
Microsoft Graph	Group.Read.All	Application
Power BI Service	Tenant.Read.All	Delegated

Power BI Tenant Settings – applied for Group containing Entra App

Tenant Setting	Value
Allow service principals to use Power BI APIs	Enabled for Specific security groups
Allow service principals to use read-only Admin APIs	Enabled for Specific security groups
Enhance admin APIs responses with detailed metadata	Enabled for Specific security groups
Enhance admin APIs responses with DAX and mashup expressions	Enabled for Specific security groups
Semantic Model Execute Queries REST API	Enabled for Specific security groups

Product Observability Impact



Decision Making

Compliance & Security

Innovation

Cost Reduction

Market Responsiveness

Conclusion

Review Systems

Review systems and processes, documenting what you're trying to achieve with each one and how they achieve it.

Critique Them

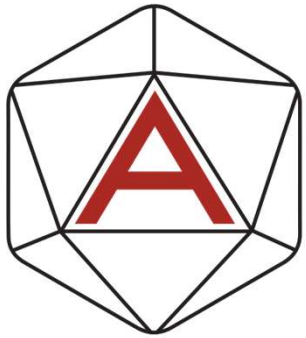
Critique each system and process you're employing. Do they answer the questions we've been asking? Are there any questions missing that need to be answered?

Observe Early

Observe early and observe often. It's much easier to build observability in from the start and continue observing and adjusting based on observed behaviour.

Observe, Orient, Decide, Act

Thanks & Code



ADVANCING ANALYTICS