

Brij Kishore Pandey
@brijpandeyji

TOP 7 DATA

TERMS

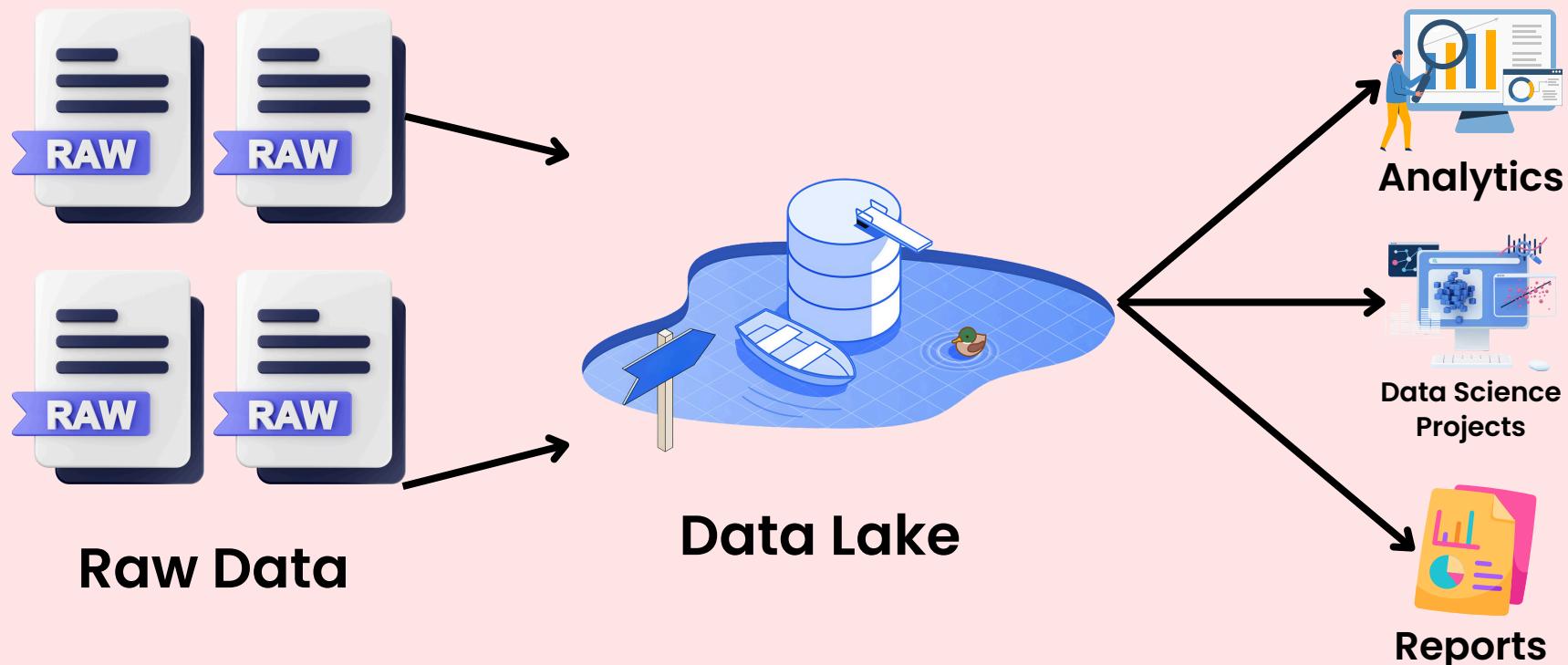
Explained



DATA LAKE

Definition:

A storage repository that holds a vast amount of raw data in its native format until it's needed.



Key Features:

Aggregates Varied Data: Combines structured, semi-structured, and unstructured data.

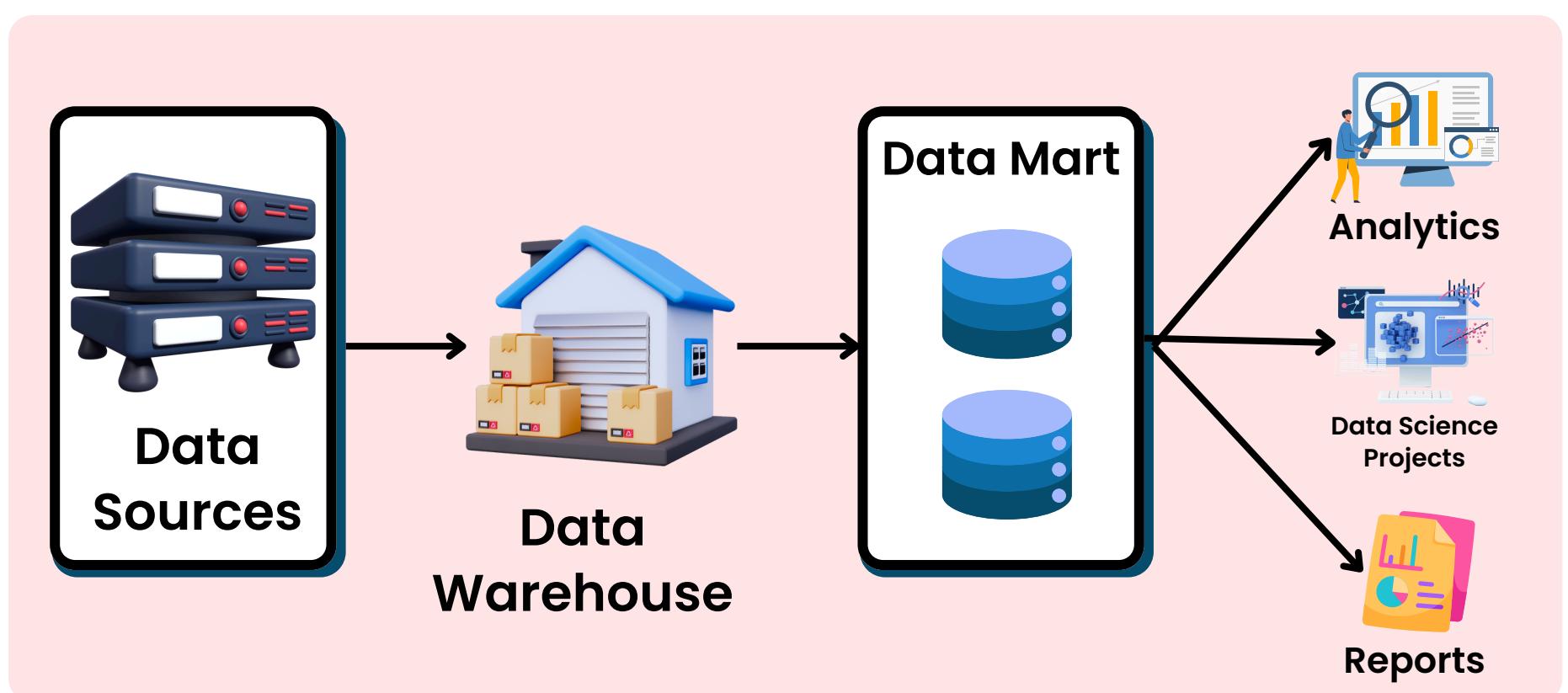
Seamless Storage: Stores data without the need for transformation.

Use Cases: Ideal for analytics, machine learning, and reporting.

DATA MART

Definition:

A subset of a data warehouse tailored for specific business departments or use cases.



Key Features:

Focused Insights:
Provides department-specific views like sales or finance.

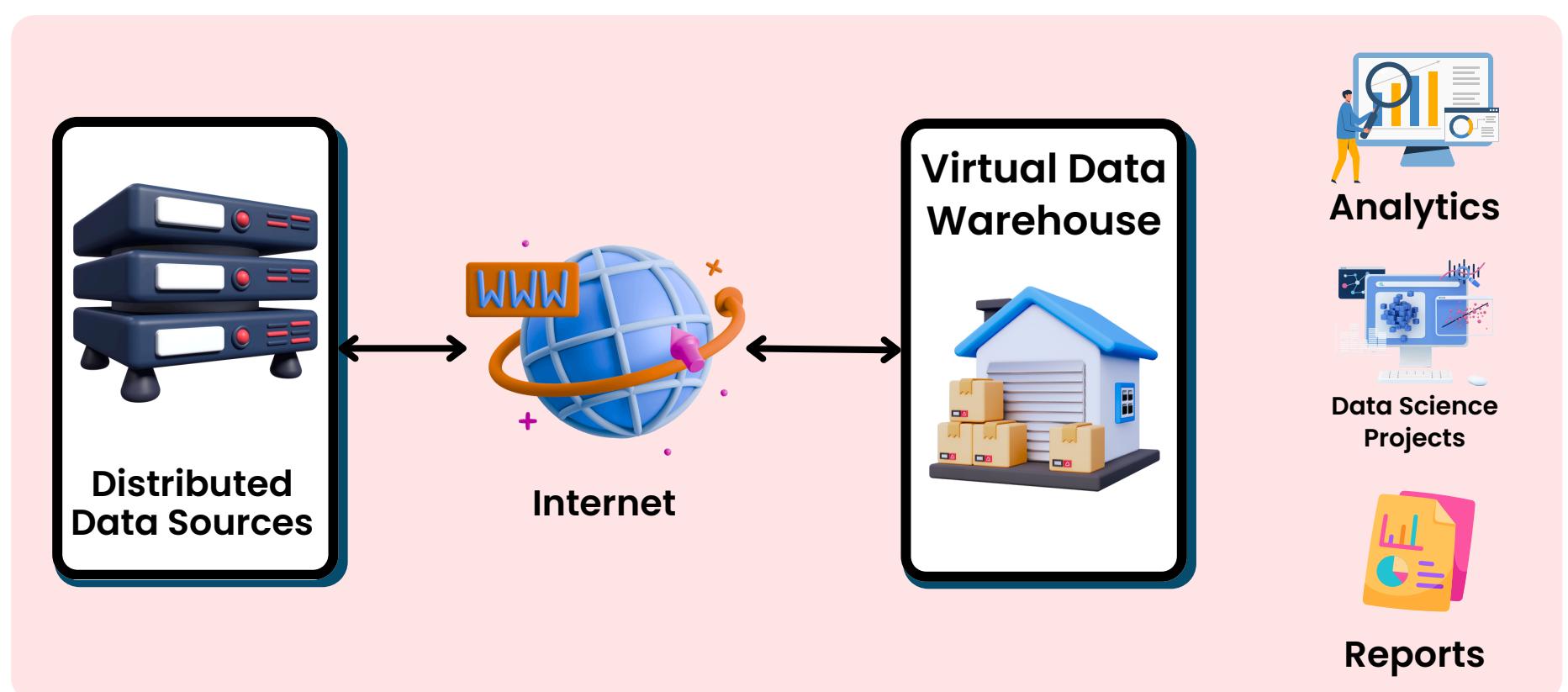
Optimized for Speed:
Data is pre-processed for faster queries and analytics.

Use Cases: Enhances efficiency for business teams by providing relevant insights.

DATA MESH

Definition:

A decentralized data architecture focusing on domain-oriented ownership.



Key Features:

Distributed Ownership:
Enables teams to manage their own data.

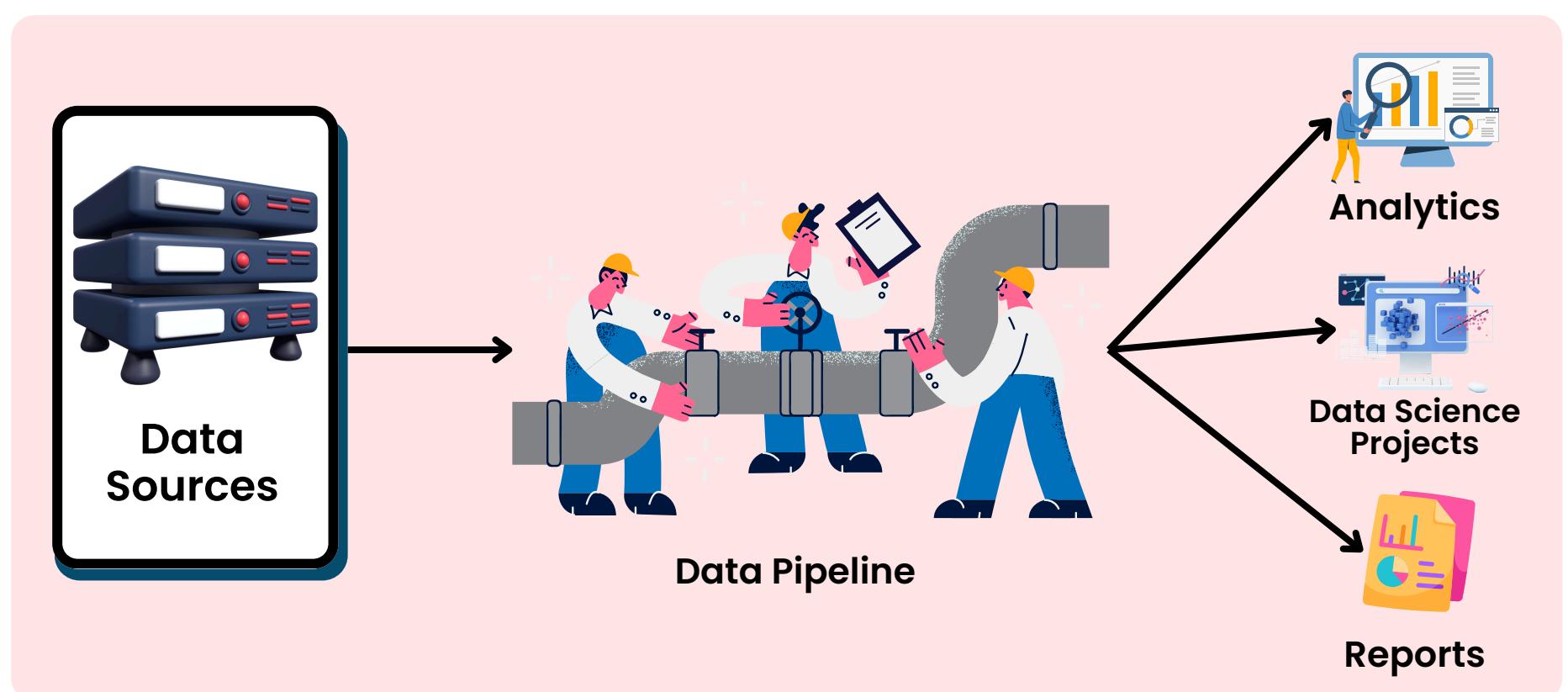
Domain-Specific Solutions: Provides tools tailored for specific needs.

Use Cases: Promotes scalability and self-service in large organizations.

DATA PIPELINE

Definition:

A set of automated processes that move, transform, or process data between systems.



Key Features:

Efficiency: Automates data workflows for seamless transfer.

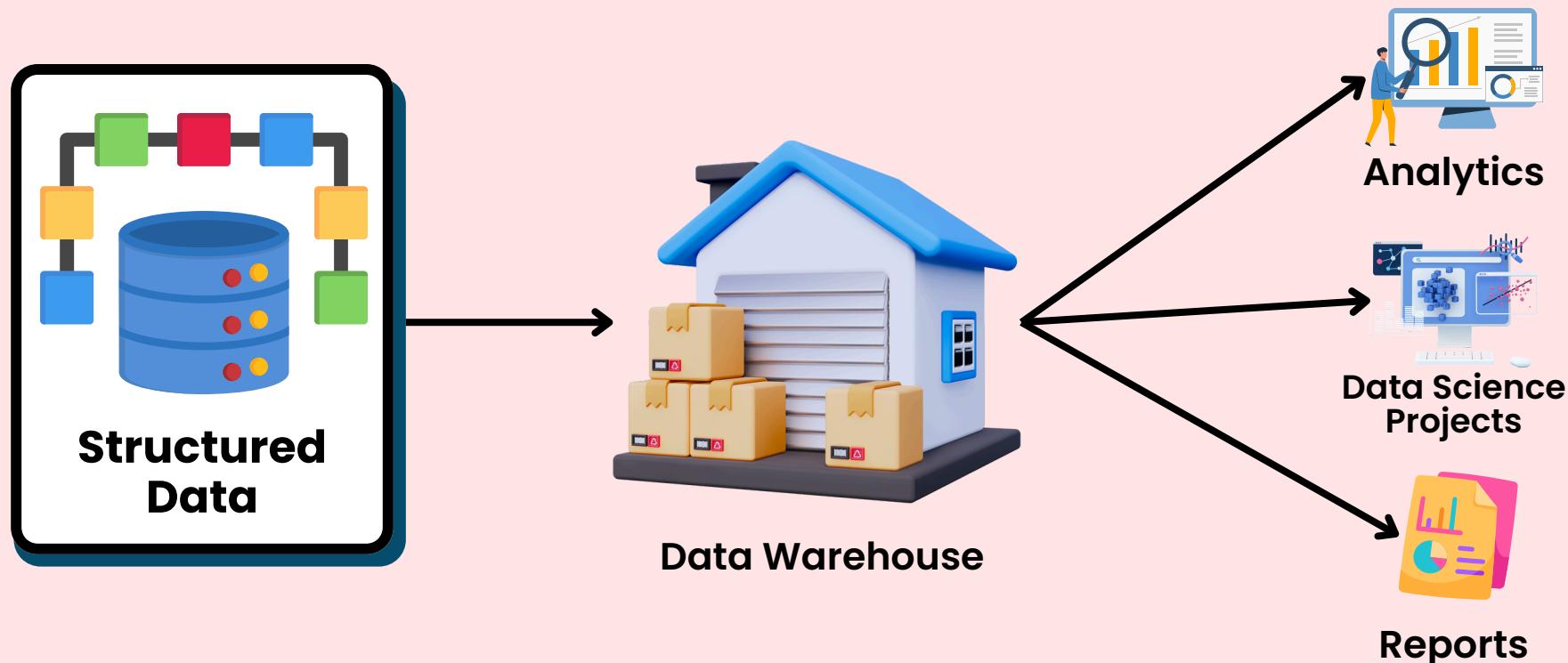
Scalability: Handles large-scale data movements effectively.

Use Cases: ETL processes, real-time analytics, and data integration.

DATA WAREHOUSE

Definition:

A centralized repository for storing structured data for analytics and reporting.



Key Features:

Centralized Storage:
Houses curated and organized datasets.

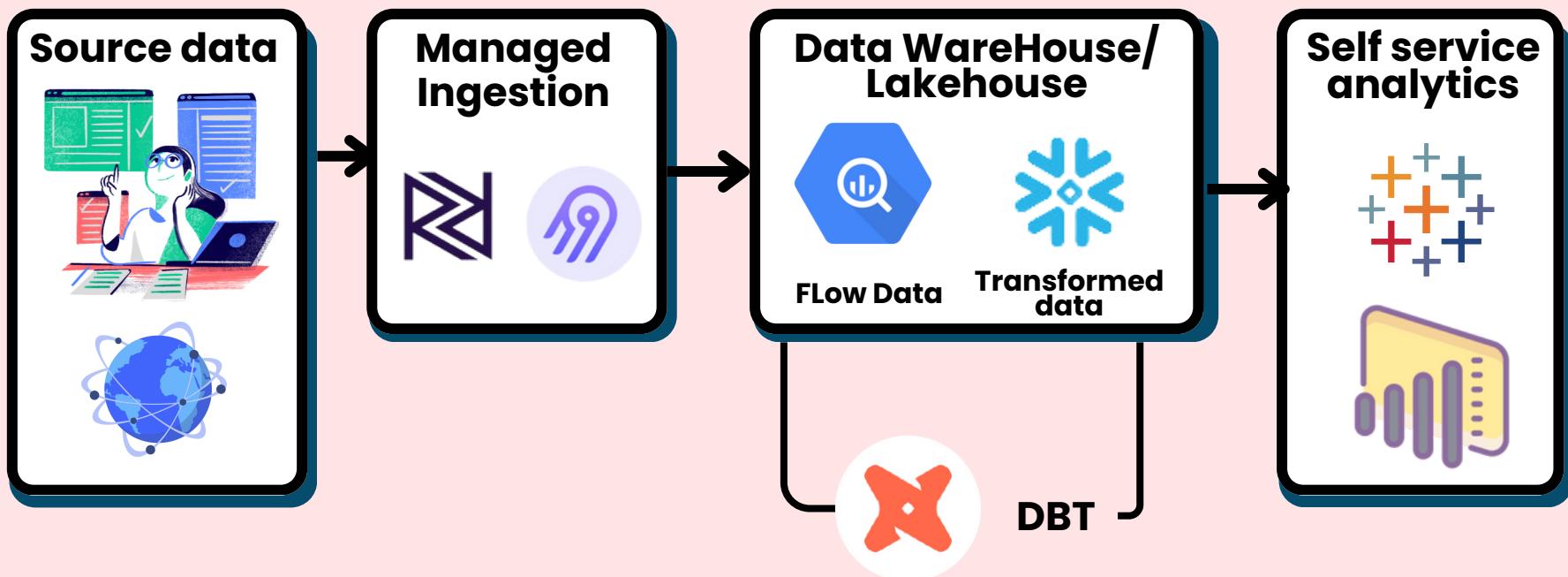
Optimized Queries:
Designed for fast querying and BI tools.

Use Cases: Business intelligence, reporting, and historical data analysis.

DATA OBSERVABILITY

Definition:

The ability to monitor and understand the health and performance of data systems.



Key Features:

Ensures Accuracy:
Tracks and resolves data quality issues.

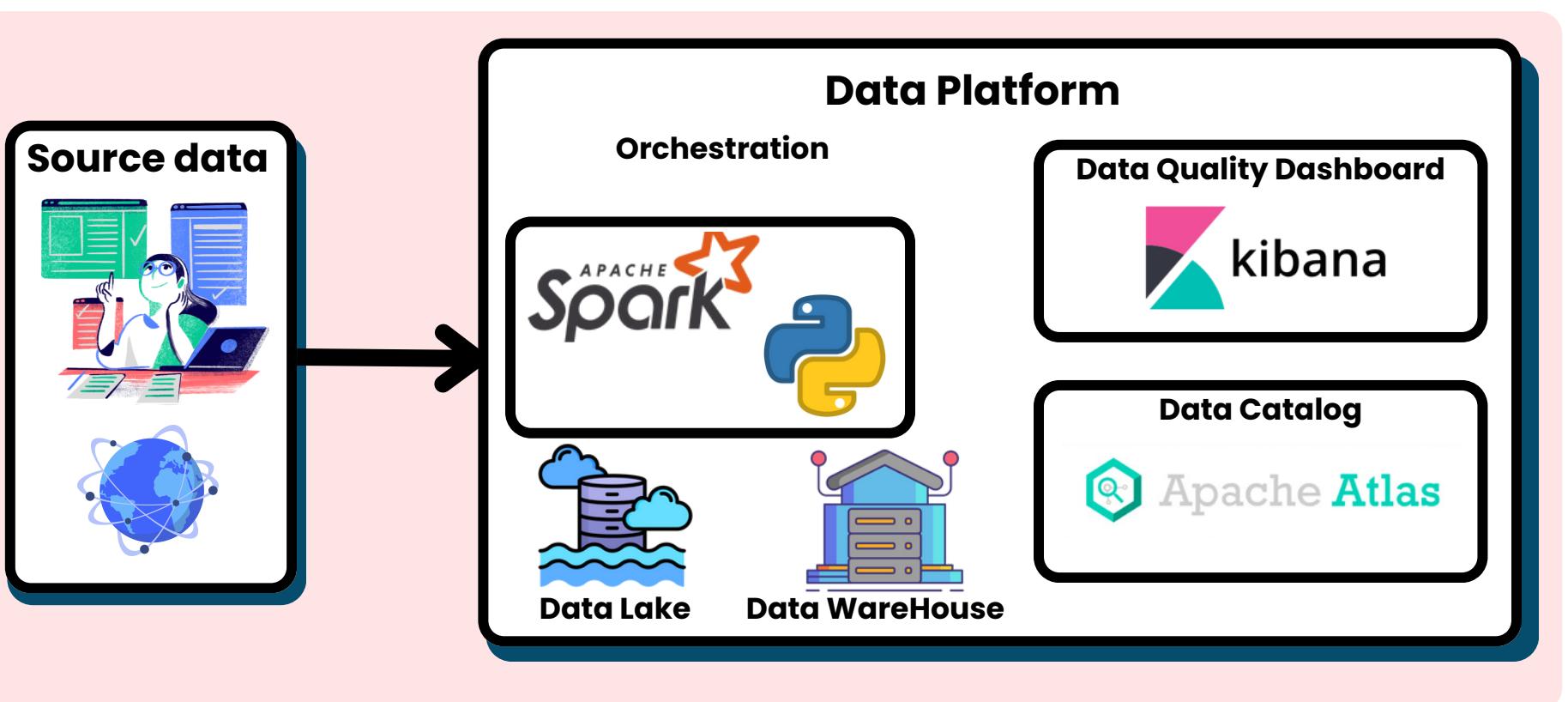
Proactive Monitoring:
Prevents data pipeline failures.

Use Cases: Data reliability for AI models, compliance, and reporting.

DATA QUALITY

Definition:

The process of ensuring that data is accurate, consistent, and reliable.



Key Features:

Real-Time Monitoring:
Tracks anomalies and fixes issues promptly.

Transparent Processes:
Builds trust in analytics and reporting.

Use Cases: Essential for compliance, predictive models, and operational decisions.



Brij Kishore Pandey

@brijpandeyji

Follow for More

