

Azure Databricks: Unlocking Unstructured Data for Analytics

Wilson Mok

Feb 16, 2026





Wilson Mok

Sr. Data architect & Consultant

 /wilson-mok

 /wilson-mok

 @the-analytics-lab



My experiences includes:

- Avanade
- CAE
- Air Canada



Expected Audience and Presentation Structure



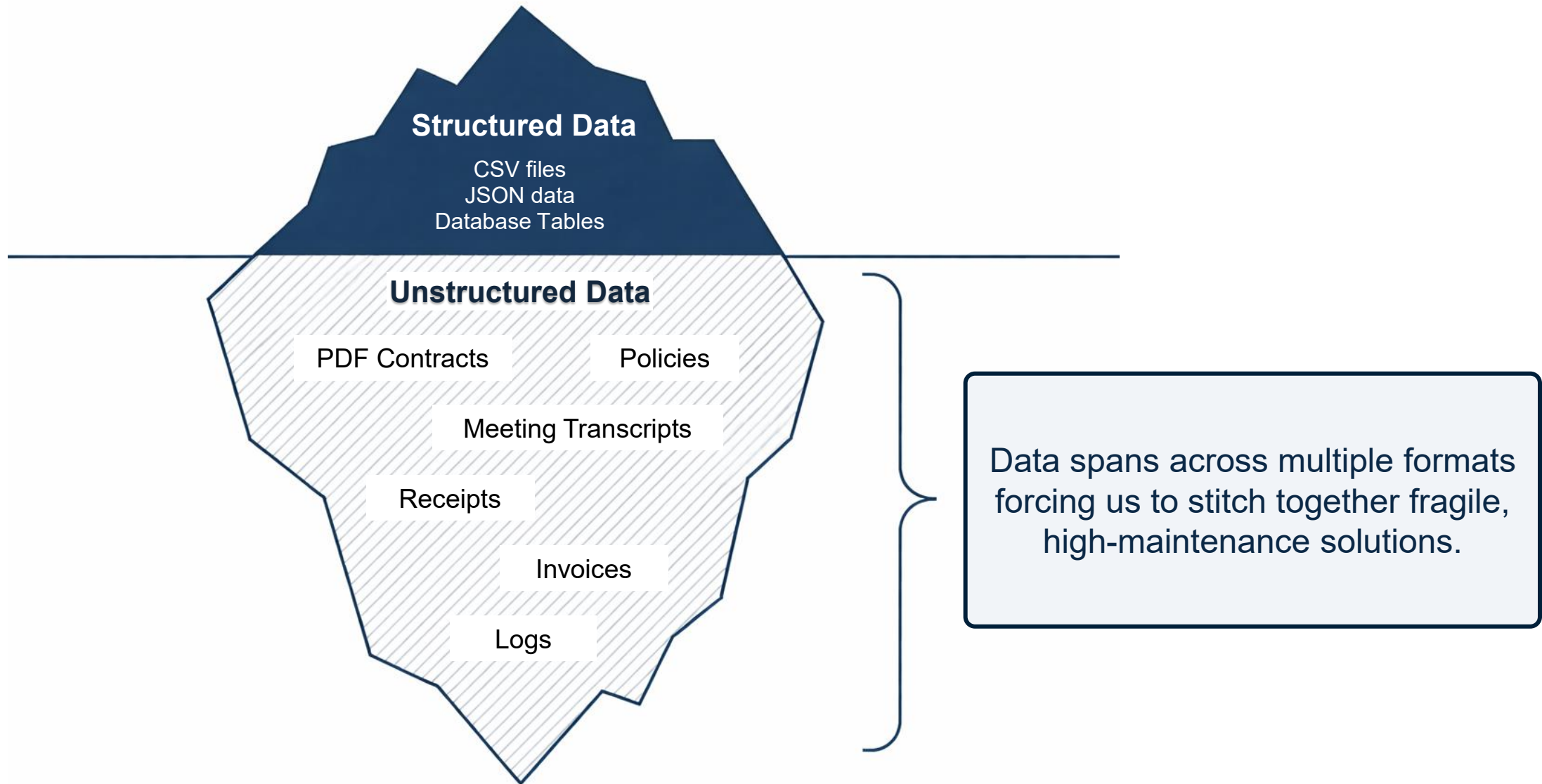
What We'll Cover:

- The challenges of working with unstructured data.
- How GenAI enables adaptive, schema-driven extraction?
- How Databricks AI-driven document parser and Agent Brick Information Extraction addresses those challenges?
- How does unstructured data fits into Medallion Architecture?

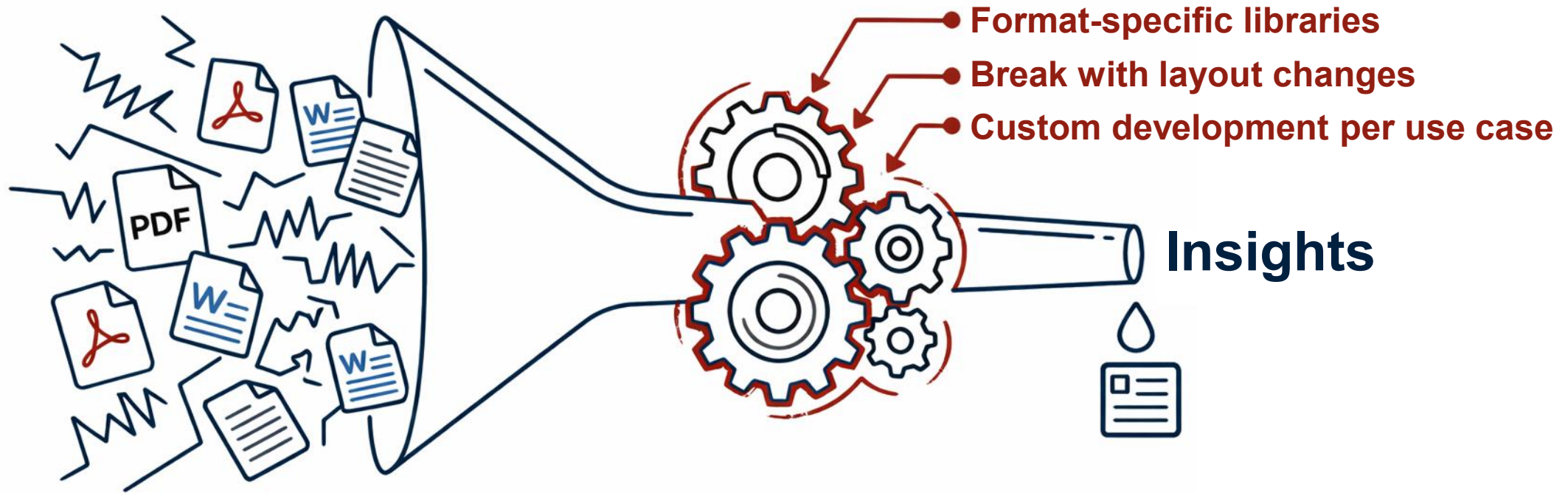


Fundamental knowledge: Basic understanding of GenAI and Data processing

Most enterprise value sits in unstructured data, but our tooling wasn't built for it



Traditional analytics wasn't designed for unstructured data



GenAI removes the schema bottleneck. Eliminating template-bound parsing and per-format pipelines.

Databricks Agent Bricks turns GenAI capability into enterprise workflow



Information Extraction

PREVIEW



Converts unstructured documents into structured, analytics-ready data without custom model training.

Knowledge Assistant



Delivers source-cited answers grounded in your internal data.

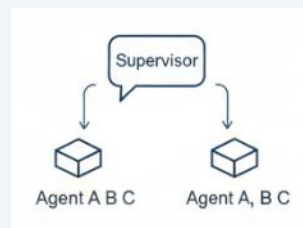
Custom LLM

PREVIEW



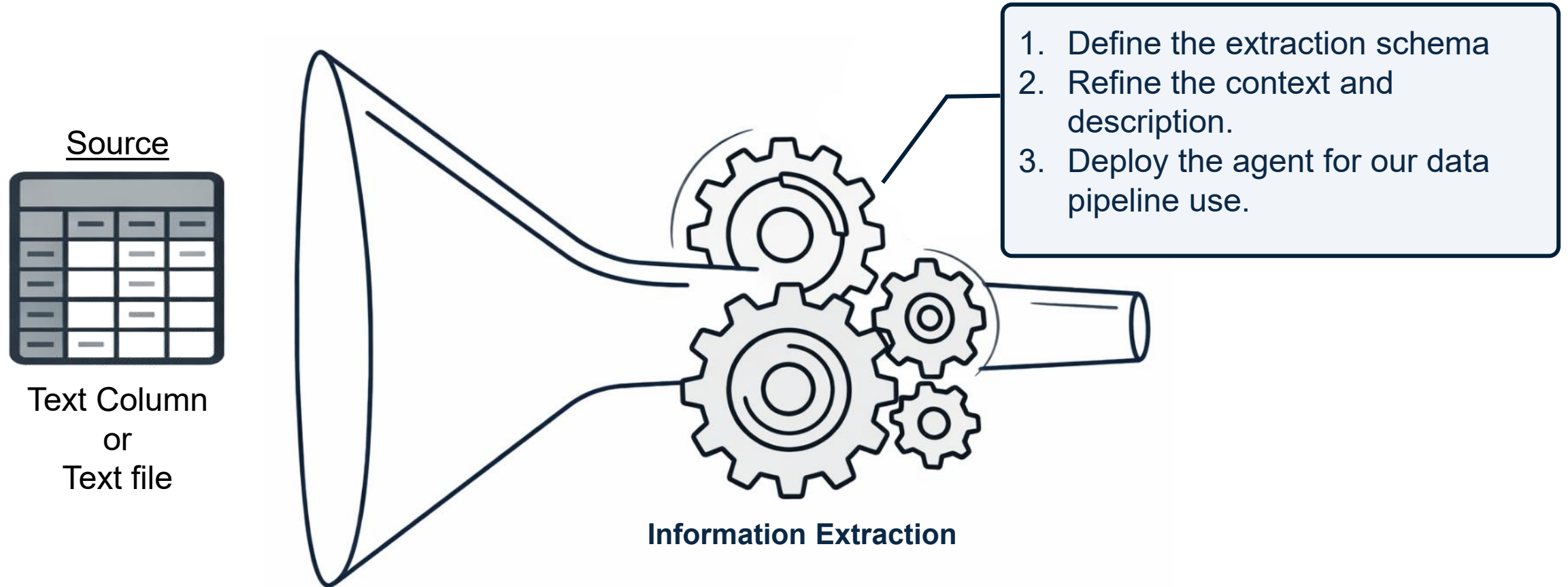
Builds domain-specific generation workflows from summarizations to structured content creation.

Multi-Agent Supervisor

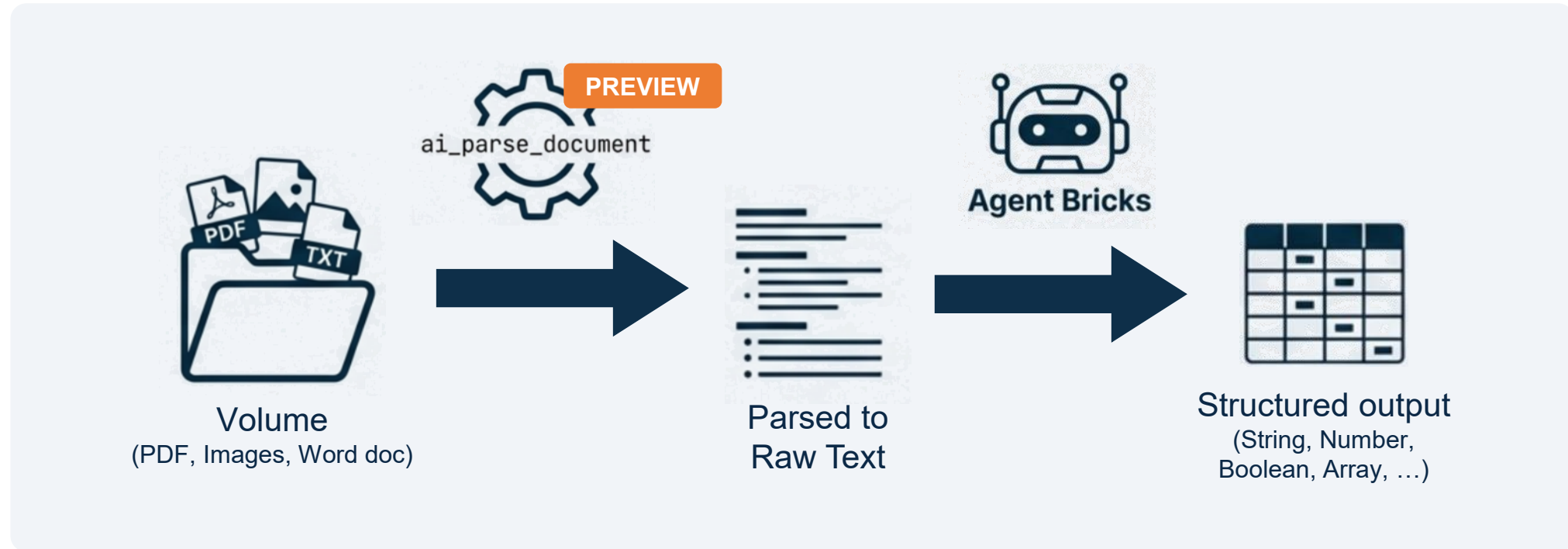


Orchestrates multiple agents to execute complex, multi-step workflows.

Information Extraction turns brittle parsing into adaptive extraction



From raw documents to extraction-ready text

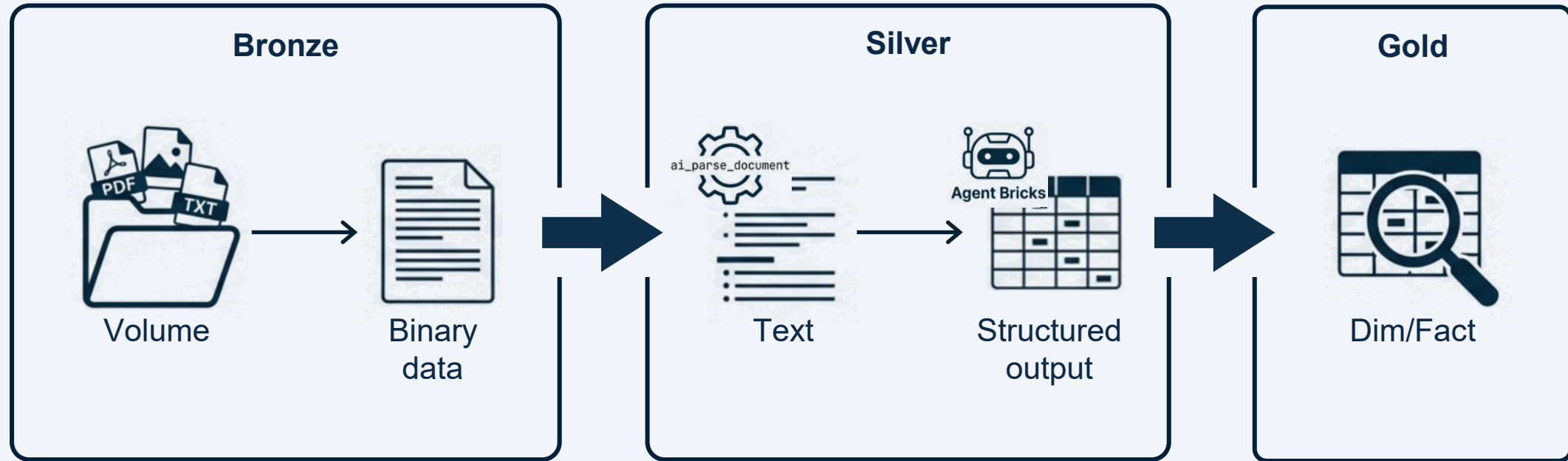


Document parsing and extraction run natively within the Databricks platform.
Eliminating external OCR pipelines and format-specific code.

This is how unstructured data becomes governed, analytics-ready enterprise data

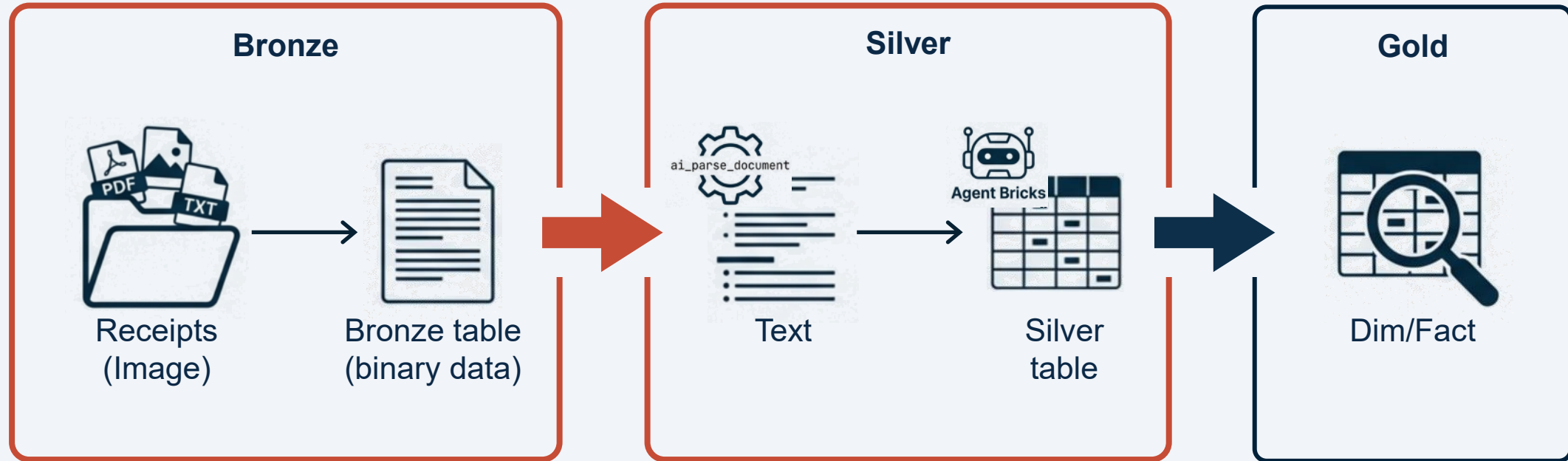


Medallion Architecture



Unstructured data follows the same governed lifecycle as structured data

Medallion Architecture



Key Takeaways: Unlocking Unstructured Data for Analytics

- GenAI enables adaptive, schema-driven extraction. Eliminating brittle, format-specific parsing pipelines.
- Agent Bricks operationalizes extraction as declarative, reusable AI agents built and deployed natively within the Databricks Intelligence platform.
- Document parsing and extraction run end-to-end inside the platform. No external tools required.
- Unstructured data follows the same governed lifecycle (Medallion architecture) as structured data.
- The code is in my Github repo:
 - <https://github.com/wilson-mok/demo/blob/main/2026/MSDEVMTL/1-Databricks-Information-Extraction/>

Thank you



LinkedIn