

Streamline Your Data. Strengthen Every Operation.

AI Data Curation & Deployment

The Challenge

A leading manufacturing company with over 45 years of operational data, set out to build an AI-powered chatbot to assist their tech support team. The goal was to provide fast, accurate answers by querying historical troubleshooting records, manuals, and pricing information. But their unstructured data - scattered across locations, formats, and file types - made that impossible.

Despite an early chatbot prototype, results were inconsistent and unreliable. The system couldn't locate or understand the right information. With 9+ million files (over 9TB of data) across plants and systems - and limited traceability across different product types - the project stalled.

The Breakthrough

This company deployed the Aparavi Data Suite (ADS) alongside the Data ToolChain (DTC) to gain control over their data and streamline the AI pipeline.

In just 5 days, they:

- ✓ Located and curated the right historical data from 9 million+ files
- ✓ Prepped that data for ingestion into their chatbot
- ✓ Built a fully functional, responsive, and accurate AI assistant
- ✓ Enabled fast iteration using a flexible pipeline with support for any model or vector store

The result was a clean, connected data foundation that saved weeks of development time, reduced manual effort, and dramatically improved chatbot performance and trust.

Results at a Glance



9M+ files scanned and sorted



AI chatbot fully operational in 5 days



Unified workflow enabling fast iteration

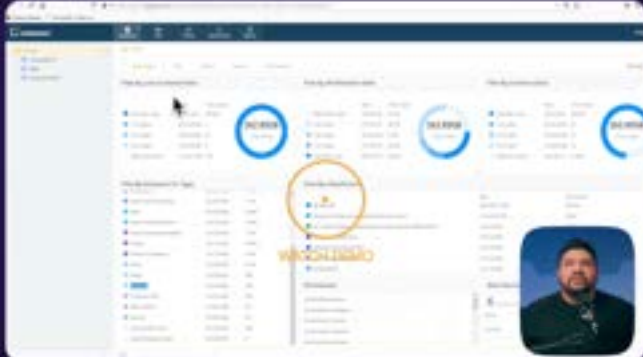


Hours of development time saved



Accurate, reliable, and traceable AI responses

[Breaking it Down]



[Data Discovery] Aparavi Data Suite

- Connect – Secure access to all your data, no need to move it.
- Scan – Automatically index files, metadata, and content.
- Normalize – Clean and standardize data formats.
- Classify – Tag by sensitivity, relevance, and content.
- Visualize – See everything clearly in one dashboard.
- Act – Quickly move, delete, or prep data for AI.
- Maintain – Keep data healthy with ongoing monitoring.

[Deployment] Data ToolChain

Once your data is clean, classified, and visible—DTC takes it from there:

- Ingest – Drag-and-drop ready files directly into your AI workflows.
- Chunk – Break data into optimized, context-aware sections.
- Embed – Generate rich vector embeddings for search or LLM use.
- Map – Align data to models or business logic with full traceability.
- Deploy – Push clean, structured data into agents, copilots, or RAG pipelines.
- Repeat – Fine-tune with feedback loops and reprocessing as needed.



