



## **GenAI Data Assistant — HR Analytics**

Industry: Enterprise Data & Operations

Customer Profile: Mid-market manufacturer with distributed ERP and CRM data sources.

Status: Production deployment since Q1 2025

### **Business Challenge**

Managers lacked technical skills to query Athena or QuickSight directly, causing delays and constant ad-hoc report requests.

### **AWS-Powered Solution**

- Amazon Bedrock (Claude 3 Instant) for natural-language to SQL translation.
- Amazon Athena + Glue Catalog for query execution and schema management.
- OpenSearch Serverless + Titan Embeddings for semantic context and metadata grounding.
- QuickSight API renders visual summaries automatically.
- Verified Permissions enforces row-level security per user role.
- SageMaker Pipeline ingests user feedback on query accuracy for continuous learning.

### **Outcomes**

#### **Metric Before After Δ**

Report creation time 2 days 15 min ↓ 88 %

Analyst workload 100 % 35 % ↓ 65 %

Query accuracy — 0.94 F1 —

Cost per report \$5.20 \$0.70 ↓ 87 %

### **Business Impact**

Line managers now self-serve data insights securely. Analysts reallocated to strategic projects; company

achieved 4.5× ROI within six months.

### **5 · Cross-Case Observations**

#### **Theme Finding Mitigation**

Bias Drift Observed in text and voice models over time. Weekly Clarify scans + auto-retraining pipelines.

Cost Spikes Token usage variance in Bedrock workloads.

FinOps Inspector Lambda detects and tunes model routing.

Safety Triggers Early voice models escalated too often. Retrained intent classifier +



contextual safety

weights.

Human

Oversight All models retain manual review path. Guardrail failures auto-route to review queue.

