

Data Engineer – Challenge

THE CHALLENGE

Legacy system & Fragmented Data: Legacy platforms create data silos, increasing inefficiency and maintenance costs.

Operational Risk & Security Vulnerabilities: Outdated systems raise security exposure and compliance risk.

Limited Scalability & Innovation: Limited scalability prevents real-time insights and adoption of AI-driven innovation.

Sluggish Time-to-Market: Slow delivery reduces competitiveness against more agile competitors.

THE SOLUTION

Modern Cloud Platform: Move to a scalable, secure, and integrated Azure platform to boost performance and flexibility.

Centralized Data Lakehouse: Create a single source of truth to unify data and enable advanced analytics at scale.

Automated Workflows & AI Integration: Automate workflows and embed AI/ML to optimize operations and decision-making.

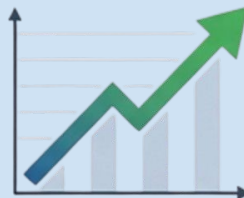
Agile Operating Model: Adopt DevOps practices to accelerate delivery, continuous improvement, and innovation.

BUSINESS IMPACT

GROWTH METRICS



Transforming data into a strategic asset to drive sustainable profitability and market leadership.



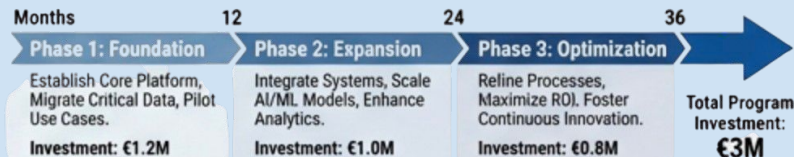
Cost Reduction
-20%
Operational Efficiency & IT Spend Optimization

Customer Retention
+10%
Enhanced Personalization & Service Speed

Risk Mitigation
90%
Reduction in Security Incidents & Compliance Gaps

Transforming data into a strategic asset to drive sustainable profitability and market leadership.

INVESTMENT & TIMELINE

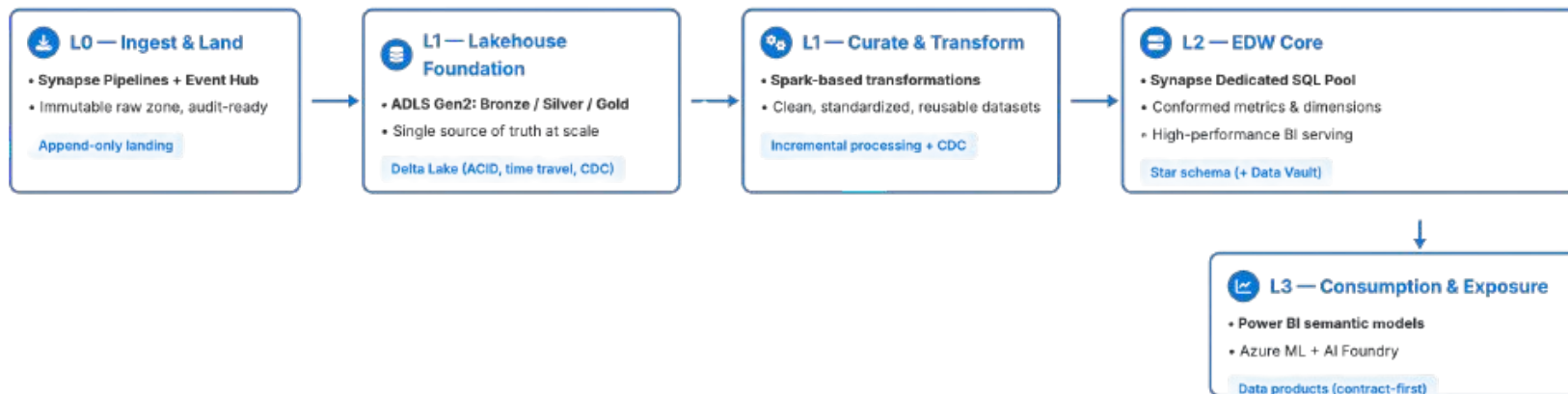
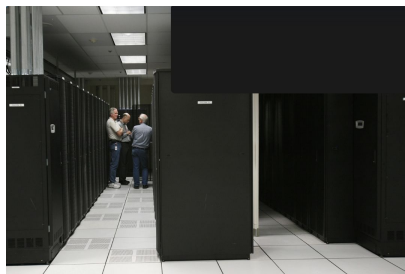


DECISION REQUIRED

€3M Investment to Secure Our Future. Approve Program Initiation.

Target EDW Architecture (Layers & Patterns)

Objective: Build a scalable, governed EDW enabling BI + AI + automation



Cross-cutting: Govern & Operate

IDMC governance (catalog, lineage, glossary) + DataOps automation (cost, reliability, security)

What each tool delivers in the Enterprise Data Warehouse



Synapse Pipelines

- Orchestration, scheduling, retries
- Parameterized domain ingestion
- Release-ready with CI/CD integration



Spark Pools

- Transformation at scale
- Incremental processing + CDC merges
- Reusable rule-based frameworks



ADLS Gen2 + Delta

- Central, low-cost storage
- Reliable tables with versioning
- Enables replay, audit and scalability



Synapse Dedicated SQL Pool

- Enterprise BI performance layer
- Governed KPI tables + aggregates
- Security and workload isolation



IDMC

- Catalog, lineage, glossary
- Stewardship + compliance control



Power BI + Azure ML

- Certified BI datasets
- AI use cases: fraud, churn, pricing optimization

Reusable patterns, standardized artifacts, and automation to reduce operational costs



Design patterns (by layer)

- **L0:** Append-only + ingestion metadata
- **L1:** Delta Lake + CDC merges + time travel
- **L2:** Optional Data Vault for audit/history
- **L2/L3:** Star schemas + aggregates for BI
- **L3:** Data products + ownership model



Standard artifacts (reusability)

- Pipeline templates (batch / CDC / streaming)
- Canonical EDW models (dim/fact/KPI)
- Domain naming + folder conventions
- Shared transformation framework (Spark)



DataOps modules (reduce run cost)

- **CI/CD:** GitHub Actions + DEV/UAT/PROD promotion
- **Automated testing:** schema + DQ + reconciliation
- **Observability:** SLA dashboards + alerting
- **Cost guardrails:** autoscale + budgets + workload mgmt
- **Governance automation:** catalog + lineage publishing