

<epam>

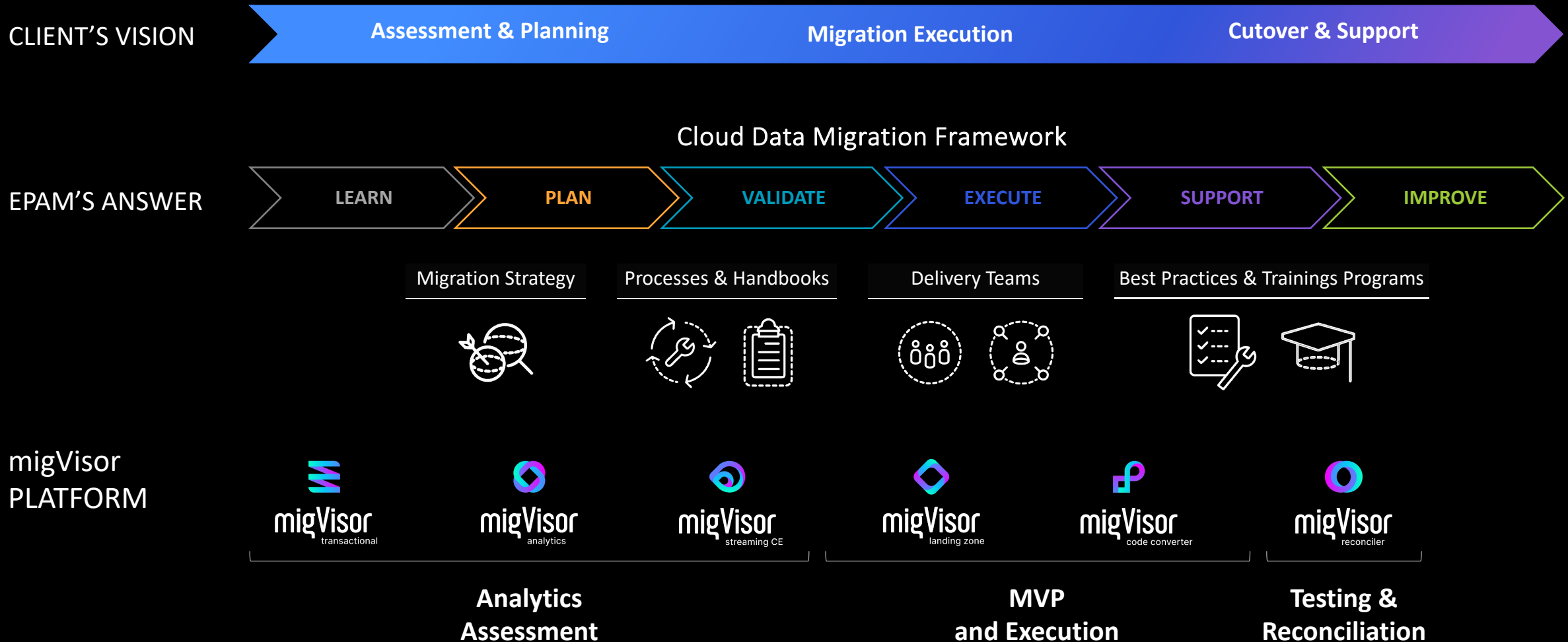
EPAM migVisor Platform

Cloud Migration Service Offering

2023



Cloud Data Migration Journey with EPAM



Accelerator Platform for Data & Analytics Migrations of Any Complexity



Key Capabilities

- Speeds up process of OLTP and general DB assessment
- Automatically scans metadata in Oracle, MSSQL, PostgreSQL, MySQL, Hbase, and MongoDB
- Analyzes complexity of db engine change and potential cloud migrations
- Applies AI and past-experience into define target sizing and accurate migration timeline
- Provides a detailed TCO report and detailed migration path
- Speeds up process of DWH ETL and Reports assessment
- Automatically scans metadata in DWHs, Reports and ETL pipelines
- Analyzes complexity of identified inventory and clusters objects
- Applies AI algorithms to identify dependencies and bring down the scope
- Provides an extendable analytics engine
- Accelerating & Streamlining a complex Migration Planning process through intuitive Admin UI
- Eliminating the barriers to buy by reducing the uncertainties surrounding the target streaming platform migration complexity and future runtime cost with a click of a button
- Approximately reducing streaming migration time and cost by more than 50% -based on previous manual migration experience-
- Scripts infrastructure deployment containing basic set of data platform services
- Automated infrastructure deployment to CSP's
- Data product framework
- CI/CD framework, security models, dynamic resource allocation and RBAC with service principals
- Demo application with synthetic data, including data lake, data transformation pipelines, data mart, semantic models and dashboards
- Leveraged within EPAM's conversion acceleration methodology
- Configurable automation tool that works with most used ETL/ELT platforms including:
 - Informatica
 - DataStage
 - Talend
 - SQL
 - Scripting Languages
- Ability to update conversion configurations to handle exception cases and iterate through ~80% automated converted code
- Schema comparison (tables, columns, partitions, DB objects)
- Statistics comparison (row count, checksum)
- Data comparison (value by column)
- Automated database scanning
- AI-driven approach for mapping tables, columns and data types
- High-performance scalable data comparison
- Several layers of reconciliation (quick, detailed, deep analysis)

Migration Technology Capabilities

	On-prem	Azure	AWS Amazon	GCP
DATA WAREHOUSE	RDBMS / STORAGE 	CLOUD DW 	CLOUD DW 	Cloud DW
UPSTREAM	ETL 	CLOUD DATA PIPELINES 	CLOUD DATA PIPELINES 	CLOUD DATA PIPELINES
DOWNSTREAM	REPORTING 	CLOUD BI 	CLOUD BI 	CLOUD BI
DATABASES	LICENSED AND NON-LICENSED DB 	CLOUD DB 	CLOUD DB 	CLOUD DB
	SOURCE	TARGET		

MIGRATION EXPERIENCE

Snowflake

150+ projects migrated from on-prem RDBMS and Hadoop to Snowflake

Databricks

100+ projects migrated from Informatica, Pentaho, SSIS, Talend and Cloudera Spark to Databricks

Data Migration CoE

- Support from Center of Excellence includes:
- Migration Framework, best practices in migration and reconsolidation
 - Migration tools and accelerators
 - Assessment for DWs and Data Lakes

WHY

- License costs reduction reusability
- Leverage fully-managed solutions
- Increase database automation levels
- Reduce commercial databases footprint
- Provide a more robust and scaled solution
- Have database-level agility

WHAT

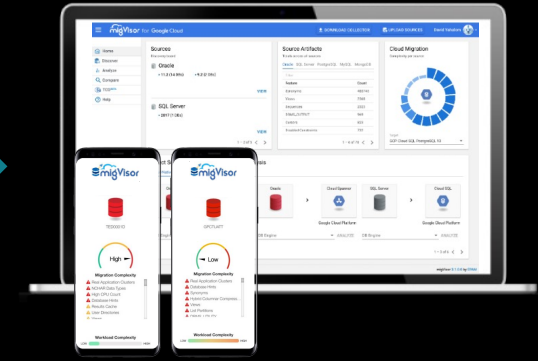
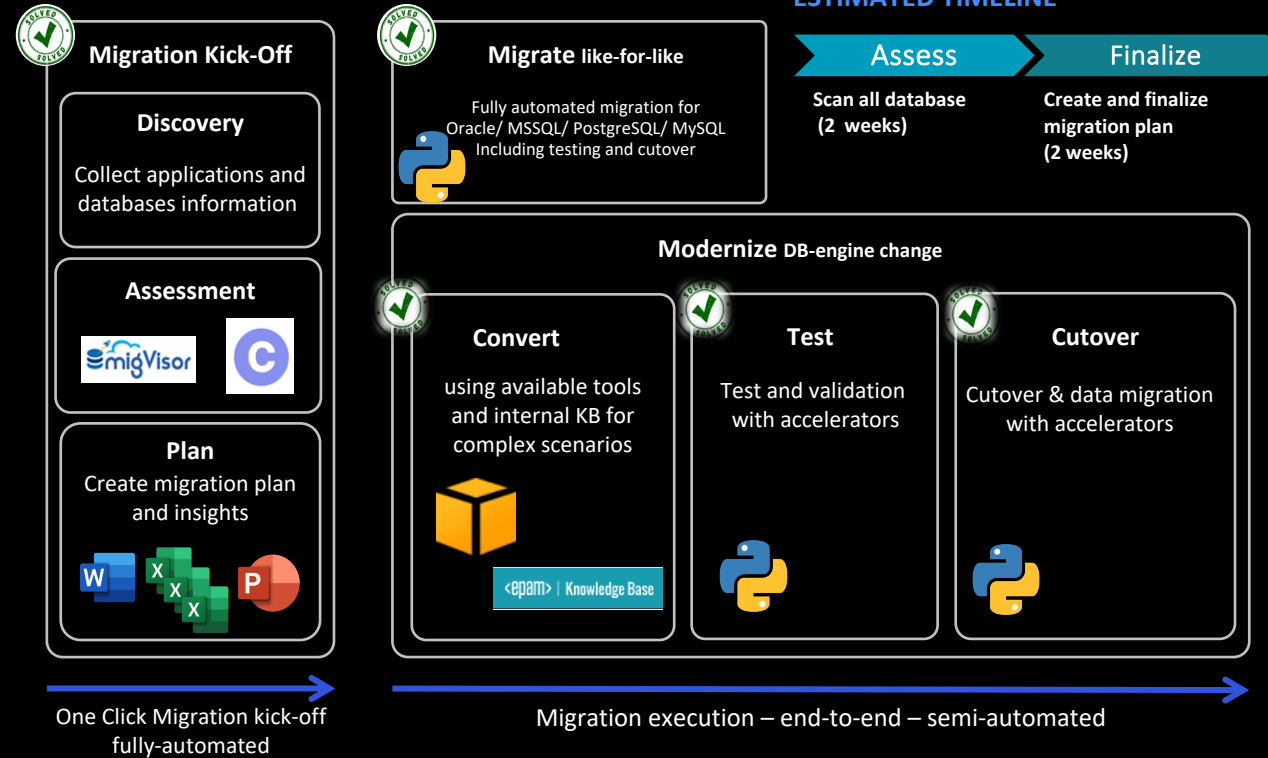
KEY CAPABILITIES

- Discover and assess your database fleet and app source code using migVisor by EPAM
- Analyze finding to create dependencies mapping and initial estimates
- Create a draft migration plan including TCO
- Confirm the plan and adjust with relevant stakeholders and non-functional requirements and constraints

Outcome:

EXECUTION-READY MIGRATION PLAN AND PROJECT JUSTIFY WITH ROI

HOW



THE migVisor WAY – IMMEDIATE ROI

Automated assessment with TCO

Better decisions

Better migrations



KEY DIFFERENTIATORS

Highly-automated approach:
Quick and accurate migration plan for your databases

Widely credible migration experience:
Google selected product for database assessments

WHY

| Greenfield start challenge

| Time and resource-consuming development of infrastructure deployment

| Legacy processes slow down modern technologies adoption and automation

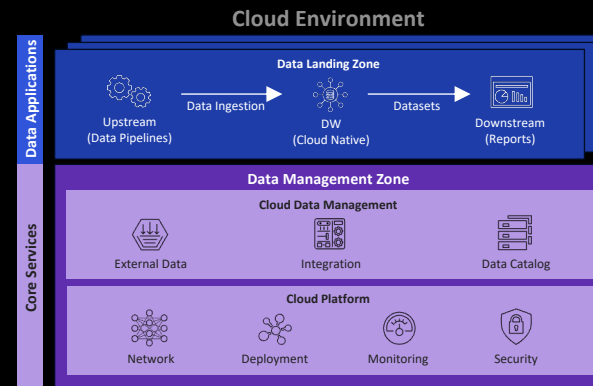
WHAT

KEY CAPABILITIES

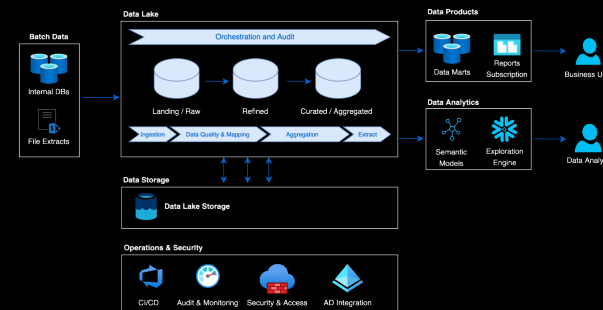
- Ready to use terraform scripts for infrastructure deployment, which contain simplified Data Platform according to EPAM Data Factory standards, including capabilities of data collection, cleansing, consolidation, transformation and aggregation
- Automated infrastructure deployment to Azure, AWS, GCP
- Best practices and an example for end-to-end data analytics solution using cloud Data Lake and Data product approaches
- Best practices and examples for CI/CD approach, security models, dynamic resource allocation and RBAC with service principals
- Implemented training case with synthetic data, including Data Lake, Data Transformation Pipelines, Data Mart, Semantic Models and Dashboards

HOW

HOW DOES IT WORK CONCEPTUALLY



Cloud Infrastructure Diagram



Data Flow Diagram

HOW TO USE IT ON A PROJECT

- 1 Deploy environment**
30 minutes for deployment
- 2 Configure environment**
1 day for environment configuration
- 3 Start Migration**
Cloud Environment is ready in 1st sprint
- 4 Scale the Platform**
Landing zone is designed for scaling

KEY DIFFERENTIATORS

| Environment set-up in 30 minutes and ready for a first POC execution

| Best practices in solution architecture, security, CI/CD, Data Quality

| 2 weeks for Onboarding and Technology adoption, based on provided demos

WHY

| High-complexity of legacy ETL data transformation pipelines in low-code tools

| Extremely time-consuming process of manual SQL and ETL conversion to pySpark

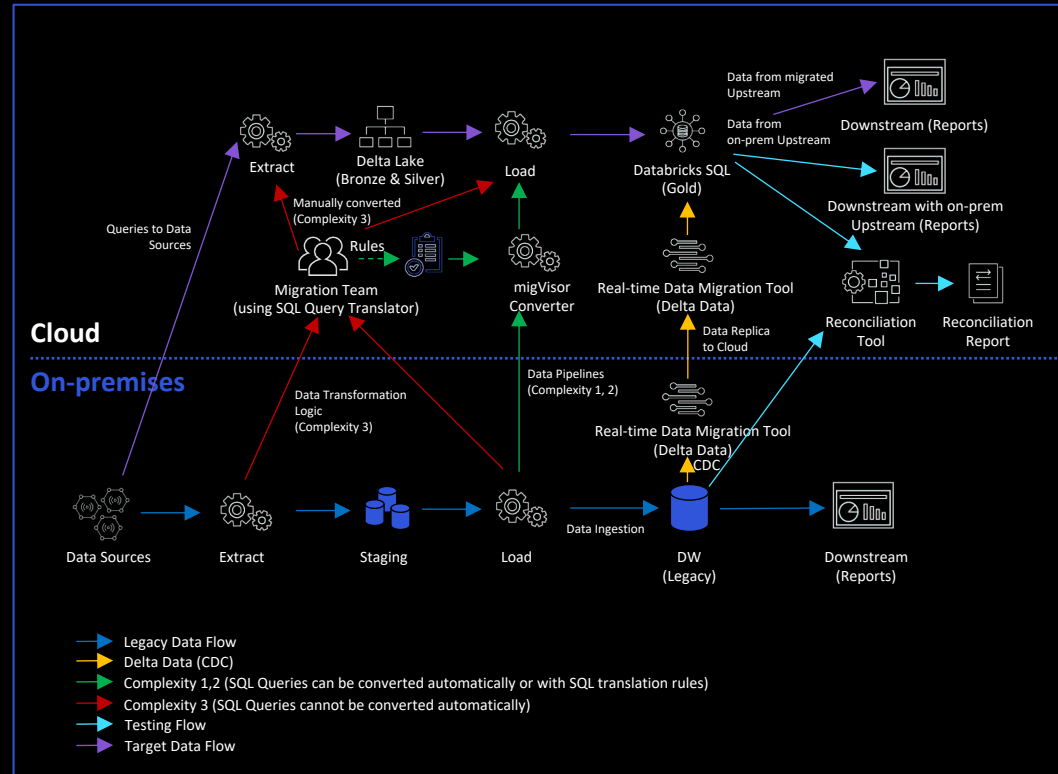
| High-complex process of data quality testing and reconciliation

WHAT

HOW

KEY CAPABILITIES

- Leveraged within EPAM's conversion methodology to accelerate ETL/ELT conversion
- Configurable automation tool that works with most used ETL/ELT platforms including:
 - Informatica
 - DataStage
 - Talend
 - SQL
 - Scripting Languages
- Ability to update conversion configurations to handle exception cases and iterate through ~80% automated converted code



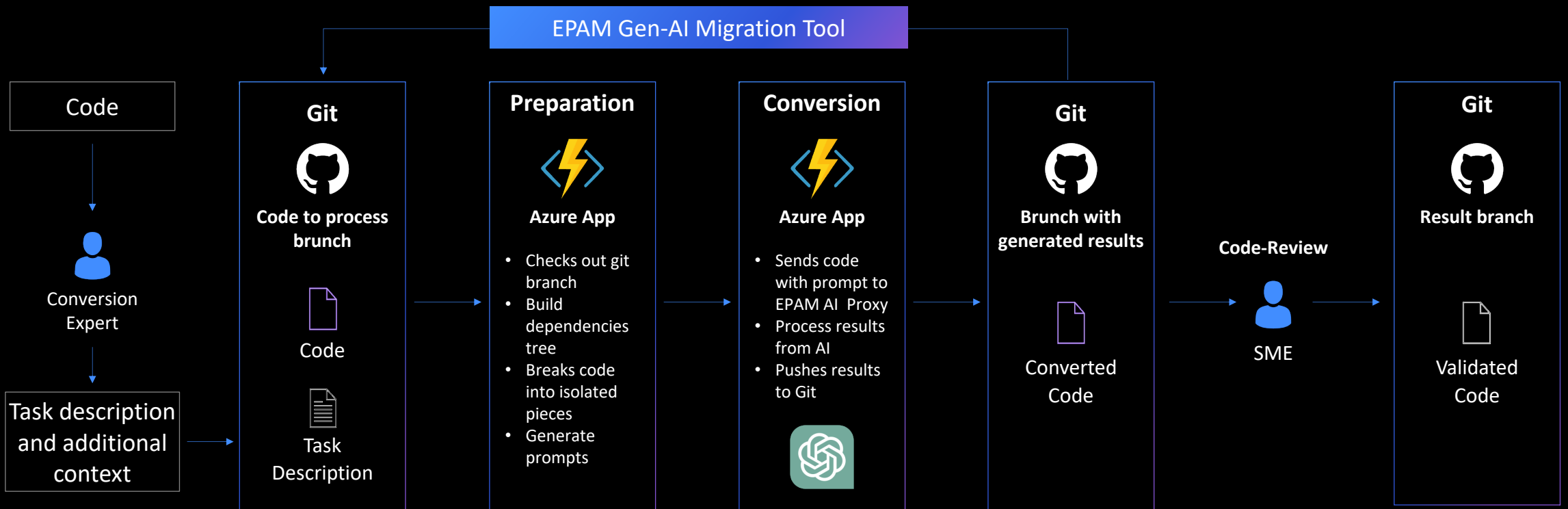
- 1 **Set up an environment** for ETL conversion from legacy low-code to pySpark
- 2 **Converter Reader Configuration** configures as per source metadata (up to 20 days depending on source)
- 3 **Writer Configuration** uses appropriate configuration as per target platform
- 4 **Convert & Iterate** perform conversion and iterate/adapt through ~80% code (convert 20% high-complexity workloads manually)

KEY DIFFERENTIATORS

| High-performance tools for automated ETL and SQL conversion from legacy low-code to pySpark

| Integrated with EPAM's framework for legacy workloads migration and reconciliation

migVisor Converter Empowered by Gen-AI



Unlock the power of seamless code processing with OpenAI, enabling effortless transformation of code, streamlined collaboration with git, and innovative integration into diverse projects.

WHY

Streaming Cloud Migration is a complex process, and full of uncertainties

Extremely time-consuming process of manual assessment and cost estimation

Lack of automated migration assessment, planning and implementation tools for a repeatable process

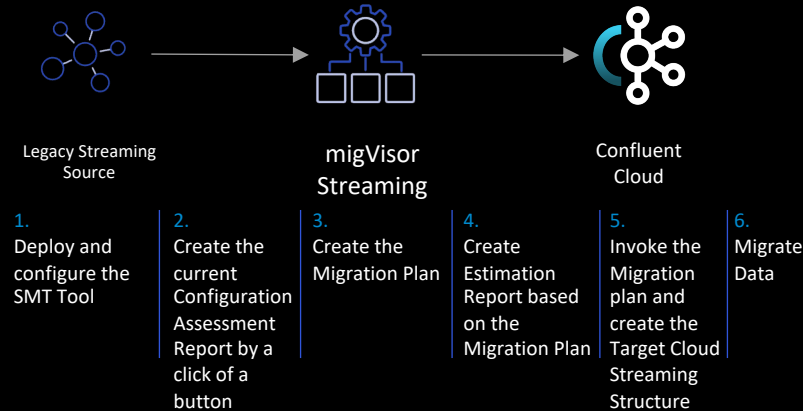
WHAT

HOW

KEY CAPABILITIES

- Accelerating & Streamlining a complex Migration Planning process through intuitive Admin UI
- Eliminating the barriers to buy by reducing the uncertainties surrounding the target streaming platform migration complexity and future runtime cost with a click of a button
- Approximately reducing streaming migration time and cost by more than 50% -based on previous manual migration experience-

HOW TO USE IT ON A PROJECT



migVisor Streaming – Confluent Edition

Data Platform Type	Data Source Name	Connectors	Master	Assessments
Apache Kafka 3.0	Default Apache Kafka Test Cluster	kafka-connect@003	kafka@20902	
Confluent Cloud 2.0	Default Confluent Cloud Test Cluster	kafka-connect@003	kafka@20902	
AWS MSK 1.0				
AWS Kinesis 1.0				
Apache Kafka 3.0				

Estimate billing for "EPAM sandbox"

Category	Value
MAX WRITE THROUGHPUT:	200 MB/s
MAX READ THROUGHPUT:	200 MB/s
DATA WRITE:	\$7041.0156125 GB-Month
DATA READ:	\$71423.046075 GB-Month
DATA STORED:	\$1.0664
TOPICS ALL:	5841
TOPICS WITH INFINITE RETENTION TIME:	43
PARTITIONS:	6381

Total Monthly Estimates

	aws marketplace	Microsoft Azure	Google Cloud
BASIC SINGLE ZONE	\$54,456.19	\$52,174.55	\$49,892.91
STANDARD MULTI ZONE	\$31,233.88	\$29,522.65	\$26,670.60
STANDARD SINGLE ZONE	\$27,241.01	\$24,959.36	\$22,677.72

KEY DIFFERENTIATORS

Intuitive Migration Assessment & Planning UI Tools

Ease of Deploy and Use

Integrated into a complete E-2-E Cloud Migration Plan

Data Reconciliation

Empowered by migVisor Reconciler

WHY

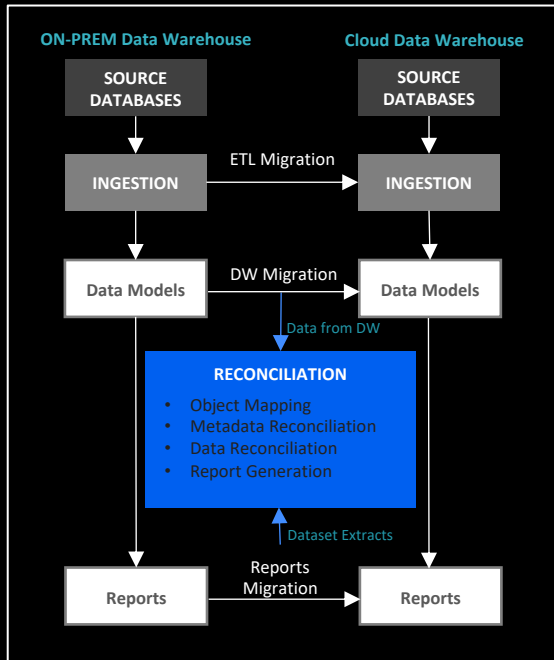
Inconsistency in scheme and database after the migration from legacy DWH

Discrepancy in audit reports between cloud and legacy BI

High complexity of Data Quality checks during the migration execution

WHAT

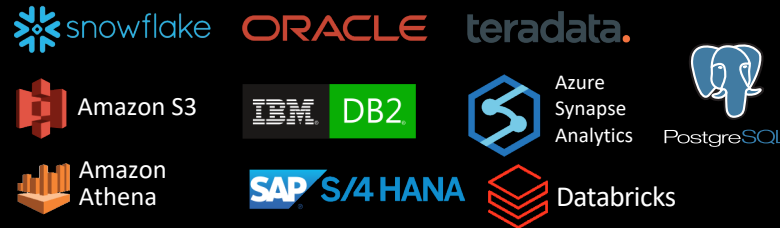
COMPLETELY INTEGRATED INTO MIGRATION PROCESS



RECONCILIATION CAPABILITIES

- Schema comparison (tables, columns, partitions, DB objects)
- Statistics comparison (Row count, checksum)
- Data comparison (value by column)
- Automated database scanning
- AI-driven approach for mapping tables, columns and data types
- High-performance scalable data comparison
- Several layers of reconciliation (quick, detailed, deep analysis)

Connector & Metadata Readers



HOW

HOW TO USE IT ON A PROJECT

- 1 Setup environment**
 - Deploy infrastructure
 - Configure connectors
 - Setup monitoring
- 2 Create mapping for databases and reports**
 - Execute automated mapping based on AI technologies
 - Review established mapping for schema and data types
- 3 Integrate with migration process**
 - Integrate with CI/CD
 - Include into regular testing
 - Add to data quality procedures
 - Add to audit process
- 4 Analyze reconciliation report**
 - Setup report subscriptions
 - Review reconciliation report and provide feedback

KEY DIFFERENTIATORS

Reconciliation on reports and data level

Automated AI-based process of database scanning and mapping

Scalable solution for large datasets

We are applying our accelerators to execute migration for industry leaders

GLOBAL FOOD COMPANY



ETL & Report Migration

Client engaged EPAM to perform an assessment of CBI & Perseus reporting tools and propose an efficient migration approach with a focus on business value.

Used EPAM's Migration Assessment Methodology to **investigate 60k legacy ETL pipelines, 37k reports and 16 data platforms** over 6 weeks.

Worked with application owners and global architecture team to **determine optimal migration path**.

GLOBAL TELECOMMUNICATION COMPANY



Rapid Discovery and Assessment for Database Migration

Driven by an urgent need to leave the on-prem data center, a multinational telecommunications, information technology, and consumer electronics company came to EPAM.

Executed **detailed assessments** for all source databases (1,000+), including PostgreSQL and MySQL.

Analyzed **additional 3,000** databases as part of the assessment.

Closed the project in 90 days, which likely **saved 1 year of expensive analysis**.

GLOBAL PETROCHEMICALS COMPANY



Continuous data reconciliation during migration

Reconcile financial data sourcing from multiple SAP ERP source systems to SAP CFIN.

Reconciliation results are shown in the Power BI Dashboard PDF/Excel reports.

Reconciler automatically mapped source and target tables performed schema conversion checks and assessed data quality post-migration. **95% of the data (300 mln records)**, was reconciled within a week. migVisor Reconciler helps speed up the reconciliation increase the reliability and accuracy of the data.

INFORMATION SERVICE, EDUCATION AND FINANCIAL COMPANIES



Code Conversion for migration streamline

Data Warehouse were migrated to a Databricks-based stack.

EPAM utilized an automated tool powered by OpenAI's Language Model (LLM), which converted **90%** of low to medium complexity code, and **50%** of highly complex code, streamlining the migration process.

The conversion significantly accelerated, reducing conversion times by **4 times** for SQL and **3 times** for SSIS components. This transformation also harnessed the capabilities of Databricks for better data management.

Thank you!

For more information, contact

Vitalii Bondarenko

Director, Data & Analytics Consulting

Vitalii_Bondarenko2@epam.com