Cloud Data Migration Offering

Service Offering Pitch

2023

Contents

AI-DRIVEN Data Cloud Migration to fully unlock efficiency of your Data driven organization

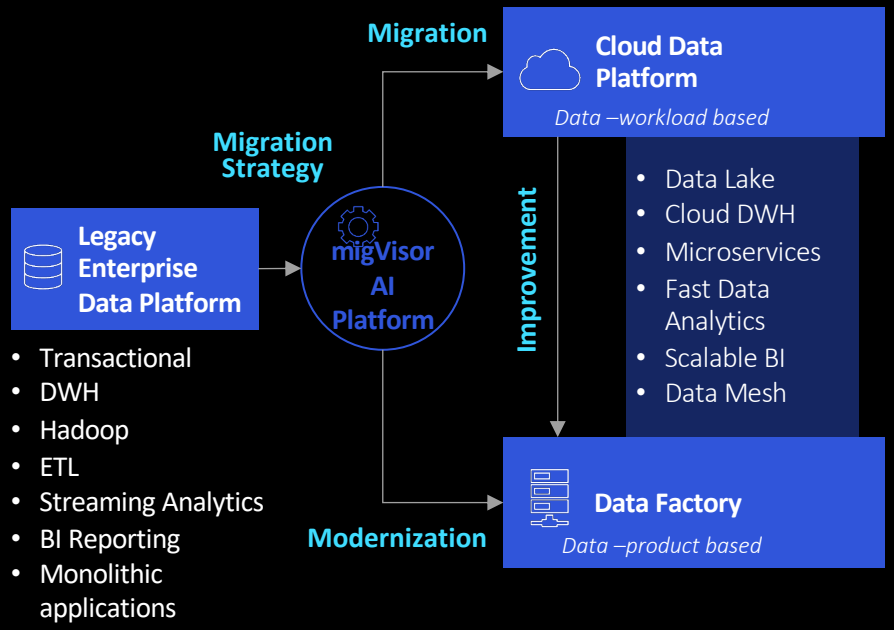
WHY

- License cost optimization
- Data Platform Modernization & Optimization
- Enable business by the Cloud Innovative Capabilities
- Support business strategy execution

AI-DRIVEN CLOUD MIGRATION PROGRAM — Pillars

WHAT

AI-Driven Data Cloud Migration program includes migration of Data Platform and key components



HOW

1 **EPAM Data Migration Framework speeds-up migration and makes it possible to have a first migration to PROD in 3 months**

2 **migVisor AI Platform - Accelerators and Tools for rapid delivery**

WHY EPAM

Vendor & Cloud Agnostic Solution

EPAM provides cloud agnostic solutions, can support, execute and consult in data migration

EPAM Data Migration Framework

Methodology, documents, tools and technical approaches for rapid and successful migration

Tools, Best Practices, Accelerators

Comprehensive toolkit library – SMART AI-driven tools, Accelerators and best-in-class tools

Partnerships

Microsoft Google databricks amazon snowflake

01

Why migrate to the Cloud?

Data Migration Technical Drivers

Data Migration process is initiated by several case scenarios and are driven by technical and business drivers

LICENSE COST OPTIMIZATION

- Reducing dependencies on Data Center infrastructure
- Migrating to cloud native products from legacy tech stack and diminish dependency on legacy software licenses
- Optimize cost strategy and scalability according to business objectives

IMPROVE, MERGE, RATIONALIZE, OPTIMIZE

- Improve operational efficiency
- Leverage cloud elasticity
- Move to open source or cloud technologies
- Get off the hook of ISVs
- Improve performance on Data Analytics
- New Data Products & Analytics Improvement enabled by Cloud Technologies (AI, Scalable Data Processing, Steaming Data Processing, etc)

ENABLE CLOUD DATA CAPABILITIES

- Business empowered by Cloud Data Platform Scalability
- Rationalize existing workloads/applications
- Enable new business use cases
- Set foundation for scalable and advanced analytics
- Merge & Acquisition with legacy Data Platforms
- Digital Transformation as a strategy for cloud migration

EPAM's Response to Key Challenges

STRATEGY

Migration to Cloud

Low cost & low efforts

Low business value

Modernization to a Cloud Data Platform

High cost & high efforts

High business value

CHALLENGES

Low adoption

LACK OF RELIABILITY, QUALITY AND FUNCTIONALITY

Over-time

UNABLE TO DECOMMISSION LEGACY SYSTEMS — LACK OF CAPABILITIES

Over-budget

LOWER THAN EXPECTED ROI

SOLUTION

SMART MIGRATION

Empowered by EPAM AI Toolset & migVisor Platform

Focus on user's functionality. Advanced dependency analysis. Scope and complexity assessment. Migration plan optimization.

02

What needs to be migrated?

EPAM Helps to Define Value and Choose the Most Relevant Migration Strategy

	Lift & Shift Migration "As Is"	Re-platform Systems & Platforms are replaced by Cloud-Native Services	Re-architect Systems replaced by Cloud-Native Services, Application Logics is redesigned
Business & Tech Value • Performance • Security • Scalability • Innovations • Foundation for Data Products development	**LOW** • Security is migrated from on-prem to Cloud without changes • Cloud Advance Analytics capabilities are not enabled • All issues in performance and scalability are migrated with legacy systems	**MIDDLE** • High Performance and Scalability • Security is integrated with cloud services • Data Platform is limited by the migrated solution, there is no foundation for data platform growth	**HIGH** • High Performance and Scalability, Security is integrated with cloud services • Data Platform foundation for further growth and new innovations enablement
Legacy & Tech Debt • Legacy Toolset • Performance issues • Low data observability • Low Data Quality	**HIGH** • All issues are migrated from on-prem to cloud with the data solutions	**MIDDLE** • Toolset is preplaced and performance is improved • Data processing approach has remained	**LOW** • Platform is replaced • Solution is optimized for cloud • Data processing is re-architected
Migration Cost • Platform modernization • Data transferring • Logic transferring, converting or re-architecting	**LOW** • Low-cost solution, only Cloud infrastructure has to be designed and implemented	**MIDDLE** • Logic conversion from legacy to Cloud-Native is the most extensive part of the scope	**HIGH** • The most expensive strategy, as the platform is redesigned from scratch
Migration Time • Solution Design • Cloud Environment Configuration • Solution Implementation • Data Transferring from on-prem • Data Reconciliation	**LOW** • Fast migration as doesn't require solution re-architecting and application code redesign • Data Transferring takes time as includes data extracting from DW and moving to Cloud environments	**MIDDLE** • Logic Conversion is a regular software development project with consists of design, implementation and testing	**HIGH** • Platform Design and Implementation is a regular software development project and takes long time
Considerations • Cases • Pros & Cons	**Cases**: Since Lift & Shift Migration Strategy provides fastest results in shortest timeframe, this approach can be considered for fast Data Center termination and historical data archiving **Pros**: Fastest results, Lowest cost **Cons**: Highest OPEX, Lowest business outcome, Security risks	• **Cases**: Preferred solution to optimize cost of ownership, introduce pay-as-you-go model when investment in modernization doesn't bring much business value from new cloud features • **Pros**: Dynamical pricing model • **Cons**: Higher cost of changes, lower time-to-market	• **Cases**: Re-architect is the best option for most innovative solutions that plan to use most advanced cloud services and CI/CD SRE practices • **Pros**: Enables cloud innovation, Lowest TCO • **Cons**: High cost, Long duration

EPAM recommends

Based on Customer's desired outcomes and constraints, EPAM will advise most relevant Cloud Data Migration Strategy.

A combination of all 3 strategies can bring the best ROI for the Customer.

EPAM Helps to Define Complexity for Data Platform Components

Upstream (ETL)
- ETL
- Integration
- Staging

DW Engine (Legacy)
- DW RDBMS
- Data Schema
- Data in DW
- SQL Logic

Downstream (Reports)
- Reporting
- Applications
- Integration

Data Warehouse Modules

Upstream: Ingestion Logic for DW with data processing and collection including incremental and stream data load

Downstream: Analytics applications including API, Reports, Dashboards and extracts of data for integration with business applications.

Migration Strategy & DW Modules	Lift & Shift Migration "As Is"	Re-platform Systems & Platforms are replaced by Cloud-Native Services	Re-architect Systems replaced by Cloud-Native Services, Application Logics is redesigned
Transactional and DW Engine RDBMS engine with logic and data	LOW COMPLEXITY • Migration "As Is", no changes	MEDIUM COMPLEXITY • Schema conversion • SQL Objects conversion • Data transferring with transformation	DATA ONLY • Data transferring from DW to Cloud storage • Data load to a new platform's storage with transformation
Upstream Ingestion Logic for DW	LOW COMPLEXITY • Migration "As Is", no changes	HIGH COMPLEXITY • Ingestion Logic from Data Sources to Raw Data Storage • Staging Storage • Transformation & Aggregation Logic • Load Logic to DW	REDESIGNED • Redesigned in a new platform
Downstream Data Consumers for DW	LOW COMPLEXITY • Migration "As Is", no changes	HIGH COMPLEXITY • Queries to DW • Data Visualization Logic	REDESIGNED • Redesigned in a new platform

EPAM recommends

1 **Re-architect for all modules**

Transactional, DW, Upstream and Downstream will be redesign according to the best practices in cloud-native platforms, security, scalability and performance.

2 *As a cheaper and faster alternative*
Database Engine Re-platform Upstream and Downstream Lift & Shift with improvements

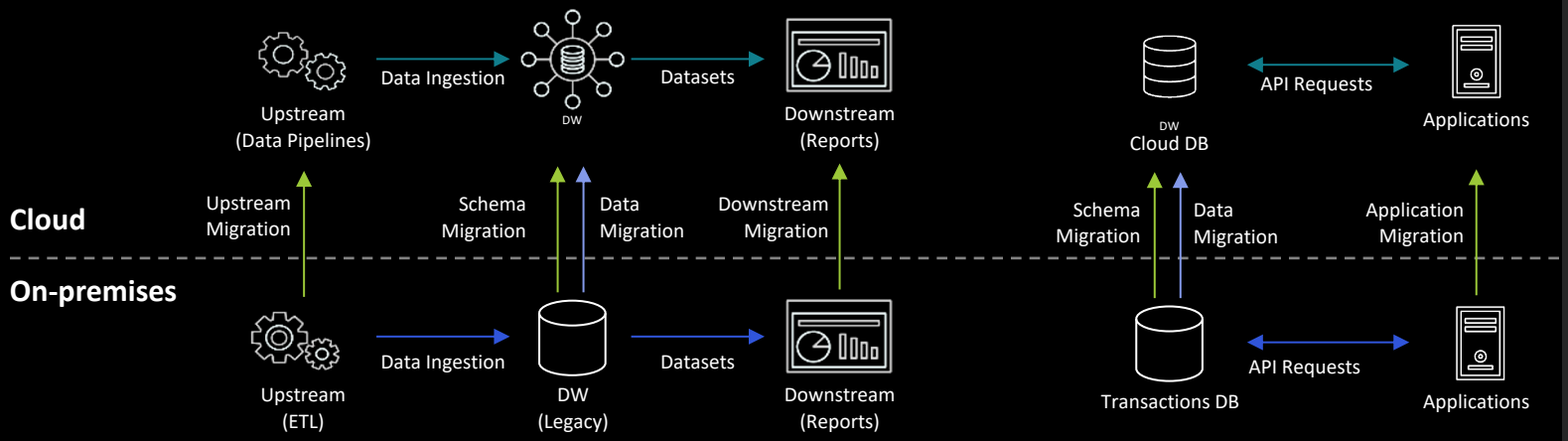
SQL Queries in Upstream and Downstream need to be converted to a new DB Engine's format. In this migration pattern the complexity of migration is lower, and, at the same time, it enables Cloud capabilities in scalability and performance on DB Engine.

EPAM helps to define optimal scope of Migration

Migration scope can be decomposed by **Applications** or **Data Workloads**

	Lift & Shift Migration ”As Is”	**Re-platform** Systems & Platforms are replaced by Cloud-Native Services	**Re-architect** Systems replaced by Cloud-Native Services, Application Logics is redesigned
By Applications Data & Logic are migrated as a one package in scope of an application	**SUPPORTED** • Migration “As Is”, no changes	**SUPPORTED** • Schema conversion • SQL Objects conversion • Data transferring with transformation	**SUPPORTED** • Data transferring from DW to Cloud storage • Data is loaded to a new storage with transformation
By Workloads Data & Logic are migrated to Cloud by workloads.	**NOT SUPPORTED** • Workloads cannot be identified as all data and logic are migrated with systems & platforms	**SUPPORTED** • Workloads are defined as set of data tables, logic, queries in Upstream and Downstream	**SUPPORTED** • Workloads are defined as set of data tables, logic, logic in Upstream and Downstream

Migration By Workloads



EPAM recommends

Gradual Migration by Workloads

Scope of migration is defined as a set of workloads. Each workload is presented as a combination of data, logic in transaction workloads and/or visualizations and reflected in tables, transformation logic in DW, logic in Upstream and Downstream.

Migration by workloads is executed as series of projects, where in each project, only one workload is migrated, tested and released to PROD environment.

Since workload provides greater isolation of data streams, this approach removes the need for intermediate applications integrations implementation and verification. Therefore, it reduces the time, cost and improves quality of migration.

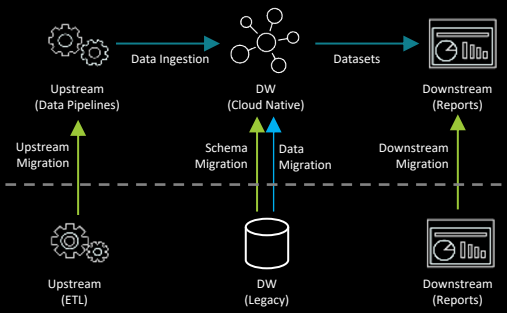
EPAM Helps to Optimize Time and Cost of Migration

Migration scope and approaches are combined into patterns

FULL MIGRATION

Migration from On-prem

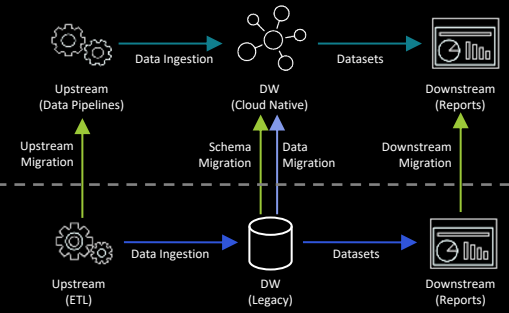
All workloads are migrated to cloud. Transactional, DW, on-prem Upstream and Downstream are switched off.



PARTIAL MIGRATION

Faster solution

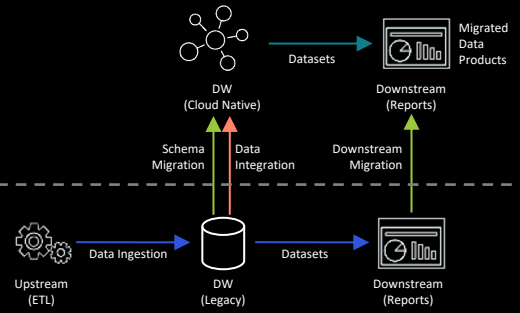
Data is separated by scope of workload that need to be migrated to cloud and some workloads that should remain on-prem.



HYBRID DOWNSTREAM

Safer & faster solution

Upstream is not migrated into cloud because complexity of Upstream migration is very high. Only selected workloads are migrated in DW and Upstream



	FULL MIGRATION	PARTIAL MIGRATION	HYBRID DOWNSTREAM
TIME	LONG-TERM	MIDDLE	SHORT-TERM
COST	HIGH	MIDDLE	LOW
COMPLEXITY	HIGH	MIDDLE	LOW
BENEFITS	HIGH	MIDDLE	LOW

EPAM recommends

Full Migration

All workloads to be migrated to cloud. DW, Upstream and Downstream are switched off on-prem after the migration

Disadvantages: long-term program, expensive, requires additional testing

Advantages: move to cloud fully, switch-off legacy on-prem

In case of cost / time constrains, high complexity other option might be consifered

02

How to migrate to the Cloud?

Cloud Data Migration Journey with EPAM

CLIENT'S VISION

Assessment & Planning → Migration Execution → Cutover & Support

EPAM'S ANSWER

Cloud Data Migration Framework

LEARN → PLAN → VALIDATE → EXECUTE → SUPPORT → IMPROVE

Migration Strategy	Processes & Handbooks	Delivery Teams	Best Practices & Trainings Programs

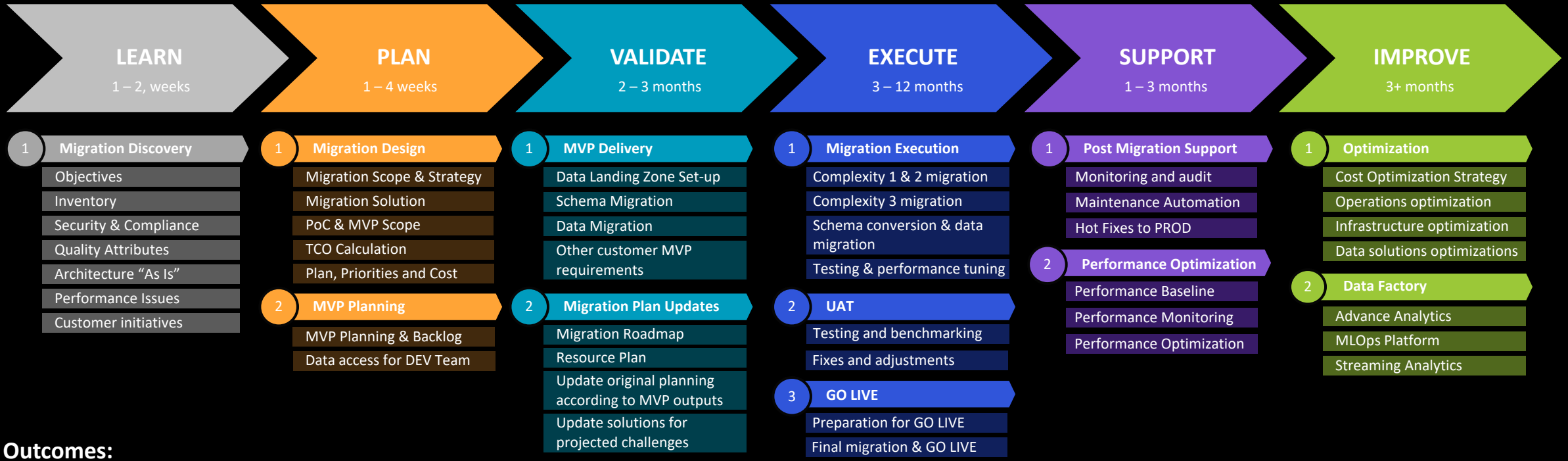
migVisor PLATFORM

migVisor transactional | migVisor analytics | migVisor streaming CE | migVisor landing zone | migVisor code converter | migVisor reconciler

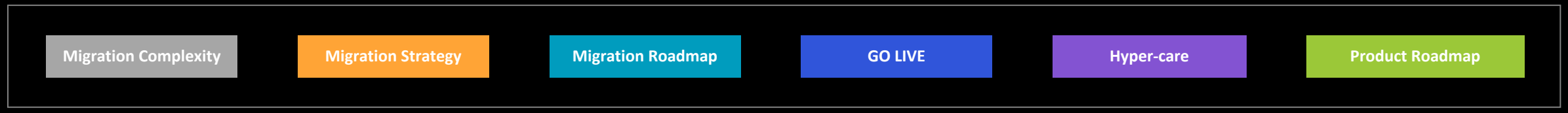
Analytics Assessment | **MVP and Execution** | **Testing & Reconciliation**

Cloud Data Migration Framework

EPAM uses cloud migration framework as integrated set of **methodologies**, **design practices**, **document templates**, **delivery processes** and **accelerators** to facilitate data migration, modernization and overcome challenges.
The framework speed-up migration and make possible to have a **first migration to PROD in 3 months as MVP**



Outcomes:



Accelerator Platform for Data & Analytics Migrations of Any Complexity

migVisor transactional

migVisor analytics

migVisor streaming CE

migVisor landing zone

migVisor code converter

migVisor reconciler

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

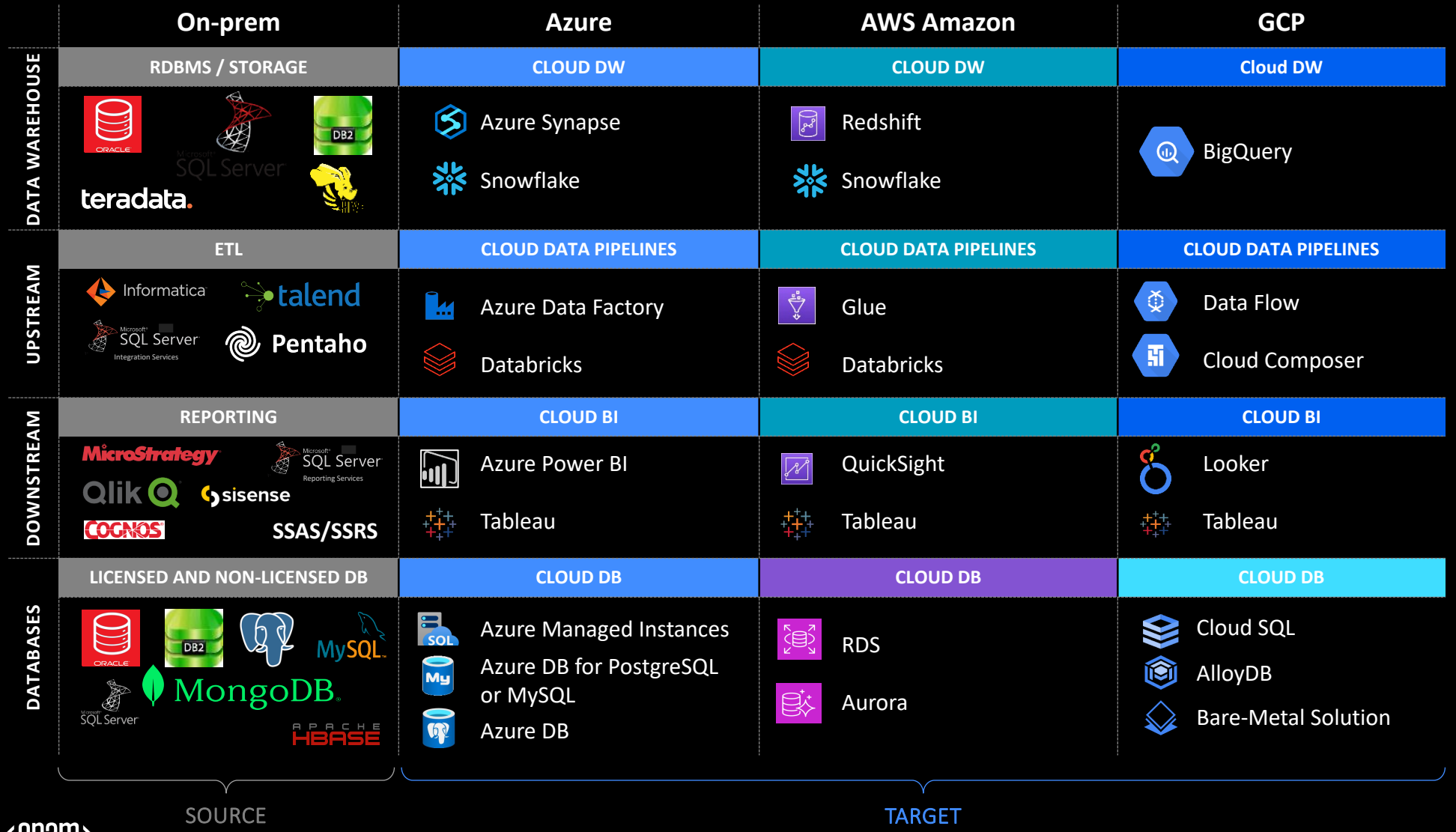
- Accelerating & Streamlining a complex Migration Planning process through intuitive Admin UI
- Eliminating the barriers to buy by reducing the uncertainties surrounding the target streaming platform migration complexity and future runtime cost with a click of a button
- Approximately reducing streaming 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



MIGRATION EXPERIENCE

Snowflake

150+ projects migrated from on-prem RDMBS 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

03

How EPAM can migrate fast, low cost and low risk?

OUR OFFER

Migration | Offering

Key Benefits

COMPLEXITY EVALUATION
EPAM provides detailed Inventory catalog. The catalog includes T-shirt complexity evaluation.

MIGRATION EFFORTS ESTIMATION
Smart Assessment provides migration scope, detailed migration efforts for each inventory item and total migration cost.

SCOPE REDUCTION RECOMMENDATIONS
AI driven dependency and usage statistics analysis reduces the scope of the migration, provides recommendations for data normalization and cost optimization

MIGRATION STRATEGY DESIGN
Includes recommendation for migration approaches (Lift&Shift, Re-platforming and Re- architecting), patterns of cut-over, tech stack selection and data quality testing approaches

MIGRATED DATA RECONCILIATION
Data Reconciliation on reports and data level. Automated AI-based process of database scanning and mapping. Scalable solution for large datasets

CODE CONVERSION
Automated ETL and SQL conversion from legacy low-code to pySpark . Integrated with EPAM's framework for legacy workloads migration and reconciliation

Smart Assessment Package

Assessment + Data Lineage Analysis

Price: **$50K**

$80K with PoC execution

Duration	**Up to 4 weeks**
Scope	• 6 interviews • 4 workshops • Manual estimation for fixed inventory scope • 10-15 cases for analyses
Deliverables	• Complexity for legacy • Migration estimates • Migration Strategy • PoC (250 p/d)

3 ppl

- Data Consultant
- Solution Architect
- Business Analyst (optional in case of complex scope)

Migration Pilot

Smart Assessment with Migration Design and Pilot Migration (MVP)

Price: **$350-450K**

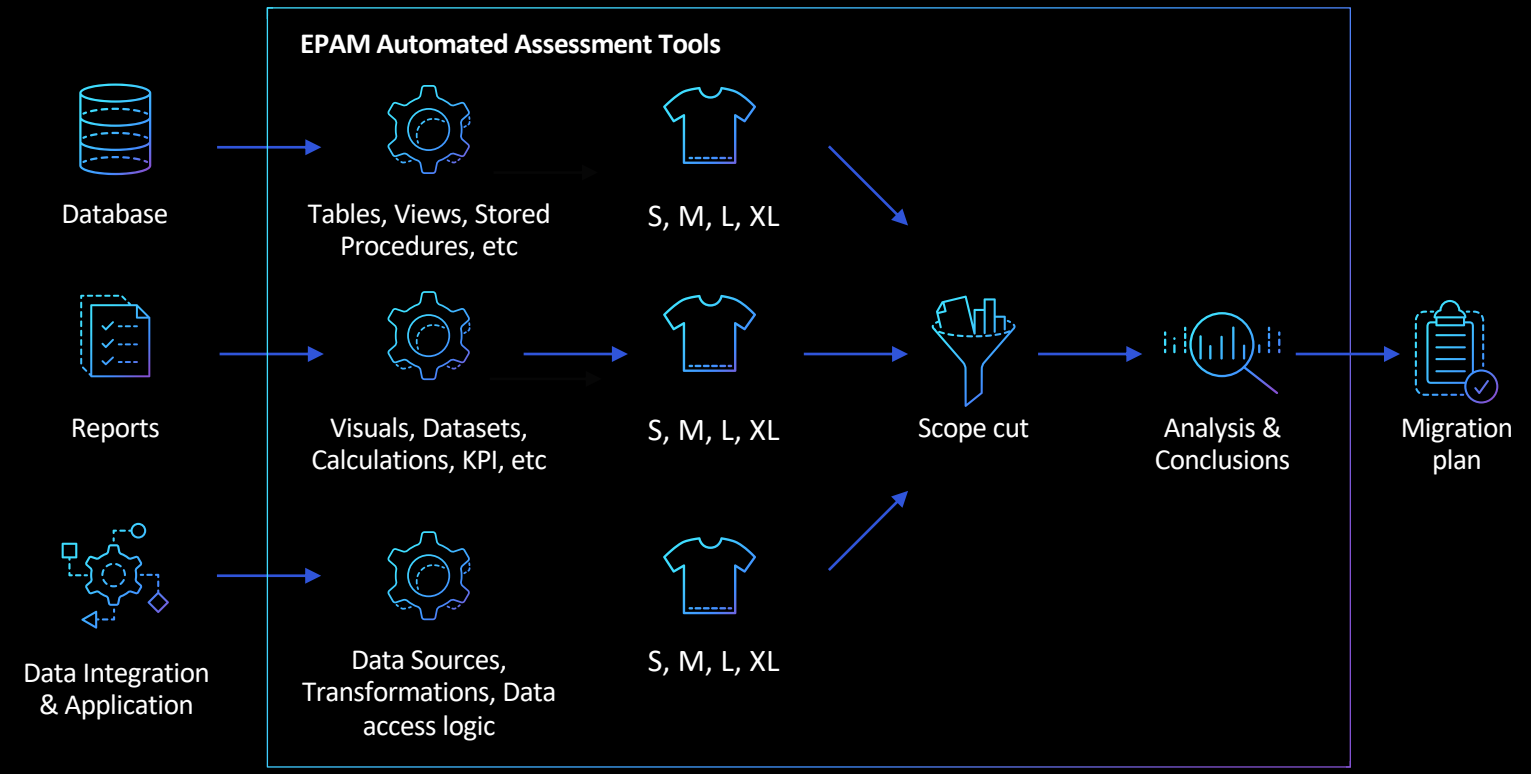
Duration	**Up to 12 weeks**
Scope	• Migration Roadmap Design • Validated migration plan • Execute MVP (Pilot Migration)
Deliverables	Migration plan & Pilot delivered

Delivery Team

- Delivery Manager
- Data Consultant
- BI Analyst
- Solution Architect
- Lead Data Engineer
- Data & BI Engineers
- DevOps
- Data Quality

Assessment| Approach

EPAM Assessment Accelerator Tool automates assessment and decreases time of estimation and complexity evaluation in 30+ times



1. Inventory with initial profiling (data sources, cubes, reports and pipelines)

2. Complexity criteria set up

3. T-shirt sizing based on defined criteria

4. Down selection of representative assets

5. Manually detailed estimation for 5% and calibration

6. Roadmap and timelines

Why Assessment Is Needed

Potential risks

1. Lack of proper planning or unmanageably long planning horizon
2. Poor assumptions and risks not addressed early enough in the process leading to missed deadlines
3. Underestimated efforts and budgets connected to the landscape complexity and size
4. Missing business case and no clear value articulation
5. Inadequate migration sequencing and missed interdependencies
6. Lack of training on the new technology and experience

Assessment | Savings

EPAM's tools for automated inventory analysis can decrease time of assessment in **3+ times** and cost in **9+ times**

	Analytics Assessment (manual)	Operational DB Assessment (manual)	Automated Assessment
Scope	**Tables: 4,000** **Views: 2,000** **Stored Procedures: 500** **Pipelines: 14,000** Reports: 15,000	DB: 400 Stored Procedures: 10K Tables: 200K Views: 100K	Applicable for both
Duration	Up to 4 months	Up to 4 months	1 month
Team	15 ppl • 1 Data Consultant • 2 Business Analysts • 1 Solution Architect • 4 Technical Leads • 7 Data Engineers	9 ppl • 1 Data Consultant • 2 Business Analysts • 1 Solution Architect • 5 Data Engineers	3 ppl • 1 Data Consultant • 1 Business Analyst • 1 Solution Architect
Activities	• Up to 50 interviews • BI Inventory review • DB & ETL Inventory review • Manual estimates for inventory items • Inventory analysis	• Up to 400 interviews • App. review • Data integration review • Manual estimates for inventory items • Inventory and dependencies analysis	• 6 general interviews • Inventory scanning and automated estimation • Estimates calibration by sampling • Inventory analysis
Deliverables	• Assessment results • Migration Strategy & Roadmap	• Assessment results • Migration Strategy & Roadmap	• Assessment results • Migration Strategy & Roadmap

$850K+

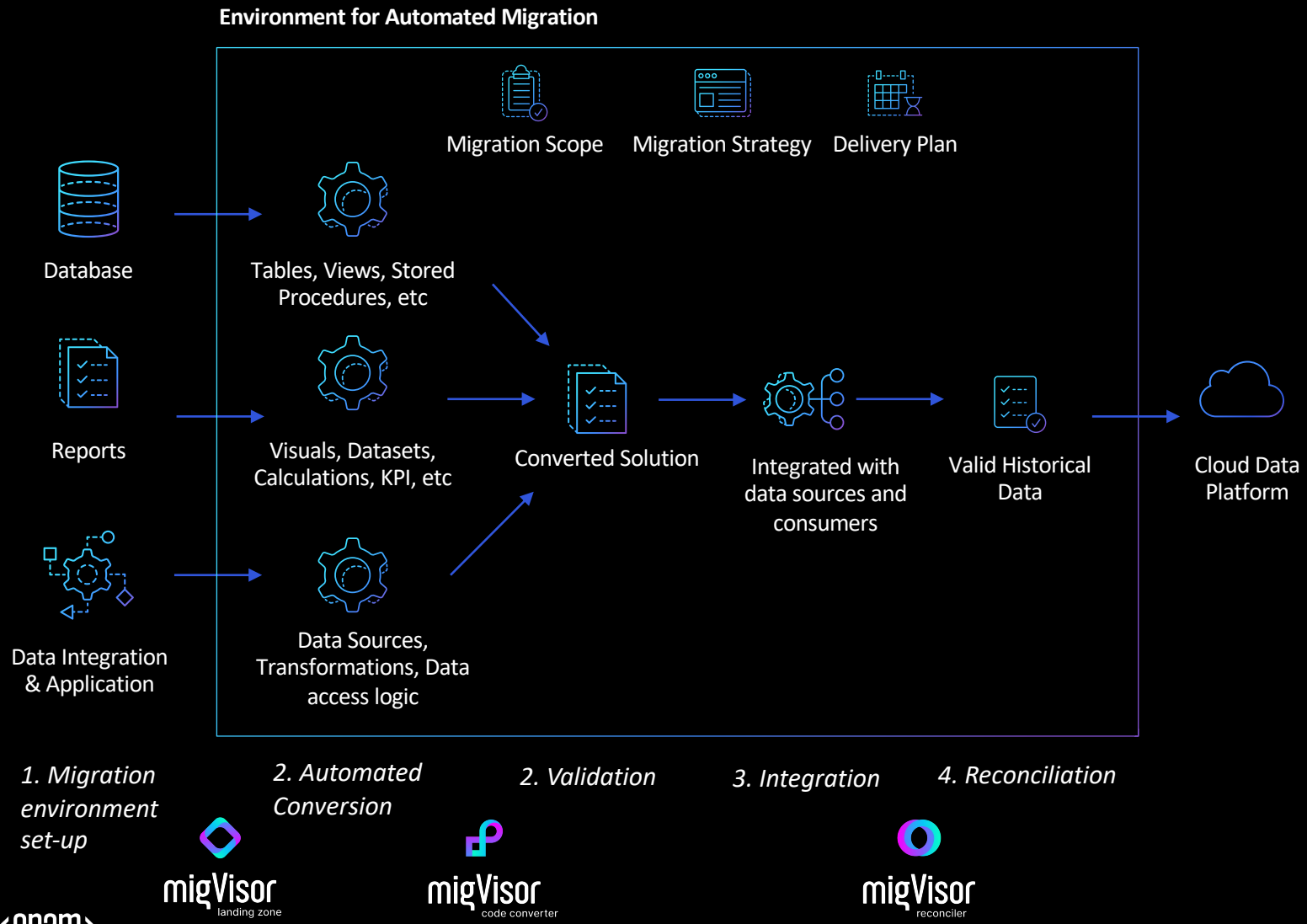
Total saving

with Automated Assessment

Migration | Automation with EPAM

Migration Objects	What is in the Scope	How we automate	Speed-up the process
Schema Migration	Table description, Views, Stored Procedures, etc	• Converts DDL scripts from legacy database to cloud database; converts tables, views, stored procedures, functions, triggers, and security. Applies converted DDL files on cloud database. • Provides reconciliation after migration of data and schema. It is configurable for different types of reconciliations (e.g., by checksum, by data types, etc.)	up to 40%
Data Migration	Historical data	• Converts scripts from legacy DW to cloud database; converts tables, views, stored procedures, functions, triggers, and security. • Provides reconciliation after migration of data and schema. It is configurable for different types of reconciliations (by checksum, by data, etc.)	over 50%
ETL Migration	Data Sources, Transformations, mappings, aggregations, data quality checks	Converts ETL transformation pipelines to Python code. Supports a configurable approach for python and PySpark code generation; coverts connectivity and mapping from/to data tables. Translates SQL Queries, expressions and aggregations into format of target platform	up to 40%
Application Migration	Transformations, Data access logic	• Converts DDL scripts from legacy database to cloud database; converts tables, views, stored procedures, functions, triggers, and security. Translates SQL Queries in data access layer for all applications. • Provides reconciliation after migration of data and schema. It is configurable for different types of reconciliations (e.g., by checksum, by data types, etc.)	up to 20%
Reports Migration	Visuals, Datasets, Calculations, KPI, etc	• Converts semantic data model from legacy reporting tools to cloud-native reporting tools. Converts relational models, calculations and security. Translates SQL Queries from the legacy database format to the cloud database format. • Automate and reduce cost of UAT, comparing data in source and target platforms	up to 60%

Migration| Approach to speed-up migration by 20-60%



How we can Automate Migration

Key highlights

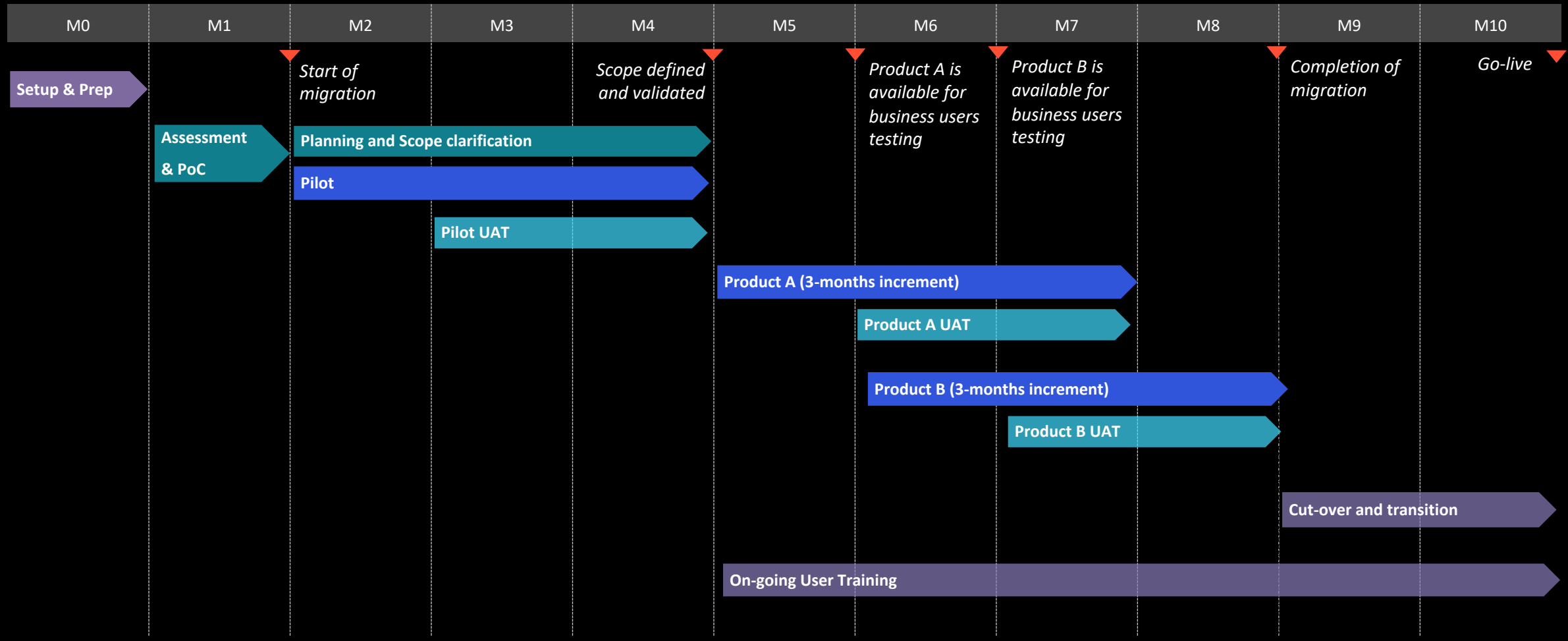
Migration Risks:

- High-complexity of legacy ETL data transformation pipelines in low-code tools increase time of the migration project
- Extremely time-consuming process of manual SQL and ETL conversion to pySpark
- High-complex process of data quality testing and reconciliation

What can be automated:

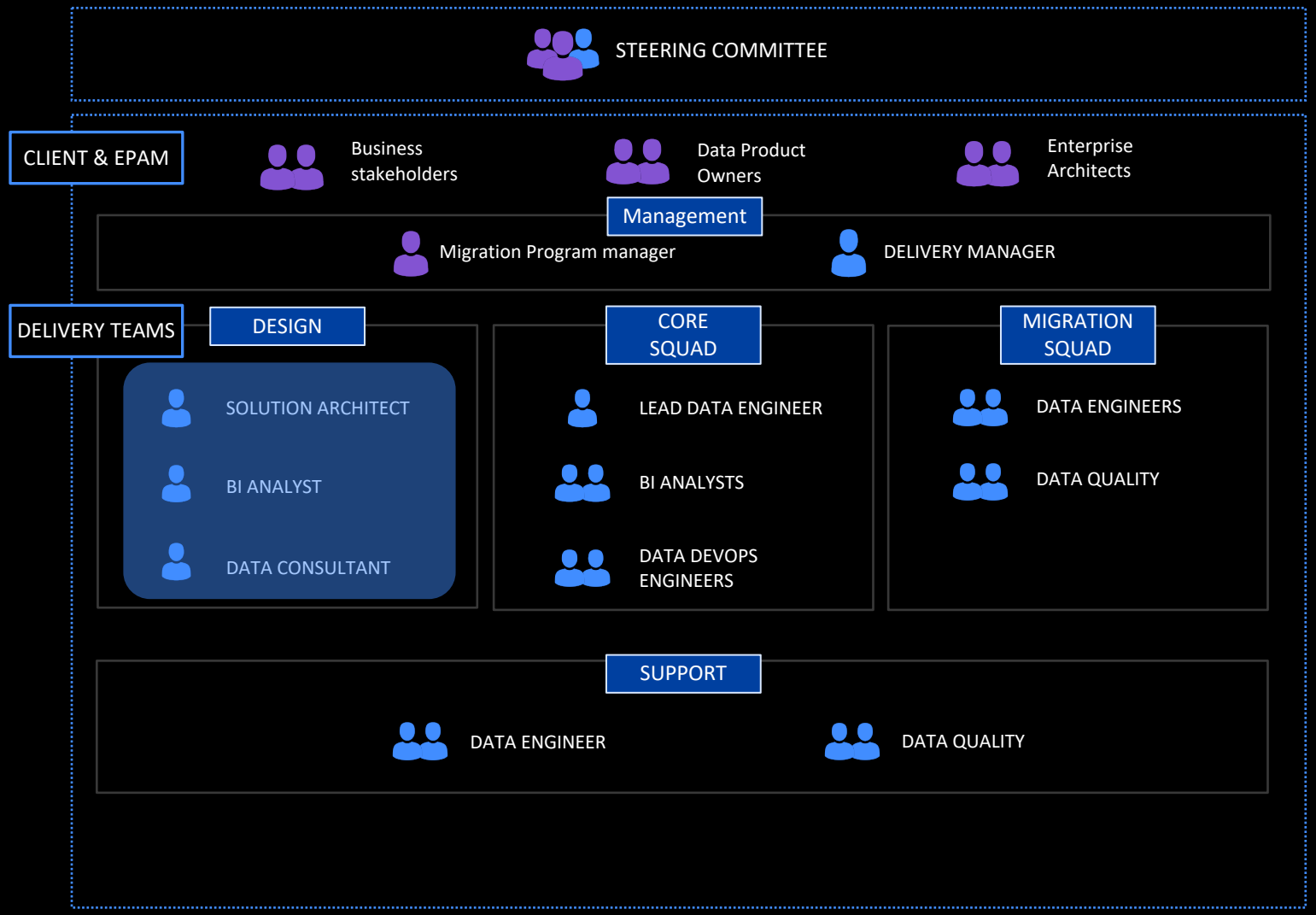
- High-performance tools for automated ETL and SQL conversion from legacy low-code to pySpark
- Integrated with EPAM's framework for legacy workloads migration and reconciliation

Migration | Indicative Delivery Timeline



* Timeline is indicative and can be changed during the migration progress according to inputs and change requests from business users, outcomes from reports validation, changes in global architecture and delivery dependencies

Migration | Indicative Team Structure



Steering committee defines strategy & oversight to deliver value. Includes project sponsors from CLIENT and EPAM.

Client and IT expertise is made up of senior domain and technology experts who provide the necessary information (AS-IS architecture, business and technical requirements, data extractions, etc.) to EPAM team

Delivery team consists of several squads for each workstream

Communication model is set according to the best practices of EPAM

CLIENT

EPAM

ASSESSMENT TEAM

04

Case Studies

Cloud Migration Track Record

ASSESSMENT EXPERIENCE

GLOBAL CPG COMPANY

CONSUMER & RETAIL

ETL & Report Migration

Assessment of reporting tools and development an efficient migration strategy with a focus on business value

GLOBAL HEALTHCARE & LIFE SCIENCE COMPANY

HEALTHCARE & LIFE SCIENCE

Assessment of Pfizer RIF Platform

Assessment for Platform's data migration to R&D Data Lake platform to reduce storage costs, enable cloud-based processing and enhance flexibility

INTERNATIONAL INFORMATION & MEADIA COMPANY

INFORMATION & MEDIA

Assessment of Informa DB for Due Diligence

Assessment of MS SQL databases aimed to assess the company's portfolio performance solutions - evaluate current state and define an approach to modernization

MIGRATION EXPERIENCE

GLOBAL LUXURY BRAND

CONSUMER & RETAIL

Analytical Platform migration from Cloudera Hadoop to Databricks Cloud

Client engaged EPAM to implement migration to Databricks to reduce operation costs as well as increase flexibility.

FINANCIAL SERVICES COMPANY

FINANCE

Migration rom Hadoop to Azure Databricks

Company approached EPAM with an ask for a Data Platform modernization from Hadoop to Azure Databricks with more than 50 TB of historical data migration.

GLOBAL CPG COMPANY

CONSUMER & RETAIL

Analytical Platform migration from Cloudera Hadoop to Databricks Cloud

EPAM performed assessment and reduced 30% of scope before the migration and migrated analytics to Azure Cloud using migVisor tools for code conversion and reconciliation

SELECTED EXAMPLE OF OUR RESULTS

ETL & Report Assessment

Business Background

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

Client Challenges

- Client has defined an overall cloud migration program. Significant part of client's data analytics was still located in historical on-premise environment.
- Global and local analytical solutions have been built upon a historical architecture and have accumulated a massive amount of data platforms over the last 10 years, including BI reports, ETL processes, and database objects
- Due to complex reporting structure, absence of owners, and historical architecture, it was a challenge to assess the migration complexity and business value

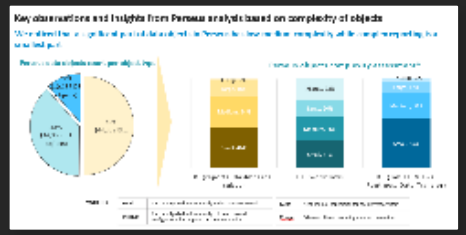
Technologies assessed

DB2, Informatica, Microstrategy, MS SQL, SSIS, SSAS, SSRS

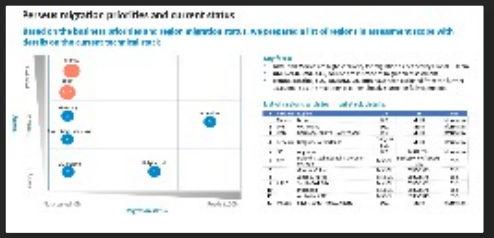
Solution Highlights

- EPAM's migVisor assessment tools were applied to **investigate 60,000 legacy ETL pipelines, 37,000 reports and 16 data platforms** over 6 weeks
- Reduced scope of migration by **70%**
- Defined complexity and estimated migration efforts for the whole program
- Worked with application owners and global architecture team to **determine optimal migration path**

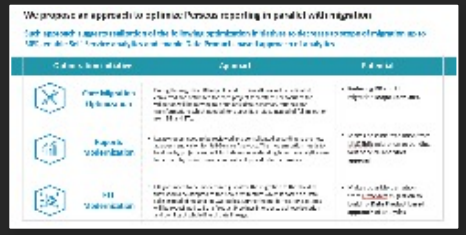
Key insights from analytical tools: usage patterns, etc.



Migration priorities report



Migration approach recommendations



Migration scope & estimates

Migration Execution

UK Analytics Migration

Scope

Client wanted to migrate UK Analytics to Azure Cloud. Legacy stack consists of DB2, Informatica, Microstrategy, MS SQL, SSIS, SSAS, SSRS.

After the assessment with migVisor tools the scope was **reduced to 150+ Informatica workflows, 1200+ SQL procedures, 1000+ DB2 tables, over 2000** Microstrategy reports.

Solution Highlights

- Instead of reports migration Self-service analytics with Data Products were implemented
- Microstrategy Reports and Informatica ETL were re-designed to Data products based on the Data lineage mapping conducted by migVisor
- Informatica workflows were converted to Python and Snowflake SQL
- DB2 DWH was migrated into Snowflake
- OLAP Cubes were converted to Snowflake views

Results

- Project was finished on time within budget
- 30% of scope was reduced before the migration
- Client successfully decommissioned legacy stack and optimized license's cost
- migVisor was used for Informatica conversion, which reduced migration time by 30%
- UAT before go-live was automated by migVisor Reconciler

Corporate Data Lake Migration

Scope

Client wanted to migrate to Azure Cloud central Data Lake based on DB2 and Informatica including integration with all corporate tools such as SAP, SF, etc.

The total scope was **38,000 tables, 92,000 views, 50,000 constrains, 10,000 stored procedures , 1000 data pipelines** . After the assessment with migVisor tools the scope was **reduced by 70%**

Solution Highlights

- Built centralized Data Lake in Azure, all data transformations were migrated to Databricks
- Final data sets were integrated with Snowflake Data marts
- migVisor Reconciler was used for the Data quality testing

Results

- EPAM has designed and built a new Data Lake and ingestion layer in Azure Cloud
- Project is still ongoing

We are applying our accelerators to execute migration for industry leaders

GLOBAL FOOD COMPANY

migVisor analytics

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

migVisor transactional

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

migVisor reconciler

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

migVisor code converter

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.

05

Appendix

Data Reconciliation

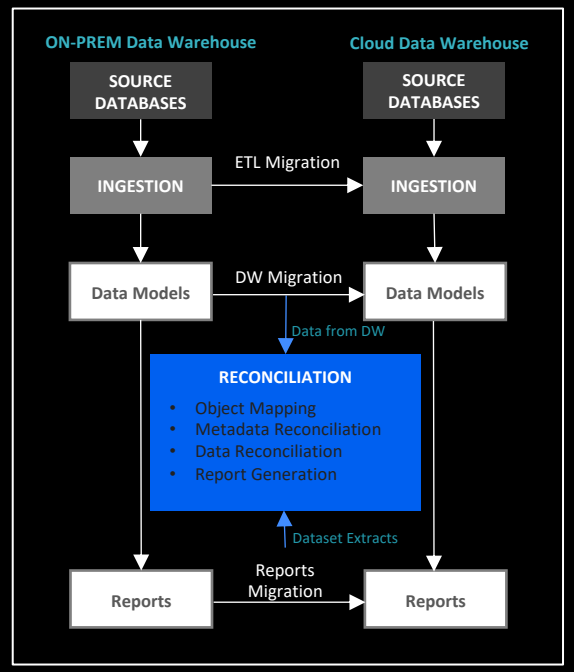
Empowered by migVisor Reconciler

WHY

- Inconsistency in scheme and database after the migration from legacy DWH
- Discrepancy in audit reports between could 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

snowflake, ORACLE, teradata., Amazon S3, IBM DB2, Azure Synapse Analytics, PostgreSQL, Amazon Athena, SAP S/4 HANA, Databricks

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

EPAM Data Landing Zone Infrastructure

Empowered by EPAM Data Landing Zone

WHY

| Greenfield start challenge

| Time and resource-consuming development of infrastructure deployment

| Legacy processes slow down modern technologies adoption and automation

WHAT

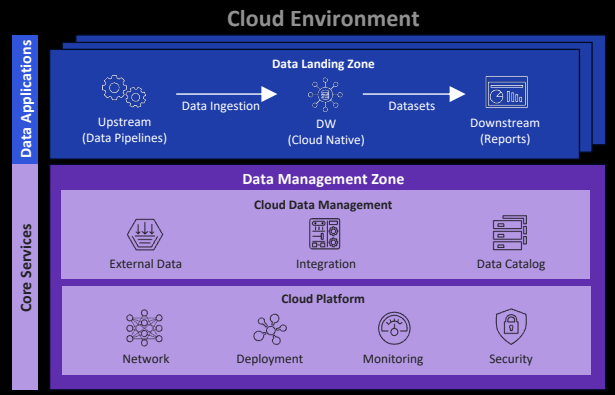
KEY CAPABILITIES

- 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

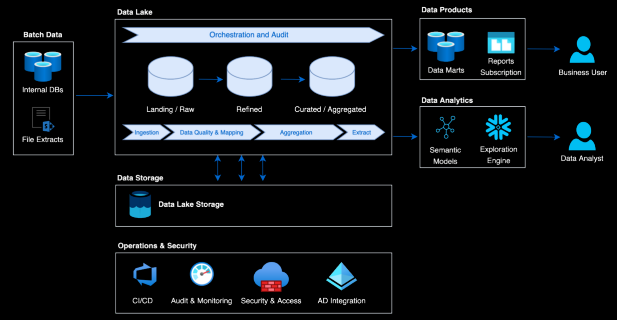
HOW

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

Analytics Migration Assessment

Empowered by migVisor Analytics

WHY

| Complex, sophisticated and long-time process of DWH, Hadoop, ETL and Reports Assessment

| Extremely time-consuming process of manual migration complexity evaluation

| Overoptimistic estimation due to lack of visibility in legacy systems complexity

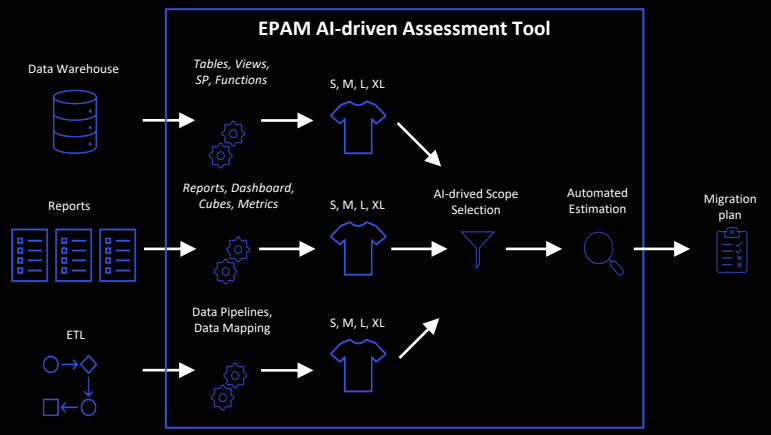
WHAT

KEY CAPABILITIES

- Speed-up process of DWH ETL and Reports assessment
- Automatically scan metadata in DWHs, Reports and ETL pipelines
- Analyze complexity of created inventory and range by S,M,L,XL
- Apply AI algorithms to select inventory scope for the Migration and divide Workloads by phases
- Create Migration Roadmap, including timeline, team composition, delivery milestones
- Provide reports with assessment analyses and deliver Migration plan

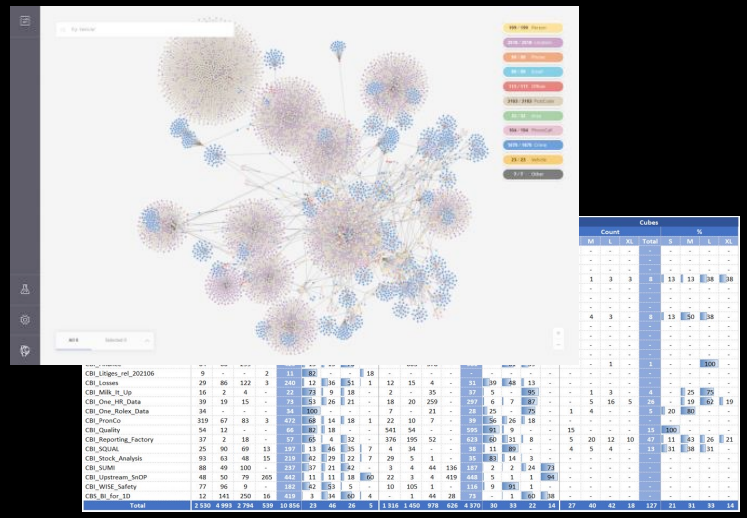
HOW

HOW TO USE IT ON A PROJECT



1. Inventory with initial profiling (data sources, cubes, reports and pipelines)
2. Complexity criteria set up
3. T-shirt sizing based on defined criteria
4. Down selection of representative assets
5. Detailed assessment, complexity estimation
6. Roadmap and timelines

AI ADVANCED ANALYTICS



KEY DIFFERENTIATORS

| Automated process of DWH ETL and Reports assessment

| Automated migration complexity evaluation

| AI-driven Migration Roadmap generation

EPAM's Workload Migration

Empowered by migVisor Converter

WHY

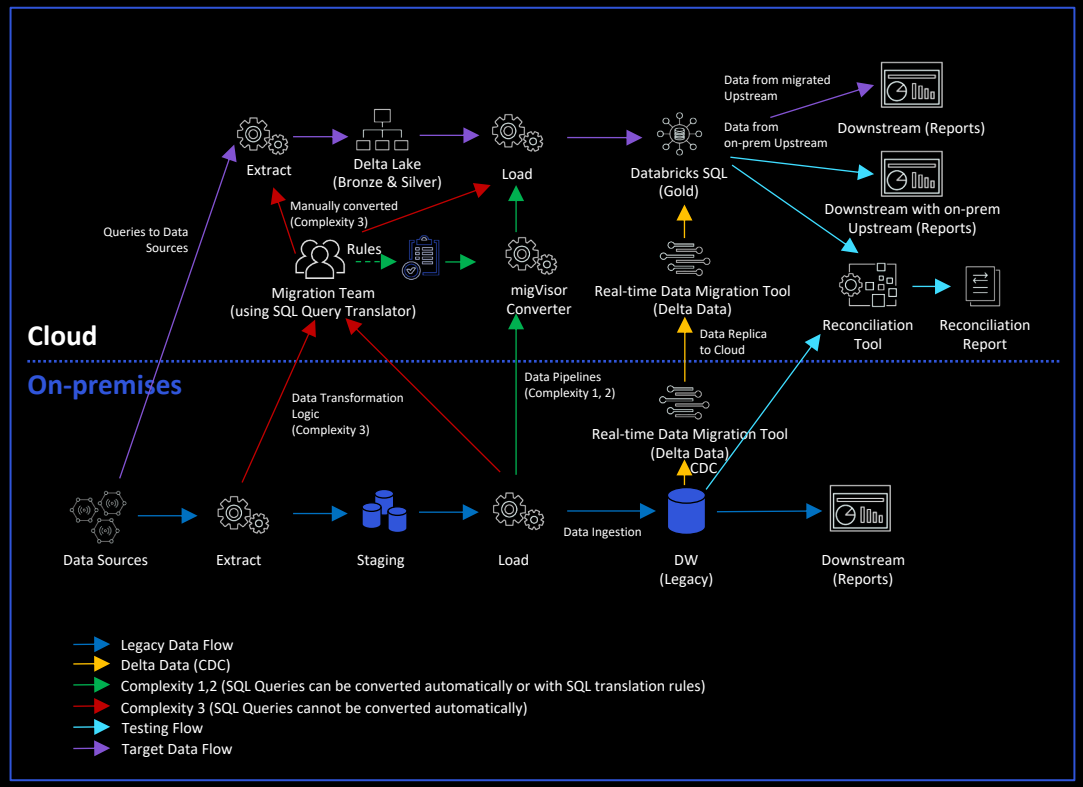
- High-complexity of legacy ETL data transformation pipelines in low-code tools
- Extremely time-consuming process of manual SQL and ETL conversion to pySpark
- High-complex process of data quality testing and reconciliation

WHAT

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

HOW

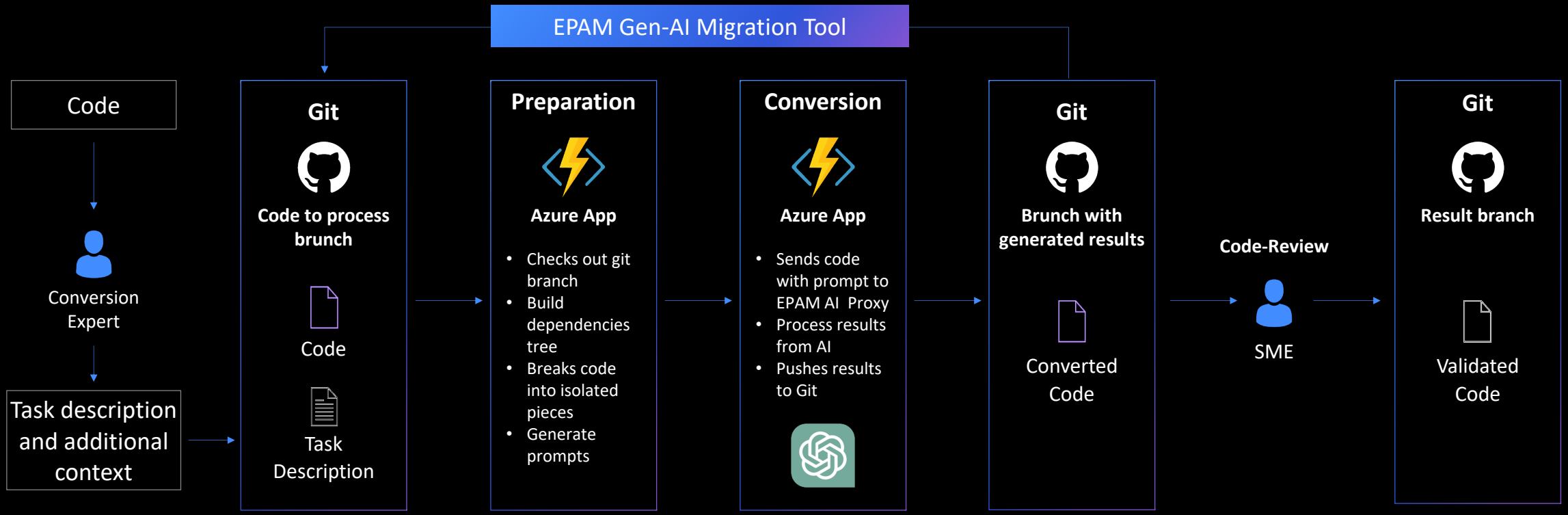


1. **Set up an environment** for ETL conversion from legacy low-code to pySpark
2. **Converter Reader Configuration** configures as per source metadata (up to 20 days depending on source)
3. **Writer Configuration** uses appropriate configuration as per target platform
4. **Convert & Iterate** perform 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.

migVisor Streaming

Empowered by migVisor Streaming

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

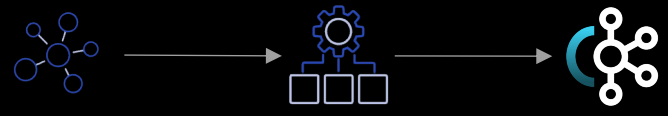
WHAT

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-

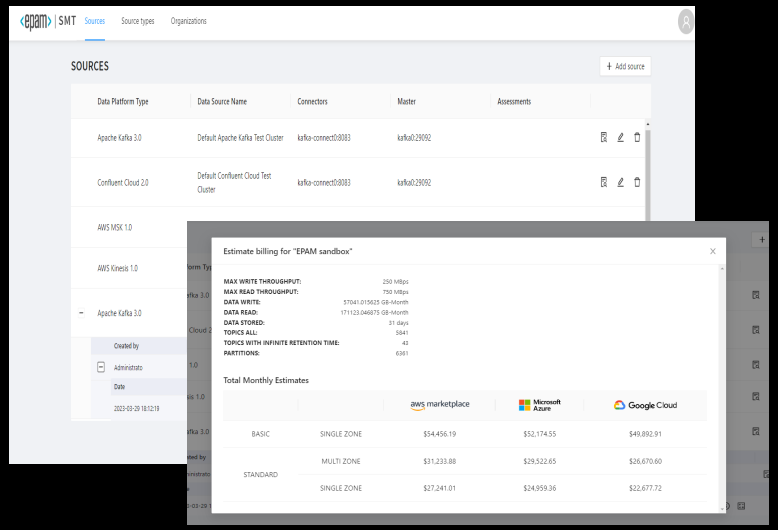
HOW

HOW TO USE IT ON A PROJECT



1. Deploy and configure the SMT Tool
2. Create the current Configuration Assessment Report by a click of a button
3. Create the Migration Plan
4. Create Estimation Report based on the Migration Plan
5. Invoke the Migration plan and create the Target Cloud Streaming Structure
6. Migrate Data

migVisor Streaming – Confluent Edition



KEY DIFFERENTIATORS

| Intuitive Migration Assessment & Planning UI Tools

| Ease of Deploy and Use

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

migVisor Transactional

Empowered by migVisor Transactional

WHY

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

WHAT

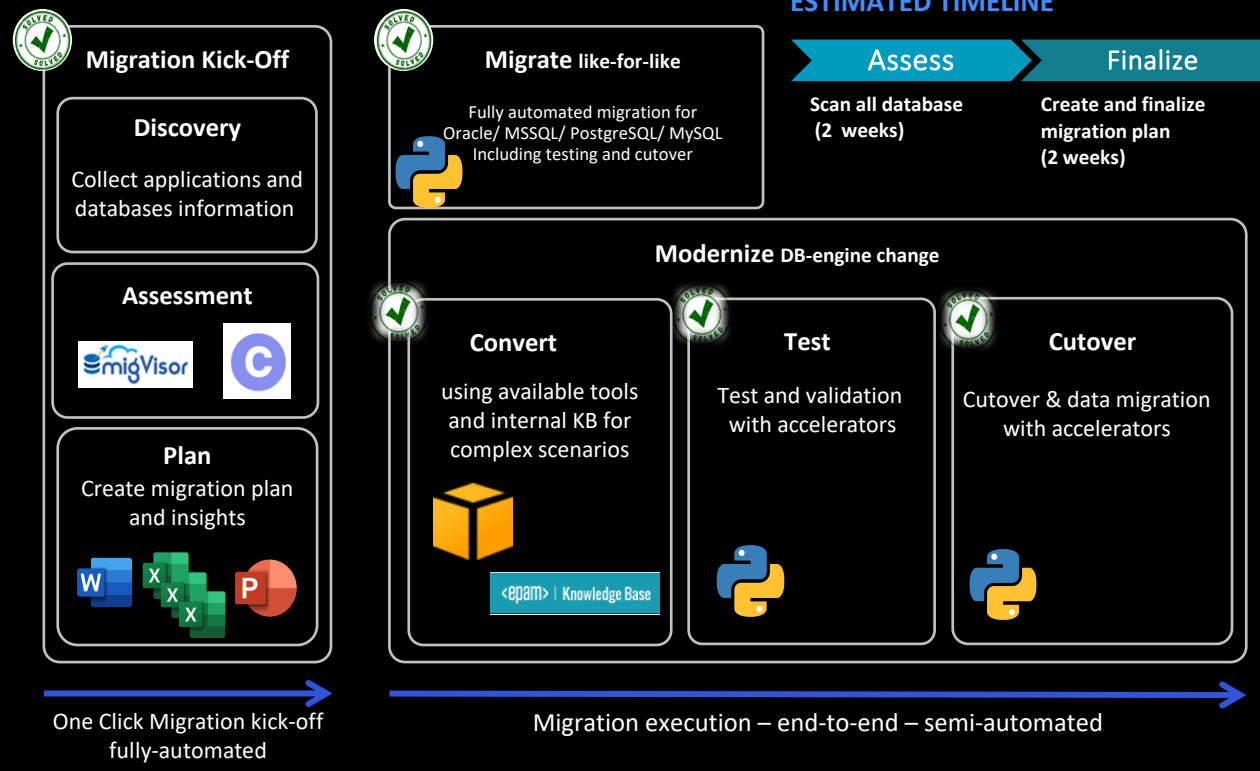
KEY CAPABILITIES

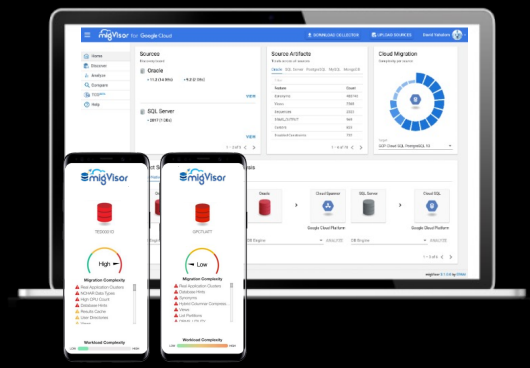
- Discover and assess your database fleet and app source code using migVisor by EPAM
- Analyze finding to create dependencies mapping and initial estimates
- Create a draft migration plan including TCO
- Confirm the plan and adjust with relevant stakeholders and non-functional requirements and constraints

Outcome:

EXECUTION-READY MIGRATION PLAN AND PROJECT JUSTIFY WITH ROI

HOW





THE migVisor WAY – IMMEDIATE ROI

Automated assessment with TCO

Better decisions

Better migrations

migVisor

KEY DIFFERENTIATORS

Highly-automated approach:

Quick and accurate migration plan for your databases

Widely credible migration experience:

Google selected product for database assessments

Thank you!

For more information, contact

Vitalii Bondarenko

Principal, Data & Analytics Consulting

Vitalii_Bondarenko2@epam.com