Cloud Data Migration
Smart Assessment

Data Platform Assessment. Service Offering

2023

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

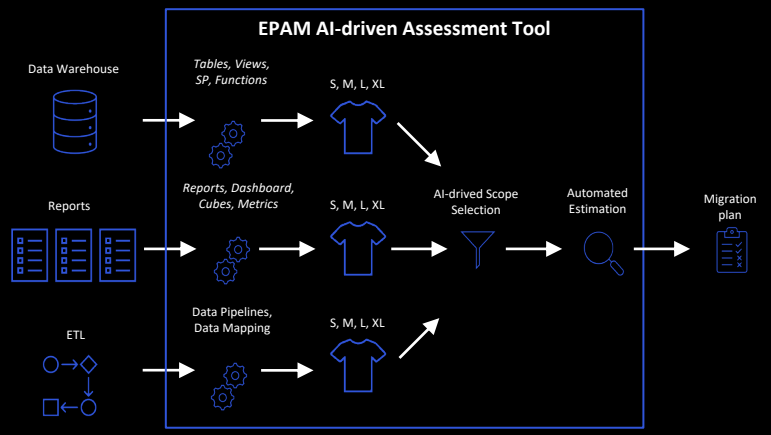
WHAT

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

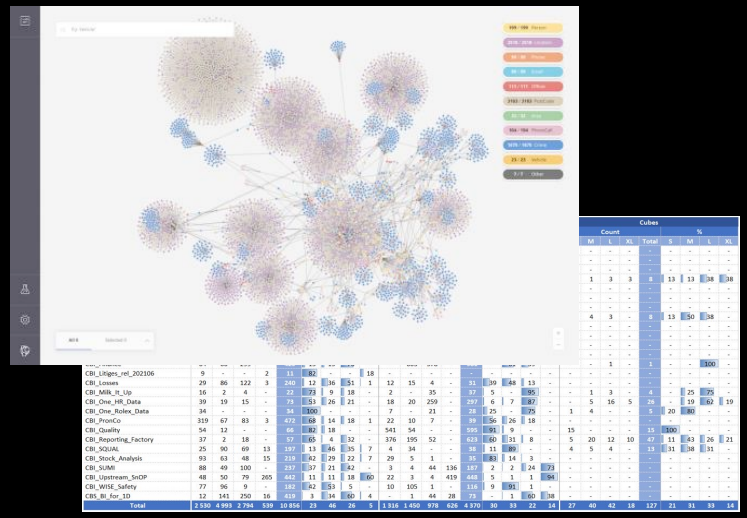
HOW

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

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 ASSESSMENT

Empowered by EPAM AI Toolset & migVisor Platform

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

Smart Assessment as a part of 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

Smart Assessment

migVisor transactional | migVisor analytics | migVisor streaming CE

migVisor landing zone | migVisor code converter

migVisor reconciler

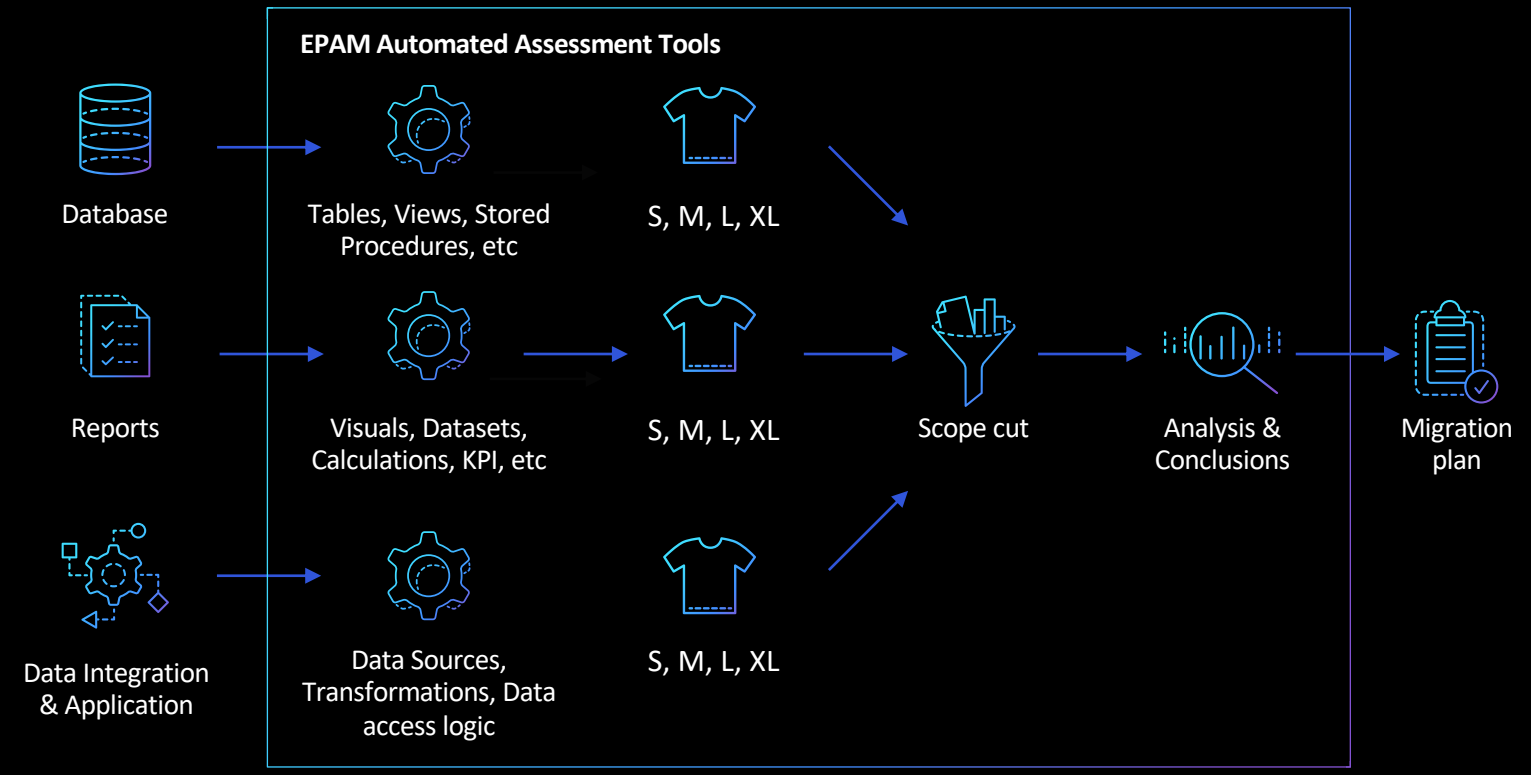
Analytics Assessment

MVP and Execution

Testing & Reconciliation

Data Migration Framework| 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

WHY| Automated Assessment Details

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 Offer

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

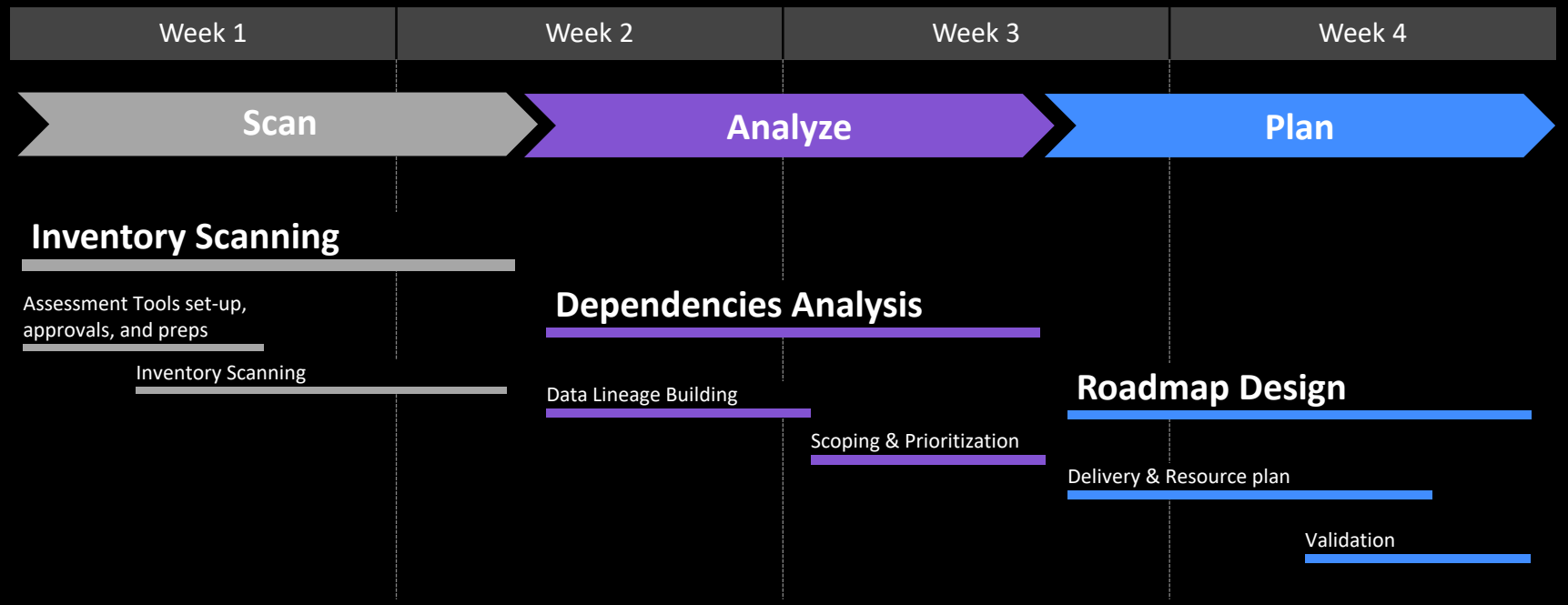
HOW| Smart Assessment Timeline

Total duration for Smart Assessment is **4 weeks** which includes Learn (Scanning) phase and Analyze phases.

All accesses to DBs, ETL pipelines and Reports need to be provided on the first day of the assessment.

Timeline doesn't include any code changes to EPAM Migration Assessment tool and assumes that scanned systems are already supported.

EPAM doesn't charge any licenses and doesn't provide the assessment tool. The tool is used by the EPAM's team only to accelerate the Smart Assessment process.



Deliverables

- Migration Scope & Strategy, incl. priorities and cost calculation
- Migration Solution, incl. target solution architecture, migration & testing approach, tools and accelerators
- Dependency analysis and scope reduction recommendations

Tools Setup

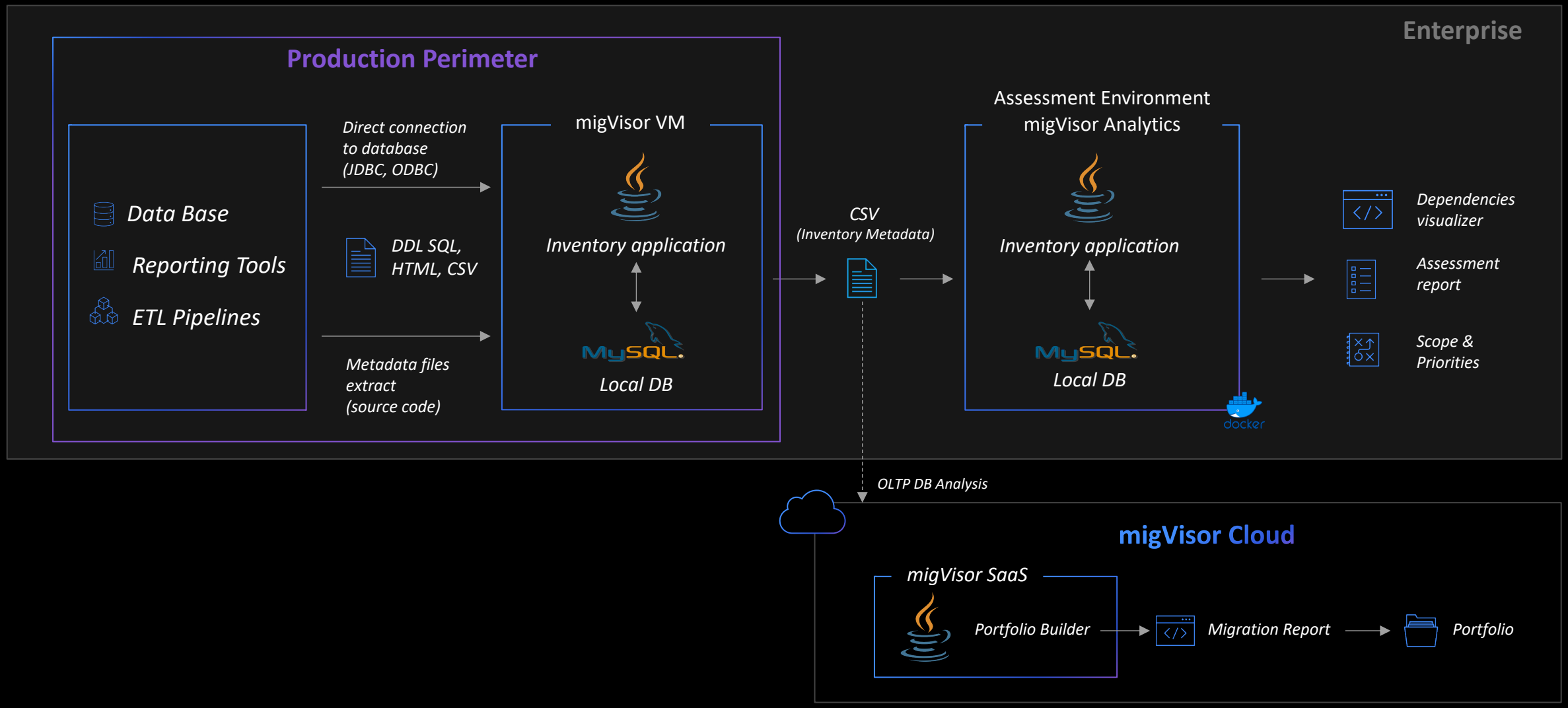
- Tools set-up should be finished during the first week of the assessment and includes deployment of a Docker container into PROD perimeter
- Read only JDBC/ODBC connections permissions to assessed DBs, extracts from Reports and ETL Source Code need to be granted
- List of DBs, ETL Instances and Reporting Tools Instances need to be provided in the first day of the assessment

HOW| Cloud Migration assessment process

ASMT Automated Assessment Tool is applied

	Presales & Preparation	Scan	Analyze	Plan & Validate (Pilot)
ACTIONS	**CLIENT ENGAGEMENT** (series of calls): • Set-up call • Present Migration service • Establish trust relationship • Collect information (identify Scope, Stakeholders and Timeline) **SCOPE DEFINITION** (series of calls) • Send Inventory Questionnaire • Set-up call and clarify answers • Identify milestones, intermediate deliverables and priorities **PROPOSAL PREPARATION** (presale team collaboration): • Prepare Proposal and SOW • Present the Proposal and clarify scope, timeline and team composition • Initiate staffing (STAFFING PLAN) • Define key stakeholders **BEFORE WE START** • Prepare and send documents request • Request necessary access rights • Prepare draft of Kick-off presentation • Prepare draft of Communication plan with meeting schedule	**WORKSHOP** (series of workshops with client) • Define scope of migration with business, incl. deadlines and business priorities • Discover existing infrastructure and define as-is architecture blueprint, incl. data flows, security compliance, performance issues and quality attributes **ASSESSMENT** (automatic inventory scanning and analyses): ASMT • Scan/ analyze inventory for all systems (DWH, Upstream, Downstream, Reporting, ML) • Provide complexity and migration estimates for scanned inventory, calibrate estimations by manual assessment	**MIGRATION DESIGN** (discovery team collaboration and presentations to client): ASMT • Define data lineage and dependencies • Design target architecture, tech stack and security approach • Execute AI driven analysis, build dependency matrix • Execute scope reduction, analyze usage statistics, eliminate duplicates and similar objects • Estimate migration efforts • Presentations to client and validation	**PILOT DELIVERY** • Set-up Data Landing Zone • Configure migration tools • Migrate schema • Migrate historical data • Migrate Data pipelines • Migrate reports • Reconciliate Data **CODE CONVERSION POC** • ETL/ELT conversion **MIGRATION PLAN** • Create Migration roadmap, define milestones, deliverables, testing and cutover timelines • Define delivery waves and team composition for each stream • Split scope by migration streams, according to priorities and expected business value • Develop Migration roadmap for Execution phase • Define resource plan
DELIVERABLES	• Short description of client needs • Questionnaire template • SOW • Proposal on Learn, Plan, Validate • Team set-up on Learn, Plan, Validate • Inventory description according to the questionnaire and conversations (to Proposal preparation) • Documents request • Access rights request • Draft of Kick-off presentation • Draft of Communication plan with meeting schedule	• Scope of migration with priorities • Inventory with complexity and migration estimates • As-is Architecture Blueprint	• Migration Scope & Strategy, incl. priorities and cost calculation • Migration Solution, incl. target solution architecture, migration & testing approach, tools and accelerators	• Migration Roadmap, incl. streams, delivery milestones, team composition, data governance approach • Validated and adjusted Migration Roadmap (incl. streams, delivery milestones, team composition, data governance approach) • Validated and adjusted resource plan for Execution phase

Assessment Tool Setup



DW Assessment Supported Platforms

DATA WAREHOUSE & DATABASE

IBM DB2
amazon EMR
Microsoft SQL Server
snowflake
HIVE
databricks
ORACLE
Amazon Redshift
teradata.

REPORTS

MicroStrategy
Microsoft SQL Server Reporting Services
Power BI

ETL

Informatica
Microsoft SQL Server Integration Services
talend

TARGET PLATFORM / CLOUD

Google Cloud
databricks
snowflake
aws

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

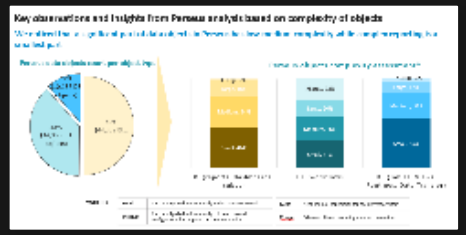
Technologies assessed

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

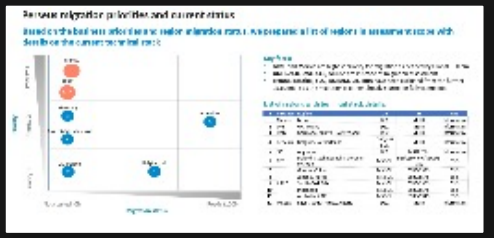
Solution Highlights

- 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**

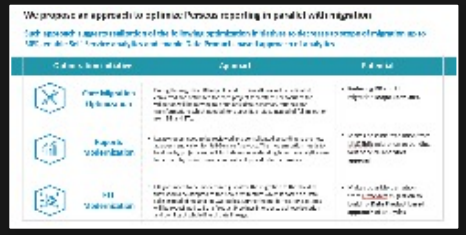
Key insights from analytical tools: usage patterns, etc.



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.

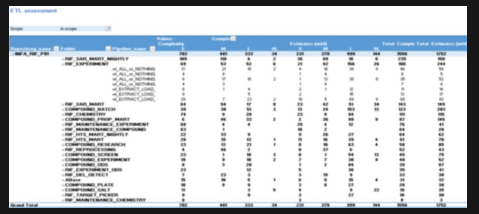
EPAM utilized an automated tool powered by OpenAI's Language Model (LLM), which converted **90%** of low to medium complexity code, and **50%** of highly complex code, streamlining the migration process.

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.

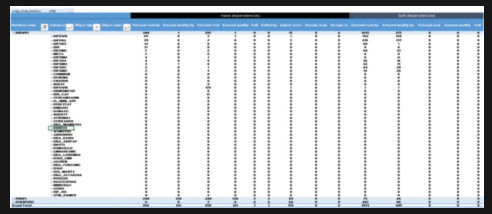
Case Study:
Assessment of Client's RIF Platform

- Pfizer was looking to migrate the Research Information Factory (RIF) Platform's data to its R&D Data Lake platform to reduce storage costs, enable cloud-based processing and enhance flexibility
- After initial Discovery using migVisor Analytics tool, we were engaged in assessment of the RIF Platform migration complexity and migration efforts estimates
- **Assessment scope:** 3 Oracle databases of ~6 Tb and ~1500 Informatica ETL jobs over 6 weeks
- **Key deliverables:**
  - → 3 Workshops, 2 interviews
  - → Complexities & Estimates
  - → Out of scope recommendations (backups, duplicates, similar SQL code)
  - → Usage statistics analysis
  - → Lineage insights (lineage data, Informatica mappings, hard and soft dependencies, linked server dependencies and stream jobs, orphans)
  - → Scope prioritization across datamarts

- Assessment report in excel
- Target architecture design calibration
- RIF Modernization & Strategy deck

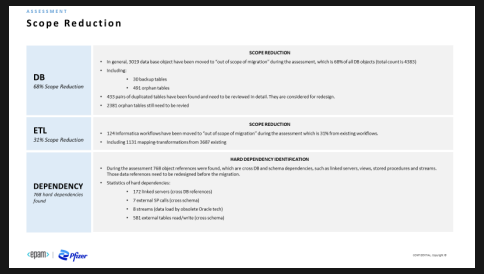


Assessment DB & ETL

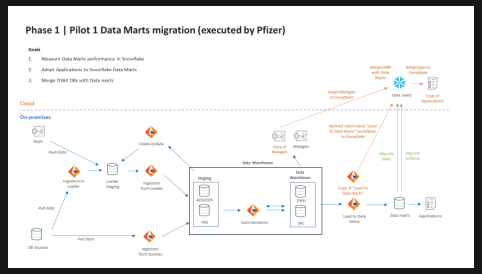


Lineage Insights

Prioritization & Estimates



Scope Reduction

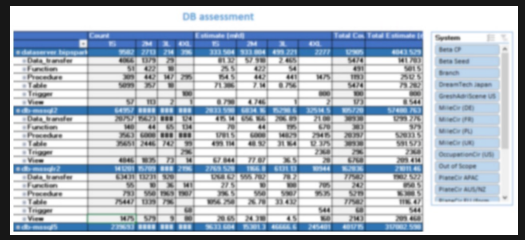


Migration Plan

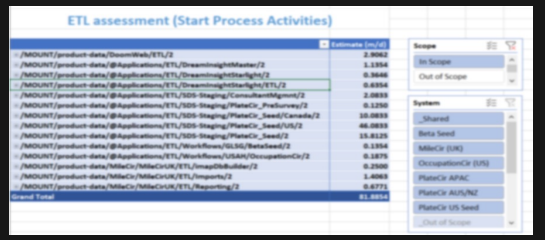
Case Study:
DB and ETL Assessment

- Client approached EPAM MigVisor Analytics team to perform an assessment of its DB and ETL to propose migration scope and estimation with a focus on business value
- MigVisor Analytics team empowered existing discovery team with capabilities of the automated data lineage analysis
- Delivered dependency analysis for a customized ETL application, which was impossible during the manual discovery
- **Assessment Scope:** 4 SQL Server 15, 1 SQL Server 12 (Spark), PostgreSQL (Branch) and ETL
- **Key Deliverables:**
  - → Assessment report in excel
  - → DB statistics with numbers of tables, data transfers, views, etc.
  - → ETL Analysis with detalization by mpd, xaml, stored procedures etc.
  - → ETL statistics
  - → Defined scope and estimates

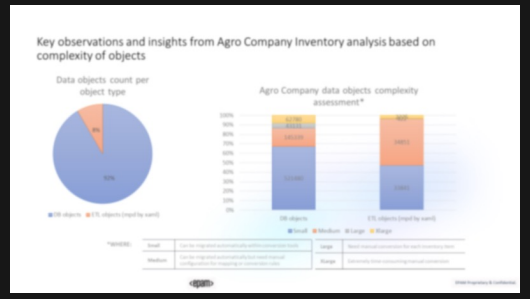
- In a week automatically generated estimates of migration from legacy Data Platform to modern Databricks Cloud Data Platform
- Performed an assessment of 6 instances and ETL in 3 weeks
- Reduced time for discovery process (3x speed up)



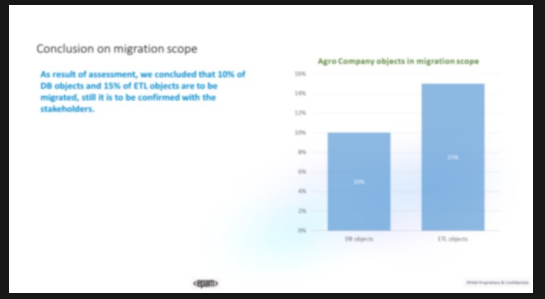
Assessment DB



Assessment ETL



Assessment insights



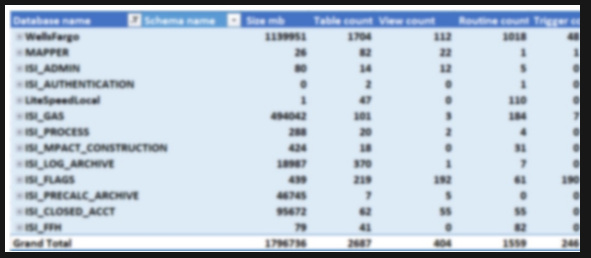
Migration scope summary

Case Study:
Assessment of Client's DB for Due Diligence

- Client approached EPAM MigVisor Analytics team to perform an assessment of its MS SQL databases as part of the larger technical Due Diligence aimed to assess the company's portfolio performance solutions - evaluate current state and define an approach to modernization
- **Assessment Scope:** 13 MS SQL DBs, ~3500 DB objects, ~1.8 Tb of data
- **Key Deliverables:**
  - → Assessment report in excel
  - → Analysis & complexities
  - → Data Size insights
  - → Lineage insights
  - → Out of scope recommendations (e.g. backups)
  - → Migration scoping & estimates

- In conclusion we discovered the extreme complexity and high dependency of legacy data platform
- Lift & Shift and Replatforming migration strategies were defined as unreasonable due to the high cost & long duration
- Assessment team proposed modernization strategy with timeline and estimates

Assessment DB



Assessment of Data Size

Assessment insights in Discovery deck

Thank you!

For more information, contact

Vitalii Bondarenko

Director, Data & Analytics Consulting

Vitalii_Bondarenko2@epam.com