PM3 Migration method and tools required for it

Contents

Purpose of the document

This document defines the assumptions on the basis of which the data migration in the FALCON project will be carried out.

The scope of this document includes:

- Presentation of a description of how to perform data migration.
- Presentation of tools used in migration
- Presentation of the expected format of input data for the migration process.
- Ensuring that the migration process will be carried out in a safe, efficient manner and in accordance with business and system requirements.
- Agreement of assumptions for migration by all parties to the project.

Data Migration Assumptions

Migration Goals

1. Supplying the WF systems of the new Falcon solution with the necessary data from the bank's systems to enable the correct operation of the solution.
2. Enabling use of the new solution in the planned functional scope from the production launch.
3. Improving data quality - as part of preparing data for migration, data can be verified, supplemented and corrected in various business and technical areas (depending on requirements and arrangements).

General assumptions

1. Data migration as one of the elements of the new solution implementation project, plays a dependent role in relation to other project streams.
2. The data migration stream can determine and plan data flows only within the scope of objects / entities occurring in the Falcon project and required by the target systems.
3. Within the data migration stream, the issues of functional gaps between the source systems and the target solution will not be resolved, while the identified discrepancies will be forwarded for analysis.
4. In the case of the production initial data migration process and general migration tests (Dress Rehearsal), the full volume of production data will be used.

5. In the case of the production incremental data migration process, the loading method (incremental / full loading) will be determined at the stage of detailed analysis of the implementation stage for each entity and target system.
6. For test environments (INT, UAT, QUA), the scope of data (including their volume) and the degree of anonymization will be determined in the migration project.
7. The Bank is responsible for the design and execution of data extraction processes from source systems.
8. The supplier is responsible for preparing migration processes and preparing the transformation of source data transferred for migration by the Bank to the migration model and then to the target model appropriate for Workflow system.
9. The bank is responsible for anonymizing data transferred for use in the migration process (development, tests).
10. A detailed description of the data required by the target systems, which must be made available by the source systems, will be developed together with the Bank and the supplier as part of the detailed analysis of the project implementation stage.
11. The migration will be carried out in the following manner:
    a. safe - The Bank will provide safe infrastructure and safe data transfer channels,
    b. effective - The supplier will conduct migration efficiency tests and ensure the optimal way of building data loading processes into source systems,
    c. compliant with the Bank's IT security requirements,
    d. compliant with business requirements - the Bank and the Supplier will determine - during the detailed analysis of the implementation stage - such scope of data, along with mappings and necessary transformations, which, when implemented by the Supplier, will meet the business requirements set for the target systems.
12. Migration tests will be performed accoring to the principles describe in the document.

Beyond the scope of data migration

- Planning tasks related to archiving data from source systems and shutting down source systems.

Approach to unmigrated data

The discovery phase of the Falcon project does not include an analysis of how to decommission the Bank's current systems.

We assume that data that has not been indicated for migration will remain available for migration if such a need arises as a result of, for example, changes in business requirements or the method of implementation during the implementation project.

Managing the data migration process

The issue of Release Management goes beyond the scope of this document and will be included in the Project Plan.

Source data to be migrated

Source data preparation

1. During the process of data extraction and preparation for migration from source systems, the Bank is responsible for:
    a. data selection
    b. data validation in terms of its correctness in relation to business requirements, in particular:
        i. supplementing missing field values that are required in target systems for a given entity
    c. data validation in terms of data accuracy, i.e. the data is consistent with the actual values
    d. data validation in terms of consistency, including:
        i. referential data integrity (foreign keys indicate the correct / existing records)
        ii. lack of contradictions in related values
        iii. data coherence / reconciliation between different source systems (e.g. which customer email address is primary)
    e. data validation in technical terms, including:
        i. data compliance with the data type
        ii. data compliance with the declared schema
        iii. key integrity (dictionaries, relations between objects)
        iv. field formatting correctness (e.g. email, phone)
        v. code consistency
        vi. size and length fields
    f. repair or removal of errors noticed during the above-mentioned data validation and verification processes
    g. checking the currency of data
    h. checking the uniqueness of data and its deduplication

2. The Bank is responsible for the proper anonymization of data - to the appropriate extent appropriate for the phase and environment (e.g. DEV, INT, UAT, QA)
3. The Bank is responsible for approving the prepared data set for migration, before forwarding it to the supplier for further processing.
4. The Bank is responsible for forwarding the data for migration in the agreed format by saving directly to the intermediate database (to the established scheme)
5. The supplier will verify the consistency of the transferred data and report detected errors
6. The Bank is responsible for analyzing and possibly repairing problems with data extracts reported by the supplier

Assumptions regarding direct transfer of data to the intermediate migration database

1. Data for migration will be transferred by the Bank via direct loading into intermediate database tables **(any other method of transferring database-to-database data to be determined during Discovery workshops, e.g. *.bacpac, *.bac)**
2. The target structures of the intermediate database for data intended for migration will be determined during the detailed analysis of the implementation stage.
3. Data transferred for migration should be: clear, uniform, consistent, matched in structure and should be free of duplicates.
4. Data transferred for migration should contain completed attributes that are not in the source systems but are required by the target data model.
5. The transferred data will contain foreign keys, enabling an unambiguous relational connection with the corresponding primary keys. Data and tables referenced by foreign keys will be transferred together with the referencing data.
6. For security reasons, access to data stored in the intermediate database will be limited to designated persons.

Scope and volume of data to be migrated

The scope and volume of data to be migrated is specified in PM2 Scope of data to be migrated

Detailed arrangements for data flows, transformations and quality control

Detailed arrangements (data flows, necessary transformations and data quality control and processing) will be the result of joint work of the Bank and the Supplier in the implementation project. At that time, the locations for storing extracts will be determined (including details regarding the names of folders and files or table schemas in the intermediate database). The implementation project will define detailed rules for mapping fields and entities and rules for verifying the technical quality of data transferred for migration before the actual process of loading them into target systems. It should be assumed that as the project progresses, data flows may change or be supplemented.

Data flows will be prepared divided into systems, within the implementation stages:

- source systems → data extraction (CSV / intermediate SQL database) [Bank responsible]
- data extraction (CSV / intermediate SQL database) → migration model -> target systems [Supplier responsible]

Data security and archiving

We assume that operations on migrated data (files) will be compliant with security guidelines, with particular emphasis on:

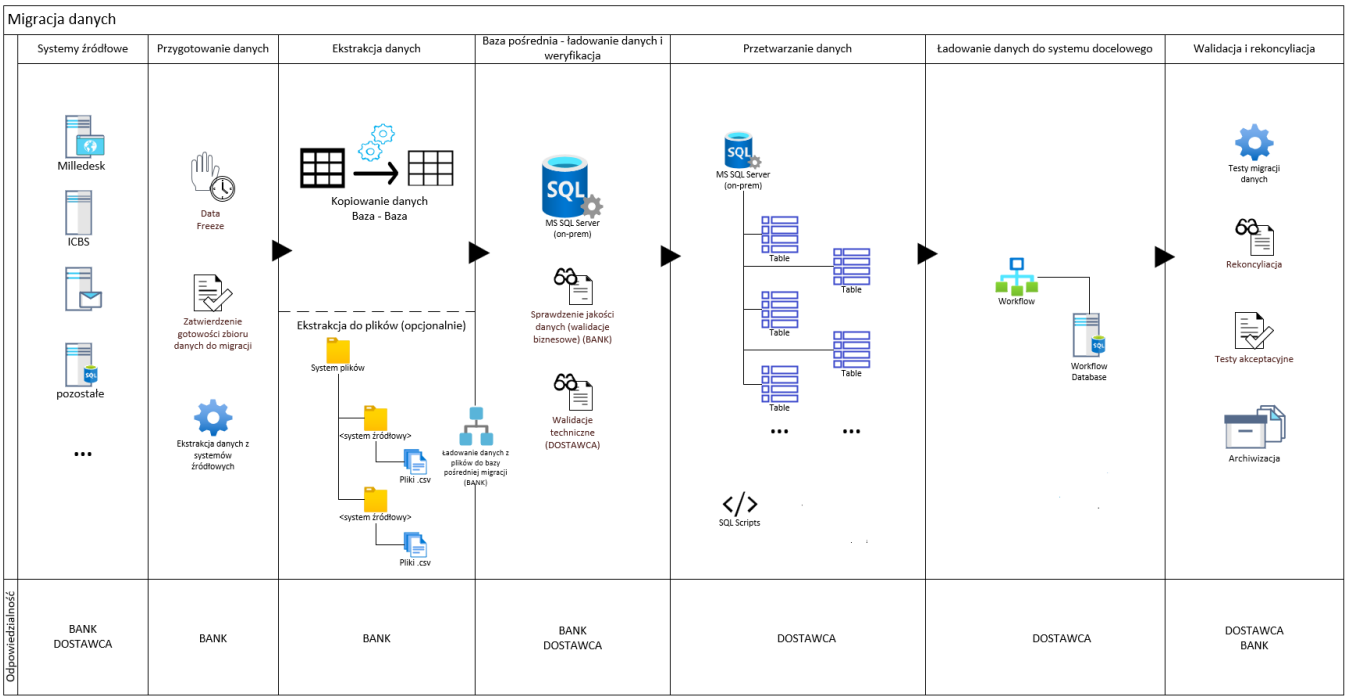
1. Data transferred between systems must be sent using encrypted connections (e.g. HTTPS, TLS), access to data should be limited to the necessary minimum.
2. Data is encrypted at rest and during transmission.
3. Access to data (Storage Account, SQL Server, etc.) is only possible via the bank's network (VPN).
4. Access to the Storage Account and the intermediate migration database should be limited to the necessary minimum - application of the least privilege principle to users and applications.
5. All operations on the Storage Account, SQL Server, Databricks should be logged - tools to be indicated by the Bank.

Data Migration Solution Architecture

For data migration new on-premise intermediate database will be setup and configured by bank.

All data prepared for migration will be copied using DB link. Systems that do not use MS SQL data will be extracted to CSV files and than loaded to intermediate database (bank)

SQL scripts or procedures or dedicated software will be used for transformation to target schema and migration to target system



The most important components presented in the diagram are:

1. File system - a place on the Bank's on-premise cluster to which data for migration will be exported from systems with which a direct database-to-database connection cannot be established (ICBS)
2. On-premise SQL Server database
    a. Target database for direct data extraction from source systems
    b. Storing and preparing data for migration
    c. Provides all MS SQL Server capabilities
    d. Database server / dedicated database for migration set up as part of the Bank's on-premise infrastructure
3. SQL scripts or dedicated program that will perform the following ETL tasks
    a. input data validation
    b. input data transformation
    c. loading data to the Workflow database
    d. verification of data consistency
    e. verification of completness of migration

Environments in the context of data migration

The migration will apply to all environments INT, UAT, QUA, PROD.

Due to the availability of the production environment for the main data migration before phase 1, the DMTEST environment is not needed.

Migration tests

Details of migration tests, verification of migrated data and assumptions regarding data reconciliation will be developed in chapter ??PM4 Propozycja faz, zakresu testów migracyjnych oraz sposobu przepięcia na nowe narzędzie.

Incremental migrations - high-level method of implementation

The exact method of implementation of incremental migration and its assumptions will be described based on the findings of the detailed migration analysis carried out at the implementation stage.

In general, incremental migration will be carried out on objects already migrated using the selected incremental key.

Incremental keys and tags allowing for precise determination of which data should be covered by incremental migration will be determined during the above detailed migration analysis and during planning of the implementation phases of the

The incremental migration will be launched in the full migration launch regime.

The ETL process should support the following scenarios (to be specified during the detailed analysis):

- Reading data from the intermediate database and reading data from the target system
- Inserting a new record into an existing entity in the target system
- Updating a record existing in the target system
- Deleting a record in the source system and transferring this information to the target system

ETL processes should be prepared in such a way that they can report how many records have been added, modified or deleted as a result of the incremental migration.

ETL processes should store information about the date/time of the last migration (in the intermediate database, in the metadata area) - this value will be used as a reference point in subsequent migrations.

The source of data for incremental migrations is the intermediate database - just like in the case of full migration.

ETL processes should enable logging of the incremental migration process and progress.

Withdrawal of migrated data from target systems

Detailed mechanisms for withdrawing migrated data packages from target systems will be developed at the implementation stage.

Migration process reporting

In the case of short migration processes (up to 6 hours), reporting will take place after the entire migration process has been completed.

In the case of long migration processes (over 6 hours), the migration process will be monitored and reported daily (possibly subject to agreement at the detailed analysis stage of the implementation phase).

The report will include, among others:

- date and time of migration start and end
- person responsible for accepting the migration process (optional, required for the PROD environment)
- person responsible for launching the migration process (optional, required for the PROD environment)
- environment
- type of migration
- information on the migration status (e.g.: planned, in progress, suspended, successfully completed, completed with errors, failure, canceled, migration verification)
- number of entities and records accepted for migration
- number of records tested (technical validation)
    - number of records accepted
    - number of records rejected
- number of records migrated
- number of records not migrated (and reason - may require additional analysis)
- results of automatic reconciliation tests
- records with errors along with indication of the type of error

Open Points

Necessity to use a dedicated environment for DMTEST data migration.

Acceptances

(acceptance by insertion ✅ by the name of the accepting person)

1. DevOps
    a. Grzegorz Majszyk ✅
2. Application teams
    a. Milledesk Corporate: Janusz Norbert Cichocki ✅
    b. Workflow Corporate:Agnieszka JAnczulewicz-Pralicz ✅
    c. ICBS
3. Architecture team Tomasz Kochanek ✅