

Data Migration Strategy

Taxi & Private hire Service System

Version 1.0

8thJuly, 2021

Confidentiality Statement

**This document contains information that is proprietary and confidential to Tata Consultancy Services Limited, which shall not be disclosed outside Transport for London, transmitted, or duplicated, used in whole or in part for any purpose other than its intended purpose. Any use or disclosure in whole or in part of this information without express written permission of Tata Consultancy Services is prohibited. Any other company and product names mentioned are used for identification purposes only and may be trademarks of their respective owners.**

**© Copyright <2021>, Tata Consultancy Services (TCS)**

TATA Code of Conduct

**We, in our dealings, are self-regulated by a code of conduct as enshrined in the ‘TATA Code of Conduct’. We request your support in helping us adhere to the code in letter and spirit. We request that any violation or potential violation of the code by any person be promptly brought to the notice of the Local Ethics Counsellor or the Principal Ethics Counsellors or the CEO of TCS. All communication received in this regard will be treated and kept as confidential.**

**Table 1: DOCUMENT REVISION LIST**

|  |  |  |  |
| --- | --- | --- | --- |
| Revision No. | Revision Date | Author | Revision Description |
| V1.0 | 22-Jun-2021 | Pranav Gandhi, Kunal Singh | This document provides on Data Migration Strategy. |
| V1.1 | 02-Jul-2021 | Pranav Gandhi, Chirag Khajuria, Kunal Singh | Changes have been incorporated as per review comments received from TfL. |
| V1.2 | 08-Jul-2021 | Pranav Gandhi, Chirag Khajuria, Kunal Singh | Changes have been incorporated as per review comments received from TfL. |

Contents

[1 Introduction 8](#_Toc76629639)

[1.1 Purpose 8](#_Toc76629640)

[1.2 List of Abbreviations 8](#_Toc76629641)

[1.3 Audience 8](#_Toc76629642)

[2 Data Migration Strategy/Approach 10](#_Toc76629643)

[2.1 Scope 11](#_Toc76629644)

[2.1.1 Inclusions 11](#_Toc76629645)

[2.1.2 Exclusions 11](#_Toc76629646)

[2.1.3 Pre-requisites 12](#_Toc76629647)

[2.2 Data Migration Approach 12](#_Toc76629648)

[2.2.1 Requirement of Source System Data and Details of Target Database 14](#_Toc76629649)

[2.2.2 Data Profiling & Cleansing Methodology 16](#_Toc76629650)

[2.2.3 Data Mapping 17](#_Toc76629651)

[2.2.4 Data Transformation 17](#_Toc76629652)

[2.2.5 Approach to migrate In-flight Data from existing Service System 17](#_Toc76629653)

[2.2.6 Data Validation & Reconciliation 18](#_Toc76629654)

[2.2.7 Data Security 18](#_Toc76629655)

[3 Data Migration Plan 20](#_Toc76629656)

[4 Data Migration Testing 22](#_Toc76629657)

[5 Go Live/Cutover 23](#_Toc76629658)

[5.1 Checkpoint Based Approach for Go Live/Cutover 23](#_Toc76629659)

[5.2 Early Booking Live (Vehicle Inspection) 26](#_Toc76629660)

[6 Risk & Mitigation Plan 28](#_Toc76629661)

[7 Responsibility Matrix & Dependencies 31](#_Toc76629662)

[8 Software, Tools and Services 33](#_Toc76629663)

Table Of Figures

[Figure 1: Data Migration Activity 10](#_Toc76629664)

[Figure 2: Data Migration Approach 12](#_Toc76629665)

[Figure 3: Data Migration Plan 20](#_Toc76629666)

[Figure 4: Phased Migration Approach 23](#_Toc76629667)

Table Of Tables

[Table 1: DOCUMENT REVISION LIST 3](#_Toc76629668)

[Table 2: List of Abbreviations 8](#_Toc76629669)

[Table 3: Data Extract Schedule 14](#_Toc76629670)

[Table 4: Data Migration activity in each phase 20](#_Toc76629671)

[Table 5: Identified Risks and Mitigations 28](#_Toc76629672)

[Table 6: Responsibility Matrix 31](#_Toc76629673)

# Introduction

## Purpose

The objective of this document is to propose the Data Migration Strategy. This approach will be further refined in subsequent phases after an analysis of the data structures and data specifications of existing TOLA system. This document covers the following:

* Data Migration Strategy/ Approach
* Data Migration Plan
* Data Migration Design
* Data Migration Testing

## List of Abbreviations

**Table 2: List of Abbreviations**

|  |  |
| --- | --- |
| Abbreviation | Expanded Form |
| TfL | Transport for London |
| TCS | Tata Consultancy Services Ltd. |
| TOLA | Taxi and Private Hire Operational Licensing Administration |
| TPH | Taxi and Private Hire |
| IT | Information Technology |
| CSV | Comma Separated Values |
| AWS | Amazon Web Services |
| VPC | Virtual Private Cloud |
| RDS | Relational Database Service |
| ETL | Extract, Transform and Load |
| CLI | Command-Line Interface |

## Audience

This document is intended to provide a high-level overview to different teams:

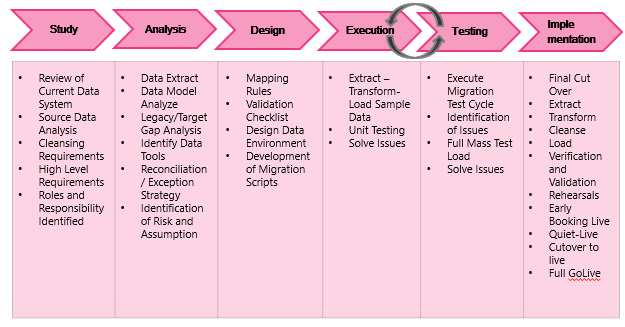
* TfL Project Management & Technical Team
* TfL Data Manager
* TCS Project team for carrying out data migration activity

# Data Migration Strategy/Approach

Data migration is required as part of introduction of a new Service System replacing the existing TOLA system. TCS will perform the following activities:

* Understanding the data definition specification of existing TOLA system
* Finalizing data to be migrated into new Service System
* Preparing mapping of data fields of existing TOLA system to new Service System
* Preparing scripts for migration
* Extraction of data
* Reporting uncleaned data to TfL
* Receive cleaned data from TfL
* Migrating cleaned data to the database of new Service System
* Develop and run reconciliation job and generate exception report
* Analyse root cause, resolve errors and modify migration scripts
* Perform testing on sample migrated data

The below Figure 1 provides the series of activities performed for Data Migration.

**Figure 1: Data Migration Activity**

## Scope

### Inclusions

Data related to TPH Operations from TOLA (existing Service System) and other required data sources is required to be migrated in new Service System. Data Migration scope shall include activities such as,

1. Assessment and identification of AS-IS data to be migrated
2. Identify and provide format for CSV files in which data from TOLA will be given. CSV files shall be Password protected and shared.
3. These CSV files will be sent to AWS S3 bucket via Secure FTP where the data will be stored by server-side encryption (AES-256).
4. Verifying format for provided in CSV files
5. Assessment, identification and design of target TO-BE database
6. Data profiling and identification of erroneous data.
7. Identification quality assessment of AS-IS data to be migrated
8. Report erroneous data to TfL and execute cleansing scripts as per inputs from TfL on erroneous data
9. Identification of data to be mapped from AS-IS database to TO-BE database.
10. Identification & development of data transformation scripts as required in TO-BE data model
11. Development of reconciliation & validation scripts of the data migrated; Corrections for errors discovered during validation process. Validation includes successful testing in ST, stage-wise
12. Testing in FAT and UAT environment
13. Migration of data from AS-IS legacy and application to the TO-BE architecture (Non-prod and Production instances)
14. Migration of documents and images the existing physical storage to the new system (TO-BE)
15. Once data from CSV files are loaded into target service system, the same will be deleted from S3 bucket and informed to TfL.

This type of data which would be migrated will be,

1. Master and Standard data
2. Configurational Data
3. Transactional Data
4. Relationships
5. Documents and Images

### Exclusions

Below are excluded from the scope of data migration activity

1. Data extraction from TOLA and preparation of required CSV files
2. Migration of data pertaining to workflow assignment – It covers such transactional data which is relevant to workflow engine of TOLA. All in-flight applications with required transactional data will be migrated.
3. These exclusions would be revised after study of data structures and data specifications of existing TOLA system and other data sources.

### Pre-requisites

Below are pre-requisites required from existing incumbent supplier.

1. Data definition specification of required data to be migrated
2. Entity-Relationship Diagram with FK and other constraints
3. Understanding of existing datasets

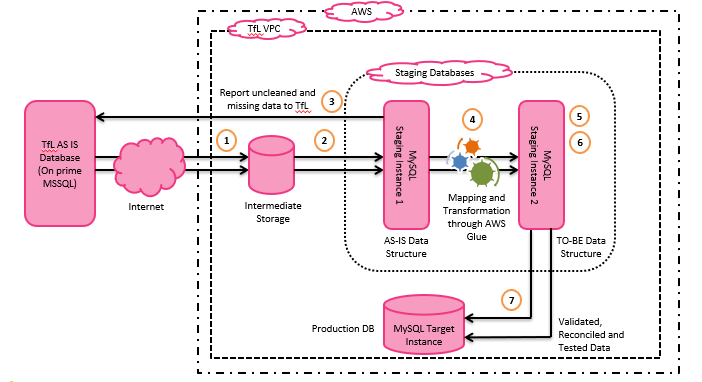
## Data Migration Approach

TCS will perform the following activities under the scope of data migration:

* Data Extraction
* Data profiling and Cleansing
* Data Mapping
* Data Transformation
* Data Validation and Reconciliation
* Data Security

TCS will follow approach as shown in Figure-2 for Data Migration.

**Figure 2: Data Migration Approach**



TCS will follow the following steps:

1. Source database CSV files (password protected) shall be uploaded into AWS S3 which is secure file storage on cloud. Some delay may be introduced when uploading the dumps to S3. This delay may be attributed due to the available network latency. The secure file transfer will be used with additional security and firewall configuration.
2. Uploaded CSV files shall be restored in an Amazon MySQL RDS instance (similar to source data structure). This instance having AI-IS data structure shall act as source (AS-IS) database on cloud. All the validation, reconciliation of data shall be done with respect to this instance. This instance shall be referred to as **‘replica-source’** instance further in this document.
3. Validation Scripts will verify data loaded in replica-source and find out uncleaned data/ missing data and report the same to TfL for Data correction. Then cleaned data will be loaded on replica-source on receipt of the same from TfL.
4. Mapping and Transformation scripts developed in AWS Glue will run and migrate data into Instance 2.
5. Cleaned data from Instance 1 is now ready to be moved to next staging instance. Instance 2 is Amazon MySQL RDS instance having data architecture designed for TO-BE business process. Data transformation and mapping shall be required when migrating to Instance 2. AWS Glue ETL shall be used for data transformation; data mapping and migration from Instance 1 to Instance 2. Transformation and mapping to be implemented shall be documented.
6. All the required data validation & data reconciliation shall be done in Instance 2. Correction of erroneous data shall be done. Validation of migrated data in Instance 2 shall include testing. Validation outcomes shall be documented and shared with TfL for review. Sign-off from TfL shall be taken for the migrated data in Instance 2.
7. Dump of reconciled and approved data from Instance 2 shall be imported to target production MySQL instance.
8. Documents available in TOLA system file storage, required as per new Service System shall be migrated to AWS S3 folder structure with defined transformation logic.
9. After migration of all data, Quiet-Live phase will start. In this phase business transaction will be performed controlled environment on production setup and with migrated production data. This will happen before 3-4 weeks before Cut Over.

### Requirement of Source System Data and Details of Target Database

#### Requirements from Source System data

The required details from source system like AS IS data, CSV formats etc. will be defined during MS2B – Detailed Design Phase. This activity is also dependent on information mentioned in Section 2.1.3.

#### Data Extract Schedule

This table contains our proposition for the data extraction schedule for the data load required from existing systems/suppliers.

**Table 3: Data Extract Schedule**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Phase** | **Sub-Phase** | **Expected Date** | **TOLA/ Source Systems** | **Count of Load** | | **Remarks** |
| **Initial** | **Delta** |  |
| Design and Development | Sprint 4 (19th Oct 2021) | Start | Test | 1 |  |  |
| Sprint 5 (9th Nov 2021) | Start | ETL Configuration and  Development on Initial Load | | |  |
| Sprint 6 (30th Nov 2021) | Start |  |
| Sprint 7 (21st Dec 2021) | Start |  |
| Sprint 8 (11th Jan 2022) | Start | Test | 1 |  |  |
| Sprint 8 (31st Jan 2022) | End | Test |  | 1 |  |
| Testing | MS4–FAT (1st Feb 2022) | Start | Test | 1 |  |  |
| MS6–E2E (30th March 2022) | Start | Test |  | 1 |  |
| MS6 – UAT (13th May 2022) | Start | Test | 1 |  |  |
| MS6–UAT (29th June 2022) | End | Test | 1 | 1 | At this stage we are taking the delta and full load considering that TCS will do reconciliation of all the incremental deltas applied on target database against this received full load. It will give confidence that all deltas have been processed accurately. |
| Performance Test (1st March 2022) | Start - 1 week | Performance / Production | 1 |  |  |
|  |
| Migration | Early Booking Live (MS7A) (8th Aug 2022) | Start - 1 week | Production | 1 |  |  |
| Early Booking Live (MS7A) (29th Aug 2022) | Start + 2 weeks | Production |  | 1 |  |
| Service Readiness (MS7B) (1st Oct 2022) | Cutover | Production | 1 | 1 | At this stage we are taking the delta and full load considering that TCS will do reconciliation of all the incremental deltas applied on target database against this received full load. It will give confidence that all deltas have been processed accurately. |
| Final Load (21st Oct 2022) | Cutover & Go Live | Production | 1 | 1 |

**Note:** The above mentioned dates may vary as per the plan.

#### Details of Target Database

Details related to structure of target Database will be added during completion of Detailed Design – MS2B

#### Change Freeze in Existing Source Systems

It is very essential to freeze/control changes in current source systems as it will impact the data migration design, reconciliation scripts, migration scripts, test cases. This will impact over all Data Migration Schedule. Considering this, TCS proposes to freeze changes to current source systems 21st December 2021

### Data Profiling & Cleansing Methodology

Data profiling and cleansing refers to the process of assessing the quality of data in source and then reporting the same to TfL. The mechanism takes into consideration incomplete data, data relevance, data accuracy and data duplication as factors for validation. Data cleansing focuses on maximizing data accuracy in a system thereby ensuring only clean & required data is moved to target instance(s). Data verification shall be integral part of the data cleansing process that checks data is available, accessible, and complete in correct format. Following data quality check will be performed in ‘replica-source’ instance (on AWS cloud platform) as part of data profiling & cleansing activity.

* Data uniqueness checks
* Inconsistencies for foreign keys across data files
* Data format & data length validations by data types e.g. date fields, numeric values
* Null/Not Null Validations
* Intra field validations
* Control/ de-normalized fields like totals/count/average, etc.
* ~~All rejected records will be shared with TfL team for review and correction. Corrected data received from TfL will be loaded in replica-source.~~ All rejected records will be reported with TfL team for review and further steps will be taken based on direction given by TfL.
* All the documents shall be scanned for any virus. Since documents with extension .gif are a security threat, hence, \*.gif files shall not be stored in new Service System file system. Available \*.gif files in TOLA system to be converted to \*.jpg, \*.jpeg format.

TCS will report AS-IS data to be cleaned to TfL or supplier of source system. This data will be cleaned at the end of source system or by TCS as per TfL feedback. This is dependent on type of erroneous data. TCS will discuss with TfL of all such possible scenarios, define and identify business rules during the design phase to perform data reconciliation. This is further described with examples below:

* 1. In any extract if we do not receive mandatory attributes for records, data against to such records need to be provided again from source system or by TfL.
  2. If we receive duplicate records, TCS will highlight such records to TfL and data cleansing will be done as per direction from TfL.
  3. In any extract, if we receive data of 100 characters length in any attribute and we have defined 80 characters length as per Data definition specification of source system, TCS will highlight such cases to TfL. Data cleansing will be done as per TfL directions or data against to such records needs to be provided again from source system.

Data cleansing will be done in 2 ways as enumerated below.

* Automatically – programs shall be executed in AWS RDS instance of cloud (replica-source).
* Manually – Any erroneous data records identified shall be shared with TfL team for review. TCS will implement the corrections suggested by TfL team with help of DML (data manipulation language) scripts.

Detailed rules for data cleansing will be discussed and finalised during design phase.

### Data Mapping

Data mapping is the process of mapping data fields from data definition specification of existing TOLA system to the data fields of target database of new Service System. After analysis of ERD and data from existing system, mapping of each data fields from existing system and new Service System table columns will be submitted to TfL for validation and review. Table structures defined as per the new Service System business process shall be considered as the base for identifying the mapping of data fields from existing Service System. Only such identified data shall be migrated. Any other available data in existing Service System which is not mapped with new Service System data shall have to be identified jointly with TfL team. These identified data shall be included for migration to cloud platform after discussion with TfL team.

Data mapping of new Service System table column fields with replica-source table column fields shall be defined in AWS - Glue ETL. In case of complex scenarios, mapping functions shall be written in Python in AWS Glue ETL thereby automating the process for entire migration.

Unstructured data such as documents, jpeg, scanned files etc. will be indexed and stored in predefined folder structure into Document Management System. Reference of each of the documents will be mapped with respective application record. This will be detailed down in AS-IS to TO-BE mapping document.

### Data Transformation

Data transformation is the process of converting data from one format or structure into another format or structure. Data transformation will form a critical step in revamping of TfL-TPH system. Data transformation shall include range of activities: converting data type, codification of data, transposing of data columns, splitting of a column into multiple columns, sorting, selection of only required columns for migration etc. as per new Service System design.

1. AWS Glue ETL shall be used for data transformation. This tool can generate its own script for simple transformation rules defined.
2. For medium & complex transformation, mapping functions written in Python shall be required to define the transformation rules in AWS Glue. This shall help automate the process of data transformation during migration process.

### Approach to migrate In-flight Data from existing Service System

This section will be described during MS2B – Detailed Design Phase. All in-flight applications with required transactional data will be migrated to new Service System.

### Data Validation & Reconciliation

A framework to reconcile the data migration will be designed and configured during the early stages of development in close collaboration with TfL and the current TOLA system vendor to address every part of the data. The framework will ensure:

* **Metadata Validation** – This will cater to cover all the data entities in the existing system that are necessary for the functioning of the system are available as part of the new system. The metadata will be categorised into configuration items, look up values, business rules, workflow hierarchy, letter / notification templates etc. In some cases, direct value to value mapping will do the validation while in some cases simulation of the system will be required which will be done as part of Site Acceptance test.
* **Data Reconciliation** – An automated data comparison tool will be built to validate that the migration is successful by means comparing data entity by entity. The tool will work by providing the data export out from the TCS Service System in the same format in which it was provided from the TOLA system. Then a comparison is run against each of the records and data attributes, as well as a count of the records. In case of any exceptions i.e. if a mismatch is reported, then those records are identified and the necessary Root Cause Analysis (RCA) is performed. Based on the RCA, the fix will be applied to the appropriate element(s). The tool will be tested for accuracy during the course of the delivery and fine-tuned for better efficiency and performance.

### Data Security

Data Security is very important aspect during Data migration activity. Following measures will be taken to ensure security and integrity of data.

* **Protection of Source data** –Source service system supplier will prepare CSV files for source data and apply password protection. The same will be shared by them to authorised team of TCS which is working from UK/ EU geography.
* **Transfer of Source data to Intermediate storage** –Source service system supplier will have 2 factor authentications i.e. credentials of VPN and credentials to send the file through SFTP. Moreover, the data will be sent to Intermediate storage from whitelisted IPs. These CSV files will be password protected. The data will be stored in AWS S3 which provides server-side encryption with AES-256 algorithm. The same will be enabled while setting up the S3 bucket.

TCS will further refined this mechanism during MS2B after Data Migration meetings with TfL or incumbent supplier.

* **Infrastructure Access** – Access to such instances or servers, where data from the source system is available, TCS will only allow access of authorised TCS personnel located in UK/EU and working on this project. The data migration activity will be conducted from UK/EU only.
* **Disposal of Data after migration to ‘replica-source’ –** After reconciliation of data at replica-source, TCS will remove the source CSV files from intermediate storage and inform TfL.

# Data Migration Plan

Data Migration is one of the workstreams which will run as a separate track during the process of Transition. The figure below provides a high-level overview of the various phases and tasks that will be involved in the Data Migration workstream and its testing.

**Figure 3: Data Migration Plan**

Timeline

Description automatically generated

The following table provides a brief description of each phase identified above:

**Table 4: Data Migration activity in each phase**

|  |  |  |  |
| --- | --- | --- | --- |
| Phase | Milestone | Key Data Migration Tasks | Deliverables/Output |
| Mobilisation | Mobilisation  (MS1) | Define data migration strategy | Draft Data Migration Strategy  **(Under Review)** |
| Design and Development | Preliminary Design (MS2A) | Refine data migration strategy  Data mapping (source and Target data models)  Agree on data sources |  |
| Detailed Design (MS2B) | Data mapping (source and Target data models)  Define data validation rules to be applied on each data type  Define handling of rejected data | Migration Strategy completed and signed off |
| Development (MS3) | Develop extraction / Transformation / Load scripts  Test Data loading  Develop test cases | Executable scripts if required by TfL |
| Testing | Factory Acceptance (MS4) | Execute data load on Test System (Test Data) | Successful data load  Data Migration Testing Plan |
| End to End Testing (MS5) | Fix any errors in data load and run data load on UAT system (Test Data) | Successful data load prior to start of joint end to end testing |
| Migration | MS7A (Early Bookings live and BCDR Testing complete) | Initial extract from Production | TOLA production data available in new TPH system |
| MS7B (Migration to live operations complete) | Incremental data loads | Final incremental data load from TOLA |

The project activities related to the migration will start from the Design and Development Phase (covering milestones MS1 to MS3B) while each of the deliverables will be tested during the Testing Phase (covering milestones MS4 to MS6) which will then be tested in the Testing phase.

# Data Migration Testing

Migration testing will be undertaken to ensure that the Migration to the TPH Service System (including all data migration) has been successful. The Data Migration testing plan will be prepared in conjunction with TfL and finalised by MS4 as per current plan.

**Continuous Testing**

The development and configuration of the TCS Service system undergo testing as part of each Sprint, Release as well as Regression test. Following testing will be conducted on **migrated test data** from TOLA system test environments.

* MS4 – Factory Acceptance Testing
* MS5 – System Integration Testing
* MS6 – End-to-End Testing & User Acceptance Testing
* Regression Testing at every release level

In case if there is any issue with test data quality of existing source system, TCS will generate the test data for testing purpose.

**Validation and Verification**

TCS will perform Data Validation and Reconciliation mentioned in section 2.2.6.

# Go Live/Cutover

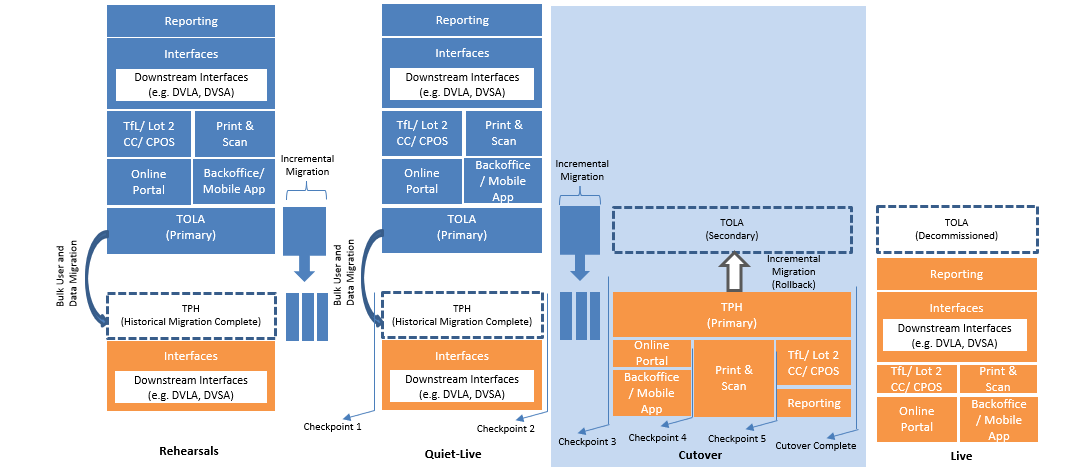
## Checkpoint Based Approach for Go Live/Cutover

The Go Live in production will be carried out in four phases as listed below

* Rehearsals
* Quiet-Live
* Cutover to Live
* Live

The diagram below is a depiction of the strategy we have devised to full the objectives of migration.

**Figure 4: Phased Migration Approach**



This approach of having multiple phases in the migration phase will:

* Ensure that there are multiple iterations of the migration activity which will help in identifying any potential issues and resolve them in time. The resolved issues will also undergo testing as part of the multiple iterations
* Provide the opportunity for business users to acquaint themselves with the system based on the real migrated data in a controlled environment, as well as replicate any known issues or complex business transactions
* Reduce the risk in actual cutover, contribute to the final, detailed cutover plan/timetable and thus ensure a smooth cutover.

1. **Rehearsals**

The Rehearsals phase will start 1 weeks after the start MS7A phase and will span about 3 weeks. This phase will ensure that all testing is signed off and that the application and integrations are being deployed on the production instance. In addition to the code and configuration migration, this phase will also ensure that migration of the data from the TOLA Live will also happen for the purpose of initial data migration test on production as well as to enable the rehearsals. The rehearsals phase will ensure the following:

* Completion of the migration of the new TPH application and the necessary configurations, such as business rules, templates, access control etc on the production platform
* Migration of the initial full data extract from the TOLA system into the TPH system for rehearsal. Data Validation and Reconciliation jobs will be run, and issues reported during this activity will be resolved.
* Migration of the delta data extracts from the TOLA system into the TPH system for rehearsal. Data Validation and Reconciliation jobs will be run, and issues reported during this activity will be resolved.

**Rollback Approach:** No rollback will be required as TOLA will still be the master and all the transactions will be happening in TOLA

1. **Quiet-Live Phase**

Quiet-Live is replicating the business transaction within controlled environment on the production setup and with migrated production data. The Quiet-Live will happen 3 to 4 weeks before Cutover and as part of the phase, all End-User Data, Application Data and Objects will be migrated into the TPH System as a bulk load using the ETLs. The point in time to be begin the migration will be achieving a ‘GO’ at Checkpoint 1 which will involve satisfaction of the below criteria:

1. **Approval by TFL, Lot 2 Supplier, Scan supplier and Print supplier** to ensure that the integration points, tools, and resources are available for the migration in this phase
2. Functional Completeness achieved by acquiring the sign-off of User Acceptance Testing and the Rehearsals. This will ensure that there are no changes in the system to be expected and the business processes and functions will be tested on the approved codebase.
3. Initial Data Extract and any subsequent Delta Extract(s) are available for migration to ensure that the data can be loaded in the system for the Quiet-Live phase

Following the successful migration of the data, a thorough validation exercise will be carried out to ensure the migration was successful using custom ‘Data Reconciliation’ framework designed and built to verify the completeness at an entity level.

During this phase the TOLA system will still be the primary while the TPH system will be connected to the other downstream interfaces, kept in a Quiet-Live state with no data transferred to downstream interfaces.

**Rollback Approach:** No rollback will be required as TOLA will still be the master and all the transactions will be happening in TOLA

1. **Cutover to Live**

Once the data migration is successful and signed-off as part of Checkpoint 2 by TFL and Lot2 Supplier certifies that all dependencies with TPH are resolved, the cutover will commence. We propose to take a ‘big bang’ approach and perform the cutover to live operations over a weekend. This will be planned as a part of the Cutover Plan which will be agreed and signed off as part of the deliverable of MS6. This approach will avoid the need of data synchronisation as well as parallel running of both TOLA and the new TPH system and the data synchronisation complexity, risks and costs associated with parallel running. As a part of the big bang approach, all the data will be migrated to the new TPH system.

The cutover activity will be done over the weekend with a planned outage and the incremental data will be loaded from the TOLA to the new TPH system. The Checkpoint 3 in the cutover strategy (mentioned in the diagram above) once crossed, will provide the ability to test the incremental data migration and the system will be made available to all the stakeholders for final validation. The incremental data will be provided from the TOLA system.

In the event of failure of any of the checkpoint 3,4 or 5, due to issue in the TPH system or with any of the other stakeholders such as the Lot2, Print or Scan Supplier, attempts will be made to fix the issue in the new TPH before the downtime. In an unforeseen situation of encountering a serious issue even after multiple rehearsals and extensive testing, if the issue cannot be resolved in the downtime (even if it needs to be extended by few hours with business approval) then we shall roll back to TOLA and plan the cutover on a future alternate date. The issue will be investigated and resolved before the next attempt. The issue could have a root cause in any system(s) or associated component. Therefore, TCS will take ownership of working with the stakeholders to ensure that the issue is resolved.

**Rollback Approach:** The TOLA system be brought back into operations and will act as the master system. The issue will be investigated and fixed, and a future cutover date will be agreed.

1. **Live**

The live operations phase will commence when the cutover has completed, and TfL has started using the TPH system for the live business transactions. At this point the TOLA system will become the candidate for the decommission process.

Once the cutover is complete, and before the TOLA is decommissioned, in the event of any issue arising, attempts will be made to fix the issue in the new TPH system. Depending on the severity, cause and the anticipated time to fix the issue, assessment will be done to check the feasibility to fallback to TOLA system. The assessment will be based how long the new TPH system has been accepting the business transactions and how much time and effort will it take to recreate all the data and transaction in the TOLA system. For the fallback to the TOLA system, the following 2 options can be exercised

Option 1: Manual extract of changes made in new system after ‘go-live’ to be entered offline in TOLA in case needed.

Option 2: Technical solution + Dual feeding.

We would like to discuss the options to fallback during the Design phase.

**Rollback Approach:** Option 1 or Option 2 based on the assessment of time and effort it will take to recreate all the data and transactions which are not in TOLA system

## Early Booking Live (Vehicle Inspection)

The Early Booking Live milestone will ensure that the customers will be able to book for the Vehicle Inspection on the new locations provided by the Lot2 Supplier in advance of the actual go-live of the Lot2 supplier Vehicle Inspection site. For the early booking option, TCS proposes the loading of Vehicle Inspection booking schedule in the existing TOLA System. This will avoid the complexity of two-way data sync between the new Service System and existing TOLA System which will be costly and will add more risk to the migration. The following are the reasons for proposing this option:

* Since the data schema related to the Inspection Locations, Inspection Bays at each location, opening times, inspection slots, is not changing, loading the additional data will be configured in the Existing Service System. The early bookings can be taken on the newly created Vehicle Inspection booking schedule based on the agreed date as per the cutover plan. All the early booking data will be migrated as part of the cutover (MS7B).
* All the inspection data will get migrated as part of the initial data migration and delta migration during cutover
* From MS7B onwards the existing Service System will not be in use anymore and its decommissioning process can be started

# Risk & Mitigation Plan

Following table shows associated risks and plan to mitigate the same.

**Table 5: Identified Risks and Mitigations**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Risk ID** | **Category** | **Risk Circumstance: "This situation exists…."** | **Risk Description: "….therefore, there is a risk that…"** | **Risk Impact: "….and the impact could be."** | **Probability** | **Impact** | **Severity** | **Mitigation Plan / Contingency** | **Owner** |
| 1 | Implementation | Performance Issues | The system will perform slow on transactions | Increased user time on transactions leading to inefficiency | L | M | M | - TCS Platform is tested for systems with far higher volume of transactions than at TfL  - Detailed Test Strategy as part of Service design and delivery will be worked out  - Test Plans and Test cases for each test phase will be mapped to requirements for full traceability | TCS |
| 2 | Implementation | Data Quality | Wrong data, missing data, misrepresented data | Inaccurate transactions and reporting | L | H | H | - An extensive exercise would be performed to understand the data quality issues through existing documentation and remediated during design of migration workflows.  - The migration workflows will be extensively tested through multiple iterations of data quality audit performed using data reconciliation frameworks.   - We shall build a robust bad record rejection ruleset to remediate each discrepancy | TCS, TfL |
| 3 | Implementation | Rejection of new system by users | Go Live will be delayed | Delay in milestone | L | H | H | - Detailed Go-live strategy will include onboarding of key stakeholders as part of core team  - Collaboration with business for scenario identification  - Organisation Change Management for consensus building | TCS, TfL |
| 4 | Implementation | Access and Connectivity Problems with Integrations | Delay in testing related to migration | Delay in milestone achievement | L | L | L | Necessary Disaster Recovery options to be triggered | TCS, TfL, Other Suppliers |
| 5 | Implementation | Data Extracts | Delay in receiving data extracts for migration | Delay in milestone achievement | L | H | H | Multiple extracts are provisioned with contingency | TCS, TfL |
| 6 | Implementation | Dynamic Changes in Source System design after sharing of Extract | Reconciliation report will fail. Changes in migration scripts. Delay in migration testing | Delay in milestone achievement | H | H | H | TCS and TfL will decide and communicate to current incumbent about change freeze in existing service systems. | TfL, TCS |
| 7 | Hypercare | Data Loss | Data might not be available in new TPH system post migration | Missing records, integrity, reputation loss | L | H | H | Reconciliation tool will be used, and verification and validation will be done on each migration with report being produced | TCS |

# Responsibility Matrix & Dependencies

The following table provides the responsibility matrix between TfL and TCS for Data Migration activities:

**Table 6: Responsibility Matrix**

|  |  |  |  |
| --- | --- | --- | --- |
| # | Activity | TfL | TCS |
| 1 | Creation of Data Migration Strategy Document | A,C | R |
| 2 | Creation of design document with mapping and transformation rules adopted for new Service System | A,C | R |
| 3 | Sharing of Data Definition Specification and ERD diagram of Source Systems | A,R | C,I |
| 4 | Table column mapping of TO-BE with AS-IS | C,I | A,R |
| 5 | Sign-off for data mapping– table column mapping of TO-BE with AS-IS | A, R | C,I |
| 6 | Sign-off for data transformation in TO-BE architecture | A, R | R |
| 7 | Sharing of CSV files from AS-IS data sources | A,R | C,I |
| 8 | Report generation of erroneous records and reporting to TfL | C,I | A,R |
| 9 | Provide correct cleaned data to be migrated | A,R | C,I |
| 10 | Data Cleansing/Loading clean data as per corrections/rules provided by TfL | A,C | R |
| 11 | Migration of Data in Instance 2 | C,I | A,R |
| 12 | Readiness of validated data in Instance 2 | I | A,R |
| 13 | Migration of Data from Instance 2 to Target instance | I | A,R |
| 14 | Reconciliation Report for Migrated Data from Target Instance | C,I | A,R |
| 15 | Sign off Migrated data on Target Instance | A,R | C,I |

The following are the dependencies and constraints:

|  |  |  |  |
| --- | --- | --- | --- |
| **Category** | **Description** | **Owner** | **Mitigation** |
| Dependencies | Support from current vendor of the TOLA system for migration and turning-off the online channel. | TfL | Early engagement with the TOLA supplier for migration |
| Dependencies | Availability TfL, Lot 2 Supplier, Print & Suppliers and current TOLA system vendor for migration planning | TfL, Lot2 Supplier, Print and Scan Supplier | Early invite and enagement. Mutiple workshops to agree plan |
| Dependencies | Availability of TfL, Lot 2 Supplier, Print & Suppliers and current TOLA system vendor for rehearsals | TfL, Lot2, Print and Scan Supplier | Multiple rounds of testing and rehearsals as part of migration |
| Dependencies | Availability of external interfaces to new service system during Cutover/Go Live | TfL | Discussion with third party interfaces to provide parallel inputs to new service system |
| Dependencies | Source Source System SMEs should be idetified and available for queries, clarifications and meetings | TfL | Disucssion with Source system supplier to be available for meetings |
| Constraints | Technical sync between TPH and TOLA system to cater for fallback post migration | TCS, TfL | Alternate approach for data sync to be defined and tested as part of Joint E2E, UAT and Site Acceptance Test |

# Software, Tools and Services

Entire migration process will be on AWS on cloud platform. Below are the services used for different migration activities.

1. AWS Glue ETL – for data mapping and transformation.
2. AWS CLI – for upload of CSV files having size greater than 160GB
3. MySQL workbench