Texas Immunization Registry Data Remediation Project

Joe Lyn Williams, Research Specialist V,
Immunization Information Systems (IIS) Unit
Immunization Section
Texas Department of State Health Services

Agenda

- Introduction
- Background
- Project Phases
- Results
- Challenges

Introduction

- **Overview**:
  - In September 2024, the Texas IIS started a data remediation project to address approximately 1.8 million duplicate client records in the registry.
  - The project dramatically dropped the percentage of duplicates from approx. 5.5% to 0.5% by September 2025.
  - This decrease was achieved over a 12-month period by partnering with our Maintenance & Operations (M&O) team to determine the necessary IT solutions.
  - While program staff worked to reduce the primary record duplicate backlog by manually merging records in the ImmTrac2 user interface.
- **Key Goals**:
  - Reduce the number of duplicate client records in the registry.
  - Implement measures to help reduce the ingestion of duplicates.
  - Create a systematic, backend merge process that continually identifies and auto-merges new duplicates.

Background

COVID-19 Public Health Emergency

- Texas law required all COVID-19 vaccine doses administered within the State be reported to the Immunization Registry within 24 hours of administration to ensure accurate tracking and follow-up.

Issue/Problem

- When multiple data exchange requests were submitted for the same client within a short interval, the requests may have been processed simultaneously across different server instances. The second request did not find an existing client in the system because the first request had not finished processing and created the client. Therefore, two duplicate primary client records ended up being created.

Impact

- After two duplicate primary client records were created due to rapid submission and ingestion, every subsequent update to the clients in question was sent to our pending client table as a possible duplicate. By the end of the PHE, multiple requests submitted created over a million pending client records within the system.

Project Phases

Phase 1

- Identify and delete exact Pending Client (possible duplicate) records.
- Have the Manual Merge Team, consisting of 36 personnel, merge the primary duplicate client records identified in our system.

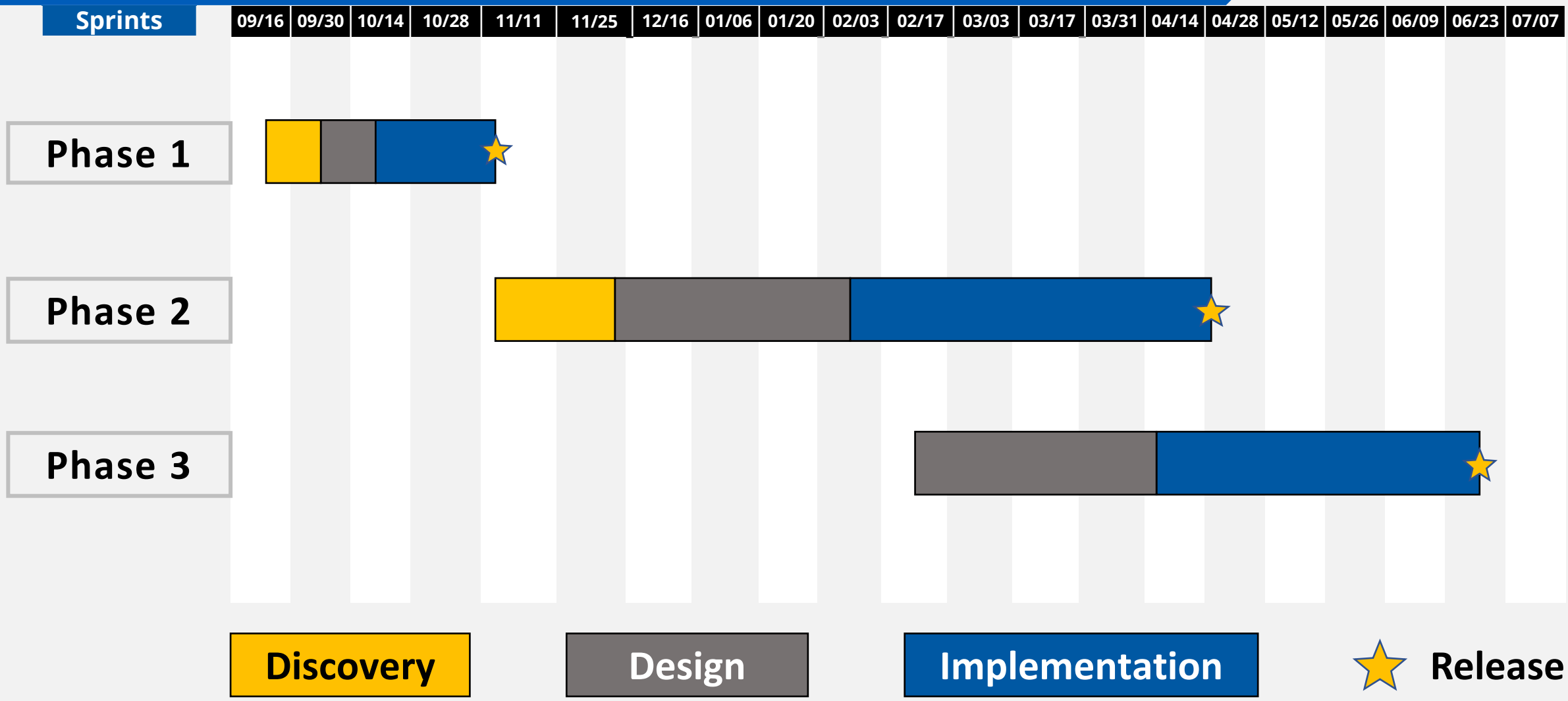
Phase 2

- Design the Auto-Merge Process.
- Set requirements for the Auto-Merge process.
- Implement the data exchange enhancement to stop the rapid ingestion of records causing duplicate records to be created.

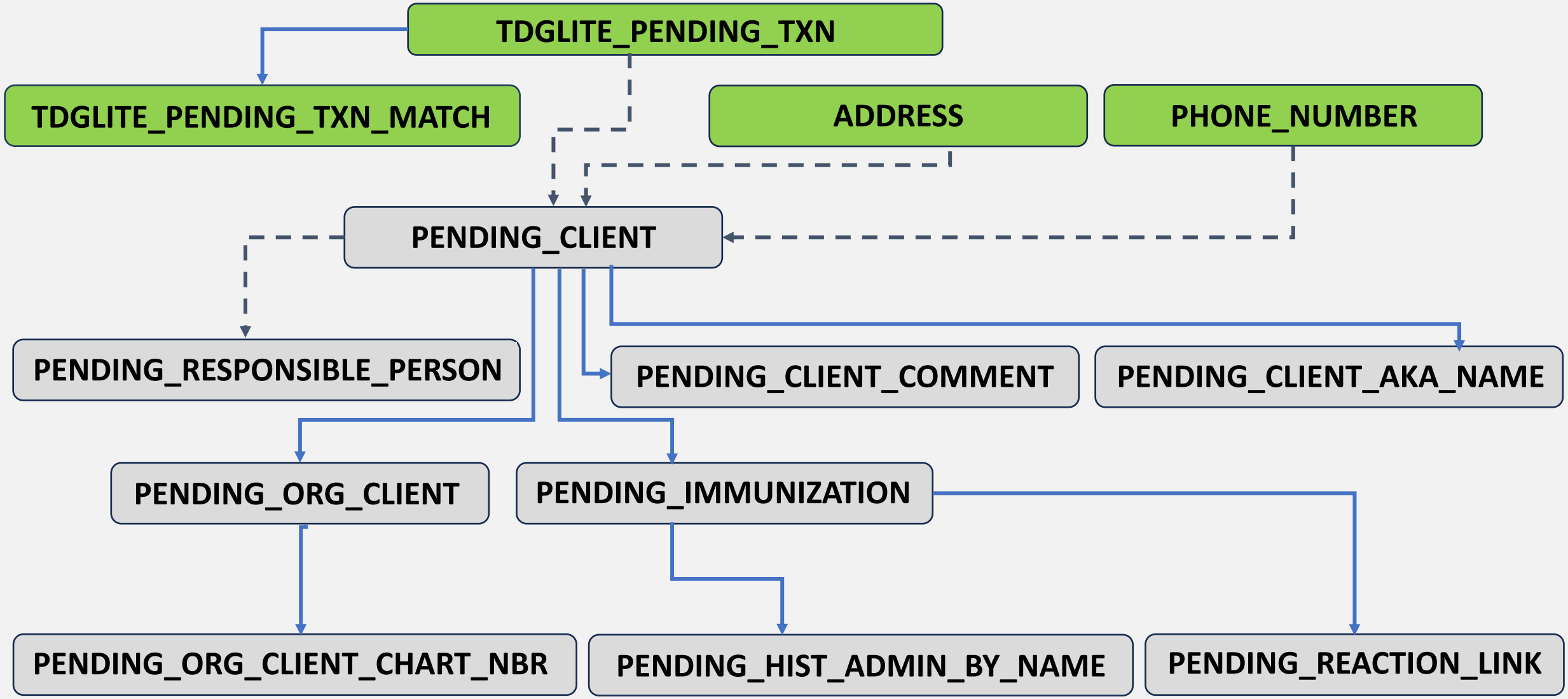
Phase 3

- Deploy Auto-Merge process.
- Identify the duplicate client record backlog and label them as one of the following three options:
  1. Merge
  2. Unpend
  3. Manual Review
- Merge or Unpend the backlog and set a schedule for the Auto-Merge to run opposite work hours during the week and on weekends so as not to put a strain on our database during work hours.

Timeline – High-Level Plan



Phase 1: Pending Client Related Tables



Phase 1: Pending Client Record Creation

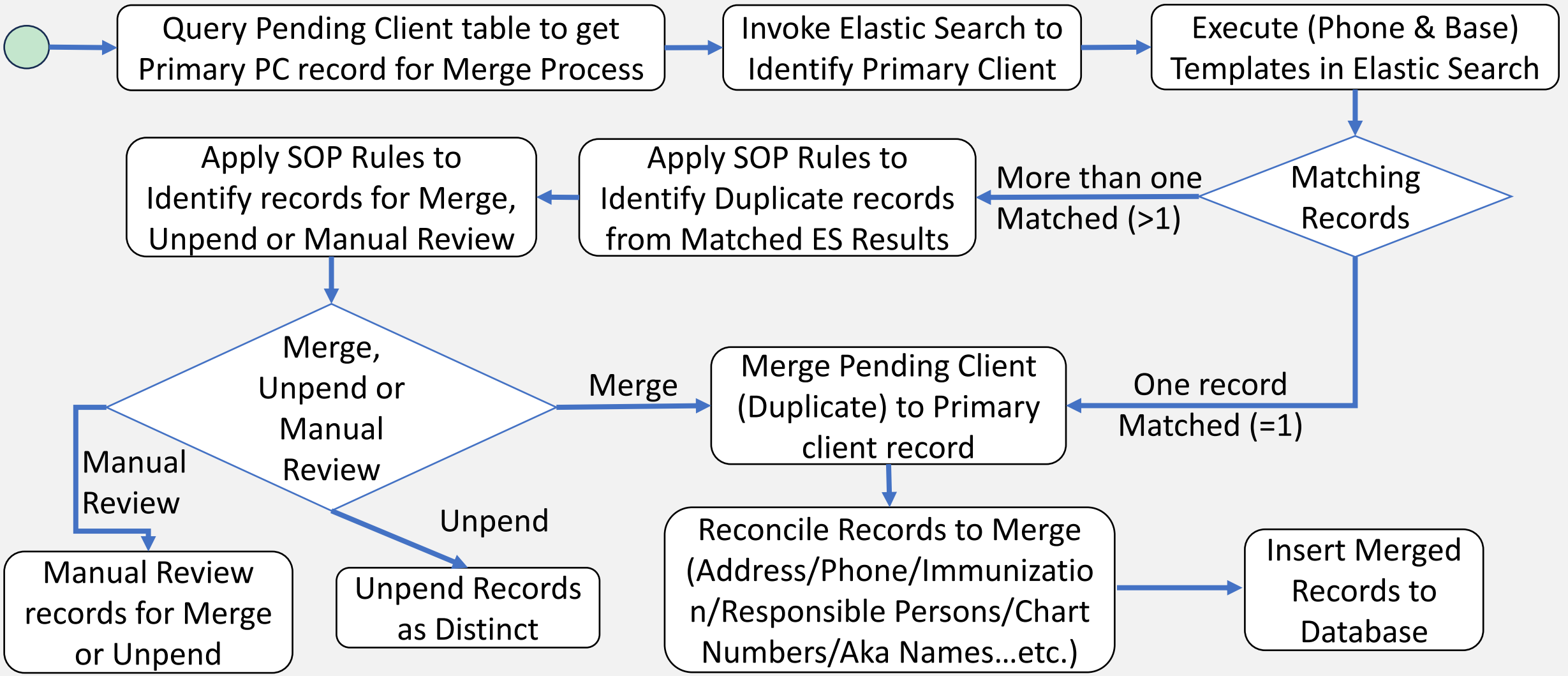
Elastic Search is our deduplication engine for ImmTrac2 and is used to match incoming client data against clients existing in the system. The high-level logic for this matching is as follows:

1. Pre-match template selection - Elastic search uses two templates to determine the parameters to use for the matching:
   a. Phone Number Template - Includes Phone Number, First Name, Date of Birth, and Middle Name.
   b. Base Template - Includes First Name, Last Name, Date of Birth, Gender, Middle Name, and Address.
2. Match using phone number template:
   a. If no phone number provided, go to next step
   b. If no match found, go to next step
   c. If single match found with match score > 37 for DX (27 for UI), return match and update client information
   d. If multiple matches found with match score > 37 for DX (27 for UI), go to next step
3. Match using base template:
   a. If no match found, create new client
   b. If single match found with match score > 37 for DX, return match and update client information
   c. If multiple matches found with match score > 37 for DX, create pending client (pending clients are not created for UI – instead, user is prompted to select one of the multiple matches)

Phase 1: Columns to Identify Exact Duplicates

Tables			
FIRST_NAME	SEX_CODE	RECORD_LOCK_ORG_ID	LANGUAGE_CODE
MIDDLE_NAME	CONSENT_ORG_ID	INDIAN	DIR_ATTRIBUTE
LAST_NAME	BIRTH_CERTIFICATE_NUMBER	ASIAN	FIRST_RESPONDER_ATTR
NAME_SUFFIX	MULTIPLE_BIRTH_COUNT	HAWAIIAN	NAME_TYPE
BIRTH_DATE	MULTIPLE_BIRTH_ORDER	BLACK	
MOTHERS_FIRST_NAME	DEATH_DATE	WHITE	**Tables/Columns**
MOTHERS_MAIDEN_LAST	DEATH_IND	OTHER	STREET_ADDRESS_LINE
RACE_CODE	COUNTY_CODE	MEDICAID_ID	OTHER_ADDRESS_LINE
ETHNICITY_CODE	DEATH_CERTIFICATE_NUMBER	CLIENT_TYPE	CITY_NAME
SSN	CREATE_ORG_ID	MOTHERS_BIRTH_DATE	ZIP_CODE
CONTACT_ALLOWED_IND	PRE_CHANGE_IND	BIRTH_FACILITY_CODE	**Tables/Columns**
CONSENT_IND	HBSAG_STATUS_CODE	BIRTH_STATE_CODE	AREA_CODE
NOTICE_COUNT	RECORD_LOCK_IND	BIRTH_COUNTY_CODE	PHONE_NUMBER

Phase 2: Auto-Merge Process Flow



Phase 2: Client Type Requirements for Merging

- Some examples of requirements created as part of the Auto-Merge Logic included:

  - The latest DATE_ENTERED values kept from all records
  - The CLIENT_TYPE follow the below priority sequence:

  IA (ImmTrac Adult) > **IC** (ImmTrac Child) > **DC** (Disaster Consented) > **DU** (Disaster Unconsented) > **QM** (Questionable Match)

  - The DIR_ATTRIBUTE follow the below priority sequence:

    Y (Yes) > **M** (Minor) > **N** (No), > Null (adults/children)

  - The FIRST_RESPONDER_ATTR follow the below priority sequence:

    Not Null value > **N** (No)

Phase 2: Data Exchange 5-Second Issue/Fix

5-Second Issue

- **Issue**: When multiple data exchange requests were submitted for the same client within a short interval, the requests may have been processed simultaneously across different server instances. The second request did not find an existing client in the system because the first request had not finished processing and created the client. Therefore, two duplicate primary client records ended up being created.
- **Resolution**: Force the duplicate requests to be processed sequentially on the same processing instance (as compared to simultaneous processing). This is achieved by:
  - Extracting the Patient Identification (PID) segment from the incoming Bidirectional (BiDX) or Health Level 7 (HL7) request.
  - Generating a hash code based on the client details in the Patient Identification (PID) segment.
  - Directing requests with the same hash code (i.e., same client information) to be queued and processed within the same instance sequentially.

Phase 2: Create Job Status Tracking Tables

Table Name	Column Name
REMOVE_PENDING_CLIENT_JOB_TABLE	REMOVE_JOB_ID
	TABLE_NAME
	JOB_START_TIME
	JOB_END_TIME
	JOB_STATUS
	D_CNT
	DEL_FROM_ID
	DEL_TO_ID
	DATE_ENTERED
	LAST_UPDATED_DATE
REMOVE_PENDING_CLIENT_EXCEPTION_LOG	EXCEPTION_ID
	EXCEPTION_DATE
	STORED_PROCEDURE
	SQL_ERROR
	ARCH_PARAMETERS

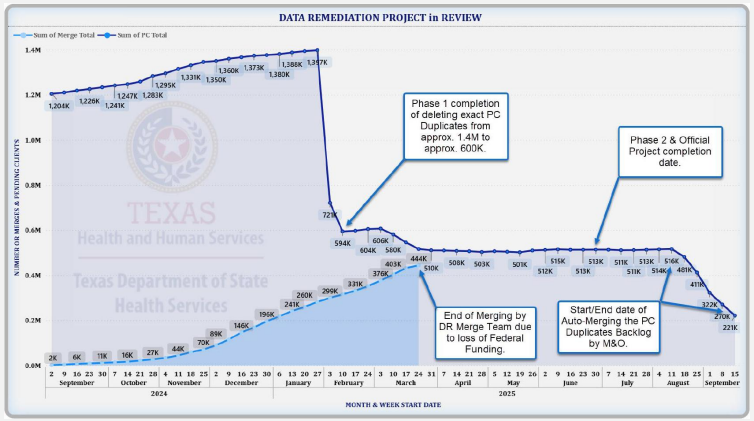
Phase 2: Create Job History Tracking Tables

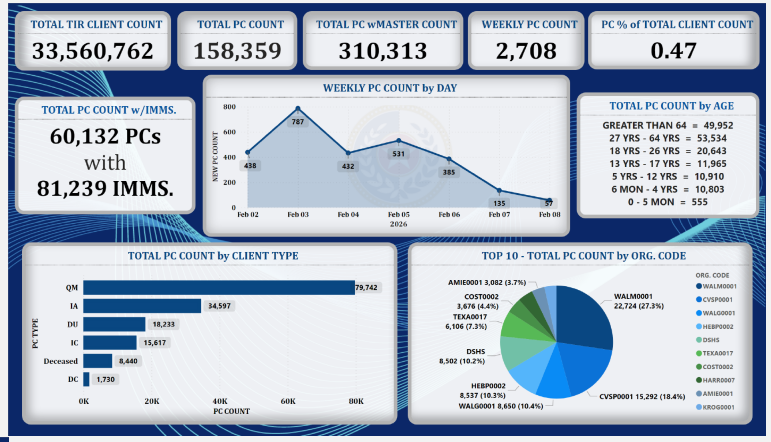
Table Name	Table Created Based On
PENDING_CLIENT_HIST	PENDING_CLIENT
PENDING_RESPONSIBLE_PERSON_HIST	PENDING_RESPONSIBLE_PERSON
PENDING_CLIENT_AKA_NAME_HIST	PENDING_CLIENT_AKA_NAME
PENDING_CLIENT_COMMENT_HIST	PENDING_CLIENT_COMMENT
PENDING_IMMUNIZATION_HIST	PENDING_IMMUNIZATION
PENDING_HIST_ADMIN_BY_NAME_HIST	PENDING_HIST_ADMIN_BY_NAME
PENDING_ORG_CLIENT_HIST	PENDING_ORG_CLIENT
PENDING_ORG_CLIENT_CHRT_NBR_HIST	PENDING_ORG_CLIENT_CHRT_NBR

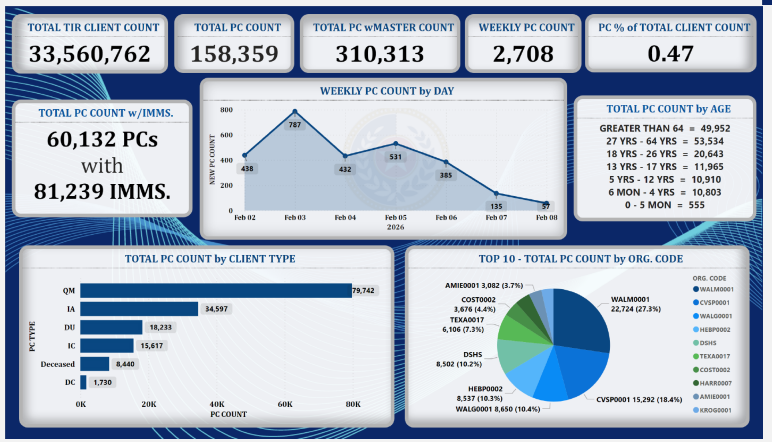
Phase 2: Create Auto-Merge Backup Tables

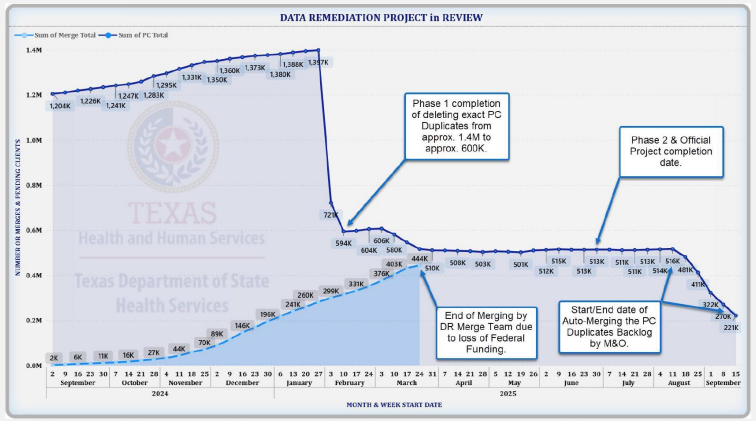
Table Name	Table Created Based On
PENDING_CLIENT_BKP	PENDING_CLIENT
PENDING_RESPONSIBLE_PERSON_BKP	PENDING_RESPONSIBLE_PERSON
PENDING_CLIENT_AKA_NAME_BKP	PENDING_CLIENT_AKA_NAME
PENDING_CLIENT_COMMENT_BKP	PENDING_CLIENT_COMMENT
PENDING_IMMUNIZATION_BKP	PENDING_IMMUNIZATION
PENDING_HIST_ADMIN_BY_NAME_BKP	PENDING_HIST_ADMIN_BY_NAME
PENDING_ORG_CLIENT_BKP	PENDING_ORG_CLIENT
PENDING_ORG_CLIENT_CHRT_NBR_BKP	PENDING_ORG_CLIENT_CHRT_NBR

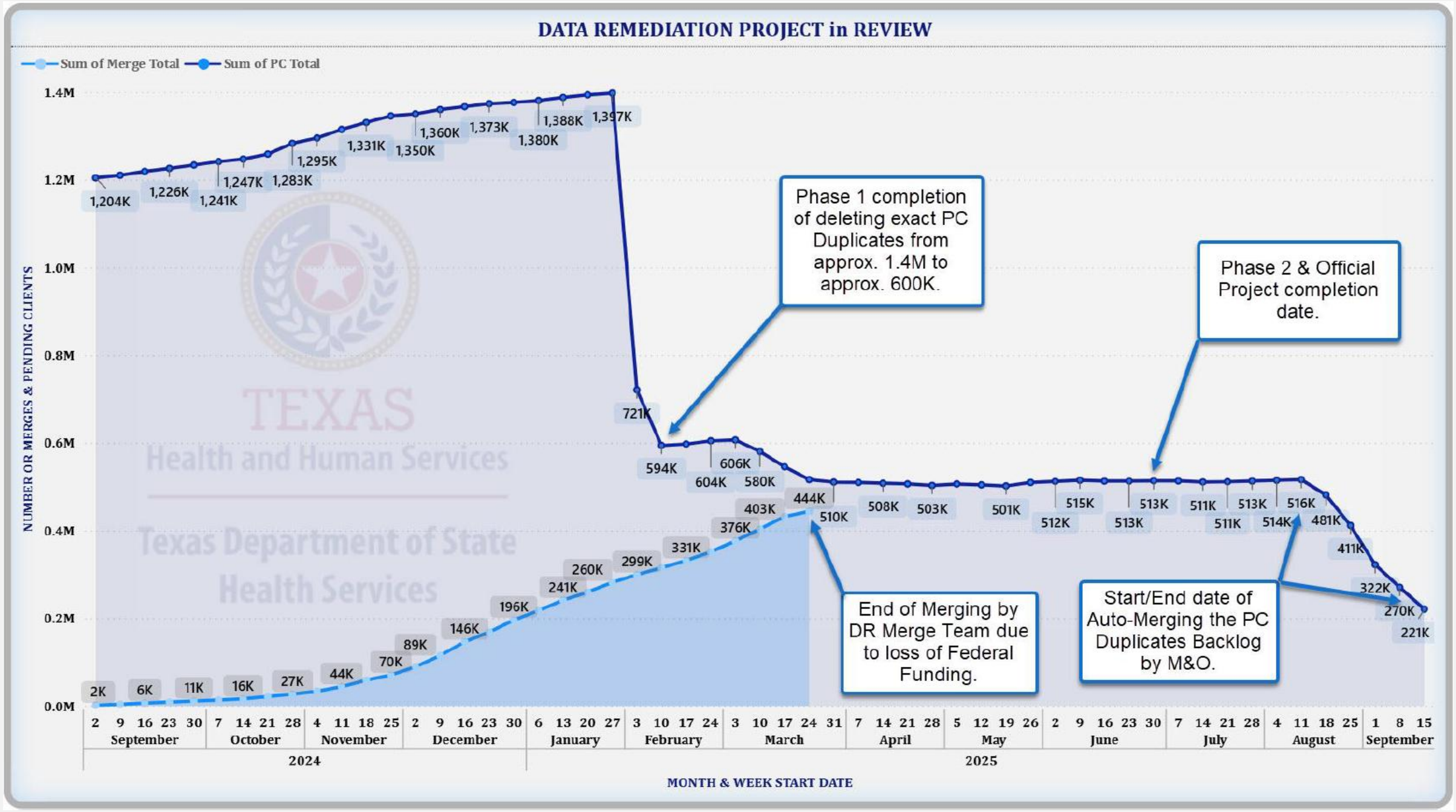
Phase 3: Deployment & Results

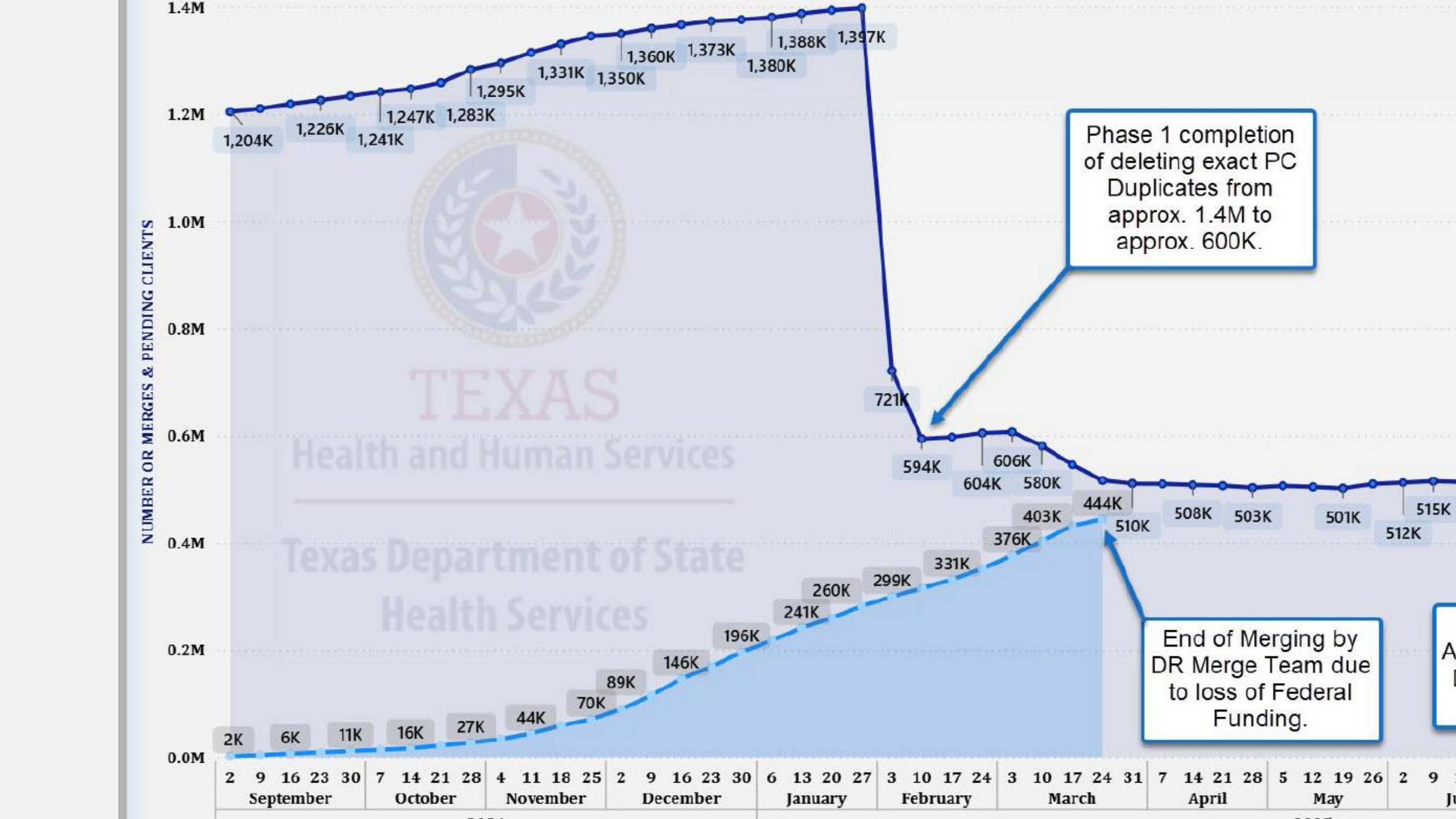


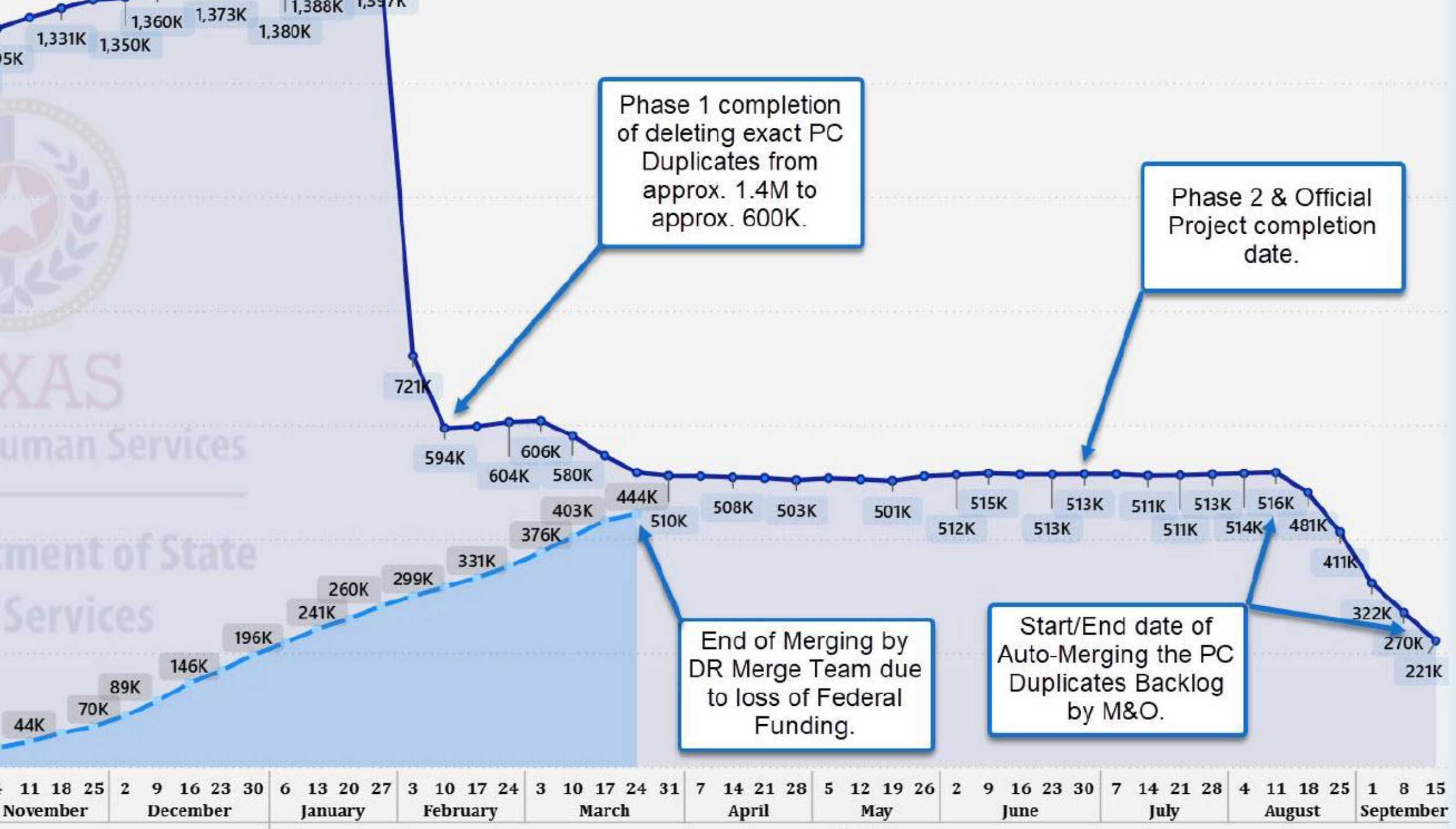


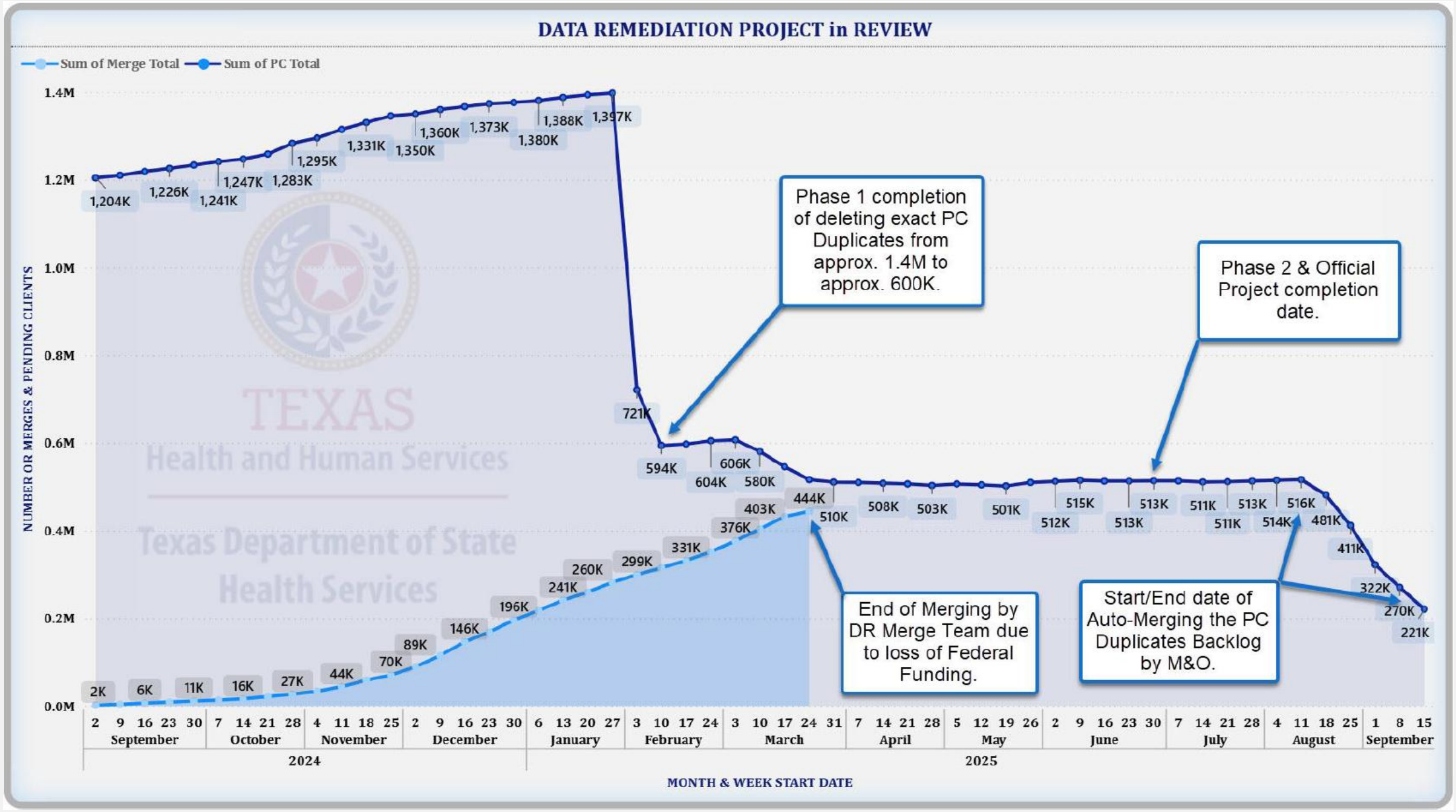





DATA REMEDIATION PROJECT in REVIEW
Sum of Merge Total
Sum of PC Total
NUMBER OR MERGES & PENDING CLIENTS
1.4M
1.2M
1.0M
0.8M
0.6M
0.4M
0.2M
0.0M
1,204K
1,226K
1,241K
1,247K
1,283K
1,295K
1,331K
1,350K
1,360K
1,373K
1,380K
1,388K
1,397K
721K
594K
606K
604K
580K
444K
403K
376K
510K
508K
503K
501K
515K
512K
513K
513K
511K
511K
513K
514K
516K
481K
411K
322K
270K
221K
2K
6K
11K
16K
27K
44K
70K
89K
146K
196K
241K
260K
299K
331K
Phase 1 completion of deleting exact PC Duplicates from approx. 1.4M to approx. 600K.
Phase 2 & Official Project completion date.
End of Merging by DR Merge Team due to loss of Federal Funding.
Start/End date of Auto-Merging the PC Duplicates Backlog by M&O.
2 9 16 23 30 7 14 21 28 4 11 18 25 2 9 16 23 30 6 13 20 27 3 10 17 24 3 10 17 24 31 7 14 21 28 5 12 19 26 2 9 16 23 30 7 14 21 28 4 11 18 25 1 8 15
September October November December January February March April May June July August September
2024
2025
MONTH & WEEK START DATE


NUMBER OR MERGES & PENDING CLIENTS
1.4M
1.2M
1.0M
0.8M
0.6M
0.4M
0.2M
0.0M
1,204K
1,226K
1,241K
1,247K
1,283K
1,295K
1,331K
1,350K
1,360K
1,373K
1,380K
1,388K
1,397K
721K
594K
604K
606K
580K
444K
510K
508K
503K
501K
512K
515K
2K
6K
11K
16K
27K
44K
70K
89K
146K
196K
241K
260K
299K
331K
376K
403K
Phase 1 completion of deleting exact PC Duplicates from approx. 1.4M to approx. 600K.
End of Merging by DR Merge Team due to loss of Federal Funding.
2 9 16 23 30 September
7 14 21 28 October
4 11 18 25 November
2 9 16 23 30 December
6 13 20 27 January
3 10 17 24 February
3 10 17 24 31 March
7 14 21 28 April
5 12 19 26 May
2 9
TEXAS
Health and Human Services
Texas Department of State Health Services


1,331K
1,350K
1,360K
1,373K
1,380K
1,388K
721K
Phase 1 completion of deleting exact PC Duplicates from approx. 1.4M to approx. 600K.
Phase 2 & Official Project completion date.
594K
604K
606K
580K
444K
510K
403K
376K
508K
503K
501K
512K
515K
513K
513K
511K
511K
513K
514K
516K
481K
411K
322K
270K
221K
44K
70K
89K
146K
196K
241K
260K
299K
331K
End of Merging by DR Merge Team due to loss of Federal Funding.
Start/End date of Auto-Merging the PC Duplicates Backlog by M&O.
November
December
January
February
March
April
May
June
July
August
September


DATA REMEDIATION PROJECT in REVIEW
Sum of Merge Total
Sum of PC Total
NUMBER OR MERGES & PENDING CLIENTS
1.4M
1.2M
1.0M
0.8M
0.6M
0.4M
0.2M
0.0M
1,204K
1,226K
1,241K
1,247K
1,283K
1,295K
1,331K
1,350K
1,360K
1,373K
1,380K
1,388K
1,397K
721K
594K
606K
604K
580K
444K
403K
376K
510K
508K
503K
501K
515K
512K
513K
513K
511K
511K
513K
514K
516K
481K
411K
322K
270K
221K
2K
6K
11K
16K
27K
44K
70K
89K
146K
196K
241K
260K
299K
331K
Phase 1 completion of deleting exact PC Duplicates from approx. 1.4M to approx. 600K.
Phase 2 & Official Project completion date.
End of Merging by DR Merge Team due to loss of Federal Funding.
Start/End date of Auto-Merging the PC Duplicates Backlog by M&O.
2 9 16 23 30 7 14 21 28 4 11 18 25 2 9 16 23 30 6 13 20 27 3 10 17 24 3 10 17 24 31 7 14 21 28 5 12 19 26 2 9 16 23 30 7 14 21 28 4 11 18 25 1 8 15
September October November December January February March April May June July August September
2024
2025
MONTH & WEEK START DATE

TOTAL TIR CLIENT COUNT
33,560,762

TOTAL PC COUNT
158,359

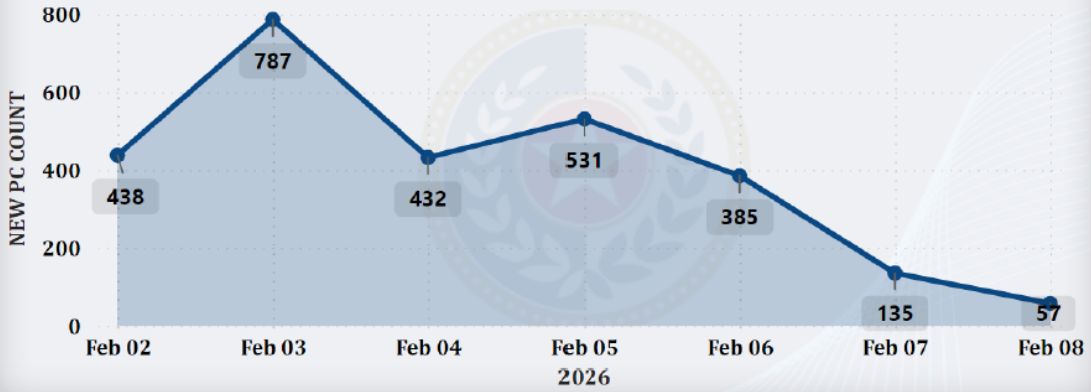
TOTAL PC wMASTER COUNT
310,313

WEEKLY PC COUNT
2,708

PC % of TOTAL CLIENT COUNT
0.47

TOTAL PC COUNT w/IMMS.
60,132 PCs
with
81,239 IMMS.

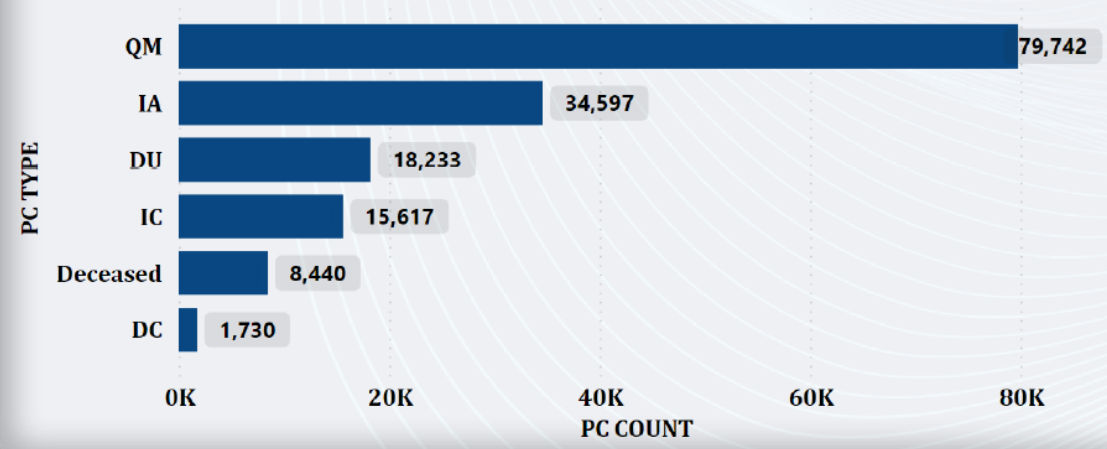
WEEKLY PC COUNT by DAY



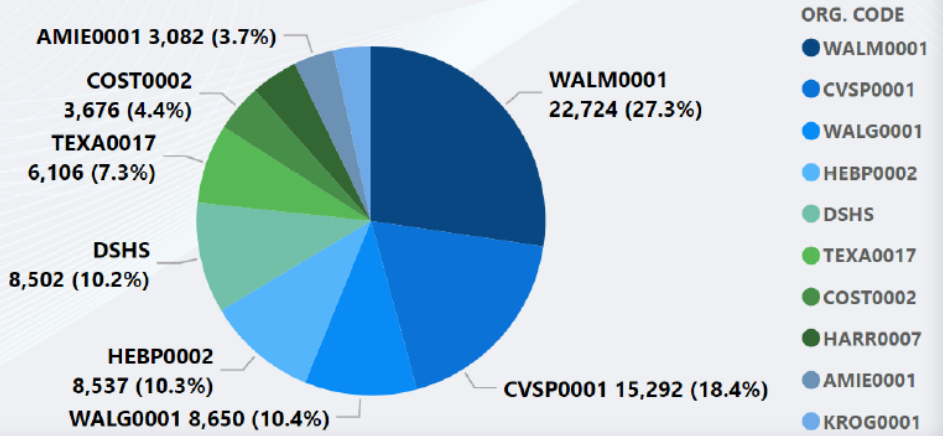
TOTAL PC COUNT by AGE

GREATER THAN 64 = 49,952
27 YRS - 64 YRS = 53,534
18 YRS - 26 YRS = 20,643
13 YRS - 17 YRS = 11,965
5 YRS - 12 YRS = 10,910
6 MON - 4 YRS = 10,803
0 - 5 MON = 555

TOTAL PC COUNT by CLIENT TYPE



TOP 10 - TOTAL PC COUNT by ORG. CODE



Challenges

- Missed requirements
- Identifying & solutioning for fringe cases
- Fixing an issue that creates another
- Implementation of rules that cause downstream issues
- Unexpected consequences to making multiple changes to database tables/columns at the same time
- Loss of project resources

Conclusion

- This project tremendously improved the accuracy of Texas IIS data as client records are now more complete.
- With the new auto-merge tool, the Texas IIS can now identify and merge duplicate client records in near real-time as they come into our registry, allowing more time to focus on records needing manual review.
- The Texas IIS continues to monitor new duplicate client records while creating plans for addressing duplicate immunizations.

Thank you!

Contact Email:
Joe.Williams1@dshs.texas.gov