INFORME TÉCNICO CONSOLIDADO FINAL

Proyecto Calidad de datos

Entidad contratante: Ministerio de Educación Nacional – MEN
Supervisor del contrato: Juan Carlos Bolívar Sandoval
Contratista: 2NV S.A.S.

Versión: v1.0

Lugar: Bogotá D.C., Colombia

Fecha: 16 de marzo de 2026

Contenido del informe

1. Resumen ejecutivo 5
1.1. Resultados por fase 6
1.2. Riesgos activos y mitigaciones 8
2. Objetivo 10
3. Metodología de proyecto 11
3.1. Marco de trabajo 11
3.2. Plan de ejecución 13
4. Fase I – Extracción de datos 15
4.1. Metodología de gestión de fase 15
4.2. Protocolo Técnico de extracción 19
4.3. Recursos y aprovisionamiento de máquina virtual 19
4.4. Informe de extracción de datos 20
4.5. Modelo de Entidad Relación (MER) 21
4.6. Inventario de BOTS y estados 21
4.7. Arquitectura de ZAT Capture 24
4.8. Resultado de extracción 25
4.9. Metodología de aplicación de cálculo de precisión 31
4.10. Ejercicio de medición con muestras 31
4.11. Plan de pruebas 32
4.12. Protocolo de validación cruzada y distribución de casos 32
4.13. Hallazgos 33
5. Fase II – Validación de consistencia de datos 36
5.1. Metodología de gestión de fase 36
5.2. Protocolo Técnico de validación 40
5.3. Mapa de Flujo y Dependencias de Validación de Datos 45
5.4. Modelo de Entidad Relación (MER) 48
5.5. Resultado de Registros Procesados 49
5.6. Metodología de aplicación de cálculo de precisión 50
5.7. Ejercicio de medición con muestras 50
5.8. Hallazgos 51

6. Fase III: Definición de metadatos y migración de la data y archivos con metadatos 55
6.1. Metodología de gestión de fase 55
6.2. Protocolo Técnico de Validación 58
6.3. Modelo de Entidad Relación (MER) 60
6.4. Informe consolidado de metadatos 72
6.5. Hallazgos 75
8. Propuesta de Mejora: Implementación de Visor de Datos para Gestión y Validación de Información. 82
8.1. Descripción General 82
8.2. Arquitectura Propuesta 82
8.3. Requerimientos de Infraestructura (Propuesta) 83
9. Anexos incluir archivos de correcciones 84

Tabla de ilustraciones

I Estructura Azure 12
II Avance Sprint 1 15
III Avance Sprint 2 16
IV Avance Sprint 3 16
V Avance Sprint 4 17
VI Avance Sprint 5 17
VII Avance Sprint 6 18
VIII Diagrama de arquitectura Fase I 19
IX Modelo entidad relación 21
X Diagrama arquitectura conceptual ZATCapture 25
XI Ilustración resultado registro calificado 27
XII Ilustración resultado instituciones acreditadas 28
XIII Ilustración resultado programas acreditados 28
XIV Ilustración resultado conjunto Sin Tipo 29
XV Ilustración resultado inspección y vigilancia 30
XVI Ilustración resultado general de extracción 31
XVII Backlog actual FASE II 39
XVIII Avance Sprint 7 39
XIX Diagrama de arquitectura Fase II 44
XX Flujo y Dependencias de Validación de Datos Fase II - Resoluciones-Auto 45
XXI Flujo y Dependencias de Validación de Datos Fase II - Informes 45
XXII Flujo y Dependencias de Validación de Datos Fase II - Planes de Visita 46

XXIII Flujo y Dependencias de Validación de Datos Fase II - Evaluación de Viabilidad ... 46
XXIV Flujo y Dependencias de Validación de Datos Fase II - Registro Calificado ... 47
XXV Flujo y Dependencias de Validación de Datos Fase II - Programas Acreditados en alta calidad ... 47
XXVI Flujo y Dependencias de Validación de Datos Fase II - Instituciones Acreditadas ... 48
XXVII Tablas intermedias y de llenado ... 49
XXVIII Backlog actual FASE III ... 57
XXIX Avance Sprint 8 ... 58
XXX [dbo].[INS_RESOLUCION_INSPECCION_VIGILANCIA] ... 61
XXXI [dbo].[INS_INFORME_INSPECCION_VIGILANCIA] ... 62
XXXII [dbo].[INS_PLAN_VISITA_INSPECCION_VIGILANCIA] ... 62
XXXIII [dbo].[INS_EVALUACION_VIABILIDAD_INSPECCION_VIGILANCIA] ... 63
XXXIV [dbo].[PRO_HISTORICO_REGISTRO_CALIFICADO] ... 64
XXXV [dbo].[PRO_HISTORICO_MUNICIPIO_RC] ... 65
XXXVI [dbo].[PRO_HISTORICO_ACREDITACION_PROGRAMA] ... 66
XXXVII [dbo].[PRO_HISTORICO_MUNICIPIO_ACREDITACION_PROG] ... 67
XXXVIII [dbo].[INS_HISTORICO_ACREDITACION_INSTITUCION] ... 68
XXXIX [dbo].[INS_HISTORICO_MUNICIPIO_ACREDITACION_INST] ... 69
XL [dbo].[PRO_INFO_RESOLUCION] ... 70
XLI [dbo].[INS_INFO_RESOLUCION] ... 71
XLII MetaData ... 72
XLIII Resolución que MDRA ... 73
XLIV Fecha Resolución ... 74
XLV Tipo de Identificación ... 74
XLVI Número de Identificación ... 75
XLVII Se encuentra PDF ... 75
XLVIII Propuesta de mejora - Arquitectura ... 83

1. Resumen ejecutivo

El presente informe consolida los avances técnicos y operativos del proyecto **"Calidad de Datos"**, incorporando los resultados obtenidos durante las tres fases del proyecto. Esto con el fin de garantizar la integridad, trazabilidad y confiabilidad de la información proveniente de los actos administrativos procesados por el proyecto.

Durante el periodo de reporte, la **Fase I** amplió su alcance mediante la incorporación de **15.927 registros adicionales** a los **45.380 documentos iniciales**, consolidando un universo actualizado de **61.307 casos**. A la fecha, se registra un procesamiento del **100% en la extracción**, en la cual **61.290 documentos fueron aptos para ser procesados equivalentes al 99,97% del total de casos**, soportados en la arquitectura combinada de automatización RPA, OCR e inteligencia artificial generativa de la plataforma **ZAT Capture**. Este ecosistema tecnológico ha permitido garantizar la estandarización y trazabilidad de los datos, asegurando la calidad del proceso de extracción.

La Fase II del proyecto **Calidad de Datos** tuvo como objetivo establecer un modelo integral de validación de consistencia que permitiera contrastar, depurar y asegurar la coherencia de la información extraída en la Fase I frente a las fuentes oficiales del Ministerio, específicamente el backup de la base de datos EDU-SUP.

Durante el periodo de ejecución, se completó el 100 % de las actividades planificadas para esta fase, enfocadas en la preparación técnica, funcional y operativa del proceso de validación. Como resultado, se habilitó una arquitectura robusta basada en Azure Databricks que permitió la consolidación de fuentes, la construcción de una base intermedia canónica y la aplicación sistemática de reglas de transformación y consistencia de datos.

La fase incluyó la definición y operacionalización de reglas de negocio para los procesos de **Inspección y Vigilancia**, sustentadas en los documentos de **Mapeo de Campos por tipo de acto administrativo** y en el **archivo de Reglas de Transformación**, garantizando la correcta armonización, normalización y validación de la información correspondiente a Registros Calificados, Programas Acreditados e Instituciones Acreditadas. Estas reglas permitieron detectar automáticamente diferencias, duplicidades y casos nuevos, clasificando los registros según su estado de consistencia.

Adicionalmente, se dejó habilitado el flujo de validación funcional mediante el diseño y validación de los mockups de la solución frontend, que permitirá a los usuarios del Ministerio visualizar, revisar y corregir los casos identificados con diferencias, asegurando trazabilidad, control de cambios y auditoría de las decisiones realizadas.

En conjunto, la Fase II consolida un modelo escalable y gobernado de validación de datos, dejando preparada la información consistente y trazable para su explotación, migración e

integración en la Fase III, y fortaleciendo la confiabilidad de los datos institucionales utilizados en los procesos de Inspección, Vigilancia y aseguramiento de la calidad en educación superior.

1.1. Resultados por fase

1.1.1. Fase I – Extracción de datos

- Se completó exitosamente la automatización de la captura y procesamiento de metadatos desde el portal **SGDEA** y **SharePoint**, con una precisión superior al **90 %**, validada mediante ejercicios de completitud ponderada.
- Se mantuvo la estabilidad operativa en la infraestructura de Azure, garantizando seguridad, rendimiento y cumplimiento regulatorio.
- Se consolidaron 61.307 casos como base maestra del proyecto, que servirán como insumo directo para los cruces de validación de la Fase II.

Conclusión Fase I:

- La Fase I demuestra un cumplimiento técnico robusto y la madurez de la arquitectura implementada, evidenciando la capacidad del modelo para escalar hacia los procesos de validación y análisis de consistencia.

1.1.2. Fase II – Validación de consistencia de datos

- Se completó la preparación técnica, funcional y operativa de la **Fase II**, alcanzando un avance del **100 %** conforme a lo planificado en el Product Backlog, mediante la habilitación de los insumos, reglas y componentes necesarios para la validación de consistencia de datos.

- Se integró la base de extracción de la **Fase I** con el **backup de la base de datos EDU-SUP**, permitiendo la consolidación de fuentes y la construcción de una **base intermedia canónica** en **Azure Databricks**, definida como punto único de verdad para el proceso de validación.
- Se definieron y aplicaron **reglas de transformación y consistencia** para los procesos de **Inspección y Vigilancia**, sustentadas en:
  - El **archivo de Mapeo de Campos** por tipo de acto administrativo (Registro Calificado, Programas Acreditados e Instituciones Acreditadas).
  - El **archivo de Reglas de Transformación de Inspección y Vigilancia**, que establece validaciones de obligatoriedad, formatos, normalización, control de duplicados y reglas condicionales.
- Se habilitó el procesamiento en **Databricks** para la:
  - Armonización de campos y catálogos.
  - Detección automática de diferencias entre fuentes.

- Clasificación de los registros mediante estados (**COMPLETO, DIFERENCIAS, NUEVO**).
- Generación de una base de casos trazables para validación funcional.

Se procesaron los registros asociados a los principales tipos documentales de Inspección, asegurando disponibilidad y trazabilidad de la información para su revisión y análisis.

Se diseñó y validó funcionalmente la **solución de verificación**, incluyendo la construcción completa de los **mockups del frontend**, que permitirán a los usuarios del Ministerio visualizar, validar y corregir los casos con diferencias, con auditoría de cambios.

Conclusión Fase II:

- La Fase II consolida un modelo sólido de validación de consistencia, soportado en una arquitectura escalable, reglas funcionales claramente definidas y un flujo controlado de revisión por parte del usuario. El proceso deja habilitada la verificación funcional de la información y asegura la consistencia y trazabilidad necesarias para la migración e integración con los sistemas institucionales.

1.1.3. Fase III-Definición de metadatos y migración de la data y archivos con metadatos

- Durante la Fase III, el contratista ejecutó la definición final, normalización y consolidación del modelo de metadatos de negocio y técnicos, asegurando su alineación con las estructuras institucionales del Ministerio de Educación Nacional y con los resultados validados en las fases previas del proyecto.
- Se realizó la preparación integral del entorno de migración, incluyendo la validación de dependencias técnicas, control de versiones de scripts y verificación previa de las estructuras de base de datos, garantizando condiciones controladas para la ejecución de los procesos de carga.
- La migración de la información se ejecutó mediante un esquema ordenado y verificable, que contempló:
- La aplicación secuencial de scripts de definición de datos (DDL) para la adecuación estructural de las tablas.
- La ejecución controlada de scripts de inserción de datos (DML), siguiendo un orden alfabético y lógico previamente definido, asegurando consistencia referencial y coherencia de los registros migrados.

- El proceso cubrió de manera integral los dominios funcionales de **Inspección y Vigilancia, Registro Calificado, Programas Acreditados en Alta Calidad e Instituciones Acreditadas**, incluyendo información histórica y operativa asociada a

resoluciones, evaluaciones, visitas, acreditaciones y su relación con entidades territoriales.

- Se implementaron mecanismos de trazabilidad completos sobre la migración, incluyendo bitácoras de ejecución, registros de eventos y validaciones post-migración, permitiendo identificar el origen, transformación y destino de cada conjunto de datos migrado.
- Se ejecutaron validaciones posteriores a la carga orientadas a confirmar:
- La integridad de los datos migrados.
- La correspondencia entre los registros fuente y destino.
- El cumplimiento del criterio de precisión del 100 % exigido para la migración certificada.
- Como parte de los entregables de la fase, el contratista suministró al Ministerio los insumos técnicos necesarios para la verificación, certificación y auditoría del proceso, incluyendo:
- Scripts de migración organizados y documentados.
- Evidencias de ejecución.
- Registros de auditoría de los procesos automatizados y manuales.
- Material audiovisual de soporte que documenta la correcta ejecución de las actividades.
- La culminación de la Fase III dejó la información completamente migrada, validada y disponible en las estructuras definitivas del modelo de datos institucional, cumpliendo con los criterios de calidad, control, trazabilidad y auditabilidad definidos contractualmente, y habilitando el uso productivo de la información sin requerir prórrogas ni ajustes adicionales.

Conclusión Fase III:

• **La Fase III** permitió cerrar el ciclo técnico y operativo del proyecto mediante la migración certificada y verificable de los datos y metadatos consolidados. El proceso se ejecutó conforme a los lineamientos contractuales y técnicos del Ministerio, garantizando la disponibilidad, integridad, trazabilidad y confiabilidad de la información, y dejando el proyecto en condiciones óptimas para su explotación institucional y cierre contractual.

1.2. Riesgos activos y mitigaciones

Dentro del proceso se identifican los siguientes riesgos, su impacto y la mitigación sugerida

Riesgo identificado	Fase impactada	Impacto	Probabilidad	Mitigación propuesta
Variabilidad en los formatos de origen (documentos no estándar, calidad irregular, estructuras atípicas)	Fase I	Alto	Medio	Ajuste de reglas OCR/ZAT Capture, reprocesamiento dirigida, validación manual para casos críticos.
Diferencias estructurales entre la base de extracción y la base EDU-SUP (nombres de columnas, tipos de datos,	Fase II	Alto	Alto	Ejecución de mesas técnicas de homologación de campos, scripts de normalización en Databricks, pruebas de consistencia previas a los cruces masivos.

estructuras no homologadas)				
Retrasos en la restauración o disponibilidad del backup EDU-SUP	Fase II	Medio	Medio	Coordinación anticipada con equipos TI del MEN, priorización de tareas dependientes y pruebas de carga incremental en ambientes controlados.
Dependencia en el aprovisionamiento de infraestructura en Azure (servicios, accesos, configuraciones, permisos)	Fase II	Medio	Alto	Gestión adelantada de accesos, validación del despliegue de Databricks y ambientes de pruebas, checklist de prerrequisitos técnicos.
Complejidad en la creación y validación de mockups (cambios funcionales solicitados por usuarios del MEN)	Fase II	Medio	Medio	Sesiones recurrentes de revisión funcional, control de versiones de mockups y cierre anticipado del diseño para evitar retrabajos.
Sobrecarga operativa en los procesos de cruzamiento en Databricks (volumen superior a lo estimado)	Fase II	Alto	Bajo	Ajustes de escalabilidad, optimización de pipelines, pruebas de rendimiento con subconjuntos representativos antes de la corrida completa.
Cambios en reglas de negocio durante las mesas de validación	Fase II	Medio	Medio	Documentación centralizada, versionamiento de reglas y validación final antes de ejecución en Databricks.
Riesgo de inconsistencia por registros duplicados o incompletos en la base intermedia	Fase II	Alto	Medio	Implementar validaciones automáticas de integridad, algoritmos de duplicación, y controles previos a la carga final.
Errores en la ejecución secuencial de scripts de migración (DDL/DML) que afecten la integridad de las tablas	Fase III	Alto	Medio	Definición y validación previa del orden de ejecución de scripts, checklist de pre-migración, ejecución en ambiente controlado y respaldo previo de las estructuras existentes.
Fallos técnicos durante la migración (interrupciones de red, sesiones, recursos de BD)	Fase III	Medio	Medio	Plan de contingencia con reintentos controlados, monitoreo de ejecución, bitácora de eventos y posibilidad de Re ejecución parcial por script.
Desalineación entre el modelo de metadatos definido y las estructuras finales en base de datos institucional	Fase III	Alto	Bajo	Revisión técnica conjunta del MER, validación previa de estructuras, pruebas unitarias de scripts DDL y aprobación técnica antes de la carga de datos.
Pérdida de trazabilidad o evidencias insuficientes para auditoría e interventoría	Fase III	Alto	Alto	Generación y entrega de logs de ejecución, bitácoras de eventos, evidencias audiovisuales y documentación técnica de soporte conforme a los requerimientos contractuales.

Cambios de último momento en requerimientos técnicos o criterios de certificación	Fase III	Medio	Bajo	Validación anticipada de criterios de cierre, formalización de entregables, control de alcance y aprobación técnica previa a la certificación.

2. Objetivo

El presente informe tiene como propósito proporcionar una visión clara, concisa y objetiva al **Ministerio de Educación Nacional (MEN)** sobre el avance técnico y operativo del proyecto **"Calidad de Datos"**, correspondiente a la ejecución integral de la **Fase I – Extracción de datos**, la **Fase II – Validación de consistencia de datos** y la **Fase III – Definición de metadatos y migración de la data y archivos con metadatos**, conforme a los lineamientos contractuales y técnicos establecidos.

De manera específica, el documento tiene como objetivos:

- **Fase I – Extracción de datos:**

  Informar sobre la implementación y operación de la arquitectura RPA/IA para la captura automatizada de metadatos desde el portal SGDEA y las carpetas institucionales de SharePoint, así como cuantificar el avance del proceso frente al universo documental actualizado. En esta fase se consolidó un total de **61.307 casos**, alcanzando un **100 % de avance en la extracción**, de los cuales **61.290 documentos (99,97 %)** fueron aptos para ser procesados, superando el umbral técnico mínimo de precisión definido por el MEN.

- **Fase II – Validación de consistencia de datos:**

  Presentar los resultados del proceso de validación de consistencia entre la base de extracción y el respaldo del modelo EDU-SUP, incluyendo la construcción de una base intermedia canónica en **Azure Databricks**, la aplicación de reglas de transformación y consistencia, y la clasificación de los registros según su estado (COMPLETO, DIFERENCIAS, NUEVO). Esta fase permitió identificar, analizar y preparar los casos que requerían validación funcional, asegurando la trazabilidad, integridad y calidad de la información antes de su migración definitiva.

- **Fase III – Definición de metadatos y migración de la data y archivos con metadatos:**

  Describir las actividades orientadas a la definición, estandarización y consolidación del modelo de metadatos, así como la ejecución controlada y verificable de la migración de los datos validados hacia las estructuras definitivas del modelo de datos institucional. Esta fase garantizó la correcta incorporación de la información correspondiente a Inspección y Vigilancia, Registro Calificado, Programas Acreditados en Alta Calidad e Instituciones Acreditadas, asegurando la integridad,

consistencia, trazabilidad y auditabilidad de la información, y dejando los datos disponibles para su uso institucional y paso a producción.

- **Gestión de riesgos y aseguramiento de la calidad:**

  Identificar los principales riesgos, hallazgos y limitaciones que pudieron afectar la calidad del dato o el desempeño del proceso en las distintas fases, así como documentar las acciones de mitigación implementadas, con el fin de asegurar el cumplimiento de los criterios técnicos, de calidad de datos y de control exigidos por el **MEN** y soportar el cierre contractual del proyecto.

3. Metodología de proyecto

3.1. Marco de trabajo

El proceso de extracción de datos se desarrolla bajo una **metodología ágil y dinámica**, orientada a garantizar entregas funcionales continuas, control de calidad y trazabilidad en cada iteración del proyecto.

Esta aproximación se implementa mediante **Sprints quincenales**, que permiten al equipo entregar incrementos medibles del proceso de extracción y reaccionar oportunamente ante los cambios o hallazgos técnicos identificados durante la ejecución.

3.1.1. Estructura de planificación en Azure DevOps

El progreso del proyecto se estructura en tres niveles jerárquicos de planificación:

- **Features:** representan los objetivos macro de cada fase, brindando una visión estratégica de alto nivel.
- **Product Backlogs:** agrupan las actividades específicas que contribuyen al logro de cada Feature.
- **Tasks:** detallan las acciones operativas ejecutadas dentro de cada Sprint.

Esta estructura ágil ha permitido mantener una ejecución continua, priorizar la resolución de incidencias en tiempo real y garantizar la alineación del proyecto con el cronograma técnico aprobado por el **Ministerio de Educación Nacional**.

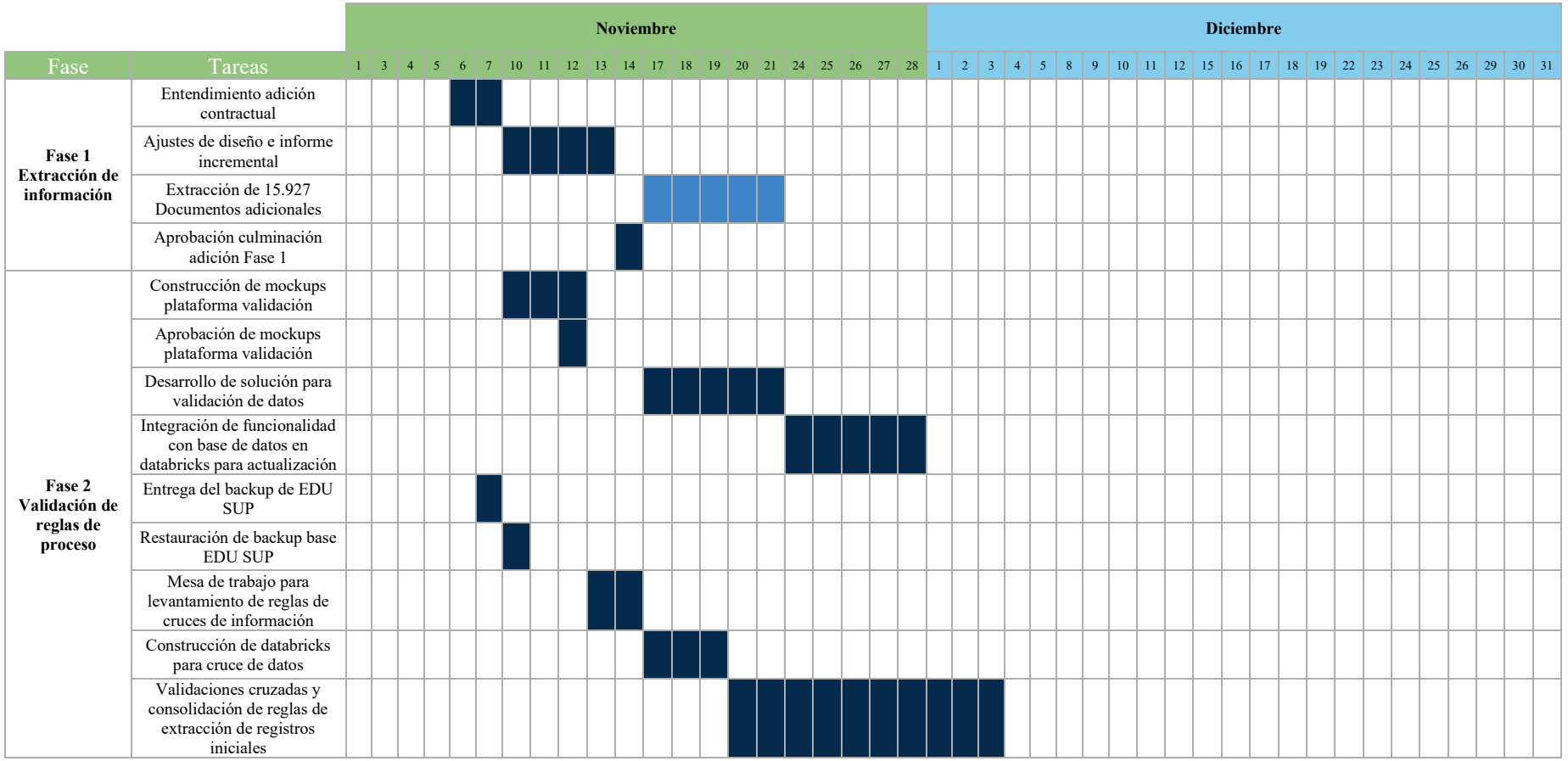
Work Item Type	Title
Epic	Calidad de datos
Feature	Fase I Extracción de datos
Product Backlo...	Entendimiento
Task	Kickoff del proyecto
Task	Firmar acuerdos
Task	Verificar acceso a usuarios recibidos
Task	Verificar acceso al portal SGDEA
Task	Participar en primera reunión de entendimiento
Task	Participar en segunda reunión de entendimiento
Product Backlo...	FASE I Extracción
Feature	Fase II Validación consisten de de data
Feature	Fase III Definición de Metadatos y migración de data

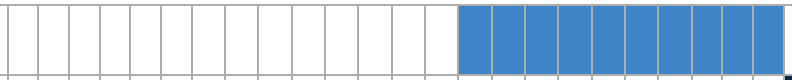
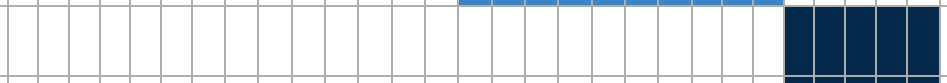
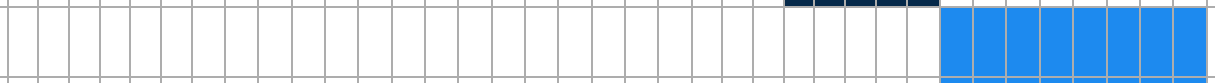
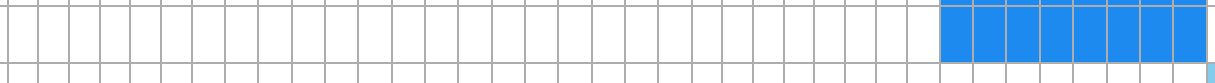
I Estructura Azure

3.2. Plan de ejecución

Para el plan de trabajo actual se incluyen las actividades de la adición de los 15.927 documentos adicionales por validar, teniendo a continuación el plan de trabajo de las actividades establecidas hasta el momento.

Se considera el avance del proyecto realizando en paralelo el desarrollo del módulo de validación y el cruce de información de las fuentes EDU-SUP y la base de extracción,



	Validaciones cruzadas y consolidación de reglas de extracción de adición	
	Desarrollo de módulo de generación de Script de actualización de información	
	Construcción de informe para validación de diferencias metadatos SGDEA vs Extracción	
Fase 3 Validación de reglas de proceso	Control de versiones en scripts ETL y reglas para actualización de metadatos.	
	Aprobación funcional de los lotes procesados adicionales	
Estabilización	Estabilización y verificación	

4. Fase I – Extracción de datos

4.1. Metodología de gestión de fase

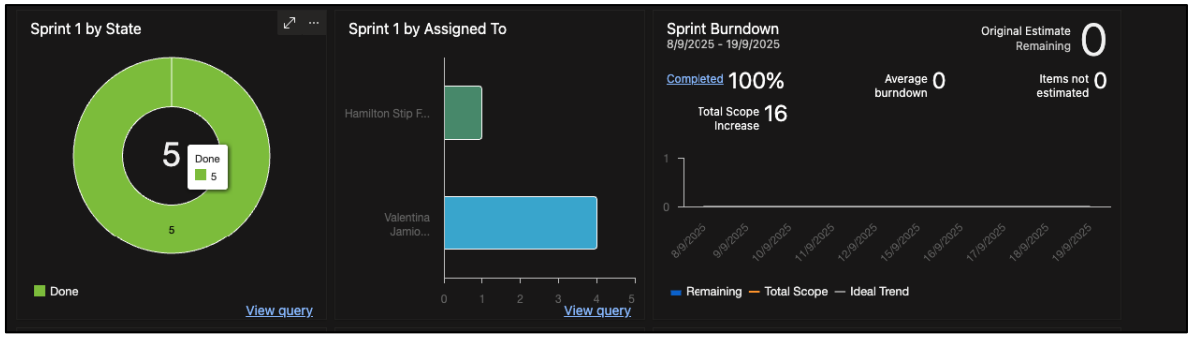
Durante la **Fase I**, se definieron dos *Product Backlogs* principales:

1. **Entendimiento:** iniciado con la reunión de *Kick-off* y sesiones colaborativas entre los equipos de RPA y ZAT, enfocadas en la comprensión integral del proceso de extracción, definición de campos y reglas de negocio.
2. **Fase de extracción:** contempla las actividades de desarrollo, pruebas y ejecución de los *BOTS RPA*, junto con las tareas de ZAT Capture, orientadas a la lectura, clasificación y análisis de calidad de los documentos PDF provenientes del SGDEA y SharePoint.

A continuación, se observa la trazabilidad del progreso por estado, responsable y desempeño general del equipo:

Durante el periodo analizado se evidencian seis sprints en seguimiento, reflejando el progreso general de las tareas asignadas, la distribución de responsables y el cumplimiento de objetivos planificados, en total para la fase 1 del proyecto se ejecutaron un total de 116 tareas

Sprint 1



II Avance Sprint 1

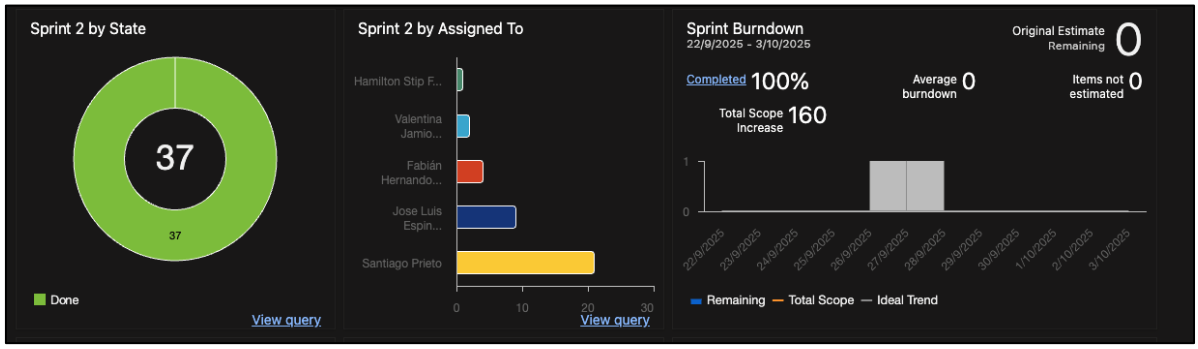
II Avance Sprint 1

Estado general: Todas las 5 tareas planificadas fueron completadas exitosamente, alcanzando un 100 % de cumplimiento.

Asignación: Las actividades estuvieron principalmente a cargo del equipo de diseño, consolidando el cierre total del sprint.

Burndown: La curva muestra una reducción completa de pendientes, sin desviaciones frente a la estimación inicial (cero tareas restantes)

Sprint 2



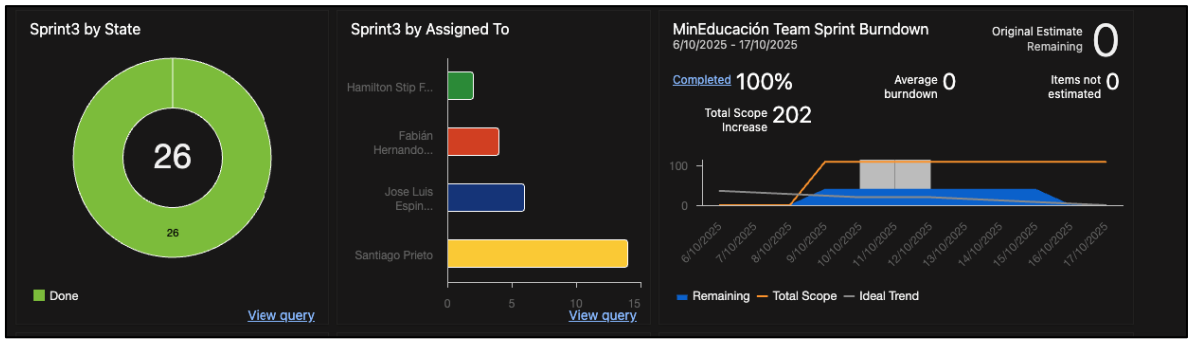
III Avance Sprint 2

Estado general: Se desarrollaron 37 tareas, de las cuales todas fueron finalizadas dentro del tiempo estimado, evidenciando nuevamente un cumplimiento del 100 %.

Asignación: La mayor carga de trabajo fue gestionada por el equipo de extracción de ZATCapture y RPA, manteniendo equilibrio entre responsables y una ejecución alineada con la planificación.

Burndown: Se observa una tendencia estable sin acumulación de backlog, lo que indica eficiencia y buena capacidad de respuesta del equipo.

Sprint 3



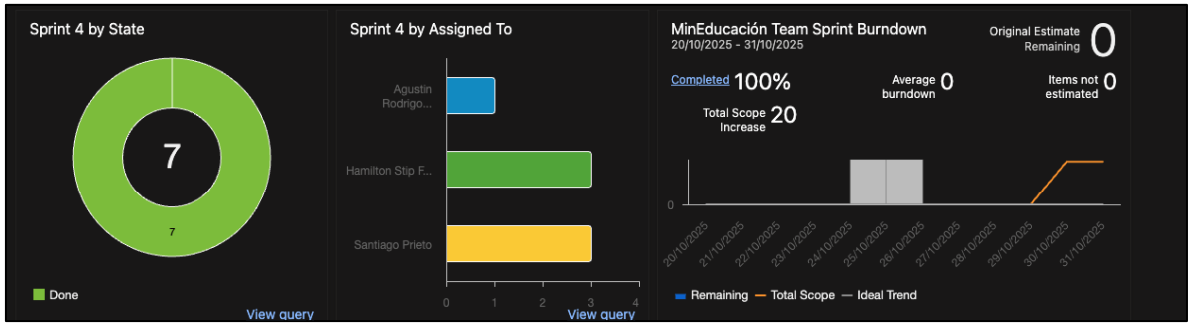
IV Avance Sprint 3

Estado general: El sprint reporta un total de 26 tareas, todas 100 % completadas, sin actividades en progreso ni pendientes por iniciar. Este cierre total refleja un desempeño óptimo del equipo frente a la planificación establecida.

Asignación: Las tareas fueron distribuidas entre los integrantes del equipo, destacándose la participación equilibrada de los responsables de extracción de ZATCapture y RPA. La ejecución evidencia una capacidad sostenida de respuesta y una asignación efectiva de cargas de trabajo.

Burndown: La gráfica de burndown muestra una ejecución lineal y controlada, con una reducción continua de las tareas hasta alcanzar cero pendientes antes del cierre del sprint. Esta tendencia indica un ritmo adecuado, una estimación precisa y un control eficiente de las actividades planificadas.

Sprint 4



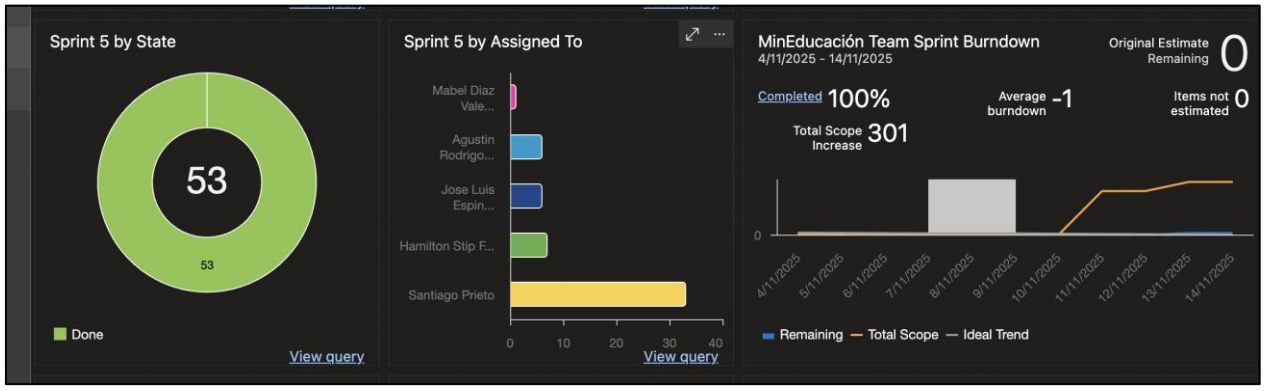
V Avance Sprint 4

Estado general: El sprint registra un total de 7 tareas, todas 100 % completadas, sin actividades en progreso ni pendientes por iniciar. El cumplimiento total del alcance confirma una ejecución controlada y alineada con la planificación técnica establecida para este ciclo.

Asignación: Las tareas fueron distribuidas entre los distintos integrantes del equipo, con participación activa de los responsables de extracción, análisis y soporte técnico. La distribución observada refleja un balance adecuado de carga y una ejecución coordinada entre los miembros del equipo, garantizando tiempos de respuesta eficientes.

Burndown: La gráfica de burndown evidencia una reducción constante del backlog hasta llegar a cero tareas pendientes, completando el objetivo del sprint sin desviaciones frente a la estimación inicial. Esta tendencia demuestra un control efectivo del ritmo de trabajo y una correcta gestión de los compromisos planificados.

Sprint 5



VI Avance Sprint 5

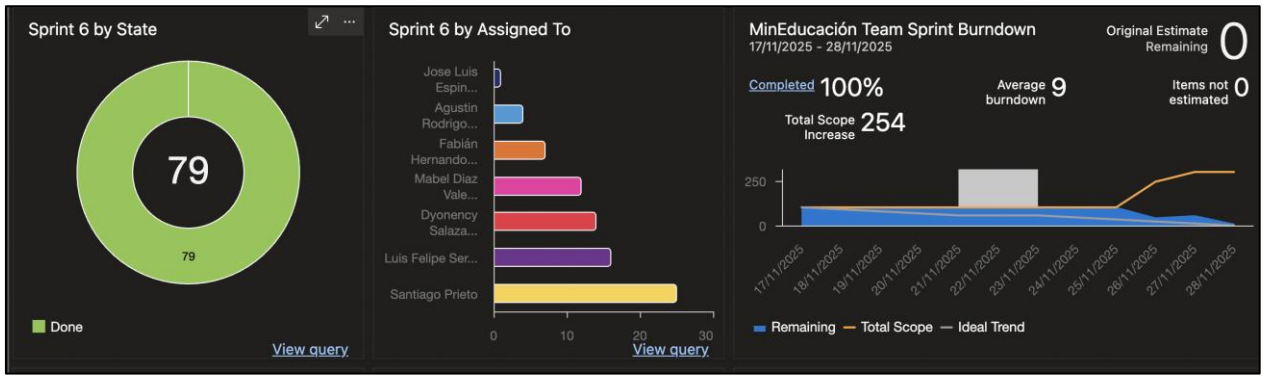
Estado general: El Sprint 5 registra un total de 53 tareas, todas ellas completadas al 100 %, sin elementos en progreso ni pendientes por iniciar. Este resultado refleja una ejecución

sobresaliente del sprint, evidenciando una alta capacidad de cumplimiento y una adecuada planificación durante el periodo evaluado.

Asignación: La distribución de actividades muestra una participación significativa del equipo, con un volumen mayor de tareas gestionado por Santiago Prieto, seguido por contribuciones de otros integrantes como Hamilton, José Luis, Agustín y Mabel. Esta distribución permitió mantener estabilidad operativa y asegurar la continuidad del flujo de trabajo requerido para cumplir con los objetivos del sprint.

Burndown: La gráfica de burndown confirma un cierre del 100 %, evidenciando la reducción total del backlog y una gestión controlada del alcance, que finalmente ascendió a 301 puntos de esfuerzo. La tendencia del burndown se mantuvo alineada con el comportamiento esperado, sin desviaciones relevantes y demostrando una efectiva utilización de la capacidad del equipo.

Sprint 6



VII Avance Sprint 6

Estado general: El Sprint 6 registra un total de 79 tareas, todas ellas completadas al 100 %, sin actividades en progreso ni pendientes por iniciar. Este desempeño evidencia una ejecución altamente efectiva y una clara capacidad del equipo para responder al incremento de alcance del ciclo, manteniendo la calidad y estabilidad operativa.

Asignación: La distribución de tareas refleja una participación balanceada entre los miembros del equipo, con un mayor volumen gestionado por Santiago Prieto, seguido de contribuciones relevantes de Luis Felipe, Dyonency, Mabel, Hernando y otros integrantes. Esta colaboración permitió sostener un flujo de trabajo continuo aun con el aumento de la carga operativa, garantizando la entrega oportuna de los compromisos.

Burndown: La gráfica de burndown confirma un cumplimiento del 100 %, mostrando una disminución progresiva de las tareas hasta alcanzar cero elementos restantes. El alcance total del sprint llegó a 254 puntos de esfuerzo, manteniéndose dentro de un comportamiento

controlado y alineado con la tendencia ideal. No se evidencian desviaciones significativas, lo que refleja una adecuada planificación y una gestión efectiva de la capacidad disponible.

4.2. Protocolo Técnico de extracción

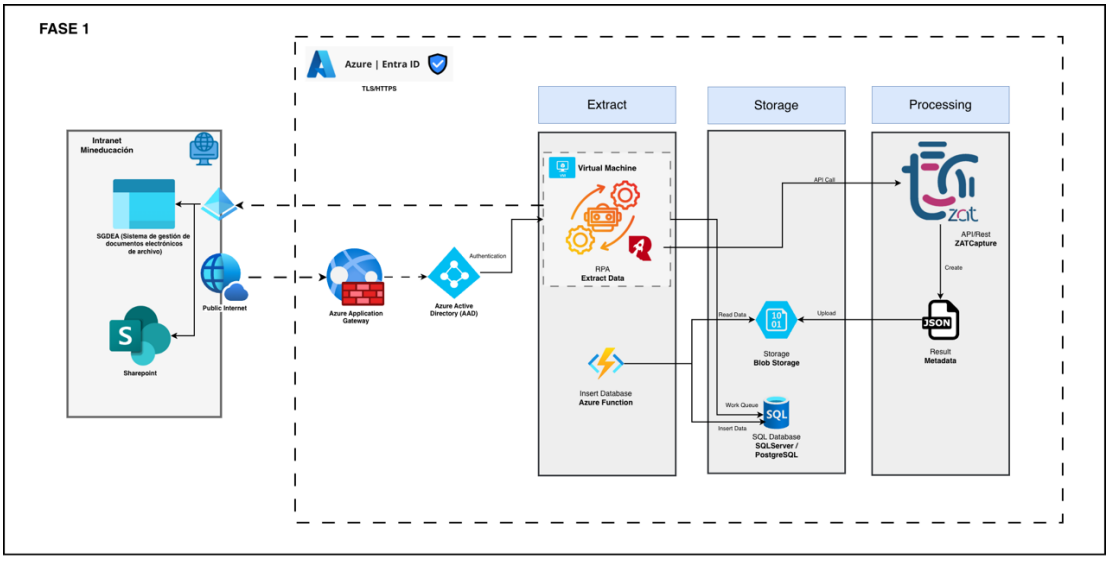
4.2.1. Arquitectura de extracción

La arquitectura de Fase I está diseñada para automatizar el ciclo completo de extracción, transformación y procesamiento de metadatos institucionales, integrando servicios de Microsoft Azure, herramientas de RPA y componentes de inteligencia artificial.

El flujo inicia desde la base de datos y culmina con la entrega estructurada de resultados en formato JSON.

1. **Consulta en base de datos:** Fase I del proceso de extracción, almacenamiento y procesamiento de datos en un entorno híbrido que conecta la intranet del Ministerio de Educación con servicios en la nube de Microsoft Azure. Desde plataformas internas como SGDEA y SharePoint, los datos se acceden a través de internet mediante un Azure Application Gateway y se autentican con Azure Active Directory (AAD).
2. **Extracción de datos:** luego, una máquina virtual con RPA en Rocketbot que extrae los datos, que se almacenan temporalmente en Blob Storage y se insertan en una base de datos SQL Server mediante funciones de Azure Function.

Resultado de extracción: finalmente, el sistema ZAT Capture procesa los datos mediante API REST y genera un resultado en formato JSON con metadatos procesados.



VIII Diagrama de arquitectura Fase I

4.3. Recursos y aprovisionamiento de máquina virtual

Con el fin de garantizar la eficiencia en el procesamiento, análisis y almacenamiento de documentos extraídos mediante flujos automatizados, se dispuso una máquina virtual con especificaciones técnicas adecuadas para soportar operaciones intensivas en lectura y clasificación de archivos PDF. La siguiente tabla detalla los recursos asignados a dicho

entorno, incluyendo características del procesador, capacidad de memoria, unidades de almacenamiento locales y rutas de red utilizadas para el acceso centralizado a los archivos. Este aprovisionamiento responde a los requerimientos funcionales del sistema de extracción implementado, asegurando estabilidad, velocidad de respuesta y trazabilidad de los datos procesados.

Parámetro	Detalle
Nombre del dispositivo (VM)	2nvmindedudevvm0
Procesador	Intel(R) Xeon(R) Platinum 8272CL CPU @ 2.60GHz
Memoria instalada (RAM)	64.0 GB
Sistema	64-bit, x64-based
Unidades locales	C: (250 GB, ~2.7 GB libres), D: Temporary (63.9 GB, 57.6 GB libres), E: disk01 (1 TB, 953 GB libres)
Recurso de red para PDFs	\\2nvmindeduevaldls.file.core.windows.net\rpafiles (montaje 'rpafiles')
Uso	Descarga, análisis y almacenamiento temporal y centralizado de PDFs

Adicionalmente, la infraestructura utilizada en **Microsoft Azure** cumple con las más estrictas normas internacionales en materia de **seguridad de la información, protección de datos personales, disponibilidad y cumplimiento regulatorio**.

De acuerdo con las certificaciones vigentes detalladas en el **Anexo 1 – "Regulaciones y certificaciones de Azure"**, la plataforma cuenta con acreditaciones como **ISO/IEC 27001, ISO/IEC 27017, ISO/IEC 27018, SOC 1, SOC 2 y SOC 3**, así como con conformidad con el **Reglamento General de Protección de Datos (GDPR)**, el **Esquema Nacional de Seguridad (ENS)** y las políticas de **confidencialidad y gestión de riesgos** de Microsoft Trust Center.

Estas certificaciones garantizan que todos los servicios empleados para el procesamiento y almacenamiento de información del proyecto —incluyendo máquinas virtuales, bases de datos, almacenamiento en Blob Storage y funciones de Azure— operan bajo protocolos de **cifrado de datos en tránsito y en reposo**, **redundancia geográfica**, **control de acceso basado en identidad (Azure Active Directory)** y **monitoreo continuo de seguridad**, asegurando la **confidencialidad, integridad y disponibilidad** de los datos institucionales del **Ministerio de Educación Nacional**.

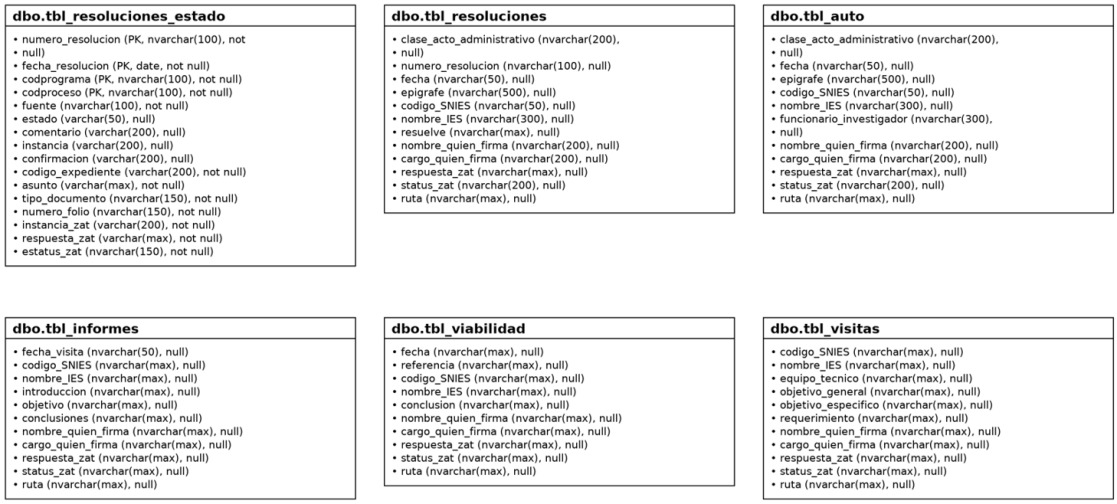
4.4. Informe de extracción de datos

Durante el proceso de extracción, se empleó la plataforma ZAT Capture, la cual integra tecnología de Reconocimiento Óptico de Caracteres (OCR) para la lectura automatizada de documentos en formato PDF. Estos archivos fueron previamente obtenidos mediante flujos

de automatización robótica de procesos (RPA) desde las fuentes oficiales del portal SGDEA y repositorios en SharePoint. Gracias a esta integración, fue posible capturar información estructurada y no estructurada con alta precisión, facilitando su clasificación, trazabilidad y análisis dentro del entorno digital del proyecto.

4.5. Modelo de Entidad Relación (MER)

El diseño del Modelo Entidad-Relación (MER) define la estructura de la base de datos necesaria para la gestión de la información capturada. Este modelo se compone de seis entidades o tablas principales: dbo.tbl_resoluciones_estado,dbo.tbl_resoluciones_auto, dbo.tbl_acto_administrativo, dbo.tbl_informes, dbo.tbl_viabilidad, y dbo.tbl_visitas. Cada tabla se encuentra definida de manera autónoma para capturar y estructurar un conjunto de datos específicos, incluyendo campos como números de resolución, fechas, códigos de identificación (IES/SNIES), responsables y las respuestas automatizadas del sistema ZAT. Es relevante señalar que, para este proceso, no se generaron relaciones explícitas entre las distintas entidades. El modelo se enfoca en asegurar la integridad individual y la definición precisa de los atributos dentro de cada tabla, lo cual facilita la clasificación, el almacenamiento y la trazabilidad independiente de la información clave del proyecto.



IX Modelo entidad relación

4.6. Inventario de BOTS y estados

A continuación, se presenta el inventario de robots ejecutados durante la Fase I, incluyendo tipo, propósito, ventana de ejecución, estado final y observaciones.

ID	Tipo	Descripción	Ventana de ejecución	Estado	Observaciones
BOT-01	Descarga	Descarga De PDF del portal SGDEA	Continuo	OK	Sin novedad
BOT-02	Descarga	Descarga De PDF del portal SGDEA	Continuo	OK	Sin novedad
BOT-03	Análisis	Envió de PDF a Zat y guardado de datos	Continuo	OK	Sin novedad

BOT-04	Análisis	Envio de PDF a Zat y guardado de datos	Continuo	OK	Sin novedad
BOT-05	Descarga	Descarga de información de SharePoint inspección y vigilancia	Continuo	OK	Sin novedad
BOT-06	Descarga	Descarga de información de SharePoint actos administrativos	Continuo	OK	Sin novedad

4.6.1. Pruebas de los BTS/BOTS

CP01: Descarga exitosa de PDF desde SGDEA

Precondiciones: Registro válido en SGDEA con enlace directo a archivo PDF accesible; credenciales y redes disponibles.
Pasos:
1)Consultar SGDEA por número/ID.
2) Validar tipo de archivo.
3) Descargar y guardar en ruta de trabajo.
4) Registrar estado en cola/log.

Resultado esperado: Archivo PDF descargado y disponible en ruta; estado 'OK'. Hash/tamaño registrado.
Resultado obtenido: Pendiente de ejecución / OK (indicar fecha y BOT).

CP02: No existe registro en SGDEA

Precondiciones: Consulta con número/ID inexistente en SGDEA.
Pasos: 1) Consultar SGDEA.
2) Evaluar respuesta 'sin resultados'.
3) Registrar evento y marcar trabajo como 'No encontrado'.
Resultado esperado: No se descarga archivo; estado 'NO_ENCONTRADO'; sin reintentos automáticos.
Resultado obtenido: Pendiente de ejecución / Resultado conforme.

CP 03: Existe registro, pero el recurso no es PDF
Precondiciones: Registro en SGDEA referencia a archivo no-PDF (p. ej., DOCX/JPG).
Pasos:
1) Consultar SGDEA.
2) Validar content-type/extension.
3) Rechazar descarga si no es PDF.
4) Registrar evento y continuar con siguiente trabajo.
Resultado esperado: No se descarga archivo; estado 'TIPO_INVALIDO'; se conserva traza.
Resultado obtenido: Pendiente de ejecución / Resultado conforme.

CP 04: análisis exitoso en ZAT (extracción de datos)
Precondiciones: PDF válido en repositorio de trabajo; motor ZAT disponible; reglas de extracción cargadas.
Pasos:
1) Encolar PDF para análisis.
2) Ejecutar motor ZAT.
3) Validar campos extraídos y esquema.
4) Persistir en DB (staging/final) y marcar 'OK'.
Resultado esperado: Campos extraídos completos según plantilla; inserción correcta en DB; estado 'OK_ANALISIS'.
Resultado obtenido: Pendiente de ejecución / OK (indicar fecha y BOT).

CP 05: Fallo de análisis en ZAT con reintento automático
Precondiciones: PDF válido, pero con patrón complejo; política de reintentos habilitada (p. ej., 3 intentos backoff).
Pasos:
1) Encolar PDF para análisis.
2) Fallo controlado en ZAT.
3) Activar reintento automático según política.
4) Si persiste, registrar 'FALLA' y mover a cola de revisión.
Resultado esperado: Reintentos ejecutados; si éxito en reintento → 'OK_ANALISIS'; si falla definitiva → 'ERROR_ANALISIS' con detalle y rastro.
Resultado obtenido: Pendiente de ejecución / Detallar número de reintentos y resultado final.

CP 06: Descarga de información SharePoint inspección y vigilancia
Precondiciones: PDF válido, pero con patrón complejo; política de reintentos habilitada (p. ej., 3 intentos backoff).
Pasos:
1) Consultar información sharepoint.
2) Validar tipo de archivo.
3) Descargar y guardar en ruta de trabajo.
4) Registrar estado en cola/log.
5) Envió información ZATCapture
Resultado esperado: Reintentos ejecutados; si éxito en reintento → 'OK_ANALISIS'; si falla definitiva → 'ERROR_ANALISIS' con detalle y rastro.
Resultado obtenido: Pendiente de ejecución / Detallar número de reintentos y resultado final.

CP 07: Descarga de información SharePoint actos administrativos
Precondiciones: PDF válido, pero con patrón complejo; política de reintentos habilitada (p. ej., 3 intentos backoff).
Pasos:

1)Consultar información sharepoint.
2) Validar tipo de archivo.
3) Descargar y guardar en ruta de trabajo.
4) Registrar estado en cola/log.
5) Envió información ZATCapture
Resultado esperado: Reintentos ejecutados; si éxito en reintento → 'OK_ANALISIS'; si falla definitiva → 'ERROR_ANALISIS' con detalle y rastro.
Resultado obtenido: Pendiente de ejecución / Detallar número de reintentos y resultado final.

4.7. Arquitectura de ZAT Capture

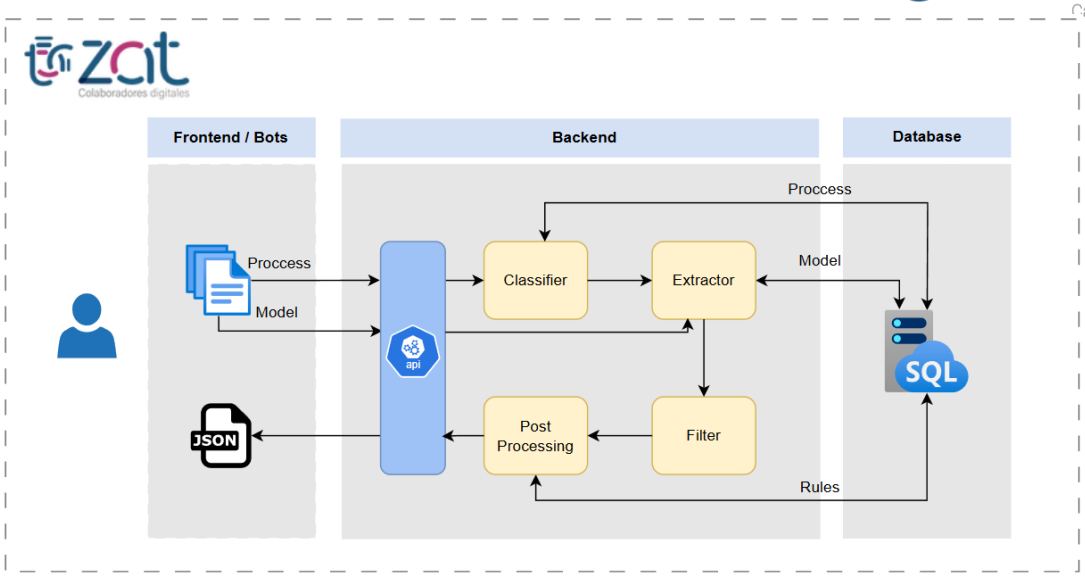
Flujo de procesamiento de datos que inicia en el **frontend**, donde usuarios o bots envían información en formato **JSON**. Esta entrada es recibida por el **backend**, que opera como núcleo de procesamiento inteligente. El primer módulo en este flujo es el **clasificador**, encargado de identificar el tipo de contenido y dirigirlo al **extractor**, que realiza la lectura y segmentación de los datos relevantes.

Posteriormente, los datos extraídos son enviados al **filtro**, el cual aplica reglas específicas para depurar o validar la información antes de su paso al módulo de **posprocesamiento**. Este último realiza ajustes finales, como normalización o enriquecimiento de datos, y puede retroalimentar al clasificador para mejorar la precisión del sistema en ciclos posteriores.

La arquitectura mantiene una interacción constante con una **base de datos SQL**, que almacena tres componentes clave:

- **Modelos** generados por el clasificador
- **Procesos** definidos por el extractor
- **Reglas** utilizadas por el filtro

Estos elementos permiten que el sistema se adapte dinámicamente a nuevos tipos de documentos o fuentes, como los PDFs extraídos mediante RPA desde portales como **SGDEA** y **SharePoint**, y procesados con tecnología **OCR** a través de plataformas como **ZAT Capture**.



X Diagrama arquitectura conceptual ZATCapture

4.8. Resultado de extracción

El proceso de extracción de datos, enmarcado en la primera etapa de la **Fase I – Extracción de datos** del proyecto, reporta un avance del **100 %** de actos administrativos planificados. Tras la ampliación del universo documental que pasó de 45.380 a **61.307 documentos** con la adición de 15.927 casos, a la fecha se tiene la totalidad de registros extraídos y procesados.

La información capturada se distribuye en las siguientes categorías: **Registro Calificado** (el componente más extenso), **Acreditación de Alta Calidad** para Programas e Instituciones, y los actos de **Inspección y Vigilancia (I. y V.)**, que incluyen documentos Sancionatorios y Medidas Preventivas. Este progreso asegura una base sólida de datos para la clasificación, el cruce con la base EDU-SUP y el análisis detallado en las etapas subsiguientes del proyecto.

Acto administrativo	Cantidad	Adición
Registro calificado	41.717	11.033
Acreditación de alta calidad Programas	2.360	3.937
Acreditación de alta calidad Instituciones	199	0
Inspección y vigilancia	1.104	851
Documento sin clasificación	0	106

Teniendo en cuenta la distribución de casos, a continuación, se presenta el detalle de la extracción:

Acto administrativo	Base	Total casos recibidos		Casos aptos para ser procesados		Casos NO aptos para ser procesados (Duplicados o Sin llave para identificación)		Casos aptos Encontrados en SGDA o SharePoint (PDF)	Casos aptos NO Encontrados en SGDA o SharePoint (PDF)	Casos aptos Encontrados en SGDA o SharePoint SIN CALIDAD	Casos aptos Encontrados en SGDA o SharePoint CON CALIDAD
		Por Base	Total	Por Base	Total	Por Base	Total				
Registro Calificado	cat_rc_e1_rc	41717	52750	41714	52738	3	12	50411	2327	0	50411
	Base_RC-E2	11033		11024		9					
Instituciones Acreditadas	cat_aac_les	199		197		2		194	3	0	194
Programas acreditados	cat_aac_prg_el	2360	6297	2360	6297	0	0	6100	197	0	6100
	Base AAC Prg-E2	3937		3937		0					
Sin tipo	Base_0106SinTipo	106		104		2		94	10	87	7
TOTAL		**59352**		**59336**		**16**		**56799**	**2537**	**87**	**56712**

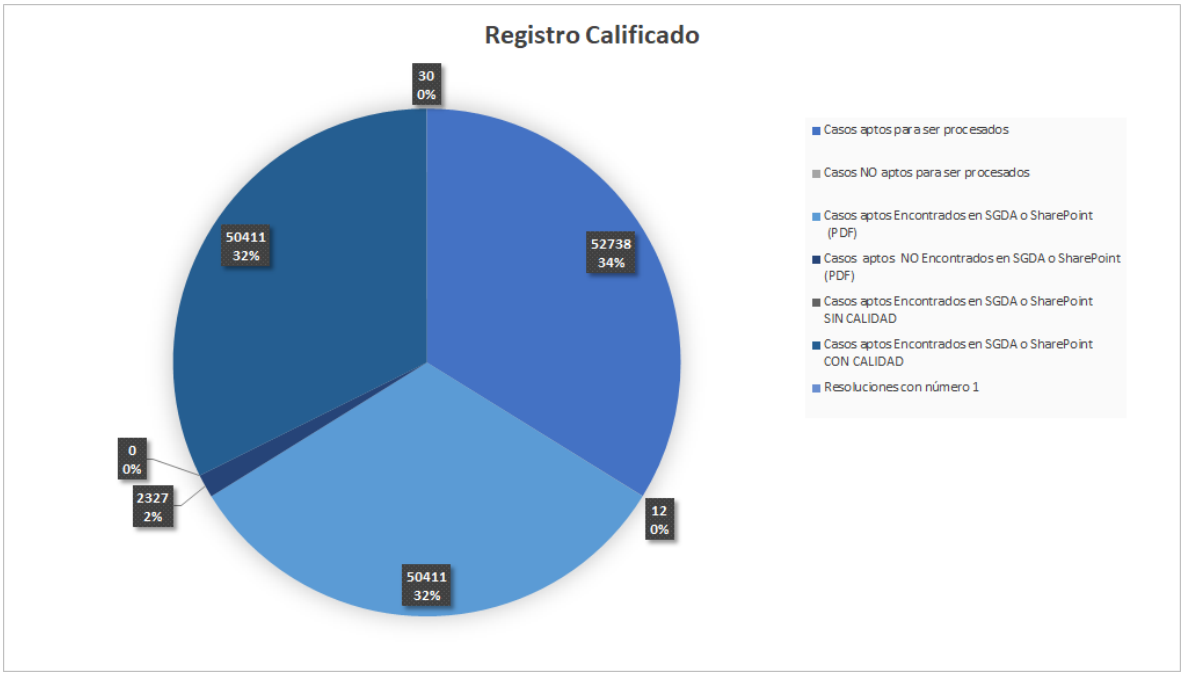
Acto administrativo	Total casos recibidos	Casos procesados exitosamente	Casos NO procesados (Archivo corrupto)
Inspección y vigilancia	1955	1954	1

TOTAL RECIBIDOS (Todos los Actos Administrativos)	61307
TOTAL APTOS PARA SER PROCESADOS (Todos los Actos Administrativos)	61290
% de representación APTOS PARA SER PROCESADOS vs alcance	99,97

Base_0106SinTipo			
Acto administrativo	Casos aptos Encontrados en SGDA o SharePoint (PDF)	Casos aptos Encontrados en SGDA o SharePoint SIN CALIDAD	Casos aptos Encontrados en SGDA o SharePoint CON CALIDAD
Instituciones Acreditadas	0	0	0
Programas acreditados	7	0	7
Registro Calificado	87	87	0
Total	**94**	87	7

Para el acto administrativo de **Registro Calificado**, se recibieron 52.750 registros (41.717 base inicial y 11.033 de adición), de los cuales 52.738 fueron aptos para procesamiento, evidenciando únicamente 12 casos no aptos por duplicidad o ausencia de llave de identificación.

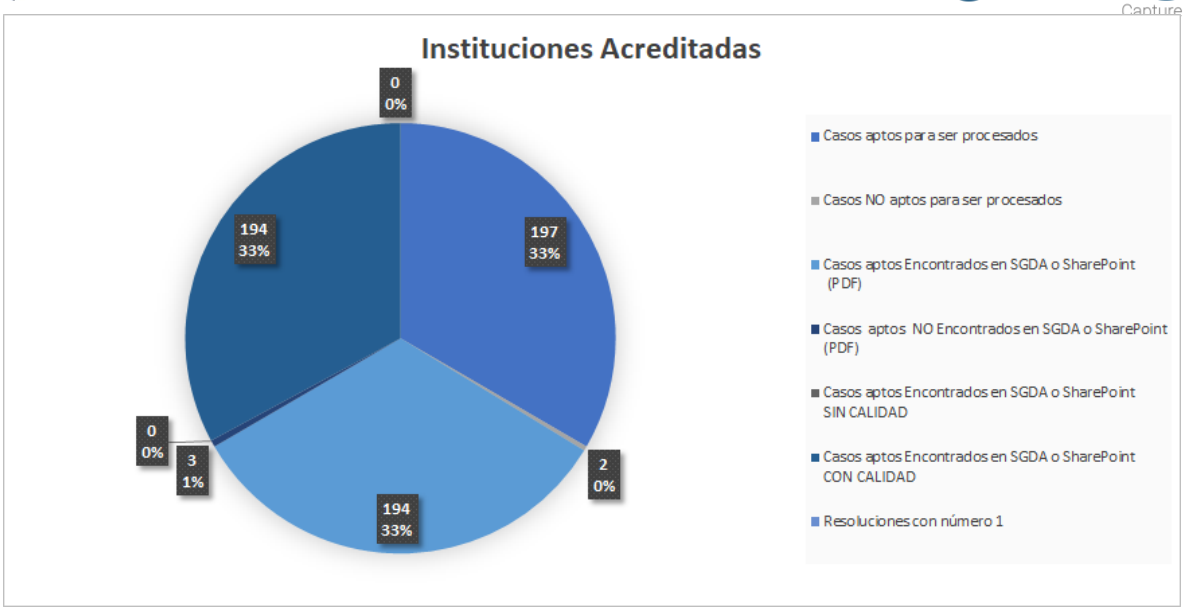
Del total apto, se localizaron exitosamente 50.411 documentos, mientras que 2.327 registros no fueron encontrados en SGDEA o SharePoint y no se identificaron casos sin calidad.



XI Ilustración resultado registro calificado

Para **Instituciones Acreditadas**, se gestionaron 199 registros totales, de los cuales 197 fueron aptos para procesamiento, identificándose solo 2 casos no aptos.

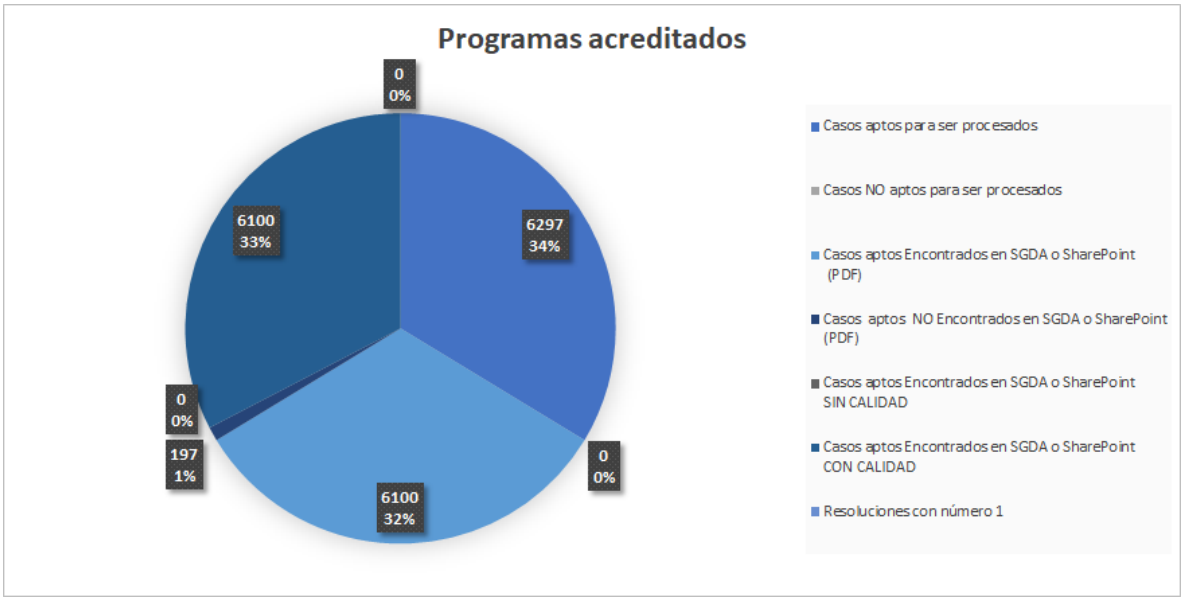
Se logró ubicar 194 documentos, mientras que 3 registros no fueron encontrados en los sistemas fuente y no se registraron casos sin calidad.



XII Ilustración resultado instituciones acreditadas

En **Programas Acreditados**, se recibieron 6.297 registros (2.360 base y 3.937 adición), todos aptos para procesamiento (0% no aptos).

Se localizaron 6.100 documentos, mientras que 197 registros no fueron encontrados en SGDEA o SharePoint y no se presentaron casos sin calidad.



XIII Ilustración resultado programas acreditados

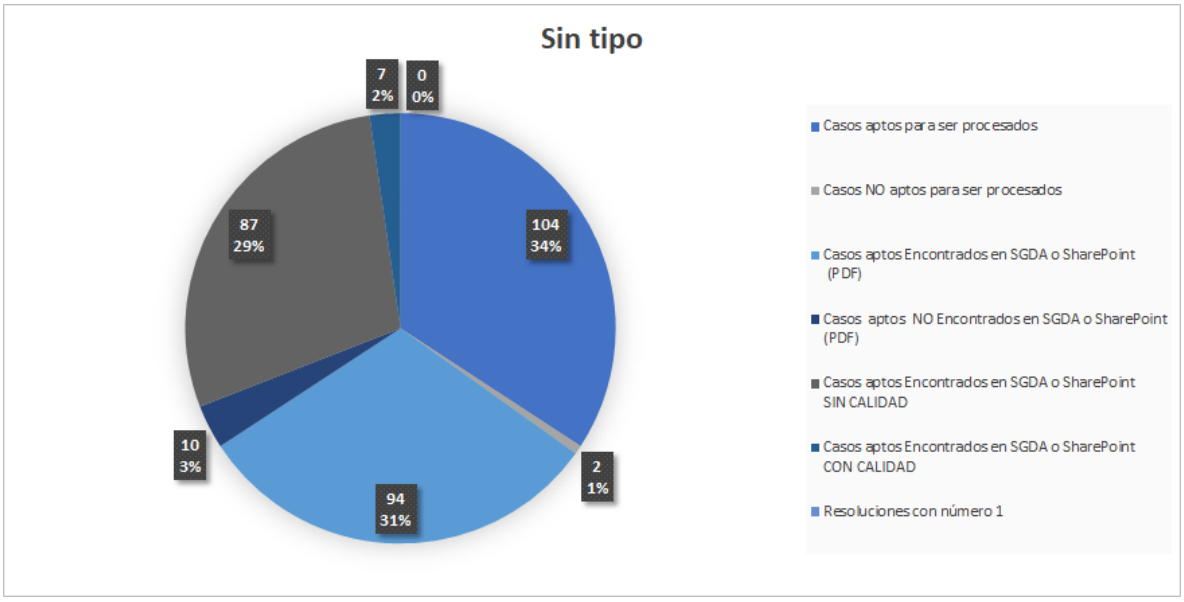
En el conjunto **Sin Tipo** (Base_0106SinTipo) se identificaron 94 registros aptos y encontrados en los repositorios (SGDEA o SharePoint). De estos, 87 casos fueron clasificados como sin calidad, mientras que 7 registros cumplen con los criterios de calidad definidos.

En el análisis por acto administrativo:

Registro Calificado concentra 87 registros encontrados, los cuales no cumplen criterios de calidad documental.

Programas Acreditados registra 7 casos, todos con calidad validada.

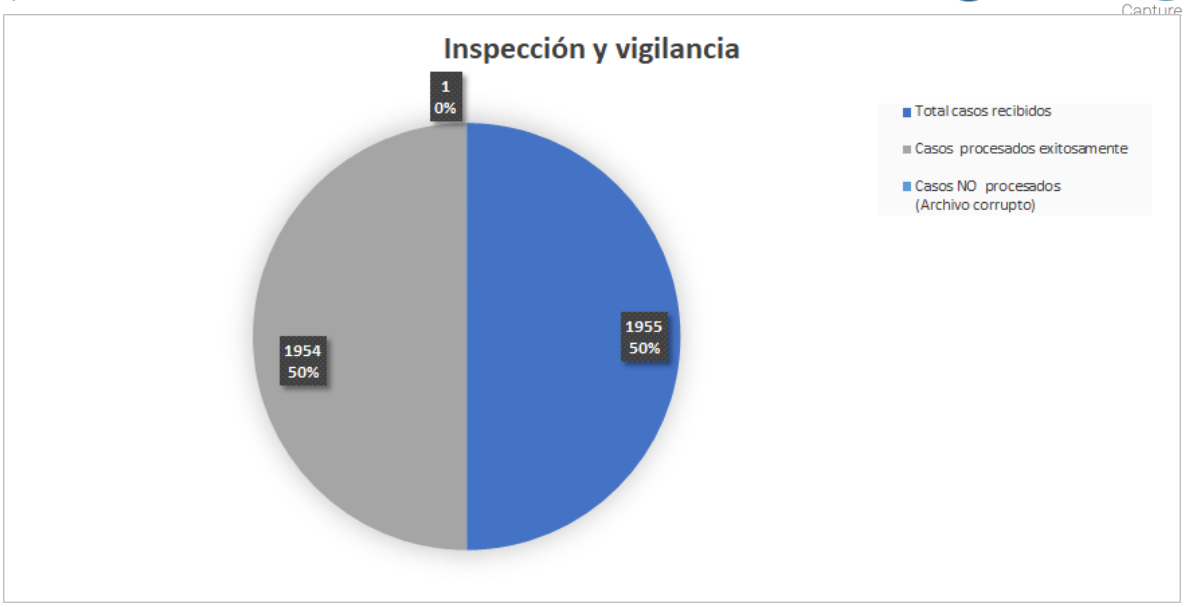
Instituciones Acreditadas no presenta registros en esta base.



XIV Ilustración resultado conjunto Sin Tipo

Para el acto administrativo de **Inspección y Vigilancia**, se recibieron un total de 1.955 registros, de los cuales 1.954 fueron procesados exitosamente, lo que representa un nivel de efectividad del 99,95% del universo total.

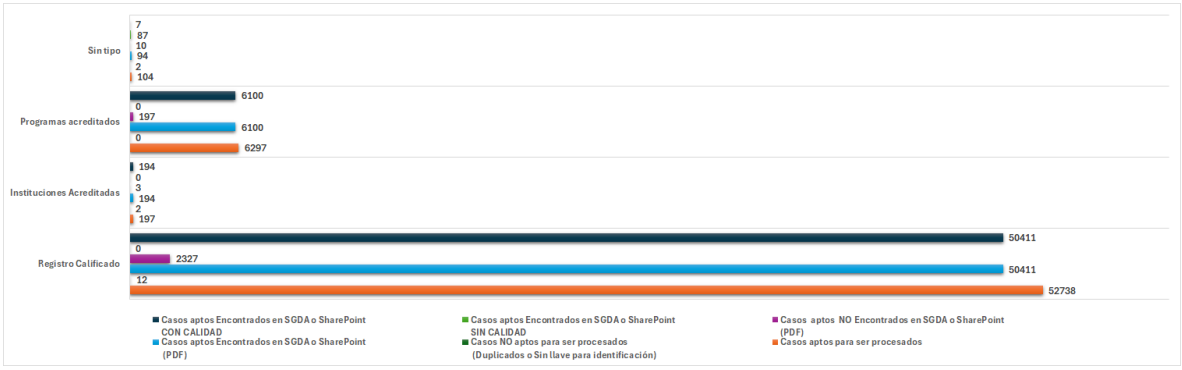
Únicamente 1 caso (0,05%) no pudo ser procesado debido a que el archivo se encontraba corrupto, situación que corresponde a un evento puntual de calidad documental.



XV Ilustración resultado inspección y vigilancia

Consolidando el avance de los documentos para los 61.307 registros totales recibidos en el alcance del proyecto, se determinó que 61.290 registros fueron aptos para ser procesados, lo que representa un 99,97% de representación frente al universo recibido, evidenciando únicamente 17 casos (0,03%) no aptos o no procesados, correspondientes a duplicidades, ausencia de llave de identificación y un archivo corrupto en el componente de Inspección y Vigilancia.

En el consolidado sin incluir Inspección y Vigilancia, se gestionaron 59.352 registros, de los cuales 59.336 fueron aptos para procesamiento, con 16 casos no aptos identificados. Por su parte, el componente de Inspección y Vigilancia registró 1.955 casos recibidos, de los cuales 1.954 fueron procesados exitosamente, presentándose únicamente un (1) archivo no procesado por encontrarse corrupto, el resultado de la extracción se encuentra dentro del **Anexo 10. Extracción completa.**





XVI Ilustración resultado general de extracción

4.9. Metodología de aplicación de cálculo de precisión

La validación del desempeño de la extracción se fundamenta en una **medición objetiva de completitud**, la cual permite evaluar la capacidad del sistema para obtener los campos requeridos de cada documento, considerando tanto los datos efectivamente extraídos como aquellos que no aplican para ciertos tipos documentales. Esta métrica integra un factor de compensación del **80 % de cumplimiento** para los campos identificados como no aplicables, reflejando así una valoración más realista del desempeño global del modelo de extracción. El cálculo se realiza mediante la siguiente expresión:

$$Completitud = \frac{Total\ de\ campos\ extraidos}{Total\ de\ campos\ esperados - Promedio\ de\ campos\ aplicables(80\%)} \times 100$$

La aplicación de esta fórmula permite **cuantificar de manera objetiva y balanceada la eficiencia del sistema**, evitando penalizar aquellos documentos que, por su naturaleza, no contienen ciertos campos. De esta forma, se garantiza que la evaluación refleje la **calidad integral del proceso de extracción** y no únicamente la presencia o ausencia literal de información.

El proyecto establece un **umbral técnico mínimo del 80 % de precisión**, que deberá alcanzarse y mantenerse durante las fases de validación, asegurando la **consistencia, trazabilidad y confiabilidad de los datos extraídos**.

4.10. Ejercicio de medición con muestras

A continuación, se presenta el **resumen consolidado del desempeño obtenido en la validación de los tres conjuntos de datos calificados**: *Registros Calificados*, *Instituciones* y *Programas Acreditados*.

Para revisar el detalle individual de cada documento, consultar el **Anexo 2. Medición de completitud**.

Conjunto	Total de documentos evaluados	Promedio de completitud ponderada (%)	Mínimo registrado (%)
Registros Calificados	128	97.64%	94.62%

Instituciones	127	96.46%	92.00%
Programas Acreditados	127	96.69%	94.0 %

Interpretación:
Los resultados evidencian un **alto nivel de completitud** en la extracción de información en los tres conjuntos de validación, con valores promedio superiores al **96 %**. Este desempeño supera ampliamente el **umbral técnico del 80 %** establecido, confirmando la **consistencia, precisión y estabilidad del modelo de extracción** en escenarios documentales heterogéneos.

Asimismo, se observa una **variación mínima entre los conjuntos**, lo cual refleja la solidez del proceso de extracción y la correcta aplicación del **factor ponderado del 80 % para los campos no aplicables**.
Estos resultados garantizan la **calidad, trazabilidad y confiabilidad** de los datos extraídos, y constituyen una base técnica sólida para la siguiente fase del proyecto.

Adicionalmente, se envía el informe detallado de extracción de la muestra, ver **Anexo 3. Extraccion Muestra_documentos,** el cual complementa este análisis y permite evidenciar el desempeño individual de los documentos validados durante el proceso.

4.11. Plan de pruebas

La validación y certificación de la precisión se realizará mediante un **Plan de Pruebas** formal, utilizando una muestra de documentos que será contrastado con la **Validación Manual/Scripts de Control.**

Criterios de Aceptación y Validación

El umbral técnico mínimo de aceptación para la fase de extracción es del 90%. El proceso solo se considera exitoso si se cumple con el siguiente criterio de manera concurrente:

- Aceptación Final: Completitud ≥ 90%

Para realizar una muestra significativa para aplicar la metodología de cálculo de precisión para la primera etapa de Fase I se tiene un avance del 75% de extracción correspondiente a un esperado de 34671 registros extraídos que se debe tener en cuenta para tomar la muestra.

4.12. Protocolo de validación cruzada y distribución de casos

Para asegurar la calidad y representatividad de los resultados obtenidos, se implementó un **protocolo de validación cruzada** que permitió contrastar los desempeños del modelo de extracción en diferentes subconjuntos de documentos.
Este enfoque garantizó la **consistencia técnica, trazabilidad y reproducibilidad** de las mediciones realizadas.

La **muestra total de 382 documentos** (distribuidos entre *Registros Calificados*, *Instituciones* y *Programas Acreditados*) fue segmentada en grupos de validación

equivalentes, con el propósito de evaluar el comportamiento del sistema frente a distintos tipos y estructuras documentales.

Cada grupo fue analizado de manera independiente, aplicando los mismos criterios de evaluación y la **fórmula de completitud ponderada (80 % para campos no aplicables)**. Posteriormente, los resultados parciales fueron integrados en un **consolidado general**, verificando la homogeneidad del desempeño y la ausencia de sesgos entre los conjuntos de prueba.

Durante el proceso se realizó una **revisión cruzada** entre el equipo técnico y los validadores del proyecto, asegurando la coherencia de los resultados y la correcta interpretación de los campos aplicables y no aplicables.
Esta metodología permitió confirmar la **estabilidad y confiabilidad del modelo de extracción**, sentando las bases para su aplicación masiva en la siguiente fase del proyecto.

4.13. Hallazgos

21 de octubre — Hallazgo: Registros no presentes en SGDEA:
Descripción: Tras concluir la extracción de datos desde SGDEA, se identifican **7.947 registros que no aparecen en el sistema**, pese a estar esperados dentro del universo de información.
Impacto: La ausencia de estos datos compromete la **integridad y completitud del dataset** requerido para los análisis y el procesamiento RPA.
Acción: Se genera y comparte un informe detallado con el listado y evidencias (*Anexo 6_informe reporte SGDEA.xlsx*).

24 de octubre — Hallazgo: Estructura documental inconsistente en SharePoint:
Descripción: Se realiza un análisis de la estructura de carpetas del SharePoint de inspección y vigilancia, evidenciando **falta de organización y ausencia de una arquitectura documental clara** para ubicar los soportes requeridos.
Impacto: La falta de estandarización dificulta la **trazabilidad, búsqueda y clasificación de documentos**, afectando la automatización del proceso.
Acción: Se documentan y comparten las **reglas de organización** propuestas para el RPA (*Anexo 5_Verificacion regla df_informes*).

25 de noviembre — Hallazgo: Inconsistencias en el tipo de trámite de registros calificados:
Descripción: Se identifican 1.233 registros de programas con registro calificado cuyo tipo de trámite no corresponde con las resoluciones asociadas, principalmente por cargas provenientes de SGDEA.
Impacto: La inconsistencia puede generar errores en la categorización, clasificación documental y decisiones posteriores, afectando la calidad de la información.
Acción: Se comparte un informe con el detalle de los casos y las inconsistencias encontradas (Anexo 7_resoluciones con tipo tramite no aplica).

27 de noviembre — Hallazgo: Registros no encontrados ni en SGDEA ni en SharePoint:
Descripción: Se evidencian **2.337 registros de la base de adición** que no se encuentran ni en el sistema SGDEA ni en los documentos del SharePoint de actos administrativos.
Impacto: La ausencia simultánea de información en ambas fuentes limita la **validación de soportes, la reconstrucción de expedientes y el alcance del procesamiento automatizado**.
Acción: Se genera y comparte el informe de detalle con la totalidad de casos encontrados (*Anexo 8_registros segundo ciclo no encontrados ni en sgdea ni en sharepoint.xlsx*).

3 de diciembre — Hallazgo: Corrección parcial de inconsistencias en registros calificados
Descripción: Tras un segundo procesamiento de los **1.233 registros con tipo de trámite no aplicable**, se identifican **1.029 casos corregidos** mediante búsqueda en el SharePoint de actos administrativos, quedando **204 registros** aún con tipo de trámite incorrecto.
Impacto: La corrección mejora la **calidad y confiabilidad del dataset**, pero los casos pendientes requieren tratamiento especial antes del procesamiento final.
Acción: Se comparte el informe consolidado del reprocesamiento realizado (*Anexo 9_registros RC tipo tramite no deseado (doble procesamiento).xlsx*).

Hallazgo: Restricciones de Acceso al Repositorio:

Descripción: Se identificó la **falta de acceso** a una porción de los registros documentales alojados en el portal de origen.

Impacto: La falta de acceso limita la completitud del universo de extracción, impide procesar expedientes completos y ralentiza la ejecución del RPA y ZAT Capture.
Acción: Se gestionaron solicitudes de acceso con los administradores de los repositorios.

Hallazgo: Formatos de Documentos No Estándar:

Descripción: Se encontraron registros que contienen documentos adjuntos en formatos diferentes a PDF.

Impacto: La presencia de documentos en formatos distintos a PDF impide que sean procesados directamente por el flujo automatizado definido (RPA/ZATCapture), obligando a incorporar pasos adicionales y afectando los tiempos y la eficiencia del proceso de extracción.

Acción: Identificar y definir una estrategia específica para el tratamiento de estos formatos no estándar.

Hallazgo: Datos de Identificación Faltantes (Código SNIES):

Descripción: Se detectó que algunos documentos no incluyen el código SNIES de las instituciones o programas.

Impacto: La ausencia del Código SNIES impide identificar correctamente la institución o programa asociado a cada documento, lo que afecta la trazabilidad y dificulta consolidar la información dentro del universo de extracción.

Acción: Para subsanar esta deficiencia, se recurrirá a la base inicial del proyecto y se realizará un cruce de información utilizando el campo Codproceso como clave de emparejamiento.

Hallazgo: Verificación de la Aplicabilidad de Nombres:

Descripción: Se identificó que los documentos contienen múltiples nombres de personas, pero no existe claridad sobre cuáles de ellos son relevantes o aplicables para la extracción como entidades de interés. Es necesario revisar el criterio de extracción para definir qué roles o menciones deben ser incluidos formalmente en los campos de datos.

Impacto: La falta de un criterio definido puede generar extracción de nombres no pertinentes o la omisión de información relevante, afectando la precisión y coherencia de los datos asociados a cada caso.

Acción: Realizar una revisión del criterio de extracción junto con el área usuaria para delimitar los roles o menciones que deben ser considerados como datos válidos y actualizar las reglas de extracción en función de esa definición

Hallazgo: Ajuste en la Extracción de I. y V.:

Descripción: Para los registros de Inspección y Vigilancia (I. y V.), se identificó la necesidad de crear un nuevo campo en la tabla de extracción con el propósito de registrar la ruta específica del documento en SharePoint, garantizando trazabilidad y referencia directa al archivo original.

Impacto: Sin este campo, no es posible mantener una trazabilidad documental completa ni acceder de forma directa al documento fuente, lo que dificulta la validación, auditoría y revisión posterior de los registros extraídos.
Acción: Crear el nuevo campo en la tabla de extracción e incorporar su diligenciamiento dentro del proceso de captura, registrando la ruta de SharePoint asociada a cada documento de I. y V.

- **Hallazgo: Aplicación de reglas de negocio de extracción:**
  Descripción: Durante el proceso de extracción se identificó que algunas de las reglas de negocio presentan ambigüedades en su interpretación y aplicación, lo que afecta la claridad del proceso y la calidad de la información extraída.
  Impacto: Las ambigüedades en las reglas pueden generar inconsistencias en los resultados de extracción, variaciones en los datos obtenidos y la necesidad de reprocesamientos, afectando la eficiencia y confiabilidad del proceso.
  Acción: Realizar mesas de trabajo para refinar y aclarar las reglas de negocio, y posteriormente reprocesar los casos afectados con el fin de asegurar la correcta aplicación de los criterios definidos.

- Hallazgo: **Documentos en repositorio diferentes al alcance:**
  Descripción: Durante el ejercicio de extracción se identificó que en el SharePoint de Inspección y Vigilancia existen carpetas y documentos que no corresponden a los establecidos dentro del alcance del contrato.
  Impacto: La presencia de documentos fuera de alcance puede generar confusión en la clasificación, aumentar el volumen de análisis y afectar la eficiencia del proceso de extracción al incluir información no pertinente.
  Acción: Con el fin de mitigar este escenario se realizó un análisis y aplicación de posibles reglas que permitan segmentar los documentos que aplican para el alcance, donde el área usuaria estuvo conforme con esta actividad.

5. Fase II – Validación de consistencia de datos

5.1. Metodología de gestión de fase

5.1.1. Avance validación de consistencia de datos

El avance reportado del **100% en la Fase II – Validación de Consistencia de Datos** se respalda en el cumplimiento exitoso de las actividades planificadas en el **Product Backlog** y la ejecución de tareas clave asociadas a la preparación técnica, funcional y operativa del entorno de validación.

Dada la naturaleza de esta fase, el progreso no solo se mide por el análisis de las diferencias, sino también por las actividades previas que aseguran una correcta ejecución del proceso. Por ello, el avance se ha evaluado en función de la construcción de capacidades, habilitación de insumos y definición de reglas y componentes funcionales necesarios.

Entre las actividades completadas que respaldan el avance del 100%, destacan:

5.1.2. Planeación Técnica, Funcional y Operativa de la Fase II (Completada)

- **Establecimiento del mecanismo de extracción de datos** desde **EDU-SUP**.
- Obtención y **restauración del backup** de la base de datos **EDU-SUP**.
- Ajuste de la **arquitectura técnica** para permitir la integración fluida entre la base de extracción (Fase I) y **EDU-SUP**.

Estas actividades han permitido disponer de los **insumos fundamentales** para la construcción de la base intermedia de validación.

5.1.3. Aprovisionamiento y Configuración de Infraestructura

- **Aprovisionamiento** y ajuste de las capacidades de la arquitectura en **Azure**, asegurando que la infraestructura esté alineada con los requerimientos de la fase de validación.

- **Configuración inicial del entorno Azure Databricks**, que será el núcleo para el procesamiento de los algoritmos de consistencia.

Este trabajo ha asegurado la **disponibilidad de los entornos operativos** necesarios para iniciar el cruce de información de manera eficiente.

5.1.4. Diseño Funcional y Construcción de Mockups de la Solución de Validación

- **Levantamiento de requerimientos funcionales** para la plataforma de validación.
- Desarrollo y finalización de la **1ra, 2da, 3ra y 4ta fase de maquetación de mockups**, incluyendo:
  - **Diseño general de la interfaz**.
  - **Estrategia de aplicación de filtros**.
  - **Mejoras y correcciones funcionales**.
  - **Culminación del prototipo final** del mockup para validación con el **MEN**.

Estas actividades representan el avance funcional más relevante, pues constituyen la base para la **interacción del usuario con la herramienta de verificación**, asegurando una plataforma intuitiva y funcional.

5.1.5. Integración y Uso de Archivos de Reglas de Transformación en Databricks

Se utilizaron los siguientes archivos para la correcta validación de los datos en la Fase II:

- **Archivo de Mapeo de Campos**: Define las correspondencias entre los campos de origen y destino para el proceso de validación, discriminando los diferentes tipos de actos administrativos, como **Registro Calificado**, **Programas Acreditados en Alta Calidad** y **Instituciones Acreditadas en Alta Calidad**.
- **Archivo de Reglas de Transformación de Inspección y Vigilancia**: Contiene las reglas de negocio y validación de los datos, aplicando criterios como la **validación de obligatoriedad**, la **normalización de formatos** (fechas, textos, códigos) y el **control de duplicados y unicidad** de los registros.

En **Azure Databricks**, se llevan a cabo las siguientes actividades clave:

- **Construcción de la base intermedia de validación**: A través de **notebooks y jobs** en Databricks, se armonizan los campos (tipos, catálogos, claves) y se resuelven duplicados. La base intermedia actúa como el **punto único de verdad** para la validación.

- **Aplicación de reglas de consistencia**: Se ejecutan las reglas de negocio definidas en los archivos de reglas de transformación (completitud, unicidad, integridad referencial, rangos y formatos). Sobre esta base, se realiza la comparación entre los datos extraídos de **EDU-SUP** y la **base de extracción**.
- **Generación de banderas de estado**: Se crean indicadores de estado (**COMPLETO, DIFERENCIAS, NUEVO**) para cada registro y se almacena una tabla de casos a validar con la traza de reglas aplicadas.

5.1.6. Visualización de Casos para Validación a través del Frontend

Los resultados obtenidos en Databricks, incluidos los casos con **diferencias** o **inconsistencias**, son expuestos a través de una **API/servicio backend** que alimenta la **interfaz frontend** para su visualización. A través del **mockup de la plataforma de validación** (ver enlace de mockups), los usuarios podrán:

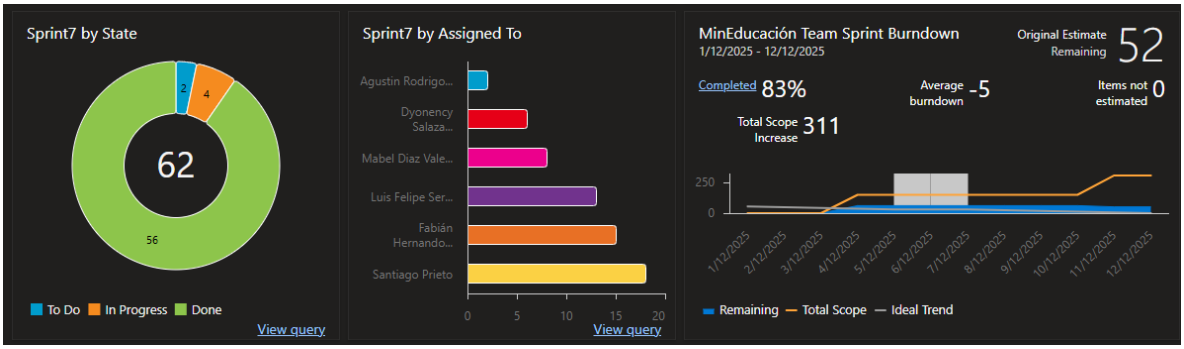
- **Verificar y validar casos con diferencias**.
- **Aplicar correcciones** a los registros detectados, permitiendo a los usuarios corregir los datos y registrar las modificaciones en la base intermedia, con un registro de auditoría de cambios (usuario, fecha, regla aplicada, valor anterior/nuevo).

Este flujo garantiza que los usuarios puedan interactuar de manera eficiente con los datos y asegurar su consistencia antes de su publicación final.

Product Backlo...	Fase II Validación datos	Committed	Business	MinEducación\Sprint 5	
Task	Establecer mecanimo de extracci◆n de datos EDU SUP	Done		MinEducación\Sprint 4	Ministerio...
Task	Obtener Backup de BD	Done		MinEducación\Sprint 5	Ministerio...
Task	Ajustar arquitectura de acuerdo a las validaciones de con...	Done		MinEducación\Sprint 5	Ministerio...
Task	Aprovisionar y ajustar capacidades con la arquitectura pr...	Done		MinEducación\Sprint 5	Ministerio...
Task	levantamiento de requerimientos para mockup	Done		MinEducación\Sprint 5	Ministerio...
Task	1ra fase maquetacion general mockup	Done		MinEducación\Sprint 5	Ministerio...
Task	2da fase planteamiento de estrategia en filtros	Done		MinEducación\Sprint 5	Ministerio...
Task	3ra fase mejora y correcciones	Done		MinEducación\Sprint 5	Ministerio...
Task	4ta fase culmunacion del mockup para entrega	Done		MinEducación\Sprint 5	Ministerio...
Task	Construcción de mockups plataforma validación	Done		MinEducación\Sprint 5	Ministerio...
Task	Aprobación de mockups plataforma validación	Done		MinEducación\Sprint 5	Ministerio...
Task	Desarrollo de solución para validación de datos	Done		MinEducación\Sprint 6	Ministerio...
Task	Integración de funcionalidad con base de datos en datab...	Done		MinEducación\Sprint 6	Ministerio...
Task	Entrega del backup de EDU SUP	Done		MinEducación\Sprint 5	Ministerio...
Task	Restauración de backup base EDU SUP	Done		MinEducación\Sprint 5	Ministerio...
Task	Mesa de trabajo para levantamiento de reglas de cruces ...	Done		MinEducación\Sprint 5	Ministerio...
Task	Construcción de databricks para cruce de datos	Done		MinEducación\Sprint 5	Ministerio...
Task	Validaciones cruzadas y consolidación de reglas de extra...	Done		MinEducación\Sprint 6	Ministerio...
Task	Validaciones cruzadas y consolidación de reglas de extra...	Done		MinEducación\Sprint 6	Ministerio...
Task	Implementación de vista de tablas (listado general)	Done		MinEducación\Sprint 6	Ministerio...
Task	Implementación de vista detallada por registro	Done		MinEducación\Sprint 6	Ministerio...
Task	Implementación de filtros en la tabla y cambios pedidos ...	Done		MinEducación\Sprint 6	Ministerio...
Task	descarga de repo y puesta en marcha en local	Done		MinEducación\Sprint 6	Ministerio E
Task	Integración con endpoints existentes (obtener registros y...	Done		MinEducación\Sprint 6	Ministerio E
Task	Implementar edición de registros	Done		MinEducación\Sprint 6	Ministerio E

XVII Backlog actual FASE II

Sprint 7



XVIII Avance Sprint 7

Estado general: El Sprint 7 contempló un total de 62 tareas, todas ellas finalizadas en su totalidad (100 % de cumplimiento), sin actividades pendientes ni en ejecución al cierre del

ciclo. Este resultado refleja una ejecución eficiente y controlada, así como la capacidad del equipo para atender el incremento de alcance del sprint sin afectar la calidad ni la estabilidad operativa de la solución.

Asignación: La asignación de tareas evidencia una distribución equilibrada del esfuerzo entre los integrantes del equipo. Se destaca una mayor participación de Santiago Prieto, acompañado por aportes relevantes de Luis Felipe, Dyonency, Mabel, Fabián Hernando y otros miembros del equipo. Esta dinámica colaborativa permitió mantener un flujo de trabajo constante, asegurar la continuidad de las actividades y cumplir oportunamente con los compromisos definidos para el sprint.

Burndown: La gráfica de burndown confirma un cumplimiento del 100 % de las tareas planificadas, mostrando una reducción progresiva y sostenida del backlog hasta alcanzar cero tareas restantes. El sprint alcanzó un total de 311 puntos de esfuerzo, manteniéndose dentro de un comportamiento estable y alineado con la tendencia ideal, sin desviaciones significativas. Este comportamiento evidencia una planificación adecuada, un uso eficiente de la capacidad y un seguimiento efectivo del avance.

5.2. Protocolo Técnico de validación

5.2.1. Arquitectura de validación

Para la correcta ejecución de la **Fase II**, se definieron y utilizaron insumos funcionales que permiten garantizar la coherencia, trazabilidad y correcta aplicación de las reglas de negocio durante el procesamiento de la información. Estos insumos constituyen la base para la construcción del mapeo de campos y las reglas de transformación para la validación de los datos.

5.2.2. Mapeo de Campos

El **Mapeo de Campos** establece la correspondencia entre los campos de origen y los campos de destino utilizados en el proceso de validación y consolidación de la información. Este archivo fue consultado de manera integral y se encuentra organizado en las siguientes pestañas:

Registro Calificado (RC y Reglas):

Define los campos asociados a los trámites de Registro Calificado, incluyendo información del programa académico, institución educativa, modalidad, estado del trámite y fechas relevantes. Adicionalmente, incorpora criterios de obligatoriedad y homologación de valores.

Programas Acreditados en Alta Calidad (AAC Programas):

Contiene el mapeo de los campos correspondientes a los programas acreditados, considerando identificadores como SNIES, tipo de acreditación, vigencia, resultados del proceso y resoluciones asociadas.

Instituciones Acreditadas en Alta Calidad (AAC Instituciones):

Incluye los campos relacionados con las instituciones acreditadas, tales como datos institucionales, tipo de acreditación, periodos de vigencia y actos administrativos asociados.

Este archivo constituye el insumo principal para asegurar la correcta interpretación semántica y estructural de la información procesada.

5.2.3. Archivo de Reglas de Consistencia y Transformación de Datos

El Archivo de Reglas de Consistencia y Transformación de Datos es el insumo técnico que consolida, describe y formaliza las reglas de extracción, limpieza, transformación, comparación y llenado aplicables a la información de actos administrativos, actas ejecutorias, informes de visitas preventivas, planes de visita, autos de trámite y documentos asociados a las funciones preventiva y sancionatoria administradas por el Ministerio de Educación.

Este archivo se utiliza como complemento al mapeo de campos y establece los criterios necesarios para asegurar la correcta integración y correspondencia entre la información extraída por ZAT y los datos almacenados en las tablas del modelo **EDU_SUP**, garantizando consistencia, trazabilidad y calidad del dato a lo largo del proceso.

- **Alcance del Archivo de Reglas**

El archivo define las reglas que deben aplicarse sobre los datos provenientes de las fuentes relevantes, incluyendo, entre otras:

Registro Calificado

- cat_rc_e1_rc
- Base_RC-E2
- Base_0106SinTipo
- tbl_resoluciones_estado
- PRO_INFO_RESOLUCION
- INS_INSTITUCION_EDUCATIVA
- PRO_PROGRAMA_ACADEMICO
- PRO_PROGRAMA_LUGDES
- PRO_HISTORICO_REGISTRO_CALIFICADO

- PRO_HISTORICO_MUNICIPIO_RC
- Tablas auxiliares (REF_MODALIDAD, tipo trámite MRD, REF_DEPARTAMENTO, REF_MUNICIPIO, archivo SOL994599, entre otras)

Programas Acreditados en Alta Calidad

- cat_aac_prg_e1
- Base_0106SinTipo
- tbl_resoluciones_estado
- PRO_INFO_RESOLUCION
- INS_INSTITUCION_EDUCATIVA
- PRO_PROGRAMA_ACADEMICO
- PRO_HISTORICO_ACREDITACION_PROGRAMA
- PRO_HISTORICO_MUNICIPIO_ACREDITACION_PROG
- Tablas auxiliares (REF_MODALIDAD, tipo trámite MRD, REF_DEPARTAMENTO, REF_MUNICIPIO, archivo SOL994599, entre otras)

Instituciones Acreditadas

- cat_aac_ies
- tbl_resoluciones_estado
- INS_INFO_RESOLUCION
- INS_INSTITUCION_EDUCATIVA
- REF_TIPO_ACREDITACION
- INS_HISTORICO_ACREDITACION_INSTITUCION
- INS_HISTORICO_MUNICIPIO_ACREDITACION_INST
- Tablas auxiliares (REF_TIPO_ACREDITACION, tipo trámite MRD, REF_DEPARTAMENTO, REF_MUNICIPIO, archivo SOL994599, entre otras)

Inspección y Vigilancia

- **Resoluciones y Autos:**

- tbl_resoluciones
- tbl_auto
- REF_CLASE_RESOLUCION_INSPECCION_VIGILANCIA
- INS_INSTITUCION_EDUCATIVA
- DBO.INS_RESOLUCION_INSPECCION_VIGILANCIA

- **Informes:**
  - tbl_informes
  - INS_INSTITUCION_EDUCATIVA
  - DBO.INS_INFORME_INSPECCION_VIGILANCIA
- **Planes de Visita:**
  - tbl_visitas
  - INS_INSTITUCION_EDUCATIVA
  - DBO.INS_PLAN_VISITA_INSPECCION_VIGILANCIA
- **Evaluación de Viabilidad:**
  - tbl_viabilidad
  - INS_INSTITUCION_EDUCATIVA
  - DBO.INS_EVALUACION_VIABILIDAD_INSPECCION_VIGILANCIA

- **Estructura de las Reglas Definidas en el Archivo**

El archivo organiza las reglas en los siguientes componentes:

1. Reglas Generales

Incluyen directrices aplicables a todos los campos, tales como limpieza de caracteres especiales, estandarización de formatos numéricos y de fecha, homologación de nombres y definición de llaves de correspondencia entre tablas.

2. Reglas de Transformación

Corresponden a la estandarización y depuración del dato extraído, incluyendo eliminación de ceros no significativos, transformación de fechas, conversión de tipos de dato, construcción de estructuras compuestas y manejo de valores múltiples.

3. Reglas de Extracción

Definen cómo obtener cada campo desde sus fuentes, especificando tabla origen, columnas, llaves requeridas y orden de validación, permitiendo complementar información faltante cuando la fuente primaria no la provee.

4. Reglas de Comparación

Establecen validaciones entre fuentes y tablas para asegurar la consistencia de la información, permitiendo identificar diferencias, inconsistencias y posibles errores en los datos procesados.

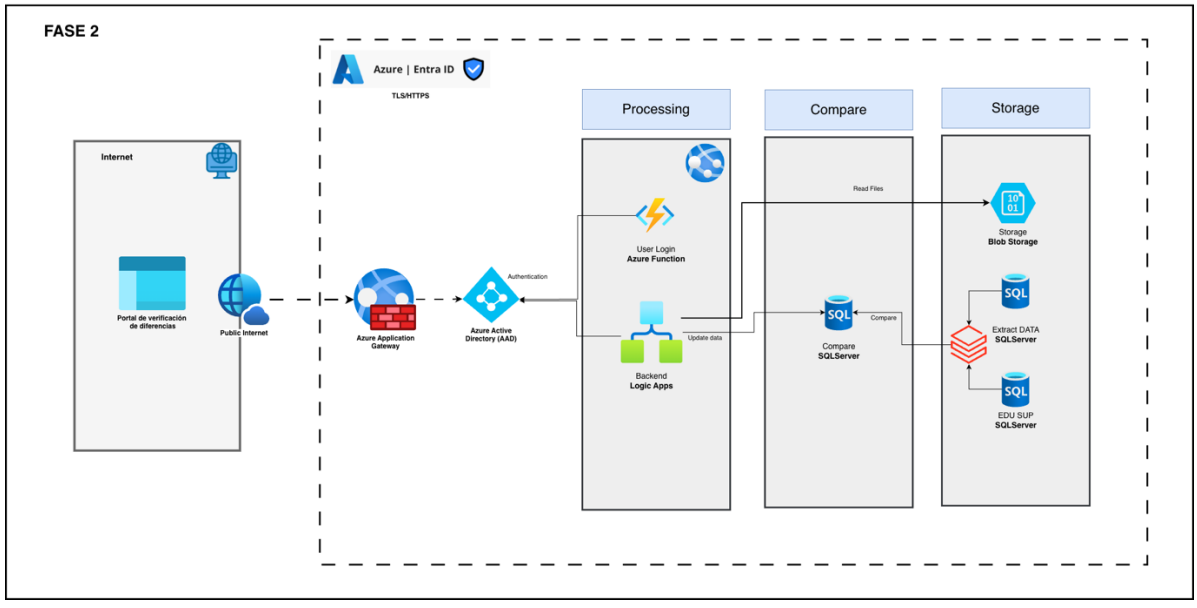
5. Reglas de Llenado

Describen los criterios para poblar las tablas destino del modelo EDU_SUP, incluyendo campos obligatorios, registros múltiples, valores estándar y decisiones lógicas según la naturaleza del dato.

- **Gestión, Control y Trazabilidad**

El Archivo de Reglas de Consistencia y Transformación es gestionado como un insumo oficial del proceso, sujeto a versionamiento y control de cambios. Cualquier ajuste debe ser validado previamente por los usuarios funcionales y reflejado en la ejecución del proceso.

El archivo se referencia como anexo del informe y constituye el soporte documental para auditorías, validaciones funcionales y futuras evoluciones del proceso, asegurando la trazabilidad completa desde la fuente original hasta el dato final consolidado.



XIX Diagrama de arquitectura Fase II

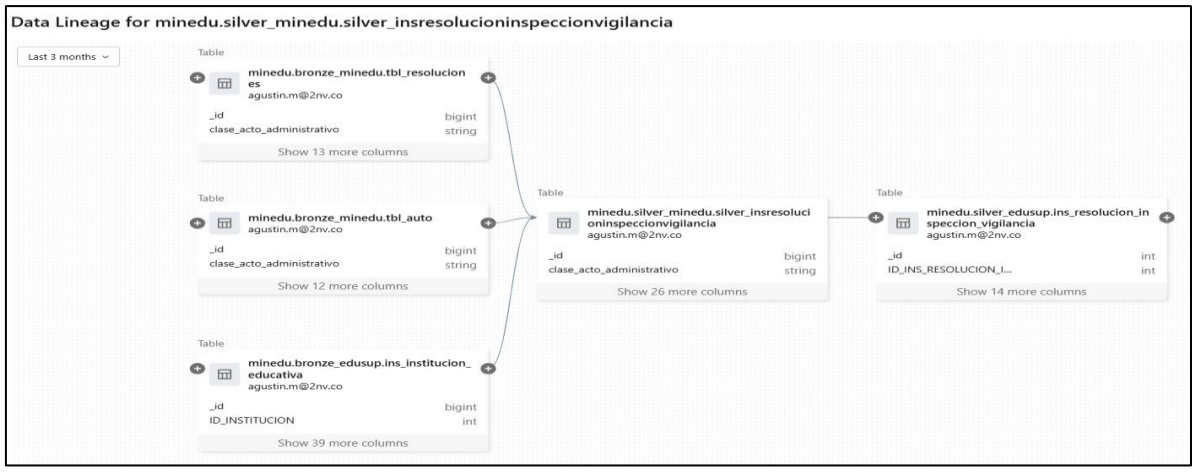
5.3. Mapa de Flujo y Dependencias de Validación de Datos

Este apartado presenta el detalle técnico del flujo de validación de datos, mostrando las conexiones entre las fuentes, las tablas intermedias y los componentes encargados de las transformaciones, proporcionando un panorama claro de las dependencias y del recorrido que sigue la información dentro de la arquitectura de validación. Estas representaciones complementan la arquitectura general descrita y facilitan la trazabilidad técnica del proceso.

A continuación, se detallan los flujos por "**Inspección y vigilancia", "Registro Calificado", "Programas Acreditados en alta calidad"** e **"Instituciones Acreditadas"**.

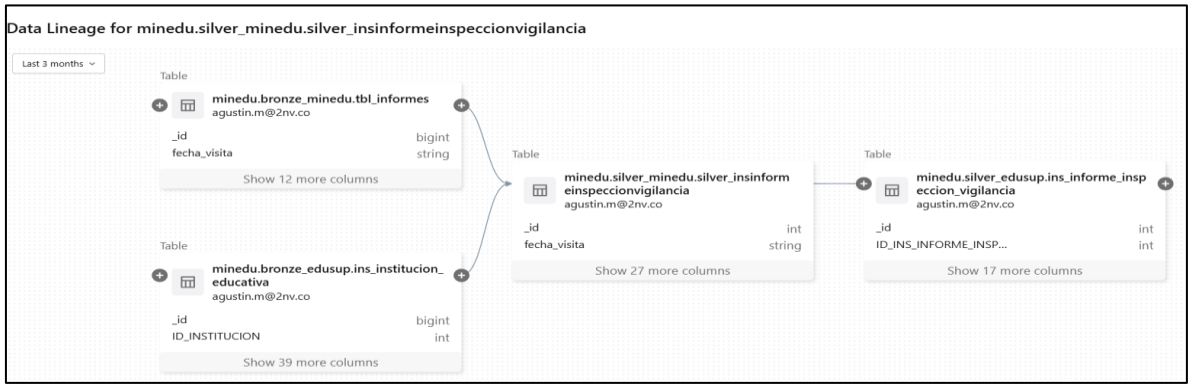
Inspección y Vigilancia

- ***Resoluciones-Auto***



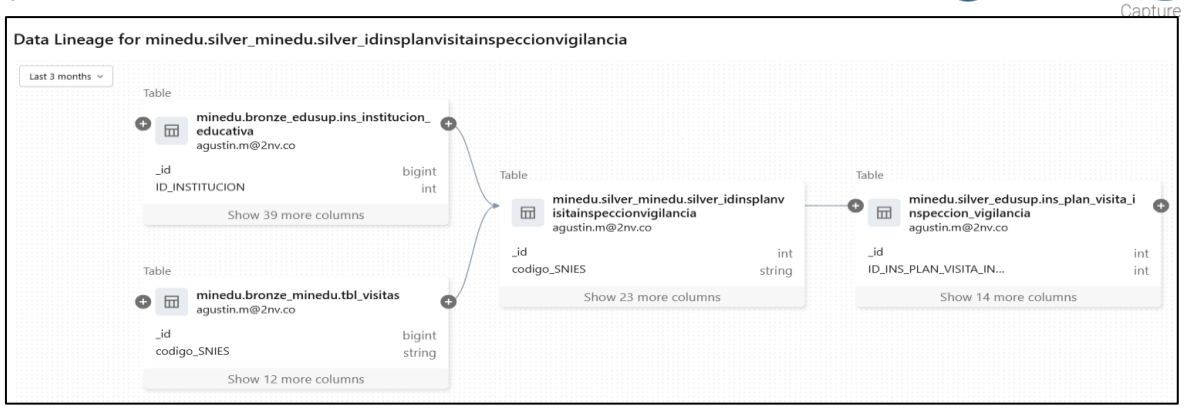
XX Flujo y Dependencias de Validación de Datos Fase II - Resoluciones-Auto

- ***Informes***



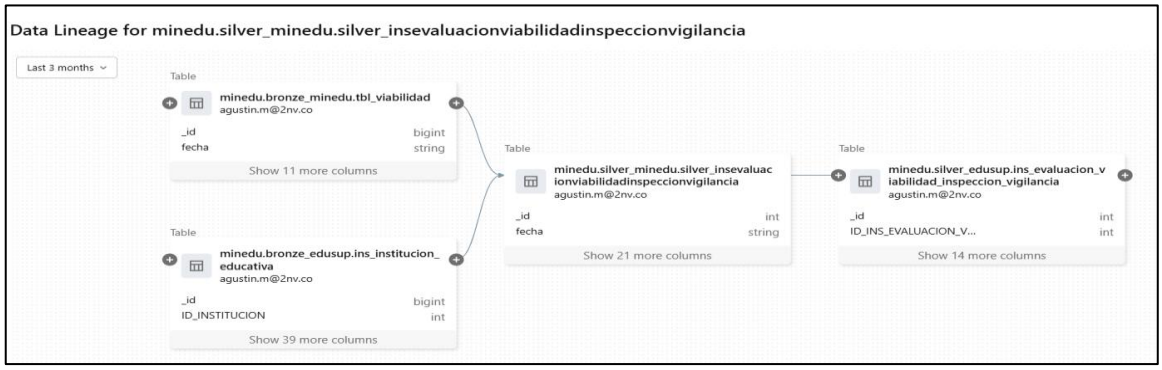
XXI Flujo y Dependencias de Validación de Datos Fase II - Informes

- ***Planes de Visita***



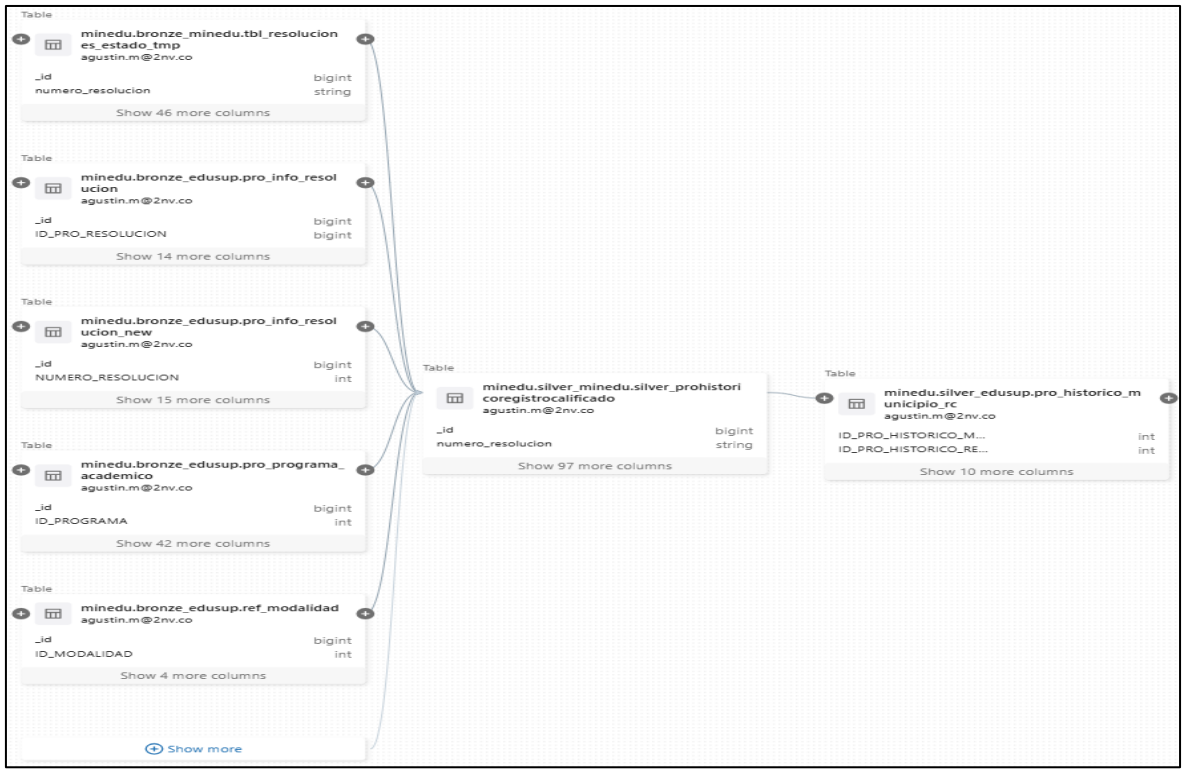
XXII Flujo y Dependencias de Validación de Datos Fase II - Planes de Visita

- ***Evaluación de Viabilidad***



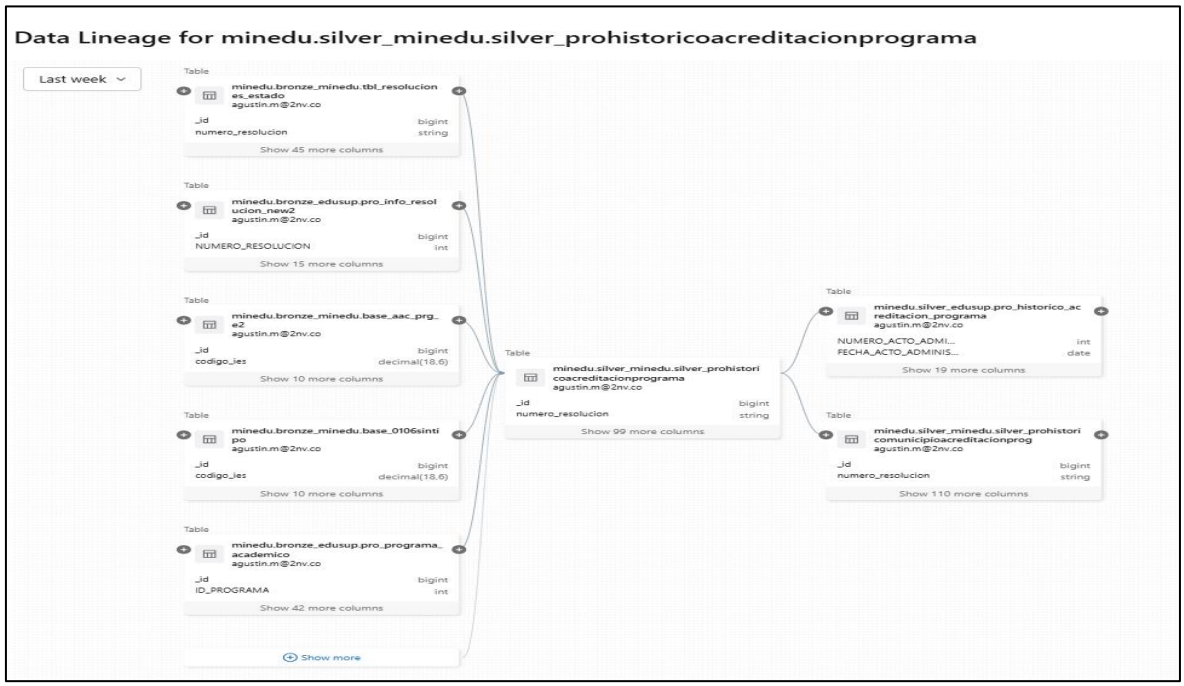
XXIII Flujo y Dependencias de Validación de Datos Fase II - Evaluación de Viabilidad

- ***Registro Calificado***



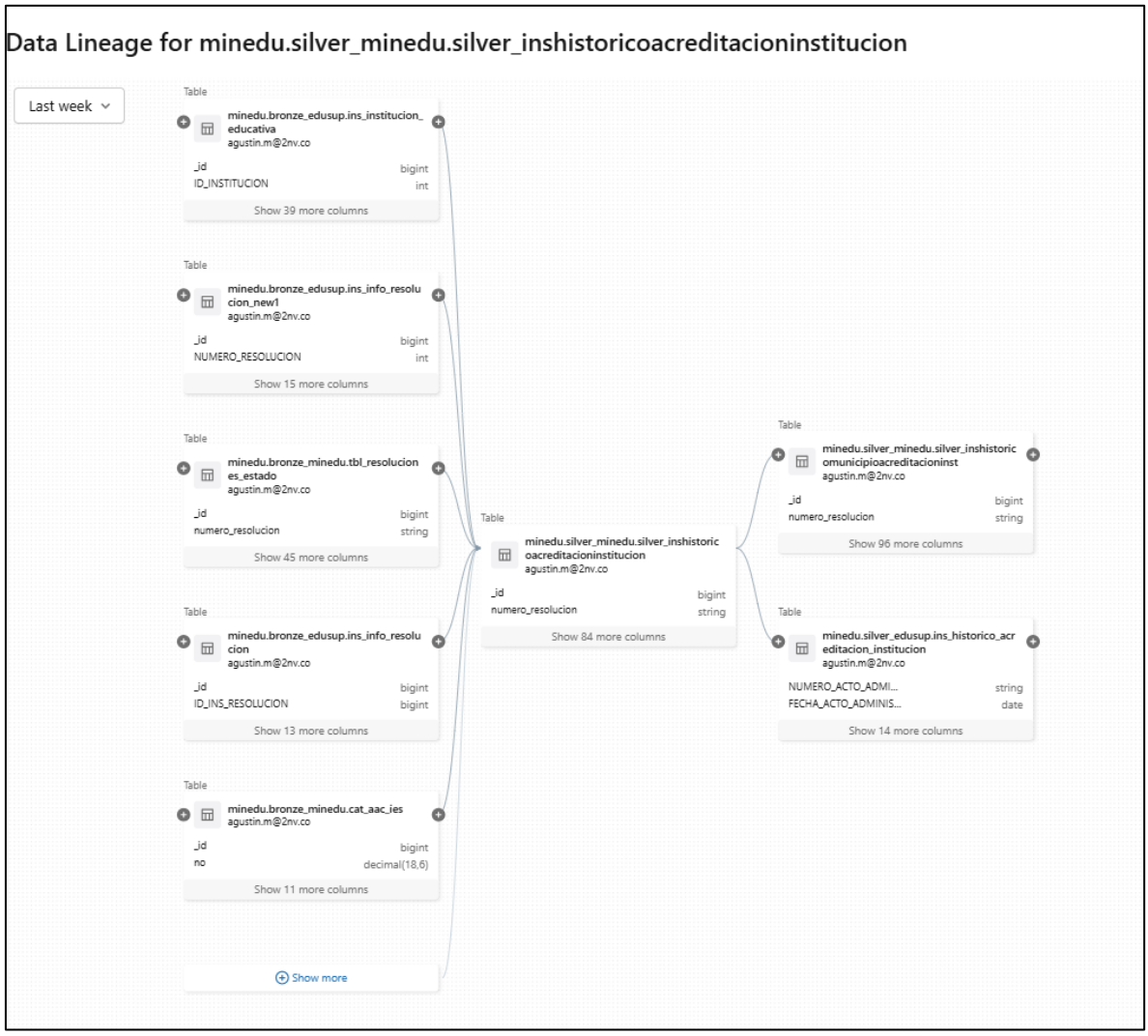
XXIV Flujo y Dependencias de Validación de Datos Fase II - Registro Calificado

- ***Programas Acreditados en alta calidad***



XXV Flujo y Dependencias de Validación de Datos Fase II - Programas Acreditados en alta calidad

- ***Instituciones Acreditadas***



XXVI Flujo y Dependencias de Validación de Datos Fase II - Instituciones Acreditadas

5.4. Modelo de Entidad Relación (MER)

El Modelo Entidad-Relación (MER) define la estructura base para el tratamiento y la administración de la información del proyecto. El diseño se organiza en tres grupos de tablas: Tablas Fuente, Tablas Intermedias y Tablas de Llenado, las cuales permiten gestionar de manera ordenada el flujo de datos desde su origen hasta su consolidación final en el módulo Edusup.

El modelo está conformado por entidades autónomas que representan los principales componentes del proceso, como resoluciones, informes, visitas, viabilidad, información institucional y registros históricos. Cada entidad fue diseñada para capturar y estructurar datos específicos, garantizando integridad, claridad en los atributos y trazabilidad durante todo el ciclo de procesamiento.

Dado el enfoque modular del proyecto, no se establecieron relaciones explícitas entre las tablas. Esta decisión responde a la necesidad de mantener independencia operativa entre los conjuntos de datos y facilitar su uso, clasificación y control en cada etapa del proceso de Extracción, Transformación y Carga ETL.

En conjunto, el Modelo Entidad-Relación (MER) proporciona una estructura robusta, flexible y alineada con los requerimientos funcionales del proyecto, asegurando una gestión eficiente y confiable de la información crítica administrada por el sistema.

Tablas Fuente: Edusup

- ins_equipo_tecnico_plan_visita_inspeccion_vigilancia
- ins_evaluacion_viabilidad_inspeccion_vigilancia
- ins_info_resolucion
- ins_info_resolucion_new1
- ins_informe_inspeccion_vigilancia
- ins_institucion_educativa
- ins_resolucion_inspeccion_vigilancia
- pro_info_resolucion
- pro_info_resolucion_new
- pro_info_resolucion_new2
- pro_programa_academico
- ref_departamento
- ref_modalidad
- ref_municipio
- ref_tipo_acreditacion

Tablas Intermedias: Silver_Minedu

- silver_insevaluacionviabilidadinspeccionvigilancia
- silver_inshistoricoacreditacioninstitucion
- silver_inshistoricomunicipioacreditacioninst
- silver_insinformeinspeccionvigilancia
- silver_insplanvisitainspeccionvigilancia
- silver_insresolucioninspeccionvigilancia
- silver_prohistoricoacreditacionprograma
- silver_prohistoricomunicipioacreditacionprog
- silver_prohistoricoregistrocalificado
- sol994599_clean

Tablas Fuente: Minedu

- base_0106sintipo
- base_aac_prg_e2
- base_rc_e2
- cat_aac_ies
- cat_aac_prg_e1
- cat_rc_e1_rc
- cat_rc_e1_xx
- sol994599
- tbl_auto
- tbl_cola_pdf_inspeccion_vigilancia
- tbl_cola_pdf_inspeccion_vigilancia_aux
- tbl_cola_trabajo_inspeccion_vigilancia
- tbl_informes
- tbl_resoluciones
- tbl_resoluciones_estado
- tbl_resoluciones_estado_20250710
- tbl_resoluciones_estado_backup
- tbl_resoluciones_estado_tmp
- tbl_viabilidad
- tbl_visitas
- tipo_tramite_mrd
- vw_resoluciones_estado_expandida
- vw_resoluciones_unificada

Tablas Llenado: Edusup

- ins_evaluacion_viabilidad_inspeccion_vigilancia
- ins_historico_acreditacion_institucion
- ins_historico_municipio_acreditacion_inst
- ins_informe_inspeccion_vigilancia
- ins_plan_visita_inspeccion_vigilancia
- ins_resolucion_inspeccion_vigilancia
- pro_historico_acreditacion_programa
- pro_historico_municipio_acreditacion_prog
- pro_historico_registro_calificado
- pro_historico_municipio_rc

XXVII Tablas intermedias y de llenado

5.5. Resultado de Registros Procesados

El proceso de registros procesados, permitió consolidar y gestionar de manera exitosa la información proveniente de las diferentes fuentes del sistema, garantizando la correcta captura, transformación y carga de los datos definidos para el proyecto, conforme a las reglas de negocio establecidas.

Como resultado de la ejecución del proceso, se procesaron los siguientes volúmenes de información por tipo de registro:

Tablas	Registros Procesados
Resolución	256
Auto	222
Informe	1063
Evaluación de Viabilidad	89
Plan Visita	324
Registro Calificado	52.738
Programas Acreditados en alta calidad	6.297
SinTipo	104
Instituciones Acreditadas	197
Total	**61.290**

Estos resultados evidencian la correcta operación del modelo de procesamiento y la estabilidad de la arquitectura implementada, así como la capacidad del sistema para gestionar distintos tipos de actos administrativos y volúmenes de información. Asimismo, se garantiza la disponibilidad, consistencia y trazabilidad de los registros para su posterior consulta, análisis y uso operativo.

5.6. Metodología de aplicación de cálculo de precisión

La validación del desempeño del proceso de extracción se fundamentó en una medición objetiva de **completitud ponderada**, orientada a evaluar la capacidad del sistema para capturar los campos definidos para cada tipo documental, considerando tanto los valores efectivamente extraídos como aquellos que, por la naturaleza del documento, no resultan aplicables. Para estos últimos, se incorporó un **factor de compensación del 80 %**, el cual permite reflejar de manera equilibrada y realista el desempeño global del modelo, evitando penalizaciones indebidas sobre documentos que legítimamente no contienen determinados atributos.

La aplicación de esta metodología permitió cuantificar de forma objetiva la eficiencia del proceso de extracción, asegurando que la evaluación se centre en la calidad integral del modelo y no únicamente en la presencia literal de información. En este marco, el proyecto definió un **umbral técnico mínimo del 80 % de precisión**, el cual constituye el criterio de aceptación para las actividades de validación, garantizando la consistencia, trazabilidad y confiabilidad de los datos extraídos durante la Fase II.

5.7. Ejercicio de medición con muestras

Como parte del proceso de validación, se realizó un ejercicio de medición basado en muestras representativas de los tres conjuntos de datos críticos del proyecto: **Registros Calificados, Instituciones Acreditadas y Programas Acreditados en Alta Calidad**. El análisis permitió

obtener indicadores consolidados de completitud ponderada, cuyos resultados se resumen a continuación:

Conjunto	Total de documentos evaluados	Promedio de completitud ponderada (%)	Mínimo registrado (%)
Registros Calificados	128	97,64%	94,62%
Instituciones	127	96,46%	92,00%
Programas Acreditados	127	96,69%	94,00%

Los resultados evidencian un **alto nivel de completitud en los tres conjuntos evaluados**, con promedios superiores al 96 %, superando de manera amplia el umbral técnico del 80 % definido para el proyecto. Asimismo, la baja dispersión entre valores mínimos y promedios refleja la **estabilidad, precisión y homogeneidad del modelo de extracción**, incluso frente a escenarios documentales heterogéneos.

5.8. Hallazgos

- **Hallazgo 1: Identificación de tablas adicionales necesarias para Registro Calificado (RC), Programas Acreditados en alta Calidad (ACC_PROG) e Instituciones Acreditadas (ACC_INS):**

  Descripción: Durante las sesiones técnicas de análisis se evidenció que, contrario a la premisa inicial en la que solo se debían cargar datos extraídos por ZAT una vez procesados y limpios a una tabla consolidada, para los conjuntos de **Registro Calificado (RC)**, **Acreditación de Programas (ACC_PROG)** y **Acreditación de Instituciones (ACC_INS)** era obligatorio alimentar **tablas adicionales** en la base EDU_SUP. Estas tablas incluyen:

  PRO_INFO_RESOLUCION
  INS_INFO_RESOLUCION
  PRO_HISTORICO_MUNICIPIO_RC
  PRO_HISTORICO_MUNICIPIO_ACREDITACION

  Esta necesidad no estaba definida al inicio de la fase.

  Impacto: La inclusión de estas tablas amplía el alcance funcional y técnico del procesamiento, afectando el diseño del modelo de integración y el volumen de datos a poblar durante la validación y migración. Requiere ajustes en reglas de negocio, mapeos y estructura del pipeline de Databricks.

  Acción: Se documentan las nuevas dependencias, se actualizan los diccionarios de datos y se incorporan las transformaciones necesarias dentro del pipeline de validación y carga hacia EDU_SUP. Se actualizan los mockups y se ajustan las reglas de consistencia correspondientes.

- **Hallazgo 2: Modificación y sustitución de campos relacionados con ciclos propedéuticos**:

  Descripción: En la tabla **EDU_SUP.DBO.PRO_HISTORICO_REGISTRO_CALIFICADO** se identificó que los campos originalmente utilizados para representar los niveles del ciclo propedéutico:

  Codigo_snies_nivel_tecnico
  Codigo_snies_nivel_tecnologico
  Codigo_snies_nivel_universitario

  no correspondían al requerimiento funcional actualizado del MEN. Tras la revisión con negocio, se definió reemplazarlos por los campos:

  NOMBRE_PROGRAMA_CICLO_PROPEDEUTICO_1
  NOMBRE_PROGRAMA_CICLO_PROPEDEUTICO_2
  CODIGO_SNIES_PROPEDEUTICO_1
  CODIGO_SNIES_PROPEDEUTICO_2

  Impacto: Este cambio modifica las reglas de mapeo y las transformaciones que dependen de los ciclos propedéuticos, impactando el diseño del modelo intermedio y la validación cruzada con la base EDU-SUP.

  Acción: Se actualizan las reglas de negocio y la documentación de mapeos. Se ajustan los notebooks de Databricks encargados de poblar la tabla y se valida nuevamente la consistencia de estos campos con la fuente funcional.

- **Hallazgo 3: Incorporación de archivos complementarios entregados por negocio:**

  Descripción: Para completar campos críticos de las tablas relacionadas con trámites y decisiones administrativas, se identificó la necesidad de incluir los archivos externos **"SOL994599"** y **"tipo tramite_MRD"**, los cuales no estaban contemplados en el diseño inicial. Estos archivos contienen la información requerida para poblar:

  TIPO_TRAMITE
  ID_TIPO_TRAMITE
  DECISION_RESUELVE
  TIPO_ACREDITACION

  Impacto: La ausencia inicial de estos insumos impedía el llenado correcto de campos obligatorios en la estructura de EDU_SUP, afectando la completitud y consistencia del modelo.

Acción: Se incorporan los archivos dentro del flujo de procesamiento, se construyen las tablas auxiliares necesarias para su lectura y se actualizan las reglas de negocio que dependen de esta información.

- **Hallazgo 4: Adición del campo NOMBRE_MUNICIPIO en PRO_HISTORICO_MUNICIPIO_RC y PRO_HISTORICO_MUNICIPIO_ACREDITACION:**

  Descripción: Durante la revisión estructural de la tabla **PRO_HISTORICO_MUNICIPIO_RC y PRO_HISTORICO_MUNICIPIO_ACREDITACION**, se identificó que el campo **NOMBRE_MUNICIPIO** no se encontraba definido, pese a ser requerido para garantizar la trazabilidad territorial de los registros históricos de Registro Calificado.

  Impacto: Sin este campo la tabla no proporciona la referencia semántica necesaria para análisis posteriores, generando riesgo de ambigüedad en la información del territorio asociado.

  Acción: Se actualiza el modelo de datos añadiendo el campo, se ajustan las reglas de negocio y se incorpora su llenado dentro del pipeline de validación y carga, utilizando las fuentes oficiales proporcionadas por el MEN.

- **Hallazgo 5: Imposibilidad de tomar fechas desde PRO_PROGRAMA_ACADEMICO por tener tipo de dato INT:**

  Descripción: Las reglas iniciales para poblar los campos:

  FECHA_INICIO
  FECHA_FIN
  FECHA_MODIFICACION

  establecían que debían derivarse desde los campos **FECHA_INICIO_RC**, **VIGENCIA_RC** y **FECHA_MODIFICACION** de la tabla **PRO_PROGRAMA_ACADEMICO**.
  Sin embargo, durante la validación se determinó que alguno de dichos campos estaba almacenado como **INT** y no como **DATETIME**, por lo que su uso directo como fechas era inválido.

  Tras consultar con negocio y revisar la semántica operativa, se definió el siguiente comportamiento:

  FECHA_INICIO = GETDATE()
  FECHA_FIN = NULL
  FECHA_MODIFICACION = NULL

Impacto: El cambio obliga a redefinir la lógica de asignación de fechas y ajustar las validaciones de consistencia asociadas a los periodos históricos.

Acción: Se actualizan las reglas de negocio y el procesamiento en Databricks, asegurando que los campos se generen bajo la estructura requerida y validando su correcta exposición en las tablas finales.

- **Hallazgo 6: Ajuste de campos y reglas para inspección y vigilancia:**

  Descripción:
  Durante el proceso de validación y diseño de reglas para los conjuntos de Inspección y Vigilancia, se identificaron dos modificaciones necesarias a petición del MEN:

  En la tabla EDU_SUP.DBO.INS_PLAN_VISITA_INSPECCION_VIGILANCIA se añadió el campo **"Equipo Técnico"**, requerido para registrar la información asociada a los profesionales participantes en cada visita.

  En la tabla EDU_SUP.DBO.INS_INFORME_INSPECCION_VIGILANCIA se modificó la regla correspondiente al campo **"Numero Radicado" (**el día **12-12-2025)**, que inicialmente:

  - Era de tipo **INT**,
  - Y debía obtenerse desde rutas de archivos alojados en carpetas institucionales, tales como:
    - *Opción 1:*
      1. C:\Users\rpa02\mineducacion.gov.co\Preventiva I & V - 2.IES Sin Medidas\4110-ITFITP-INST TOLIM. DE FORMA. TÉCN PROFES\2.Informes\Info 2020_ER_193256 21_08_2020\Observaciones MEN\2020_ER_**206096** 02_09_2020\3. COMPONENTE FINANCIERO\OBSERVACIONES No. 13, 14, 15 Y 18 COMPONENTE FINANCIERO.pdf
    - *Opción 2:*
      1. C:\Users\rpa02\mineducacion.gov.co\Preventiva I & V - 2.IES Sin Medidas\4108-ESCUELA TECNOLÓGICA INSTITUTO TÉCNICO CENTRAL\2. Informes\Info 2024-ER-**0685155** 27_12_2024\Respuesta observaciones 2025-ER-0143726\Rta_Requerimiento_MEN.pdf

  Sin embargo, tras la solicitud del MEN, la regla fue reemplazada por una nueva instrucción funcional que:

  - Cambia el tipo de dato a **VARCHAR**,

- Y establece que el dato a extraer es **Numero Radicado** = **2024_ER_0685155**

Impacto:
Estos cambios modifican el comportamiento esperado de llenado y validación del campo "Numero Radicado", impactando las reglas de consistencia, los procesos de transformación y la lógica implementada en los pipelines de procesamiento.

Acción:
Se actualizan las reglas de negocio, diccionarios de datos y transformaciones asociadas. Se ajustan los notebooks y procesos de Databricks para soportar el nuevo tipo de dato y las nuevas reglas de extracción. Se valida nuevamente el comportamiento del llenado del campo "Numero Radicado" bajo los patrones definidos y se incorpora el campo "Equipo Técnico" en las rutas de carga y validación correspondientes.

6. Fase III: Definición de metadatos y migración de la data y archivos con metadatos

6.1. Metodología de gestión de fase

La **Fase III** corresponde a la etapa final del proyecto y tiene como objetivo la **consolidación definitiva de la información**, el **registro de metadatos de negocio**, la **migración de datos y documentos** a la base de datos definida por el Ministerio de Educación Nacional, y el **cierre técnico y administrativo del proyecto**.

Durante esta fase, el contratista adelantó el registro de los metadatos de negocio asociados a los actos administrativos previamente analizados y validados en las Fases I y II, así como la preparación, migración y consolidación de la información, garantizando la integridad, trazabilidad y consistencia de los datos.

La gestión de la Fase III se fundamentó en actividades de control, aprobación funcional, monitoreo de desempeño y cierre, asegurando la correcta transición de los resultados del proyecto hacia su uso institucional.

6.1.1. Avance de la Fase III – Migración y consolidación

El avance de la **Fase III** se reporta en un **100 %**, sustentado en el cumplimiento de las actividades planificadas en el Product Backlog y en la ejecución integral de las tareas técnicas, funcionales y administrativas definidas para esta etapa.

El progreso de esta fase se evaluó considerando no solo la migración efectiva de la información, sino también la implementación de controles de versionamiento, validación funcional de los lotes procesados, consolidación del universo total de datos y el cierre formal del proyecto.

Entre las actividades completadas que respaldan el avance del 100 %, se destacan:

Control de versiones y gestión técnica

- Implementación de control de versiones sobre scripts ETL, reglas de negocio y configuraciones asociadas a la actualización de metadatos.
- Gestión controlada de cambios, garantizando trazabilidad técnica y auditabilidad de los ajustes realizados sobre los lotes adicionales procesados.
- Aseguramiento de alineación entre las reglas aplicadas en Fase III y las definiciones establecidas en las Fases I y II.

Estas actividades permitieron mantener la estabilidad del proceso y evitar desviaciones no controladas durante la consolidación final.

Aprobación funcional y control de calidad de lotes adicionales

- Ejecución del proceso de validación funcional de los lotes adicionales procesados.
- Verificación del cumplimiento de criterios de calidad, consistencia, completitud y trazabilidad.
- Formalización de la aprobación funcional previa a la consolidación definitiva de la información.

Este control aseguró que únicamente información validada y conforme a los lineamientos definidos fuera incorporada al universo final del proyecto.

Consolidación del universo total y migración de la información

- Consolidación del universo total de información, integrando el conjunto original de documentos con los registros adicionales incorporados durante el proyecto.
- Registro de evidencias de trazabilidad extremo a extremo, permitiendo identificar origen, transformaciones aplicadas y estado final de cada registro.
- Migración de los datos y documentos con sus respectivos metadatos a la base de datos institucional definida por el MEN.

Este proceso garantizó la integridad de la información y su disponibilidad para explotación institucional.

Monitoreo final y métricas de desempeño

- Monitoreo final de métricas de desempeño del proceso, incluyendo tiempos de procesamiento, precisión de la información, cobertura y estabilidad operativa.
- Validación del cumplimiento de los objetivos técnicos del proyecto frente a los indicadores definidos.

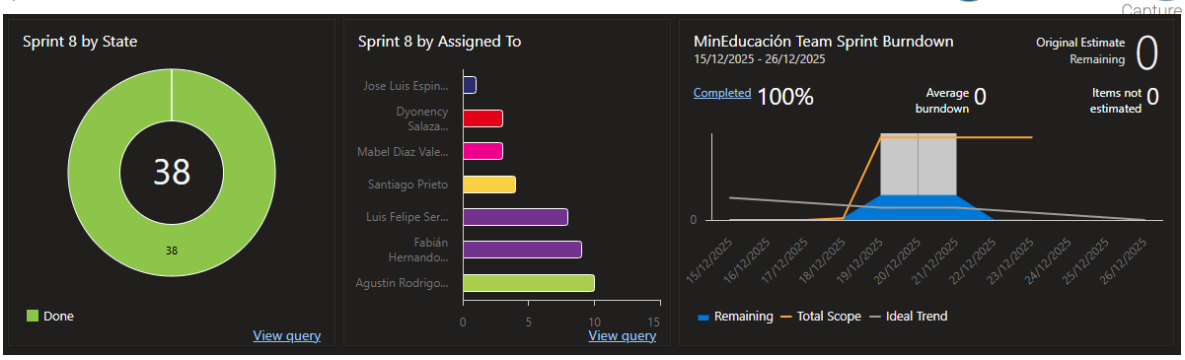
- Documentación de los resultados obtenidos como evidencia del impacto y desempeño del modelo implementado.

Este flujo garantiza que los usuarios puedan interactuar de manera eficiente con los datos y asegurar su consistencia antes de su publicación final.

ID	Work Item Type	Title	Assigned To	State	Tags
1422	Task	Ajustes endpoints Datos	Luis Felipe Serna	Done	Ministerio E
1423	Task	Ajustes de filtros para casos especiales	Luis Felipe Serna	Done	Ministerio E
1424	Task	Definicion de casos de alerta	Luis Felipe Serna	Done	Ministerio E
1425	Task	Revision Casos Extraños. documentos que no se deben mostrar	Luis Felipe Serna	Done	Ministerio E
1426	Task	Definicion de casos de alerta	Luis Felipe Serna	Done	Ministerio E
1427	Task	Consulta Segunda DB para alertas	Luis Felipe Serna	Done	Ministerio E
1428	Task	reajuste de endpoints para envio de alertas RC	Luis Felipe Serna	Done	Ministerio E
1429	Task	reajuste de endpoints para envio de alertas AAC	Luis Felipe Serna	Done	Ministerio E
1430	Task	Convertir y ocultar data de EduSup para comparar con Zat	Fabián Hernando Carri...	Done	Ministerio E
1431	Task	Configurar tablas necesarias para editar	Fabián Hernando Carri...	Done	Ministerio E
1432	Task	Corrección conteo de coincidencias entre zat y EduSup	Fabián Hernando Carri...	Done	Ministerio E
1433	Task	Convertir la data en texto legible y normalizado	Fabián Hernando Carri...	Done	Ministerio E
1434	Task	Habilitar menús desplegables en columnas específicas que permitan filtrar l...	Fabián Hernando Carri...	Done	Ministerio E
1435	Task	Modificar la visualiación de las tablas relacionadas solo para edición y no p...	Fabián Hernando Carri...	Done	Ministerio E
1436	Task	Mostrar los inputs de EduSup que no existen	Fabián Hernando Carri...	Done	Ministerio E
1437	Task	Modificar el estado de comparación para realizar filtros y obtener mayor ac...	Fabián Hernando Carri...	Done	Ministerio E
1438	Task	Implementar porcentaje de coincidencia del campo y valor previo a la imple...	Fabián Hernando Carri...	Done	Ministerio E
1439	Task	Conexion de tablas acreditacion programa y registro calificado al sistema	Jose Luis Espinel Cruz	Done	Ministerio E
1444	Task	Sesión Archivos reglas MEN (interno) Gereneación de INFORME FASE II	Dyonency Salazar Mu...	Done	Ministerio E
1445	Task	Sesión Archivos reglas MEN (interno) Gereneación de INFORME FASE II	Mabel Diaz Valencia	Done	Ministerio E
1446	Task	Sesión Archivos reglas MEN (interno) Gereneación de INFORME FASE II	Dyonency Salazar Mu...	Done	Ministerio E
1447	Task	Sesión Archivos reglas MEN (interno) Gereneación de INFORME FASE II	Mabel Diaz Valencia	Done	Ministerio E
1448	Task	Sesión Archivos reglas MEN (interno) Gereneación de INFORME FASE II	Dyonency Salazar Mu...	Done	Ministerio E
1449	Task	Sesión Archivos reglas MEN (interno) Gereneación de INFORME FASE II	Mabel Diaz Valencia	Done	Ministerio E
1450	Task	Sesión de transformación y cruce resoluciones	Santiago Prieto	Done	Ministerio E
1451	Task	Sesión de transformación y cruce resoluciones	Santiago Prieto	Done	Ministerio E
1452	Task	Sesión de transformación y cruce resoluciones	Santiago Prieto	Done	Ministerio E

XXVIII Backlog actual FASE III

Sprint 8



XXIX Avance Sprint 8

Avance Sprint 8

Estado general: El Sprint 8 contempló un total de 57 tareas, las cuales fueron ejecutadas en su totalidad, sin actividades pendientes ni en estado de ejecución al cierre del ciclo. Este resultado evidencia una gestión efectiva del sprint y un control adecuado de las actividades asociadas a la Fase III – Validación de Consistencia de Datos, asegurando la continuidad operativa y el cumplimiento de los objetivos definidos para la fase.

Asignación: La asignación de tareas presentó una distribución balanceada entre los integrantes del equipo. Se destaca una mayor concentración de actividades en Santiago Prieto, complementada por aportes relevantes de Luis Felipe, Dyonency, Mabel, Fabián Hernando y otros miembros del equipo. Esta articulación permitió atender de manera coordinada las actividades técnicas, funcionales y de validación, garantizando la correcta ejecución de las tareas propias de la fase.

Seguimiento y control (Burndown): El seguimiento del sprint refleja un cumplimiento del 100 % de las tareas planificadas, evidenciando una reducción progresiva del backlog hasta alcanzar cero actividades restantes. Este comportamiento confirma una planificación adecuada, un uso eficiente de la capacidad del equipo y un seguimiento continuo del avance, en línea con las prácticas definidas para la gestión de la Fase III y los lineamientos del documento de calidad de datos

6.2. Protocolo Técnico de Validación

Durante la Fase III, el Contratista implementó y documentó el Protocolo Técnico de Validación para la migración de datos y documentos con metadatos, entregando al Ministerio de Educación Nacional los informes técnicos de migración que acreditan la trazabilidad completa del proceso de transferencia de información, la bitácora de eventos, los resultados de pruebas unitarias, pruebas integrales y la validación post-migración, certificando un cumplimiento del 100 % en la precisión de los datos migrados. Como parte integral de los entregables, se suministraron los scripts de migración organizados en dos (2) grupos: (i) scripts DDL, correspondientes a modificaciones estructurales de las tablas (creación y ajuste de columnas y tipos de datos), los cuales deben ser ejecutados previamente por el Ministerio en el orden definido en los nombres de los archivos; y (ii) scripts de carga de información, organizados en secuencia alfabética, que contienen los INSERT a cada una de las tablas y deben ejecutarse respetando dicho orden para garantizar la consistencia e integridad de la

información. Adicionalmente, se entregó el log de auditoría completo que registra los ajustes, validaciones y procesos automatizados realizados sobre los datos, incluyendo los eventos generados por plataformas RPA, componentes de inteligencia artificial y actuaciones de usuarios autorizados, constituyendo evidencia verificable del cumplimiento técnico, funcional y contractual de la fase.

6.2.1. Preparación de Insumos para Migración

- Consolidar el conjunto final de metadatos validados en Fase I y Fase II, correspondientes a los actos administrativos de:
    - Inspección y Vigilancia
    - Registro Calificado
    - Programas Acreditados en Alta Calidad
    - Instituciones Acreditadas en Alta Calidad
- Validar que la información a migrar corresponda exclusivamente a los registros aprobados para Fase III, sin pendientes de corrección funcional.
- Confirmar el alcance de las doce (12) tablas objetivo sobre las cuales se realizará la migración:
    - INS_RESOLUCION_INSPECCION_VIGILANCIA
    - INS_INFORME_INSPECCION_VIGILANCIA
    - INS_PLAN_VISITA_INSPECCION_VIGILANCIA
    - INS_EVALUACION_VIABILIDAD_INSPECCION_VIGILANCIA
    - PRO_HISTORICO_REGISTRO_CALIFICADO
    - PRO_HISTORICO_MUNICIPIO_RC
    - PRO_HISTORICO_ACREDITACION_PROGRAMA
    - PRO_HISTORICO_MUNICIPIO_ACREDITACION_PROG
    - INS_HISTORICO_ACREDITACION_INSTITUCION
    - INS_HISTORICO_MUNICIPIO_ACREDITACION_INST
    - PRO_INFO_RESOLUCION
    - INS_INFO_RESOLUCION

6.2.2. Entrega y Organización de Scripts de Migración

- Entregar al Ministerio el paquete scripts conteniendo la totalidad de scripts requeridos para la ejecución de la Fase III.
- Clasificar los scripts en dos (2) grupos claramente diferenciados:

Scripts DDL (Estructura de datos)

- Scripts destinados a la modificación de la estructura de las tablas (creación de columnas, ajuste de tipos de datos, adecuaciones de esquema).

- Estos scripts deben ejecutarse obligatoriamente en primer lugar.
- El orden de ejecución está determinado por el nombre de cada archivo.
- La ejecución será realizada por el Ministerio de Educación Nacional o por el equipo técnico que este designe.
- La correcta ejecución de estos scripts habilita la estructura necesaria para la carga de información.

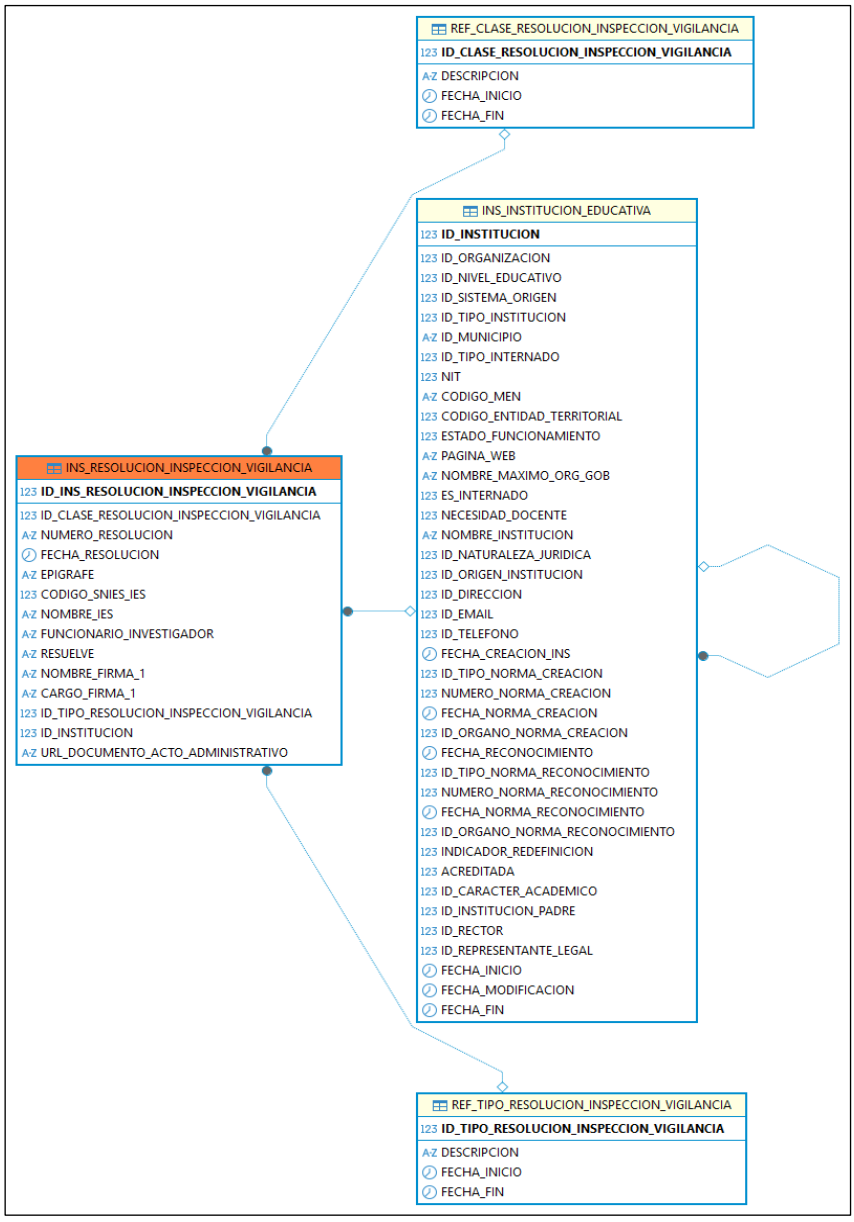
Scripts de carga de datos (INSERT)

- Scripts que contienen las sentencias INSERT para poblar cada una de las doce (12) tablas.
- Se encuentran organizados en orden alfabético, el cual define la secuencia obligatoria de ejecución.
- Solo podrán ejecutarse una vez finalizada y validada la ejecución de los scripts DDL.
- Garantizan la carga controlada, íntegra y trazable de los metadatos migrados.

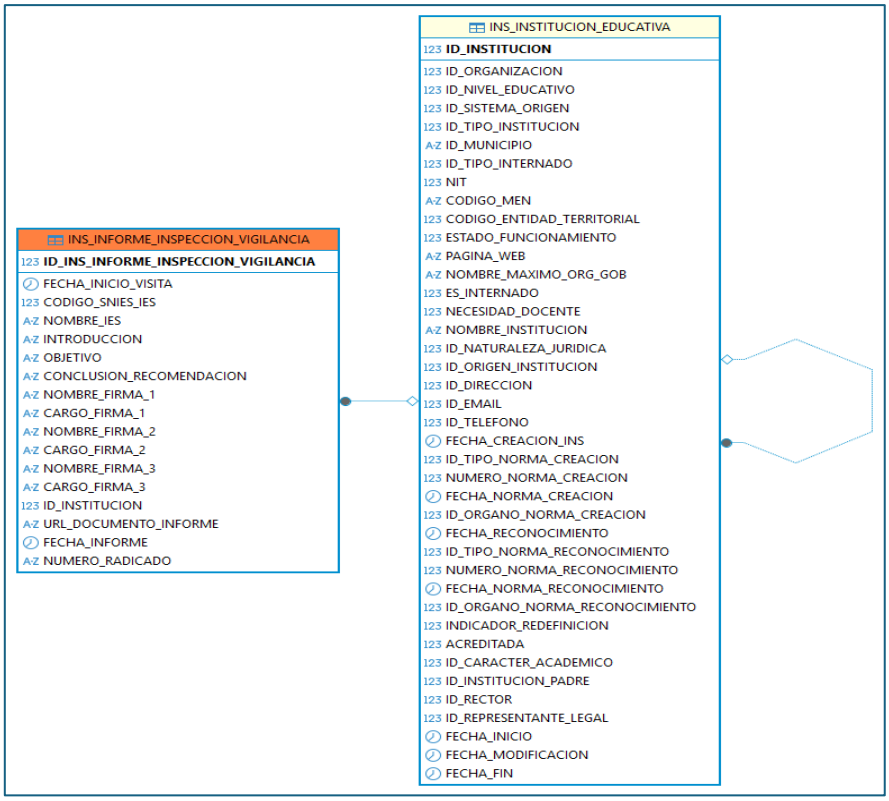
6.3. Modelo de Entidad Relación (MER)

El Modelo de Entidad-Relación (MER) establece la estructura de la base de datos que soporta la gestión y trazabilidad de la información asociada a los procesos de Inspección y Vigilancia, Registro Calificado, Programas Acreditados en Alta Calidad e Instituciones Acreditadas, mediante doce (12) tablas definidas de forma autónoma. El diseño prioriza la integridad individual de los datos, el almacenamiento histórico y la correcta clasificación de resoluciones, informes, evaluaciones, planes de visita y acreditaciones, permitiendo cargas

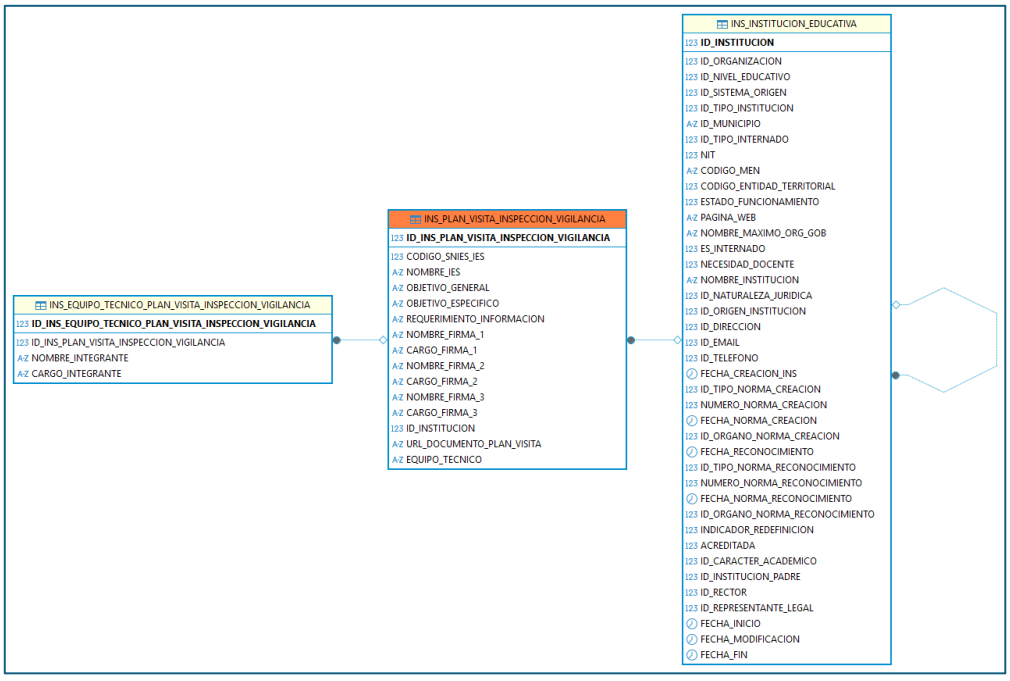
incrementales, reprocesamientos controladas y validaciones independientes, en alineación con los requerimientos de control, auditoría y seguimiento institucional.



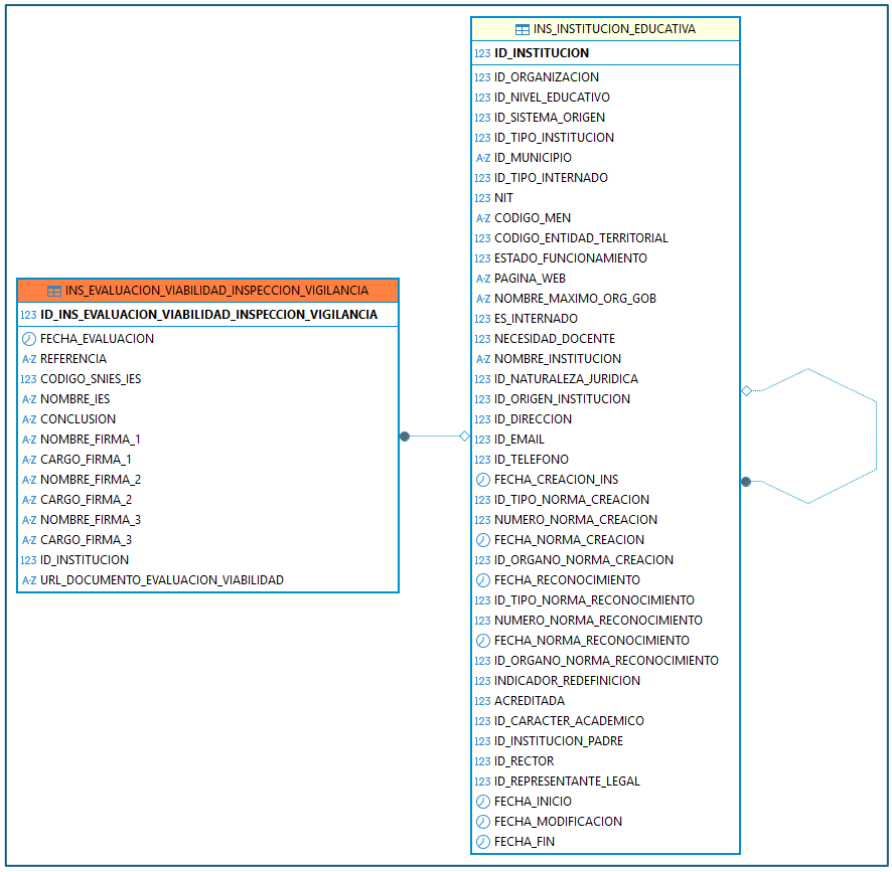
XXX [dbo].[INS_RESOLUCION_INSPECCION_VIGILANCIA]



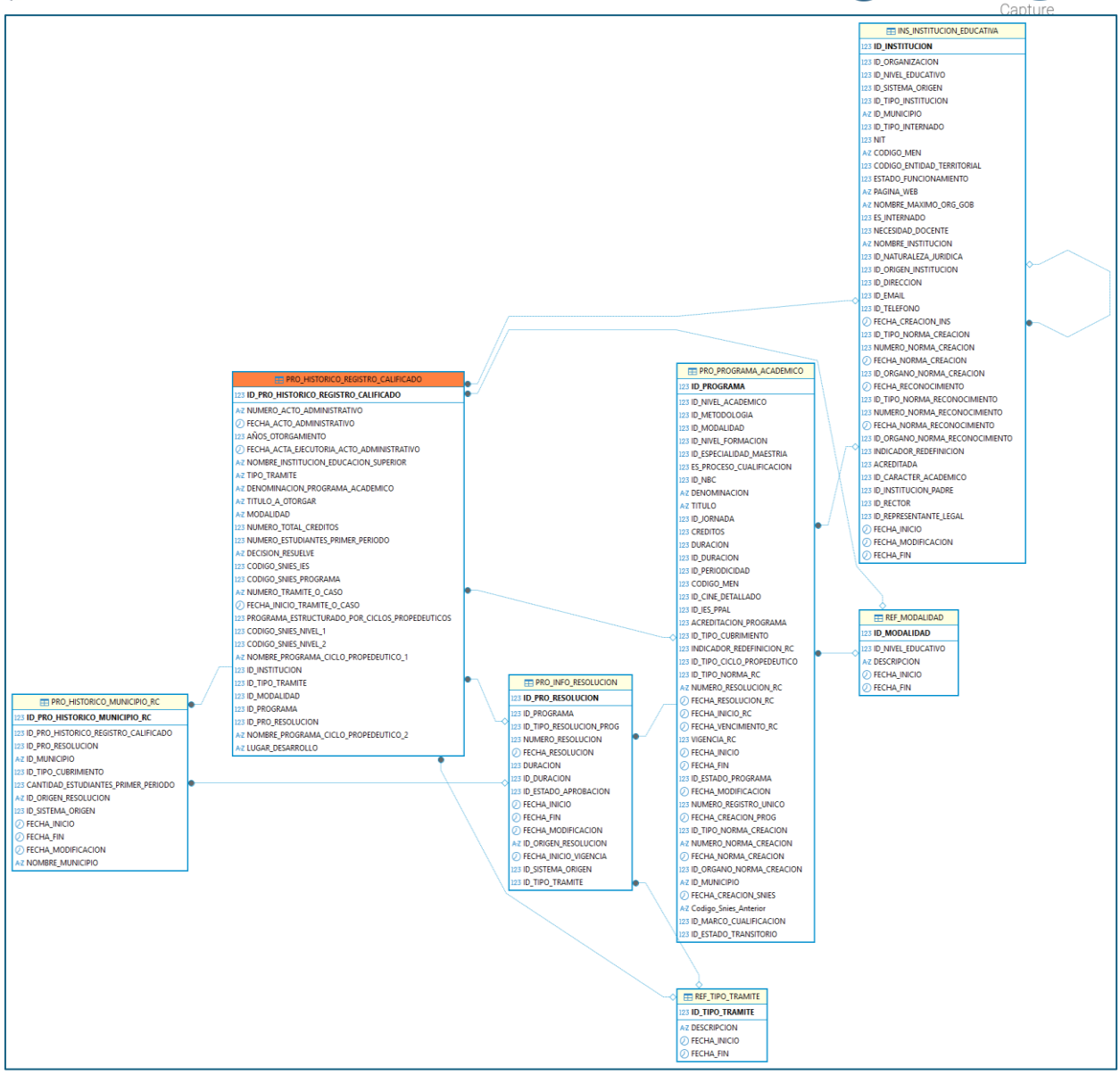
XXXI [dbo].[INS_INFORME_INSPECCION_VIGILANCIA]



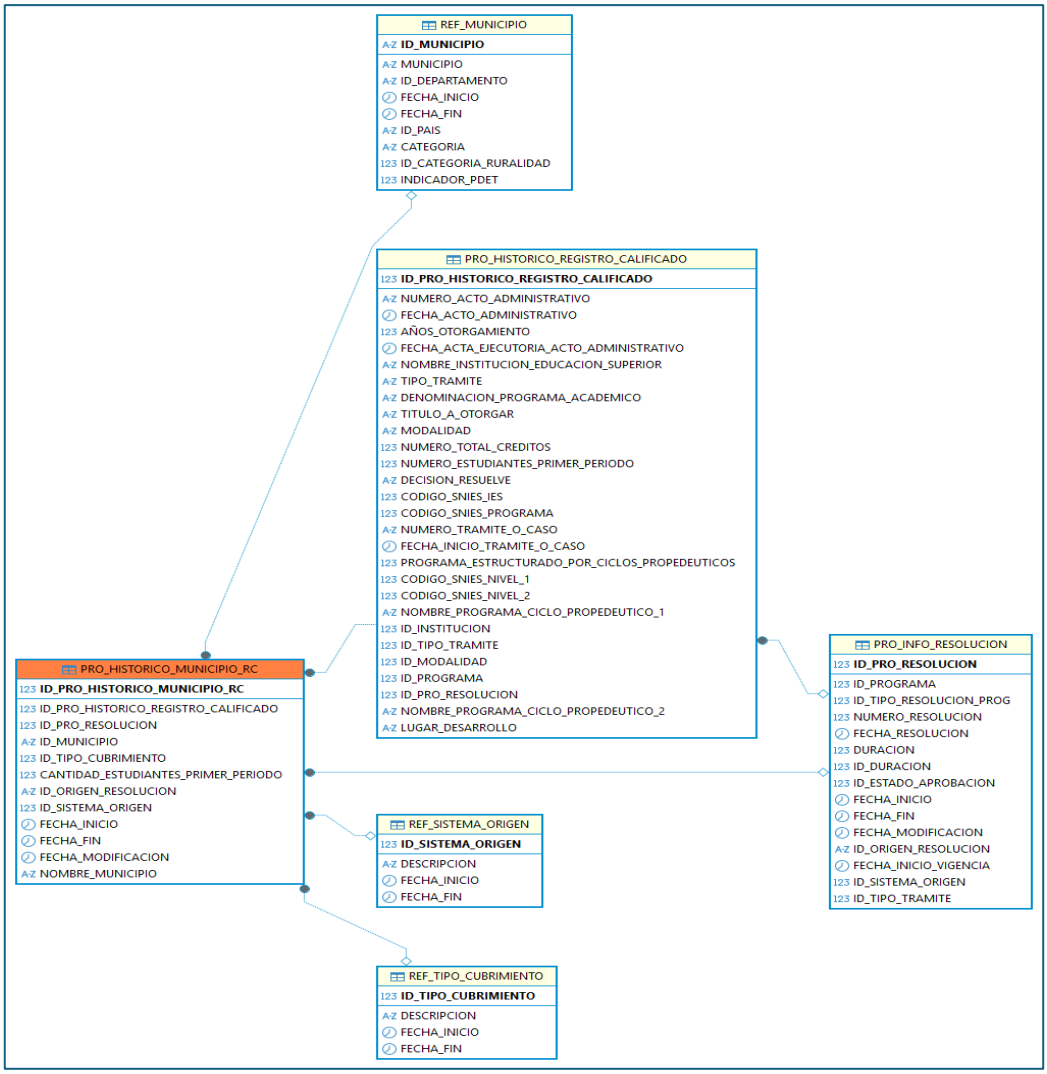
XXXII [dbo].[INS_PLAN_VISITA_INSPECCION_VIGILANCIA]



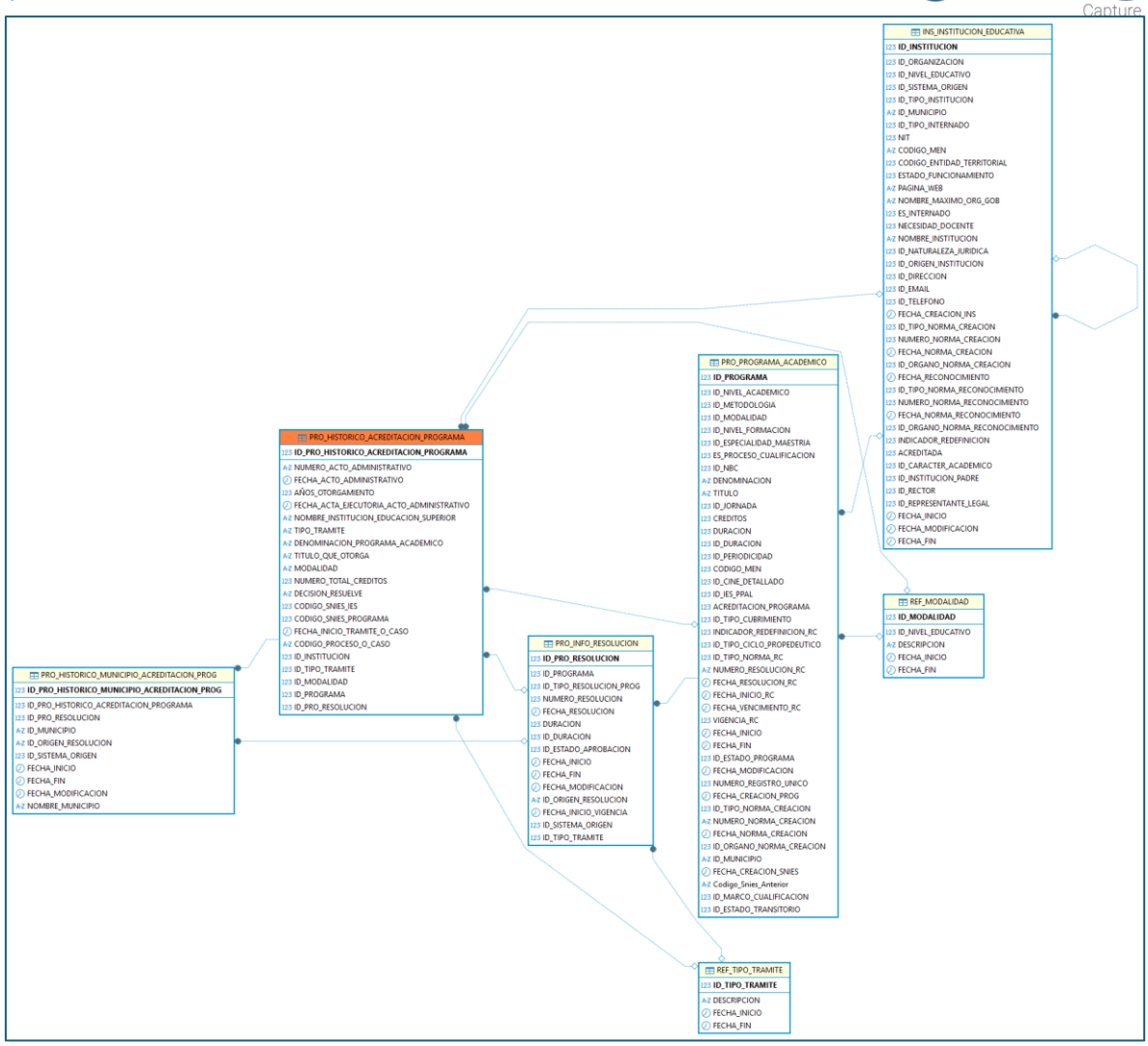
XXXIII [dbo].[INS_EVALUACION_VIABILIDAD_INSPECCION_VIGILANCIA]



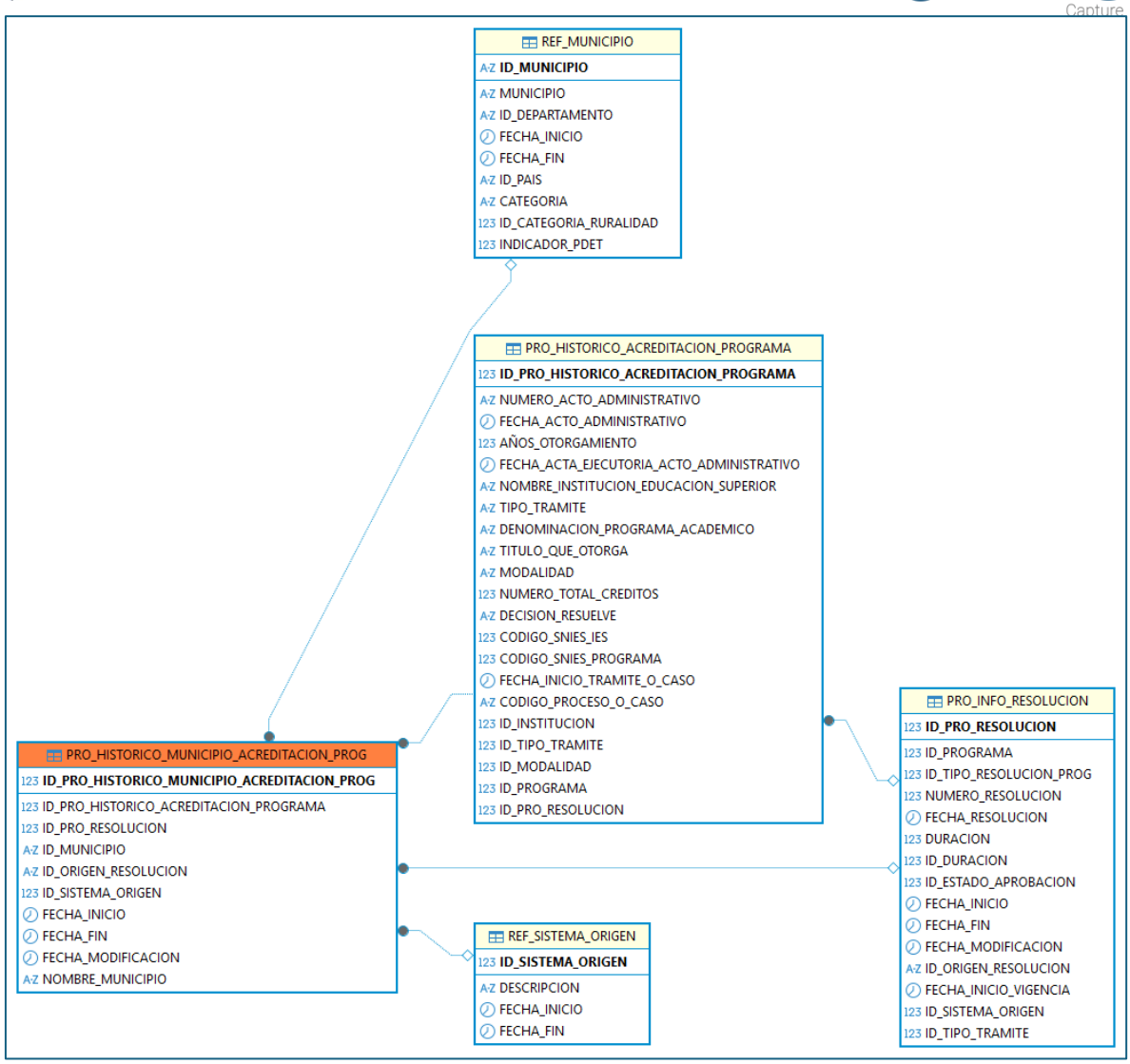
XXXIV [dbo].[PRO_HISTORICO_REGISTRO_CALIFICADO]



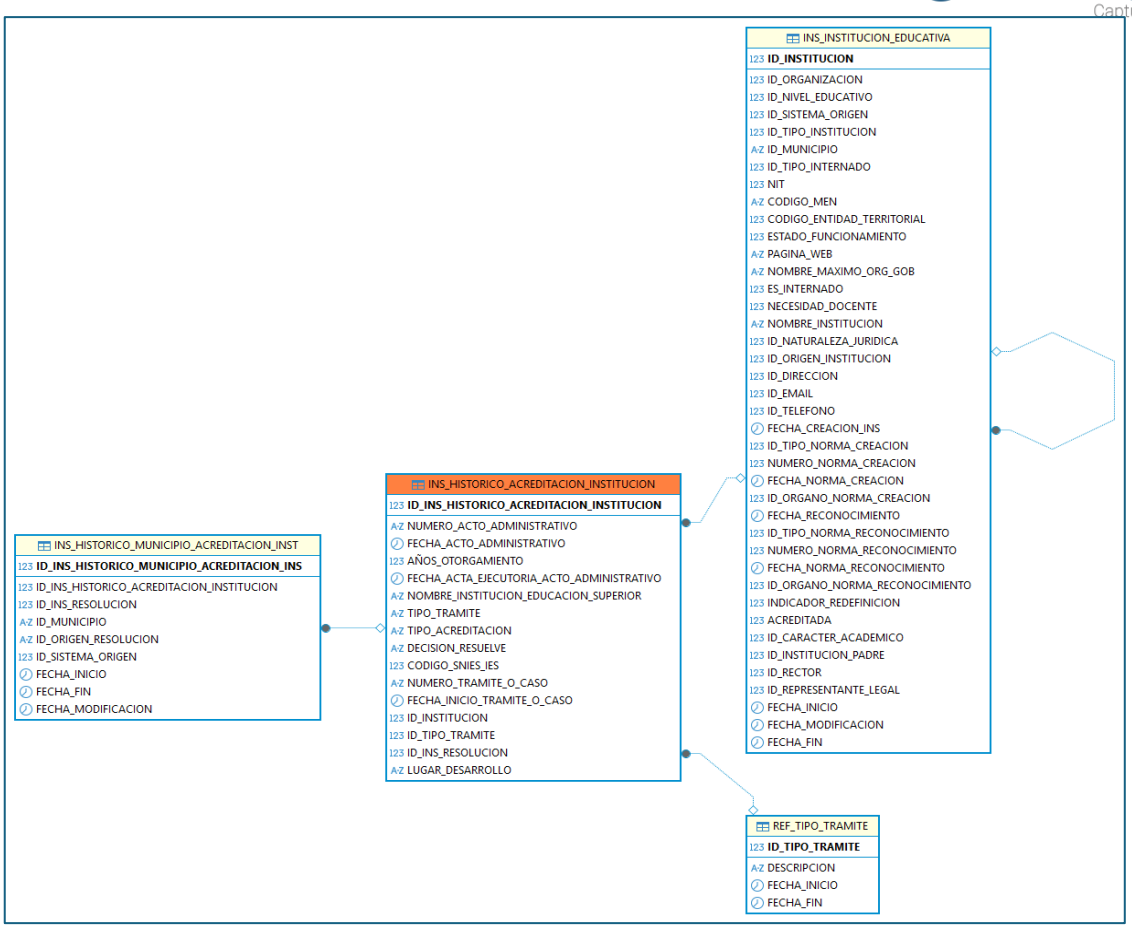
XXXV [dbo].[PRO_HISTORICO_MUNICIPIO_RC]



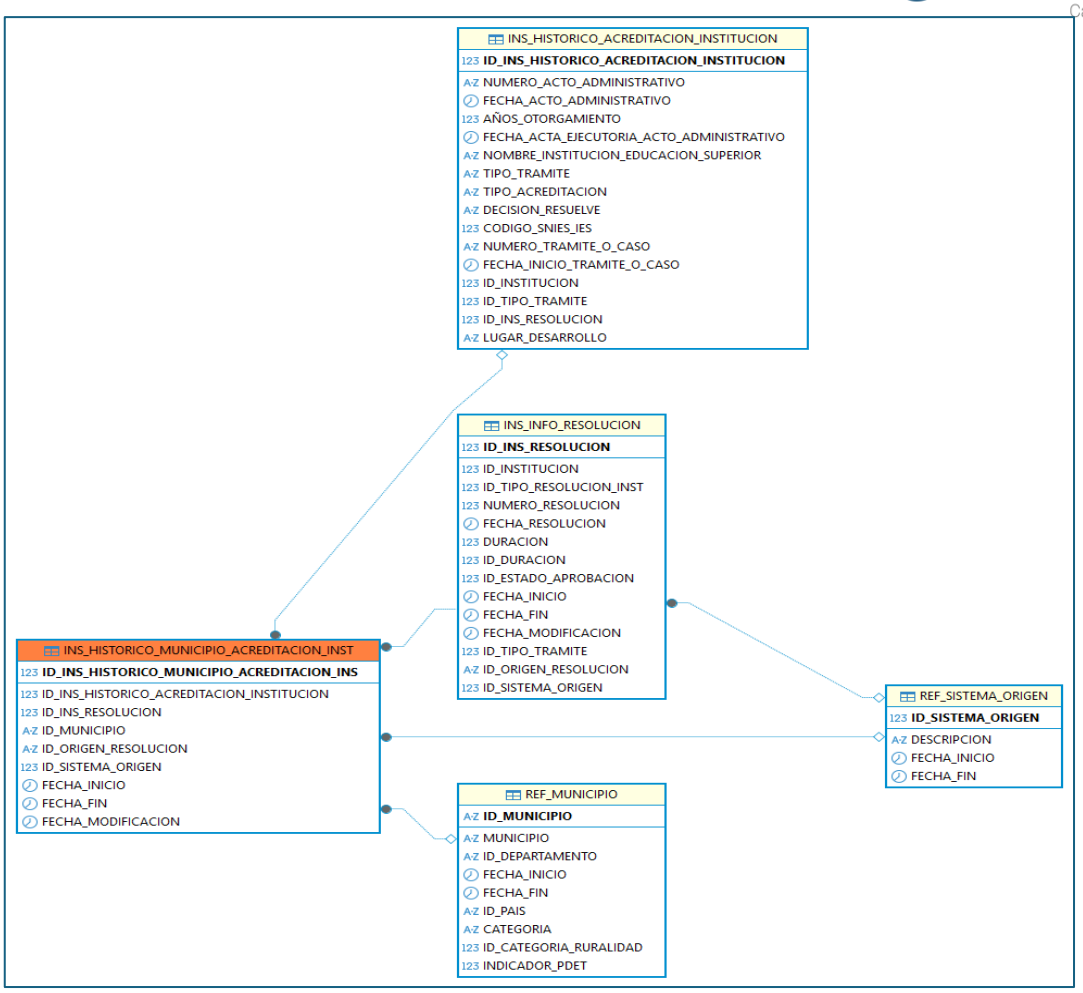
XXXVI [dbo].[PRO_HISTORICO_ACREDITACION_PROGRAMA]



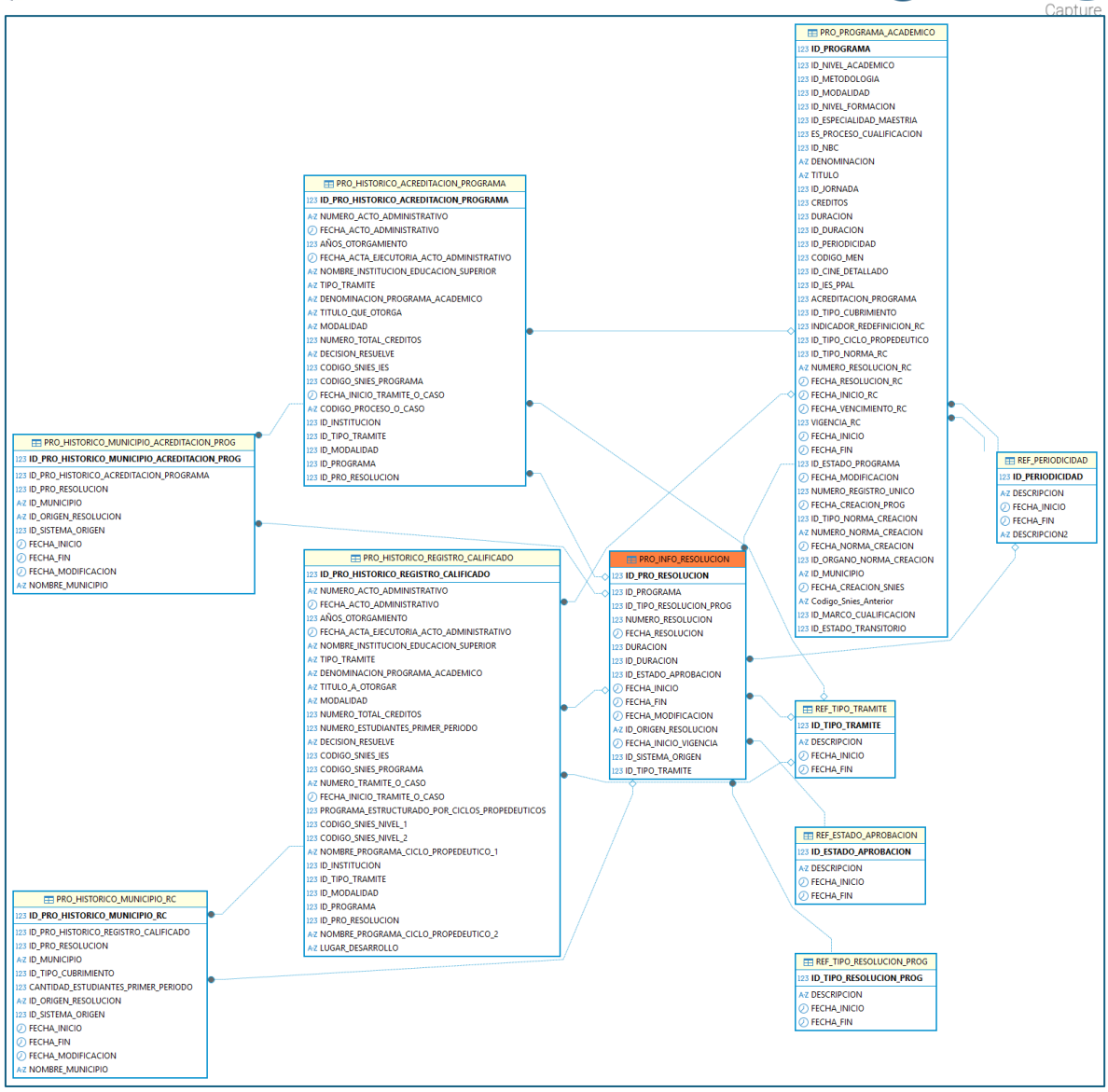
XXXVII [dbo].[PRO_HISTORICO_MUNICIPIO_ACREDITACION_PROG]



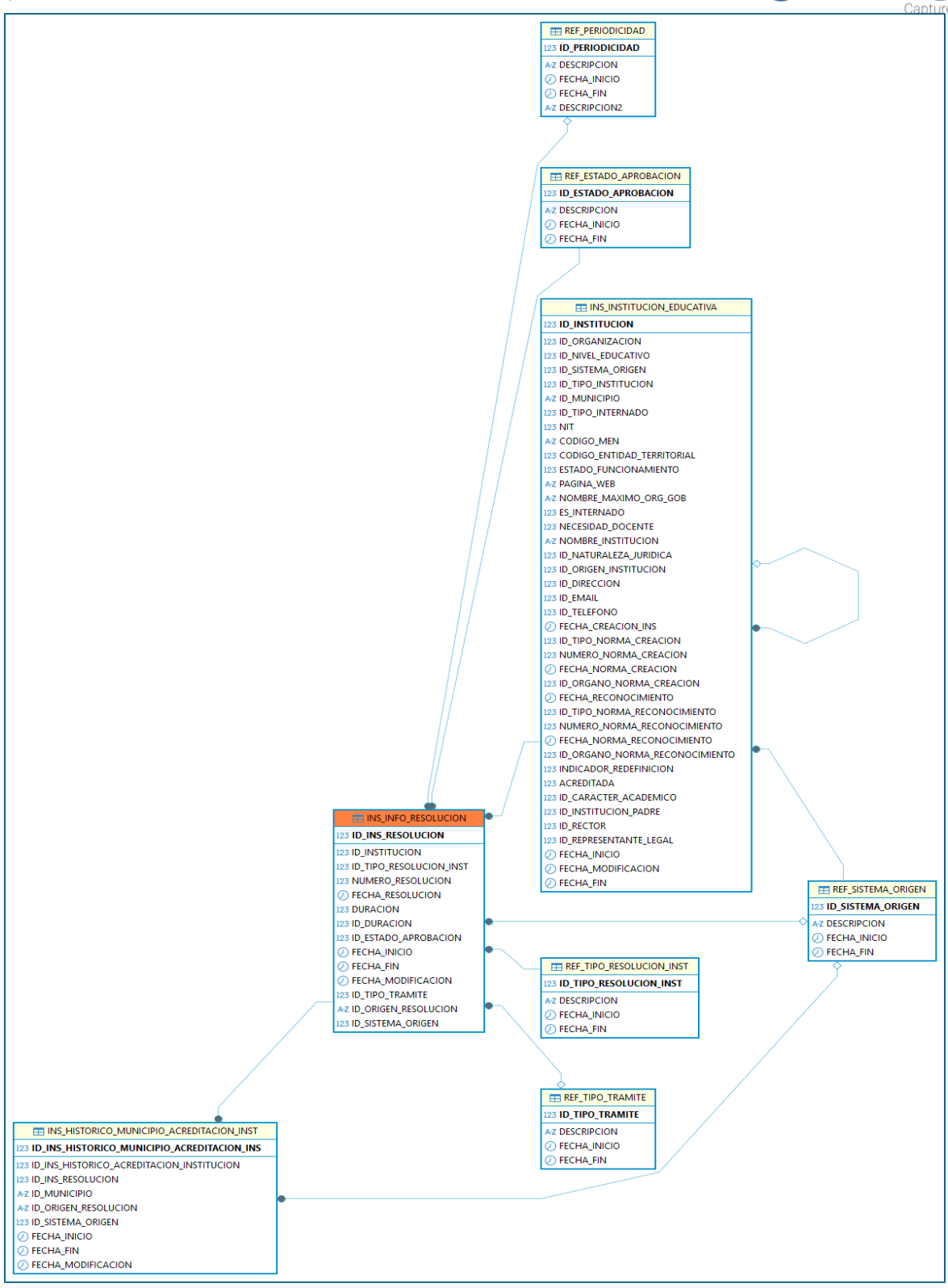
XXXVIII [dbo].[INS_HISTORICO_ACREDITACION_INSTITUCION]



XXXIX [dbo].[INS_HISTORICO_MUNICIPIO_ACREDITACION_INST]



XL [dbo].[PRO_INFO_RESOLUCION]



XLI [dbo].[INS_INFO_RESOLUCION]

6.3.1. Ejecutar los scripts de INSERT conforme al orden alfabético definido, asegurando:

- Correspondencia entre registros fuente y registros cargados.
- Ausencia de pérdidas o duplicidades.
- Integridad referencial a nivel de cada tabla.

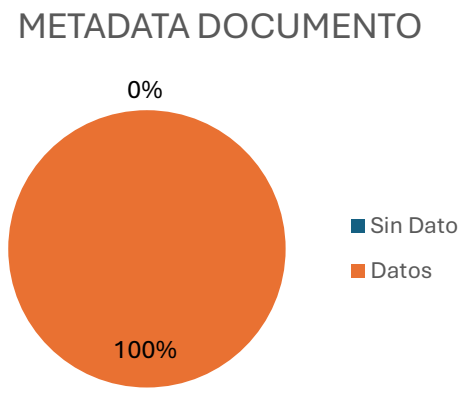
6.4. Informe consolidado de metadatos

Con base en los resultados obtenidos en la extracción de metadatos del portal SGDEA se presentó el informe teniendo en cuenta la estructura solicitada para garantizar el Control de calidad de datos, donde se logró obtener la siguiente cantidad de campos extraído.

METADATA DOCUMENTO

- Numero De Documento
- Fecha Del Documento
- Asunto
- Tipo Del Documento
- Numero De Folios

De los campos correspondientes anteriores a Metadata de documento se logró una extracción total de los campos.



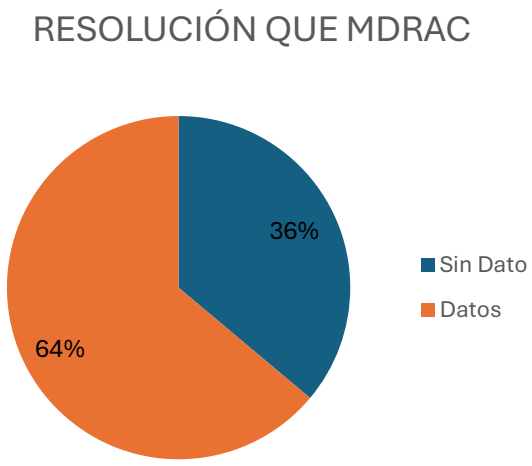
XLII MetaData

METADATA DE RESOLUCIONES

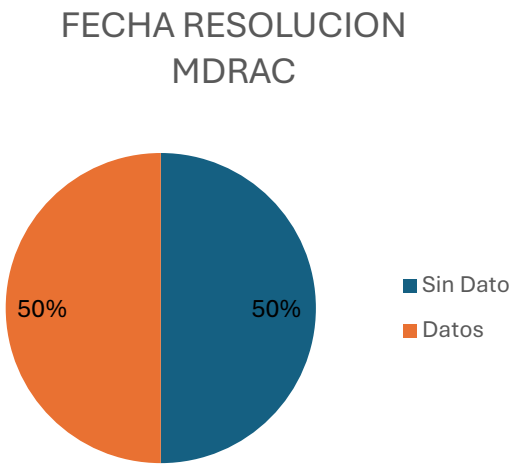
- Dependencia Origen (nombre dependencia que firma el acto administrativo)
- Usuario Originador (nombre de la persona que firma el acto administrativo)
- Lugares Descritos (todas las ciudades que se nombran en el acto administrativo)
- Resolución Que Mdrac (si nombra otro acto o modifica deroga o resuelve se debe colocar el número de esta)

- Fecha Resolución Mdrac
- Tipo De Identificación (se coloca tipo de identificación de todas las personas que nombran en el acto si se encuentran)
- Nombres Y Apellidos (se colocan todos los nombres y apellidos de todas las personas que nombran en el acto)
- Institución (se colocan todos los nombres de las IES de todas las que nombran en el acto)
- Número De Identificación (se colocan todos los números de identificación de todas las personas que nombran en el acto si se encuentran).
- Se encuentra PDF (se coloca SI o NO, de acuerdo a la identificación de PDF en el portal de SGDEA).

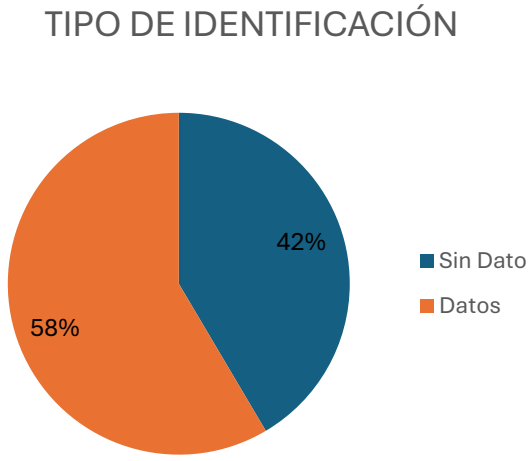
De los campos correspondientes anteriores a Metadata de resoluciones se logró una extracción total de los campos, teniendo en cuenta que los campos relacionados en las siguientes graficas se evidencia que no se encontraron todos los valores correspondientes en el SGDEA.



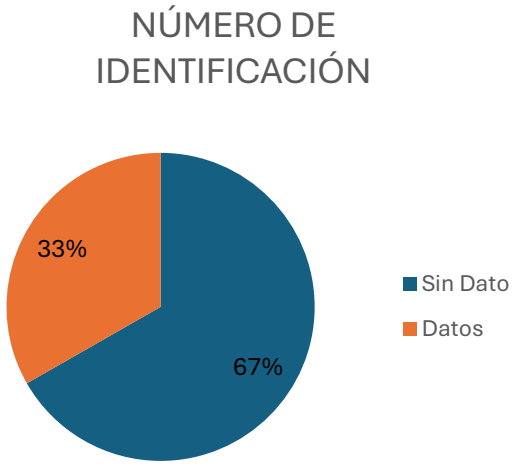
XLIII Resolución que MDRA



XLIV Fecha Resolución



XLV Tipo de Identificación



XLVI Número de Identificación



XLVII Se encuentra PDF

Anexo 16_ Control Calidad Data-SGDEA

6.5.Hallazgos

- **Hallazgo 1: Identificación de dependencias técnicas para la ejecución controlada de scripts de migración (Fase III).**
  Descripción:
  Durante la preparación y entrega de los scripts de migración correspondientes a la Fase III, se evidenció la capacidad de memoria para la correcta ejecución del proceso

depende de una secuencia estricta entre scripts DDL y DML, así como de la existencia previa de estructuras, permisos y datos base en la base EDU_SUP. Si bien los scripts fueron entregados organizados, estas dependencias no estaban explícitamente contempladas como prerrequisitos técnicos en la planeación inicial de la fase.

Impacto:

Una ejecución fuera de secuencia o sin el cumplimiento de los prerrequisitos técnicos puede generar errores en la carga, inconsistencias parciales de información o reprocesos innecesarios, afectando la trazabilidad y confiabilidad del proceso de migración de datos.

Acción:

Se documenta de manera detallada el orden de ejecución de los scripts, se incorpora un checklist de prerrequisitos técnicos y se incluye esta información en el Manual de Uso y Protocolo de Migración, con el fin de garantizar una ejecución controlada y auditable en los ambientes del MEN.

- **Hallazgo 2: Dependencia institucional para la ejecución y evidencia de los casos de prueba de migración (Fase III).**

  Descripción:

  Se identificó que, aunque los casos de prueba funcionales y técnicos fueron diseñados y entregados conforme al estándar MEN para validar la migración de datos, la ejecución efectiva de los mismos y la generación de evidencias finales dependen exclusivamente de la disponibilidad y programación de los ambientes institucionales del MEN.

  Impacto:

  Esta dependencia puede generar desfases entre la entrega técnica y la certificación formal de la migración, impactando los tiempos de cierre de la fase y la consolidación de evidencias requeridas para auditoría o interventoría.

  Acción:

  Se entrega el conjunto completo de casos de prueba documentados, incluyendo criterios de aceptación, datos esperados y evidencias requeridas, y se acuerda con el MEN un esquema de acompañamiento técnico para soportar la ejecución y validación en ambiente institucional.

- **Hallazgo 3: Riesgo de adopción parcial del Manual de Uso y Protocolo de Migración (Fase III).**

  Descripción:

  Durante la socialización del Manual de Uso y del Protocolo de Migración se identificó que la correcta operación del sistema ZATDATASYNC y la ejecución

controlada de reprocesos y migraciones futuras dependen de la apropiación efectiva del documento por parte de los equipos técnicos y funcionales del MEN.

Impacto:

Una adopción parcial del manual puede incrementar el riesgo operativo en futuras ejecuciones, reprocesos o ajustes al modelo de datos, afectando la estandarización, el control y la continuidad del proceso.

Acción:

Se consolida el manual como documento único de referencia, se incluyen diagramas de flujo, responsabilidades por rol y controles de validación, y se recomienda su uso obligatorio como insumo base para cualquier ejecución o reproceso asociado a la migración de datos.

- **Hallazgo 4: Necesidad de consolidación de evidencias de trazabilidad extremo a extremo (Fase III).**

  Descripción:

  Se evidenció que las evidencias del proceso de migración se generan en distintos momentos y actores del proceso (scripts, ejecución, resultados de pruebas y validaciones funcionales), lo que implica la necesidad de una consolidación centralizada para efectos de auditoría y control

- **Impacto:**

  La dispersión de evidencias puede incrementar el esfuerzo requerido para atender requerimientos de entes de control, auditorías internas o revisiones posteriores, afectando la eficiencia del proceso de verificación.

- **Acción:**

  Se define un esquema de consolidación de evidencias que integra scripts ejecutados, resultados de pruebas, logs y actas de validación, alineado con los lineamientos de trazabilidad y control del MEN.

- **Hallazgo 5: Ajustes estructurales y solicitudes de modificación sobre metadatos (Fase III).**

  Descripción:

  Durante la ejecución de la Fase III, el Ministerio realizó un seguimiento detallado sobre un conjunto específico de registros migrados, identificando la necesidad de efectuar ajustes estructurales y funcionales en la información consolidada.

  Entre las solicitudes recibidas se encuentran:

  - **Eliminación de registros.**
    - Primera corrección:
      - Resolución: 14 casos
      - Auto: 14 casos

- Informe: 442 casos
- Viabilidad: 14 casos
- Plan visita: 37 casos

- Segunda corrección:
  - Resolución: 11 casos
  - Auto: 17 casos
  - Informe: 21 casos
  - Viabilidad: 12 casos
  - Plan visita: 10 casos
- Tercera corrección:
  - Resolución: 0 casos
  - Auto: 0 casos
  - Informe: 1 caso
  - Viabilidad: 0 casos
  - Plan visita: 0 casos
- Cuarta corrección:
  - Resolución: 4 casos
  - Auto: 0 casos
  - Informe: 0 casos
  - Viabilidad: 0 casos
  - Plan visita: 0 casos

- **Modificación de datos de columnas existentes**, tales como Nombre Universidad, Investigador, Fecha Visita y Código SNIES.
  - Primera corrección:
    - Resolución: 55 casos
    - Auto: 30 casos
    - Informe: 246 casos
    - Viabilidad: 77 casos
    - Plan visita: 60 casos
  - Segunda corrección:
    - Resolución: 0 casos
    - Auto: 0 casos
    - Informe: 788 casos
    - Viabilidad: 2 casos
    - Plan visita: 1 caso
  - Tercera corrección:
    - Resolución: 0 casos
    - Auto: 0 casos
    - Informe: 1 caso

- Viabilidad: 0 casos
- Plan visita: 0 casos

- Cuarta corrección:
  - Resolución: 0 casos
  - Auto: 0 casos
  - Informe: 0 casos
  - Viabilidad: 0 casos
  - Plan visita: 0 casos

- **Adición y actualización de valores de nuevas columnas en el modelo de datos,** tales como "ID preventiva o sancionatoria", Preventiva o sancionatoria", "Proceso" y "Fecha de Visita".
  - Primera corrección:
    - **Adición**: Se solicita creación de columnas ID preventiva o sancionatoria" y Preventiva o sancionatoria" para Resolución.
  - Segunda corrección:
    - Resolución: 24 casos actualizados
    - Auto: 2 casos actualizados
    - Informe:
      - **Adición**: Se solicita creación de columna "Proceso".
    - Viabilidad:
      - **Adición**: Se solicita creación de columna "Proceso".
    - Plan visita:
      - **Adición**: Se solicita creación de columnas "Proceso" y "Fecha de Visita".
  - Tercera corrección:
    - Resolución: 0 casos
    - Auto: 0 casos
    - Informe: 0 casos
    - Viabilidad: 0 casos
    - Plan visita: 319 casos actualizados
  - Cuarta corrección:
    - Resolución: 8 casos actualizados
    - Auto: 0 casos
    - Informe: 0 casos
    - Viabilidad: 0 casos
    - Plan visita: 0 casos

Impacto:

Las solicitudes de modificación sobre registros previamente procesados implicaron

la revisión y ajuste de estructuras de datos, reglas de transformación y scripts de procesamiento asociados al flujo de migración.

Estas modificaciones generaron la necesidad de realizar reprocesamientos controlados sobre casos específicos, así como validaciones adicionales para garantizar la consistencia y trazabilidad de la información dentro del modelo de datos consolidado.

Acción:

Con el fin de gestionar adecuadamente estos ajustes, se implementó un proceso de seguimiento y control basado en ciclos formales de observación y corrección. Para cada ciclo se elaboró un **informe de observaciones** y su correspondiente **informe de corrección**, en los cuales se registraron los cambios solicitados, el tratamiento aplicado y la validación de los resultados obtenidos.

Estos documentos se incluyen como anexos del presente informe, con el fin de garantizar la trazabilidad del proceso de revisión y de las modificaciones realizadas durante la Fase III a mayor detalle *(Ver anexos 17 – 24).*

- **Hallazgo 6: Identificación de registros con estado "CON_INCONSISTENCIAS", en el FrontEnd para Programas Acreditados en Alta Calidad y Registros Calificados (Fase III).**

  Descripción:

  Durante la revisión de la información cargada en el FrontEnd correspondiente a los módulos de Programas Acreditados en Alta Calidad y Registros Calificados, se identificó la presencia del estado “CON_INCONSISTENCIAS” en algunos registros. Este estado puede generarse por múltiples razones asociadas a diferencias o inconsistencias en los datos migrados, tales como variaciones en el nombre de la institución, denominación o título del programa, tipo de modalidad, años de otorgamiento, entre otros campos relevantes. Dado que estas inconsistencias requieren validación funcional o ajustes que no pueden resolverse de manera automatizada, los registros identificados son consolidados en listados que se remiten al Ministerio para su revisión y corrección manual.

  Impacto:

  La existencia de registros con estado “CON_INCONSISTENCIAS” puede limitar la disponibilidad de información validada en el sistema y requerir actividades adicionales de revisión y corrección por parte de los equipos funcionales del MEN.

  Acción:

  Se consolida y entrega al Ministerio el listado de registros que presentan el estado “CON_INCONSISTENCIAS”, con el fin de facilitar su revisión y corrección manual por parte de los responsables funcionales. Este procedimiento permite asegurar la depuración progresiva de los registros y contribuir a la consistencia final de la información cargada en el sistema.

Estos documentos se incluyen como anexos del presente informe, con el fin de visualizar el detalle de cada caso reportado *(Ver anexos 25 y 26)*.

7. Conclusiones

- El proyecto **"Calidad de Datos"** se ejecutó de manera satisfactoria a través de las Fases I, II y III, cumpliendo los lineamientos técnicos, funcionales y contractuales definidos por el Ministerio de Educación Nacional (MEN), y garantizando la calidad, trazabilidad y confiabilidad de la información gestionada a lo largo del proceso.
- En la **Fase I – Extracción de datos**, se implementó de forma exitosa una arquitectura RPA/IA para la automatización de la captura y procesamiento de metadatos desde el portal SGDEA y las carpetas institucionales de SharePoint. La solución demostró estabilidad operativa sobre la infraestructura de Azure y alcanzó niveles de precisión y completitud superiores al umbral técnico establecido, validado mediante ejercicios de medición ponderada. Como resultado, se consolidaron **61.307 casos** como base maestra del proyecto, constituyéndose en el insumo fundamental para las etapas posteriores de validación y análisis.
- La **Fase II – Validación de consistencia de datos** permitió consolidar un modelo robusto de control y verificación de la información, mediante la integración de la base de extracción de la Fase I con el respaldo de la base EDU_SUP. Se construyó una base intermedia canónica en **Azure Databricks**, definida como punto único de verdad, sobre la cual se aplicaron reglas de transformación, homologación y consistencia para los procesos de Inspección y Vigilancia, Registro Calificado y Acreditaciones. El procesamiento incluyó la detección automática de diferencias, la clasificación de registros por estado y la generación de casos trazables para validación funcional. Adicionalmente, se diseñó y validó la solución de verificación mediante mockups de frontend, habilitando un flujo controlado de revisión y corrección por parte de los usuarios del MEN, con auditoría de cambios.
- En la **Fase III – Definición de metadatos y migración de la información**, se ejecutó la estandarización del modelo de metadatos y la migración efectiva de los datos validados hacia las estructuras definitivas del modelo de datos institucional. El proceso se realizó de manera controlada y verificable, mediante la ejecución de ajustes estructurales y cargas ordenadas de información, garantizando la integridad, consistencia y trazabilidad de los registros migrados. La migración cubrió información histórica y operativa asociada a los dominios funcionales del proyecto, dejando los datos disponibles en condiciones de uso institucional y cumpliendo los criterios de calidad, control y auditabilidad establecidos.
- Como resultado integral, el proyecto deja habilitada una solución técnica y operativa madura, escalable y alineada con los lineamientos del MEN, que asegura la calidad de los datos, soporta su explotación institucional y permite el cierre contractual en condiciones satisfactorias.

8. Propuesta de Mejora: Implementación de Visor de Datos para Gestión y Validación de Información.

8.1. Descripción General

Como parte de las oportunidades de mejora identificadas durante el desarrollo del proyecto, se propone la implementación de un **visor web para la visualización, consulta y edición controlada de la información procesada**.

Este visor permitiría facilitar las actividades de revisión, validación y gestión de los registros asociados a **los actos administrativos**, centralizando la información en una única herramienta de consulta.

La solución propuesta estaría compuesta por los siguientes componentes:

- **Frontend:** interfaz web para la visualización y edición de datos.
- **Backend (API):** componente encargado de la lógica de negocio y la conexión con la base de datos.
- **Base de Datos:** repositorio para el almacenamiento de los datos históricos procesados.
- **Servidor de Documentos:** componente destinado al almacenamiento y visualización de los documentos asociados.

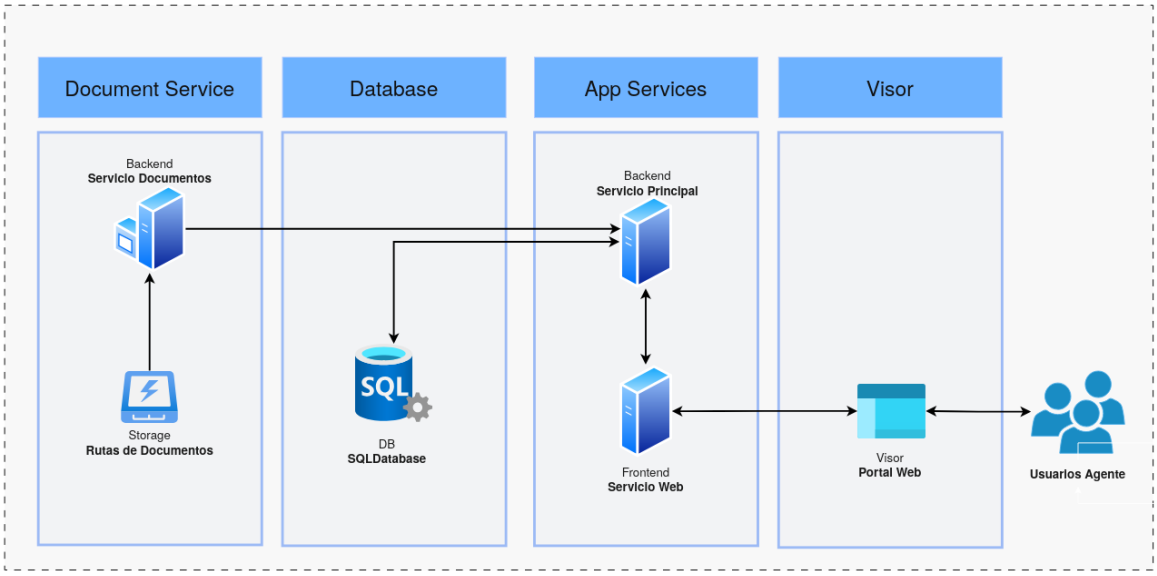
8.2. Arquitectura Propuesta

El visor estaría compuesto por un servicio de backend encargado de la lógica de negocio y un frontend responsable de la capa de presentación y experiencia de usuario.

Ambos componentes se encontrarían desacoplados, siguiendo buenas prácticas de arquitectura que permiten mejorar la escalabilidad, el mantenimiento y la evolución independiente de cada capa del sistema.

Adicionalmente, se propone un servicio independiente para la gestión de documentos, el cual se desplegaría en el mismo servidor donde se almacenan los archivos y tendría como responsabilidad exponerlos de manera controlada.

En este esquema, el backend actuaría como proxy hacia el servidor de documentos, centralizando el acceso a través de una única API. Esto permitiría gestionar de manera unificada los mecanismos de autenticación, autorización y control de seguridad.



XLVIII Propuesta de mejora - Arquitectura

8.3. Requerimientos de Infraestructura (Propuesta)

- **Cómputo:** Para el despliegue del visor se estiman los siguientes recursos:
    - **Backend**
        - 1 VM o Azure App Service
        - App Service – Basic Tier (B2 / B3)
        - 2 cores
        - 3.5 GB RAM
        - 10 GB almacenamiento
    - **Frontend**
        - 1 VM o App Service independiente
        - App Service – Basic Tier (B2 / B3)
        - 2 cores
        - 3.5 GB RAM
        - 10 GB almacenamiento
    - **Servidor de Documentos**
        - 1 VM dedicada
        - Tipo D2s v5
        - 2 cores
        - 8 GB RAM
        - Disco adjunto de 1 TB para almacenamiento de documentos.

- **Base de Datos:** La solución podría utilizar una base de datos relacional desplegada en máquina virtual, utilizando motores como:

- **SQL Server**
- **PostgreSQL**

La base de datos permitiría almacenar los registros procesados y gestionar su consulta a través del visor.

Nota: *La implementación de este visor corresponde a una propuesta de mejora derivada del proyecto de calidad de datos y no hace parte del alcance desarrollado durante las fases ejecutadas.*

9. Anexos incluir archivos de correcciones

A continuación, se detallan los anexos del informe:

N° de Anexo	Nombre del archivo	Fuente / Ubicación	Estructura	Descripción del contenido
Anexo 1	**Regulaciones y certificaciones de seguridad en Microsoft Azure**	Microsoft – *Azure Compliance Offerings*, agosto de 2025	Documento técnico oficial emitido por Microsoft. Contiene secciones clasificadas por tipo de certificación (ISO, SOC, GDPR, ENS, entre otras).	Detalla las normas internacionales de seguridad, privacidad y cumplimiento aplicables a los servicios de Azure utilizados en el proyecto (máquinas virtuales, almacenamiento, base de datos, y funciones de Azure). Garantiza la trazabilidad normativa y respaldo en materia de protección de datos institucionales.
Anexo 2	**Medición completitud casos**	*Resumen_campos_registros*, *resumen_instituciones*, *resumen_programas_acreditados*	Archivos en formato Excel con tres hojas principales: **Registros Calificados**, **Instituciones Acreditadas**, y **Programas Acreditados**.	Contiene los resultados de las pruebas de completitud ponderada realizadas sobre muestras de documentos, incluyendo métricas de precisión, campos extraídos y validación manual. Permite verificar la consistencia técnica y el cumplimiento del umbral de calidad del 90 %.
Anexo 3	**Extracción Muestra documentos**	Base de datos extracción Fase I	Base en formato CSV con estructura tabular: ID Documento, Tipo de Acto, Campos extraídos, Estado del proceso, Fecha de procesamiento y Resultado OCR.	Muestra representativa de documentos procesados durante la Fase I, utilizada para los ejercicios de validación y comparación. Permite auditar la calidad de extracción y validar los resultados reportados en el informe.

Anexo 4	**Base Completa extracción Fase I**	Base de datos extracción Fase I	Base consolidada en formato CSV con el total de casos registrados dentro de la BD.	Contiene el universo total de datos procesados durante la Fase I, incluyendo los campos estructurados extraídos por los BOTS y ZAT Capture. Representa la fuente oficial de datos del proyecto para los cruces de consistencia de la Fase II.
Anexo 5	**Logs de ejecución**	Registros de logs de ejecución de RPAs	Registro de eventos de los logs operativos de los RPAs encargados de la extracción de la información.	Contiene el registro de las actividades que ejecutan las automatizaciones, el registro de eventos (Exitosos, fallidos) que presenta la automatización.
Anexo 6	**Informe reporte SGDEA**	SGDEA PDF corrupto y SGDEA sin registro	Resolución, código y fecha.	Contiene registros que no aparecen en el sistema, pese a estar esperados dentro del universo de información.
Anexo 7	**Resoluciones con tipo tramite no aplica**	Inconsistencias en el campo tipo tramite	Código proceso, Fuente y Tipo tramite.	Contiene los registros de programas con registro calificado cuyo tipo de trámite no corresponde con las resoluciones asociadas, principalmente por cargas provenientes de SGDEA.
Anexo 8	**Registros segundo ciclo no encontrados ni en sgdea ni en sharepoint**	Registros con estado de PDF no existe	Número resolución, Fecha resolución, Cod programa, Cod proceso, Fuente y Estado.	Contiene registros de la base de adición que no se encuentran ni en el sistema SGDEA ni en los documentos del SharePoint de actos administrativos.
Anexo 9	**Registros RC tipo tramite no deseado (doble procesamiento).**	Registros con tipo de trámite incorrecto	Número resolución, fecha resolución, Cod programa, Cod Proceso.	Contiene registros con tipo de trámite no aplicable.

Anexo 10	**Extracción completa**	Registros procesados en la etapa completa de extracción	Tabla con resumen de datos, detalle para Registro Calificado, Programas Acreditados, Instituciones Acreditadas, Actos Administrativos Sin Tipo e Inspección y Vigilancia.	Contiene el detalle de la extracción completa para Registro Calificado, Programas Acreditados, Instituciones Acreditadas, Actos Administrativos Sin Tipo e Inspección y Vigilancia.
Anexo 11	**MapeoCamposBID_HistoricoProgramas**	Reglas que detalla la extracción detalla las reglas aplicadas a RC Reglas, ACC_Prog, AAC_Ins	Archivos en formato Excel con tres hojas principales: Registros Calificados, Instituciones Acreditadas, y Programas Acreditados.	Contiene las reglas de negocio, validaciones y criterios de extracción aplicables a los procesos de Registro Calificado (RC), Acreditación de Programas (ACC_Prog) y Acreditación de Instituciones (AAC_Ins). Contiene comentarios relacionados con ajustes de reglas y hallazgos identificados durante la extracción.
Anexo 12	**REGLAS TRANFORMACIÓN INSPECCION_VIGILANCIA**	Reglas que detalla la extracción detalla las reglas aplicadas a INFORMES, PLANES DE VISITA, EVALUACION DE VIABILIDAD	Archivos en formato Excel con tres hojas principales: **Registros Calificados**, **Instituciones Acreditadas**, y **Programas Acreditados**.	Contiene las reglas de negocio y validaciones aplicadas al proceso de Inspección y Vigilancia.
Anexo 13	**Scripts**	Scripts de Migración Fase III	Repositorio de entrega técnica del proyecto.	Conjunto de scripts SQL organizados en dos grupos: scripts DDL (ajustes estructurales) y scripts DML (INSERT), con ejecución secuencial definida por nomenclatura.
Anexo 14	**Casos de Pruebas**	Casos de Prueba – Sistema ZATDATASYNC	Evidencias de pruebas individuales.	Archivo en formato Excel estructurado según el estándar de pruebas MEN, con identificación del caso, precondiciones, pasos, resultados esperados y resultados obtenidos.
Anexo 15	**Manual de Uso**	Manual de Uso – Protocolo de Migración y Sistema ZATDATASYNC	Documentación operativa del proyecto.	Documento en formato Word estructurado en dos secciones principales: (i) Manual de uso del Protocolo de Migración y (ii) Manual de uso del Sistema ZATDATASYNC.

Anexo 16	**Control Calidad Data-SGDEA**	Informe de extracción de Metadata de SGDEA	Documentación operativa del proyecto.	Documento en formato Excel estructurado con Metadata de documento y Metadata de resolución correspondiente a lo encontrado en SGDEA.
Anexo 17	**Data_20260128 revisado I&V 1ra observaciones 12_02_26**	Archivo #1 con modificaciones de información de los casos procesados para Inspección y vigilancia	Archivos en formato Excel con hojas separadas por cada ítem correspondiente a Inspección y Vigilancia, incluyendo las columnas asociadas a la información de los casos procesados.	Contiene las observaciones y ajustes solicitados por el Ministerio, entre los que se incluyen: cambios en el nombre de columnas, eliminación de casos que no corresponden al acto administrativo respectivo, modificación del nombre de universidades o instituciones, actualización de información en las columnas Investigador y Código SNIES, así como la adición de nuevas columnas relacionadas con Proceso y Fecha de visita.
Anexo 18	**1ra_correccion_ 13_02_2026**	Archivo #1 de respuesta a observaciones sobre casos procesados de Inspección y Vigilancia	Archivo en formato Excel que mantiene la estructura de hojas separadas por cada ítem correspondiente a Inspección y Vigilancia, incluyendo las columnas asociadas a la información de los casos procesados.	Contiene la respuesta consolidada a las observaciones realizadas por el Ministerio, incorporando los ajustes solicitados.
Anexo 19	**Data 2da observaciones 19_02_2026**	Archivo #2 con modificaciones de información de los casos procesados para Inspección y vigilancia	Archivos en formato Excel con hojas separadas por cada ítem correspondiente a Inspección y Vigilancia, incluyendo las columnas asociadas a la información de los casos procesados.	Contiene las observaciones y ajustes solicitados por el Ministerio, entre los que se incluyen: cambios en el nombre de columnas, eliminación de casos que no corresponden al acto administrativo respectivo, modificación del nombre de universidades o instituciones, actualización de información en las columnas Investigador y Código SNIES, así como la adición de nuevas columnas relacionadas con Proceso y Fecha de visita.
Anexo 20	**2da correción_24_02 _2026**	Archivo #2 de respuesta a observaciones sobre casos procesados de Inspección y Vigilancia	Archivo en formato Excel que mantiene la estructura de hojas separadas por cada ítem correspondiente a Inspección y Vigilancia, incluyendo las columnas asociadas a la información de los casos procesados.	Contiene la respuesta consolidada a las observaciones realizadas por el Ministerio, incorporando los ajustes solicitados.

Anexo 21	**Data_inspeccion_vigilancia 3ra observaciones 24_02_2026**	Archivo #3 con modificaciones de información de los casos procesados para Inspección y vigilancia	Archivos en formato Excel con hojas separadas por cada ítem correspondiente a Inspección y Vigilancia, incluyendo las columnas asociadas a la información de los casos procesados.	Contiene las observaciones y ajustes solicitados por el Ministerio, entre los que se incluyen: cambios en el nombre de columnas, eliminación de casos que no corresponden al acto administrativo respectivo, modificación del nombre de universidades o instituciones, actualización de información en las columnas Investigador y Código SNIES, así como la adición de nuevas columnas relacionadas con Proceso y Fecha de visita.
Anexo 22	**3ra_correcion_24_02_2026**	Archivo #3 de respuesta a observaciones sobre casos procesados de Inspección y Vigilancia	Archivo en formato Excel que mantiene la estructura de hojas separadas por cada ítem correspondiente a Inspección y Vigilancia, incluyendo las columnas asociadas a la información de los casos procesados.	Contiene la respuesta consolidada a las observaciones realizadas por el Ministerio, incorporando los ajustes solicitados.
Anexo 23	**Data_inspección_ 25_02_2026**	Archivo #4 con modificaciones de información de los casos procesados para Inspección y vigilancia	Archivos en formato Excel con hojas separadas por cada ítem correspondiente a Inspección y Vigilancia, incluyendo las columnas asociadas a la información de los casos procesados.	Contiene las observaciones y ajustes solicitados por el Ministerio, entre los que se incluyen: cambios en el nombre de columnas, eliminación de casos que no corresponden al acto administrativo respectivo, modificación del nombre de universidades o instituciones, actualización de información en las columnas Investigador y Código SNIES, así como la adición de nuevas columnas relacionadas con Proceso y Fecha de visita.
Anexo 24	**Data_inspección_ 25_02_2026_4ta_corrección (1)**	Archivo #4 de respuesta a observaciones sobre casos procesados de Inspección y Vigilancia	Archivo en formato Excel que mantiene la estructura de hojas separadas por cada ítem correspondiente a Inspección y Vigilancia, incluyendo las columnas asociadas a la información de los casos procesados.	Contiene la respuesta consolidada a las observaciones realizadas por el Ministerio, incorporando los ajustes solicitados.
Anexo 25	**PRO_INCONSISTENCIAS**	Listado de registros con estado “CON_INCONSISTENCIAS” identificados en el FrontEnd para Programas Acreditados en Alta Calidad	Archivo en formato Excel que consolida los casos procesados que tuvieron inconsistencias.	Contiene los respectivos campos de validación de los registros que presentan inconsistencias detectadas durante la revisión del proceso de migración y validación de datos.

Anexo 26	**RC_INCONSISTENCIAS**	Listado de registros con estado "CON_INCONSISTENCIAS" identificados en el FrontEnd para Registro Calificado	Archivo en formato Excel que consolida los casos procesados que tuvieron inconsistencias.	Contiene los respectivos campos de validación de los registros que presentan inconsistencias detectadas durante la revisión del proceso de migración y validación de datos.