|  |
| --- |
| detalle de manos con tijeras, marcadores, trabajandoINFORME CALIDAD DE DATOS |

|  |  |  |
| --- | --- | --- |
| JURISPRUDENCIA | **06 de noviembre de 2023** |  |
| Correo electrónico: biblioteca\_areatecnica@jusformosa.gob.ar | Dirección de Biblioteca e Informática Jurisprudencial – Área Técnica |  |

CONTENIDO

[INFORME CALIDAD DE DATOS 3](#_Toc150245092)

[OBJETIVO 3](#_Toc150245093)

[ALCANCE 3](#_Toc150245094)

[ESQUEMA DE DATOS 3](#_Toc150245095)

[TABLA DE FALLOS 1](#_Toc150245096)

[TABLA DE JURISPRUDENCIA 2](#_Toc150245097)

[ANÁLISIS DE DATOS 1](#_Toc150245098)

[VALORES NULOS 1](#_Toc150245099)

[VALORES DUPLICADOS 1](#_Toc150245100)

[VALORES VACIOS 1](#_Toc150245101)

[RECOMENDACIONES TÉCNICAS 1](#_Toc150245102)

[SOBRE LA ENTIDAD DE FALLOS: 1](#_Toc150245103)

[ Cambiar tipo de dato en FechaFallo 1](#_Toc150245104)

[ Eliminar la columna Resolucion 1](#_Toc150245105)

[ Considerar la columna Expediente como campo no nulo 1](#_Toc150245106)

[ Actualizar la columna Tipo valores nulos por 0 1](#_Toc150245107)

[ Registros con los campos Partes y Fallo vacíos 1](#_Toc150245108)

[ Valores duplicados en FechaFallo y fecord 2](#_Toc150245109)

[ Adicionar fecha de última modificación 2](#_Toc150245110)

[ Auditar cambios en los fallos 2](#_Toc150245111)

# INFORME CALIDAD DE DATOS

## OBJETIVO

Presentar un análisis de calidad de datos basados en los procesos de ETL (Extracción, Transformación y Carga de Datos), EDA (Análisis Exploratorio de Datos) y testeo de los tiempos de respuesta de las consultas. Para lo cual se tomó como objeto de estudio la base de datos de jurisprudencia, almacenada en una base de datos relacional: Mysql versión 8.0.21

## ALCANCE

Este informe es presentado al Departamento de Jurisprudencia para su análisis y oportunamente corrección de los datos si así lo considerasen necesario.

## ESQUEMA DE DATOS

Antes de ofrecer un informe de la calidad de los datos analizados es necesario describir su estructura de datos para dar al lector un panorama más preciso del contexto analizado.

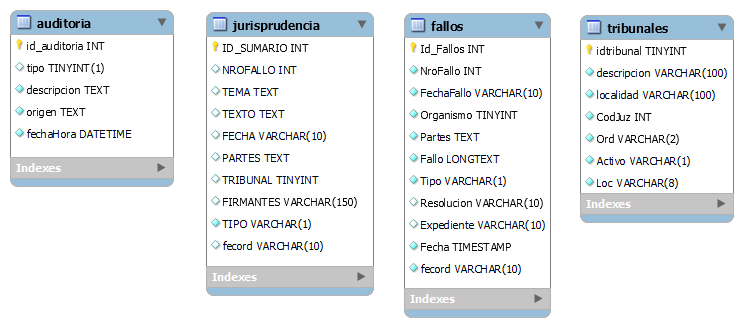


Figura 1. Diagrama Entidad Relación de Jurisprudencia

En las descripciones de las tablas se adicionan algunos comentarios acerca de la calidad de datos y más adelante en recomendaciones técnicas se ofrecen buenas practicas a implementar sobre los datos.

### TABLA DE FALLOS

Tabla 1. Entidad de Fallos

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nombre campo | Tipo de campo | Longitud del campo | ¿Es dato obligatorio? | Observaciones |
| Id\_Fallos | int | 11 | si | Clave primaria. |
| NroFallo | int | 11 | si | Número del fallo. |
| FechaFallo | varchar | 10 | si | Fecha del fallo. Tipo de dato no adecuado. |
| Organismo | tinyint | 2 | si | Clave referencial a la tabla de tribunales. |
| Partes | text |  | si | Carátula del fallo. Se observan campos vacíos. |
| Fallo | longtext |  | si | Texto plano del fallo. |
| Tipo | varchar | 1 | si, pero tiene registros vacios cuando el tipo es 0 sentencia. | Según comentario de tabla se nutre de valores 0 sentencia y 1 auto interlocutorio. |
| Resolucion | varchar | 10 | no | Campos vacíos. |
| Expediente | varchar | 10 | no | Campos vacíos. |
| Fecha | timestamp |  | si | Fecha de sistema. Fecha de alta del registro. |
| fecord | varchar | 10 | si | Esta fecha se corresponde y es la misma que la fecha del fallo. Se observa campos duplicados entre FechaFallo y fecord. |

### TABLA DE JURISPRUDENCIA

Tabla 2. Entidad de Jurisprudencia

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nombre campo | Tipo de campo | Longitud del campo | ¿Es dato obligatorio? | Observaciones |
| ID\_SUMARIO | int | 11 | si | Clave primaria. |
| NROFALLO | int | 11 | si | Número del fallo. Dato que ya se encuentra en la tabla de fallos. |
| TEMA | text |  | si | Descriptores del sumario. |
| TEXTO | text |  | si | Texto del sumario. |
| FECHA | varchar | 10 | si | Fecha del sumario. Tipo de dato no adecuado. |
| PARTES | text |  | si | Carátula del fallo. Dato que ya se encuentra en la tabla de fallos. No se tiene campos separados por actor, demandado y sobre o tema. |
| TRIBUNAL | tinyint |  | si | Clave referencial a la tabla de tribunales. No se usa claves foráneas solo referencial. Sin embargo, algunos registros están con el valor 0. |
| FIRMANTES | varchar | 150 | si | Cadena con los nombres de los firmantes. |
| TIPO | varchar | 1 | si, pero tiene registros vacios cuando el tipo es 0 sentencia. | Según comentario de tabla se nutre de valores 0 sentencia y 1 auto interlocutorio. |
| fecord | varchar | 10 | si | Esta fecha es la misma que el campo Fecha. Campos duplicados. |

### TABLA DE TRIBUNALES

Tabla 3. Tabla de Tribunales

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nombre campo | Tipo de campo | Longitud del campo | ¿Es dato obligatorio? | Observaciones |
| idtribunal | tinyint |  | si | Clave primaria. |
| descripcion | varchar | 100 | si | Nombre del tribunal. |
| localidad | varchar | 100 | si | Nombre de la localidad donde Se encuentra el tribunal. Hay registros vacios. |
| CodJuez | int |  | si | Código de Juzgado. |
| Ord | varchar | 2 | si | Orden para ordenar los tribunales. |
| Activo | varchar | 1 | si | Indica si el registro está activo o no. |
| Loc | varchar | 8 | si | Campo redundante, almacena la localidad en formato abreviado FSA. |

## ANÁLISIS DE DATOS

No presentan valores duplicados a nivel de registros, pero sí se observan valores nulos en las columnas de **Resolucion** y **Expediente** en los fallos, que implica un 11,64% para el total de fallos consultados al día de la fecha 18.108 registros. En estas mismas columnas se observan registros con valores vacíos siendo el 88,35%, por lo que podemos sostener que estas columnas en casi el 100% (18.107 registros) tienen sus campos con datos vacíos o nulos.

Siguiendo con la entidad de los fallos el campo **Tipo** determina si el registro es sentencia o AI, en este caso los registros en vacío son Sentencias. Por último, se debe revisar aquellos registros que no poseen **Partes** y **Fallo** cargados en sus registros. -

En la revisión de esos datos con Partes en valor vacío se observan los registros de la Figura 2. Consulta fallos con columna Partes en valores . En relación a la columna Fallos se visualiza la Figura 3. Consulta fallos con columna Fallo en valores vacíos. Véase las recomendaciones técnicas para estos casos.

Observándose la entidad de jurisprudencia (también denominada sumarios)

### VALORES NULOS

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Entidad | Número de registros | Número de columnas | Nulos totales (en todas las columnas) | Número de columnas con valores nulos | Cantidad máxima de nulos encontrados sobre columna | Porcentaje de cantidad máxima de nulos encontrados sobre columna | Columnas con nulos |
| fallos | 18108 | 11 | 4216 | 2 | 2108 | 11,64% | Resolucion (2108), Expediente (2108) |
| jurisprudencia | 6683 | 10 | 659 | 1 | 659 | 9,86% | fecord(659) |
| tribunales | 37 | 7 | 0 | 0 | 0 | 0,00% | - |

### VALORES DUPLICADOS

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Entidad | Total de registros duplicados | Porcentaje de duplicados | Columnas con duplicados | Proporción de duplicados |
| fallos | 0 | 0% | - | 0 |
| jurisprudencia | 0 | 0% | - | 0 |
| tribunales | 0 | 0% | - | 0 |

### VALORES VACIOS

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Entidad | Vacíos totales (en todas las columnas) | Número de columnas con valores vacíos | Cantidad máxima de vacíos encontrados sobre columna | Porcentaje de cantidad máxima de vacíos encontrados sobre columna | Columnas con valores vacíos |
| fallos | 49569 | 5 | 15999 | 88,35% | Partes(12), Fallo(14), Tipo (17545), Resolucion(15999), Expediente(15999) |
| jurisprudencia | 7704 | 4 | 6415 | 95,98% | TEMA(1), PARTES(9), TIPO(6415), fecord(1279) |
| tribunales | 4 | 1 | 4 | 10,81% | localidad(4) |

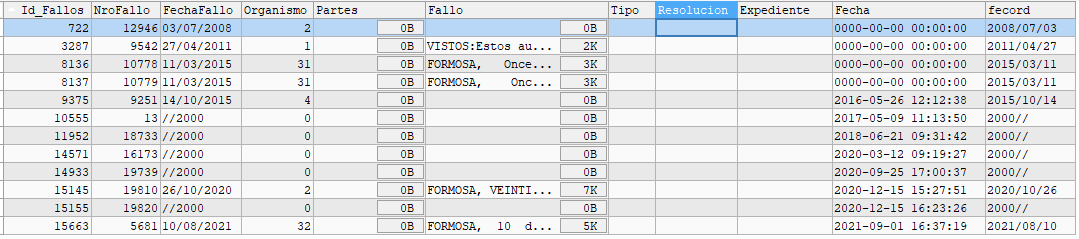


Figura 2. Consulta fallos con columna Partes en valores vacíos

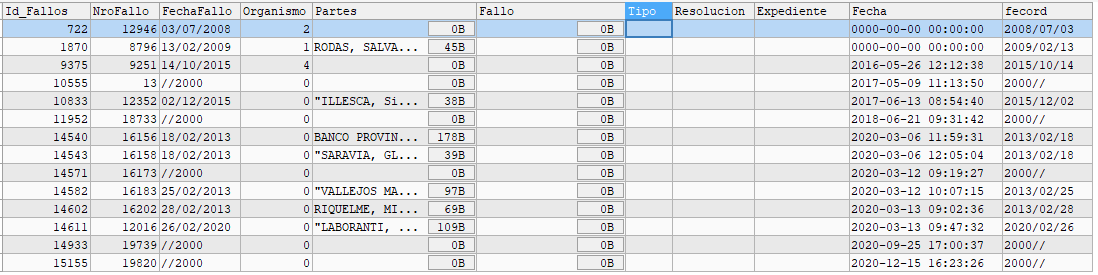


Figura 3. Consulta fallos con columna Fallo en valores vacíos

# RECOMENDACIONES TÉCNICAS

## SOBRE LA ENTIDAD DE FALLOS:

## Cambiar tipo de dato en FechaFallo

Cambiar a tipo date para FechaFallo, ya que el tipo de dato varchar(10) no es el adecuado.

## Eliminar la columna Resolucion

Dado que solo un registro cuenta con esos campos completos, además Resolucion almacena el número y año del fallo, datos que se hallan en las columnas de NroFallo y FechaFallo.

## Considerar la columna Expediente como campo no nulo

Se entiende que esta columna almacena el número y año del expediente donde tiene lugar el fallo, considerar obligatorio la carga de este campo para los futuros registros.

## Actualizar la columna Tipo valores nulos por 0

Es mejor asignar un valor a este campo antes que quede como dato vacío para marcar sus registros Tipo Sentencia.

## Registros con los campos Partes y Fallo vacíos

Dada la Figura 2. Consulta fallos con columna Partes en valores y Figura 3. Consulta fallos con columna Fallo en valores vacíos, se recomienda eliminar los registros con **FechaFallo** en **“//2000”**, también eliminar aquellos con los campos **Partes** y **Fallos** en vacío, ya que se consideran registros erróneos o que no aportan valor.

Los demás registros que tienen un valor ya sea en Partes o en Fallo se pueden mantener completando el campo vacío. El campo Partes almacena la caratula del expediente y Fallo el texto de la sentencia o AI, por lo cual es fundamental completar estos registros cuyos campos presentan valores vacíos.

## Valores duplicados en FechaFallo y fecord

Esta fecha es la misma que la fecha del fallo. Se observa campos duplicados entre FechaFallo y fecord, se puede eliminar **fecord**.

## Adicionar fecha de última modificación

Es necesario agregar un campo de última modificación del registro.

## Auditar cambios en los fallos

Auditar cambios en el registro, ya sea por triggers, procedimientos, la forma que resulte más oportuna.

|  |
| --- |
| **1) Cambiar el tipo de dato de la fecha del sumario** |
| **2) Se debe exigir la carga obligatoria del campo Tipo.** |
| **3) Resolver los registros que tienen tribunal en cero.** |
| **4) Agregar fecha de sistema de ultima modificacion del registro.** |
| **5) Corregir el dato vacio en el campo fecord.** |
| **6) suprimir el campo PARTES y NROFALLO, esta info ya se encuentra en la tabla de fallos** |