



UNIVERSIDAD DE MÁLAGA



Trabajo Almacenes de Datos

Diseño y Explotación de un almacén de UCI Sanitaria

Realizado por
De Pablo Diego y Soriano Juan

Profesor encargado:
Luque Baena Rafael Marcos

Departamento
Lenguajes y Ciencias de la Computación

MÁLAGA, noviembre de 2024



UNIVERSIDAD
DE MÁLAGA



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA INFORMÁTICA
ESTUDIANTES DE INGENIERÍA BIOINFORMÁTICA

Diseño y Explotación de un almacén de UCI Sanitaria

Almacenes de Datos

Realizado por
De Pablo Diego y Soriano Juan

Profesor encargado:
Navas Luque Baena Rafael Marcos

Departamento
Lenguajes y Ciencias de la Computación

UNIVERSIDAD DE MÁLAGA
MÁLAGA, NOVIEMBRE DE 2024

Contents

1	Introducción	3
2	Objetivos	3
3	Descripción de los datos utilizados (eICU)	3
4	Diseño conceptual	4
4.1	Tablas	5
4.1.1	Diagnosis	5
4.1.2	respiratoryCharting	6
4.1.3	respiratoryCare	6
4.1.4	apacheApsVar	7
4.1.5	pacienteIngresado	7
5	Diseño Lógico - Copo de nieve	8
5.0.1	FK	9
6	Dificultades encontradas	10
7	Conclusiones	10

1 Introducción

El diseño y desarrollo de **almacenes de datos** es crucial para el análisis clínico, mejorando la toma de decisiones médicas y la calidad de atención a los pacientes. Para este proyecto, se utilizará una visión parcial de un almacén de datos basado en la *Base de Datos de Investigación Colaborativa eICU*, la cual contiene una vasta cantidad de información sobre los ingresos en **Unidades de Cuidados Intensivos (UCI)** en diversos hospitales de los Estados Unidos.[1]

El enfoque de este trabajo está dirigido específicamente a los **pacientes con problemas respiratorios**. Estos casos son de especial relevancia en entornos de cuidados críticos, ya que las afecciones respiratorias representan una de las principales causas de ingreso en las UCI. Por lo cual un almacén de datos centrada en estas patologías sigue representando una gran fuente de información y con mucho interés para hacer un análisis de datos. [4, 5].

En este trabajo, el **hecho principal** está constituido por los ingresos en la UCI, complementado por un conjunto de dimensiones que permiten analizar diversas variables. La selección de las tablas y atributos más relevantes de la base de datos eICU será el mayor reto para construir un **almacén de datos eficiente y de interés analítico**, que no solo facilite la consulta de información clave, a su vez soporte y facilite tanto la investigación clínica como la optimización de los protocolos de tratamiento.

2 Objetivos

El objetivo de este trabajo es desarrollar un **diseño conceptual y lógico** de un almacén de datos centrado en los pacientes ingresados en la **UCI** con afecciones respiratorias, seleccionando las partes más relevantes de la base de datos para el análisis de estas patologías.

Se procederá a la restauración de dicha base de datos en *SQL Server*, para realizar una selección de las tablas más relevantes que contribuyan a la construcción del almacén de datos centrado en estos pacientes. En cada una de las tablas seleccionadas, se identificarán los atributos más significativos para el análisis clínico y el seguimiento de los pacientes con problemas respiratorios.

Una vez completada la selección de dimensiones y tablas de hechos, se diseñará el modelo conceptual utilizando la herramienta *draw.io*, y posteriormente se desarrollará el modelo lógico mediante un diagrama de base de datos en el entorno de *SQL Server*.

3 Descripción de los datos utilizados (eICU)

La base de datos *eICU-CRD*, desarrollada por el Philips eICU Research Institute (eRI), contiene datos desidentificados de pacientes ingresados en Unidades de Cuidados Intensivos (UCI). Esta base de datos, que forma parte del programa de telesalud en cuidados críticos de Philips, documenta detalles sobre diagnósticos, tratamientos, pruebas y resultados.

La *eICU-CRD* recopila una amplia cantidad de datos clínicos con fines de investigación, permitiendo estudiar los resultados de los pacientes, identificar tendencias

clínicas y evaluar protocolos de mejores prácticas utilizados en distintos centros de atención médica. Esta base de datos tiene como fin facilitar la investigación para mejorar la calidad de la atención en UCI.

La *eICU-CRD* contiene un total de 31 tablas, que agrupan 391 columnas de datos y un total de 457,325,320 filas. Esta vasta cantidad de información abarca diversos aspectos clínicos, como datos demográficos de los pacientes, información clínica obtenida durante el ingreso, pruebas y diagnósticos realizados, tratamientos y terapias administradas, así como la evolución y resultados durante la estancia en la UCI. Dada la magnitud de los datos. En este trabajo, se ha optado por simplificar el almacén, enfocándose en aquellos datos más relevantes para el análisis de pacientes ingresados en la UCI con problemas respiratorios. A continuación, se presentan las 9 tablas seleccionadas del almacén original:

- **ApacheApsVar:** Contiene diversas métricas utilizadas para evaluar la gravedad de la enfermedad de los pacientes al ingreso en la UCI como parte del sistema APACHE.
- **Admissiondx:** Contiene el diagnóstico principal por el cual el paciente fue admitido en la UCI, de acuerdo con los criterios de puntuación APACHE.
- **Allergy:** Contiene información sobre las alergias del paciente.
- **Diagnosis:** Contiene los diagnósticos documentados para cada paciente. Permite identificar las enfermedades documentadas durante la estancia en UCI, así como el momento en que se registraron.
- **Hospital:** contiene información sobre los hospitales participantes en el programa y sus pacientes vinculados.
- **Medication:** Contiene las órdenes activas de medicamentos para los pacientes.
- **Patient:** Contiene los datos demográficos del paciente, así como los detalles de su admisión y alta del hospital y la UCI.
- **RespiratoryCare:** Contiene información sobre el cuidado respiratorio.
- **RespiratoryCharting:** Contiene datos sobre la configuración respiratoria, como el tipo de gráfico y valores respiratorios.

Además de reducir el número de tablas también se encontró oportuno reducir el número de atributos de las mismas, al contener información poco esencial o que no se ve utilidad para las siguientes fases del trabajo. (Vea la figura 1 para detallar las tablas y sus atributos de manera global)

4 Diseño conceptual

El diseño conceptual define las entidades, relaciones y requisitos de información de manera abstracta, sin detalles técnicos. Se enfoca en representar las necesidades y cómo organizar los datos para soportar la toma de decisiones.

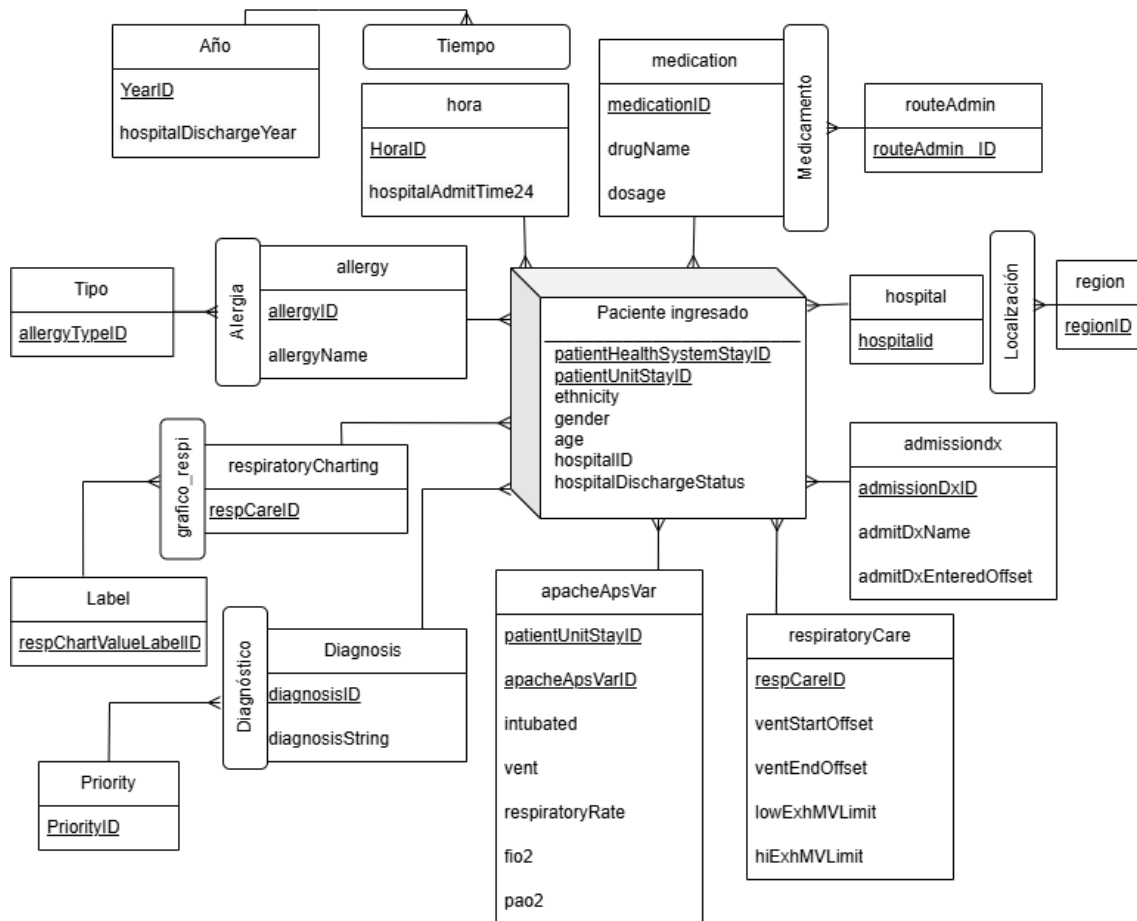


Figure 1: Diseño conceptual

4.1 Tablas

4.1.1 Diagnósis

- **Relevancia:** La tabla **Diagnósis** es clave para analizar diagnósticos de enfermedades respiratorias en pacientes de UCI, permitiendo filtrar y clasificar las condiciones por severidad.
- **Selección de atributos:**
 - **patientUnitStayID:** Relaciona el diagnóstico con un paciente específico en la UCI.
 - **diagnosisID:** Clave primaria para diferenciar cada diagnóstico.
 - **diagnosisString:** Descripción completa del diagnóstico para consultas específicas.
 - **diagnosisPriority:** Indica la prioridad del diagnóstico (Primario, Mayor, Otro).

No se han incluido atributos como **diagnosisOffset** e **ICD9Code** debido a que nuestro análisis se centra más en la naturaleza y prioridad del diagnóstico que en el tiempo específico de entrada o el código ICD-9. [2]

4.1.2 respiratoryCharting

- **Relevancia:** La tabla `RespiratoryCharting` es fundamental para el monitoreo de los valores respiratorios de los pacientes en UCI, especialmente en aquellos con enfermedades respiratorias graves.
- **Selección de atributos:**
 - `patientUnitStayID`: Asocia los datos respiratorios con un paciente específico.
 - `respCareID`: Identificador único del registro respiratorio.
 - `respChartValueLabel`: Describe el tipo de valor respiratorio (ej. HR, I:E Ratio), útil para categorizar los datos.

Decidimos que datos como `respChartOffset` y `respChartValue` añadían detalles precisos, por lo que priorizamos la relevancia clínica general sobre estos detalles puntuales, ya que el objetivo es estudiar tendencias y no el seguimiento minuto a minuto. [2]

4.1.3 respiratoryCare

- **Relevancia:** La tabla `RespiratoryCare` es clave para el análisis de los cuidados respiratorios proporcionados a los pacientes en la UCI, ya que permite evaluar las intervenciones de ventilación mecánica y otras terapias respiratorias.
- **Selección de atributos:**
 - `patientUnitStayID`: Asocia el cuidado respiratorio a un paciente específico.
 - `respCareID`: Identificador único para cada intervención de cuidado respiratorio.
 - `ventStartOffset`: Refleja el inicio de la ventilación, crucial para analizar la relación entre el inicio de la intervención y la evolución del paciente.
 - `ventEndOffset`: Indica el final de la ventilación, permitiendo estudiar la duración de las intervenciones respiratorias.
 - `lowExhMVLimit`: Establece el límite inferior del volumen minuto expiratorio, importante para evaluar la efectividad de la ventilación.
 - `hiExhMVLimit`: Define el límite superior del volumen minuto expiratorio, igualmente relevante para evaluar la capacidad de los sistemas respiratorios.

Hemos excluido atributos como `airwayType` y `cuffPressure` porque nos centramos en los parámetros que permiten observar el impacto de la ventilación mecánica en la evolución de los pacientes y no tanto en los detalles técnicos de cada intervención.[2]

4.1.4 apacheApsVar

- **Relevancia:** La tabla `apacheApsVar` es esencial para el cálculo del **Acute Physiology Score (APS) III**, un sistema ampliamente utilizado para evaluar la gravedad de la enfermedad de los pacientes al ingreso en la UCI. Esta puntuación es parte del sistema **APACHE** para predecir los resultados de los pacientes críticos.
- **Selección de atributos:**
 - `patientUnitStayID`: Relaciona a cada entrada de paciente en la UCI con su respectivo registro en la tabla de pacientes.
 - `apacheApsVarID`: Clave primaria que identifica de manera única cada conjunto de variables de **APACHE APS**.
 - `intubated`: Indica si el paciente fue intubado al momento de obtener el peor valor de gasometría (ABG), crucial para evaluar la necesidad de intervención respiratoria.
 - `vent`: Indica si el paciente fue ventilado al momento de registrar la peor frecuencia respiratoria, reflejando la gravedad de la insuficiencia respiratoria.
 - `respiratoryRate`: Refleja la frecuencia respiratoria más baja durante el período de **APACHE**, utilizada para medir la función respiratoria del paciente.
 - `fio2`: Mide la fracción inspirada de oxígeno, importante para evaluar la insuficiencia respiratoria del paciente y su respuesta al tratamiento con oxígeno.
 - `pao2`: Mide la presión parcial de oxígeno en sangre, un indicador clave de la gravedad de la hipoxia en los pacientes.

Atributos como la puntuación de Glasgow (GCS), los valores de creatinina, glucosa, o hematocrito, aunque son importantes para el cálculo del puntaje **APACHE**, no están tan directamente relacionados el análisis de los problemas respiratorios. [2]

4.1.5 pacienteIngresado

- **Relevancia:** La tabla `pacienteIngresado` es el hecho que almacena información demográfica de los pacientes y detalles relacionados con sus ingresos y egresos del hospital y la UCI. Es fundamental para el análisis de la estancia en la UCI y la hospitalización, proporcionando una visión global de la evolución clínica y de los tiempos de atención.
- **Selección de atributos:**
 - `patientUnitStayID`: Clave primaria que identifica de manera única la estancia del paciente en la UCI. Relaciona con el registro de `patient` mediante este identificador.
 - `patientHealthSystemStayID`: Identificador del ingreso hospitalario, vincula las estancias hospitalarias de un paciente durante un mismo periodo.

- **ethnicity**: Etnia del paciente, una variable importante para el análisis demográfico y la disparidad en la atención.
- **gender**: Género del paciente, relevante para estudios epidemiológicos y de salud pública.
- **age**: Edad del paciente, crucial para la estratificación del riesgo y el análisis de comorbilidades.
- **hospitalID**: Identificador único del hospital, esencial para la segmentación por centro de atención.
- **hospitalDischargeStatus**: Estado del paciente al momento del alta hospitalaria, indicando si está vivo, fallecido, o en otro estado.
- **hospitalAdmitTime24**: Hora exacta del ingreso al hospital, importante para el análisis temporal de la atención.
- **hospitalDischargeYear**: Año de alta hospitalaria, relevante para la evaluación de tendencias a lo largo del tiempo.
- **UniquePID**: Identificador único del paciente, utilizado para distinguir a los pacientes en el sistema.

Algunos atributos no se han incluido debido a la duplicación o irrelevancia en el contexto del hecho pacienteIngresado. Por ejemplo, hospitalAdmitTime24 y hospitalDischargeYear ya cubren la información temporal necesaria, por lo que atributos como unitAdmitTime24 y unitDischargeTime24 no son necesarios en este nivel, ya que hacen referencia a eventos más específicos de la UCI, no del ingreso hospitalario. Igualmente, wardID y unitType se han excluido porque son detalles más relevantes para el contexto de la unidad de UCI, no para el hecho principal del paciente ingresado. [2]

5 Diseño Lógico - Copo de nieve

Para la elaboración del modelo lógico, nos fundamentamos en el modelo conceptual previamente diseñado:

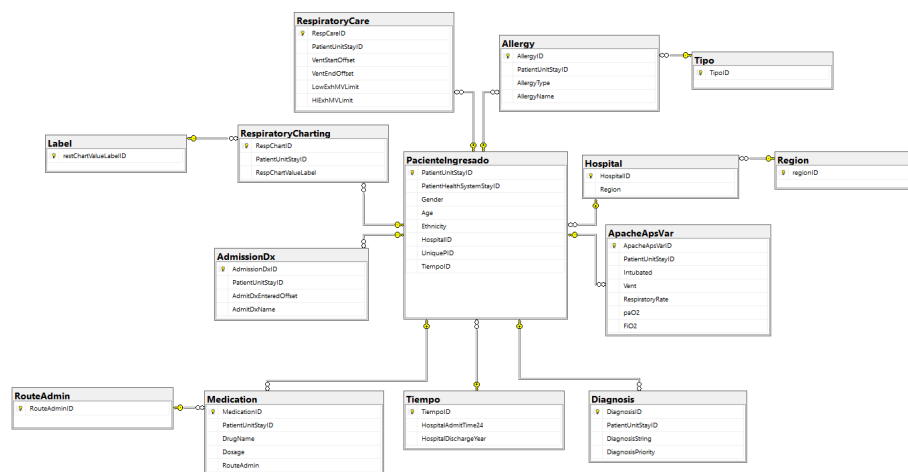


Figure 2: Diseño lógico

En el modelo lógico del almacén de datos, se ha implementado la normalización para organizar eficientemente las tablas y reducir la redundancia. Se ha propuesto una normalización en tablas, como PacienteIngresado, Medication, RespiratoryCharting, entre otras, menos la dimensión tiempo que se mantiene denormalizada.

5.1 FK's

- **PatientUnitStayID** en:

- RespiratoryCare → PacienteIngresado.PatientUnitStayID
- RespiratoryCharting → PacienteIngresado.PatientUnitStayID
- AdmissionDx → PacienteIngresado.PatientUnitStayID
- ApacheApsVar → PacienteIngresado.PatientUnitStayID
- Medication → PacienteIngresado.PatientUnitStayID
- Diagnosis → PacienteIngresado.PatientUnitStayID

Esto vincula las tablas que contienen información específica sobre los cuidados y condiciones del paciente con su estancia en UCI.

- **HospitalID** en:

- PacienteIngresado → Hospital.HospitalID

Relaciona cada estancia del paciente con el hospital correspondiente.

- **Region** en:

- Hospital → Region.regionID

Define la región en la que se encuentra el hospital.

- **RouteAdmin** en:

- Medication → RouteAdmin.RouteAdminID

Especifica la vía de administración de cada medicamento.

- **RespChartValueLabel** en:

- RespiratoryCharting → Label.restChartValueLabelID

Define el tipo de valor en el gráfico de respiración.

- **TiempoID** en:

- PacienteIngresado → Tiempo.TiempoID

Permite asociar información de tiempo como la admisión y el alta.

- **TipoID** en:

- Allergy → Tipo.TipoID

Clasifica el tipo de alergia mediante una FK hacia Tipo.

En lugar de incluir un ID genérico y un atributo que describa detalles (como en Region), optamos por almacenar el valor descriptivo directamente en regionID, eliminando la necesidad de otro campo y simplificando el modelo.

6 Dificultades encontradas

Principalmente la complejidad de la base de datos eICU, que contiene numerosas tablas y atributos, y ha requerido un análisis exhaustivo para identificar las tablas y campos más relevantes para un modelo de datos centrado en pacientes con enfermedades respiratorias. También hemos tenido que lidiar con problemas de permisos al visualizar el modelo relacional en SQL Server, lo que requirió modificar la autorización del propietario de la base de datos para acceder a los diagramas de relación.

7 Conclusiones

Todo el proyecto está accesible en github [3]

References

- [1] Chamsi Bah, Sultan Alharthi, and Ashraf El Metwally. “Clinical data warehousing: A review of current systems and future directions”. In: *Journal of Healthcare Engineering* 2017 (2017), pp. 1–11. DOI: 10.1155/2017/8326740.
- [2] eICU Collaborative Research Database. *eICU Collaborative Research Database*. <https://eicu-crd.mit.edu/about/eicu/>. Accessed: 2024-11-14. 2024. URL: <https://eicu-crd.mit.edu/about/eicu/>.
- [3] Diegodepab. *Almacén UCI Sanitaria*. https://github.com/Diegodepab/almacen_UCI_Sanitaria. Accessed: 2024-11-14. 2024. URL: https://github.com/Diegodepab/almacen_UCI_Sanitaria.
- [4] Christopher Pekar, Sven Gordan, and Ewan Goligher. “Epidemiology of respiratory failure in the intensive care unit: A review”. In: *Critical Care* 25.1 (2021), pp. 1–9. DOI: 10.1186/s13054-021-03772-y.
- [5] Jean-Louis Vincent, Yasser Sakr, and V Marco Ranieri. “The epidemiology of respiratory failure in the ICU”. In: *Chest* 129.1 (2006), pp. 90–99. DOI: 10.1378/chest.129.1.90.



UNIVERSIDAD
DE MÁLAGA

| **uma.es**

E.T.S. DE INGENIERÍA INFORMÁTICA

E.T.S de Ingeniería Informática
Bulevar Louis Pasteur, 35
Campus de Teatinos
29071 Málaga