



PRÁCTICA #3

Bases de Datos Distribuidas

Descripción breve

En esta práctica se fragmentará la base de datos covidHistorico por regiones y se distribuirán en 4 nodos, incluyendo uno con MySQL. Se modificarán consultas distribuidas para ejecutarse desde cualquier nodo SQL Server

- Urrutia González Brenda
 - Olea García Alan
- Juárez Anguiano Mario Alexis

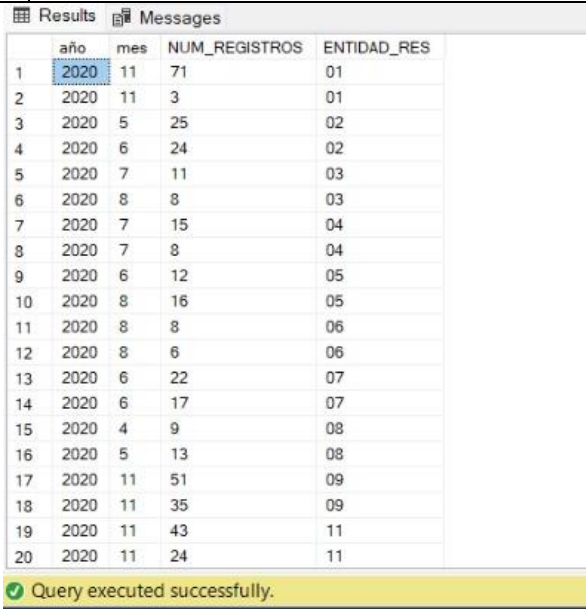
Equipo 03
20/06/2025

Consulta 3.	Listar el porcentaje de casos confirmados en cada una de las siguientes morbilidades a nivel nacional: diabetes, obesidad e hipertensión		
Requisitos:	<ul style="list-style-type: none">• Calcular el porcentaje nacional de casos confirmados para las morbilidades de diabetes, obesidad e hipertensión.• Considerar solo los casos clasificados como "confirmados" en el campo CLASIFICACION_FINAL (valores 1, 2, 3).• Utilizar servidores vinculados para obtener los datos de las diferentes regiones distribuidas entre los nodos (AZURE-ALAN, AZURE-MARIO, AZURE-BRENDA, MYSQL_ALAN)		
Significado de valores de los catálogos:	<ul style="list-style-type: none">• HIPERTENSION: Si tiene valor '1', el paciente presenta la enfermedad, de lo contrario es '0'.• OBESIDAD: Si tiene valor '1', el paciente presenta la enfermedad, de lo contrario es '0'.• DIABETES: Si tiene valor '1', el paciente presenta la enfermedad, de lo contrario es '0'. <p>CLASIFICACION_FINAL:</p> <ul style="list-style-type: none">• 1: Confirmado por asociación clínica-epidemiológica.• 2: Confirmado por dictaminación médica.• 3: Confirmado por laboratorio.		
Responsable:	Alan Olea García.		
Comentarios:	<ul style="list-style-type: none">✓ Se utiliza la función CASE para evaluar cada morbilidad y contar los casos confirmados de cada una de ellas.✓ Las consultas se ejecutan desde servidores distribuidos utilizando servidores vinculados para acceder a los datos de las diferentes regiones.✓ Se calcula el porcentaje nacional para cada morbilidad en función de los casos confirmados por región.		
	Porcentaje_DIABETES	Porcentaje_HIPERTENSION	Porcentaje_OBESIDAD
1	9.62	12.79	10.57

Consulta 4.	Listar los municipios que no tengan casos confirmados en todas las morbilidades: hipertensión, obesidad, diabetes, tabaquismo.																														
Requisitos:	<ul style="list-style-type: none">Filtrar los municipios que no presentan casos confirmados de hipertensión, obesidad, diabetes y tabaquismo.Consultar los datos distribuidos desde los nodos correspondientes.																														
Significado de los valores de los catálogos:	<ul style="list-style-type: none">HIPERTENSION: Si tiene valor '1', el paciente presenta la enfermedad, de lo contrario es '0'.OBESIDAD: Si tiene valor '1', el paciente presenta la enfermedad, de lo contrario es '0'.DIABETES: Si tiene valor '1', el paciente presenta la enfermedad, de lo contrario es '0'.TABAQUISMO: Si tiene valor '1', el paciente presenta la enfermedad, de lo contrario es '0'.MUNICIPIO_RES: Municipio de residencia del paciente.																														
Responsable:	Alan Olea García.																														
Comentarios:	<ul style="list-style-type: none">✓ La consulta evalúa las condiciones de cada morbilidad para excluir aquellos municipios sin casos confirmados.✓ Se realiza la consulta de manera distribuida, accediendo a datos de múltiples nodos.																														
<table><tr><td></td><td>entidad_res</td><td>MUNICIPIO_RES</td></tr><tr><td>1</td><td>05</td><td>018</td></tr><tr><td>2</td><td>05</td><td>030</td></tr><tr><td>3</td><td>14</td><td>072</td></tr><tr><td>4</td><td>14</td><td>093</td></tr><tr><td>5</td><td>14</td><td>098</td></tr><tr><td>6</td><td>19</td><td>006</td></tr><tr><td>7</td><td>25</td><td>001</td></tr><tr><td>8</td><td>25</td><td>006</td></tr><tr><td>9</td><td>30</td><td>174</td></tr></table>			entidad_res	MUNICIPIO_RES	1	05	018	2	05	030	3	14	072	4	14	093	5	14	098	6	19	006	7	25	001	8	25	006	9	30	174
	entidad_res	MUNICIPIO_RES																													
1	05	018																													
2	05	030																													
3	14	072																													
4	14	093																													
5	14	098																													
6	19	006																													
7	25	001																													
8	25	006																													
9	30	174																													

Consulta 5.	Listar los estados con más casos recuperados con neumonía.
Requisitos:	<ul style="list-style-type: none"> Filtrar los casos recuperados con neumonía para determinar cuáles son los estados con mayor cantidad de casos. Consultar los datos distribuidos por nodos vinculados.
Significado de valores de los catálogos:	<ul style="list-style-type: none"> NEUMONIA: Si tiene valor '1' es que tiene la enfermedad, de lo contrario '0' ENTIDAD_RES: Nos da el municipio de residencia del paciente
Responsable:	Brenda Urrutia
Comentarios:	<ul style="list-style-type: none"> ✓ Se utiliza un JOIN con la tabla cat_entidades para obtener el nombre del estado correspondiente a cada entidad. ✓ Los datos se obtienen de los nodos distribuidos para realizar un análisis completo a nivel nacional.

<div>89 %</div> <div>Results Messages</div> <table> <tr> <th>entidad</th><th>Casos_Recuperados_Totales</th></tr> <tr><td>1 CIUDAD DE MEXICO</td><td>46545</td></tr> <tr><td>2 MÉXICO</td><td>45599</td></tr> <tr><td>3 PUEBLA</td><td>16971</td></tr> <tr><td>4 VERACRUZ DE IGNACIO DE LA LLAVE</td><td>11936</td></tr> <tr><td>5 GUANAJUATO</td><td>11356</td></tr> <tr><td>6 BAJA CALIFORNIA</td><td>9519</td></tr> <tr><td>7 JALISCO</td><td>9480</td></tr> <tr><td>8 HIDALGO</td><td>9312</td></tr> <tr><td>9 NUEVO LEÓN</td><td>8926</td></tr> <tr><td>10 MICHOACÁN DE OCAMPO</td><td>7562</td></tr> <tr><td>11 CHIHUAHUA</td><td>7261</td></tr> <tr><td>12 SONORA</td><td>7096</td></tr> <tr><td>13 SINALOA</td><td>6952</td></tr> <tr><td>14 QUERÉTARO</td><td>6537</td></tr> <tr><td>15 GUERRERO</td><td>6430</td></tr> <tr><td>16 TABASCO</td><td>5673</td></tr> <tr><td>17 SAN LUIS POTOSÍ</td><td>5502</td></tr> <tr><td>18 MORELOS</td><td>4944</td></tr> <tr><td>19 COAHUILA DE ZARAGOZA</td><td>4880</td></tr> <tr><td>20 TAMAULIPAS</td><td>4267</td></tr> <tr><td>21 OAXACA</td><td>3959</td></tr> <tr><td>22 QUINTANA ROO</td><td>3622</td></tr> </table> <div>Query executed successfully.</div>		entidad	Casos_Recuperados_Totales	1 CIUDAD DE MEXICO	46545	2 MÉXICO	45599	3 PUEBLA	16971	4 VERACRUZ DE IGNACIO DE LA LLAVE	11936	5 GUANAJUATO	11356	6 BAJA CALIFORNIA	9519	7 JALISCO	9480	8 HIDALGO	9312	9 NUEVO LEÓN	8926	10 MICHOACÁN DE OCAMPO	7562	11 CHIHUAHUA	7261	12 SONORA	7096	13 SINALOA	6952	14 QUERÉTARO	6537	15 GUERRERO	6430	16 TABASCO	5673	17 SAN LUIS POTOSÍ	5502	18 MORELOS	4944	19 COAHUILA DE ZARAGOZA	4880	20 TAMAULIPAS	4267	21 OAXACA	3959	22 QUINTANA ROO	3622
entidad	Casos_Recuperados_Totales																																														
1 CIUDAD DE MEXICO	46545																																														
2 MÉXICO	45599																																														
3 PUEBLA	16971																																														
4 VERACRUZ DE IGNACIO DE LA LLAVE	11936																																														
5 GUANAJUATO	11356																																														
6 BAJA CALIFORNIA	9519																																														
7 JALISCO	9480																																														
8 HIDALGO	9312																																														
9 NUEVO LEÓN	8926																																														
10 MICHOACÁN DE OCAMPO	7562																																														
11 CHIHUAHUA	7261																																														
12 SONORA	7096																																														
13 SINALOA	6952																																														
14 QUERÉTARO	6537																																														
15 GUERRERO	6430																																														
16 TABASCO	5673																																														
17 SAN LUIS POTOSÍ	5502																																														
18 MORELOS	4944																																														
19 COAHUILA DE ZARAGOZA	4880																																														
20 TAMAULIPAS	4267																																														
21 OAXACA	3959																																														
22 QUINTANA ROO	3622																																														

Consulta 7.	Para el año 2020 y 2021 cuál fue el mes con más casos registrados, confirmados, sospechosos, por estado registrado en la base de datos.
Requisitos:	<ul style="list-style-type: none"> • Determinar el mes y año con la mayor cantidad de registros de casos por estado. • Considerar únicamente los años 2020 y 2021. • Diferenciar entre casos confirmados (CLASIFICACION_FINAL: 1, 2, 3) y sospechosos (CLASIFICACION_FINAL: 6). • Utilizar la función ROW_NUMBER() para asignar un ranking a cada mes dentro de cada estado, ordenando por el número de registros en orden descendente. • Filtrar para obtener solo los dos meses con más casos por estado (Rank <= 2).
Significado de valores de los catálogos:	<ul style="list-style-type: none"> • ENTIDAD_RES: Código de la entidad federativa del paciente. • FECHA_INGRESO: Fecha de registro del paciente. • CLASIFICACION_FINAL: <ul style="list-style-type: none"> - 1: Confirmado por asociación clínica-epidemiológica. - 2: Confirmado por dictaminación médica. - 3: Confirmado por laboratorio. - 6: Caso sospechoso sin confirmación definitiva.
Responsable:	Mario Alexis Juarez Anguiano
Comentarios:	<ul style="list-style-type: none"> ✓ Se utiliza un CTE (Common Table Expression) para organizar los datos y calcular el total de registros por mes y estado. ✓ La función ROW_NUMBER() se emplea para asignar un número de fila a cada mes, ordenado por el número de registros en orden descendente. Esto permite identificar los dos meses con más casos por estado. ✓ La consulta se realiza en múltiples nodos utilizando OPENQUERY para acceder a las bases de datos distribuidas.
	

Conclusiones

Primeramente, la forma en la que se fragmentó la tabla fue en 4 nodos, de los cuales 3 se almacenaron en SQL y otro en MySQL.

La forma de distribución de los nodos fue de la siguiente manera (sacada de Wikipedia):



Las primeras dos regiones (noroeste y noreste) fueron almacenadas en el nodo del compañero Alan, las dos siguientes (occidente y oriente) en el nodo SQL de la compañera Brenda, las regiones centro norte y centro sur fueron almacenadas en el nodo del compañero Mario y finalmente las últimas regiones se almacenaron en el nodo MySQL.

En esta práctica se abordaron aspectos clave relacionados con la fragmentación de bases de datos distribuidas y la ejecución de consultas SQL en servidores vinculados. Aplicamos los conocimientos adquiridos sobre bases de datos distribuidas, utilizando técnicas avanzadas como la consulta de datos de múltiples nodos y la optimización de consultas para obtener resultados más precisos y eficientes, así como se trabajó en la forma de optimizar las consultas para reducir el tiempo de respuesta:

1. **Optimización de las consultas UNION ALL:** Al consolidar múltiples operaciones UNION ALL en consultas más eficientes y eliminar declaraciones redundantes de DISTINCT, se redujo significativamente la complejidad y se mejoró el rendimiento de las consultas. Este enfoque simplificó el proceso de recuperación de datos al reducir el código repetitivo y asegurar un filtrado eficiente.
2. **Resolución de conflictos de collation:** Un problema común encontrado fue el conflicto de collation, que se resolvió mediante el uso de la cláusula COLLATE en las columnas involucradas en las operaciones de comparación. Esto aseguró que todos los datos

comparados entre diferentes fuentes, como servidores vinculados y bases de datos, fueran consistentes en su collation, evitando errores y facilitando el proceso de comparación.

3. **Manejo de OPENQUERY y servidores vinculados:** Durante el proceso, se utilizó OPENQUERY para obtener datos de servidores externos, pero se encontraron algunos problemas relacionados con el proveedor OLE DB y la configuración de los servidores. Las soluciones involucraron simplificar las consultas y asegurarse de que los servidores vinculados estuvieran correctamente configurados, así como solucionar problemas de conexión.

Para finalmente analizar la optimización de nuestras consultas, comparamos el uso de OPENQUERY con los procedimientos almacenados de 4 pasos, y nos dimos cuenta que cada enfoque tiene ventajas y desventajas dependiendo del escenario.

El uso de OPENQUERY resultó ser más eficiente para consultas simples, ya que permite ejecutar las consultas directamente en el servidor vinculado sin necesidad de mover grandes volúmenes de datos entre servidores. Sin embargo, su flexibilidad es limitada y no permite realizar operaciones dinámicas o manejar lógica compleja antes de devolver los resultados.

Por otro lado, los procedimientos almacenados de 4 pasos ofrecen una mayor flexibilidad y control sobre la ejecución de las consultas, permitiendo manipular datos, realizar transformaciones y manejar parámetros dinámicos de SQL Server. Sin embargo, este enfoque conlleva una mayor complejidad y, si no está bien optimizado, puede impactar negativamente el rendimiento, especialmente en consultas que involucran grandes volúmenes de datos o múltiples pasos de procesamiento.