

Prueba Técnica – Ingeniero de Datos

Sara Tobón Restrepo

Notas adicionales

Punto 1

Para el desarrollo de la prueba, en un inicio se decidió usar la capa gratuita de Azure como solución en la Nube, se decide crear un grupo de recursos que contiene un servicio de Synapse, se avanza en el desarrollo, pero eventualmente la cuenta es bloqueada, se intenta nuevamente crear otra cuenta con capa gratuita y no se permite, aunque se use un correo nuevo y una tarjeta de crédito nunca usada para este beneficio. Se adjuntan pruebas.

Microsoft AzureActualización

Buscar recursos, servicios y documentos (G+)

sa.to.re@hotmail.comDIRECTORIO PREDETERMINADO

Inicio > Azure Synapse Analytics >

Creación de un área de trabajo de Synapse

\*Datos básicos

\*Seguridad

Redes

Etiquetas

Revisión y creación

Cree un área de trabajo de Synapse para desarrollar una solución de análisis empresarial con solo unos clics.

Detalles del proyecto

Seleccione la suscripción para administrar los recursos implementados y los costos. Use los grupos de recursos como carpetas para organizar y administrar todos los recursos.

Suscripción \*

Azure subscription 1

Grupo de recursos \*

rg-employees

Crear nuevo

Grupo de recursos administrado

employees

Detalles del área de trabajo

Asigne un nombre al área de trabajo, seleccione una ubicación y elija un sistema de archivos principal de Data Lake Storage Gen2 para que sirva como ubicación predeterminada de los registros y la salida del trabajo.

Nombre del área de trabajo \*

ws-employees

Región \*

East US

Seleccionar Data Lake Storage Gen2 \*

De la suscripción

Manualmente a través de la URL

Nombre de la cuenta \*

(Nueva) datalakeemployees1

Crear nuevo

Nombre del sistema de archivos \*

(Nueva) fs-employees

Crear nuevo

Asignarme el rol Colaborador de datos de Storage Blob en la cuenta de Data Lake Storage Gen2 para consultarlo de forma interactiva en el área de trabajo.

De forma automática, se concederá acceso a los datos de identidad del área de trabajo a la cuenta de Data Lake Storage Gen2 especificada con el rol Colaborador de datos de Storage Blob. Para permitir que otros usuarios usen esta cuenta de almacenamiento después de crear el área de trabajo, realice estas tareas:

Asignar a otros usuarios el rol Colaborador en el área de trabajo

Asignar a otros usuarios el rol Roles RBAC de Synapse pertinente usando Synapse Studio

Asignar a su usuario u otros usuarios el rol Colaborador de datos de Storage Blob en la cuenta de almacenamiento

Más información

Revisar y crear

< Anterior

Siguiente: Seguridad >

Ilustración 1 Creación de recurso Synapse

Ilustración 2 Mensaje de expiración

## No es apto para Evaluación gratuita.

No se preocupe, ya que seguirá pudiendo usar Azure con el sistema de pago por uso, sin compromisos iniciales. Puede cancelarlo en cualquier momento.

[Regístrese para obtener Azure con los precios de pago por uso.](#)

[Ponerse en contacto con soporte técnico](#)

*Ilustración 3Intento de crear segunda cuenta*

Se opta por cambiar de nube, creando una cuenta en AWS. Por limitaciones de tiempo, se decide para el ejercicio crear una base de datos SQL usando AWS RDS, ya que la naturaleza de los datos del desafío permitía esta elección.

### Ilustración 4 Servicios RDS

Ilustración 5 Creación servicio RDS

Se presentan varios inconvenientes para conectarse a la base de datos usando la librería pyodbc, se decide cambias a la librería sqlalchemy para tratar de solucionar estos problemas y porque nos permite directamente cargar un dataframe, al persistir los errores, se vuelve nuevamente a la primera opción y por las limitaciones de la librería se carga en la base de datos el vsc generado a partir del dataframe.

	employee_name	employee_email	phone	job_title	hire_date	
1	Esteban Alberto Contreras Gómez	vortega@example.org	4194854	Veterinario - 11	2015-08-26	
2	Orlando Orlando Arroyo	anagiraldo@example.org	+57 318 300 85 74	Administrador de bases de datos - 11	2017-10-25	
3	José Enrique Blanco Martínez	ramosronald@example.org	(+57)3106775262	Asistente fisioterapeuta - 11	2016-06-04	
4	Hernán Bravo Pulido	castellanosvirginia@example.net	+573018900595	Conserje - 11	2011-06-23	
5	Javier Johan Chávez Mora	ospinadiego@example.net	3113503905	Grabador de datos - 05	2010-10-12	
6	Carolina Esperanza Camacho	melissavelandia@example.net	+573252399247	Suboficial de las fuerzas armadas - 05	2012-05-09	
7	Myriam Espitia	danielcontreras@example.org	573164457499	Guardián de prisión - 05	2020-12-27	
8	Yesid Torres	galvisgeovanny@example.com	+57 303 275 77 46	Administrador de sistemas - 05	2014-08-05	
9	Juan Camargo Valderrama	silvaadriana@example.com	576085652563	Cantante - 73	2009-07-19	
10	Cecilia Ospino Bravo	jazminvalencia@example.com	(+57)6098873921	Tramper - 73	2009-12-02	
11	Inés Dora Gil Zapata	alexander93@example.org	603 670 45 52	Cantante - 18	2023-09-05	
12	Blanca Myriam Aguilar	jose97@example.org	+57 318 924 71 71	Operador de instalaciones de vidriería y cerámica - 18	2011-01-21	
13	Leydi Carmen Sánchez Calderón	mmontoya@example.net	+573117554144	Técnico de la Web - 18	2011-06-17	
14	Ruth Patricia Camacho	yadira37@example.org	01 8001 580 586	Director financiero - 18	2007-07-25	
15	Yaneth Portilla	npadilla@example.net	3172694966	Compositor - 68	2010-06-24	
16	Estela Liliana Cano	manuelgonzalez@example.com	573170516407	Instalador y reparador en tecnología de la información y las comunicaciones - 68	2009-06-25	
17	Adriana Isabel Quintero	valentinahernandez@example.net	(+57)3018801069	Ensamblador de maquinaria mecánica - 68	2014-08-02	
18	María Paola Galvis	molinaerika@example.net	+57 324 620 19 32	Ensamblador de equipos electrónicos - 68	2015-12-25	
19	Rafael Jhon Ojeda Gutiérrez	narvaezadriana@example.net	576037591836	Peluqueros - 50	2017-02-28	
20	Claudia Rita Carvajal	rosaperez@example.com	328 226 54 45	Maleter - 50	2024-01-10	
21	Andrea Lara	andres49@example.org	(+57) 314 693 77 09	Curador de museos - 17	2015-06-21	
22	Vanessa Alejandra Moreno	aldair19@example.com	+573258366838	Peón forestales - 17	2020-10-25	
23	Brayan Aurelio Fonseca	jhoncardenas@example.org	(+57)6097941609	Mecánico y reparador de máquinas agrícolas e industriales - 17	2019-01-27	
24	Sonia Medina Molina	cardonaangie@example.net	579 45 94	Agente de compras - 68	2013-04-28	
25	Edith Zuluaga Gómez	eliasosorio@example.org	573193126748	Cartógrafo - 68	2022-07-08	
26	Adriana Magnolia Ramírez Monroy	suarezalejandra@example.com	+576067181764	Diseñador de bases de datos - 68	2017-02-13	
27	Sonia Lucía Ibarra	xgarcia@example.org	018002938185	Regulador y operador de máquinas y herramientas - 25	2008-03-16	
28	Andrés Ángel Hernández	correaeduardo@example.org	573124437757	Peón de carga - 25	2008-02-05	
29	Fredy Sebastián Martínez Castillo	osalazar@example.org	573135997055	Técnico de laboratorio médico - 25	2010-11-30	
30	Blanca Torres Ramírez	rosalba40@example.net	(+57)3004377555	Técnico de telecomunicaciones - 25	2016-01-27	
31	Álvaro Gaviria	aviladaniela@example.net	(+57) 305 569 60 69	Profesor de idiomas - 18	2022-09-13	

Ilustración 6 Conexión a dbeaver con RDS

Para el desafío 5 se decide usar una lambda function para posteriormente complementar el desafío 6, se presentan inconvenientes para desplegarse desde visual studio code, se crea desde la interfaz pero presenta problemas por la librería pyodbc por no encontrarse el módulo, se crea un servicio en S3 para alojar la librería guardado como un zip y cargarse como una capa en la lambda function y tampoco lo lee, se decide cargar el zip también directamente en la lambda y persiste el error, por el tiempo disponible se descarta la idea y se crea una API REST usando Flask. Se adjunta evidencia.

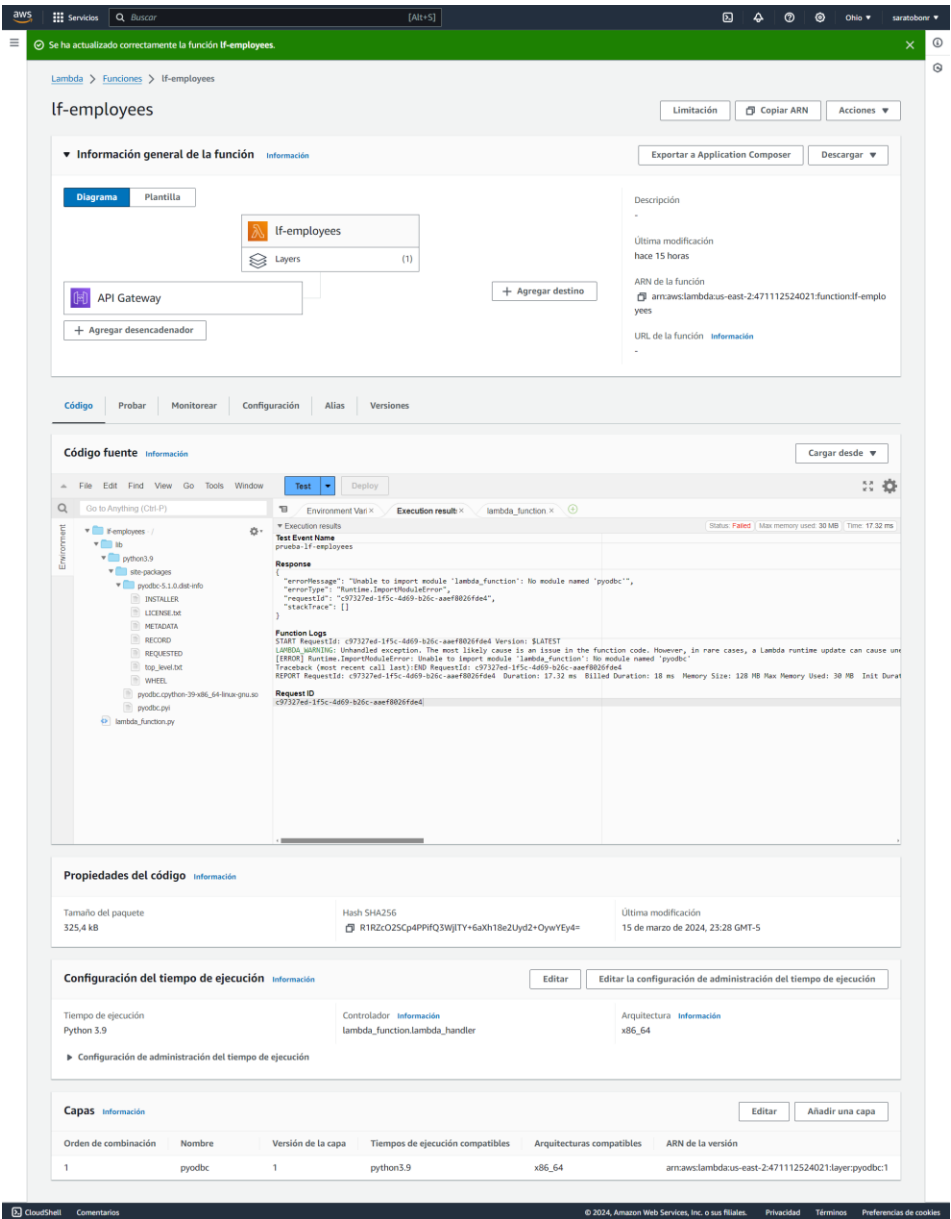


Ilustración 7 Lambda function problema librería pyodbc

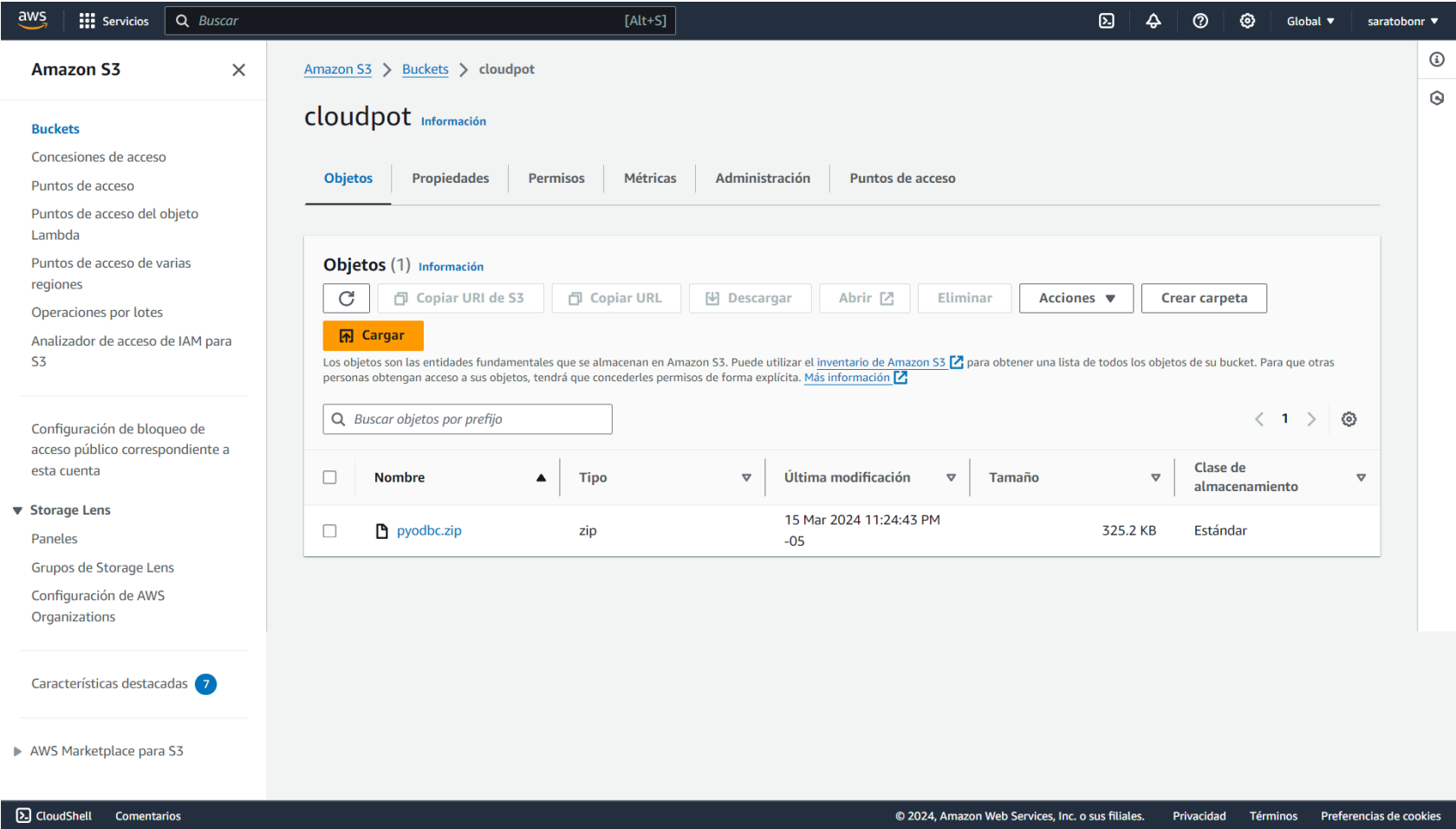


Ilustración 8 S3 con la librería en un comprimido

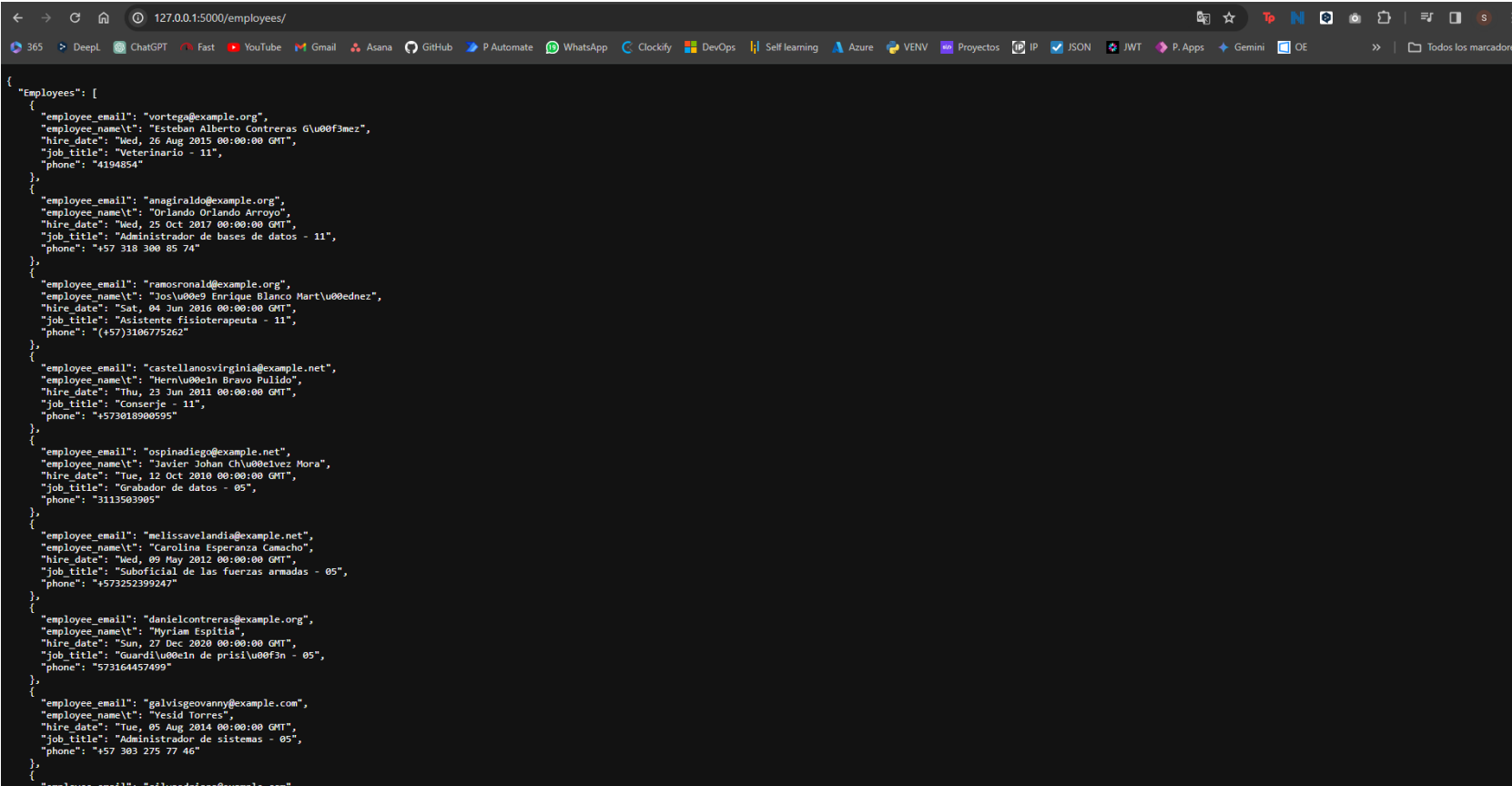


Ilustración 9 API REST en local

Para el desafío 6 se intenta desplegar la API y un Frontend básico desde AWS usando Beanstalk pero se descarta la idea porque los servicios ya implementados empiezan a generar cobros. Se adjunta evidencia.

## Listado de Empleados

*Ilustración 10 Resultado Frontend local consumiendo API en local*

Administración de facturación y costos

Inicio

Nuevo

Introducción

Nuevo

Facturación y pagos

Facturas

Pagos

Créditos

Pedidos de compra

Análisis de costos

Cost Explorer

Nuevo

Informes guardados del Explorador de costos

Detección de anomalías de costos

Nivel gratuito

Exportaciones de datos

Nuevo

Organización de costos

Cost Categories

Comentarios

Buscar

[Alt+S]

Descargar todo en CSV

Imprimir

Periodo de facturación: marzo de 2024

saratobonr

Hora de actualización de la página: el sábado, 16 de marzo de 2024 a las 22:41:21 GMT-5

Resumen de la factura estimada de AWS

Información

Cargos totales e información de pago

ID de cuenta

471112524021

Periodo de facturación

1 de marzo - 31 de marzo de 2024

Estado de la factura

Pendiente

Proveedor de servicios

Amazon Web Services, Inc.

Total en USD

2,00 USD

Total general estimado:

2,00 USD

► Información de pago

Información

Costo estimado más alto por proveedor de servicios

Información

Visualización Amazon Web Services, Inc.

Amazon Web Services, Inc.

Mayor gasto en servicios

1,88 USD

Tendencia en comparación con el mes anterior

Mayor gasto en la región de AWS

2,00 USD

Tendencia en comparación con el mes anterior

*Ilustración 11 Evidencia costos*

Punto 2:

Se empieza con la conexión siguiendo las recomendaciones y el primer paso es analizar la información para una primera familiarización, se toman medidas sobre la data (véase los comentarios en la pestaña ocultas llamada Notas del Power BI. Se adjunta evidencia.

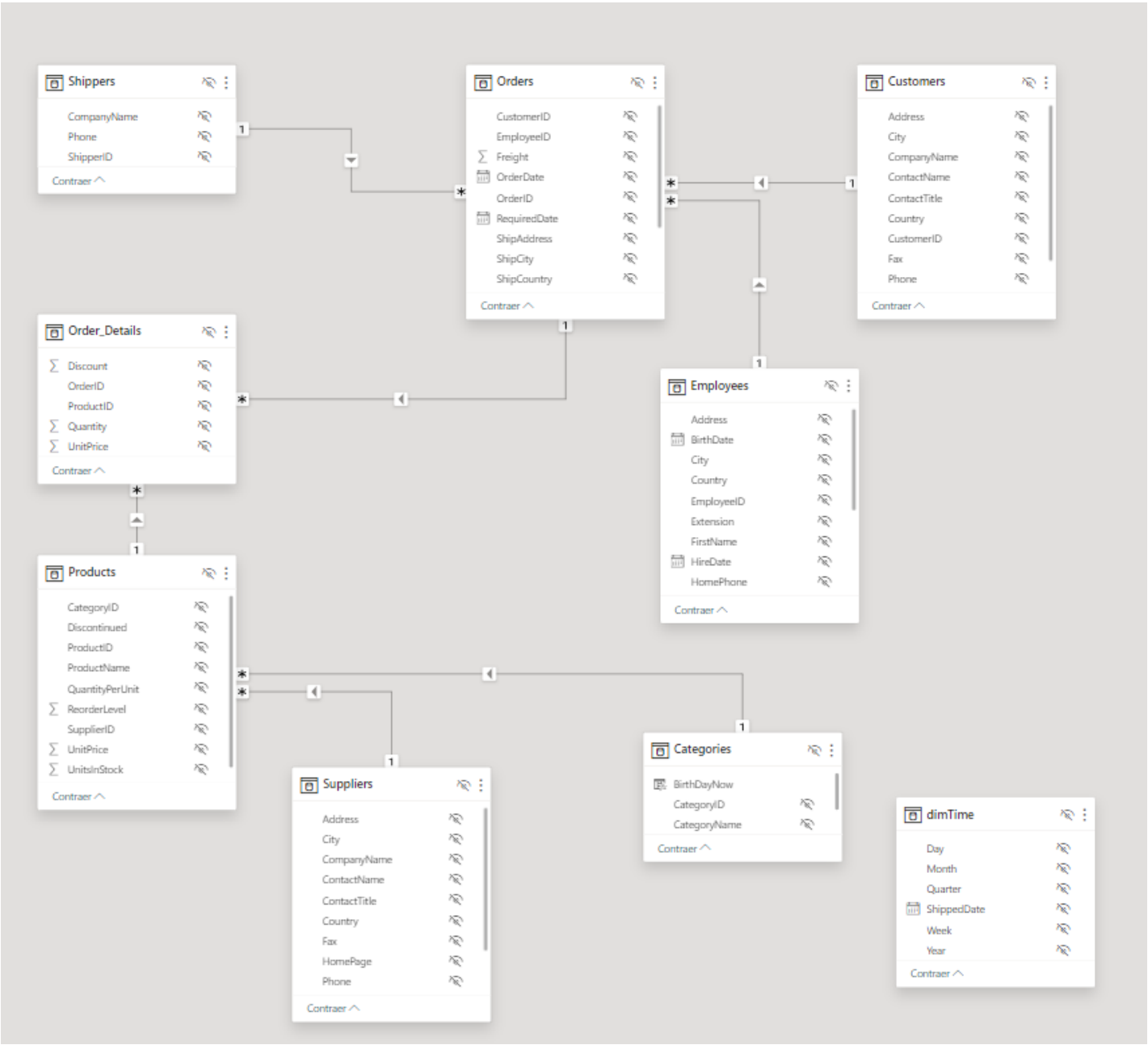


Ilustración 12 Modelo original

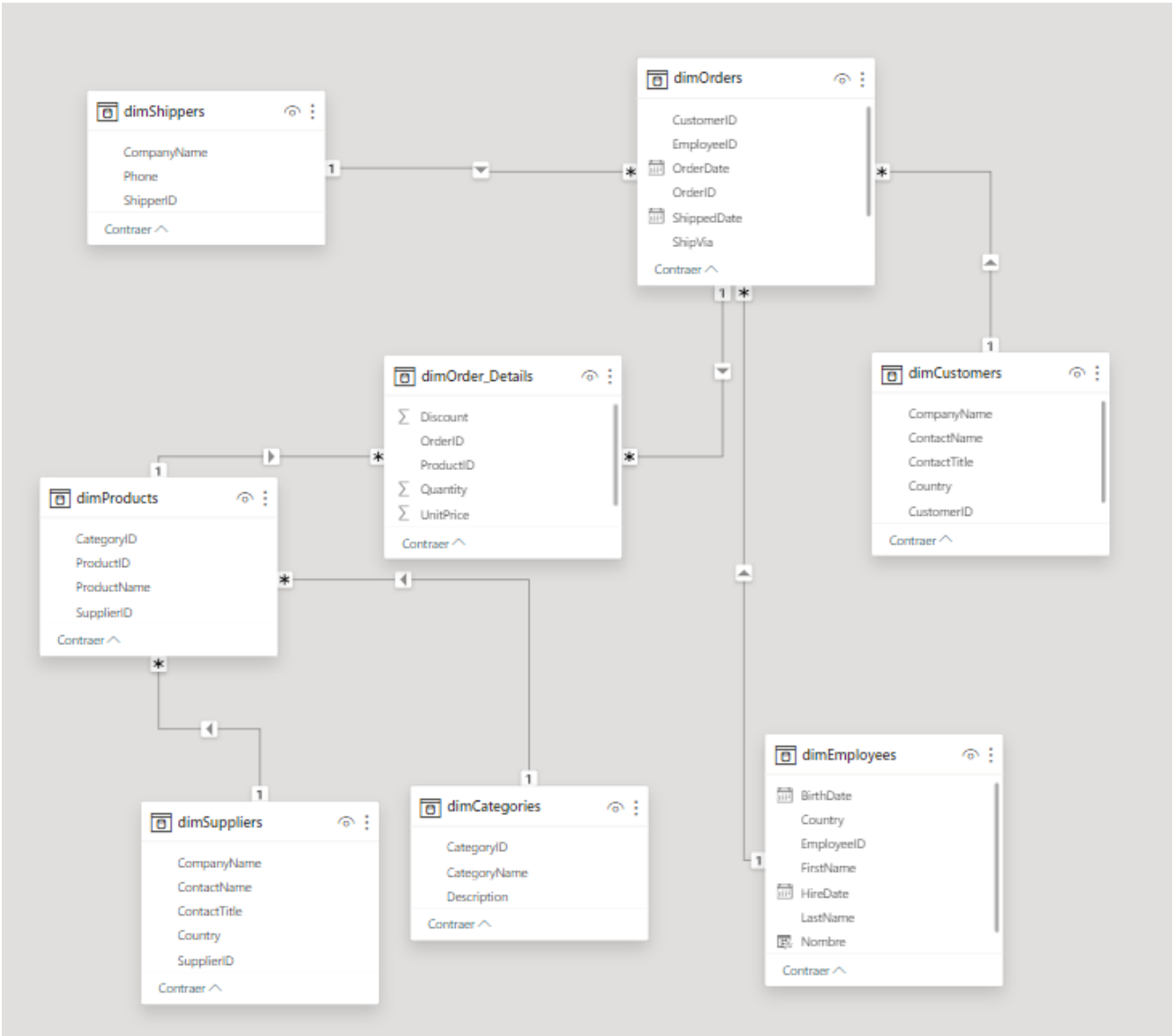


Ilustración 13 Modelo dimensional elegido