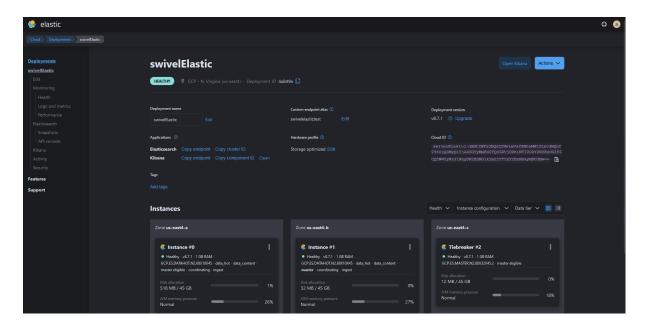
# Actividad de aprendizaje 3: Definición de un sistema de recuperación de información por medio de Elasticsearch

**Deployment**: swivelElastic



El deployment de Elasticsearch está siendo hosteada en la nube de Google Cloud Platform, sin embargo, la administración y manejo sigue siendo desde el servicio de Elastic Cloud. La decisión de utilizar GCP sobre otros proveedores de nube recae en la conveniencia y practicidad de poder acceder a más créditos gratuitos, y así tener un límite más grande para efectuar pruebas y levantar los recursos necesarios bajo una sola cuenta.

Las instancias usadas son aquellas que utilizan los recursos más mínimos, es decir se optó por el espacio de almacenamiento mínimo, y la capacidad de memoria mínima para las instancias de elasticsearch y kibana.

Dentro del cluster de Elastic, solamente se tiene un solo índice, este tiene como propósito guardar el catálogo de autos de la plataforma, y al mismo tiempo alimentar la búsqueda de este mismo catálogo con su habilidad de indexación y clasificación de palabras.

Los campos que se están guardando no solo son elementos descriptivos de un auto, sino que además se está guardando información relevante acerca del contexto del auto, tal como la dirección de la agencia y detalles de plazos y financiamiento en caso de su compra.

En la creación de un conjunto de autos cualitativo, se generaron varios scripts para la limpieza de datos, al igual que la completud de todos los campos en formato y contenido de cada entrada.

Adicionalmente, se automatizó la generación de descripciones de los autos en base a sus características como vehículo, las cuales nos permiten alimentar el sistema de recuperación de Elasticsearch usando lenguaje natural. Estas descripciones son breves pero sustanciosas de aproximadamente 350 palabras.

A partir de un conjunto vasto de 1000 entradas en el sistema de Elasticsearch se comenzó a realizar pruebas y queries con el fin de encontrar un balance entre la recuperación de autos y la precisión de la información que se consulta. Fue a través de una mezcla de diferentes parámetros que optamos por utilizar una búsqueda a varios campos, buscando que se encontrara un "match" con por lo menos dos palabras del query.

# Consultas de Lenguaje Natural al sistema de Elasticsearch:

Con un primer acercamiento a un query optimizado, se decidió hacer 20 consultas al índice de autos con la siguiente estructura de consulta.

1.

Prompt: Auto deportivo con cámara trasera ('took' 3, 'timed out': False, 'shards': ('total': 1, 'successful': 1, 'skipped': 0, 'failed': 0}, 'hits': ('total': {'value': 555, 'relation': 'eq'}, 'max\_score': 1 0.247161, 'hits': [{'\_index': 'autos', \_id': 'lakvcIgBUZmpqf-95hSE', '\_score': 10.247161, '\_ignored': ['colores.keyword', 'extras.keyword', 'descripcion.keyword'], '\_source': {'marca': 'Mazda', 'modelo': 'Mazda 2', 'año': 2019, 'precio': 286999.0, 'transmision': 'Automático', 'rendimiento': 18.5, 'combustible': 'Gasolina', 'litros' ': '1.5', 'motor': 'Inline-5', 'tipo vehiculo': 'Deportivo', 'pasajeros': 2, 'gerente\_id': '64756196b3044350ba23113', 'colores': "[{'nombre': 'Negro Noche', 'valor hex adecimal': '#F2F2F', 'imagenes': ['https://images.prd.kavak.io/eyJidWNrZXQi0iJPXZNaylpbWFnZXMhUTRyMiSQJOZXMhWTMMM]Y2L0VYVEVSSU9SLWZybz50012kZWBpbG90FmWhcioXMhiZWJOZXMhJTCyMhiSqvCaylogiolymore': 'Negro Azabache', 'valor hexadecimal': '#800900', 'imagenes': ['https://images.prd.kavak.io/eyJidWNrZXQi0iJrYXZhaylpbWFnZXMhUCJrZXki0iJpbWFnZXMhVTMMM]Y2L0VYVEVSSU9SLWZybg50012kZVBpbG90TmVhci0XNJMZNJUZUAUMTCyMiSqcGciLCJlZGlocyIGeyJXZXWpmUllonsid2lkdGgi0JUDWCwiaGVpzZh0IjozMTB97K0e-']}, '(nombre': 'Marron Oscuro', 'valor hexadecimal': '#884513', 'imagenes': ['https://images.prd.kavak.io/eyJidWNrZXQi0JJrYXZhaylpbWFnZXMiLCJrZXki0JpbWFnZXMhUTMMJY2L0VYVEVSSU9SLWZybG905012kZVBpbG90TmVhci0XNJMZNJUZUAUMTCyMiSqcGciLCJlZGlocyIGeyJXXNJpemUllonsid2lkdGgi0JUMCwiaGVpz2h0Ijoz MTB97K0e-']}, '(nombre': 'Marron Oscuro', 'valor hexadecimal': '#884513', 'imagenes': ['https://images.prd.kavak.io/eyJidWNrZXQi0JJrYXZhaylpbWFnZXMiLCJrZXki0JpbWFnZXMJCJloleyJGeyJZXJXNpemUllonsid2lkdGgi0JUMCWlaGVpZXNDjoJDJYXZNJAylpbWFnZXMJCJrZXXNidJpbWFnZXMJCJrZXXNidJpbWFnZXMJCJrZXXNidJpbWFnZXMJCJrZXXNidJpbWFnZXMJCJrZXXNidJpbWFnZXMJCJrZXXNidJpbWFnZXMJCJrZXXNidJpbWFnZXMJCJrZXXNidJpbWFnZXMJCJrZXXNidJpbWFnZXMJCJrZXXNidJpbWFnZXMJCJrZXXNidJpbWFnZXMJCJrZXXNidJpbWFnZXMJCJrZXXNidJpbWFnZXMJCJrZXXNidJpbWFnZXMJCJrZXXNidJpbWFnZXMJCJrZXXNidJpbWFnZXMJCJrZXXNidJpbWFnZXMJCJrZXXNidJpbWF

2.

took': 3, 'timed out': False, 'shards': ('total': 1, 'successful': 1, 'skipped': 0, 'failed': 0), 'hits': {'total': {'value': 300, 'relation': 'eq'}, 'max score': 120428, 'hits': {{'index': autos', 'id': 'rōkvcIgBUzmpof-92hT'., 'score': 13.20428, 'jonored': ['colores keyword', 'extras keyword', 'descripcion keyword'], 'so c': ('marca': 'Mazda', "nodelo: 'Mazda' 3', 'anö': 2017, 'precio': 28999.0, 'transion: 'Automatico, 'rendimento': 9.6, 'combustible': 'Gascipcion keyword'], 'so c': ('marca': 'Mazda', "nodelo: 'Mazda' 3', 'anö': 2017, 'precio': 28999.0, 'transion: 'Automatico, 'rendimento': 9.6, 'combustible': 'Gascipcion', 'yalor hexadeci', 'imagenes': ['https://images.prd.kavak.io/ey]idWhrZX0i0i]YYXZhay]pbWFnZXWhILD!ZXK0J]pbWFnZXWMJMMDY2L0YYVEX5U95LWZyb250U2lkZVBpbG90TWhcioxNhjgwMD13
EwMZMkLmpvZkciLCllZGlocyIdeyJgzXNpembiJonsid2lkdGgi0jUdMcwiacVpZzh0IjozMTB9fX0=']}]', 'caracteristicas: "['Faros LED', 'Sensor de estacionamiento', 'Asientos con memoria', 'Pantalla táctil']', 'extras': '[('titulo': 'Sistema de Asistencia de Mantenimiento de Carril', 'precio': 7500.0, 'descripcion': 'Nyuda antener el vehículo en el centro de Larril de manera automatica'), {'titulo': 'Sistema de Reconocimento de Señales de Tráfico', 'precio': 7500.0, 'descripcion': 'Origoni': 'Alerta al conductor cuando se produce un cambio de carril involuntario')]', 'enganche': '[15, 25, 30, 35]', 'plazos': '('12': 4.0, '24': 4.1, '36': 4.2, '4': 4.4, '6': 4.6," '('totos 3d': '['thtps://mages.prd.kawki.o/eyJidMYPXX0ilol/PXXNPJWFNZXMVIJCAYZXWiJDYBVZVBJVSUSWLZVByBDG97MVhcioxN.WDI3MTEWMZWLIDMZDVBVZUSVBVLZVBVBLXVBJPSUSWLZVBJVSUSWLZVBJPSUSWLZVBJVSUSWLZVBJPS

rompt: Ford azul familiar
'took': 4, 'timed out': False, 'shards': {'total': 1, 'successful': 1, 'skipped': 0, 'failed': 0}, 'hits': {'total': {'value': 9, 'relation': 'eq'}, 'max scr
5631, 'hits': {\'index': 'autos', 'id': 'gqkwcIgBUZmpqf-9TBUC', 'score': 8.056631, 'ignored': ['colores.keyword', 'descripcion.keyword'], 'source': {'ma
'd', 'modelo': 'Escape', 'año': 2013, 'precio': 234999.0, 'transmision': 'Automatico', 'rendimiento': 9.8, 'combustible': 'Gasolina', 'litros': '2,5', 'motor'
ocharged', 'tipo vehiculo': 'Familiar', 'pasajeros': 5, 'gerente\_id': '6475d196b304a50ba23a1131', 'colores': "['nombre': 'Gris Azudado' 'valor hexadecimal';
', 'imagenes': ['https://images.prd.kavak.io/eyJidWnrZX010117YXZhay1pbWFnZXMLC17ZXX1011pbWFnZXMMJAyNJg4L0VYVEVSSU9SLWZyb250U2kZVBpb690TmVc10XNJYZMDk200gz/
cii.CollZclocyGeyJyZXNpemU10nsid2lkdGg10jU0MCwiaGVp2ZxD15J0XMTB9TX0=']}, 'nombre': 'Gris Claro', 'valor hexadecimal': '#030333', 'Imagenes': ['https://images.prd.kavak.io/eyJidWnrZX01011rYXZhay1pbWFnZXMMJAyNJg4L0VYVEVSSU9SLWZyb2DSULXVByBb690TmVhc10XNJYZMDk2D0gzMDF1LmpxZWc1.CollZoflocyTeyJXXDpemU10nsid2lkdGg10jU0MCwiaGVp2ZxD10J0XTB9TX0=']}, 'ionbur': 'Rojo Oscuro', 'valor hexadecimal': '#0303050', 'magenes': ['https://images.prd.kavak.io/eyJidWnrZX01011rYXZhay1pbWFnZXMLOTAXX1011pbWFnZXMMJAyNJg4L0VYVEVSSU9SVEXVPSS09SUZXVPSS0

5.

npt: Auto negro o blanco para familias grandes
opk': 4, 'timed out': False, 'shards': ('total': 1, 'successful': 1, 'skipped': 0, 'failed': 0), 'hits': ('total': {'value': 894, 'relation': 'eq'}, 'max\_scc
16, 'hits': [{' index': 'autos', 'id': 'AkkvcIgBUZmppf-98xWZ', 'score': 7.3646, 'ignored': ['colores.keyword', 'extras.keyword', 'descripcion.keyword'], '
("marca': 'Chevrolet', 'modelo': 'Malibu', 'año': 2015, 'precio': 262999.0, 'transmision': 'Automático', 'rendimiento': 12.0, 'combustible': 'Gasolina', 'litr
, 'motor': 'Inline-5, 'tipo\_vehiculo: 'SUV', 'pasajeros': 7, 'gerente\_id': '6475d196b0304450ba2a3113', 'colores': "[{'nombre': 'Marrón Oscuro', 'valor'#884513', 'imagenes': ['https://images.prd.kavak.io/ey]idWhrZXX01011rYXZhay1pbWFnZXMHICO17ZXK1011pbWFnZXMMIDOSHUZYbZS01021kZVBpb690TmWhci0xNlY12
jbu0mpwZWcilcO12Glocy16dvyJZXNpemblionsid2lkdogi0jUMDVcia6VpZZDR1jozMTB9TX0e-11, 'nombre': 'Marrón Oscuro', 'valor-lexedecimal': "#884513', 'imagenes': ['https://image.sprd.kavak.io/ey]idWhrZX01011rYXZhay1pbWFnZXMHICO17ZXK1011pbWFnZXMMTD9TMPXMHICO17ZXK1011pbWFnZXMMTD9TMPXMHICO17ZXK1011pbWFnZXMMTD9TMPXMHICO17ZXK1011pbWFnZXMMTD9TMPXMHICO17ZXK1011pbWFnZXMMMTG9MTAwL0VVEVSSU9SLWZyb259U21kZVBpb690TmWhci0xNlY3D107MPXQwMJUGHmpXMC11011ZGlocy16eyJyZXNpemblionsid2lkdogi0jUMPCwia6VpZZb1jozMTB91
'caracteristicas': "['Control de tracción', 'Sistema de arranque remoto']", 'extrac': "[{'titulo': 'Asistente de Arranque en Pendientes', 'precio': 5500.0, 'descripcion': 'Activa y desact
uces del vehículo automáticamente según las condiciones de luz'), {'titulo': 'Encendido Automático de Luces', 'precio': 3500.0, 'descripcion': 'Activa y desact
uces del vehículo automáticamente según las condiciones de luz'), {'titulo': 'Encendido Automático de Luces', 'precio': 3500.0, 'descripcion': 'Activa y desact
uces del vehículo automáticamente según las condiciones de luz'), {'titulo': 'Encendido Automático de Luces', 'precio': 3500.0, 'descripcion': 'Activa y desact
uces del vehículo automáticamente según las condiciones de luz'), {'titul

6.

Prompt: Quiero un auto para uso en carretera y ciudad, que sea económico y que tenga un buen rendimiento de combustible.

('took': 5, 'timed out': False, 'shards': {'total': 1, 'successful': 1, 'skipped': 0, 'falled': 0}, 'fhits': {'total': {'value': 1000, 'relation': 'eq'}, 'max\_score': 11.034745, 'hits': {'index': autos', 'id': 'CakvclgBUZmpqf-99hVO', 'score': 11.034745, 'inored': ['colores.keyword', 'extras.keyword', 'descripcion.keyword'], 'source': {'motor': 'Vla', 'tipo whiculo': 'Scádn', 'pasajeros': 6, 'gerente id': '647519600304305ba23all3'; 'colores': "{\tombre': 'Negro Noche', 'valor hexadecimal': '#2FFFFFF', 'imagenes': ['thtps://images.prd.kavak.io/ey]idWhrZXQidiJrYXZhaylpbWFnZXMLQJTZXKNOiJpbWFnZXWMjNYDDEW.BVYVEVSSU9SLWZyb2S0UZ\kZVBpbG99TmVhcioXNj;cODEMZwkWTAYLmpXwGwcilcillZiGoleyGeyJzXVMppemUlonsidQxGdfgGj0jUBWCmadyVZzADJjozMTB9fXe='], ('nombre': 'Negro Noche', 'valor hexadecimal': '#FFFFAFF', 'imagenes': ['https://images.prd.kavak.io/ey]idWhrZXQidiJrYXZhaylpbWFnZXWhJKyDbGS9TmVhcioXNj;vaMgyDyDBWFnZXWhJKyDbGS9TmVhcioXNj;vaMgyDyDBWFnZXWhJKyDbGS9TmVhcioXNj;vaMgyDyDBWFnZXWhJKyDbGS9TmVhcioXNj;vaMgyDyDBWFnZXWhJKyDbGS9TmVhcioXNj;vaMgyDyDBWFnZXWhJKyDbGS9TmVhcioXNj;vaMgyDyDBWFnZXWhJKyDBGS9TmVhcioXNj;vaMgyDyDBWFnZXWhJKyDBGS9TmVhcioXNj;vaMgyDyDBWFnZXWhJKyDBGS9TmVhcioXNj;vaMgyDyDBWFnZXWhJKyDBGS9TmVhcioXNj;vaMgyDyBWFnZXWhJKyDBGS9TmVhcioXNj;vaMgyDyBWFnZXWhJKyDBGS9TmVhcioXNj;vaMgyDyBWFnZXWhJKyDBGS9TmVhcioXNj;vaMgyDyBWFnZXWhJKyDBGS9TmVhcioXNj;vaMgyDBWFnZXWhJKyDBGS9TmVhcioXNj;vaMgyDBWFnZXWhJKyDBGS9TmVhcioXNj;vaMgyDBWFnZXWhJKyDBGS9TmVhcioXNj;vaMgyDBWFnZXWhJKyDBGS9TmVhcioXNj;vaMgyDBWFnZXWhJKyDBGS9TmVhcioXNj;vaMgyDBWFnZXWhJKyDBGS9TmVhcioXNj;vaMgyDBWFnZXWhJKyDBGSPJKyDBGS9TmVhcioXNj;vaMgyDBWFnZXWhJKyDBGSPJKyDBGS9TmVhcioXNj;vaMgyDBWFnZXWhJKyDBGSPJKyDBGS

Prompt: Dame un auto que sea de color rojo y que sea de la marca Ford. Del 2016 ('took': 1, 'timedourt': False, 'shards': ('total': 1, 'skuped': 0, 'failed': 0}, 'hits': ('total': ('value': 1000, 'relation': 'eq'), 'max score': 11.783677, 'hits': [{'index': 'autos', 'id': 'RôkwcIgBUZmpqf-9ChVz', 'score': 11.783677, 'inpored': ['colores, keyword', 'extras, keyword', 'descripcion, keyword'], 'source': ('marca': 'Ford', 'modelo': 'Expedition', 'ano': 2015, 'precio': 443999 0, 'transmision': 'Automático', 'rendimiento': 9.2, 'combustible': 'Gasolina', 'litr os': '4', 'motor': 'V10', 'tipo vehiculo': 'Familiar', 'pasajeros': 5, 'gerente\_id': '6475d196b0304a56ba2a3113', 'colores': "[('nombre': 'Gris Claro', valor\_hexadectal': '#030303', 'unagenes': ['https://images.prd.kavak.io/ey)idWhrZX0101JrYXZhaylpbWfnZXMhICL7ZKkiO1JpbWfnZXMhVHXSMDgxL0VYVEVSSU9SLWZyb2S09SLWZyb2S09SLWZyb2S09SLWZyb2S09SLWZyb2S09SLWZyb2S09SLWZyb2S09SLWZyb2S09SLWZyb2S09SLWZyb2S09SLWZyb2S09SLWZyb2S09SLWZyb2S09InWhrZXMiDiJDWFnZXMhVHXMDBMFNZWHICLOTZXKiO1JpbWfnZXMhIMPXMADMBYVEVSSU9SLWZyb2S00ZlkZYbpBG96TmWhcioXhjgyMDY4MZAOMJUSLmpxZwciLClZGlogyIGeyJZXNpemBiOnsidZlkdGgj0JUMCwiaGVZZhdIjoxMTBFX0='], 'nombre': 'Blanco Perla', 'valor hexadectamal': '##BF8FF', 'imagenes': ['https://images.prd.kavak.io/eyJidWhrZX0i0iJDWTAZMYVEVSSU9SLWZyb2S00ZlkZYBpbG99TmWhcioXhjgyMDY4MZAOMJUSLmpxZwciLClZGlogyIGeyJZXNpemBiOnsidZlkdGgj0JUMCwiaGVZZhGIJoxMTBFX0='])", 'cractarctristicas': '"|'Sistema de sonido premium', 'tuces automáticas', 'Sistema de navegación'!", 'extras': '['titulo': 'Enendido Automáticas', 'Sistema de navegación'!", 'extras': '['tit

R

9.

pte/../.debugpy/launcher 34141 -- /mnt/C038AED238AEC72E/Repository/Universidad/6toSemestre/TC3005B/GenerateDummyData/createSpanishIndex.py
pt: Autos compactos para llevar a la familia de vacationes.
8801. hits: [{' inded out: 'False, ' shards: {' total': |, 'successful': 1, 'skipped': 0, 'failed': 0}, 'hits': {' total': {' value': 1000, 'relation': 'eq'}, 'max\_score':
8801. hits: [{' index': 'autos', 'iof: 'LKkwclgBUZnpqf'.9oxf7', 'score': 9.568861, 'ajonored': [' clores.keyword', 'extras.keyword', 'descripcion.keyword', 'descripcion' 'kayoord', 'descripcion' 'descripcion' 'kayoord', 'descripcion' 'ka

adapter/../../debugpy/launcher 36421 -- /mnt/C038AED238AEC72E/Repository/Universidad/6toSemestre/TC3905B/GenerateDummyData/createSpanishIndex.py
rompt: Automóvil para climas frios y buen sonido.
'took': 4, 'timed out': Faise, 'sharfs': ('total': 1, 'successful': 1, 'skipped': 0, 'failed': 0), 'hits': ['total': ['value': 999, 'relation': 'eq'], 'max\_scorr
7659087, 'hits': [['indek': 'autos', 'id': 'OKKwcIgDUZmpqf-9BRXg', 'score': 6.7659087, 'ignored': ['colores.keyword', 'descripcion.keyword'], 'descripcio

# 12.

E/TC3005B; /usr/bin/env /home/andres/miniconda3/envs/databases/bin/python /home/andres/.vs.code/extensions/ms-python.python.python.2023.8.0/pythonFileS/liOp bin/env /home/andres/miniconda3/envs/databases/bin/python /home/andres/.vscode/extensions/ms-python.python-2023.8.0/pythonFiles/lib/python/debug y/launcher 43595 -- /mnt/C038AED238AEC72E/Repository/Universidad/6toSemestre/TC3005B/GenerateDummyData/createSpanishIndex.py y/adapter/../../debugpy/launcher 44395 -- /mnt/C038AED238AEC7ZE/Repository/Universidad/6toSemestre/TC30058/GenerateDummyData/createSpanishIndex.py
Prompt: Carro para escuchar musica y viajar con amigos y para ir a la escuela.
['took': 6, 'timed out': False, 'shards': ('total': 1, 'successful': 1, 'skiped': 9, 'failed': 9), 'hits': ('total': ('value': 1000, 'relation': 'eq'), 'max score':
1.589243, 'ingore': ('index': 'wiltos', 'id': 'wakvc[98Umpqf-9gRWM'] - 'score': 11.592943, 'ingore': (Fores keyword', 'extras keyword', 'descripcion. keyword', 'score': Itop vehiculo': 'Sedan', 'pasajeros: 6, 'gerente id': 6475d195b0304450ba23a113', 'colores: "[('tombre': Gris Medio', 'valor hexadecimal'; '#808080; 'limagenes: ['https://images.prd.kavak.a/e/ey]didMlrZQ1011/YZD1Ay1pbMfnZWMHIQ17XKH011pbMfnZWMHIVMALDWYVESSU9SLWZ}b959012kZ%pb590

#### 14.

Prompt: Quiero un coche hibrido que sea de color rojo y que sea de la marca Toyota.

'totok': 8, 'timed out': 'false,' shards': ('total': 1, 'successful': 1, 'skipped': 9, 'failed': 0), 'hits': ('total': ('value': 1000, 'relation': 'eq'), 'max score': 13.2096, 'hits': [('index': 'autos', 'id': 'MakwclgBUZmpqf-9AwNt', 'score': 13.2096, 'intramission': 'Automático' 'rendimiento': 31.3 'combustible': 'Gasolina' 'litros': '1.5', 'motor': 'Boxer-4', 'tipo vehiculo': 'Hatchback', 'pasajeros': 6, 'gerente id': '6475410560304a508a2a3113' (colores': "[('nombre': 'Gris Medio', 'valor hexadec imal': '#860808', 'imagenes': ['https://images.prd.kavak.io/eyjidWhrZXQiloljrYXZhaylpbWhrZXQHLClJEGlecyImpxZWciLClJEGlecyImpxZWciLClJEGlecyImpxZWciLClJEGlecyImpxZWciLClJEGlecyImpxZWciLClJEGlecyImpxZWciLClJEGlecyImpxZWciLClJEGlecyImpxZWciLClJEGlecyImpxZWciLClJEGlecyJEdployArgpbG9TmWhcloXHjUNDY Mages.prd.kavak.io/eyjidWhrZXQJJDbWrAzyJbWhrZXQHDpWhrZXHLCJTZKGlecyJEdployArgpbG9TmWhcloXHjUNDY Mages.prd.kavak.io/eyjidWhrZXQJDlojlrYZXphgpidiployArgpbG9TmWhcloXHjUNDYWciLClJEGlecyJEdployArgpbG9TmWhcloXHjUNDYWciLClJEGlecyJEdployArgpbG9TmWhcloXHjUNDYWciLClJEGlecyJEdployArgpbG9TmWhcloXHjUNDYWciLClJEGlecyJEdployArgpbG9TmWhcloXHjUNDYWciLClJEGlecyJEdployArgpbG9TmWhcloXHjUNDYWciLClJEGlecyJEdployArgpbG9TmWhcloXHjUNDYWciLClJEGlecyJEdployArgpbG9TmWhcloXHjUNDYWciLClJEGlecyJEdployArgpbG9TmWhcloXHjUNDYWciLClJEGlecyJEdployArgpbG9TmWhcloXHjUNDYWciLClJEGlecyJEdployArgpbG9TmWhcloXHjUNDYWciLClJEGlecyJEdployArgpbG9TmWhcloXHjUNDYWciLClJEGlecyJEdployArgpbG9TmWhcloXHjUNDYWciLClJEGlecyJEdployArgpbG9TmWhcloXHjUNDYWciLClJEGlecyJEdployArgpbG9TmWhcloXHjUNDYWciLClJEGlecyJEdployArgpbG9TmWhcloXHjUNDYWciLClJEGlecyJEdployArgpbG9TmWhcloXHjUNDYWciLClJEGlecyJEdployArgpbG9TmWhcloXHjUNDYWciLClJEGlecyJEdployArgpbG9TmWhcloXHjUNDYWciLClJEGlecyJedployArgpbG9TmWhcloXHjUNDYWciLClJEGlecyJedployArgpbG9TmWhcloXHjUNDYWciLClJEGlecyJedployArgpbG9TmWhcloXHjUNDYJEGlecyJedployArgpbG9TmWhcloXHjUNDYJEGlecyJedployArgpbG9TmWhcloXHjUNDYJEGlecyJedployArgpbG9TmWhcloXHjUNDYJEGlecyJedployArgpbG9TmWhcloX

#### 15.

Prompt: Ecologico, con diseño moderno, que sea adapte a la ciudad y que sea fácil de estacionar.

('took': 8, timed out': false, 'shards': ('total': 1, 'successful': 1, 'skipped': 8, 'falled': 8), 'hits': ('total': ('value': 999, 'relation': 'eq'), 'may score': 16, 899227, 'ignored': ('colores keyword', 'doscripcion keyword'), 'source': ('marca': 'Toyota', 'mode': ('Toyota', 'mode': ('Colores keyword', 'doscripcion keyword'), 'source': ('marca': 'Toyota', 'mode': ('Toyota', 'mode': ('Colores keyword', 'doscripcion keyword'), 'source': ('marca': 'Toyota', 'mode': ('Toyota', 'mode': ('Toyota'), 'mode'

## 18.

('took': 2, 'timed out': False, 'shards': ('total': 1, 'successful': 1, 'skipped': 9, 'failed': 9}, 'hits': ['total': {'value': 999, 'relation': 'eq'}, 'max\_score': 1.815291, 'hits': [['index': 'autos', 'id': '8akwcigBungd-98Rb', 'score': 11.815291, 'ignored': ['colores.keyword', 'extras.keyword', 'descripcion.keyword'] to colored to the colored to

20.

mpt: De carreras, con motor de alto rendimineto, deportivo

posible carreras, con motor de alto rendimineto; de carreras, com control carreras, control carreras, com control carreras, contro

#### Aportaciones del grupo:

#### Andres Briseño:

Con el fin de poblar la base de datos de Elasticsearch se tomó un conjunto de 1000 autos previamente obtenido y se limpiaron los campos para poder tener un esquema estandarizado de los autos. Esta limpieza se realizó a través de un script de python, leyendo un archivo csv, modificándolo según fuera necesario. Adicionalmente a esto, se logró automatizar la generación de descripciones de los autos implementando el API de OpenAI chat, en el cual se utilizaban las características como parámetros principales. Finalmente, con un conjunto completo en cuanto a sus campos, se creó un script que suba este csv a la instancia de Elastic Cloud y se poblara el índice. Posteriormente, en equipo, se realizaron varias consultas para poder optimizar el tipo de query que podría devolver los mejores resultados.

#### Gerardo Gutierrez:

Buscando la estandarización de un esquema para el índice de Elasticsearch se trabajó en consonancia con los desarrolladores de back-end para asegurar la compatibilidad entre formatos y calidad de la información que se incluiría en el conjunto de datos del índice. También se desplegó una instancia de Elasticsearch hospedada en GCP mediante el servicio Elastic Cloud, y se generaron pruebas de conexión y llaves o api Keys para poder utilizar la instancia en calidad de desarrollo constante. Finalmente, se redactó un reporte de acuerdo a la actividad de clase y del proceso que se realizó para completar esta tarea de integración de un sistema de recuperación; En equipo, una vez que se pobló el índice con un conjunto de autos, se realizarán varias consultas distintas con el fin de buscar un query que fuera óptimo para regresar un balance entre cantidad de resultados y la calidad de estos.

#### Karla Mondragón:

El área principal de nuestra aplicación en la que Elasticsearch se integra es la búsqueda y filtros dentro del catálogo. La búsqueda dentro del catálogo es parte fundamental del flujo del comprador ya que es por medio de ella y los filtros que se utilizan (también dentro del catálogo) que el usuario comprador puede acceder y acotar los diferentes autos que existen registrados dentro de la plataforma y realmente encontrar un vehículo que se ajuste a su gusto y necesidades. Es esta posibilidad la que le brinda valor a la plataforma, una herramienta que le permita al usuario buscar libremente de acuerdo a sus requisitos, ya sea color, marca, modelo, año, transmisión, precio, entre otros campos, que se utilizan para describir y elegir un automóvil. Tanto la barra de búsqueda (que utiliza Elasticsearch) como algunos filtros que también se pueden encontrar en la parte principal del landing page con un conjunto de prompts dinámicos, sirven como ejemplo de lo que un usuario puede buscar dentro de la plataforma así llamando la atención del usuario a uno de los features principales de la aplicación.

#### Mateo Herrera Lavalle:

Siendo el punto principal de nuestra aplicación el realizar una búsqueda sencilla e intuitiva dentro de un catálogo de autos, las funcionalidades que Elasticsearch proveen se volvieron parte importante dentro del desarrollo del proyecto. En específico, se buscó hacer uso de Elasticsearch como motor de búsqueda de lenguaje natural. Para esto, dentro de un índice de Elasticsearch, que se encuentra alojado en Google Cloud Platform (GCP) se guardaron

entidades "auto" que contienen datos generales del automóvil en cuestión, como lo son la marca, el modelo, el color, el precio, las cuales son útiles si se desea realizar un filtrado de información. Además de esto se guardó un campo llamado descripción. Este campo es el más importante para la funcionalidad de búsqueda con lenguaje natural, ya que consiste de un string continuo y largo en donde se describe al automóvil usando lenguaje coloquial. Esto permite que se puedan realizar queries en donde se busque una cadena de texto y se regresen los documentos que tengan coincidencias parciales o totales con lo ingresado. Para esto, la entrada se divide en "tokens" y utilizando una métrica de ranking, comúnmente "TF-IDF", se regresan los documentos que contengan un mejor grado de similitud. Es por esto que la descripción completa y amplia era de gran importancia, hecho para el que se usó el API de Chat-GPT como soporte para su automatización.

Más aún, se buscó un cierto grado de realismo en los datos que se utilizaron dentro de la plataforma, por lo que técnicas de scraping fueron utilizadas para recopilar información de sitios con mercados similares. En base a estos, se generó una amplia gama de dummy data que ya se encuentra en la nube.

#### Luis Javier Karam:

Se utilizó Elasticsearch en nuestro proyecto con la finalidad de mejorar la búsqueda de automóviles en nuestro catálogo, esto es ya que Elasticsearch abre la posibilidad de realizar búsquedas rápidas y precisas en una gran cantidad de datos. Además, proporciona funciones más avanzadas como filtrado, clasificación y agregación de resultados, lo cual mejora la experiencia del usuario al encontrar automóviles en nuestro catálogo. La capacidad de ElasticSearch para indexar y buscar documentos en tiempo real también garantiza que los resultados sean siempre actualizados y siempre se reflejen los cambios más recientes a la hora de realizar la búsqueda, también cabe mencionar que no solo se guarda la información de los autos sino también información de las agencias y los locales en donde se encuentran para mejorar y darle más flexibilidad a la búsqueda. El uso de Elasticsearch en nuestra barra de búsqueda y filtros de catálogo nos ayudó a mejorar la experiencia de usuario a la hora de realizar búsquedas ya que este es capaz de tomar lenguaje natural y regresar resultados correspondientes, lo cual proporciona una experiencia más personalizada para cada usuario.

# Ana Paula Katsuda Zalce:

En cuanto al uso que se le está dando a Elasticsearch para el presente proyecto, se realiza la integración de la misma con el principal objetivo de mantener un manejo adecuado del catálogo de autos para la aplicación. Además de permitir el guardado de dichos autos, Elasticsearch contribuye a la búsqueda efectiva de los mismos, utilizando sus distintas características tales como la clasificación de palabras. En específico, es posible comprender la implementación que se le da mediante los datos que se almacenan: se tienen las características generales de los autos (tales como marca, modelo y año), se manejan datos tales como los distintos colores que pueden estar disponibles para cada automóvil, sus imágenes e incluso los plazos para calcular las mensualidades. Adicional a los datos básicos, los autos guardados contienen información relevante respecto a las agencias en donde se encuentran y la adquisición de autos sea más sencilla (al identificar a los proveedores).

Agregando a las posibilidades de clasificación de palabras, el campo que permite la búsqueda por lenguaje natural es el de descripción, el cual contiene una cadena en donde se especifica el autor y sus características de manera que se puedan encontrar coincidencias posteriormente. Definitivamente, el uso de Elasticsearch brinda una mayor capacidad a nuestra aplicación, pues facilita y potencializa una de las principales funcionalidades de la misma: la búsqueda que se adapte a las necesidades del cliente.

# Regina Rodríguez Sánchez:

En nuestra aplicación el uso de Elasticsearch creemos que es fundamental para mejorar la experiencia del usuario para la búsqueda de un auto dentro del catálogo. El uso que se le está dando dentro de la aplicación es para que los usuarios puedan buscar el auto que desean encontrar de una forma más descriptiva que solo por métodos de búsqueda tradicionales como sería que solo incluyera una búsqueda por medio de filtros. Para nuestro proyecto se implementaron ambos, una búsqueda por medio de filtrado y también agregamos una barra de búsqueda, que se encuentra en la vista del catálogo y en la landing page del comprador, que recupera la información con Elasticsearch. Además, como se mencionó previamente los campos que se están utilizando para alimentar el motor de búsqueda, no solo son del auto en sí, sino de información general como datos de la agencia o el financiamiento, para que el usuario pueda buscar entre los autos no solo el modelo y color que le guste, sino también el que se ajuste a su presupuesto y agencia que prefiera y que lo pueda hacer a través de frases o palabras en "lenguaje natural".

## Alfredo Jeong Hyun Park:

El uso de Elasticsearch dentro de nuestra aplicación es para el rápido acceso y extracción de datos del catálogo, la razón para el uso de Elasticsearch se basa en las funcionalidades que se incluyen como la búsqueda por filtrado y especialmente la búsqueda a través de lenguaje regular. Por lo cuál la decisión de utilizar Elasticsearch brinda además un valor mayor ya que el usuario puede buscar su coche deseado sin saber el nombre del modelo y/o marca, así brindando una excelente experiencia de usuario debido a su búsqueda personalizada y flexible.

Elasticsearch se encuentra montado en el servidor de Google Cloud Platform y almacena datos del catálogo de autos, lo que abarca el modelo, marca, color, precio y año. Además almacena datos relevantes de la agencia en los cuáles se pueden encontrar los autos seleccionados para que el usuario tenga una mayor facilidad para la adquisición de su auto.

#### Andrea Serrano Diego:

Elasticsearch mediante el pre entrenamiento de los datos de los vehículos como el modelo, marca, color, precio, año, etc. Se integra la búsqueda y los filtros dentro del catálogo en nuestra aplicación. La búsqueda y los filtros dentro del catálogo es una parte importante de la experiencia del comprador porque a través de ella el usuario puede acceder y acotar los diferentes vehículos registrados en la plataforma con el objetivo de encontrar un automóvil que realmente se ajuste a sus gustos y necesidades. Esta oportunidad es lo que hace que la plataforma tenga valor. Un ejemplo de lo que un usuario puede buscar se encuentra en la página principal con un conjunto de instrucciones dinámicas. Esto invita visualmente al usuario a una de las características primordiales de la aplicación. El despliegue está siendo

alojado en Google Cloud Platform, sin embargo, la gestión y operación sigue siendo desde el servicio de Elastic Cloud.

## Iwalani Amador Piaga:

Basado en los requerimientos del cliente, se implementó el uso de Elasticsearch principalmente para ofrecer al usuario una interacción de búsqueda más "natural" durante el proceso de compra de un vehículo y la búsqueda dentro del catálogo. Para ello cada vehículo dentro del catálogo está clasificado (para obtener una busqueda por filtros) y además para alimentar la búsqueda se agregan descripciones del vehículo y situaciones aplicables que pueden relacionarse al vehículo en cuestión, por ejemplo, "auto de ciudad en morelos", "carro familiar de 4 cilindros", etc, sin la necesidad de especificar un modelo o tipo de vehículo. Esta implementación aporta más valor comercial y empresarial, por mejorar el servicio ofrecido en la plataforma, atendiendo a la necesidad del cliente/usuario cuando no reconoce todas las características técnicas del vehículo que busca. Por otro lado, sin ser menos importante, los filtros se conforman de elecciones de acuerdo a una clasificación de palabras; marcas, modelos, años, precios, colores, etc.

Visualmente para el usuario, dentro de la aplicación, el uso de Elasticsearch se ve en la barra de búsqueda (dentro del catálogo y la página principal de la plataforma), que recibe los inputs de las búsquedas para reflejar los resultados filtrados de acuerdo a las coincidencias. De momento en Elasticsearch se realiza una alimentación de información semi-automatizada, pero definitivamente a futuro como proyecto, la integración y alimentación (ahora si) completamente automatizada, puede crear una búsqueda natural más acertada e incluso encontrando similitudes, por ejemplo en otro idioma, pensando en una comercialización expansiva de la plataforma.

#### Emiliano Cabrera Ruiz:

El propósito principal tras la utilización de Elasticsearch para la funcionalidad de búsqueda es la posibilidad de facilitar la búsqueda para el usuario al permitirle usar lenguaje natural. Existen palabras que no son tan fáciles o eficientes de implementar como filtros a base de elecciones binarias, cómo lo pueden ser los modelos y colores de vehículos. Elasticsearch permite que el usuario describa sus necesidades de la misma manera que lo haría al platicar con otra persona, dándole más libertad al expresarse y más opciones al buscar. Otra ventaja que tiene es que al ampliar el *pool* de campos o palabras disponibles para utilizar como condición se pueden agregar más formas de encontrar entradas, por ejemplo partes vehiculares, planes de financiamiento, o incluso descripciones únicas para cada agencia. En conclusión, Elasticsearch le da al usuario una experiencia más natural y amigable al navegar por la plataforma, y al mismo tiempo vuelve la búsqueda más flexible y adaptable.

# Andrew Dunkerley Vera:

La distribución en Elasticsearch se realizó considerando los campos más importantes para describir los automóviles de manera efectiva. Estos campos clave incluyen la marca, el modelo, el año y el precio, que permiten a los compradores filtrar y seleccionar vehículos según sus preferencias y presupuesto. Otros aspectos relevantes son la disponibilidad para pruebas, el tipo de combustible, el rendimiento, la capacidad de pasajeros y la transmisión. Además, los datos de la agencia, como la ubicación y el nombre, son fundamentales para brindar a los compradores información relevante sobre la accesibilidad y proximidad de la

agencia. En resumen, esta distribución en Elasticsearch facilita la búsqueda y selección de automóviles, ofreciendo a los usuarios una experiencia más personalizada y adecuada a sus necesidades.

Adicionalmente, el uso de Elasticsearch en este proyecto proporciona una búsqueda y consulta eficiente, permitiendo a los usuarios encontrar rápidamente los automóviles deseados gracias a la capacidad de indexación y la velocidad de recuperación de datos que ofrece esta plataforma.

#### **Tonatiuh Reyes:**

Se decidió implementar ElasticSearch ya que permite hacer uso del lenguaje natural para la búsqueda en la base de datos. Esta búsqueda la usamos en el catálogo de autos ya que es un elemento que contiene una extensa cantidad de datos, información e incluso una gran variedad de palabras. Con el uso del lenguaje natural, la búsqueda dentro de este bloque de información se facilita ya que permite el uso de términos sencillos que se usan en nuestra comunicación con el entorno día a día. Esto a su vez permite al usuario tener una experiencia con la plataforma más cómoda ya que no tiene que remitirse a aprender el sistema de búsqueda que se pueda implementar sino que simplemente puede usar su propio habla.

#### Andreina Sananez:

Como se ha hecho notar a lo largo de la realización del presente proyecto, la funcionalidad de búsqueda por medio de lenguaje natural del catálogo de autos de la plataforma, es un elemento crucial para el cliente el cual le da una gran importancia al hecho de que los usuarios tengan la mejor experiencia posible al momento de buscar y poder encontrar su auto deseado con una gran facilidad. Ahora bien, para implementar lo anterior se decidió utilizar Elasticsearch la cual es una tecnología de búsqueda y análisis de texto capaz de manejar considerables cantidades de datos en tiempo real y ofrecer búsquedas rápidas y precisas. Particularmente, se utilizó un único índice para almacenar el catálogo de autos en la plataforma cuyo despliegue está alojado en la nube de Google Cloud Platform (GCP) el cual es reconocido por su escalabilidad, rendimiento y seguridad, lo que significa que podemos confiar en que nuestra solución de búsqueda estará disponible y funcionará de manera eficiente incluso en momentos de alta demanda.

De manera más específica, en cuanto a la implementación se almacenaron entidades de autos con sus características intrínsecas como lo son marca, color, modelo y año, así como otra información relevante en cuanto su contexto como la dirección de la agencia, detalles de plazos y financiamiento. Asimismo, como parte de las áreas más importantes de este avance se encuentra la automatización de la generación de las descripciones de los autos para poder alimentar el sistema de recuperación de Elasticsearch por medio de lenguaje natural, esto ya que es gracias a dichas descripciones que describen al auto en lenguaje coloquial que el motor de Elasticsearch logra encontrar coincidencias con el texto ingresado referente a la búsqueda. Es por ello, que el uso de Chat-GPT para la generación de las mismas fue de gran ayuda para obtener descripciones extensas y completas que mejoraran la precisión de las búsquedas.

# Grupo 501 - Documento de base de datos

En resumen, sin lugar a duda Elasticsearch, combinado con GCP y nuestra implementación estratégica del mismo, es un componente vital en la entrega de una experiencia de compra de autos en línea completa y efectiva. Esto debido a que hace posible la realización de búsquedas rápidas y precisas, que mejoran la experiencia del usuario y aumentan la eficiencia operativa de la plataforma.