

UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA



PROYECTO INTEGRADOR

DOCENTE:

ING. DANILO JARAMILLO

POR:

ALBERTO HERRERA NICÓLAS SALINAS

JOSUÉ XAVIER PARDO CHAMBA

ANTONELLA DEL CISNE PARRA CARRIÓN

FECHA: 22/07/2025

ASIGNATURA: PROGRAMACIÓN AVANZADA

Contenido

Introducción	3
Repositorio GitHub:.....	3
Datos Base:.....	3
Datos Complementarios.....	6
Registro de Datos Administrativos	6
Datos de ciclo académico, id provincias, id cantones e id parroquias:	6
Registro Administrativo del año lectivo (para obtener una descripción más detallada de los datos)	6
Diseño lógico relacional	7
Diccionario de datos.....	7
Normalización:	9
Tabla: MINEDUC_RegistrosAdministrativos	9
Tabla: RegistroAdministrativoHistorico	9
Tabla: INEVAL_SerEstudiante	9
Script SQL	9
Herramientas utilizadas	10
Sistema operativo Ubuntu	10
¿Qué es Zeppelin?	10
Configuración del entorno	11
Instalación de la Máquina Virtual	11
Permisos de usuario de Ubuntu	13
Instalación de mysql	13
Instalación de Zeppelin.....	14
Instalación de java.....	14
Instalación de Python	15
Instalación de Spark	15
Instalación de Scala	17
Interprete Zeppelin configuración de MySQL, Spark, Python.....	18
Configuración de MySQL en Interprete de Zeppelin	18
Configuración de Spark en Interprete de Zeppelin.....	19
Configuración de Python en Interprete de Zeppelin	19
Análisis a realizar	19
Visualización de consultas.....	19
Consulta 1.....	19

Consulta 2.....	20
Consulta 3.....	21
Consulta 4.....	21
Consulta 5.....	21
Consulta 6.....	22
Conclusiones	22
Bibliografía	23

Introducción

En el ámbito de la investigación educativa, las estadísticas desempeñan un papel esencial para comprender y analizar el funcionamiento del sistema escolar de un país. Los registros administrativos del Ministerio de Educación constituyen una fuente primaria de información confiable y actualizada sobre las instituciones, el personal docente, los estudiantes y la infraestructura educativa. En este sentido, la gestión y análisis de estadísticas permanentes permiten identificar tendencias, evaluar políticas públicas y formular estrategias orientadas a mejorar la calidad del sistema educativo. El presente análisis se basa en los datos recopilados durante los años 2022 y 2023, con el objetivo de ofrecer una visión integral del estado de la educación en Ecuador, a partir de variables clave como modalidad, sostenimiento, área geográfica, jornada, entre otras.

Repositorio GitHub:

Repositorio: <https://github.com/CisneParra/Programacion-Avanzada.git>

Datos Base:

Los datos base contienen información detallada proveniente de registros administrativos y encuestas aplicadas en el ámbito educativo y socioeconómico, correspondientes al periodo 2022-2023. Estos datos permiten analizar de manera integral las condiciones estructurales de las instituciones educativas del país, así como los resultados de aprendizaje y características demográficas de los estudiantes. Además, contribuyen a la generación y gestión de estadísticas permanentes relevantes para la formulación de políticas públicas y la evaluación del sistema educativo nacional.

En particular, los registros contienen datos sobre la oferta educativa por establecimiento, el desempeño estudiantil según nivel educativo y quintil socioeconómico, la infraestructura institucional, y las condiciones del entorno familiar. Esta información, obtenida por entidades como el Ministerio de Educación y el INEVAL, proporciona una base sólida para el análisis técnico y territorial de la educación en Ecuador.

A continuación, se presenta una descripción de las principales variables incluidas en los datos:

Año Lectivo	Periodo escolar al que pertenece el registro.
AMIE	Código único de identificación de la institución educativa.
Nombre Institución	Denominación oficial del establecimiento educativo.
Zona / Provincia / Cantón / Parroquia	División político-administrativa en la que se ubica la institución.
Escolarización / Tipo Educación / Nivel Educación	Características pedagógicas del servicio educativo ofertado.
Sostenimiento	Fuente de financiamiento del establecimiento (fiscal, fiscomisional, municipal o particular).
Área	Indica si la institución se encuentra en zona urbana o rural.
Régimen Escolar:	Calendario académico aplicado (Costa-Galápagos o Sierra-Amazonía).
Modalidad / Jornada	Tipo de enseñanza (presencial, virtual, etc.) y horario (matutino, vespertino, etc.).
Tenencia del inmueble / Forma Acceso	Información sobre la propiedad del local educativo y su accesibilidad.
Docentes y Administrativos	Cantidad de personal femenino y masculino que labora en la institución.
Estudiantes por nivel y género	Número de estudiantes por año de estudio, diferenciados por sexo, incluyendo indicadores de promoción, abandono, y no actualización.
ciclo / grado	Nivel educativo y año escolar del estudiante encuestado.
estado_eval	Estado del proceso de evaluación del estudiante.
codigo / amie	Identificadores únicos del estudiante y de la institución educativa.
tp_sexo	Sexo del estudiante (femenino o masculino).
tp_area	Área de localización del plantel (urbana o rural).

A1																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo																		
Año Lectivo</																		

2_MINEDUC_RegistrosAdministrativos_2022-2023FILA															Guardado en Este PC															Buscar														
Archivo Inicio Insertar Dibujar Disposición de página Fórmulas Datos Revisar Vista Automatizar Ayuda															Comentarios Compart																													
Pegar Fuente Alineación Número Estilos Celdas															Complementos Analizar datos Ghostwriter Autopilot																													
Aptos Narrow 11 A A' General Dar formato como tabla Estilos de celda															Insertar Eliminar Formato Ordenar y filtrar Buscar y seleccionar Complementos																													
Año Lectivo															Complementos																													
D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U																											
1	Provincia	Cod_Provincia	Cantón	Cod_Cantón	Parroquia	Cod_Parroquia	Escolarización	Tipo Educación	Nivel Educativo	Sostenimiento	Área	Régimen	Esco	Jurisdicción	Modalidad	Jornada	Tenencia de la Forma																											
2	6	AZUAY	1 CUENCA	101	BAÑOS	10151	Escolarizada	Educación Esj EGB y Bachille	Fiscal	Rural	Sierra	Intercultural	Presencial	Matutina	Propio	Terrestre	Accesibilidad																											
3	1	CARCHI	4 ESPEJO	40353	SAN ISIDRO	40353	Escolarizada	Educación Esj EGB y Bachille	Fiscal	Rural	Sierra	Intercultural	Presencial	Matutina	Propio	Terrestre																												
4	4	SANTO DOMINGO	23 LA CONCORD	2302	LA CONCORD	230250	Escolarizada	Educación Esj Inicial, Educación	Fiscal	Rural	Costa	Intercultural	Presencial	Matutina	Propio	Terrestre																												
5	1	ESMERALDAS	8 ATACAMES	806	TONSUPA	80654	Escolarizada	Educación Esj Inicial, Educación	Fiscal	Rural	Costa	Intercultural	Presencial	Matutina	Propio	Terrestre																												
6	1	ESMERALDAS	8 QUININDÉ	804	MALIMPIA	80453	Escolarizada	Educación Esj EGB y Bachille	Fiscal	Rural	Costa	Intercultural	Presencial	Matutina	Propio	Terrestre																												
7	8	GUAYAS	9 GUAYAQUIL	901	POSORIA	90156	Escolarizada	Educación Esj Inicial, Educación	Fiscal	Rural	Costa	Intercultural	Presencial	Matutina	Propio	Terrestre																												
8	1	IMBABURA	10 IBARRA	1001	SAN ANTONIO	100157	Escolarizada	Educación Esj Inicial, Educación	Fiscal	Rural	Sierra	Intercultural	Presencial	Matutina	Propio	Terrestre																												
9	7	LOJA	11 SARAGURO	1111	LUZHUAPA	111153	Escolarizada	Educación Esj Educación Básica	Fiscal	Rural	Costa	Intercultural	Presencial	Matutina	Propio	Terrestre																												
10	5	LOS RÍOS	12 PUEBLOVIEJO	1204	SAN JUAN	120452	Escolarizada	Educación Esj Educación Básica	Fiscal	Rural	Costa	Intercultural	Presencial	Matutina	Propio	Terrestre																												
11	2	PICHINCHA	17 MEJÍA	1703	ALOAG	170351	Escolarizada	Educación Esj EGB y Bachille	Fiscal	Rural	Sierra	Intercultural	Presencial	Matutina	Propio	Terrestre																												
12	1	SUCUMBIOS	21 GONZALO PÍZ	2102	GONZALO PÍZ	210252	Escolarizada	Educación Esj Inicial y EGB	Fiscal	Rural	Sierra	Intercultural	Presencial	Matutina	Propio	Terrestre																												
13	2	ORELLANA	22 ORELLANA	2201	SAN LUIS DE A	220161	Escolarizada	Educación Esj Inicial, Educación	Fiscal	Rural	Sierra	Intercultural	Presencial	Matutina	Propio	Terrestre																												
14	6	AZUAY	1 CUENCA	101	CHAUCHA	10153	Escolarizada	Educación On Educación Básica	Fiscal	Rural	Costa	Intercultural	Presencial	Matutina	Propio	Terrestre																												
15	6	AZUAY	1 CUENCA	101	QUINGEO	10161	Escolarizada	Educación On Inicial y EGB	Fiscal	Rural	Sierra	Intercultural B	Presencial	Matutina	Propio	Terrestre																												
16	6	AZUAY	1 CUENCA	101	QUINGEO	10161	Escolarizada	Educación On Inicial y EGB	Fiscal	Rural	Sierra	Intercultural B	Presencial	Matutina	Propio	Terrestre																												
17	6	AZUAY	1 CUENCA	101	SANTA ANA	10164	Escolarizada	Educación On Inicial y EGB	Fiscal	Rural	Sierra	Intercultural B	Presencial	Matutina	Propio	Terrestre																												
18	6	AZUAY	1 CUENCA	101	TURI	10169	Escolarizada	Educación On Inicial y EGB	Fiscal	Rural	Sierra	Intercultural B	Presencial	Matutina	Propio	Terrestre																												
19	6	AZUAY	1 CUENCA	101	TURI	10169	Escolarizada	Educación On Inicial y EGB	Fiscal	Rural	Sierra	Intercultural B	Presencial	Matutina	Propio	Terrestre																												
20	6	AZUAY	1 CUENCA	101	TURI	10169	Escolarizada	Educación On Inicial y EGB	Fiscal	Rural	Sierra	Intercultural B	Presencial	Matutina	Propio	Terrestre																												
21	6	AZUAY	1 CUENCA	101	SANTA ANA	10164	Escolarizada	Educación On Inicial y EGB	Fiscal	Rural	Sierra	Intercultural B	Presencial	Matutina	Propio	Terrestre																												
22	6	AZUAY	1 GUALACEO	103	SAN JUAN	10357	Escolarizada	Educación On Inicial y EGB	Fiscal	Rural	Sierra	Intercultural B	Presencial	Matutina	Propio	Terrestre																												
23	6	AZUAY	1 GUALACEO	103	SAN JUAN	10357	Escolarizada	Educación On Inicial y EGB	Fiscal	Rural	Sierra	Intercultural B	Presencial	Matutina	Propio	Terrestre																												
24	6	AZUAY	1 GUALACEO	103	ZHIDMAD	10358	Escolarizada	Educación On Educación Básica	Fiscal	Rural	Sierra	Intercultural B	Presencial	Matutina	Propio	Terrestre																												
25	6	AZUAY	1 NABÓN	104	EL PROGRESO	10452	Escolarizada	Educación On Educación Básica	Fiscal	Rural	Costa	Intercultural	Presencial	Matutina	Propio	Terrestre																												
26	6	AZUAY	1 SIGSIG	109	LUDO	10954	Escolarizada	Educación On Inicial y EGB	Fiscal	Rural	Sierra	Intercultural B	Presencial	Matutina	Propio	Terrestre																												
27	6	AZUAY	1 OÑA	110	SUSUDEL	11051	Escolarizada	Educación On Inicial y EGB	Fiscal	Rural	Sierra	Intercultural B	Presencial	Matutina	Propio	Terrestre																												
28	6	AZUAY	1 OÑA	110	SUSUDEL	11051	Escolarizada	Educación On Inicial y EGB	Fiscal	Rural	Sierra	Intercultural B	Presencial	Matutina	Propio	Terrestre																												

Estos datos, junto con los análisis que se realicen, serán almacenados en un repositorio de GitHub para facilitar su gestión, revisión y colaboración entre investigadores y especialistas interesados en el tema. El repositorio permitirá mantener un registro histórico de los cambios realizados, fomentando así la transparencia y la reproducción de los resultados obtenidos.

Datos Complementarios

Registro de Datos Administrativos

Datos obtenidos de <https://educacion.gob.ec/base-de-datos> (se extrajo de los años 2022-2023)

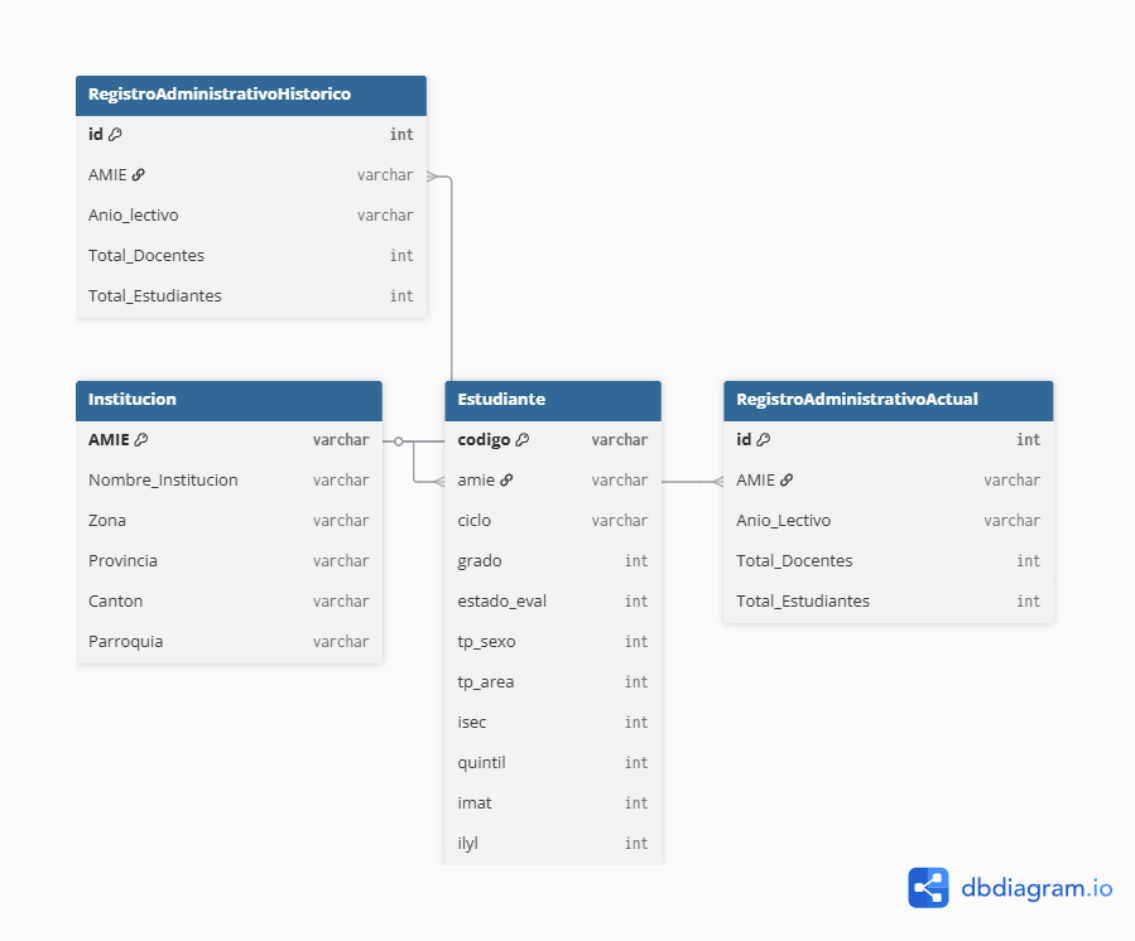
Datos de ciclo académico, id provincias, id cantones e id parroquias:

Datos obtenidos de <https://www.datosabiertos.gob.ec/dataset/ser-estudiante/resource/367ed0a9-f912-4a77-82c1-248432b2cfb3>

Registro Administrativo del año lectivo (para obtener una descripción más detallada de los datos)

Datos obtenidos de <https://datosabiertos.gob.ec/dataset/registro-de-matricula-mineduc/resource/59db7bb5-eb6b-4a7a-a06a-d1f93ac7506a>

Diseño lógico relacional



Diccionario de datos

Tabla	Columna	Tipo	Descripción
MINEDUC_RegistrosAdministrativos	AMIE	varchar	Código único de la institución educativa (clave primaria)
MINEDUC_RegistrosAdministrativos	anio_lectivo	varchar	Período académico (ej: "2022-2023 Fin")
MINEDUC_RegistrosAdministrativos	Provincia	varchar	Provincia de la institución
MINEDUC_RegistrosAdministrativos	Canton	varchar	Cantón de la institución
MINEDUC_RegistrosAdministrativos	Total_Estudiantes	int	Número total de estudiantes
MINEDUC_RegistrosAdministrativos	Total_Docentes	int	Número total de docentes

RegistroAdministrativoHistorico	AMIE	varchar	Código único de la institución educativa (clave primaria)
RegistroAdministrativoHistorico	anio_lectivo	varchar	Período académico histórico
RegistroAdministrativoHistorico	Provincia	varchar	Provincia de la institución
RegistroAdministrativoHistorico	Canton	varchar	Cantón de la institución
RegistroAdministrativoHistorico	Total_Estudiantes	int	Número total de estudiantes históricos
RegistroAdministrativoHistorico	Total_Docentes	int	Número total de docentes históricos
INEVAL_SerEstudiante	id	int	Identificador único (clave primaria, autoincremental)
INEVAL_SerEstudiante	AMIE	varchar	Código de la institución educativa (clave foránea)
INEVAL_SerEstudiante	ciclo	varchar	Ciclo académico evaluado (ej: "2022-2023")
INEVAL_SerEstudiante	grado	int	Grado evaluado
INEVAL_SerEstudiante	estado_eval	int	Estado de la evaluación
INEVAL_SerEstudiante	isec	int	Indicador de desempeño: ciencias sociales
INEVAL_SerEstudiante	imat	int	Indicador de desempeño: matemáticas
INEVAL_SerEstudiante	ilyl	int	Indicador de desempeño: lenguaje y literatura
INEVAL_SerEstudiante	icn	int	Indicador de desempeño: ciencias naturales
INEVAL_SerEstudiante	ies	int	Indicador de desempeño: estudios sociales

Normalización:

Tabla: MINEDUC_RegistrosAdministrativos

AMIE	anio_lectivo	Provincia	Canton	Total_Estudiantes	Total_Docentes
01H01829	2022-2023 Fin	AZUAY	CUENCA	120	10

Tabla: RegistroAdministrativoHistorico

AMIE	anio_lectivo	Provincia	Canton	Total_Estudiantes	Total_Docentes
01B00002	2009-2010 Inicio	AZUAY	CUENCA	1473	96

Tabla: INEVAL_SerEstudiante

id	AMIE	ciclo	grado	estado_eval	isec	imat	ilyl	icn	ies
1	07H00020	2022- 2023	3	2	4	18	17	17	17

En este modelo se integran datos provenientes de tres archivos CSV relacionados con registros educativos en Ecuador. Se han creado tablas adicionales para mantener la integridad y coherencia de la información, así como relaciones entre ellas a través de claves primarias y foráneas.

Script SQL

Para este proyecto, el proceso de integración y análisis de los datos educativos se realizó en varias etapas utilizando principalmente Zeppelin y MySQL. El flujo de trabajo fue el siguiente:

Primero, se cargaron los tres archivos CSV principales al entorno de Zeppelin:

2_MINEDUC_RegistrosAdministrativos_2022-2023Fin.csv

registro-administrativo-historico_2009-2024-inicio.csv

ineval_serestudiante2022_2023_2023diciembre.csv

En Zeppelin, usando Spark con Scala, se realizaron las siguientes tareas:

Carga de los datos: se importaron los archivos CSV usando la opción para leer encabezados y para inferir tipos de datos.

Limpieza de datos: se eliminaron o transformaron las columnas innecesarias, se renombraron campos para uniformizar nombres entre los distintos archivos y se convirtieron los valores vacíos o inconsistentes a nulos donde fue necesario.

Unión de tablas: los datasets se integraron usando campos comunes, principalmente el código AMIE de las instituciones educativas, generando un archivo consolidado con

las columnas clave: AMIE, anio_lectivo, Provincia, Canton, Total_Estudiantes, Total_Docentes.

Una vez generada la tabla final y exportada como CSV limpio, se continuó el proceso en MySQL:

Creación de la tabla en MySQL: se creó la tabla union_educacion con las columnas clave y sus tipos de datos correspondientes.

Carga masiva de datos: se utilizó el comando LOAD DATA INFILE para importar el archivo consolidado a la base de datos, realizando ajustes para manejar correctamente los valores nulos y los formatos de los datos.

Todo lo antes mencionado se realizo en el gitpod que sus imágenes se pasaron al github correspondiente link: <https://github.com/CisneParra/Programacion-Avanzada.git>

Herramientas utilizadas

Sistema operativo Ubuntu

Para nuestra máquina virtual utilizamos el ISO de Ubuntu 20.04.5 el cual instalaremos en la máquina virtual de Oracle

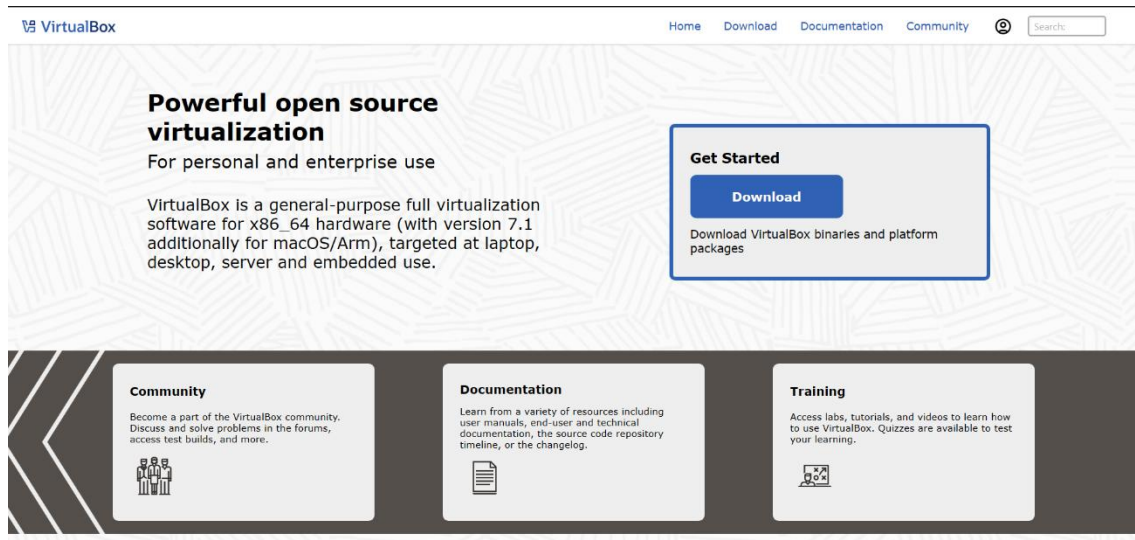
¿Qué es Zeppelin?

Zeppelin es un cuaderno web interactivo que permite a los usuarios el análisis de datos, visualización, consultas, entre otros, soportando múltiples lenguajes de programación como Python, Spark, SQL y más. Siendo estas herramientas necesarias para el proceso de análisis que vamos a realizar a continuación

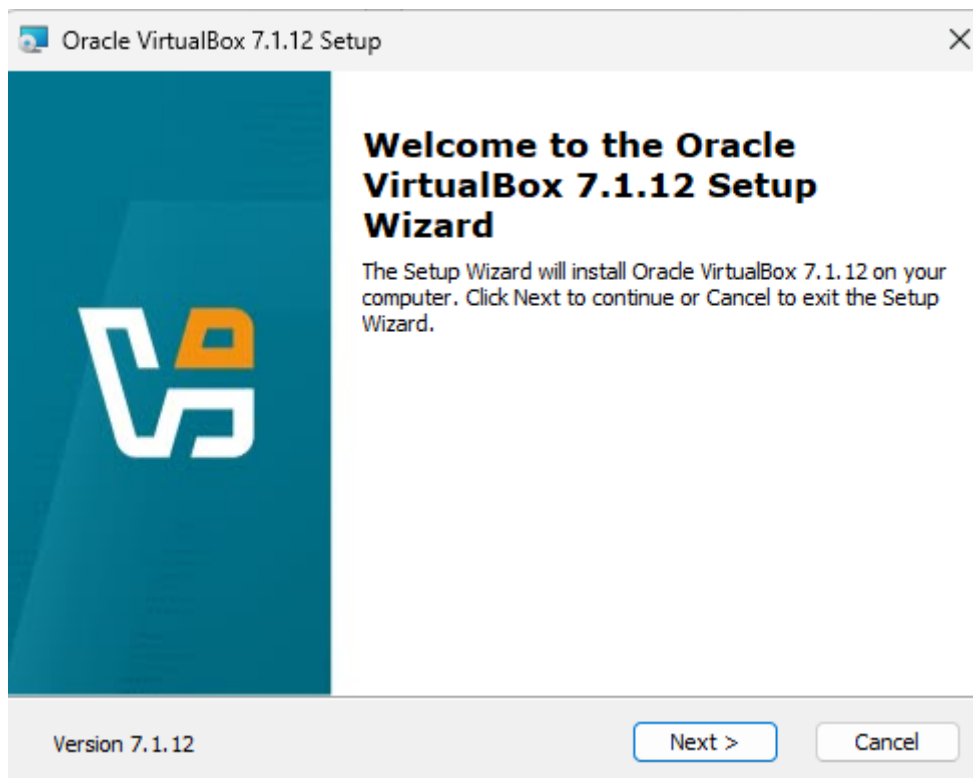
Configuración del entorno

Instalación de la Máquina Virtual

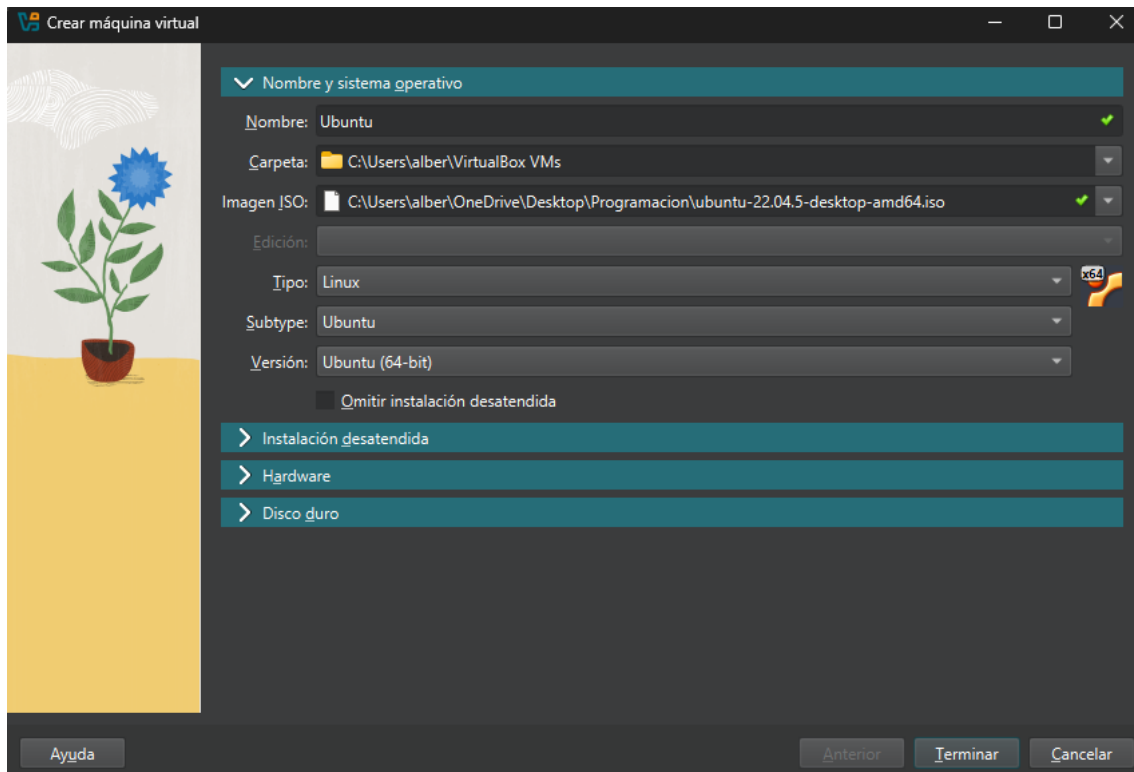
Primeramente, descargamos el instalador de <https://www.virtualbox.org>, seleccionando nuestro sistema operativo



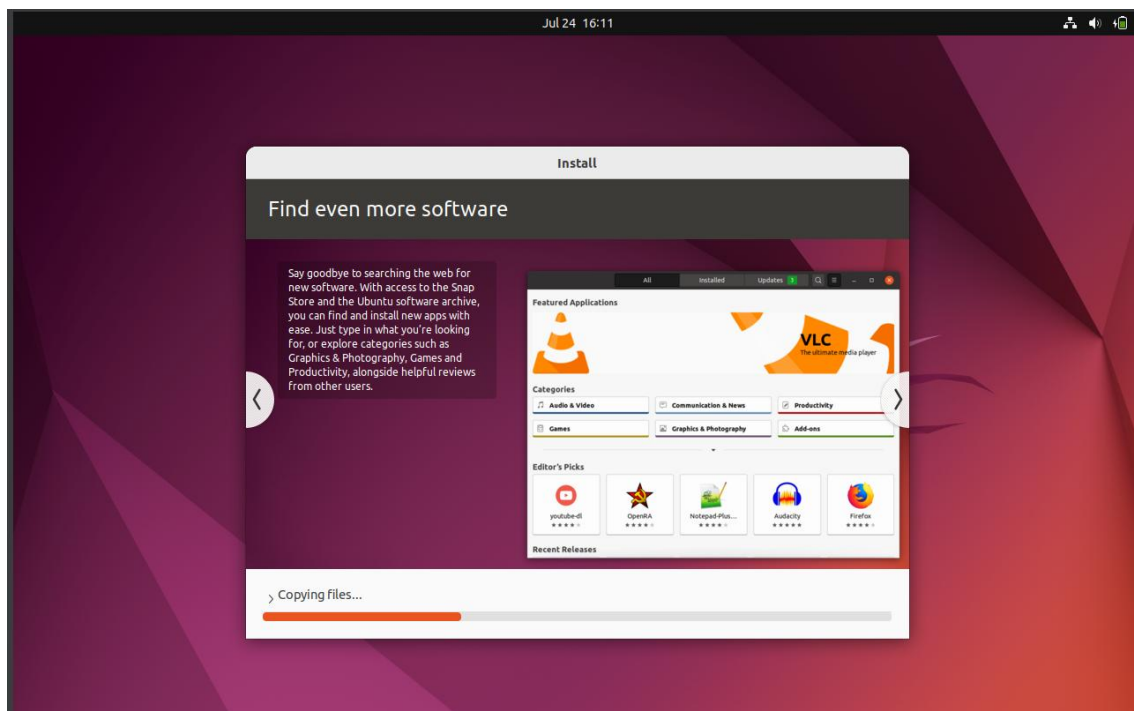
Realizamos la respectiva instalación de la maquina virtual



Una vez con el programa ya instalado se lo abre y se seleccionamos la opción de añadir, después se le pone un nombre a la máquina virtual, se selecciona el iso a utilizar (Se lo consigue en <https://ubuntu.com/download/desktop>) y se define una contraseña y le damos a Terminar



Una vez configurado se ejecutará la máquina virtual, y el proceso de instalación de Ubuntu iniciará



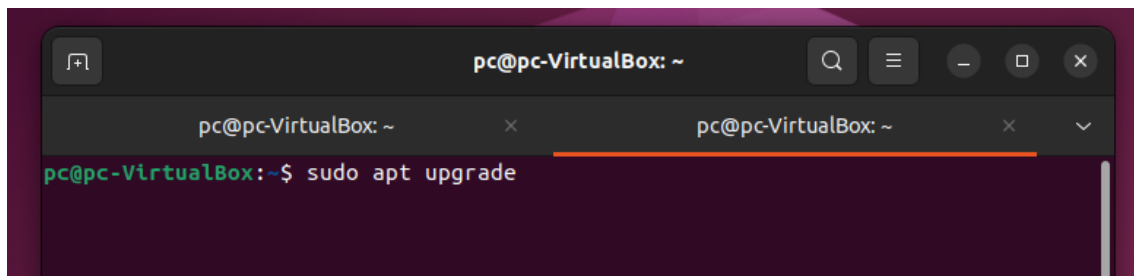
Permisos de usuario de Ubuntu

Guarda los cambios realizados en el archivo sudoers en nano. Para ello, presiona Ctrl + O, luego presiona Enter para confirmar el nombre del archivo y finalmente, presiona Ctrl + X para salir de nano.

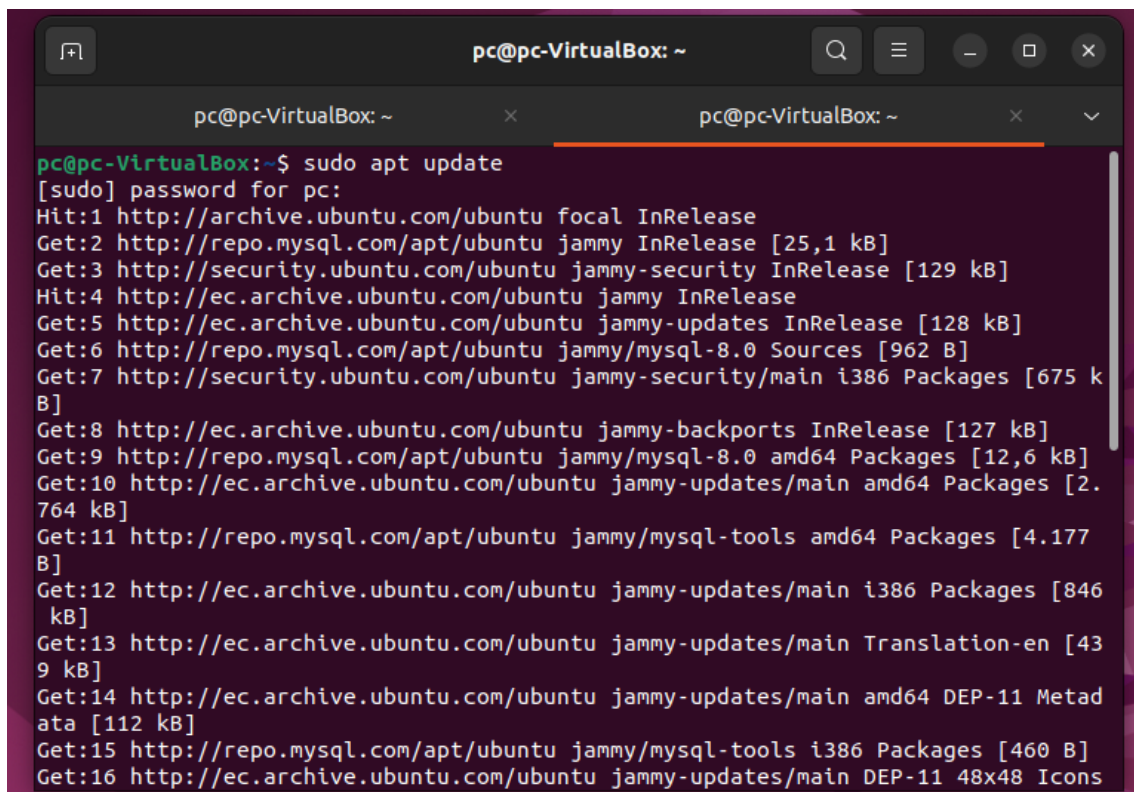
Una vez que hayas completado estos pasos, el usuario "nuevo_usuario" tendrá privilegios de superusuario y podrá usar el comando sudo para ejecutar comandos con privilegios elevados.

Instalación de mysql

Para la instalación de mySql primeramente tendremos que actualizar los paquetes para tener las versiones más actuales

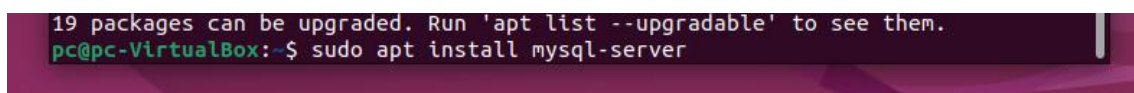


```
pc@pc-VirtualBox: ~  
pc@pc-VirtualBox:~$ sudo apt upgrade
```



```
pc@pc-VirtualBox:~$ sudo apt update  
[sudo] password for pc:  
Hit:1 http://archive.ubuntu.com/ubuntu focal InRelease  
Get:2 http://repo.mysql.com/apt/ubuntu jammy InRelease [25,1 kB]  
Get:3 http://security.ubuntu.com/ubuntu jammy-security InRelease [129 kB]  
Hit:4 http://ec.archive.ubuntu.com/ubuntu jammy InRelease  
Get:5 http://ec.archive.ubuntu.com/ubuntu jammy-updates InRelease [128 kB]  
Get:6 http://repo.mysql.com/apt/ubuntu jammy/mysql-8.0 Sources [962 B]  
Get:7 http://security.ubuntu.com/ubuntu jammy-security/main i386 Packages [675 kB]  
Get:8 http://ec.archive.ubuntu.com/ubuntu jammy-backports InRelease [127 kB]  
Get:9 http://repo.mysql.com/apt/ubuntu jammy/mysql-8.0 amd64 Packages [12,6 kB]  
Get:10 http://ec.archive.ubuntu.com/ubuntu jammy-updates/main amd64 Packages [2.764 kB]  
Get:11 http://repo.mysql.com/apt/ubuntu jammy/mysql-tools amd64 Packages [4.177 B]  
Get:12 http://ec.archive.ubuntu.com/ubuntu jammy-updates/main i386 Packages [846 kB]  
Get:13 http://ec.archive.ubuntu.com/ubuntu jammy-updates/main Translation-en [439 kB]  
Get:14 http://ec.archive.ubuntu.com/ubuntu jammy-updates/main amd64 DEP-11 Metadata [112 kB]  
Get:15 http://repo.mysql.com/apt/ubuntu jammy/mysql-tools i386 Packages [460 B]  
Get:16 http://ec.archive.ubuntu.com/ubuntu jammy-updates/main DEP-11 48x48 Icons
```

Luego ejecutamos lo siguiente y se descargara e instalara mysql



```
19 packages can be upgraded. Run 'apt list --upgradable' to see them.  
pc@pc-VirtualBox:~$ sudo apt install mysql-server
```

Se podrá comprobar la correcta instalación con el comando

```
pc@pc-VirtualBox:~$ mysql --version
mysql Ver 8.0.43 for Linux on x86_64 (MySQL Community Server - GPL)
```

Instalación de Zeppelin

De la pagina oficial de zeppelin (<https://www.apache.org/dyn/closer.cgi/zeppelin/zeppelin-0.12.0/zeppelin-0.12.0-bin-all.tgz>) Conseguimos el link para descargar desde la terminal con el comando wget

```
pc@pc-VirtualBox:~$ wget https://dlcdn.apache.org/zeppelin/zeppelin-0.12.0/zeppelin-0.12.0-bin-all.tgz
```

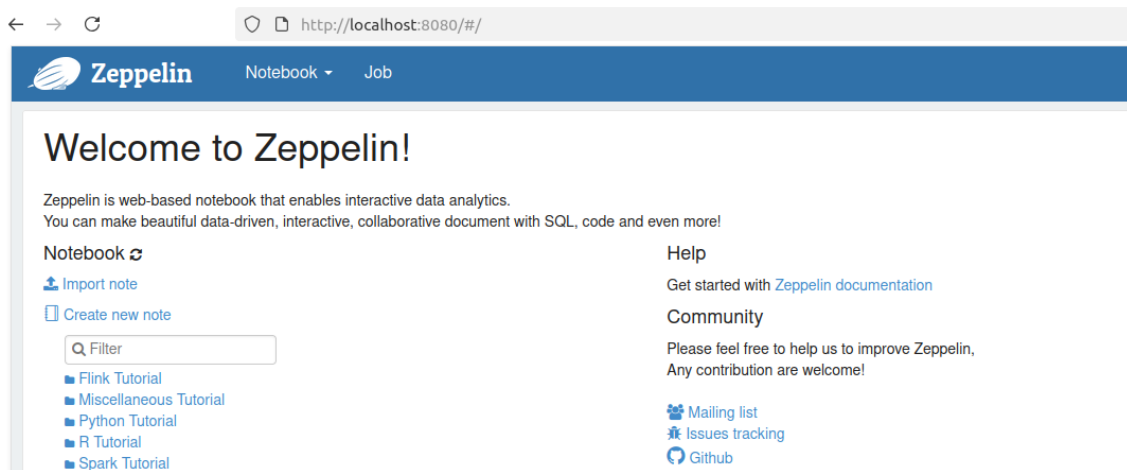
Y se empezara a descargar. Despues tendremos que descomprimirlo con el comando tar -xvzf y el nombre del archivo

```
pc@pc-VirtualBox:~$ tar -xvzf zeppelin-0.12.0-bin-all.tgz
```

Para ejecutar Zeppelin tendremos que dirigirnos a la carpeta bin del archivo que acabamos de descomprimir podemos hacerlo con cd zeppelin-0.12.0-bin-all y ejecutaremos

```
pc@pc-VirtualBox:~/zeppelin-0.12.0-bin-all$ bin/zeppelin-daemon.sh start
Please specify HADOOP_CONF_DIR if USE_HADOOP is true
Log dir doesn't exist, create /home/pc/zeppelin-0.12.0-bin-all/logs
Pid dir doesn't exist, create /home/pc/zeppelin-0.12.0-bin-all/run
Zeppelin start [ OK ]
```

Podemos encontrarlo alojado en <http://localhost:8080/>



Instalación de java

Para la instalación de java primeramente se actualizarán los paquetes como se hizo previamente y se ejecutara el siguiente

```
pc@pc-VirtualBox:~$ sudo apt install default-jdk
```

y se comprobara con el comando java -version su correcta instalación

```
pc@pc-VirtualBox:~$ java -version
openjdk version "17.0.15" 2025-04-15
OpenJDK Runtime Environment (build 17.0.15+6-Ubuntu-0ubuntu122.04)
OpenJDK 64-Bit Server VM (build 17.0.15+6-Ubuntu-0ubuntu122.04, mixed mode, sharing)
```

Instalación de Python

Para la instalación de python primeramente se actualizarán los paquetes como se hizo previamente y se ejecutara el siguiente código

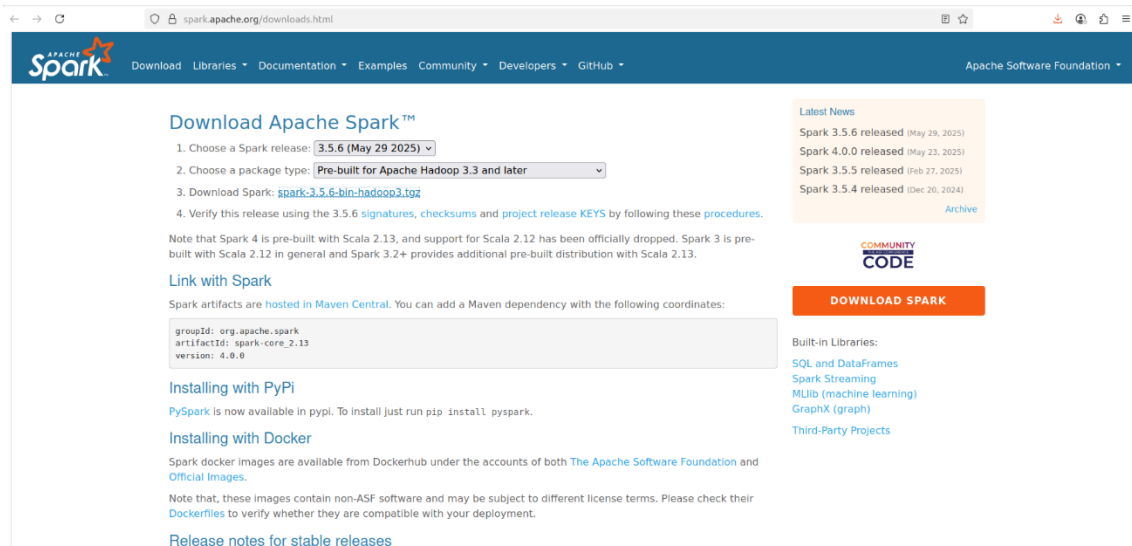
```
pc@pc-VirtualBox:~$ sudo apt install python3 python3-pip -y
Reading package lists... Done
```

y para comprobar su correcta instalación se ejecuta el código `python3 --version`

```
pc@pc-VirtualBox:~$ python3 --version
Python 3.10.12
pc@pc-VirtualBox:~$
```

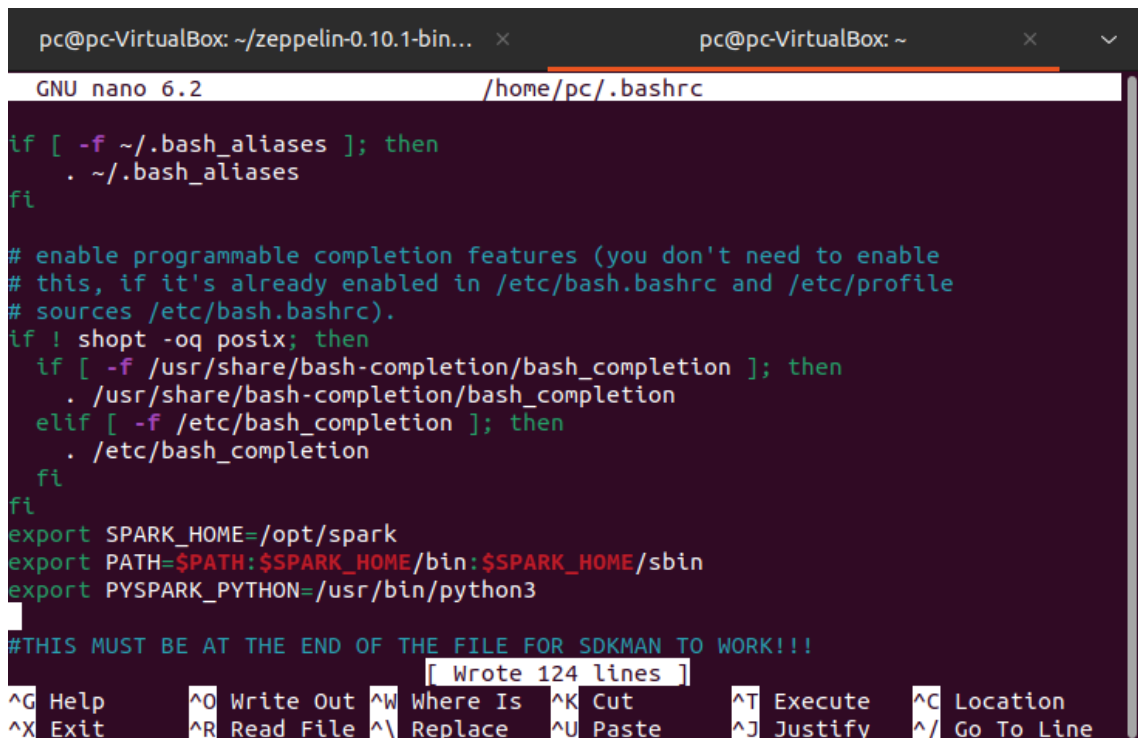
Instalación de Spark

Para la instalación de spark en la página oficial (<https://spark.apache.org/downloads.html>) obtendremos el link, en el cual seleccionaremos la versión a instalar (Teniendo en cuenta las versiones compatibles con nuestro Zeppelin)

The screenshot shows the Apache Spark download page. The header includes the Spark logo and navigation links: Download, Libraries, Documentation, Examples, Community, Developers, and GitHub. The main content area is titled "Download Apache Spark™" and contains a step-by-step guide: 1. Choose a Spark release (3.5.6 (May 29 2025)), 2. Choose a package type (Pre-built for Apache Hadoop 3.3 and later), 3. Download Spark (spark-3.5.6-bin-hadoop3.tgz), and 4. Verify this release using the 3.5.6 signatures, checksums and project release KEYS by following these procedures. Below this, there is a note about Scala versions and a section for linking with Spark, showing Maven coordinates. There are also sections for installing with PyPi and Docker. On the right side, there is a "Latest News" section with release dates for Spark 3.5.6, 4.0.0, 3.5.5, and 3.5.4, and a "COMMUNITY CODE" section with a "DOWNLOAD SPARK" button and links to built-in libraries and third-party projects.

Y dando click en el hipervínculo nos llevara a la pagina para copiar el enlace web

Utilizamos el comando nano ~/.bashrc para especificar las variables del entorno



```
pc@pc-VirtualBox: ~/zeppelin-0.10.1-bin... x pc@pc-VirtualBox: ~ x v
GNU nano 6.2 /home/pc/.bashrc

if [ -f ~/.bash_aliases ]; then
    . ~/.bash_aliases
fi

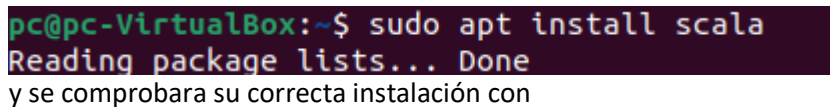
# enable programmable completion features (you don't need to enable
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
# sources /etc/bash.bashrc).
if ! shopt -oq posix; then
    if [ -f /usr/share/bash-completion/bash_completion ]; then
        . /usr/share/bash-completion/bash_completion
    elif [ -f /etc/bash_completion ]; then
        . /etc/bash_completion
    fi
fi
export SPARK_HOME=/opt/spark
export PATH=$PATH:$SPARK_HOME/bin:$SPARK_HOME/sbin
export PYSPARK_PYTHON=/usr/bin/python3

#THIS MUST BE AT THE END OF THE FILE FOR SDKMAN TO WORK!!!
Wrote 124 lines
^G Help ^O Write Out ^W Where Is ^K Cut ^T Execute ^C Location
^X Exit ^R Read File ^\ Replace ^U Paste ^J Justify ^_ Go To Line
```

Guardamos los cambios con control + O y salimos con Control + X y recargamos los datos con source .bashrc

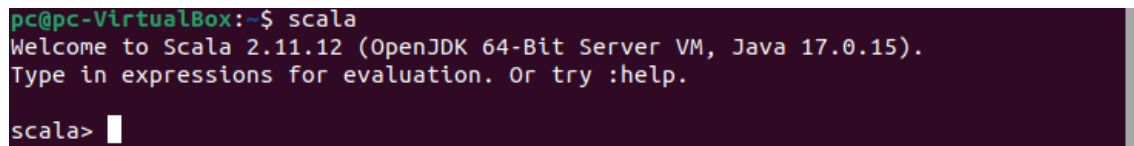
Instalación de Scala

Para la instalación de scala primeramente se actualizarán los paquetes como se hizo previamente y se ejecutara el siguiente código



```
pc@pc-VirtualBox:~$ sudo apt install scala
Reading package lists... Done
```

y se comprobara su correcta instalación con



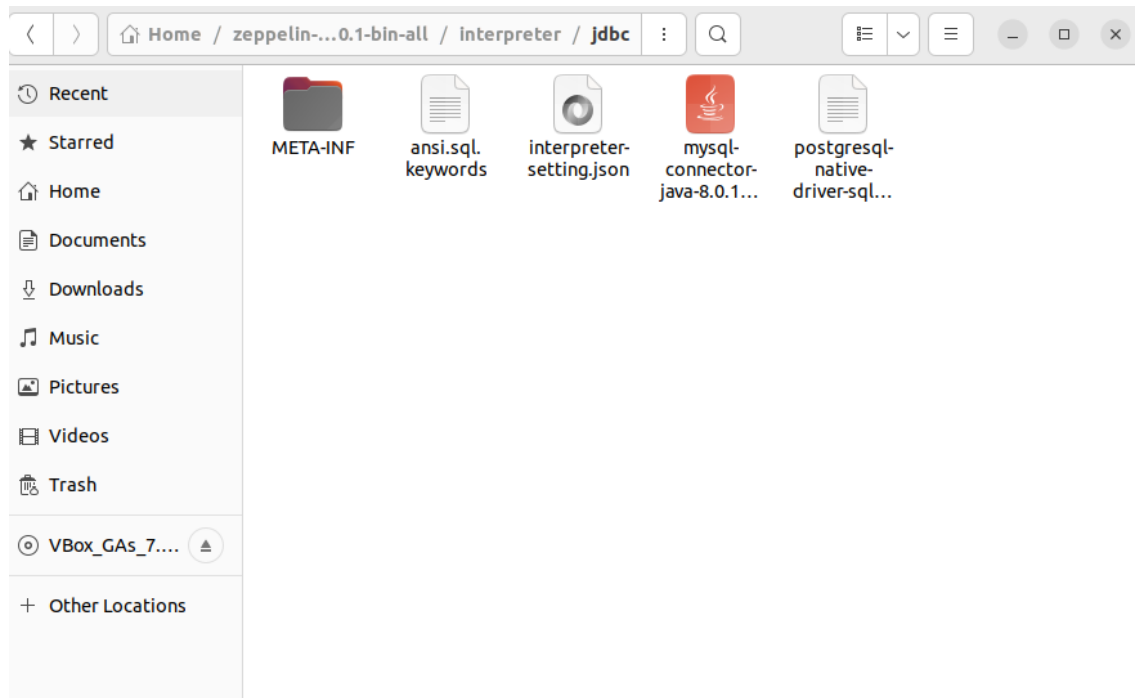
```
pc@pc-VirtualBox:~$ scala
Welcome to Scala 2.11.12 (OpenJDK 64-Bit Server VM, Java 17.0.15).
Type in expressions for evaluation. Or try :help.

scala> 
```

Interprete Zeppelin configuración de MySQL, Spark, Python

Configuración de MySQL en Interprete de Zeppelin

Para configurar correctamente se debe descargar el .jar de mysql-connector (Se puede encontrar en el siguiente link <https://repo1.maven.org/maven2/com/mysql/mysql-connector-j/>) a este jar se lo ubica en la carpeta de interpreter/jdbc en la carpeta descomprimida de zeppelin



y configuramos las propiedades en zeppelin como la url de la base de datos, el usuario y su respectiva contraseña además del controlador que se utilizara y su respectivo artefacto

Zeppelin Notebook - Job

jdbc %jdbc

Option

The interpreter will be instantiated locally in shared process

☐ Connect to existing process

☐ Set permission

Properties

Name	Value	Description
default.url	jdbc:mysql://localhost:3306/ec	The URL for JDBC.
default.user	root	The JDBC user name
default.password	***	The JDBC user password
default.driver	com.mysql.cj.jdbc.Driver	JDBC Driver Name
default.completer.ttlInSeconds	120	Time to live sql completer in seconds (-1 to update everytime, 0 to disable update)
default.completer.schemaFilters		Comma separated schema (schema - catalog - database) filters to get metadata for completion. Supports % symbol is equivalent to any set of characters. (ex. prod_v_%,public%,info)
default.precode		SQL which executes while opening connection
default.statementPrcode		Runs before each run of the paragraph, in the same connection
common.max_count	1000	Max number of SQL result to display.
zeppelin.jdbc.auth.type		If auth type is needed. Example: KERBEROS
zeppelin.jdbc.auth.kerberos.proxy.enable	true	When auth type is Kerberos, enable/disable Kerberos proxy with the login user to get the connection. Default value is true.
zeppelin.jdbc.concurrent.use	true	Use parallel scheduler
zeppelin.jdbc.concurrent.max_connection	10	Number of concurrent execution

default.completer.ttlInSeconds	120	Time to live sql completer in seconds (-1 to update everytime, 0 to disable update)
default.completer.schemaFilters		Comma separated schema (schema = catalog = database) filters to get metadata for completion s. Supports % symbol is equivalent to any set of characters. (ex. prod_v_%,public%,info)
default.precode		SQL which executes while opening connection
default.statementPrecode		Runs before each run of the paragraph, in the same connection
common.max_count	1000	Max number of SQL result to display.
zeppelin.jdbc.auth.type		If auth type is needed, Example: KERBEROS
zeppelin.jdbc.auth.kerberos.proxy.enable	true	When auth type is Kerberos, enable/disable Kerberos proxy with the login user to get the connection. Default value is true.
zeppelin.jdbc.concurrent.use	true	Use parallel scheduler
zeppelin.jdbc.concurrent.max_connection	10	Number of concurrent execution
zeppelin.jdbc.keytab.location		Kerberos keytab location
zeppelin.jdbc.principal		Kerberos principal
zeppelin.jdbc.interpolation	false	Enable ZeppelinContext variable interpolation into paragraph text
zeppelin.jdbc.maxConnLifetime	-1	Maximum of connection lifetime in milliseconds. A value of zero or less means the connection has an infinite lifetime.
zeppelin.jdbc.maxRows	1000	Maximum number of rows fetched from the query.
zeppelin.jdbc.hive.timeout.threshold	60000	Timeout for hive job timeout
zeppelin.jdbc.hive.monitor.query_interval	1000	Query interval for hive statement
Dependencies		
Artifact	Exclude	
mysql:mysql-connector-java:8.0.17		

Configuración de Spark en Interprete de Zeppelin

Para spark debemos tener el Spark_Home y spark.master con estas variables

spark %spark, %sql, %dyspark, %byspark, %r, %dr, %shiny, %kotlin

Option
The interpreter will be instantiated Globally in shared process
☐ Connect to existing process
☐ Set permission

Properties

Name	Value	Description	Action
SPARK_HOME	<input type="text" value="/opt/spark"/>	Location of spark distribution	<input type="button" value="✕"/>
spark.master	<input type="text" value="local[*]"/>	Spark master uri. local yarn-client yarn-cluster spark master address of standalone mode. ex) spark://master_host:7077	<input type="button" value="✕"/>
spark.submit.deployMode	<input type="text"/>	The deploy mode of Spark driver program, either "client" or "cluster". Which means to launch driver program locally ("client") or remotely ("cluster") on one of the nodes inside the cluster.	<input type="button" value="✕"/>
spark.app.name	<input type="text"/>	The name of spark application.	<input type="button" value="✕"/>
spark.driver.cores	<input type="text" value="1"/>	Number of cores to use for the driver process, only in cluster mode.	<input type="button" value="✕"/>
spark.driver.memory	<input type="text" value="1g"/>	Amount of memory to use for the driver process, i.e. where SparkContext is initialized, in the same format as JVM memory strings with a size unit suffix ("K", "M", "G" or "T") (e.g. 512m, 2g).	<input type="button" value="✕"/>
spark.executor.cores	<input type="text" value="1"/>	The number of cores to use on each executor	<input type="button" value="✕"/>
spark.executor.memory	<input type="text" value="1g"/>	Executor memory per worker instance. ex) 512m, 32g	<input type="button" value="✕"/>
spark.executor.instances	<input type="text" value="2"/>	The number of executors for static allocation.	<input type="button" value="✕"/>

Configuración de Python en Interprete de Zeppelin

Para Python debemos tener estas variables de este modo

python %python, %dypython, %sql, %conda, %docker

Option
The interpreter will be instantiated Globally in shared process
☐ Connect to existing process
☐ Set permission

Properties

Name	Value	Description
zeppelin.python	python	Python binary executable path. It is set to python by default (assume python is in your \$PATH)
zeppelin.python.maxResult	1000	Max number of dataframe rows to display.
zeppelin.python.useIPython	true	Whether use IPython when it is available in %python
zeppelin.python.launch.timeout	30000	Time out for ipython launch
zeppelin.python.grpc.message_size	33554432	grpc message size, default is 32M

Análisis a realizar

Visualización de consultas

Consulta 1

Contar el número total de filas (registros) que existen en la tabla union_educacion de la base de datos MySQL.

```
mysql> SELECT COUNT(*) FROM union_educacion;
```

```
+-----+  
| COUNT(*) |
```

```
+-----+  
| 378029 |
```

```
+-----+
```

```
1 row in set (0.01 sec)
```

Consulta 2

¿Cuántas instituciones educativas distintas existen en cada provincia del país, y cuáles son las provincias con mayor y menor número de instituciones?

```
mysql> SELECT Provincia, COUNT(DISTINCT AMIE) as Instituciones
```

```
-> FROM union_educacion
```

```
-> GROUP BY Provincia
```

```
-> ORDER BY Instituciones DESC;
```

Provincia	Instituciones
GUAYAS	5492
MANABI	4640
PICHINCHA	3546
LOS RIOS	1838
ESMERALDAS	1634
LOJA	1595
AZUAY	1361
CHIMBORAZO	1345

Consulta 3

¿Cuál es el total de docentes registrados en cada cantón del país, y cómo se distribuyen los cantones según el número de docentes?

```
mysql> SELECT Canton, SUM(Total_Docentes) as Docentes  
      -> FROM union_educacion  
      -> GROUP BY Canton  
      -> ORDER BY Docentes DESC;
```

Canton	Docentes
QUITO	524783
GUAYAQUIL	517479
CUENCA	140076
SANTO DOMINGO	89393
AMBATO	81720
PORTOVIEJO	75413
RIOBAMBA	70805
LOJA	68702
MACHALA	62354
MANTA	59098
ESMERALDAS	58662
IBARRA	52506
LATAKUNGA	10866

Consulta 4

¿Qué diferentes periodos académicos (años lectivos) existen en la tabla de registros educativos?

```
mysql> SELECT DISTINCT anio_lectivo FROM union_educacion;
```

anio_lectivo
2022-2023 FIN
2009-2010 INICIO
2010-2011 INICIO
2011-2012 INICIO
2012-2013 INICIO
2013-2014 INICIO
2014-2015 INICIO
2015-2016 INICIO
2016-2017 INICIO
2017-2018 INICIO

Consulta 5

¿Cuáles son los registros educativos que tienen valores nulos (vacíos) en el número total de estudiantes o docentes?

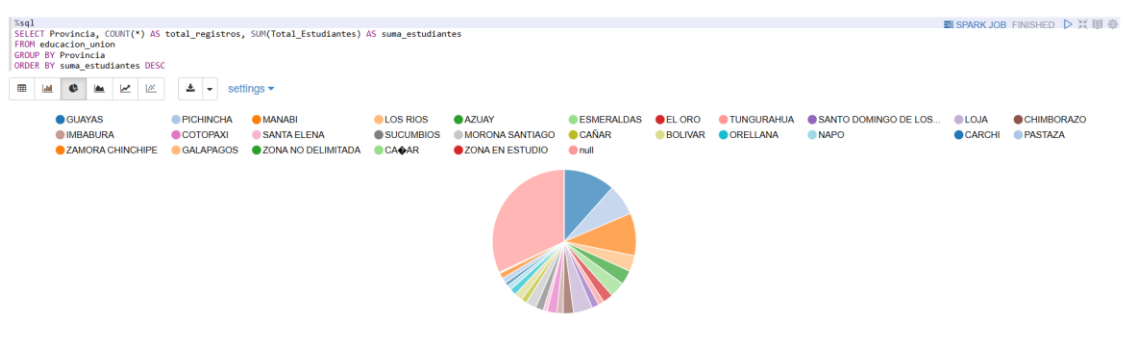
```
mysql> SELECT * FROM union_educacion WHERE Total_Estudiantes IS NULL OR Total_Docentes IS NULL LIMIT 10;
```

AMIE	anio_lectivo	Provincia	Canton	Total_Estudiantes	Total_Docentes
07H00020	NULL	NULL	NULL	NULL	NULL
07H00020	NULL	NULL	NULL	NULL	NULL
07H00020	NULL	NULL	NULL	NULL	NULL
07H00020	NULL	NULL	NULL	NULL	NULL
07H00020	NULL	NULL	NULL	NULL	NULL
07H00020	NULL	NULL	NULL	NULL	NULL
07H00020	NULL	NULL	NULL	NULL	NULL
07H00020	NULL	NULL	NULL	NULL	NULL
07H00020	NULL	NULL	NULL	NULL	NULL
07H00020	NULL	NULL	NULL	NULL	NULL

10 rows in set (0.15 sec)

Consulta 6

¿Cuál es la cantidad total de registros y la suma de estudiantes por provincia?



Conclusiones

La integración de los registros administrativos del Ministerio de Educación y las evaluaciones de INEVAL para los periodos 2022-2023 permitió construir un dataset educativo robusto y coherente. A través de la normalización relacional y la eliminación de dependencias, se garantizó la integridad de los datos y se facilitaron consultas complejas que combinan información de matrícula, desempeño estudiantil y características institucionales.

El uso de Apache Zeppelin, en conjunto con Spark/Scala y MySQL, en un entorno Ubuntu 20.04.5 instalando desde VirtualBox o Gitpod, proporcionó un flujo de trabajo reproducible y colaborativo. La lectura, limpieza y unión de los archivos CSV principales mediante Zeppelin agilizaron el proceso de preparación de datos, mientras que la posterior carga en MySQL a través de LOAD DATA INFILE aseguró un almacenamiento eficiente y listo para el análisis.

El análisis inicial confirma la viabilidad de este marco para explorar indicadores como distribución de estudiantes y docentes por provincia, comparativa histórica de matrículas y niveles de logro en matemáticas y lenguaje. Este enfoque sienta las bases para ampliar el estudio a nuevos años y variables, apoyando la toma de decisiones en políticas educativas basadas en evidencia.

Bibliografía

Cordero, P. (2020, 29 septiembre). *Como instalar Ubuntu en VirtualBox | Oficina de software libre*. <https://osl.ugr.es/2020/09/29/como-instalar-ubuntu-en-virtual-box/>

Kaplarevic, V. (2024, 10 octubre). How to Install Spark on Ubuntu. Knowledge Base By phoenixNAP. <https://phoenixnap.com/kb/install-spark-on-ubuntu>

Apache Zeppelin | Cloudera. (2022, 21 noviembre). Cloudera. <https://es.cloudera.com/products/open-source/apache-hadoop/apache-zeppelin.html#:~:text=Zeppelin%20es%20una%20moderna%20plataforma,de%20datos%20cada%20vez%20mayor.>

The Apache Software Foundation. (s. f.). Apache Zeppelin 0.12.0 Documentation: Apache Spark Interpreter for Apache Zeppelin. <https://zeppelin.apache.org/docs/latest/interpreter/spark.html>

OpenAI. (2024). ChatGPT (GPT-4o) [Modelo de lenguaje]. <https://chat.openai.com/>

Ubuntu. (s.f.). Acerca de Ubuntu. Recuperado el 21 de julio de 2025, de <https://ubuntu.com/about>