

TEMARIO DEFINITIVO

De principiante a Data Scieclist Experto

Nivel Principiante

1. Introducción a las Bases de Datos

1. Conceptos Básicos

- Qué es una base de datos
- Diferencias entre bases de datos relacionales y no relacionales
- Arquitectura de bases de datos

2. SQL Básico

- Instalación de un sistema de gestión de bases de datos (MySQL, PostgreSQL)
- Sintaxis básica de SQL
 - SELECT, INSERT, UPDATE, DELETE
- Tipos de datos en SQL
- Cláusulas WHERE, ORDER BY, y LIMIT

3. Modelado de Datos

- Entidades y relaciones
- Diagramas entidad-relación (ERD)
- Normalización de bases de datos (1NF, 2NF, 3NF)

4. Herramientas y Entornos

- Uso de herramientas de gestión de bases de datos (phpMyAdmin, pgAdmin)
- Instalación y configuración de un entorno de desarrollo

2. Introducción a la Ciencia de Datos

a. Conceptos Básicos

- Qué es la ciencia de datos
- Ciclo de vida de un proyecto de ciencia de datos
- Roles en ciencia de datos (data scientist, data engineer, data analyst)

2. Python para Ciencia de Datos

- Instalación de Python y Jupyter Notebook
- Librerías fundamentales: NumPy, Pandas, Matplotlib
- Conceptos básicos de programación en Python

3. Análisis Exploratorio de Datos (EDA)

- Importación y exportación de datos (CSV, Excel, SQL)
- Limpieza de datos (manejo de valores nulos, duplicados)
- Visualización de datos básicos (gráficos de barras, líneas, dispersión)

Nivel Intermedio

3. Bases de Datos Relacionales Avanzadas

1. Consultas SQL Avanzadas

- Joins (INNER, LEFT, RIGHT, FULL)
- Subconsultas y consultas correlacionadas
- Funciones agregadas (SUM, AVG, COUNT, etc.)
- Índices y optimización de consultas

2. Administración de Bases de Datos

- Creación y gestión de usuarios y permisos
- Backup y restauración de bases de datos
- Transacciones y bloqueo de registros

3. Introducción a ORMs

- Conceptos de ORM (Object-Relational Mapping)
- Uso de SQLAlchemy con Python
- ORM en Django

4. Bases de Datos NoSQL

1. Conceptos Básicos de NoSQL

- Diferencias con bases de datos SQL
- Tipos de bases de datos NoSQL (documento, clave-valor, columnares, grafos)

2. MongoDB

- Instalación y configuración
- Operaciones CRUD en MongoDB
- Índices y agregaciones en MongoDB

3. Redis

- Instalación y configuración
- Operaciones básicas (SET, GET, INCR, etc.)
- Uso como caché en aplicaciones web

5. Ciencia de Datos Intermedia

1. Análisis de Datos con Pandas

- Manipulación de DataFrames
- Operaciones avanzadas con Pandas (merge, groupby, pivot)

2. Visualización de Datos Avanzada

- Gráficos avanzados con Matplotlib y Seaborn
- Creación de dashboards con Plotly y Dash

3. Estadística Básica y Probabilidad

- Descriptiva: media, mediana, moda, desviación estándar
- Inferencial: distribuciones, pruebas de hipótesis, intervalos de confianza

4. Introducción al Machine Learning

- Algoritmos supervisados vs. no supervisados
- Regresión lineal y logística
- K-means clustering

Nivel Avanzado

6. Bases de Datos Distribuidas y Big Data

1. Apache Hadoop

- Conceptos básicos y arquitectura
- HDFS (Hadoop Distributed File System)
- MapReduce

2. Apache Spark

- Instalación y configuración
- RDDs (Resilient Distributed Datasets)
- Spark SQL y DataFrames

3. Bases de Datos Distribuidas

- Cassandra
- HBase

7. Machine Learning Avanzado

1. Algoritmos Avanzados

- Árboles de decisión y random forests
- Support Vector Machines (SVM)
- Redes Neuronales y Deep Learning (con TensorFlow y Keras)

2. Evaluación y Validación de Modelos

- Métodos de validación cruzada
- Métricas de evaluación (precisión, recall, F1-score, AUC-ROC)

3 Ingeniería de Características

- Selección y extracción de características
- Técnicas de reducción de dimensionalidad (PCA, LDA)

8. Data Engineering y MLOps

1. Pipelines de Datos

- Concepto y diseño de ETL (Extract, Transform, Load)
- Herramientas de orquestación (Apache Airflow)

2. MLOps

- Automatización del ciclo de vida de modelos
- Despliegue y monitoreo de modelos
- CI/CD para ciencia de datos

Nivel Experto

9. Deep Learning y Redes Neuronales Avanzadas

1. ****Conceptos Avanzados de Deep Learning****

- Redes convolucionales (CNN)
- Redes recurrentes (RNN, LSTM)
- Redes generativas adversarias (GANs)

2. Procesamiento de Lenguaje Natural (NLP)

- Tokenización y embeddings (Word2Vec, GloVe)
- Modelos de lenguaje (BERT, GPT)

3. Visión por Computador

- Detección y segmentación de objetos
- Transfer learning y modelos preentrenados

10. Arquitectura y Escalabilidad de Datos

1. Diseño de Arquitecturas de Datos

- Data lakes y data warehouses
- Arquitecturas Lambda y Kappa

2. Seguridad y Gobernanza de Datos

- Protección de datos sensibles
- Cumplimiento de normativas (GDPR, CCPA)

3. Estrategias de Almacenamiento y Recuperación de Datos

- Sharding y particionamiento de bases de datos
- Mantenimiento y optimización de grandes volúmenes de datos

Recursos Adicionales

- Libros:

- "Python for Data Analysis" de Wes McKinney
- "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" de Aurélien Géron
- "Database System Concepts" de Abraham Silberschatz, Henry F. Korth, y S. Sudarshan

- Cursos en línea:

- Coursera, Udacity, edX para cursos específicos de ciencia de datos y bases de datos
- MOOCs de universidades (MIT, Stanford)

- Comunidades y Foros:

- Stack Overflow
- Reddit (subreddits como r/datascience, r/MachineLearning, r/database)
- Grupos de Meetup para profesionales de datos
- Contribuciones a proyectos de código abierto en GitHub

Este temario te llevará desde lo básico hasta lo más avanzado en bases de datos y ciencia de datos, cubriendo todos los aspectos necesarios para convertirte en un experto en estos campos.