

UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA

DESARROLLO Y GESTIÓN DE SOFTWARE



**EXTRACCIÓN DE CONOCIMIENTO EN BASES DE
DATOS**

**REPORTE DE INVESTIGACIÓN DE LOS LENGUAJES Y
BIBLIOTECAS PARA ANÁLISIS Y PROCESAMIENTO DE
DATOS**

PRESENTA:

ANGEL RICARDO CHAVEZ ZARAGOZA

DOCENTE:

ING. LUIS ENRIQUE MASCOTE CANO

Chihuahua, Chih., 23 de septiembre de 2025

Contenidos

Objetivo	3
Introducción	3
Python	3
SQL	4
SAS	4
Scala.....	5
MATLAB.....	5
Conclusión	6
Referencias.....	6

Objetivo

Investigar y documentar los principales lenguajes de programación y sus bibliotecas o frameworks asociados que se utilizan habitualmente para la manipulación, análisis y modelado de datos en proyectos de IA, ML, DM y Big Data.

Introducción

En la actualidad, el análisis y procesamiento de datos se ha convertido en una de las actividades más relevantes dentro de la ciencia, la industria y la vida cotidiana. Con el crecimiento exponencial de la información generada por usuarios, empresas y dispositivos conectados, surge la necesidad de contar con herramientas que permitan organizar, interpretar y transformar los datos en conocimiento útil para la toma de decisiones. En este contexto, los lenguajes de programación y sus bibliotecas especializadas desempeñan un papel fundamental, ya que facilitan desde la limpieza y manipulación de grandes volúmenes de datos hasta la aplicación de modelos de inteligencia artificial (IA), aprendizaje automático (ML), minería de datos (DM) y Big Data.

Conocer los lenguajes y frameworks más utilizados en este campo resulta esencial por varias razones. En primer lugar, permiten seleccionar la herramienta más adecuada según el tipo de proyecto o problema a resolver. Además, proporcionan acceso a comunidades activas de desarrolladores y documentación que favorece la resolución de problemas y la innovación. Finalmente, la elección de un lenguaje y sus bibliotecas asociadas no solo impacta en la eficiencia del análisis, sino también en la escalabilidad y calidad de los resultados obtenidos. Por ello, la investigación y documentación de estas tecnologías es un paso necesario para comprender el panorama actual del análisis de datos y su aplicación en diversos ámbitos profesionales y académicos.

Python

Descripción general:

- Paradigma: Interpretado, tipado dinámico.
- Ámbito de uso: Ciencia de datos, inteligencia artificial, machine learning, automatización.

Bibliotecas clave:

- **pandas:** Manipulación de datos mediante DataFrames.
- **NumPy:** Operaciones matemáticas y manejo de arreglos multidimensionales.

- **scikit-learn:** Modelado predictivo y algoritmos de machine learning.

Mini-ejemplo “Hola datos”:

```
import pandas as pd

# Cargar archivo CSV
df = pd.read_csv("datos.csv")

# Mostrar primeras filas
print(df.head())
```

SQL

Descripción general:

- Paradigma: Lenguaje declarativo, especializado en bases de datos.
- Ámbito de uso: Consulta, manipulación y gestión de datos relacionales.

Bibliotecas clave:

- **MySQL / PostgreSQL:** Motores de bases de datos ampliamente usados.
- **SQLite:** Base de datos ligera y embebida.
- **Microsoft SQL Server:** Orientado a soluciones empresariales.

Mini-ejemplo “Hola datos”:

```
-- Cargar y mostrar datos
SELECT *
FROM ventas
LIMIT 5;
```

SAS

Descripción general:

- Paradigma: Lenguaje propietario, orientado a análisis estadístico.
- Ámbito de uso: Estadística avanzada, análisis de negocios, biomedicina.

Bibliotecas clave:

- **SAS/STAT:** Modelos estadísticos.
- **SAS/ETS:** Series de tiempo y econometría.
- **SAS/GRAFH:** Visualización de datos.

Mini-ejemplo “Hola datos”:

```
PROC IMPORT DATAFILE="datos.csv" OUT=work.datos DBMS=csv REPLACE;
RUN;

PROC PRINT DATA=work.datos (OBS=5);
RUN;
```

Scala

Descripción general:

- Paradigma: Compilado, tipado estático, multiparadigma (funcional y orientado a objetos).
- Ámbito de uso: Big Data, procesamiento distribuido.

Bibliotecas clave:

- **Apache Spark:** Procesamiento distribuido y análisis de big data.
- **Breeze:** Álgebra lineal y estadísticas.
- **Akka:** Computación concurrente y distribuida.

Mini-ejemplo “Hola datos”:

```
import org.apache.spark.sql.SparkSession

val spark = SparkSession.builder.appName("HolaDatos").getOrCreate()

val df = spark.read.option("header","true").csv("datos.csv")
df.show(5)
```

MATLAB

Descripción general:

- Paradigma: Interpretado, tipado dinámico, orientado a cálculos numéricos.
- Ámbito de uso: Procesamiento de señales, matemáticas avanzadas, ingeniería.

Bibliotecas clave:

- **Statistics and Machine Learning Toolbox:** Modelos estadísticos y de ML.

- **Deep Learning Toolbox:** Redes neuronales y entrenamiento de modelos.
- **Image Processing Toolbox:** Análisis y manipulación de imágenes.

Mini-ejemplo “Hola datos”:

```
datos = readtable("datos.csv");

% Mostrar primeras filas
head(datos,5)
```

Conclusión

Cada lenguaje y biblioteca analizado aporta un valor particular al campo del análisis de datos. Python destaca por su versatilidad y comunidad, SQL sigue siendo indispensable para manejar bases de datos, SAS es fuerte en análisis estadístico profesional, Scala brilla en big data gracias a Spark, y MATLAB sobresale en el ámbito científico e ingenieril. La elección del lenguaje dependerá del tipo de proyecto, los recursos disponibles y los objetivos que se persigan. Conocer estas opciones permite a los profesionales de datos elegir la mejor herramienta para resolver problemas de manera eficiente.

Para proyectos de análisis ligero (exploración de datos, pruebas, investigación académica): Python y SQL son las mejores opciones por su facilidad, versatilidad y rapidez en el desarrollo. MATLAB también puede ser útil en contextos académicos.

Para proyectos de producción a gran escala (big data, sistemas empresariales, entornos distribuidos): Scala con Spark es la alternativa más adecuada por su eficiencia y escalabilidad, mientras que SAS puede ser recomendable en entornos corporativos que ya cuentan con su infraestructura.

Referencias

¿Qué es SQL? - Explicación de lenguaje de consulta estructurado (SQL) - AWS. (n.d.). Amazon Web Services, Inc. <https://aws.amazon.com/es/what-is/sql/>

Deep Learning toolbox. (n.d.). <https://la.mathworks.com/products/deep-learning.html>

Introducción. (n.d.). Scala Documentation. <https://docs.scala-lang.org/es/tour/tour-of-scala.html>

Librerías en Data Science / Interactive Chaos. (n.d.).

<https://interactivechaos.com/es/manual/tutorial-de-python/librerias-en-data-science>

SAS Help Center: Acceso a librerías SAS. (2025, August 13).

https://documentation.sas.com/doc/es/workbenchcdc/v_001/workbenchgs/n0s1c9n4mbxa/vgn1cvrv0ztb5o54.htm