

Herramientas de Gestión de Datos

MySQL

- Sistema de **gestión de bases de datos relacional** (RDBMS) de código abierto popular
- Utiliza el lenguaje de consulta estructurado (SQL) para gestionar y almacenar datos.
- Usos comunes:
 - Aplicaciones web
 - Almacenamiento de datos
 - Comercio electrónico

PostgreSQL

- Sistema de **gestión de bases de datos relacional** (RDBMS) poderoso y de código abierto
- Enfatiza la extensibilidad y el cumplimiento de SQL.
- Ofrece características avanzadas como:
 - Soporte para JSON
 - Búsqueda de texto completo
 - Datos espaciales

Apache CouchDB

- Base de datos **NoSQL** orientada a documentos
- Utiliza JSON para almacenar datos
- Altamente escalable
- Tolerante a fallos

- Fácil de usar

MongoDB

- Base de datos **NoSQL** orientada a documentos
- Almacena datos en un JSON flexible
- Proporciona:
 - Escalabilidad
 - Alta disponibilidad
 - Distribución de datos
- Adecuada para aplicaciones web modernas que manejan grandes volúmenes de datos no estructurados

Apache Cassandra

- Base de datos **NoSQL** orientada a documentos, altamente escalable y distribuida
- Puede manejar grandes cantidades de datos estructurados y no estructurados en muchos servidores comunes.
- Ofrece:
 - Alta disponibilidad
 - Tolerancia a fallos
 - Niveles de consistencia ajustables
- Adecuada para aplicaciones críticas para la misión

Sistema de Archivos Distribuido de Hadoop (HDFS)

- Diseñado para trabajar con grandes conjuntos de datos como Apache Hadoop en un entorno de computación distribuida
- Procesamiento de datos de alto rendimiento al dividir archivos en bloques (por defecto 128MB), y estos bloques se distribuyen entre múltiples DataNodes
- Los datos se replican en diferentes DataNodes, asegurando alta disponibilidad y tolerancia a fallos

- Escalable y eficiente

Ceph

- Plataforma de almacenamiento definida por software, gratuita y de código abierto, adecuada para entornos de nube híbrida
- Diseñada para centros de datos modernos
- Proporciona un sistema de almacenamiento unificado, altamente escalable, que se puede utilizar para almacenamiento de objetos (como AWS S3), almacenamiento en bloques (como discos virtuales para máquinas virtuales) y almacenamiento de archivos (como NFS) bajo un único sistema unificado
- Alto rendimiento, disponibilidad y fiabilidad

Elasticsearch

- Principalmente un motor de búsqueda y herramienta de análisis RESTful distribuido
- Basado en la biblioteca Lucene.
- Búsqueda de texto completo, análisis de datos en tiempo real
- Altamente escalable
- Fácil de usar
- Potentes capacidades de consulta
- Indexación de datos en tiempo real para una rápida recuperación de documentos.

Herramientas de Integración y Transformación de Datos

Apache Airflow

- Plataforma de código abierto para la creación, programación y monitoreo de flujos de trabajo de manera programática
- Creada originalmente por Airbnb
- Permite a los usuarios definir y ejecutar flujos de trabajo complejos
- Soporte para:
 - Dependencias de tareas
 - Paralelismo
 - Manejo de errores

Kubeflow

- Un kit de herramientas de aprendizaje automático de código abierto que permite la ejecución de pipelines de ciencia de datos sobre Kubernetes.
- Proporciona una plataforma para construir, desplegar y gestionar flujos de trabajo de aprendizaje automático de extremo a extremo a gran escala
- Soporte para:
 - Entrenamiento distribuido
 - Servicio de modelos
 - Ajuste de hiperparámetros

Apache Kafka

- Plataforma de streaming distribuido que permite a las aplicaciones publicar, procesar y suscribirse a flujos de registros en tiempo real
- Creado originalmente en LinkedIn.
- Es escalable, tolerante a fallos y de alto rendimiento
- Adecuado para construir aplicaciones críticas para la misión y que requieren intensivo uso de datos

Apache NiFi

- Una plataforma de integración de datos de código abierto que permite a los usuarios automatizar el flujo de datos entre sistemas
- Proporciona una interfaz de usuario basada en la web para diseñar y gestionar flujos de datos

- Soporte para:
 - Enrutamiento de datos
 - Transformación
 - Enriquecimiento
 - Entre otras capacidades

Apache Spark SQL

- Un módulo en el ecosistema de Spark que proporciona una interfaz de programación para trabajar con datos estructurados utilizando:
 - SQL
 - Data frames
 - Datasets
- Soporta una amplia gama de fuentes de datos y proporciona un rendimiento optimizado para tareas complejas de procesamiento de datos.

Node-RED

- Una herramienta de programación visual de código abierto para conectar dispositivos de hardware, APIs y servicios en línea.
- Permite a los usuarios crear flujos de mensajes impulsados por eventos.
- Bajo consumo de recursos, que incluso se ejecuta en dispositivos pequeños como una Raspberry Pi.
- Soporte para:
 - Transformación de datos
 - Filtrado
 - Agregación

Herramientas de Visualización de Datos

PixieDust

- Biblioteca de código abierto para crear visualizaciones de datos interactivas y exploratorias en Python y cuadernos Jupyter
- Proporciona una variedad de visualizaciones integradas y conectores de datos
- Soporte para personalización y extensibilidad a través de bibliotecas de terceros

Matiz

- Interfaz web de código abierto para analizar y visualizar grandes conjuntos de datos en Apache Hadoop
- Ofrece una experiencia fácil de usar para explorar datos y crear visualizaciones
- No se requieren habilidades de programación; se pueden crear visualizaciones a partir de consultas SQL

Kibana

- Herramienta de visualización de datos de código abierto que permite a los usuarios interactuar con sus datos a través de una interfaz basada en la web.
- Comúnmente utilizada con Elasticsearch para analizar y visualizar grandes conjuntos de datos.

Apache Superset

- Una aplicación web de inteligencia empresarial moderna y lista para empresas que facilita la visualización y exploración de grandes conjuntos de datos.
- Ofrece un conjunto rico de opciones de visualización de datos, que incluye:
 - Gráficos
 - Tablas
 - Mapas
 - Análisis geoespacial
 - Procesamiento de datos en tiempo real

Herramientas de Despliegue de Modelos

Apache PredictionIO

- Servidor de aprendizaje automático de código abierto construido sobre una infraestructura escalable y distribuida
- Permite a los desarrolladores construir, evaluar y desplegar rápidamente motores predictivos para varios casos de uso, tales como:
 - Recomendación
 - Clasificación
 - Agrupamiento

Kubernetes

- Plataforma de código abierto para la orquestación de contenedores
- Lanza, escala y gestiona automáticamente aplicaciones en contenedores
- Ofrece características como:
 - Escalado automático
 - Auto-reparación
 - Balanceo de carga
- Permite la gestión y orquestación de contenedores en numerosos hosts

Apache Seldon

- Plataforma de código abierto para implementar y gestionar modelos de aprendizaje automático en Kubernetes
- Proporciona una manera de:
 - Servir modelos a gran escala
 - Automatizar flujos de trabajo de implementación de modelos

- Monitorear el rendimiento de los modelos implementados en tiempo real

MLeap

- Biblioteca de código abierto para serializar y deserializar modelos de aprendizaje en un archivo multiplataforma
- Ofrece a los usuarios la capacidad de exportar modelos de diferentes bibliotecas y marcos de aprendizaje automático, como:
 - Spark
 - Scikit-learn
 - TensorFlow
- Los implementa en entornos de producción de alto rendimiento y baja latencia

TensorFlow Lite

- Herramienta de código abierto para ejecutar modelos de aprendizaje automático en dispositivos móviles y embebidos
- Permite una inferencia efectiva en plataformas móviles y embebidas
- Soporta una variedad de aceleradores de hardware como:
 - CPUs
 - GPUs
 - ASICs personalizados

Red Hat OpenShift

- Marco de aplicación de contenedores basado en Kubernetes
- Con características como automatización, escalabilidad y seguridad
- Ofrece un método para crear, implementar y gestionar aplicaciones en contenedores

TensorFlow Serving

- Utilidad de código abierto que sirve modelos de aprendizaje automático en entornos del mundo real

- Soporta interfaces HTTP y gRPC para ofrecer predicciones
- Proporciona alta escalabilidad y despliegue y gestión de modelos de TensorFlow con baja latencia

TensorFlow.js

- Biblioteca de código abierto para construir y desplegar modelos de aprendizaje automático en JavaScript
- Te permite entrenar y ejecutar modelos directamente en el navegador o en Node.js
- Soporta una amplia gama de arquitecturas de modelos, incluyendo redes neuronales, árboles de decisión y k-vecinos más cercanos

Herramientas de Monitoreo y Evaluación de Modelos

ModelDB

- Plataforma de código abierto para gestionar modelos de aprendizaje automático y experimentos
- Proporciona una forma de rastrear y reproducir experimentos, versionar modelos y colaborar con miembros del equipo

Prometheus

- Sistema de monitoreo disponible de forma gratuita que recopila y almacena métricas en tiempo real desde diferentes fuentes
- Te permite visualizar y establecer alertas sobre la salud y el rendimiento de sistemas y aplicaciones
- Soporta una variedad de métodos de recopilación de datos, como puntos finales HTTP, exportadores y agentes

IBM AI Fairness 360

- Conjunto de herramientas de código abierto para detectar y mitigar sesgos en modelos de aprendizaje automático
- Proporciona una forma de medir la equidad y el sesgo de los modelos, así como un conjunto de algoritmos para mitigar el sesgo y crear modelos más justos

IBM AI Explainability 360

- Conjunto de herramientas de código abierto para explicar el comportamiento y las decisiones de los modelos de aprendizaje automático.
- Proporciona una forma de medir la explicabilidad e interpretabilidad de los modelos, así como un conjunto de algoritmos para generar explicaciones y visualizaciones del comportamiento del modelo.

IBM Adversarial Robustness 360 Toolbox

- Biblioteca gratuita y de código abierto para proteger modelos de aprendizaje automático de ataques adversariales
- Incluye un método para medir la robustez y vulnerabilidad del modelo
- Incluye un conjunto de algoritmos para mejorar la robustez del modelo y detectar ejemplos adversariales

Herramientas de Desarrollo y Ejecución de Código

Jupyter IDE

- Esfuerzo de código abierto
- Soporta:

- Julia
 - Python
 - Desarrollo en R con Jupyter Notebook
 - JupyterLab
 - JupyterHub
- Crea y comparte documentos que contienen:
 - Código en vivo
 - Ecuaciones
 - Visualizaciones
 - Texto narrativo
- JupyterLab incluye:
 - Organización personalizada de cuadernos
- JupyterHub extiende todas estas capacidades a la empresa

RStudio

- Para desarrolladores
- IDE gratuito y de código abierto
- Diseñado para gestionar y ejecutar código R
- Funciona en todas las plataformas
- Incluye:
 - Control de versiones
 - Capacidades de gestión de proyectos

Microsoft Visual Studio

- Un IDE que soporta una variedad de lenguajes de programación, incluyendo:
 - C
 - C++
 - C++/CLI
 - Visual Basic.NET
 - C#
 - F#
 - JavaScript
 - TypeScript

- XML
 - XSLT
 - HTML
 - CSS
- Usando complementos, soporta:
 - Python
 - Ruby
 - Node.js
 - M
 - Otros lenguajes

PyCharm

- Principalmente un entorno IDE basado en suscripción
- Ofrece más de 16 herramientas adicionales para asistencia en codificación, pruebas y desarrollo web
- Soporta el desarrollo científico con integración de IPython y soporte para Matplotlib y NumPy
- También ofrece un IDE gratuito, de código abierto y basado en la comunidad con capacidades limitadas

Spyder

- IDE gratuito y de código abierto basado en Python, diseñado por y para científicos, ingenieros y analistas de datos.
- Ofrece una combinación única de herramientas de desarrollo completas para:
 - Edición avanzada
 - Análisis
 - Depuración
 - Perfilado
 - Capacidades de visualización

Anaconda Navigator

- Navegador basado en GUI de código abierto que soporta el desarrollo en Python e integra con:
 - Eclipse y PyDev
 - IDLE
 - IntelliJ
 - Microsoft Visual Studio Code (VS Code)
 - Ninja IDE
 - PyCharm
 - Python para Visual Studio Code
 - Herramientas de Python para Visual Studio (PTVS)
 - Spyder
 - Sublime Text
 - Wing IDE

Herramientas de Gestión de Activos de Código

Git

- Sistema de control de versiones de código abierto para rastrear cambios en el código y colaborar entre desarrolladores
- Proporciona una manera de gestionar y organizar los cambios en el código, colaborar en el desarrollo del código y mantener un historial de revisiones del código

GitLab

- Gestor de repositorios Git basado en la web
- Proporciona una plataforma DevOps completa para:
 - Gestión de código fuente
 - Integración y despliegue continuos

- Monitoreo
- Permite a los equipos colaborar en:
 - Desarrollo de código
 - Automatización de procesos de construcción y despliegue
 - Seguimiento de métricas y rendimiento a lo largo de todo el ciclo de vida del desarrollo de software

GitHub

- Servicio de alojamiento de repositorios Git basado en la web que proporciona una plataforma para que los desarrolladores colaboren en código y gestionen proyectos de software.
- Permite a los usuarios:
 - Crear, bifurcar y contribuir a proyectos de código abierto.
 - Rastrear cambios en el código.
 - Gestionar problemas.
 - Solicitudes de extracción.

Bitbucket de Atlassian

- Servicio de alojamiento de repositorios Git basado en la web
- Proporciona una plataforma para que los desarrolladores colaboren en el código y gestionen proyectos de software, con características como:
 - Solicitudes de extracción
 - Revisión de código