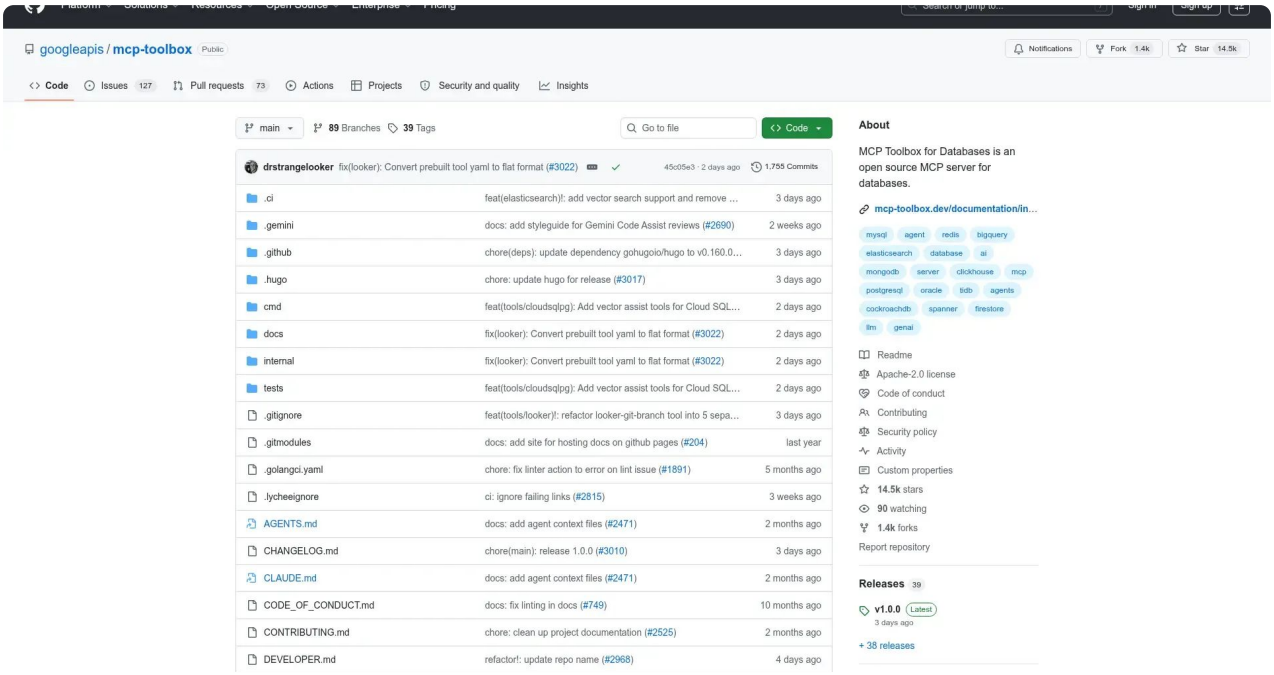

MCP Toolbox for Databases

Herramienta de código abierto diseñada para ingenieros de datos y desarrolladores que necesitan conectar agentes de IA con bases de datos empresariales. Permite que LLMs e IDEs interactúen con activos de datos estructurados mediante lenguaje natural, eliminando el código boilerplate. Es ideal para transformar esquemas complejos en herramientas utilizables por IA de forma segura, ofreciendo control total sobre las consultas SQL ejecutadas y facilitando la integración con ecosistemas cloud.

Visitar Sitio Oficial Preguntar a ChatGPT Preguntar a Claude Preguntar a Grok

Contenido del Dossier

INFORMACIÓN DE LA HERRAMIENTA

Qué y para quién es
MCP Toolbox for Databases (anteriormente Gen AI Toolbox) es un servidor de código abierto basado en el Model Context Protocol (MCP) de Google. Su función principal es actuar como puente entre agentes de IA, IDEs y aplicaciones, permitiendo que estos interactúen directamente con bases de datos empresariales mediante lenguaje natural. Está diseñado para profesionales de IT, ingenieros de datos y desarrolladores que buscan integrar capacidades de IA generativa con sus activos de datos estructurados de forma segura y eficiente.

Principal ventaja profesional
Permite transformar bases de datos complejas en herramientas utilizables por IA en cuestión de minutos, eliminando la necesidad de escribir código "boilerplate" para conectar LLMs con el almacenamiento de datos, garantizando al mismo tiempo el control sobre las consultas ejecutadas.

Para quién no es
No es adecuado para usuarios finales sin conocimientos técnicos o empresas que no tengan una infraestructura de base de datos mínima. Profesionales que busquen soluciones de análisis visual tipo BI sin integración con agentes de IA o LLMs encontrarán la herramienta demasiado orientada al desarrollo.

Funcionalidades clave
- Acceso "Out-of-the-box" a bases de datos: Herramientas preconfiguradas para explorar esquemas y ejecutar SQL.
- Framework de herramientas personalizadas: Definición de lógica de negocio propia, consultas estructuradas y búsqueda semántica mediante archivos YAML.
- Soporte Multi-Base de Datos: Compatible con ecosistemas Google Cloud (BigQuery, AlloyDB, Cloud SQL, Spanner) y externos (PostgreSQL, MySQL, MongoDB, Snowflake, Oracle, entre otros).
- Observabilidad integrada: Soporte nativo para OpenTelemetry que permite monitorizar métricas y trazas de las interacciones IA-Base de Datos.
- Seguridad avanzada: Integración con IAM (Identity and Access Management) y soporte para autenticación empresarial.

Precios
- Versión gratuita: La herramienta es Open Source bajo licencia Apache 2.0. No tiene coste de licencia por uso.
- Versión gestionada: Google Cloud ofrece una versión administrada de servidores MCP cuyo coste depende del consumo de recursos en la plataforma Google Cloud.

Perfil del usuario
- Empresas con infraestructuras de datos en la nube o híbridas que deseen implementar asistentes de IA internos.
- Departamentos de arquitectura de datos e ingeniería de software.
- Listado de perfiles profesionales:
- Ingenieros de Machine Learning / GenAI.
- Desarrolladores Backend.
- Administradores de Bases de Datos (DBA).
- Arquitectos de Soluciones Cloud.

Nivel técnico requerido
- Nivel técnico para su uso: Medio. Requiere saber formular consultas o interactuar con agentes de IA.
- Nivel técnico para instalación/configuración: Alto. Es necesario manejo de terminal, archivos de configuración YAML, Docker o entornos de ejecución como Node.js/Go.
- Conocimientos necesarios: SQL, gestión de variables de entorno y fundamentos de seguridad en bases de datos.

Ejemplos de uso profesional
- Exploración de datos en lenguaje natural: Un desarrollador puede preguntar a su IDE "muéstrame las tablas de ventas del último trimestre" sin salir del entorno de código.
- Automatización de generación de código: Generación de esquemas y scripts de migración basados en la

estructura real de la base de datos detectada por la IA.
- Agentes de soporte técnico: Creación de bots que consulten bases de datos de clientes para resolver incidencias en tiempo real de forma segura.

Uso y distribución
- Versión web: Dispone de una Toolbox UI para gestión visual.
- Versión escritorio: Integración nativa con IDEs compatibles con MCP (como Claude Desktop o extensiones de VS Code).
- CLI: Herramienta de línea de comandos para ejecución directa.
- Contenedores: Imagen oficial en Docker para despliegues escalables.
- SDKs: Disponibles para Python, JavaScript/TypeScript, Go y Java.

Open Source
Licencia Apache 2.0. Repositorio gestionado por Google APIs en GitHub.

Integraciones
- Facilidad de integración: Nivel programador (Full code / Config-driven).
- API propia: Protocolo MCP (Model Context Protocol).
- Integraciones nativas: LangChain, LlamaIndex, Genkit y Gemini CLI.
- Ejemplos de integración: Conexión de Claude Code a una instancia de PostgreSQL o uso de Gemini para consultar BigQuery de forma directa mediante el protocolo MCP.

Notas finales
información legal, licencias , contratos
Se distribuye bajo la licencia Apache 2.0, lo que permite su uso comercial, modificación y distribución gratuita, siempre que se mantengan los avisos de copyright y de licencia correspondientes.

Otros
El proyecto ha sido renombrado recientemente de genai-toolbox a mcp-toolbox para alinearse plenamente con el estándar Model Context Protocol.

Para más información:
- Sitio web oficial: https://mcp-toolbox.dev
- Documentación y configuración: https://googleapis.github.io/genai-toolbox/
- Github: https://github.com/googleapis/mcp-toolbox
- Discord: https://discord.gg/google-cloud-dev

CONSEJOS DE IMPLANTACIÓN

Aplicación profesional
Empresas tecnológicas, sectores con grandes volúmenes de datos estructurados (Fintech, Retail, Logística) y equipos de desarrollo de software. El presupuesto de licencia es inexistente al ser código abierto, pero requiere inversión en infraestructura cloud (Google Cloud, AWS, Azure) y tiempo técnico de configuración. Los puntos clave son la democratización del acceso a datos para desarrolladores y la seguridad en la ejecución de consultas mediante LLMs.

Madurez digital requerida
- Usuarios: Desarrolladores y analistas de datos con capacidad para interactuar con agentes de IA y manejo básico de SQL. No es apto para usuarios de negocio sin acompañamiento técnico.
- Empresa: Organizaciones con arquitecturas de datos maduras (Data Warehouses o bases de datos relacionales bien estructuradas) y una estrategia activa de implementación de IA Generativa.

Plan orientativo de implantación
Pasos necesarios y estimaciones
- Tiempos estimados de despliegue: Entre 2 y 4 semanas para una implementación funcional estable.
- Evaluación inicial (3-5 días): Auditoría de fuentes de datos, permisos de lectura y definición de casos de uso (ej. asistente SQL para el equipo de desarrollo).
- Configuración técnica (5-7 días): Despliegue del servidor MCP mediante Docker o Node.js, configuración de variables de entorno y conexión con bases de datos (BigQuery, PostgreSQL, etc.).
- Prueba de concepto (1 semana): Validación de la precisión de las respuestas del LLM sobre esquemas reales en entornos de sandbox.
- Formación y despliegue (3-5 días): Capacitación técnica sobre el uso del protocolo en IDEs (Claude Desktop, VS Code) y monitorización inicial.

Necesidades de formación del equipo
Capacitación en el estándar Model Context Protocol (MCP), creación de prompts para consultas de datos, configuración de archivos YAML para lógica de negocio y gestión de observabilidad con OpenTelemetry.

Perfiles necesarios
- Perfiles técnicos necesarios: Ingeniero de Datos, Ingeniero de Prompting/IA, Administrador de Sistemas (DevOps).
- Personal externo recomendado: Consultores especializados en arquitectura de IA Generativa si la infraestructura de datos es legada o compleja.

Retorno de la inversión
- Tiempos: Reducción de hasta un 60% en el tiempo dedicado a la escritura manual de consultas SQL complejas y exploración de esquemas.
- Cómo medirlo, KPIs: Número de consultas exitosas realizadas por la IA sin intervención humana, reducción del tiempo de resolución de tickets técnicos y latencia de respuesta en aplicaciones que consumen datos.

Otros
El cambio de marca de Gen AI Toolbox a MCP Toolbox responde a la estandarización de la industria hacia el protocolo abierto impulsado por Anthropic y adoptado por Google, lo que garantiza interoperabilidad con herramientas futuras fuera del ecosistema Google. La seguridad debe gestionarse estrictamente mediante roles IAM para evitar que el LLM ejecute acciones de escritura o borrado no deseadas.

PREGUNTAS FRECUENTES

¿Qué es MCP Toolbox for Databases y cuál es su función principal?

Es un servidor de código abierto desarrollado por Google, basado en el Model Context Protocol (MCP), que actúa como puente técnico entre agentes de Inteligencia Artificial e infraestructuras de datos. Su función es permitir que modelos de lenguaje (LLMs) e IDEs interactúen con bases de datos empresariales mediante lenguaje natural, eliminando la necesidad de desarrollar código puente manual.

¿Bajo qué tipo de licencia se distribuye este software?

La herramienta se distribuye bajo la licencia Apache 2.0. Esto implica que es software de código abierto (Open Source), permitiendo su uso profesional y comercial, así como su modificación y distribución gratuita, siempre que se respeten los avisos de autoría originales.

¿Es posible descargar el código fuente desde repositorios públicos?

Sí, el proyecto es totalmente accesible y se encuentra alojado en GitHub, específicamente dentro del repositorio oficial de Google APIs bajo el nombre 'mcp-toolbox'. Antes de su cambio de nombre, el proyecto se localizaba como 'genai-toolbox'.

¿Qué bases de datos son compatibles con esta tecnología?

Ofrece soporte multibase de datos que incluye ecosistemas de Google Cloud como BigQuery, AlloyDB, Cloud SQL y Spanner, además de sistemas externos ampliamente utilizados en la industria como PostgreSQL, MySQL, MongoDB, Snowflake, Oracle y SQL Server.

¿Cómo gestiona la seguridad y la privacidad de los datos corporativos?

La herramienta integra medidas de seguridad avanzadas mediante la compatibilidad con IAM (Identity and Access Management) y sistemas de autenticación empresarial. Al ser un servidor que el profesional despliega en su propia infraestructura o nube, permite mantener el control sobre quién accede a los datos y qué consultas se ejecutan.

¿Cuál es el coste de implementación para una empresa?

La versión autogestionada no tiene costes de licencia por ser Open Source. No obstante, si se opta por la versión administrada dentro de Google Cloud, el coste estará sujeto al consumo de recursos y servicios de dicha plataforma. El profesional debe considerar también los costes operativos de la infraestructura donde se despliegue el contenedor.

¿Qué nivel de conocimientos técnicos se requiere para su puesta en marcha?

Para la instalación y configuración se requiere un perfil técnico alto, con experiencia en el manejo de terminales, archivos YAML, Docker y entornos como Node.js o Go. Para su uso operativo diario, el nivel es medio, requiriendo conocimientos de SQL y gestión de variables de entorno.

¿Es compatible con los principales frameworks de desarrollo de IA?

Sí, presenta integraciones nativas con ecosistemas de desarrollo líderes como LangChain, LlamaIndex, Genkit y Gemini CLI, facilitando la incorporación de datos estructurados en flujos de trabajo de IA existentes.

¿Ofrece capacidades de monitorización para entornos de producción?

Sí, cuenta con soporte nativo para OpenTelemetry. Esto permite a los ingenieros de datos y arquitectos de soluciones obtener métricas precisas y trazas de las interacciones entre los agentes de IA y las bases de datos, asegurando la observabilidad del sistema.

¿Para qué perfiles profesionales está desaconsejada esta herramienta?

No es una solución orientada a usuarios finales sin formación técnica ni a empresas que busquen exclusivamente herramientas de Business Intelligence (BI) visual. Su enfoque es estrictamente de ingeniería y desarrollo para la integración de agentes de IA con almacenamiento de datos.

CONTRATOS Y CONDICIONES

Principales recomendaciones
- Instalar y ejecutar el servidor MCP en un entorno controlado (preferiblemente VPC o contenedores aislados) para evitar la exposición directa de las bases de datos a internet.
- Aplicar el principio de "mínimo privilegio" en las credenciales de base de datos vinculadas: usar usuarios con permisos de solo lectura o limitados a procedimientos almacenados específicos para mitigar riesgos de inyección SQL mediante IA.
- Configurar filtros de seguridad en los archivos de definición YAML para restringir el acceso a tablas que contengan datos personales o sensibles.
- Supervisar las trazas de OpenTelemetry para auditar qué consultas genera la IA y quién las activa, garantizando la trazabilidad exigida por la normativa europea.
- Validar que las integraciones con LLMs de terceros (como Claude o servicios externos) no impliquen el envío de datos de producción para el entrenamiento de modelos sin un anexo de tratamiento de datos (DPA) firmado.

Ley de Inteligencia Artificial (AI Act)
- El MCP Toolbox actúa como un componente de infraestructura (middleware). No se clasifica per se como un sistema de IA de alto riesgo, pero su uso en sectores críticos (recursos humanos, banca, infraestructuras críticas) obliga a la empresa usuaria a realizar una evaluación de impacto.
- El desarrollador debe garantizar la transparencia: si el Toolbox se usa para alimentar un chatbot que interactúa con humanos, debe informarse claramente que la respuesta es generada por una IA basándose en datos reales.
- Al permitir que la IA ejecute acciones sobre bases de datos, aumenta el riesgo de "comportamiento no deseado". La empresa es legalmente responsable de las decisiones o acciones ejecutadas a través de este puente tecnológico.

Privacidad y protección de datos
- Responsabilidades: La empresa española actúa como Responsable del Tratamiento. Google, al ser software de código abierto ejecutado localmente o en su nube, solo actúa como Encargado si se contrata la versión gestionada en Google Cloud.
- Ubicación de los datos: Al ser un servidor que se despliega en la infraestructura propia, los datos no salen de la ubicación elegida por la empresa (por ejemplo, región de Madrid en Google Cloud o servidores locales).
- Transferencia internacional: El uso del software Open Source no implica transferencia de datos, pero la conexión con modelos de IA externos (como Gemini o Claude) sí puede implicar la salida de metadatos o fragmentos de datos a EE.UU. Se requiere verificar la adhesión al Marco de Privacidad de Datos UE-EE.UU. de los proveedores de LLM.
- Derechos ARCO: La arquitectura debe permitir la localización, rectificación o supresión de datos personales mediante las consultas ejecutadas por el protocolo MCP en cumplimiento con el RGPD.

Propiedad intelectual
- Propiedad de datos: Los datos almacenados y consultados pertenecen íntegramente a la empresa u organización usuaria.
- Propiedad del resultado/procesamiento: Los esquemas, lógicas de negocio definidas en YAML y las consultas optimizadas generadas por el uso de la herramienta pertenecen a la empresa. Al ser licencia Apache 2.0, las modificaciones al código fuente original deben conservar los avisos de autoría de Google, pero pueden usarse comercialmente.

Usos y prohibiciones
- Usos prohibidos: No debe utilizarse para eludir medidas de seguridad de bases de datos o para realizar scraping masivo de datos protegidos por derechos de propiedad intelectual de terceros sin autorización.
- Usos admitidos: Integración en entornos de desarrollo, automatización de consultas internas y soporte técnico mediante agentes de IA.

Seguridad y certificaciones
- Seguridad: Soporta autenticación mediante IAM (Identity and Access Management), lo que permite alinear el acceso a los datos con la política de seguridad corporativa.
- Certificaciones: La herramienta es compatible con entornos que cumplen SOC2, ISO 27001 e HIPAA si se despliega sobre la infraestructura de Google Cloud que ya cuenta con estas certificaciones. Al ser Open

Source, la certificación final depende de la implementación del usuario.

Otros
- Cambio de marca: El paso de genai-toolbox a mcp-toolbox implica un cambio en las rutas de los repositorios; es vital actualizar las dependencias para recibir parches de seguridad críticos.
- Licencia Apache 2.0: Es una licencia permisiva que protege a la empresa española de reclamaciones de patentes, otorgando libertad total de integración en productos propietarios siempre que se cumplan los requisitos de atribución.

Fuentes consultada:
- Sitio web oficial
- Repositorio de código y Licencia Apache 2.0
- Documentación técnica oficial
- Términos de servicio de Google Cloud (para versión gestionada)
- Seguridad y cumplimiento en Google Cloud