

Aplicaciones LLM

Arquitectura de Microservicios

Problema

- Alojamiento de una Aplicación LLM en producción.
- La Arquitectura Monolítica presenta varios problemas:
 - Difícil de mantener.
 - Difícil debugging.
 - Difícil de escalar.
 - Uso ineficiente de recursos.

Solución

- Arquitectura de Microservicios.
 - Alojarse el programa en varias partes separadas.
 - Docker.
 - Microservicios.
 - Kubernetes.

Proceso

- No es sencillo.
- Normalmente no corre a cargo del desarrollador de la Aplicación LLM.
- En este capítulo nos limitaremos a este enfoque conceptual.

Otros

- Más info en la documentación de Docker.