

Infraestructura en producción



Infraestructura para producción

- Es una consideración que debes tener en mente desde el inicio del proyecto
- Se refiere al hardware y software que soporta la ejecución del modelo en producción



Cron

- Herramienta de consola disponible en sistemas basados en Unix: Mac y Linux
- Agenda trabajos para ser ejecutados a fechas e intervalos definidos
- Usado comúnmente para tareas administrativas o de mantenimiento



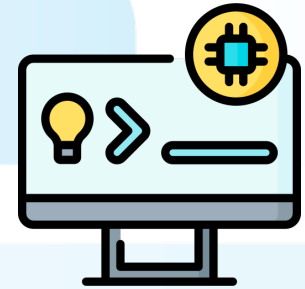
Cron

- Una vez al día, de lunes a viernes a las 10:30 de la noche:

- `30 10 * * 1-5`

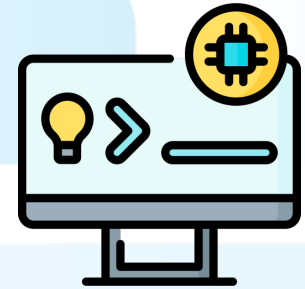
- Al minuto 10, después de las 2, 6, 10, 14, 18 y 22 en los días 8-14 del mes de diciembre:

- `10 2,6,10,14,18,22 8-14 12 *`



Ventajas de Cron

- **Sencillos de crear y mantener:** no se necesitan conocimientos avanzados para crear uno
- **Universalidad:** están disponibles en múltiples plataformas



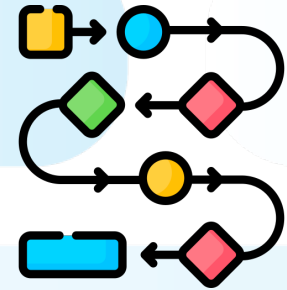
Desventajas de Cron

- **Se ejecutan en una sola computadora:** son locales; coordinar ejecuciones distribuidas se vuelve muy complicado
- **No es monitoreable:** tienen opciones limitadas para ser monitoreados y encontrados



Workflow managers

- Permiten definir tareas programadas, con dependencias entre ellas
- Se generan pipelines de ML
- Útiles para programar ejecuciones de inferencia en lotes



Workflow managers



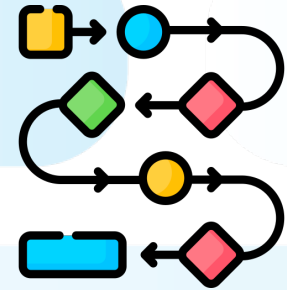
Apache
Airflow



PREFECT

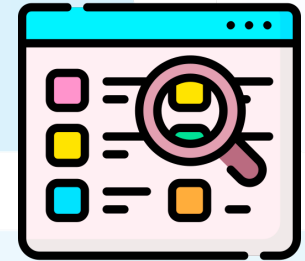


dagster



Ventajas de Workflow Managers

- **Mejor monitoreo y administración:**
usualmente son aplicaciones que tienen una interfaz gráfica
- **Más comprensivos:** permiten la orquestación de pipelines más complejos de datos



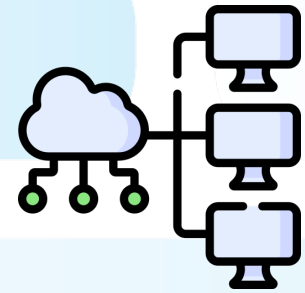
Desventajas de Workflow Managers

- **Curva de aprendizaje:** es otra herramienta que hay que dominar para manejar el sistema completo
- **Infraestructura extra:** añade un componente extra al sistema que hay que mantener



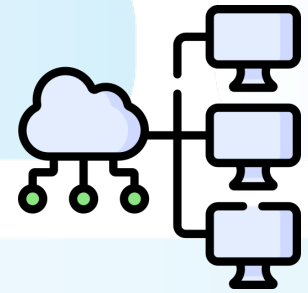
Microservicios

- Los microservicios son pequeñas aplicaciones que realizan una única tarea.
- Un modelo de machine learning puede ser desplegado como un microservicio.
- El microservicio recibe peticiones, procesa la información y entrega una predicción al usuario.

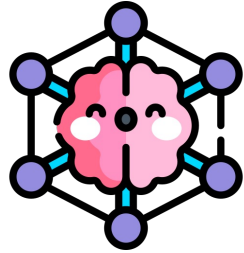


Microservicios

- En machine learning, un microservicio actúa como un servidor web que siempre está a la escucha de peticiones.
- El microservicio es comúnmente utilizado para inferencia en tiempo real o en línea.



Ejemplo de microservicios



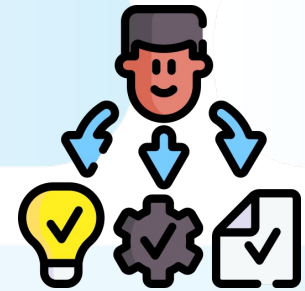
Ventajas de Microservicios

- El microservicio puede atender a múltiples usuarios simultáneamente y puede ser escalado según la demanda.



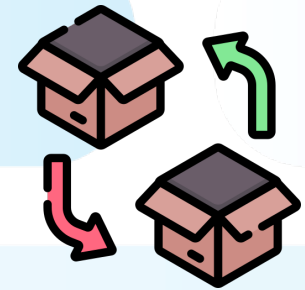
Ventajas de Microservicios

- El modelo de machine learning puede ser consultado por cualquier usuario o servicio que necesite hacer inferencias.



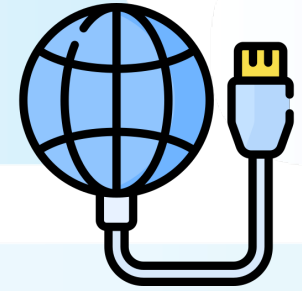
Ventajas de Microservicios

- Es posible actualizar el modelo creando una nueva versión y redirigiendo el tráfico hacia ella.



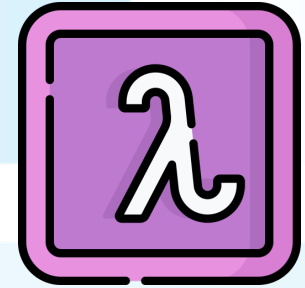
Desventajas de Microservicios

- Un microservicio es otra pieza del sistema que hay que mantener
- La comunicación entre cliente y servidor se vuelve tu problema, es importante para una buena calidad del servicio



Serverless / Stateless

- Una forma de poner en producción un modelo es a través de funciones sin estado.
- Las funciones sin estado solo existen mientras están siendo usadas.
- Las funciones sin estado son ideales para montar modelos de inferencia online.



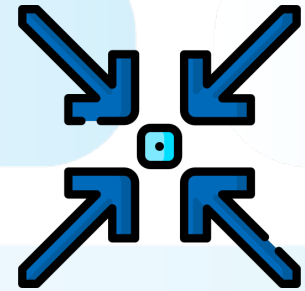
Ventajas Serverless

- Un despliegue *serverless* permite reducir los costos operativos y la manutención al no requerir un servidor permanentemente en línea.
- Los modelos desplegados así pueden escalar fácilmente para ser llamados miles de veces simultáneamente sin afectar su desempeño.



Desventajas Serverless

- Impone restricciones al modelo: debe ser pequeño y tener una rápida capacidad de inicio.
- Es necesario considerar el problema del "cold start".
- Depurar código desarrollado bajo este paradigma puede ser complicado.



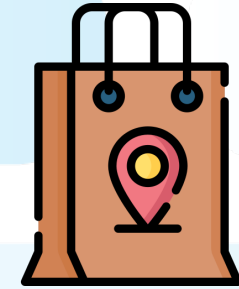
On-edge

- Implica distribuir modelos directamente en dispositivos de usuarios finales.
- Se utiliza en entornos industriales, con sistemas embebidos: sensores, microchips, dispositivos personales.
- Los datos se procesan localmente, en lugar de enviarlos a un servidor remoto o a la nube.



On-edge

- Implica distribuir modelos directamente en dispositivos de usuarios finales.
- Se utiliza en entornos industriales, con sistemas embebidos: sensores, microchips, dispositivos personales.



Ventajas *On-edge*

- La experiencia de los usuarios mejora con esta forma de despliegue.
- Se evita el envío de información sensible a través de la red.
- Se evita el mantenimiento de un servicio extra.



Desventajas *On-edge*

- Presenta complicaciones con falta actualizaciones del modelo por parte del usuario final.
- Requiere hardware especializado para problemas complejos, lo cual puede comprometer el desempeño.

