1. **El nombre de la investigación**

Análisis y Predicción del Consumo de Recursos en Entornos de Computación en la Nube.

Data set: [GitHub - alibaba/clusterdata: cluster data collected from production clusters in Alibaba for cluster management research](https://github.com/alibaba/clusterdata)

1. **La descripción de la base de datos a utilizar.**

Alibaba Cluster Trace Dataset, contiene registros reales de uso de recursos a gran escala, lo que nos permitirá entrenar modelos predictivos basados en patrones temporales, tipos de cargas de trabajo y escalabilidad horizontal. Al analizar estos datos, buscamos anticipar picos de demanda, ajustar automáticamente la asignación de workers en entornos en la nube y garantizar una transición fluida hacia la nube sin sacrificar estabilidad ni eficiencia económica.

1. **El motivo de la selección de los datos**

Actualmente estamos implementando una migración de sistemas donde todos nuestros procesos van a cambiar a un entorno en la nube y al ser escalable en recursos (CPU, memoria, workers) permitirá que la optimización de costos y el rendimiento sean críticos. Este proyecto surge de la necesidad de predecir y gestionar eficientemente el consumo de recursos en un entorno dinámico, donde la sobre provisión o subutilización de capacidades impacta directamente en la facturación operativa y la calidad del servicio.

1. **Una breve introducción del contexto**

En la era de la transformación digital, la adopción de entornos cloud escalables ha revolucionado la gestión de recursos informáticos. Sin embargo, este paradigma introduce desafíos críticos: cómo equilibrar la demanda fluctuante de recursos (CPU, memoria, workers) con la eficiencia operativa y económica. En el entorno empresarial su aplicación facilita la toma de decisiones basadas en evidencia: desde ajustar dinámicamente la escalabilidad de workers en plataformas como en la nube hasta evitar costosos escenarios de sobre provisión.

1. **Línea de tiempo de investigación**

|  |  |
| --- | --- |
| **Fecha** | **Tarea** |
| 27-mar-25 | * Selección del dataset. |
| 29-mar-25  30-mar-25 | * Planteamiento del problema * Análisis y selección de modelo predictivo * Limpieza de los datos * (EDA) Análisis exploratorio de datos * Preprocesamiento de datos * Aplicación de modelos predictivos |
| 31-mar-25 | * Realizar pruebas. |
| 5-abril-25 | * Inicio del documento del proyecto |
| 6-abril-25 | * Finalizar documento del proyecto. * Generar material apoyo para la presentación. |