



# Universidad Autónoma del Estado de México

## UAEM

### *Diplomado Superior en Ciencia y Analítica de Datos* *Práctica 5.* *Apache Spark: Operadores RDD en juegos Olímpicos*

**Unidad de Aprendizaje:** Big Data

**Fecha de Entrega:** 27 de abril 2024 antes de las 9:00 hrs

**Docente:** Dra. en C. Ing. Angélica Guzmán Ponce

---

## Objetivo

Utilizar las transformaciones y acciones en RDD vistas en clase para realizar un análisis estadístico y descriptivo de los resultados de los eventos olímpicos.

## Actividades

### 1. Carga de datos

Carga los archivos **deportista.csv**, **deportista2.csv**, **países.csv**, **evento.csv**, **juegos.csv** y **resultados.csv** en sus respectivos RDDs. Realiza las transformaciones necesarias para trabajar con ellos.

### 2. Deportistas y Países

Asocia cada deportista con su país de origen.

### 3. Eventos y Juegos

Agrupar los eventos por juego y contar cuántos eventos se llevaron a cabo en cada edición de los juegos.

### 4. Relación entre eventos, deportes y resultados

Cuenta el número de medallas ganadoras por deporte y por juego, considera las siguientes puntuaciones:

$$\text{valoresMedallas} = \{ 'Gold':7, 'Silver':5, 'Bronze':4 \}$$

Analiza cómo ha variado la cantidad de atletas y países participantes a lo largo de los juegos.

### 5. Entrega

- Deberás crear un notebook de Google Colab que contenga las respuestas a los ejercicios.

- El notebook debe estar bien organizado, con secciones claramente comentadas para cada ejercicio y los pasos que estás realizando. La claridad y la correcta documentación son tan importantes como la exactitud de tu código.
- Requisitos del Notebook:
  - Nombre del Notebook: Tu notebook de Google Colab debe seguir el formato de nomenclatura:  
**Apellidos\_Nombre\_PracticaRDDs.ipynb**  
Para facilitar su identificación y calificación.
  - Estructura y Comentarios: Cada ejercicio debe estar en una sección separada y claramente identificada con una celda de texto que contenga un título utilizando Markdown. Por ejemplo, usa **### Ejercicio Análisis de Eventos y Juegos**.
  - Código Comentado: Dentro de cada sección, proporciona comentarios en el código que expliquen cada operación PySpark que realices. Los comentarios deben ser concisos pero informativos para indicar qué y por qué estás realizando cada paso.
  - Al final del notebook, agrega una celda de texto con tus reflexiones sobre los resultados del análisis y cualquier observación interesante que hayas descubierto.

Criterio	Porcentaje	
Código	60 %	Tu código debe ejecutarse sin errores y producir los resultados esperados para cada ejercicio.
Comentarios	30 %	Tus comentarios deben explicar claramente tu razonamiento y las operaciones realizadas.
Reflexión y Análisis	10 %	Tus conclusiones y análisis deben demostrar comprensión de los datos y de los resultados obtenidos.

Tabla 1: Evaluación y Criterios de Calificación