

Operaciones de agregación

Usando Pymongo

Definición.

Se llama operaciones de agregación a aquellas que procesan múltiples documentos y regresan resultados “**computados**”, por computar entendemos que se realizan cálculos u operaciones sobre el conjunto de datos.

Usos

La agregación se usa principalmente para:

- Agrupar datos de múltiples documentos.
- Realizar operaciones sobre los datos agrupados para obtener un único resultado (Promedio de valores, desviación estándar, sumas, etc.).
- Analizar cambios en los datos a través del tiempo.

¿Cómo se pueden hacer las agregaciones?

1. A través de pipelines (pilas de trabajos consecutivos).
2. Métodos de un sólo propósito, simples pero carentes del poder de un pipeline.

Pipelines de agregación

Un pipeline de agregación consiste de una o, más etapas de procesamiento de documentos:

- Cada etapa realiza una operación en los documentos ingresados.
- Los documentos resultantes son pasados a la siguiente etapa.
- Un pipeline de agregación puede devolver múltiples resultados para grupos de documentos, como total, promedio, etc.

Agregación de un sólo propósito

Los métodos de agregación de un sólo propósito trabajan en una sola colección, estos métodos como mencionamos con anterioridad son simples, pero limitados.

Métodos de agregación de un sólo propósito

- **estimatedDocumentCount():** Regresa una aproximación de la cantidad de documentos en una colección.
- **count():** Regresa la cuenta de documentos en una colección.
- **distinct():** Regresa un array de documentos que tienen un valor distinto para el campo especificado.

Algunas etapas del pipeline

El pipeline tiene diversas etapas posibles, a continuación se describen algunas de éstas opciones

- **Match:** Filtrado de documentos por diferentes atributos
- **Group:** Agrupamiento de documentos por diferentes atributos.
- **Unwind:** Desempaquetado de datos anidados.
- **Project:** Creación o eliminación de campos.
- **Sort:** Ordenamiento de datos por uno o más parametros.
- **Limit:** Limitación de resultados devueltos.
- **Count:** Regresa la cuenta de documentos en el stage

Algunas etapas del pipeline

- **Merge:** Inserta los resultados en una colección, en la misma base de datos, o en otra base de datos. (Debe ser el último paso del pipeline)
- **Out:** Similar a **Merge**, pero con menos opciones.
- **Lookup:** Ejecuta una unión externa izquierda ($A \cup (A \cap B)$?) con una colección dentro de la misma base de datos.

Operadores aritméticos que se pueden usar en el pipeline.

- **abs:** Valor absoluto
- **add:** Añade números para regresar la suma o fechas y números (tratados como ms)
- **ceil:** Regresa el menor entero mayor o igual que el valor dado.
- **divide:** Regresa el resultado de dividir el primer número por el segundo.
- **exp:** Eleva **e** al exponente dado.
- **floor:** Regresa el mayor entero menor o igual al valor dado.
- **In:** Regresa el Logaritmo Natural del valor dado.
- **log:** Regresa el Logaritmo del valor dado en la base dada.
- **log10:** Regresa el Logaritmo del valor dado en base 10.
- **mod:** Regresa el residuo de la división de el primer número entre el segundo.

Operadores aritméticos que se pueden usar en el pipeline.

- **multiply:** Regresa el producto de todos los parámetros entregados.
- **pow:** Eleva un número a la potencia dada.
- **round:** Redondea un número a un entero, o a una cantidad especificada de decimales.
- **sqrt:** Calcula la raíz cuadrada del valor dado.
- **subtract:** Regresa el resultado de sustraer el primer valor del segundo.
- **trunc:** Lo mismo que **round**.