

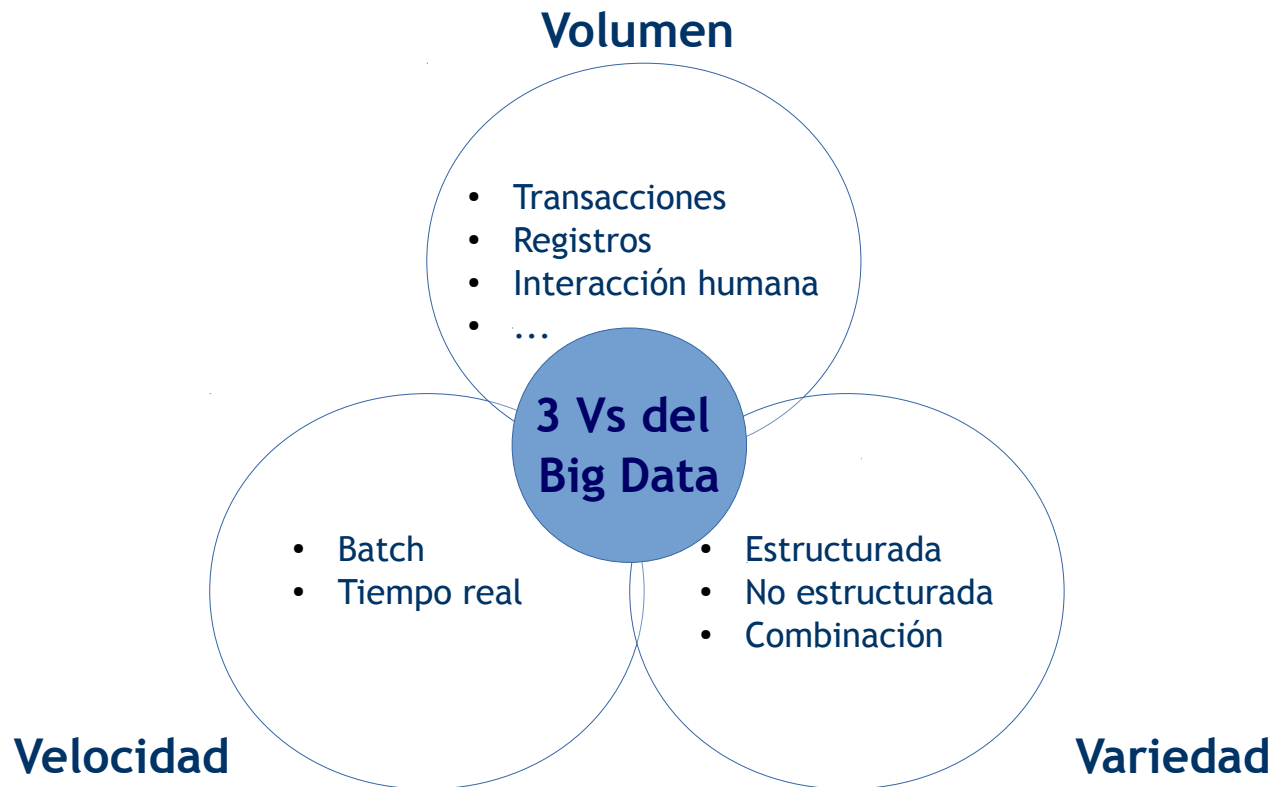
Big Data



www.logitravel.com

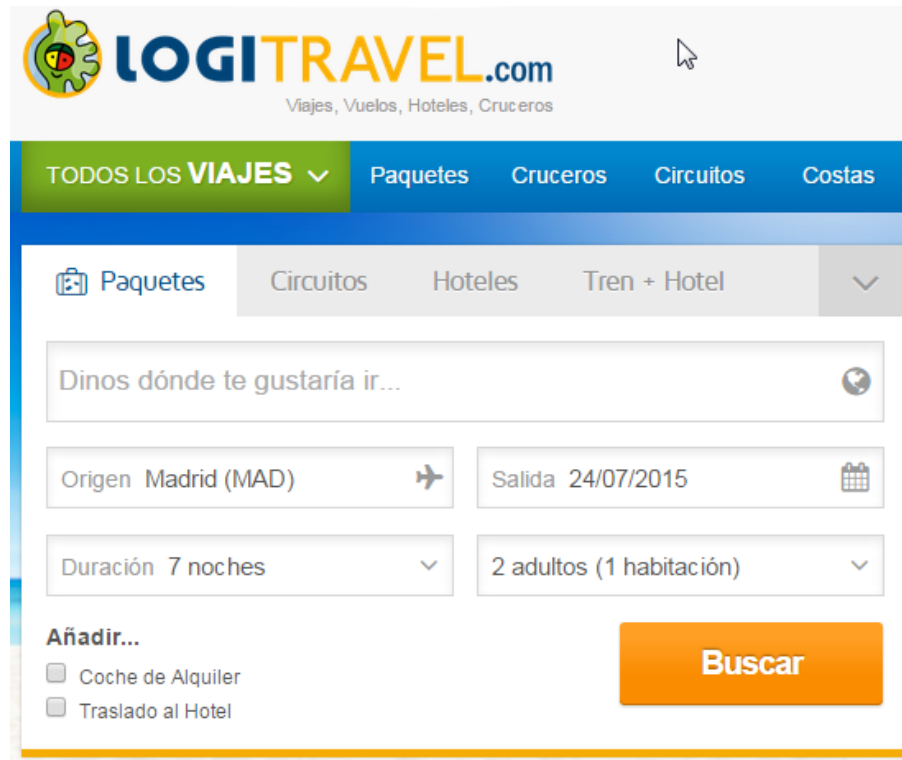
Definición

Se podría definir Big Data de manera simple como el conjunto de técnicas que nos permite **capturar, transformar, almacenar, analizar y visualizar** un enorme volumen y variedad de información de manera rápida.



Aplicaciones en Logitravel

- Ofrecer una experiencia personalizada para cada usuario que navega en nuestra web.




The screenshot displays the Logitravel.com website interface. At the top, the logo features a green alien head with a gear-like helmet next to the text "LOGITRAVEL.com" in blue and orange, with the tagline "Viajes, Vuelos, Hoteles, Cruceros" below it. A navigation bar includes "TODOS LOS VIAJES" (highlighted in green) and links for "Paquetes", "Cruceros", "Circuitos", and "Costas". Below this, a sub-navigation bar shows "Paquetes" (selected), "Circuitos", "Hoteles", and "Tren + Hotel". The main search form contains a text input field "Dinos dónde te gustaría ir..." with a globe icon. Below it are two rows of input fields: "Origen Madrid (MAD)" with an airplane icon and "Salida 24/07/2015" with a calendar icon. The next row has "Duración 7 noches" with a dropdown arrow and "2 adultos (1 habitación)" with a dropdown arrow. At the bottom left, under "Añadir...", there are checkboxes for "Coche de Alquiler" and "Traslado al Hotel". A large orange "Buscar" button is positioned at the bottom right of the form.




Aplicaciones en Logitravel

- Presentar las mejores ofertas de nuestros productos.



Viajes

al mejor precio




Riviera Maya 7 noches

Vuelo+Hotel+Traslado ★★★★★

Todo incluido

Desde

830€




Punta Cana - Bávaro 7 noc...

Vuelo+Hotel+Traslado ★★★★★

Todo incluido

Desde

702€




Tenerife 7 noches

Vuelo+Hotel ★★★★★ Desayuno

Desde

408€




Lanzarote 7 noches

Vuelo+Hotel ★★★★★ Sólo Alojamiento

Desde

409€



Hoteles

Ofertas de Hoteles en Julio

Roma

Acropoli ★★★★★ Alojamiento y Desayuno

desde

16€

Madrid

Chamartín ★★★★★ Alojamiento y Desayu...

desde

24€

Londres

Royal National ★★★★★ Alojamiento y De...

desde

67€

París

Victoria ★★★★★ Alojamiento y Desayuno

desde

38€

Barcelona

Senator Barcelona Spa Hotel ★★★★★ Só...

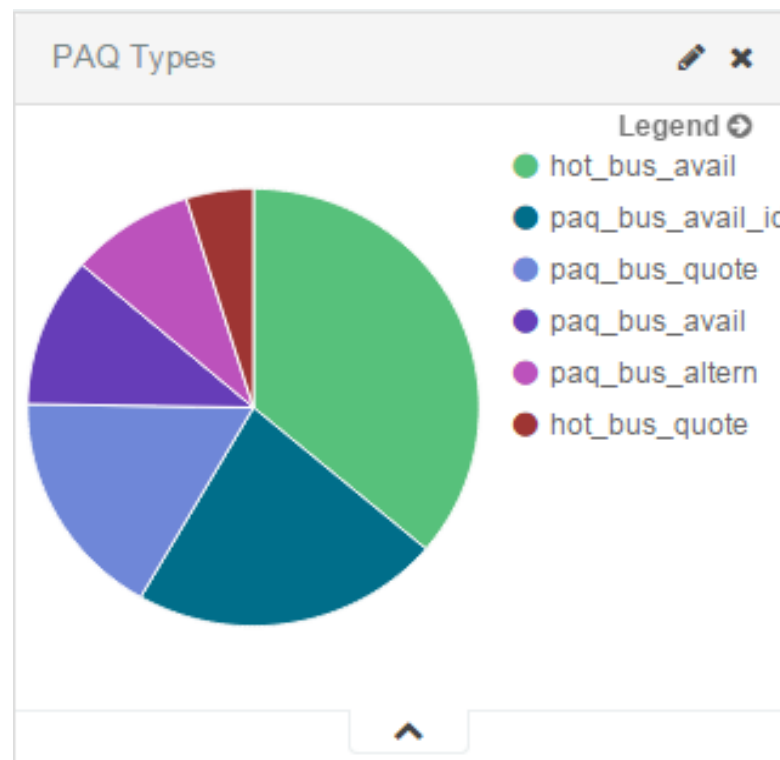
desde

30€



Aplicaciones en Logitravel

- Analizar el funcionamiento de nuestras aplicaciones.



Herramientas

Procesamiento



Visualización



Almacenamiento



MongoDB: Características

- Base de datos NoSQL. Las bases de datos están compuestas por colecciones. No siguen ningún esquema.
- Nos permite llevar a cabo inserciones más rápidamente que en SQL. No esperamos confirmación de si la inserción ha ido bien o no.
- Las operaciones de consulta por campos indexados son muy rápidas. Pese a persistir los datos en disco, mantiene toda la información que puede en memoria.
- Trabaja con formato BSON, tanto para guardar los documentos como para definir las consultas.



MongoDB: Formato BSON

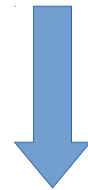
- BSON: Binary JSON. Documentos tipo JSON serializados en codificación binaria.

```
{
  "_id" : ObjectId("55a8b244390bc5206627b55c"),
  "SearchDate" : ISODate("2015-07-14T10:18:54.514Z"),
  "Code" : 1,
  "Name" : "Hotel Valparaíso",
  "Category" : 3,
  "Board" : "SA",
  "City" : 86,
  "Popularity" : 6,
  "Price" : 51
}
```



MongoDB: Recuperaciones

```
db.queueCollection.find({"city": 8727,  
"pop": 3})
```



```
...  
DBObject query = new BasicDBObject("city", 8727);  
query.put("pop", 3);  
this.queueCollection.find(query);  
...
```



MongoDB: Inserciones

```
db.finalEventsCollection.insert({"text": "Test text"})
```



```
...  
BasicDBObject dbo = new BasicDBObject("text", "Test text");  
this.finalEventsCollection.insert(dbo);  
...
```

Para insertar o actualizar un documento completo podemos usar la instrucción **save**. En caso que el documento a guardar tenga un **_id**, si este ya existe en la colección, se sobrescribirá. Si no existe, se insertará.



MongoDB: Otras operaciones

- Remove. Permite eliminar elementos de una colección que cumplan ciertos criterios.
- Update. Permite modificar completamente o parcialmente elementos de una colección que cumplan determinados criterios.
- FindAndModify. Permite, en una sola transacción, encontrar un elemento, modificarlo en la colección, y devolver el elemento original encontrado.
- FindAndRemove. Permite, en una sola transacción, encontrar un elemento, borrarlo de la colección, y devolverlo. Muy útil para cuando tratamos las colecciones como una estructura de cola.



MongoDB: Ejercicio

- Ejercicio 1: usando Robomongo, insertar documento con la siguiente estructura en la base de datos test, colección hotel_prices_test:

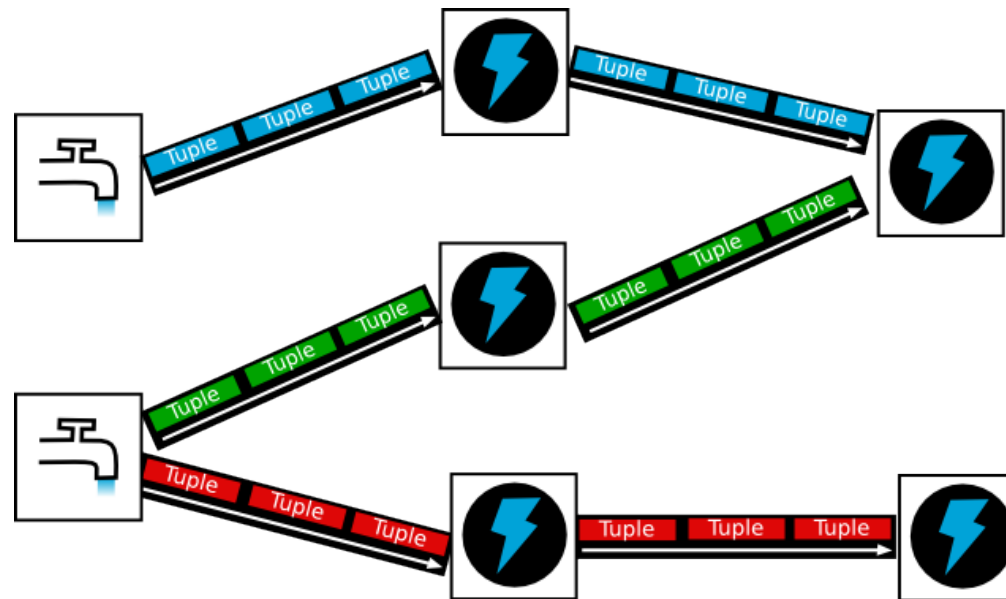
```
{  
  "hotel" : "Hotel 2",  
  "date" : ISODate("2015-07-04T15:01:04.201Z"),  
  "board" : "SA",  
  "city" : 2,  
  "category" : 5,  
  "pop" : 8,  
  "price" : 200,  
  "currency" : "EUR"  
}
```

- Ejercicio 2: Cambiar el campo hotel del documento insertado usando el método save. Deberéis pasar el documento completo actualizado al método.



Storm

Storm es un framework que nos provee un sistema de computación distribuida en tiempo real. Nos permite procesar streams de información de forma muy sencilla, además de garantizar que la información será siempre procesada.



Storm: Términos básicos

- Topología. Es el componente que construimos usando el framework Storm. Una topología es el software que procesará los flujos de información, y está compuesta por dos tipos de componentes: Spouts y Bolts.
- Spout. Es un tipo de componente cuya única misión es leer el flujo de datos (por ejemplo leer las filas de una colección de MongoDB), y *emitir* tan rápido como sea posible esa información para que empiece a ser procesada.
- Bolt. Este componente recibe tuplas de otros componentes, las procesa, y puede *emitir* el resultado a otros componentes de la topología, o por ejemplo, guardar el resultado final en una base de datos.

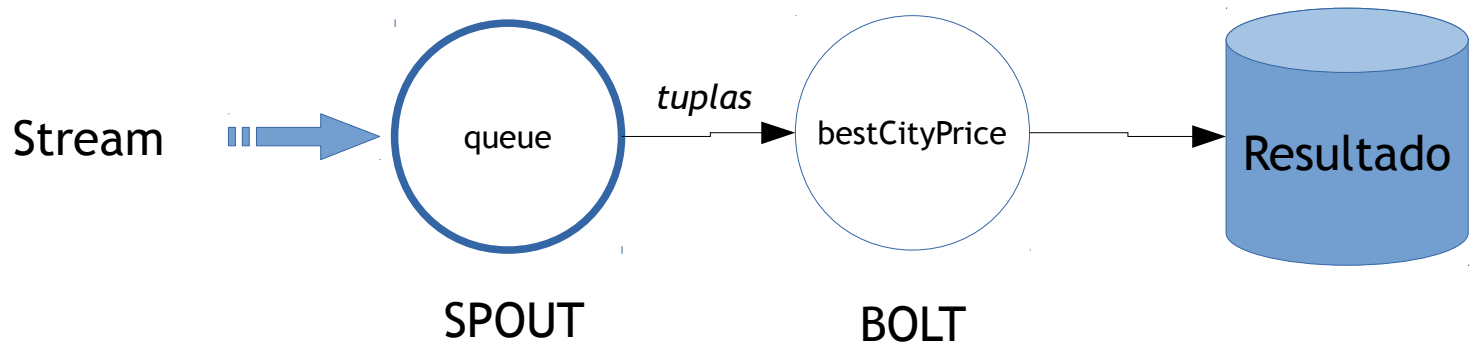


Storm: Ejemplo de topología básica

```
public static void main( String[] args )
{
    TopologyBuilder builder = new TopologyBuilder();

    // Set the spout in the topology
    builder.setSpout("queue", new PricesQueue());

    // Build your topology here
    builder.setBolt("bestCityPrice", new BestCityPrice())
        .shuffleGrouping("queue");
}
```

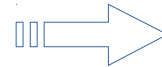


Storm: Ejercicio

Construir una topología que, tomando como stream de entrada la colección hotel_searches de la base de datos searches, genere una colección final donde cada documento represente el mejor precio del hotel más popular de una ciudad.

```
{
  "_id" : ObjectId("55a8b244390bc5206627b55c"),
  "SearchDate" : ISODate("2015-07-14T10:18:54.514Z"),
  "Code" : 1,
  "Name" : "Hotel Valparaiso",
  "Category" : 3,
  "Board" : "SA",
  "City" : 86,
  "Popularity" : 6,
  "Price" : 51
}
```

Tupla a procesar



```
{
  "_id" : 86,
  "SearchDate" : ISODate("2015-07-14T10:18:54.514Z"),
  "Name" : "Hotel Valparaiso",
  "Category" : 3,
  "Board" : "SA",
  "Popularity" : 6,
  "Price" : 51
}
```

Resultado

