

NoSQL - Document Databases

Gerardo Rossel



DEPARTAMENTO
DE COMPUTACION

2017

Document Database

Definición

Document Database

Es una base no-relacional que almacena los datos como documentos estructurados.

El concepto principal es el **documento**

- Las BD almacena y recupera documentos.
- Los documentos pueden ser XML, **JSON**, BSON, etc

Ejemplos

MongoDB, RavenDB, eXist, CouchDB, CouchBase

XML vs Json

```
<order id="1234">
  <customer id="52">Adam Fowler</customer>
  <items>
    <item qty="2" id="456" unit_price="2.00" price="4.00">Hammer</item>
    <item qty="1" id="111" unit_price="0.79" price="0.79">Hammer Time</item>
  </items>
  <delivery_address lon="-43.24" lat="54.12">
    <street>Some Place</street>
    <town>My City</town>
    . . . . .
  </delivery_address>
</order>
```

```
{
  "orderid": 1234,
  "Customer": { "id": 52, "Nombre": "Jhon Doe" },
  "items": [ { "qty": 2, "id": 456, "unit_price": 2, "price": 4 },
    { "qty": 1, "id": 111, "unit_price": 0.79, "price": 0.79 }
  ],
  "delivery_address": {
    "lon": -43.24, "lat": 54.12,
    "street": "Some Place",
    "ciudad": "My Ciudad"
  }
}
```

Colecciones

- MongoDB agrupa documentos en colecciones.
 - No es necesario que tengan la misma estructura.
 - Decisión de diseño: ¿como agrupar documentos en colecciones?
- RavenDB
 - Una colección es una forma de hablar de todos los documentos que comparten una misma marca de tipo
 - Cuando se utiliza la API de .Net la marca de tipo se realiza automáticamente inferida del nombre de la clase del objeto que se guarda.
- RethinkDB
 - Los documentos se agrupan en "tablas"
 - Hay "streams" y "selections"
- CouchDB/CouchBase
 - Identificar tipo por un *doc_type*.
 - CouchBase: *Data Bucket*(no es lo mismo que colecciones)

Diseño

Desnormalización

```
{  
  order_item_ID : 834838,  
  order_ID: 8827,  
  quantity: 3,  
  cost_per_unit: 8.50,  
  product_ID: 3648  
}
```

```
{  
  product_ID: 3648,  
  product_description: "1 package laser printer paper.  
    100% recycled.",  
  product_name : "Eco-friendly Printer Paper",  
  product_category : "office supplies",  
  list_price : 9.00  
}
```

Desnormalización

```
{  
  order_item_ID : 834838,  
  order_ID: 8827,  
  quantity: 3,  
  cost_per_unit: 8.50,  
  product_ID: 3648  
}
```

```
{  
  product_ID: 3648,  
  product_description: "1 package laser printer paper.  
    100% recycled.",  
  product_name : "Eco-friendly Printer Paper",  
  product_category : "office supplies",  
  list_price : 9.00  
}
```

Desnormalizacion

Desnormalización

```
{  
  order_item_ID : 834838,  
  order_ID: 8827,  
  quantity: 3,  
  cost_per_unit: 8.50,  
  product_ID: 3648  
}
```

```
{  
  product_ID: 3648,  
  product_description: "1 package laser printer paper.  
    100% recycled.",  
  product_name : "Eco-friendly Printer Paper",  
  product_category : "office supplies",  
  list_price : 9.00  
}
```



Desnormalizacion

Desnormalización

```
{  
  order_item_ID : 834838,  
  order_ID: 8827,  
  quantity: 3,  
  cost_per_unit: 8.50,  
  product_ID: 3648  
}
```

```
{  
  product_ID: 3648,  
  product_description: "1 package laser printer paper.  
    100% recycled.",  
  product_name : "Eco-friendly Printer Paper",  
  product_category : "office supplies",  
  list_price : 9.00  
}
```



Desnormalizacion

```
{  
  order_item_ID : 834838,  
  order_ID: 8827,  
  quantity: 3,  
  cost_per_unit: 8.50,  
  product :  
  {  
    product_description: "1 package laser printer  
      paper. 100% recycled.",  
    product_name : "Eco-friendly Printer Paper",  
    product_category : "office supplies",  
    list_price : 9.00  
  }  
}
```

Desnormalización

● ¿Cuanto desnormalización es demasiado?

- Generar facturas y remitos para los clientes (95 %)
- Generar reportes para la gerencia (5 %)

```
{
  order_item_ID : 834838,
  order_ID: 8827,
  quantity: 3,
  cost_per_unit: 8.50,
  product :
    {
      product_description: "1 package laser printer
        paper. 100% recycled.",
      product_name : "Eco-friendly Printer Paper",
      product_category : "office supplies",
      list_price : 9.00
    }
}
```

```
{
  product_description: "1 package laser printer paper.
    100% recycled.",
  product_name : "Eco-friendly Printer Paper",
  product_category : 'office supplies',
  list_price : 9.00
}
```

Desnormalización

● ¿Cuanto desnormalización es demasiado?

- Generar facturas y remitos para los clientes (95 %)
- Generar reportes para la gerencia (5 %)

```
{
  order_item_ID : 834838,
  order_ID: 8827,
  quantity: 3,
  cost_per_unit: 8.50,
  product :
  {
    product_description: "1 package laser printer
    paper. 100% recycled.",
    product_name : "Eco-friendly Printer Paper",
    product_category : "office supplies",
    list_price : 9.00
  }
}
```

```
{
  product_description: "1 package laser printer paper.
  100% recycled.",
  product_name : "Eco-friendly Printer Paper",
  product_category : 'office supplies',
  list_price : 9.00
}
```




```
{
  order_item_ID : 834838,
  order_ID: 8827,
  quantity: 3,
  cost_per_unit: 8.50,
  product_name : "Eco-friendly Printer Paper"
}
```

Documentos mutables

```
{  
  truck_id: 'T87V12',  
  time: '08:10:00',  
  date : '27-May-2015',  
  driver_name: 'Jane Washington',  
  fuel_consumption_rate: '14.8 mpg',  
  ...  
}
```

Documentos mutables

```
{  
  truck_id: 'T87V12',  
  time: '08:10:00',  
  date : '27-May-2015',  
  driver_name: 'Jane Washington',  
  fuel_consumption_rate: '14.8 mpg',  
  ...  
}
```



```
{  
  truck_id: 'T87V12',  
  date : '27-May-2015',  
  driver_name: 'Jane Washington',  
  operational_data:  
    [  
      {time : '00:01',  
       fuel_consumption_rate: '14.8 mpg',  
       ...},  
      {time : '00:04',  
       fuel_consumption_rate: '12.2 mpg',  
       ...},  
      {time : '00:07',  
       fuel_consumption_rate: '15.1 mpg',  
       ...},  
      ...]  
    ]  
}
```

Documentos mutables

```
{
  truck_id: 'T87V12',
  time: '08:10:00',
  date : '27-May-2015',
  driver_name: 'Jane Washington',
  fuel_consumption_rate: '14.8 mpg',
  ...
}
```

```
{
  truck_id: 'T87V12',
  date : '27-May-2015',
  driver_name: 'Jane Washington',
  operational_data:
    [
      {time : '00:01',
        fuel_consumption_rate: '14.8 mpg',
        ...},
      {time : '00:04',
        fuel_consumption_rate: '12.2 mpg',
        ...},
      {time : '00:07',
        fuel_consumption_rate: '15.1 mpg',
        ...},
      ...]
}
```

200 Embedded
Documents with
Default Values

```
{truck_id: 'T8V12'
 date: '27-May-2015'
 operational_data:
  [[{time: '00 : 00',
      fuel_consumption_rate: 0.0}
    {time: '00 : 00',
      fuel_consumption_rate: 0.0}
    .
    .
    .
    {time: '00 : 00',
      fuel_consumption_rate: 0.0}
  ]
}
```

Considerar el ciclo de vida

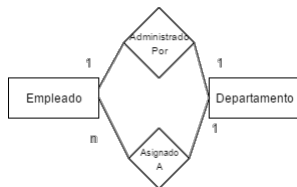
Modelo Conceptual -> DID -> Documentos

- DER - Modelo conceptual de alto nivel.
- DID (Modelo/Diagrama de Interrelación de Documentos).
- JSON Schema: especificación de la estructura de los documentos.

Embeber o Referenciar

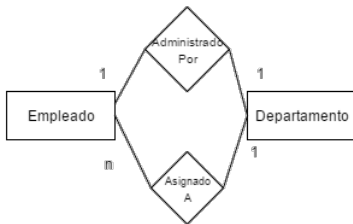
Una de las decisiones más importantes es si embeber o referenciar, lo que determinará el grado de desnormalización de los documentos

Cardinalidad 1 a N / 1 a 1

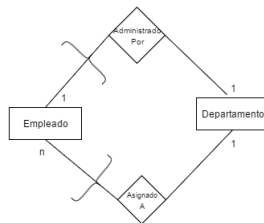


- Embeber el departamento en el empleado
- Embeber los empleados en el departamento
- Referenciar los empleados y embeber el departamento en empleado.
- Referenciar de ambos lados
- Embeber de ambos lados
- etc, etc...

Cardinalidad 1 a N / 1 a 1

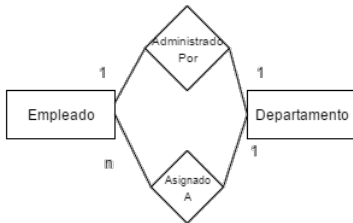


Modelo conceptual: DER

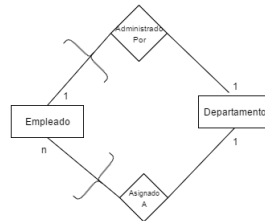


DID: Alternativa 1 Todo Embebido en Depto

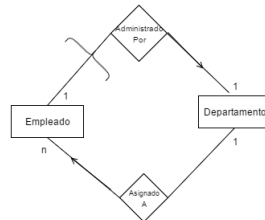
Cardinalidad 1 a N / 1 a 1



Modelo conceptual: DER

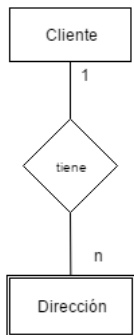


DID: Alternativa 1 Todo Embebido en Depto

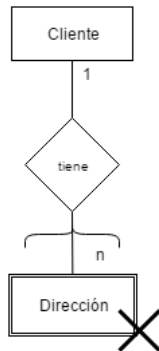
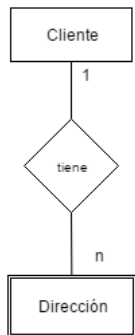


DID: Alternativa 2 - Embeber sólo Gerente.

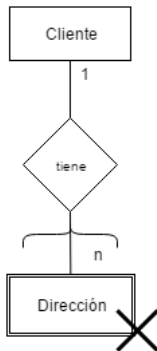
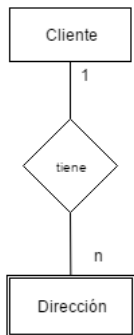
1 a N



1 a N

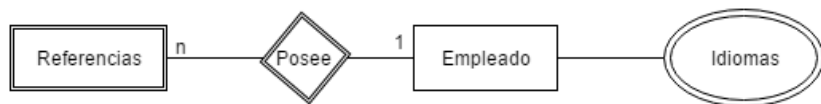


1 a N



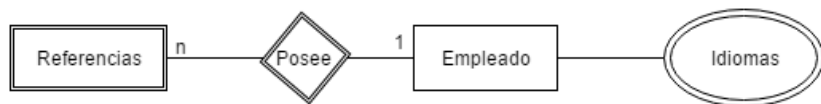
```
{  
  {  
    "cliente_id": 76123,  
    "nombre": "Acme Data Modeling  
Services",  
    "tipo_de_cliente": "business",  
    "direcciones": :  
      [  
        {calle: "San Martin 2222",  
ciudad: "Caseros",  
provincia: "Buenos Aires",  
codigo_postal: 99076} ,  
        {calle: "9 de Julio 2223",  
ciudad: "CABA",  
codigo_postal: 01097}  
      ]  
  }  
}
```

1 a N

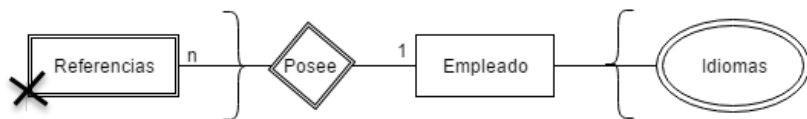


DER. Empleados Idiomas y Referencias

1 a N



DER. Empleados Idiomas y Referencias



DID. Empleados Idiomas y Referencias

Cardinalidad M a N



DER. M a N

Cardinalidad M a N



DER. M a N



DID. M a N con referencias

Cardinalidad M a N



DER. M a N



DID. M a N con referencias

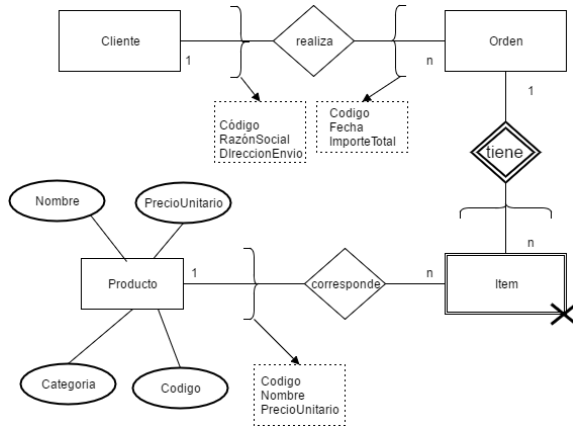
```

{
  { courseID: 'C1667',
    title: 'Introduction to Anthropology',
    instructor: 'Dr. Margret Austin',
    credits: 3,
    enrolledStudents: ['S1837', 'S3737', 'S9825' ...
      'S1847'] },
  { courseID: 'C2873',
    title: 'Algorithms and Data Structures',
    instructor: 'Dr. Susan Johnson',
    credits: 3,
    enrolledStudents: ['S1837', 'S3737', 'S4321', 'S9825'
      ... 'S1847'] },
  { courseID: 'C3876',
    title: 'Macroeconomics',
    instructor: 'Dr. James Schulen',
    credits: 3,
    enrolledStudents: ['S1837', 'S4321', 'S1470', 'S9825'
      ... 'S1847'] },
  ...
}
  
```

```

{
  {studentID: 'S1837',
    name: 'Brian Nelson',
    gradYear: 2018,
    courses: ['C1667', 'C2873', 'C3876']},
  {studentID: 'S3737',
    name: 'Yolanda Deltor',
    gradYear: 2017,
    courses: ['C1667', 'C2873']},
  ...
}
  
```

Desnormalización parcial



Desnormalización

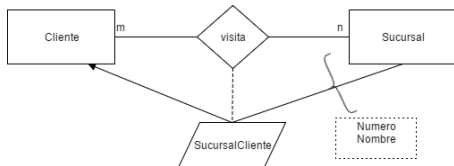
JSON Schema para Documento Orden

```
"Orden": { "type": "object",
  "properties": {
    "OrdenID": { "type": "integer" },
    "Fecha": { "type": "string", "format": "date-time" },
    "ImporteTotal": { "type": "integer" },
    "Cliente": {
      "type": "object",
      "properties": {
        "Codigo": { "type": "integer" },
        "RazonSocial": { "type": "string" },
        "DireccionEnvio": { "type": "string" }
      }
    },
    "ItemsOrden": {
      "type": "Array",
      "items": {
        "type": "object",
        "properties": {
          "Cantidad": { "type": "integer" },
          "Codigo": { "type": "integer" },
          "Nombre": { "type": "string" },
          "PrecioUnitario": { "type": "string" }
        }
      }
    }
  }
}
```

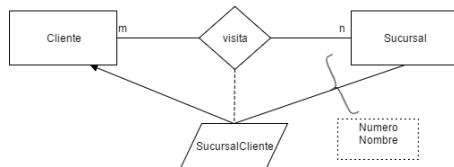
Uso de Documentos Auxiliares



Uso de Documentos Auxiliares

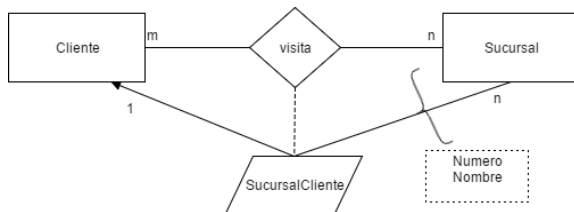


Uso de Documentos Auxiliares

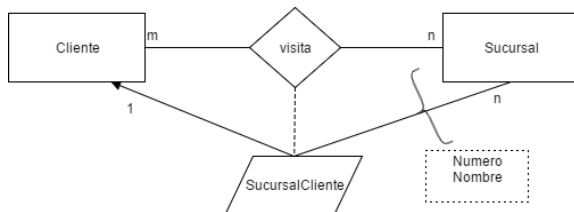


```
"SucursalCliente": { "type": "object",  
  "properties": {  
    "ClienteID": { "type": "integer" },  
    "Numero": { "type": "integer" },  
    "Nombre": { "type": "string" }  
  }  
}
```

Uso de Documentos Auxiliares

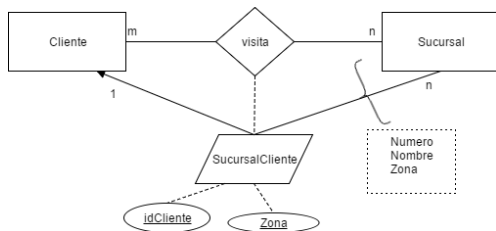


Uso de Documentos Auxiliares

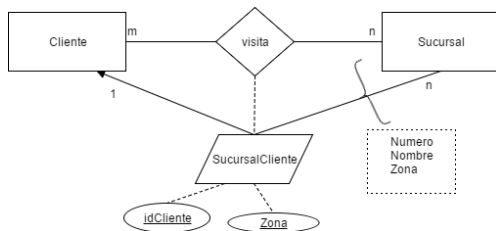


```
"SucursalCliente": { "type": "object",  
  "properties": {  
    "ClienteID": { "type": "integer" },  
    "Sucursales": {  
      "type": "Array",  
      "items": { "type": "object",  
        "properties": {  
          "Codigo": { "type": "integer" },  
          "Nombre": { "type": "string" },  
        }  
      }  
    }  
  }  
}
```

Uso de Documentos Auxiliares



Uso de Documentos Auxiliares



```

"SucursalCliente": { "type": "object",
  "properties": {
    "ClienteID": { "type": "integer" },
    "Zona": { "type": "string" },
    "Sucursales": { "type": "Array",
      "items": { "type": "object",
        "properties": {
          "Codigo": { "type": "integer" },
          "Nombre": { "type": "string" },
        }
      }
    }
  }
}
  
```

Bibliografía

- *NoSQL for Mere Mortals* - Dan Sullivan
- *NoSQL Distilled. A Brief Guide to the Emerging World of Polyglot Persistence* - Pramod J. Sadalage y Martin Fowler
- *Diseño de Bases de Datos Basadas en Documento: Modelo de Interrelación de Documentos* Gerardo Rossel y Andrea Manna
- *Mastering RethinkDB* - Shahid Shaikh
- *MongoDB Applied Design Patterns* - Rick Copeland
- *CouchDB- The Definitive Guide* - J. Chris Anderson, Jan Lehnardt, Noah Slater
- *RavenDB in Action* - Itamar Syn-Hershko