estudios de postgrado en computación

Bases de datos avanzadas

Universidad de Los Andes

Postgrado en Computación

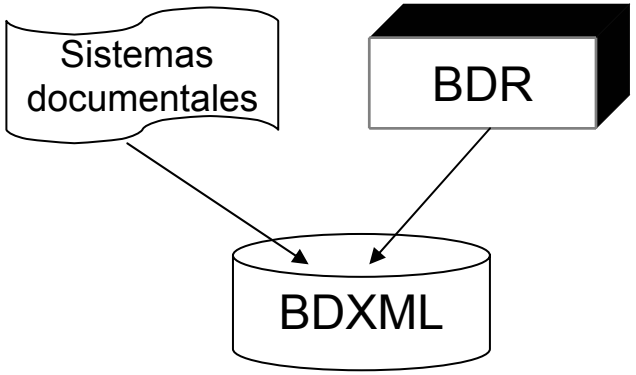
Prof. Isabel M. Besembel Carrera

Unidad II. Sesión 11 y 12. BD en XML.

UNIVERSIDAD DE LOS ANDES

Introducción …

- Convergencia de los sistemas documentales y bases de datos
- Publicación de los datos de una BD en web a través de XML
- Enfoques:
  - ✓ Middleware: uso de un SGBD relacional u orientado por objetos para almacenar los elementos simples del documento XML
  - ✓ XML nativo: almacenar el documento XML completo junto con índices de acceso especiales para acelerar su acceso

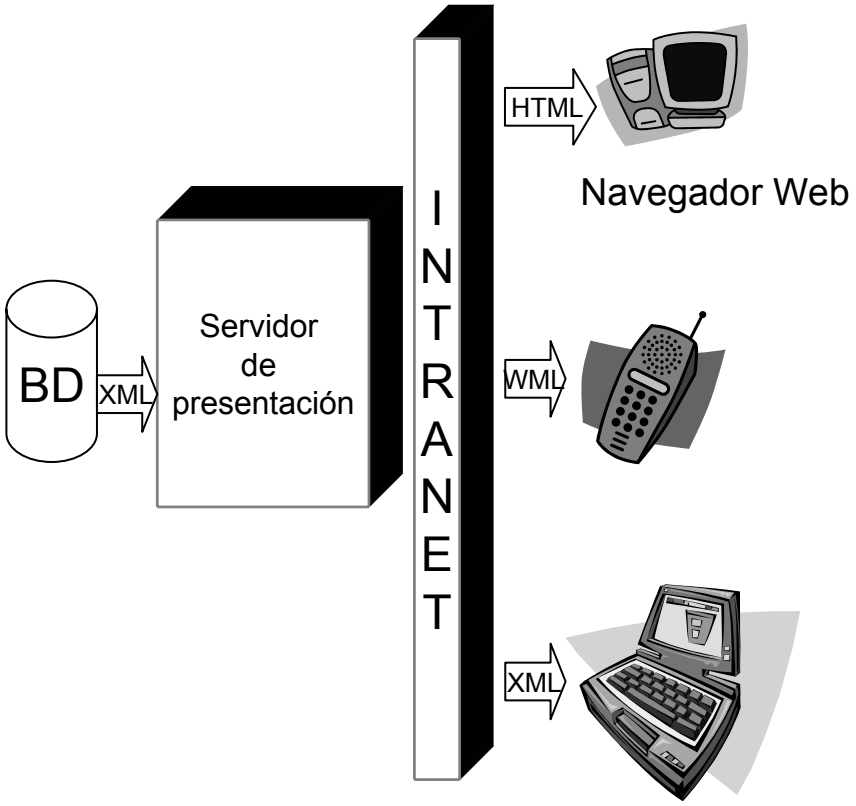


Enfoque 1: integración de XML en el sistema mediante un mapeo XML ↔ BD apropiado
Enfoque 2: necesidad de desarrollar un nuevo tipo de dato para documentos y además un SGBD completo orientado a la manipulación de documentos SGBDXML.

Lenguajes de manipulación de datos: SQL3, 1999 con funciones XML o XQuery

UNIVERSIDAD DE LOS ANDES

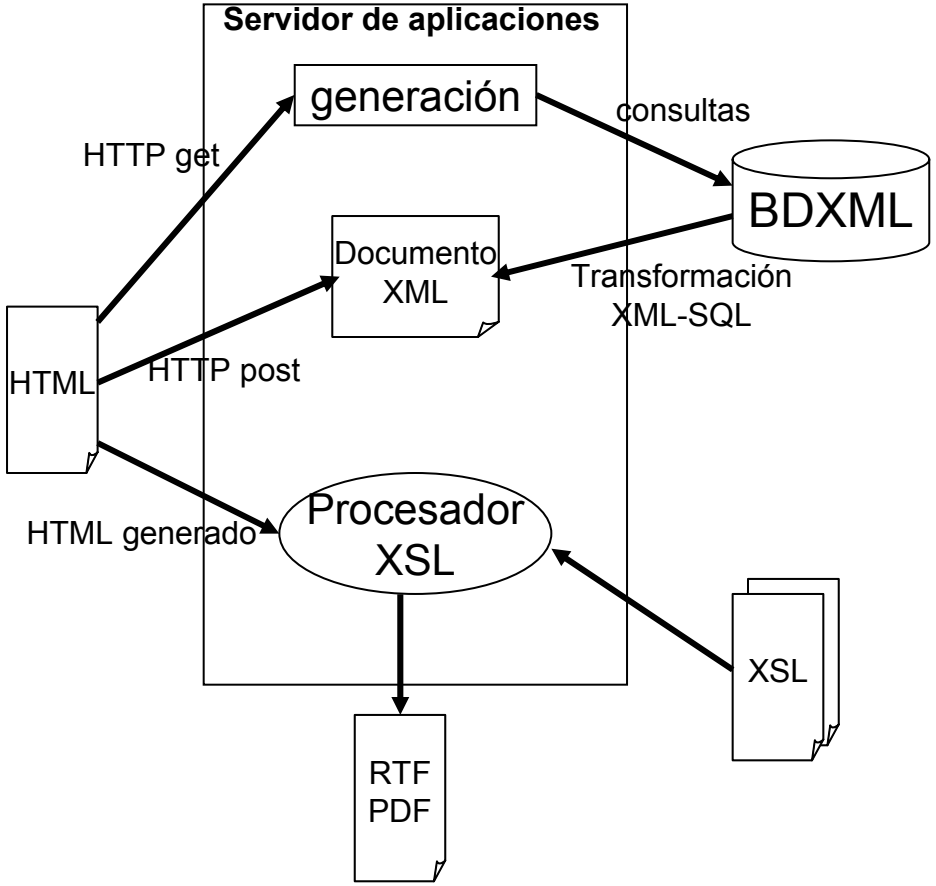
Aplicaciones …



- Sitio Web XML
  - ✓ Sitio web dinámico: varias arquitecturas posibles (scripts CGI, servlet, activeX, EJB, ASP o JSP) generan dinámicamente el HTML a partir de los datos de la BD
  - ✓ Servidor de presentación necesario para asociar la hoja de estilo con los documentos. Las hojas de estilo se aplican a los datos extraídos de la BD sobre el servidor de presentación por un procesador XSL

UNIVERSIDAD DE LOS ANDES

Aplicaciones …

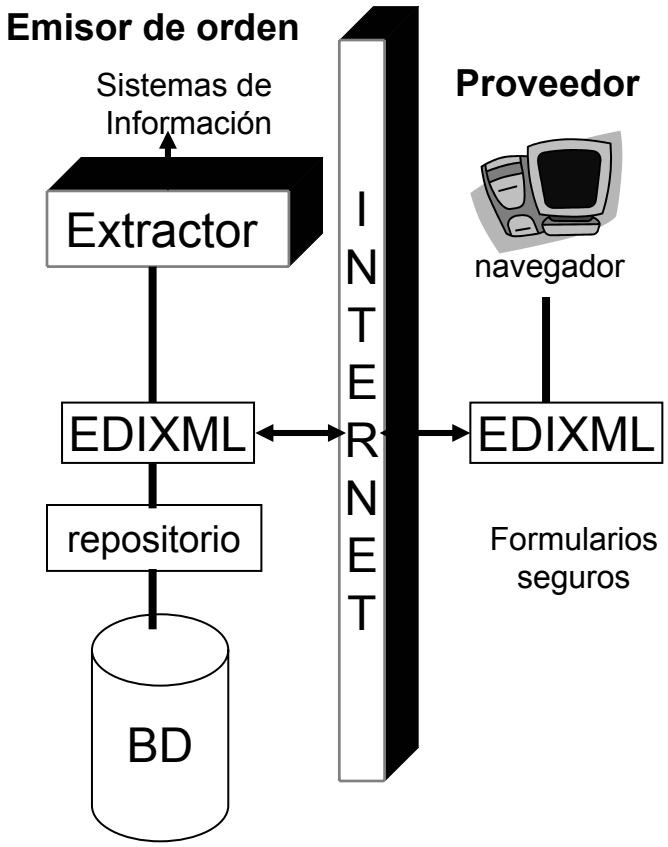


- Gestión de documentos semi-estructurados
  - ✓ BDR con una capa de mapeo XML o una BDXML
  - ✓ Manejo transaccional con un servidor de aplicaciones
  - ✓ Presentación en HTML con la ayuda de hojas de estilo procesados en el servidor de aplicaciones
  - ✓ Posibilidad de imprimir documentos en RTF o PDF

UNIVERSIDAD DE LOS ANDES

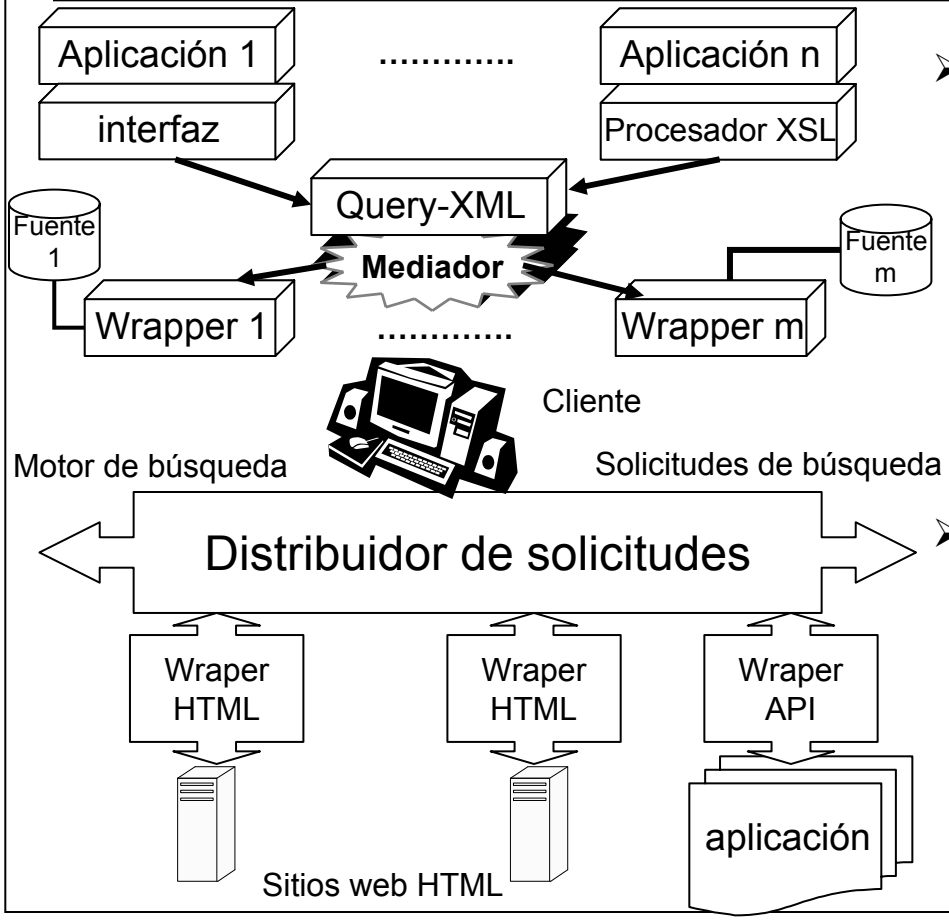
Aplicaciones …

- **Sitio de comercio electrónico**
  - ✓ B2B intercambio de mensajes del tipo pedido, recepción, entrega, factura, etc.
  - ✓ Mensajes XML según el estándar cXML o ebXML (actualmente en desarrollo)



UNIVERSIDAD DE LOS ANDES

Aplicaciones …

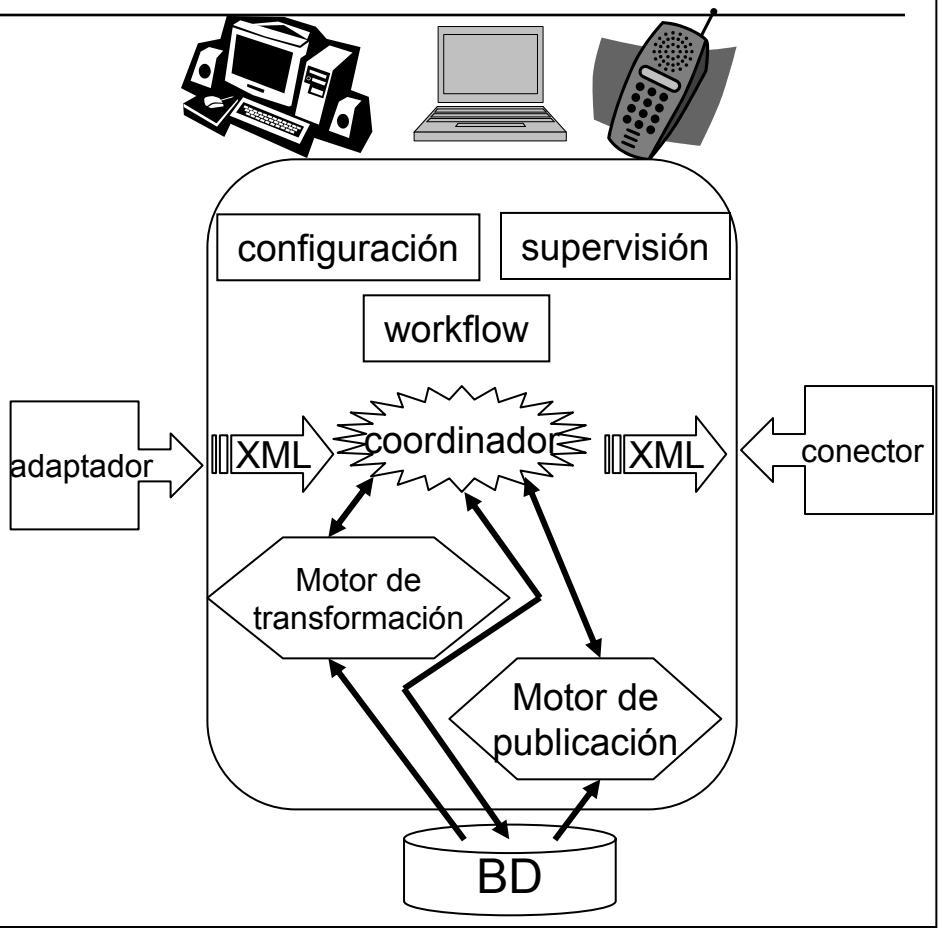


- Consulta de datos heterogéneos
  - ✓ XML es apropiado ya que es universal, más rico semánticamente e integra los conceptos jerárquicos, relacional y objetos
- Motor de búsqueda XML
  - ✓ Conversión HTML-XML según plantillas
  - ✓ Mejora de las respuestas a las búsquedas efectuadas por el motor

UNIVERSIDAD DE LOS ANDES

Aplicaciones

- Integración de aplicaciones o intercambio de datos en las empresas
  - ✓ EAI (Enterprise Application Integrator)
  - ✓ Integración de XML como vehículo de intercambio en un EAI reduce el número de conectores, mejora la supervisión y está soportado por estándares reconocidos.



UNIVERSIDAD DE LOS ANDES

Modelo de datos …

- Modelo XML con mayor poder de expresión que el relacional
- Modelo 1: traducción en atributos (Attribute translation)
  - ✓ Cada tabla es una marca que está compuesta de un elemento para cada tupla que es una marca cuyos atributos son las columnas de la tabla
    - o Ejemplo:

Empresa	nombre	direccion	telefono	codigoPostal	ciudad
	Quincalla El Palacio	Av. Principal. Calle 34. Nro.102	0223-235 7124	4034	Valencia
	Supermercado El Callao	Av. 8. Calle 45. Nro. 58	0283-240 5567	4065	Valera

```
<Empresa>
  <Tupla nombre="Quincalla El Palacio" direccion="Av. principal. Calle 34. Nro. 102"
          telefono="0223-235 7124" codigoPostal="4034" ciudad="Valencia"/>
  <Tupla nombre="Supermercado El Callao" direccion="Av. 8. Calle 45. Nro. 58"
          telefono="0283-240 5567" codigoPostal="4065" ciudad="Valera"/>
</Empresa>
```

UNIVERSIDAD DE LOS ANDES

Modelo de datos …

- Modelo 2: traducción en elementos (element translation)
  - ✓ Cada tabla es una marca que está compuesta de un elemento para cada tupla que es una marca y de elementos anidados de primer nivel representando cada atributo o columna de la tabla
  - ✓ Ejemplo:

```
<Empresa>
  <Tupla>
        <nombre>Quincalla El Palacio</nombre>
        <direccion>Av. principal. Calle 34. Nro. 102</direccion>
        <telefono>0223-235 7124</telefono>
        <codigoPostal>4034</codigoPostal>
        <ciudad>Valencia</ciudad>
  </Tupla>
</Empresa>
```

UNIVERSIDAD DE LOS ANDES

Ejercicio: Realizar la transformación a esquema relacional de umlEmpresa y luego usar el modelo 1 y el modelo 2 para obtener el documento XML

Modelo de datos

- Ambos métodos son equivalentes
- Es posible aplicar uno u otro para publicar resultados de consultas SQL en XML
- Enfoques mixtos son posibles, pero siempre el resultado es un árbol XML plano (árbol con un nivel) y se pierde el poder expresivo de XML
- Herramientas:
  - ✓ XML SDK de Oracle (traducción en elementos)
  - ✓ XML-DBMS de Ronald Bourret
  - ✓ XML en SQL Server de Microsoft
- Si el documento XML es plano se puede hacer transformación inversa, sino se transforma con XSL en dos pasos:
  - ✓ Una para extraer y almacenar los datos en la BD
  - ✓ Otra para anidar o aplanar los documentos XML

UNIVERSIDAD DE LOS ANDES

Modelo de esquemas …

- Normalmente es imposible guardar un documento XML en una única tabla
- Problema de transformación de un esquema relacional a un esquema XML y viceversa es un **problema complejo**
- Características del modelo XML
  1. Descriptores opcionales en el DTD
  2. Esquema que soporta tipos de datos simples y complejos, flexibles e irregulares
  3. Datos autodescriptivos marcados y atributos de esquema y de DTD
  4. Enlaces de tipo hipertexto representando relaciones N:M
  5. Uso intensivo de la composición por agregación con los constructores sequence, choice y all
  6. Tipos de datos variados y extensibles (entero, real, texto, fecha, identificador, etc.)

UNIVERSIDAD DE LOS ANDES

Modelo de esquemas …

- Hoy día el modelo seleccionado es el de esquemas XML
- Bosque (repository) XML: Colección de documentos XML de igual naturaleza asociados a un esquema y almacenados en conjunto
  - ✓ Cada documento es un árbol
- BDXML: es un conjunto de bosques XML

```
<?xml version="1.0" ?>
<!-- archivo: hoteles1.xml          -->
<!--                                 D A T O S                    -->
<Guia region="Oriental" version="1.0">
  <Hotel categoria="*****">
                <nombre>LagunaMar</nombre>
                <direccion>
                        <Av>Via Playa El agua</Av>
                </direccion>
                <telefono>0271-274 8234</telefono>
                <precio>
                        <HabSimple>65.000</HabSimple>
                        <HabDoble>80.000</HabDoble>
                </precio>
                <ciudad>Margarita</ciudad>
  </Hotel>
  <Hotel categoria="****">
                <nombre>Hesperia Playa El Agua</nombre>
                <direccion>
                        <Av>Via Playa El agua</Av>
                </direccion>
                <telefono>0271-256 8345</telefono>
                <precio>
                        <HabSimple>55.000</HabSimple>
                        <HabDoble>70.000</HabDoble>
                </precio>
                <ciudad>Margarita</ciudad>
  </Hotel>
```

UNIVERSIDAD DE LOS ANDES

Modelo de esquemas …

```
<Posada categoria="***">
        <nombre>La Comadre</nombre>
        <direccion>
                <Av>2</Av>
                <Nro>43-15</Nro>
        </direccion>
        <telefono>0285-287 7764</telefono>
        <precio>
                <HabSimple>20.000</HabSimple>
                <HabDoble>40.000</HabDoble>
        </precio>
        <ciudad>Cumana</ciudad>
  </Posada>
  <Posada categoria="*">
        <nombre>El Compai/nombre>
        <direccion>
                <Av>6</Av>
                <Nro>87-18</Nro>
        </direccion>
        <telefono>0286-276 7512</telefono>
        <precio>
                <HabDoble>45.000</HabDoble>
        </precio>
        <ciudad>Maturin</ciudad>
  </Posada>
</Guia>
```

```
<?xml version="1.0" ?>
<!-- archivo: hoteles.xml        -->
<!--                             D A T O S                    -->
<Guia region="Andes" version="1.0">
  <Hotel categoria="***">
        <nombre>Prado Rio</nombre>
        <direccion>
                <Av>2</Av>
                <Nro>55</Nro>
        </direccion>
        <telefono>0274-244 6683</telefono>
        <precio>
                <HabSimple>35.000</HabSimple>
                <HabDoble>55.000</HabDoble>
        </precio>
        <ciudad>Merida</ciudad>
  </Hotel>
  <Hotel categoria="****">
        <nombre>Park hotel</nombre>
        <direccion>
                <Av>Gonzalo Picon</Av>
                <Nro>67</Nro>
        </direccion>
        <telefono>0274-263 8256</telefono>
        <precio>
                <HabSimple>45.000</HabSimple>
                <HabDoble>65.000</HabDoble>
        </precio>
        <ciudad>Merida</ciudad>
  </Hotel>
```

UNIVERSIDAD DE LOS ANDES

Modelo de esquemas …

```
<Hotel categoria="**">
        <nombre>Oviedo</nombre>
        <direccion>
                <Av>3</Av>
                <Nro>3-43</Nro>
        </direccion>
        <telefono>0274-252 8745</telefono>
        <precio>
                <HabDoble>35.000</HabDoble>
        </precio>
        <ciudad>Merida</ciudad>
 </Hotel>
 <Posada categoria="***">
        <nombre>El encuentro</nombre>
        <direccion>
                <Av>2</Av>
                <Nro>17-15</Nro>
        </direccion>
        <telefono>0274-244 6234</telefono>
        <precio>
                <HabSimple>20.000</HabSimple>
                <HabDoble>40.000</HabDoble>
        </precio>
        <ciudad>Trujillo</ciudad>
 </Posada>
 <Posada categoria="****">
        <nombre>El remanzo</nombre>
        <direccion>
                <Av>Tulio Febres Cordero</Av>
                <Nro>23-18</Nro>
        </direccion>
        <telefono>0274-252 5412</telefono>
        <precio>
                <HabSimple>30.000</HabSimple>
                <HabDoble>45.000</HabDoble>
        </precio>
        <ciudad>Trujillo</ciudad>
  </Posada>
</Guia>
```

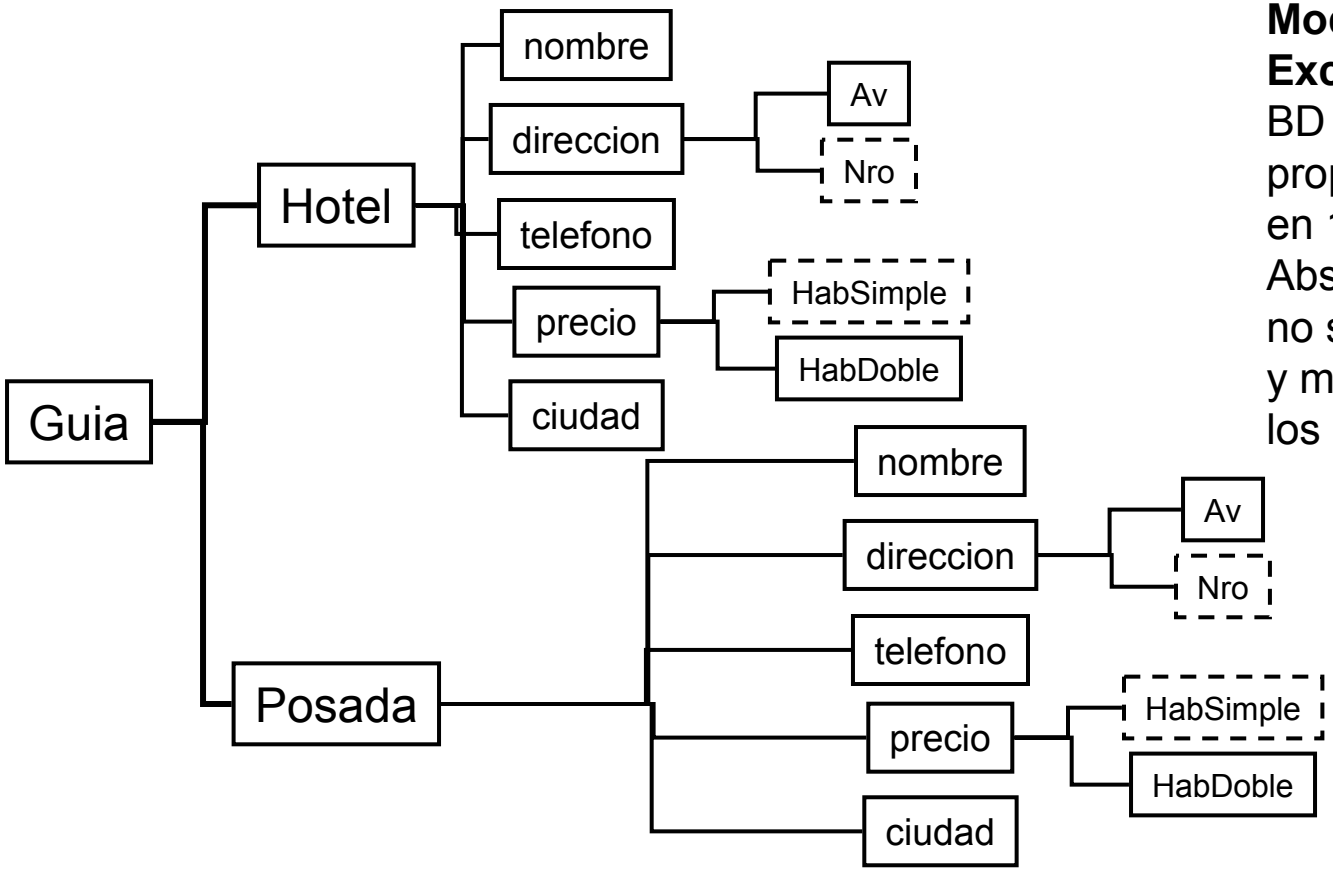
UNIVERSIDAD DE LOS ANDES

Modelo de esquemas …

- Guía de datos: esquema generado a partir de un conjunto de documentos mediante la unión de los árboles de estructura que describen todos los caminos posibles en la colección y colocando los tipos de datos como texto
- Construcción:
  - ✓ Aislar todos los caminos completos en el documento. Un camino completo es una secuencia de marcas desde la raíz hasta las hojas
  - ✓ Eliminar todos los caminos o subcaminos incluidos en los caminos encontrados, guardando los caminos distintos
  - ✓ Construir el grafo donde los caminos completos son todos los caminos distintos guardados

UNIVERSIDAD DE LOS ANDES

Modelo de esquemas



Modelo OEM (Object Exchange Model): para BD semi-estructuradas, propuesto por J. Widom en 1994.
Absorbido por XML, ya no se utiliza, más flexible y menos complejo que los esquemas XML

UNIVERSIDAD DE LOS ANDES

Lenguajes de consulta para XML

- Desde 1996 se ha investigado para proponer tal lenguaje
- Ejemplos:
  - ✓ LOREL, STRUQL (AT&T), UnQL (P. Buneman), YATL (INRIA), SgmlQL.
- Fundamentados en enfoques diferentes, pero todos usan
- Mezcla de predicados sobre los metadatos (marcas) y los datos (valores)
- Expresiones de caminos
- Permiten seleccionar subgrafos y crear nuevos grafos a partir de los datos extraídos

UNIVERSIDAD DE LOS ANDES

Objetivos

- Selección de árboles por criterios múltiples
- Posibilidad de realizar las operaciones de los tipos básicos
- Cuantificación universal y existencial de las variables
- Combinación de datos desde los documentos
- Ordenamiento de los resultados
- Anidamiento de las consultas
- Posibilidad de usar agregados y funciones asociadas
- Tratamiento de jerarquías y de secuencias
- Agregación de datos desde los documentos
- Preservación de las estructuras
- Construcción de nuevas estructuras

UNIVERSIDAD DE LOS ANDES

LOREL …

- Construido en Stanford 1996 para consultar grafos OEM con OQL
- Basado en expresiones de caminos simples, secuencia de marcas separadas por punto en vez de / del XPath
  - ✓ Ejemplo: Guia.Hotel.direccion.Av
- Uso de variables asociadas a expresiones de caminos
  - ✓ Ejemplo: $h asignada a Guia.Hotel para barrer los hoteles
  - ✓ Consulta: Listar los hoteles de Margarita

    Select Guia.Hotel.nombre

    Where Guia.Hotel.ciudad = “Margarita”
  - ✓ Consulta: Listar los nombres de las posadas en la misma avenida del hotel Prado Rio

    Select $p.nombre, $p.telefono

    From Guia.Hotel $h, Guia.Posada $p

    Where $h.direccion.Av = $p.direccion.Av and $h.nombre = “Prado Rio”

UNIVERSIDAD DE LOS ANDES

LOREL

- Permite usar expresiones de caminos generalizadas, que son aquellas donde se pueden sustituir marcas por expresiones o símbolos que sirven de joker:
  - ✓ No importa cual subcadena %
  - ✓ Disyunción de marcas |
  - ✓ Opcionalidad ?
  - ✓ Joker #
  - ✓ Operador de repetición * (0..n)
- Ejemplo:

```
Select Hotel.#
From Guia.Hotel $h
Where $h.direccion.(Av | Nro) = “%15%”
```

UNIVERSIDAD DE LOS ANDES

SQLX

- Desarrollado en 1998 en el Laboratorio PRiSM-Versalles,Francia
- Expresiones de navegación basadas en XPath
  - ✓ Consulta: Listar los hoteles de Margarita

    Select $h/nombre

    From /Guia/Hotel $h

    Where $h/ciudad = “Margarita”
  - ✓ Consulta: Listar los nombres de las posadas en la misma avenida del hotel Prado Rio

    Select $p/nombre, $p/telefono

    From /Guia/Hotel $h, /Guia/Posada $p

    Where $h/direccion/Av = $p/direccion/Av and $h/nombre = “Prado Rio”

UNIVERSIDAD DE LOS ANDES

XML-QL …

- Propuesto por AT&T en 1997 basado en STRUQL y adaptado a XML
- Selecciones: basadas en plantillas de búsqueda, donde las variables $N y $T recogen los datos encontrados

```
Select <Posada>
                    <nombre>$N</nombre>
                    <telefono>$T</telefono>
                    <ciudad>Trujillo</ciudad>
</Posada> IN Guia
```

- Posibilidad de encontrar los opcionales con

```
Select <Hotel>
            <nombre>Oviedo</nombre>
            <telefono>$T</telefono>
            <precio>
                        <HabDoble>[$HD]</HabDoble>
            </precio>
</Hotel> IN Guia
```

UNIVERSIDAD DE LOS ANDES

XML-QL …

- Construcción de resultados: la cláusula where expresa los criterios de búsqueda y el select especifica la construcción del grafo resultado

```
Where <Posada>
                <nombre>$N</nombre>
                <telefono>$T</telefono>
                <ciudad>Trujillo</ciudad>
   </Posada> IN Guia
Select <BB>
                <name>$N</name>
                <phone>$T</phone>
   </BB>
```

Permite obtener los nombres y teléfonos de las posadas de Trujillo y componer el árbol resultado con las marcas en inglés

UNIVERSIDAD DE LOS ANDES

XML-QL ...

- Productos (joins): mediante la reutilización de variables en las condiciones

```
Where <Hotel>
            <nombre>Prado Rio</nombre>
            <direccion>
                        <Av>$a</Av>
            </direccion>
   </Hotel> IN Guia
   <Posada>
            <nombre>$N</nombre>
            <direccion>
                        <Av>$a</Av>
            </direccion>
  </Posada> IN Guia
Select <resultado>
            <nombre>$N</nombre>
     </resultado>
```

Listar los nombres de las posadas en la misma avenida del hotel Prado Rio

UNIVERSIDAD DE LOS ANDES

XML-QL …

- Variables para marcas: permite consultar los metadatos
- Consulta: Encontrar todos los nombres y teléfonos de los hoteles 3 estrellas en Margarita

```
Where <Hotel @categoria="***">
            <$a>$V</$a> $a IN {nombre, telefono}
            <ciudad>Margarita</ciudad>
   </Hotel> IN Guia
Select <resultado>
            <$a>$V</$a>
   </resultado>
```

UNIVERSIDAD DE LOS ANDES

XML-QL

- Expresiones regulares: expresiones de caminos con . y operadores como: | (opción) * (cualquiera) +(concatenación)
- Consulta: encontrar los teléfonos de todos los hoteles o posadas en Margarita

```
Where <*>
        <ciudad>Margarita</ciudad>
        <telefono>$T</telefono>
  </*> IN Guia
Select <resultado>$T</resultado>
```

Resumen: select <plantilla> regresa el subárbol que responde

Where <plantilla> select <plantilla> filtro de búsqueda con where y construcción del resultado con el select

UNIVERSIDAD DE LOS ANDES

XQL

- Variante de XPath propuesta por Microsoft, pero abandonada a favor de XQuery
- Poca legibilidad y no hay forma de reestructurar los resultados
- Consulta: Encontrar todos los nombres de los hoteles 3 estrellas en Margarita

  /Guia/Hotel?(@categoria[text() = “***”]) /nombre??/ciudad[text() = “Margarita”]

  / nivel siguiente a partir de la raíz o el nodo actual

  // todos los niveles a partir de la raíz o el nodo actual

  * Designa una marca cualquiera

  @att el atributo de nombre att

  @* un atributo cualquiera

  [criterio de filtrado]

  ? Indicar el nodo resultado

  ?? La raíz de un subárbol resultado

UNIVERSIDAD DE LOS ANDES

XQuery …

- Propuesto a W3C en 2001 por IBM y algunos autores de XML/QL
- Palabras claves se escriben en minúsculas (en azul)
- Fuertemente tipeado y basado en programación funcional
- Generalización de XPath y superconjunto de SQL
- Funciones:
  - ✓ Proyección de árboles en subárboles, asegurada por XPath
  - ✓ Selección de árboles y subárboles usando predicados sobre los valores de las hojas (<, >, =, !=, <=, >=)
  - ✓ Toda operación válida sobre un tipo de dato se puede usar para componer una consulta. Tiene reglas específicas para inferir tipos de datos.
  - ✓ Soporta funciones documentales como contains para búsqueda

UNIVERSIDAD DE LOS ANDES

XQuery …

- ✓ Consultas con variables asociadas a árboles o para iterar sobre colecciones de árboles. La cuantificación universal es la opción por omisión
- ✓ Posibilidad de combinar los árboles extraídos usando join de árboles
- ✓ Reordenamiento de árboles y de los resultados según valores de los elementos en forma ascendente o descendente
- ✓ Anidamiento de consultas para indicar una reestructuración de los resultados, permitiendo así la reconstrucción de jerarquías
- ✓ Cálculo sobre colecciones particionadas o no a través de count, sum, max, min y avg
- ✓ Posibilidad de incluir cualquier función de usuario a condición que se respete el tipo de dato
- ✓ Uso de los operadores /, //, *, etc. de XPath

UNIVERSIDAD DE LOS ANDES

Consultas XPath…

- Expresiones de caminos interpretadas como consulta aplicada a un documento o a una colección de documentos cuyo resultado es un conjunto de subárboles seleccionados en cada documento
- Ejemplo:
  - ✓ document(“http://www.guiaHotelera.ve”)//Hotel/nombre/text()
  - ✓ collection(“Guia”)//Hotel/nombre/text()
- Expresiones FLWR (flower) for … let … where … return …; aplicadas a una o varias colecciones de árboles definidos en el bloque for que permite regresar los árboles construidos por el bloque return a partir de los árboles memorizados por el bloque let y/o seleccionados por el bloque where

UNIVERSIDAD DE LOS ANDES

Consultas …

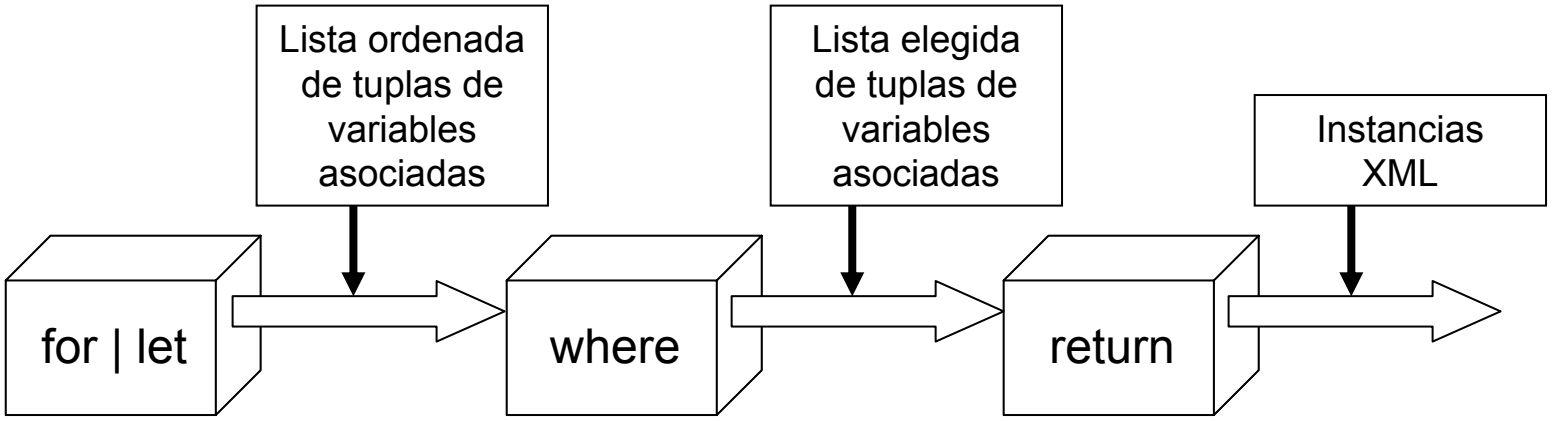
- Forma general:

  for $<var> in <bosque> [, $<var> in <bosque>] … // iteración

  let $<var> := <subárbol> // asignación

  where <condición> // elección

  return <resultado> // construcción



UNIVERSIDAD DE LOS ANDES

Consultas …

- Restricciones: $<variable>/<expresión de camino> Θ <constante>
- Productos (join): $<variable>/<expresión de camino> Θ $<variable>/<expresión de camino>
- Operadores: Θ ={ <, >, =, !=, <=, >= }, empty para probar si un elemento está vacío, contains para probar palabras claves, etc.
- Consulta 1: liste los nombres de los hoteles en Margarita

  for $H in collection(“Guia”)/Hotel

  where $H/ciudad = “Margarita”

  return $H/nombre

  Equivalente en XPath a

  collection(“Guia”)/Hotel[ciudad = “Margarita”]/nombre

UNIVERSIDAD DE LOS ANDES

Consultas …

- Anidamiento de consultas: en el for (para definir variables en árboles calculados), where (para calcular los valores de los predicados) o return (para definir los documentos anidados).

Listar los nombres de las posadas en la misma avenida del hotel Prado Rio

```
for   $H in collection("Guia")/Hotel,
      $P in collection("Guia")/Posada
where  $H//direccion/Av = $P//direccion/Av and $H/nombre = "Prado Rio"
return <resultado>
            <nombre> {$H/nombre/text()} </nombre>
            <telefono> {$H/telefono/text()} </telefono>
      </resultado>
```

UNIVERSIDAD DE LOS ANDES

Consultas …

- Encontrar todos los nombres y direcciones de los hoteles 3 estrellas en Margarita

```
for   $H in collection("Guia")/Hotel
where  $H/@categoria = "***"
return <resultado>
                {$H/nombre}
                <direccion> {$H/ direccion//text()} </ direccion >
        </resultado>
```

Aplanamiento en forma de texto de la dirección

- ¿Cuántos hoteles hay en los documentos Guia?

```
for   $H := collection("Guia")/Hotel
return <nombreHotel> {count ($H) } </nombreHotel>
```

UNIVERSIDAD DE LOS ANDES

Consultas

- Listar las direcciones de los hoteles cuyo nombre contiene la palabra “hotel”

```
for    $H := collection(“Guia”)/Hotel
where contains ($H//nombre, “hotel”)
return <resultado>
                    {$H/nombre}
                    <direccion> {$H/ direccion//text()} </ direccion >
       </resultado>
```

- Encontrar todos los valores de los atributos de los hoteles

```
for    $H in collection(“Guia”)/Hotel
return <res>
                    {for  $A  in  $H//@*     return $A}
       </res>
```

UNIVERSIDAD DE LOS ANDES

Hacia un algebra para XML

- Algebra Xquery: Conjunto de operadores elementales, cada uno permite la transformación de una o varias colecciones de árboles XML en una colección de árboles, el conjunto permite la representación de los planos de ejecución calculando la respuesta a toda consulta Xquery.
- Necesita la especificación
  - ✓ del sistema de tipos y sus operadores,
  - ✓ además de las reglas de inferencia de tipos
- Resultados preliminares están en el documento que define la semántica formal de Xquery (www.w3.org/TR/2001/WD-query-semantics-200100607

UNIVERSIDAD DE LOS ANDES

Operadores

- Asignación y construcción: let $var := expresión
- Proyección: $var/xpath
- Acceso a los datos: $var/xpath/data()
- Iteración: for $var IN seq return expresión
- Selección: for $var IN seq where cc return expresión
- Cuantificación existencial: some $var in seq satisfies cc
- Cuantificación universal: any $var in seq satisfies cc
- Producto: for $var1 in seq1, $var2 in seq2 where cc return expresión
- Ordenamiento: expresión1 sortby expresión2
- Funciones integradas: distinct-value, unordered, parent, ref, deref, index, before, after, avg, min, max, sum, count
- Definición de funciones: define function nom ([type:var]*) returns colección