# UF3 Persistència en SGBD-XML Annex 02

# Introducció al SGBD-XML BaseX

Isidre Guixà i Miranda Institut Milà i Fontanals d'Igualada



19 de febrer de 2013 Isidre Guixà i Miranda Institut Milà i Fontanals C/. Emili Vallès, 4 08700 - Igualada

Aquest material està protegit sota llicència de <u>Creative Commons BY-NC-SA 3.0</u> i és la maqueta dels materials en desenvolupament per l'Institut Obert de Catalunya, que una vegada publicats seran accessibles a l'apartat recursos de la web <u>http://ioc.xtec.cat</u>

En cas de suggeriment i/o detecció d'error, podeu informar a iguixa@xtec.cat

#### Castellano

Este material está protegido bajo licencia <u>Creative Commons BY-NC-SA 3.0</u> y es la maqueta de los materiales en desarrollo por el Institut Obert de Catalunya, que una vez publicados seran accesibles en el apartado recursos de la web <u>http://ioc.xtec.cat</u>

En caso de sugerencia y/o detección de error, dirigir-se a iguixa@xtec.cat

# Euskara

Material hau <u>Creative Commons BY-NC-SA 3.0</u> babesa eta eredu materialak garatzen Institut Obert de Catalunya, behin argitaratutako web <u>http://ioc.xtec.cat</u> Iradokizuna bada eta / edo errore detekzioa, <u>iguixa@xtec.cat</u> zinen joan

#### Galego

Este material está protexido baixo <u>Creative Commons BY-NC-SA 3.0</u> eo modelo é o desenvolvemento de materiais para o Institut Obert de Catalunya, que unha vez publicado será accesible a sección de recursos do <u>http://ioc.xtec.cat</u> web. Se suxestión e / ou detección de erros, indica <u>iguixa@xtec.cat</u>

Els continguts de la unitat formativa "Persistència en BD-XML natives" fan referència, en múltiples ocasions, al SGBD-XML *BaseX* i, en conseqüència, es fa altament recomanable tenir-lo instal·lat i així poder comprovar la correctesa del codi desenvolupat.

El motiu d'aquest annex és facilitar una introducció a *BaseX*, cobrint les necessitats bàsiques d'instal·lació, configuració i primeres passes en un SO Windows. Per dur a terme la instal·lació en altres SO o per aprofundir en aquest SGBD, cal acudir a la pàgina web oficial del producte (<u>http://basex.org/</u>).

# Instal·lació

Per procedir a la instal·lació en un SO Windows, descarreguem de la pàgina web oficial, la darrera versió disponible per a Windows, que en el moment de redactar aquest annex és la 7.1 (2012/02/07) i es facilita en el fitxer executable BaseX71.exe. Per si la pàgina web oficial ja no proporciona aquesta versió, se us facilita una còpia <u>aquí</u>.

Tot i que el procés d'instal·lació és molt simple, no hem de deixar de posar atenció als missatges i recomanacions. El procés permet indicar:

- el directori on instal·lar el programa,
- el directori on resideixin les dades,
- la contrasenya per l'usuari administrador "admin" (que per defecte és "admin")
- els ports que utilitza *BaseX* (que per defecte són 1984 pel servidor i 8984 pel servidor HTTP que incorpora)

En finalitzar la instal·lació observarem, a l'arbre de programes de Windows, la carpeta anomenada Basex amb els següents accessos directes:

- *BaseX Client*, per accedir a una consola textual amb la que treballar amb *BaseX*, amb el servidors de *BaseX* engegats.
- BaseX Documentation, per accedir, via web, a la wiki oficial de BaseX.
- *BaseX Gui*, per accedir a una aplicació gràfica amb la que treballar amb *BaseX* i que no precisa tenir els servidors *BaseX* engegats.
- *BaseX Server (Start)*, per posar en marxa el SGBD *BaseX* (pel port 1984 si no s'ha canviat) i també el servidor HTTP (pel port 8984 si no s'ha canviat).
- BaseX Server (Stop), per aturar els dos servidors engegats amb l'opció anterior.
- *BaseX Standalone*, per accedir a una consola textual amb la que treballar amb *BaseX* localment, sense necessitat de tenir els servidors de *BaseX* engegats.
- Uninstall BaseX, per desinstal·lar el programari instal·lat.

# **Primeres passes**

El primer que se'ns passa pel cap una vegada instal·lat un SGBD-XML és poder-hi crear bases de dades i/o col·leccions on incorporar fitxers XML per procedir a la seva gestió via el llenguatge XQuery.

*BaseX* ens facilita dues interfícies per aconseguir-ho: la consola textual (client o *standalone)* per treballar amb comandaments i una aplicació client gràfica.

Nosaltres utilitzarem l'aplicació gràfica BaseX Gui, que, en posar-la en marxa, té una

#### aparença com mostra la imatge:

BaseX 7.1	
Database Editor View Nodes Options Help	
x 🗠 📾 x 📤 💠 🔶 💭 🎧 🗗 📾 & 🗄 🏛 🖬	
Command -	
Editor	
file 🕅 🕂	
<b>О</b> К	
Toxt	Quany Info
	Query mo
ок	13 MB

En ella observem tres parts clarament diferenciades.

*BaseX* permet gestionar diferents bases de dades i en cadascuna d'elles tenir-hi documents. Per poder començar a treballar, ens cal seleccionar una base de dades, via l'opció *Database* | *Open* o el botó Opens a database. La primera vegada que executem aquesta opció se'ns informarà que no existeix cap base de dades i si volem procedir a la seva creació.

Procedim a crear una nova base de dades, de nom mondial. Alerta! La finestra de diàleg que ens permet crear la base de dades, bé amb uns valors per defecte, que ens poden portar problemes. Fixem-nos en els apartats emmarcats en vermell a la següent imatge:

Create Database	X
General Parsing Indexes Full-Text	
Input file or directory: C:\BDXML\BaseX\ Skip corrupt (non-well-formed) files	Browse
Parse files in archives Name of database:	
Input format: File patterns (separated by commas):	
XML     *.xml       Add remaining files as raw files	
8 Please enter a database name.	
OK	Cancel

Si executem la creació de la base de dades amb la informació que proposa *BaseX*, se'ns crearà una base de dades, amb el nom que indiquem, i l'emplenarà amb tots els documents XML (apartat *File patterns*) que trobi dins el directori C:/BDXML/BaseX (apartat *Input file or directory*) tenint en compte també totes les subcarpetes i arxius contenidors d'arxius (apartat *Parse files in arxives*). Com que resulta que la instal·lació de *BaseX* deixa algun document XML a disc, aquests serien incorporats directament a la base de dades, fet que no volem.

En consequència, deixarem en blanc l'apartat *Input file or directory* i la base de dades se'ns crearà buida, com seria d'esperar. Anomenem-la mondial. La interfície ens avisa, amb un missatge a la part baixa de la finestra que *An empty database will be created*.

Atenció! En la versió 7.1 de *BaseX* i en les predecessores, la creació d'una base de dades no comprova l'existència d'una base de dades amb idèntic nom i, en cas d'existir, la sobreescriu. Per molt informàtics, aquest funcionament no és adequat i així s'ha comunicat als responsables de *BaseX* que han recollit la recomanació, per properes versions, de sobreescriure si i només si s'indica explícitament. Mentrestant haurem de ser conscients d'aquest inadequat funcionament i comprovar, abans, si ja existeix una base de dades amb nom idèntic a la que volem crear.

Una vegada s'ha creat la base de dades, el contingut de la nostra aplicació canvia i observem 4 parts clarament diferenciades, com mostra la imatge:

😹 BaseX 7.1 - mondial		
Database Editor View Nodes Options Help		
x 🗁 🖾 X 🚔 💠 🔶 🍋 🕼 🕼 🕼 🛤 🖬 🖬	0 Results	
Search 🗸 🔳		
	mondial	
Editor		
file 🔀 🕁		
lext a	Query Info	
	Iotal Time: 83.77 ms	
	Command: SET ADDARCHIVES true ADDARCHIVES: ON	
	Database 'mondial' created in 83.72 ms.	
db:open("mondial", "mondial")	30 MB	

Disposem de 4 zones:

- Superior esquerra, anomenada Query Editor, per poder introduir instruccions XPath i XQuery.
- Superior dreta, anomenada *Map visualization*, que visualitza la base de dades oberta. En aquests moments només en veiem el seu nom mondial sense cap contingut, donat que la base de dades és buida. En aquest apartat veurem gràficament tots els continguts de la base de dades.
- Inferior esquerra, anomenada *Text view*, on es visualitzen els resultats de la darrera instrucció *XPath* o *XQuery* executada per nosaltres (zona *Editor*) a la base de dades.
- Inferior dreta, anomenada *Query info panel*, on es visualitza el procés d'execució que ha seguit *BaseX* a la última instrucció executada.

Les BD que emmagatzemen documents XML, acostumen a facilitar el concepte de col·lecció per agrupar documents XML en funció de la informació que contenen. Així, les col·leccions juguen dins les BD-XML el paper de les taules dins les BD relacionals. Les col·leccions poden, a la vegada, contenir altres col·leccions. La versió de *BaseX* que acabem d'instal·lar no implementa el concepte de col·lecció. En canvi, permet crear i gestionar diverses bases de dades i en diversos punts de la documentació oficial utilitza el terme col·lecció per referir-se a base de dades.

Una vegada tenim una base de dades, ens cal incorporar-hi documents XML, fet que

6

podem assolir mitjançant l'opció Database | Properties | Resources. A nosaltres ens interessa introduir el contingut del fitxer factbook.xml que el procés d'instal·lació de *BaseX* ha deixat en el directori /etc de l'aplicació. Aquest fitxer és una versió del document mondial.xml corresponent a la base de dades *Mondial* assequible <u>aquí</u>. Nosaltres instal·larem la versió que facilita *BaseX*.

En el moment d'afegir un document XML a una base de dades, *BaseX* ens permet canviar el nom amb el què el document quedarà inserit a la base de dades. Així, la imatge següent mostra què cal introduir per a que el document factbook.xml quedi a la base de dades enregistrat com mondial.xml:

	General Resources Names Path Summary Indexes Full-Text General Parsing Input file or directory: C:/BDXML/BaseX/etc/factbook.xml Skip corrupt (non-well-formed) files Parse files in archives Target path: mondial.xml Input format: File patterns (separated by commas):	
/ Filter Clear	Implified format.       File patterns (separated by commas).         XML       *.xml         Add remaining files as raw files             Add       Optimize	

Una vegada inserit un document a la base de dades, el panell Text view mostra el contingut de la base de dades en format text, mentre que el panell *Map visualization* el mostra de forma gràfica, on es pot distingir els diversos nivells de niament si es tracta d'un document petit, cosa que no succeeix en el document que acabem d'introduir.

*BaseX* també facilita altres visualitzacions (*Tree, Plot, Folder, Table* i *Explorer*) accessibles pel menú *View* i també per botons de la botonera superior. Així mateix, tots els panells es poden activar i desactivar.

Proveu de crear altres bases de dades i afegir-hi documents XML petits, per comprovar la potència de visualització de *BaseX*.

# Consultes XPath i XQuery

L'aplicació gràfica proporcionada per *BaseX* facilita la possibilitat de practicar amb els llenguatges *XPath* i *XQuery*, a través del panell *Query Editor*.

Simplement cal introduir l'expressió a avaluar i prémer el botó *Executes the query* de la part inferior dreta del mateix panel, per procedir a l'execució de l'expressió, i obtindrem el resultat en els diversos panells activats (*Text, Map visualization, Query info panel,...*). Vegem alguns exemples, no pas com a pràctica dels llenguatges *XPath* i *XQuery*, sinó com a pràctica de llur execució dins *BaseX*.

Els exemples que segueixen, pressuposen que tenim oberta la base de dades mondial anteriorment instal·lada i que conté un únic document mondial.xml. Posteriorment s'indica com actuar en cas que la base de dades contingui més d'un document o que la base de dades estigui tancada.

1. Mostrar el contingut del document mondial.xml de la base de dades mondial (que caldrà tenir oberta):

/mondial

La següent imatge mostra el resultat d'aquesta execució:

🚟 BaseX 7.1 - mondial			
Database Editor View Nodes Options Help			
x 🗠 🖽 x 🔺 💠 🔶 🔊 🕼 🗷 🔚 🎄 🗄 🖽 🗖 🖸	1 Results		
Search v			
Editor	mondial xml           mondial         C.         C.         C.         C.         C.         C.         C.         C.         C.         O.         OT.           C.         CO.         CO.         C.         C.         C.         C.         O.         OT.         OT.		
<pre>     Text</pre>	CC. Query Info Total Time: 84.95 ms Query: /mondial Timing: - Parsing: 0.4 ms - Compiling: 4.47 ms - Evaluating: 1.45 ms - Printing: 78.62 ms - Total Time: 84.95 ms Result: - Hit(s): 1 Item - Updated: 0 Items - Printed: 1318 KB Query plan:		
db:open("mondial", "mondial.xml")	39 MB		

- Mostrar els noms de tots els països existents a la base de dades mondial:
   Tenim diverses possibilitats per obtenir la informació demanada:
  - /mondial/country/name
  - //country/name

El resultat de l'execució és similar a:

```
<name>Albania</name>
<name>Andorra</name>
<name>Austria</name>
<name>Belarus</name>
<name>Belgium</name>
<name>Bosnia and Herzegovina</name>
...
```

El panell *Query info* ens informa que la consulta facilita 239 resultats. En canvi, si executem //country, ens informa que la consulta facilita 231 resultats. Això és possible per què si investiguem una mica veurem que hi ha països que els consta dos noms... Per tant, ens cal refinar les consultes anteriors per obtenir únicament el primer dels noms que pugui tenir un país:

- /mondial/country/name[1]
- //country/name[1]

Si el que es desitjava és obtenir únicament el text existent entre les etiquetes <name> i </name>, calia executar:

- /mondial/country/name[1]/text()
- //country/name[1]/text()

Cal tenir en compte que la base de dades mondial que estem utilitzant es basa en el contingut del document /etc/factbook.xml que incorpora el procés d'instal·lació de *BaseX* i que pel món en volten moltes versions diferents. La versió vigent sempre la podrem descarregar d'aquí.

Per finalitzar aquest exemple, comentar que la següent expressió seria un error, doncs facilitaria els continguts de totes les etiquetes name, és a dir, noms de país, noms de ciutats,... doncs dins la base de dades mondial hi ha informació diversa sota etiquetes name:

- //name
- 3. Mostrar la quantitat d'habitants del país anomenat Spain a la base de dades mondial.xml:

De fet, per poder dissenyar qualsevol consulta sobre un document XML ens cal conèixer l'estructura del document. *BaseX* ens facilita la possibilitat de veure l'estructura del document via l'opció Database | Properties. Si hi naveguem, ens trobarem amb una finestra amb diferents pestanyes. A la pestanya Names hi veurem una llista dels elements i atributs continguts a la base de dades, però sense els nivells d'imbricació. A la pestanya Path Summary és on ens podrem fer una idea de l'esquema de la base de dades, després d'executar l'opció Optimize situada sota la pestanya Resources. La imatge següent ens mostra el contingut de la pestanya Path Summary després d'haver passat el procés d'optimització:



Observem que en aquesta versió de la base de dades mondial, el nombre d'habitats d'un país està sota l'atribut population de l'element country. Cal anar amb compte, doncs en altres versions de la base de dades mondial, aquesta informació resideix en un element dins el node country.

Sabent això, i tenint la base de dades mondial oberta, la consulta a executar és:

//country[name="Spain"]/@population

4. Mostrar, a partir de la base de dades mondial, el nom de tots els països acompanyat de la seva població.

En aquest cas es tracta d'una consulta XQuery, amb una instrucció FLWOR:

```
for $i in //country
return concat($i/name[1],"-",$i/@population)
```

Observem que és necessari indicar name [1] donat que aquesta base de dades incorpora, per alguns països, dos noms i, en tal situació, la funció concat () detecta un error (XPTY0004).

Una altra possibilitat, per obtenir el resultat amb millor format, és:

for \$i in //country

```
return <pais>{concat($i/name[1],"-",$i/@population)}</pais>
```

5. Mostrar, a partir de la base de dades mondial, el nom de tots els països acompanyat de la seva població, ordenada per la població en ordre descendent.

```
for $i in //country
order by number($i/@population) descending
return <pais>{concat($i/name[1],"-",$i/@population)}</pais>
```

Fixem-nos que a l'hora d'ordenar cal utilitzar la funció number () per a ordenar la població segons ordre numèric, doncs del contrari s'ordenaria segons l'ordre lexicogràfic.

6. Mostrar, a partir de la base de dades mondial, el nom del país que tingui més població.

```
let $maxNombreHabitants:=max(//country/number(@population))
for $i in //country
where $i/number(@population)=$maxNombreHabitants
return $i/name/text()
```

# **Col·leccions i documents**

El SGBD-XML *BaseX* no facilita la possibilitat de definir col·leccions dins una base de dades i utilitza el terme "col·lecció" com a sinònim de "base de dades".

En tenir una base de dades oberta, les instruccions XQuery que s'executen ho fan sobre tots els documents continguts a la base de dades. Així, les instruccions dels exemples anteriors s'executen sobre tots els documents existents a la base de dades. En el nostre cas, si només teníem introduït a la base de dades mondial el document mondial.xml, és clar que totes les instruccions s'han executat sobre aquest únic document. Però... com hem d'actuar si tenim, a la base de dades, varis documents i només volem executar la consulta sobre algun d'ells?

En aquest cas cal utilitzar la funció collection(), que ens retorna els documents existents a la base de dades activa, en una instrucció FLWOR amb clàusula where per filtrar el document al que interessa accedir:

```
for $doc in collection()
where document-uri($doc)=<nomDocument>
return <consultaQueCorrespongui>
```

Per altra banda, *BaseX* ens permet efectuar consultes sobre una base de dades no activa (no oberta). En aquest cas, haurem d'utilitzar la funció collection() indicant, via paràmetre. el nom de la base de dades amb la que volem treballar:

collection(<nomBaseDeDades>)/<consultaQueCorrespongui>

Si a més d'indicar la base de dades, cal indicar el document amb el què treballar, utilitzarem:

for \$doc in collection(<nomBaseDeDades>)

```
where document-uri($doc) = < nomDocument>
return < consultaQueCorrespongui>
```

Per finalitzar, si volem obtenir un llistat de les bases de dades (col·leccions) existents, disposem de la funció <u>db:list()</u>. Així, per exemple, si volem obtenir una llista de totes les bases de dades amb els seus documents, escriuríem:

```
for $i in db:list()
return
<collection nom="{$i}">
    {
    for $doc in collection($i)
    return <document>{document-uri($doc)}</document>
    }
</collection>
```

El resultat seria similar a:

```
<collection nom="mondial">
<document>mondial.xml</document>
<document>...</document>
</collection>
<document>...</document>
</collection>
...
```

# Instruccions Update

El SGBD *BaseX* facilita una completa implementació de <u>XQuery Update Facility (XQUF)</u> recomanat per W3C. Així doncs, *BaseX* ens permet inserir, eliminar, reemplaçar i reanomenar nodes dels documents XML.

XQUF no facilita instruccions específiques per gestionar transaccions i s'entén que cada execució d'una instrucció XQUF és una transacció. En cas que vulguem executar en una transacció diverses instruccions, cal posar-les en seqüència, separades per una coma.

Els exemples que segueixen, pressuposen que tenim oberta la base de dades mondial anteriorment instal·lada i que conté un únic document mondial.xml

1. Inserir, després del continent Àsia, el node següent:

```
<institut nom="IOC">IOC</institut>
```

Per poder efectuar aquesta inserció ens cal mirar el document mondial.xml i observar com conté la informació.

```
<mondial>
  <continent id="f0_119" name="Europe"/>
  <continent id="f0_123" name="Asia"/>
  <continent id="f0_126" name="America"/>
  <continent id="f0_129" name="Australia/Oceania"/>
  <continent id="f0_132" name="Africa"/>
  ...
```

Ara ja podem procedir a inserir el nou node:

```
insert node <institut nom="IOC">IOC</institut>
after //continent[@name="Asia"]
```

# O també:

```
insert node <institut nom="IOC">IOC</institut>
before //continent[@name="America"]
```

#### En qualsevol cas, el resultat és:

```
<mondial>
<continent id="f0_119" name="Europe"/>
<continent id="f0_123" name="Asia"/>
<institut nom="IOC">IOC</institut>
<continent id="f0_126" name="America"/>
<continent id="f0_129" name="Australia/Oceania"/>
<continent id="f0_132" name="Africa"/>
...
```

2. Substituir el valor del node que acabem d'inserir pel valor "Institut Obert de Catalunya"

```
replace value of node //institut[@nom="IOC"]
with "Institut Obert de Catalunya"
```

# El resultat és:

```
<mondial>
<continent id="f0_119" name="Europe"/>
<continent id="f0_123" name="Asia"/>
<institut nom="IOC">Institut Obert de Catalunya</institut>
<continent id="f0_126" name="America"/>
<continent id="f0_129" name="Australia/Oceania"/>
<continent id="f0_132" name="Africa"/>
...
```

3. Substituir el node institut amb atribut nom="IOC" pel node:

```
<cicle>
   <codi>DAM</codi>
   <nom>Desenvolupament d'aplicacions multiplataforma</nom>
</cicle>
```

#### Instrucció:

```
replace node //institut[@nom="IOC"]
with
<cicle>
     <codi>DAM</codi>
     <nom>Desenvolupament d'aplicacions multiplataforma</nom>
```

M06. Accés a dades

</cicle>

#### Resultat:

```
<mondial>
  <continent id="f0_119" name="Europe"/>
  <continent id="f0_123" name="Asia"/>
   <cicle>
        <codi>DAM</codi>
        <nom>Desenvolupament d'aplicacions multiplataforma</nom>
   </cicle>
   <continent id="f0_126" name="America"/>
   <continent id="f0_129" name="Australia/Oceania"/>
   <continent id="f0_132" name="Africa"/>
   ...
```

#### 4. Reanomenar els nodes cicle per cf:

rename node //cicle as "cf"

# Resultat:

```
<mondial>
  <continent id="f0_119" name="Europe"/>
  <continent id="f0_123" name="Asia"/>
  <cf>
      <codi>DAM</codi>
      <nom>Desenvolupament d'aplicacions multiplataforma</nom>
  </cf>
  <continent id="f0_126" name="America"/>
  <continent id="f0_129" name="Australia/Oceania"/>
  <continent id="f0_132" name="Africa"/>
  ...
```

5. Eliminar els nodes cf:

```
delete node //cf
```

## Resultat:

```
<mondial>
  <continent id="f0_119" name="Europe"/>
  <continent id="f0_123" name="Asia"/>
  <continent id="f0_126" name="America"/>
  <continent id="f0_129" name="Australia/Oceania"/>
  <continent id="f0_132" name="Africa"/>
  ...
```

# Accés als documents XML via protocol WebDAV

El SGBD *BaseX* proporciona el protocol d'accés WebDAV, fet que possibilita que el repositori de documents XML del SGBD *BaseX* pugui ser mapat com una unitat de xarxa i, en conseqüència, fer-lo accessible com una carpeta més del sistema d'arxius del sistema

14

#### operatiu.

La majoria de sistemes operatius moderns proporcionen suport per WebDAV, fet que vol dir que incorporen un client WebDAV amb el que podem accedir al servidor que facilita el protocol d'accés WebDAV.

El SGBD BaseX proporciona accés WebDAV via la URL http://localhost:8984/webdav

Per a comprovar que el servidor WebDAV de *BaseX* està en funcionament, podeu provar a introduir l'anterior URL en un navegador; el servidor WebDAV de *BaseX* us demanarà l'usuari i contrasenya i posteriorment us mostrarà una pàgina en blanc, per què aquest servidor WebDAV no mostra els continguts en el navegador. Si no arriba a demanar l'usuari i contrasenya és que no teniu el servidor WebDAV de *BaseX* en funcionament.

Windows 7/XP incorporen un client WebDAV. A l'annex <u>Utilització del client WebDAV de</u> <u>Windows 7</u> s'explica la seva configuració.

Hi ha, però, mots clients WebDAV en el mercat. Algunes conegudes eines d'edició XML, com *oXygen XML Editor* o *XML Spy*, també n'incorporen un.

Com a clients WebDAV gratuïts trobem NetDrive, AnyClient, EngInSite Freeway, Glace Space, entre altres. A l'annex <u>Utilització del client WebDAV del programari NetDrive</u> s'explica la instal·lació i configuració del client WebDAV NetDrive, que permet l'accés als servidors WebDAV via carpetes del sistema d'arxius. Cal tenir en compte que alguns clients WebDAV només permeten transferència de fitxers, de forma similar al clients FTP.