

XQuery your Filesystem

An XML database as Filesystem in Userspace

Alexander Holupirek, Christian Grün, Marc H. Scholl

Demo Setup

- ① Import file hierarchy into XML database
- ② Extract metadata based on file types
- ③ Store BLOBs in backing store
- ④ Mount database as filesystem in userspace
- ⑤ Use XPath/XQuery on filesystem data
- ⑥ Use conventional OS applications

XML Filesystem View

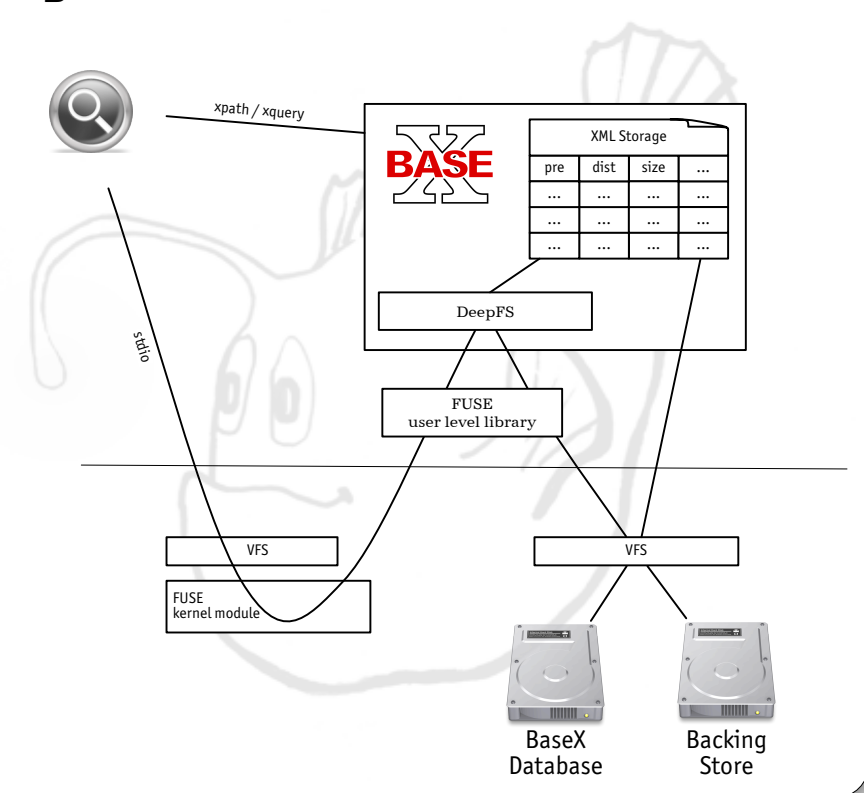
Via Extractors we store queryable metadata for specific file types in XML.

```
<?xml version="1.0" encoding="UTF-8"?>
<deepfs mountpoint="/mnt/deepfs" backingstore="/var/tmp/deepfs/demo"
<dir name="data" suffix="" size="112185" mtime="20623227">
  <file name="kmsk.xml" suffix="xml" size="11695" mtime="20623230">
    <content>
      <!--
      <regions>
        <item id="item0">
          <location>United States/Location
            <quantity>14/quantity
            <name>Dutecous nine eighteen/namc
              <payment>Creditcard/payment
            </item>
          </regions>
        </content>
      </file>
    <dir name="docs" suffix="" size="67955" mtime="20623204">
      <file name="services" suffix="txt" size="3311" mtime="20623204">
        <content>
          # Network services, Internet style
          # Note that it is presently the policy of IANA to assign a single well-known
          # port number for both TCP and UDP; hence, most entries here have two entries
          # even if the protocol doesn't support UDP operations.
          # The latest IANA port assignments can be gotten from
          # http://www.iana.org/assignments/port-numbers
          # The Well Known Ports are those from 0 through 1023.
          # The Registered Ports are those from 1024 through 49151.
          # The Dynamic and/or Private Ports are those from 49152 through 65535.
          # FreeBSD: sct/et/services,v 1.89 2002/12/17 23:59:10 Eric S
          # From: #services 5.8 (Berkeley) 5/9/91
          # WELL KNOWN PORT NUMBERS
          #
          #Routing Table Maintenance Protocol
          rtmp 1/udp #TCP Port Service Multiplexer
          tcpmux 1/udp
          netjs-2f/content
        </dir>
      <file name="images" suffix="jpg" size="3311" mtime="20623201">
        <content>
          <!--
          <img type="jpg">
            <width>364/width
            <height>399/height
          </img>
        </content>
      </file>
    <dir name="mail" suffix="" size="25316" mtime="20624481">
      <file name="1.mbox" suffix="mbox" size="8932" mtime="20623192">
        <content>
          <!--
          <date time="2051779"/>
          <from>eric@math.spbu.ru [mailto:eric@math.spbu.ru]
          <to>dbwoid@ca.wisc.edu [mailto:dbwoid@ca.wisc.edu]
          <subject>[DBWoid] ICDT/EDBT 2009 - Call for Participation/</subject>
          <content-type>text/plain;content-type
          <section-type
          <!--
          *** CALL FOR PARTICIPATION ***
          <!--
          12th International Conference on Database Theory (ICDT 2009)
          The 12th International Conference on Extending Database
          Technology (EDBT 2009)
          St. Petersburg, Russia, 23-26 March 2009
          http://www.math.spbu.ru/edbticdt/
        </section>
      </file>
    <dir name="music" suffix="" size="894304" mtime="20623204">
      <file name="05 Like A Rolling Stone.mp3" suffix="mp3" size="894304">
        <content>
          <!--
          <audio type="mp3">
            <codec>MPEG-1 Layer 3/Codec
            <bitrate>192/kbit/s
            <sample-rate>44100/Samplerate
            <encoding>Stereo
            <emphasis>None/Emphasis
            <mode>Stereo/Mode
            <extended-headers>
              <comment>372/seconds
            </extended-headers>
            <id3 version="3.0" extended-headers="false" experimental="false">
              <title>Like A Rolling Stone/Title
              <person>Bob Dylan/Person
              <composer>Bob Dylan/Composer
              <album>Greatest Hits/Album
              <track>5/16/Track
              <year>1965/Year
              <genre>Rock/Genre
              <comment>94/Comment
            </id3>
          </audio>
        </content>
      </file>
    </dir>
  </deepfs>

```

General approach:
 All queryable data in the filesystem shall be stored in the database.
 As such, querying the filesystem is querying XML.
 (Following Kristoffer Rose's XML World View)

System Architecture



EDBT/ICDT 2009 joint conference
 March 23-26 2009, Saint-Petersburg, Russia

Universität
 Konstanz

