Seismological software developments at LMU Munich: Python & ObsPy

Robert Barsch, Tobias Megies

Department für Geo- and Umweltwissenschaften (Geophysik)
Ludwig-Maximilians-Universität München

2011-03-08

SeisHub

Native, document-centric XML database

- RESTful Web service (HTTP, HTTPS),
- Standard relational database as back-end
- Both worlds: SQL for querying and manipulating data and any standard connected to XML, e.g. XSLT or XSD
- Not restricted to seismology at all

Extended to a "classical" seismic database

- Index of local, file-based waveform archive (MiniSEED, GSE2, SAC, ...)
  - ► Meta Data: Gaps, overlaps, quality and timing information
  - ► Waveform previews (30s)
- XML resource types for handling inventory data (XML-SEED) and events (QuakeML based)
- Remote waveforms access (ArcLink)

SeisHub: Technical Details

- Python-based, standalone web service
- Platform independent, open source (GPL)
- Implementation of various web protocols, like HTTP, SSH, SFTP
- Plug-in architecture: Dynamic discovering and loading of modules and support for Python .egg files
- Development remarks:
  - ▶ Test-driven development proven software, so far $\Rightarrow$ ca. 250 test cases
  - ▶ Well-documented source code
  - ▶ Subversion
  - ▶ Trac: ticket system and project wiki

SeisHub: Database Design

Data storage

- Primary data $\Rightarrow$ file system
  - ▶ Continuous waveform archive (MiniSEED, GSE2, SAC ...)
  - ▶ Other data via (GeoTIFF, GPS time series, etc.) file system
- Meta Data $\Rightarrow$ Web service on top of a XML/relational database hybrid
  - ▶ Data is packed into a XML document $\Rightarrow$ Data structure is within the document, no need for a predefined database schema
  - ▶ XML resources are archived into a BLOB field
  - ▶ Only searchable values are indexed
  - ▶ Pointers to primary data

SeisHub: Database Design

Data access

- HTTP/HTTPS: REST web service
  - ▸ XML documents have a fixed resource identifier (URL's)
  - ▸ Data transformation via XML Style Sheets on request (?output=...)
  - ▸ Data validation via Schema (XML Schema, RelaxNG, Schematron) on resource upload
  - ▸ Document properties like related meta data or indexes
- SFTP: XML documents mapped into a virtual file system

SeisHub: Database Design

- Indexing
  - Generated using a XPath expression, type and additional options
  - Simple creation + reindexing via web interface
  - Various build-in types (datetime, bool, numeric, double, float, etc..)
  - ProcessorIndex: custom processing
- Searching
  - XPath-like query on XML catalog object (restricted to indexes)
  - SQL on database object
- Mapper: predefined queries & output format bound to an fixed URL
- FileSystemResource: integrates a file system directory (read only)

SeisHub: Advantages

Technical:

- Sharing data over the network, but no firewall problems (HTTP / HTTPS)
- License free, open source, internet standards
- Platform independent
- Most basic client is a standard browser
- XML:
  - ▶ Data validation on upload (XML schemas)
  - ▶ Data transformation on request (XML stylesheets)
- Querying: SQL or XPath

Scientist:

- May modify there data provided as XML document at any time without corrupting the underlying database
- May dynamically add or delete search indexes, schemas and stylesheets

SeisHub: Disadvantages

Technical:

- Slower than "common" solutions
  - ▸ XML parsing during validation and indexing
  - ▸ Data overhead (XML verbosity)
- Infrastructure

Scientist:

- Seismologist != IT nerds