

Introduction to Python Persistence / Serialization

Christopher Barker

UW Continuing Education / Isilon

September 05, 2012

Table of Contents

- 1 Review/Questions
- 2 Serialization / Persistence
- 3 Python Specific Formats
- 4 Interchange Formats
- 5 DataBases
- 6 Object-Relational Mappers
- 7 Object databases

Review of Previous Class

Lightning talks today: Chris and John

- doctests
- unittests
- profiling

Homework review

Homework notes:

Anyone add some doctests or unittests to his project?

Anyone time or profile their project?

Did you find some bottlenecks??

Serialization

I'm focusing on methods available in the Python standard library

Serialization is the process of putting your potentially complex (and nested) python data structures into a linear (serial) form .. i.e. a string of bytes

The serial form can be saved to a file, pushed over the wire, etc.

Persistence

Persistence is saving your python data structure(s) to disk – so they will persist once the python process is finished.

Any serial form can provide persistence (by dumping/loading it to/from a file), but not all persistence mechanisms are serial (i.e RDBMS)

Python Literals

Putting plain old python literals in your file

Pickle

Pickle is a binary format for python objects

Shelve

A shelf is a persistent, dictionary-like object

JSON

JSON (JavaScript Object Notation) is a subset of JavaScript syntax used as a lightweight data interchange format.

XML

XML is a standardized version of SGML, designed for use as a data storage/interchange format.

NOTE: HTML is also SGML, and modern versions conform to the XML standard.

XML looks like:

....

XML in the python std lib

`xml.dom:`

`xml.sax:`

`xml.parsers.expat:`

`xml.etree:`

anydbm

anydbm is a generic interface to variants of the DBM database (clones of the bsd dbm system)

sqlite

SQLite is a C library that provides a lightweight disk-based single-file database that doesn't require a separate server process and allows accessing the database using a nonstandard variant of the SQL query language

SQLite is very broadly used as as an embedded databases for storing application-specific data etc. (Picassa, Apple tools, on iOS/Android, small spatial databases, etc....

DB-API

The DB-API spec (PEP 249) is a specification for interaction between Python and Relational Databases. Support by a large number of third-party Database drivers:

- MySQL
- PostgreSQL
- Oracle
- MSSQL (?)
-

<http://www.python.org/dev/peps/pep-0249>

Object Databases

Python objects:

- ZODB
- Durus

NoSQL

Map-Reduce, etc.... What to say here???

Document-Oriented Storage:

- MongoDB (BSON interface, JSON documents)
- CouchDB (Apache):
 - JSON documents
 - Javascript querying (MapReduce)
 - HTTP API

Review/Questions
Serialization / Persistence
Python Specific Formats
Interchange Formats
DataBases
Object-Relational Mappers
Object databases

LAB



Lightning Talk

Lightning Talk:

Chris

Lightning Talk

Lightning Talk:

Peter

First Topic

A topic

some code example

Review/Questions
Serialization / Persistence
Python Specific Formats
Interchange Formats
DataBases
Object-Relational Mappers
Object databases

LAB



Review/Questions
Serialization / Persistence
Python Specific Formats
Interchange Formats
DataBases
Object-Relational Mappers
Object databases

Wrap up



Homework

Send me a copy of your project: due next Wed.

Keep learning about and using Python