

Start of P6 thoughts

Mathias Claus Jensen

February 15, 2016

1 Initial Project Idea

The following is a summary of the project proposal.

There are currently no way of doing automated tests for the python package pygrametl. It is suggested that in this project it should be investigated how to do regression testing for pygrametl. Some requirements for the test are that they should be fast, and should be able to run on atleast Linux, Mac OSX and Windows. It is suggested to use SQLite for creating and tearing down databases. It is suggested that a succesfull implementation of automatic tests for pygrametl could lead to a module that can assist users in testing of their programs.

This summary suggests several areas of interest, such as what is *pygrametl* and how can automated tests be implemented, what is *automated tests* and *regression testing*, how do we make sure that it runs and fast and is portable and lastly how could this lead to a module that can assist users in testing their own *pygrametl* programs.

This initial idea could also be split up into three sequential process.

- First: Describe theoraticaly how to test ETL systems efficiantly
- Second: Implement this for pygrametl
- Third: Further develop this implementation into a module for users to use

The second and third step could be solved in one step, it could be viable to build a module for users to use that could used to do automatic tests for pygrametl.

2 A few clarification regarding terminology

In the above summary of the given problem, some terminology was used, as we are new to this shit, this section will serve to help clarifying some of it, i hope.

- **ETL:** Extract-Transform-Load. You extract data from homogenous or heterogenous data sources, transform it to a proper structure/format for querying and analysis and then load that stuff into the target (database, datawarehouse, wateva). The thing about ETLs is that they are good at getting data from a lot of different sources and then transforming that into a uniform structure that is usable. [1]
- **Regression Testing:** Testing to see that changes in the system doesn't introduce non-wanted behaviour. The test first strategy is a form of regression testing and it fits well with agile development. Usually regression testing is done by the use of unit, functional or non-functional tests. [2]
- **Functional and Non-Functional Tests:** Functional tests are tests on which a larger parts of a systems functionality is tested. Non-functional tests are tests that test the systems not functional requirements, such as run-time, system loads, etc. [3, 4]
- **Manuel Tests or Test Automation:** Test automation is the opposite of manual testings, in which you as a tester do tests of your code. Test automation is having software that runs your tests for you. It is very related to Regression testing, as you typically do a bunch of regression tests and then use them with test automation. An example of some software that does automated testing would be the test suite in visual studio. [5, 6]

References

- [1] Wikipedia article on ETL, https://en.wikipedia.org/wiki/Extract,_transform,_load, 2016.
- [2] Wikipedia article on Regression Testing, https://en.wikipedia.org/wiki/Regression_testing, 2016.
- [3] Wikipedia article on Functional Tests, https://en.wikipedia.org/wiki/Functional_testing, 2016.

- [4] Wikipedia article on Non-Functional Tests,
https://en.wikipedia.org/wiki/Non-Functional_testing, 2016.
- [5] Wikipedia article on Manual Testing,
https://en.wikipedia.org/wiki/Manual_testing, 2016.
- [6] Wikipedia article on Test Automation,
https://en.wikipedia.org/wiki/Test_automation, 2016.