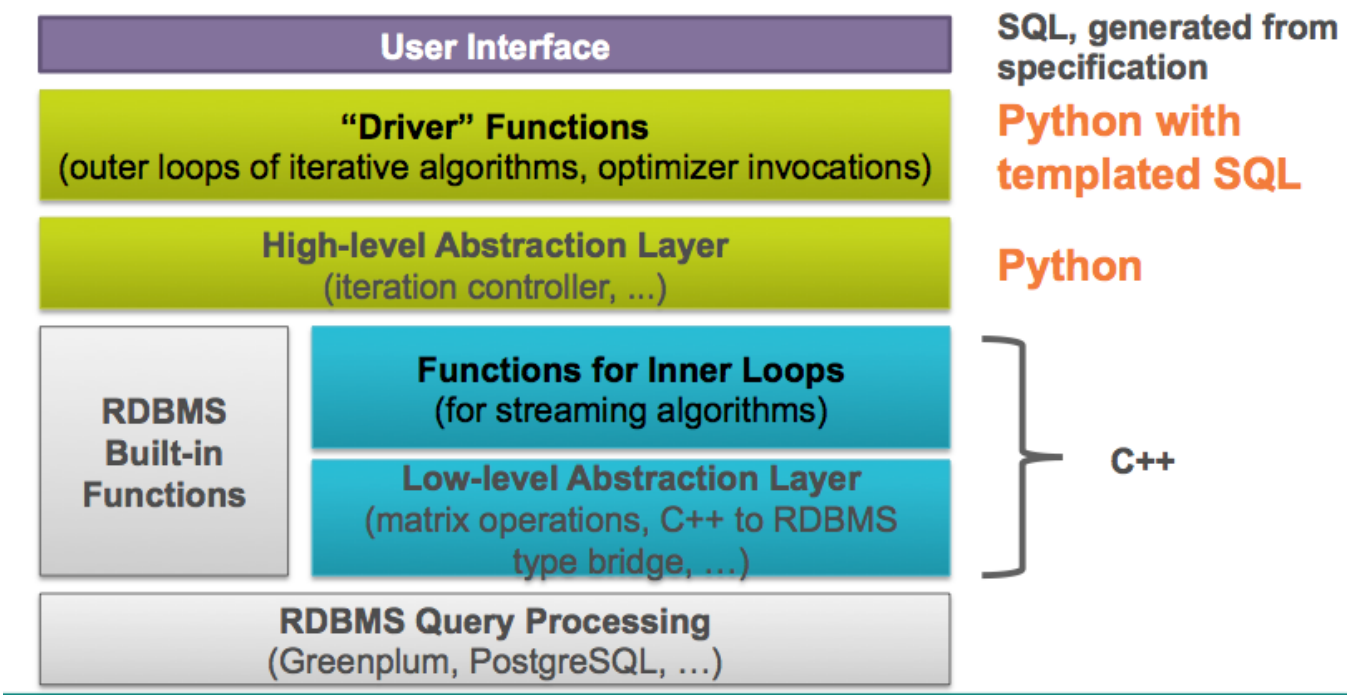
Architecture



You can think of MADlib® as having the following major components:

1. Python driver functions
2. C++ implementations functions
3. C++ database abstraction layer

Below is a brief explanation of each of these.

1. Python Driver Functions

The driver functions are mostly located in the subdirectories under https://github.com/apache/madlib/tree/master/src/ports/postgres/modules

These functions are the main entry point from user input and are largely responsible for the flow control of the algorithms. Generally, the implementation consists of validating input parameters, executing SQL statements, evaluating the results and potentially looping to execute more SQL statements until some convergence criteria has been hit.

2. C++ Implementation Functions

Mostly located under https://github.com/apache/madlib/tree/master/src/modules

These functions are the C++ definitions of the core functions and aggregates needed for particular algorithms. These are implemented in C++ rather than Python for performance reasons.

3. C++ Database Abstraction Layer

Mostly located under https://github.com/apache/madlib/tree/master/src/dbal

and

https://github.com/apache/madlib/tree/master/src/ports/postgres/dbconnector

These functions attempt to provide a programming interface that abstracts all the Postgres internal details away and provides a mechanism whereby MADlib can support different backend platforms and focus on the internal functionality rather than the platform integration logic.