



使用 **luiti** 来构建数据仓库

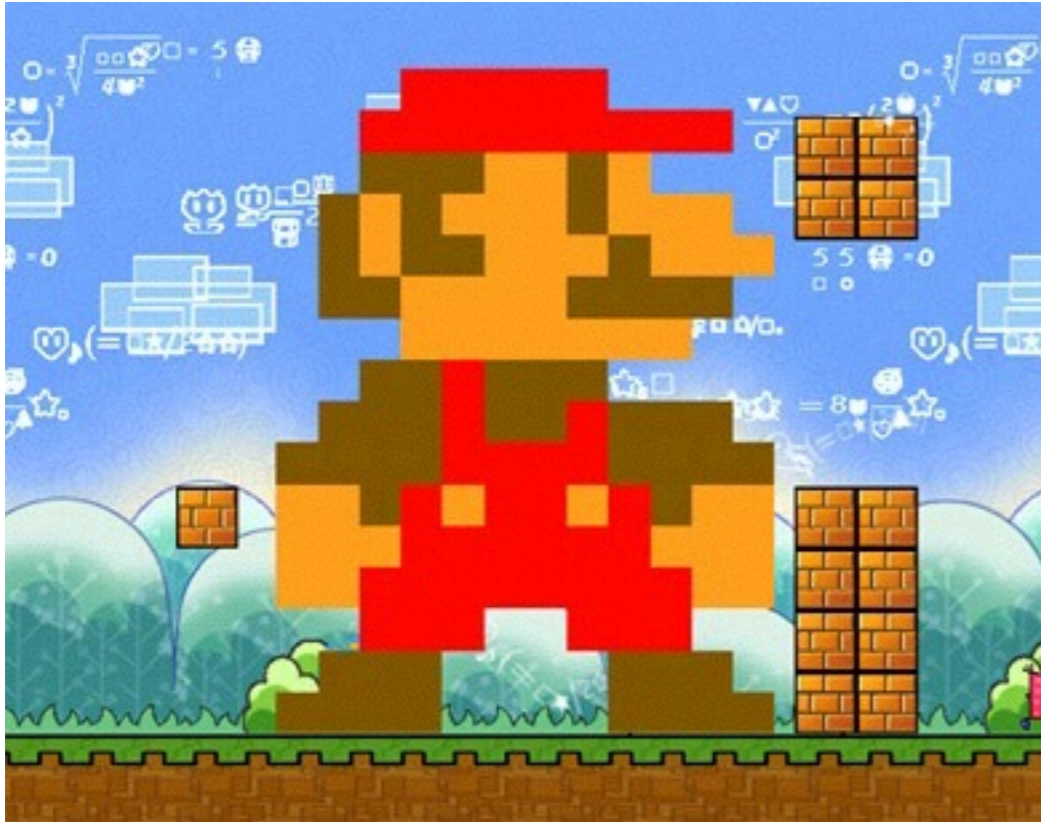
<http://github.com/luiti>

David Chen



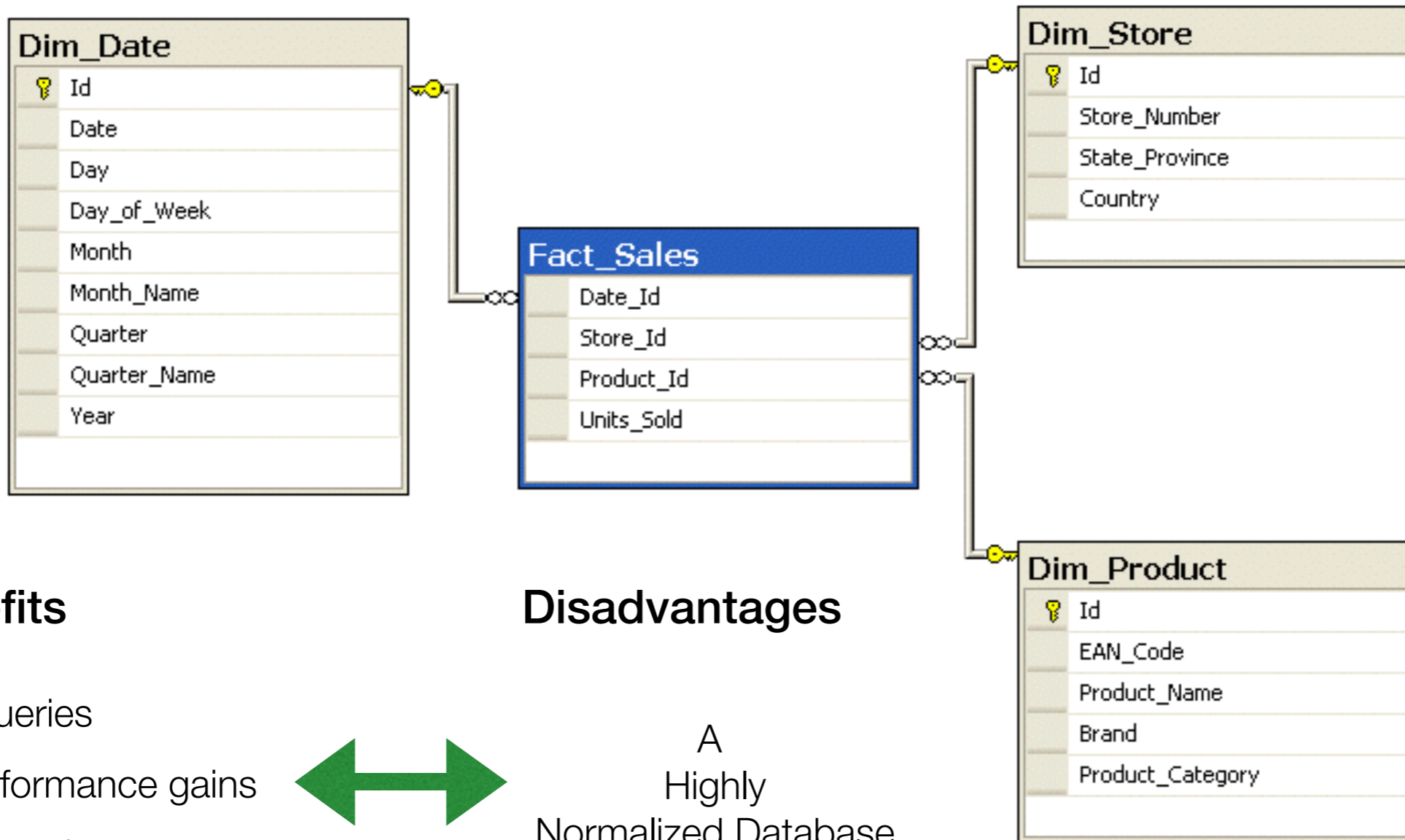
mvj3.com

Luiti is built on top of **pygame**



Luigi was built at  **mainly**

Star Schema



Benefits

- 1. Simpler queries
- 3. Query performance gains
- 4. Fast aggregations



Disadvantages

A
Highly
Normalized Database

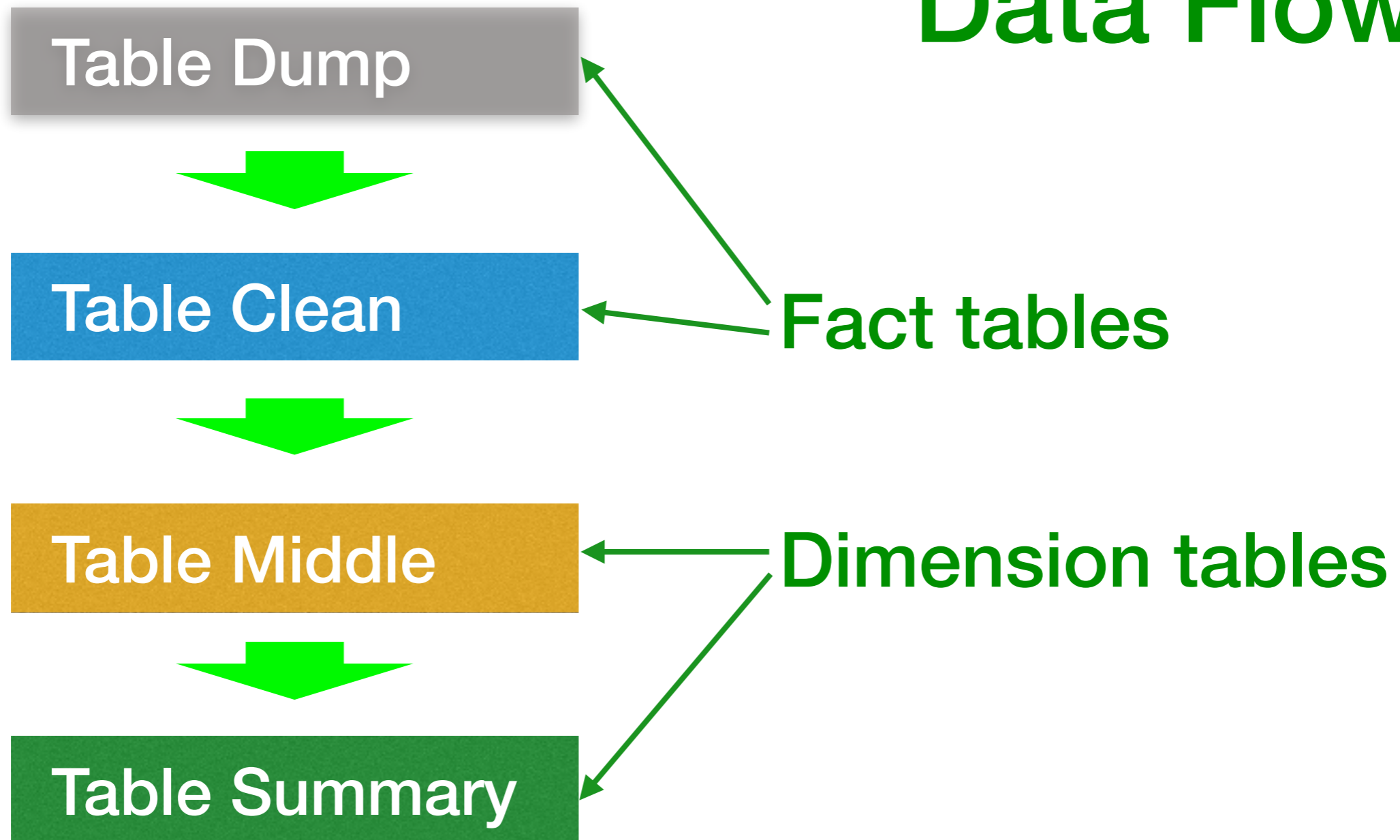
Ideal mode?!

https://en.wikipedia.org/wiki/Star_schema

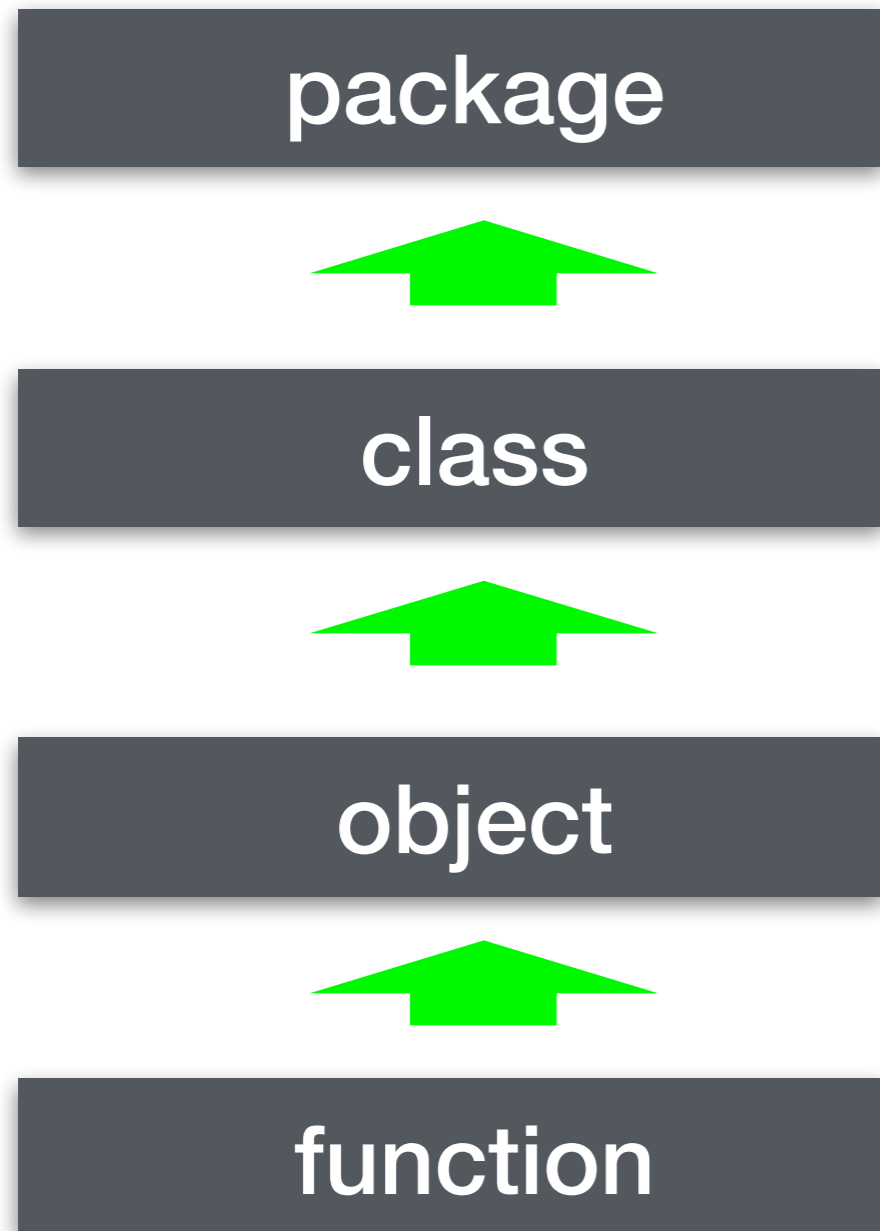
<http://emhughes.com/arbitrary/observation-starfish-cool/>

Hierarchical Data Warehouse

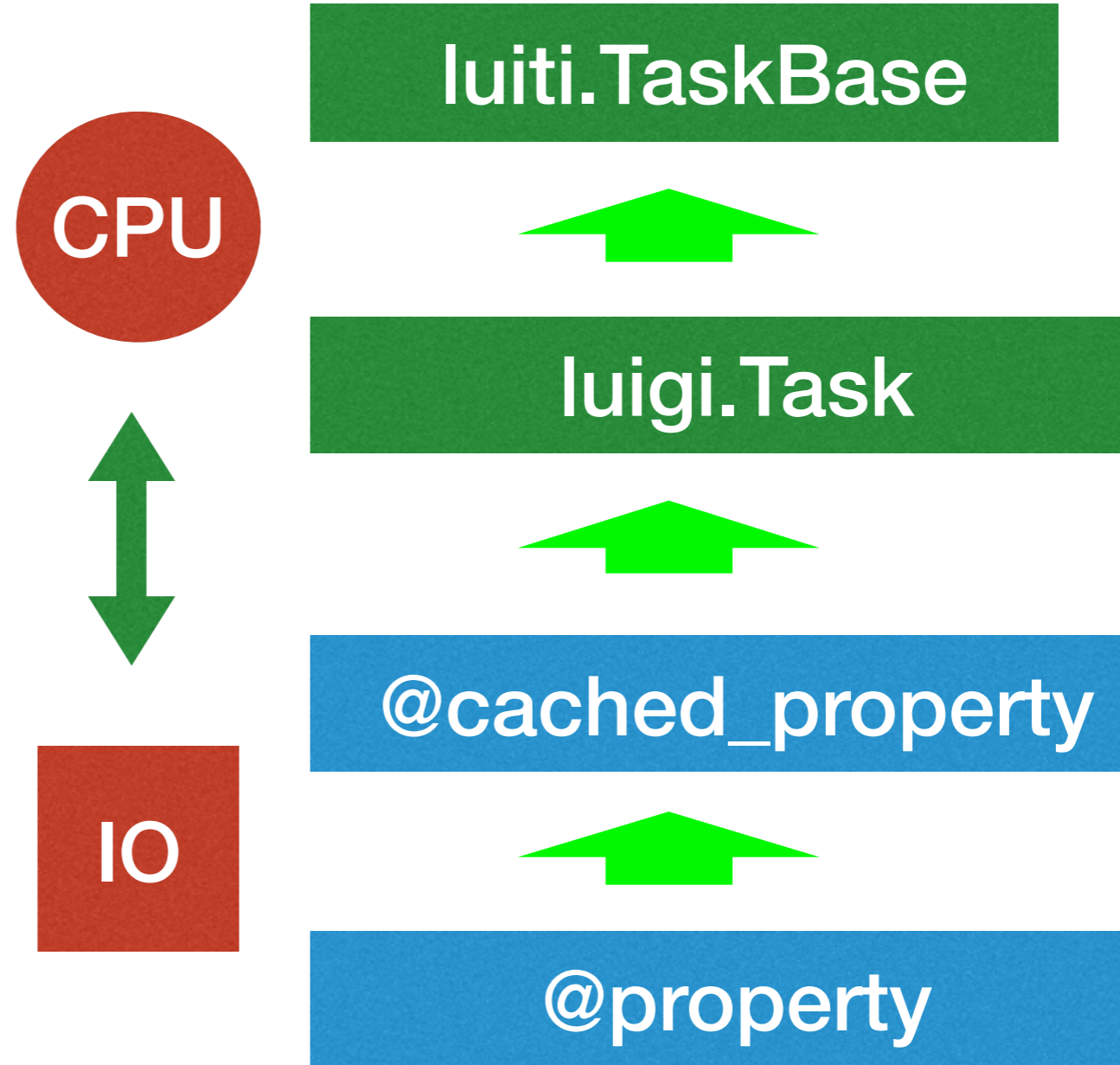
Data Flow



Context

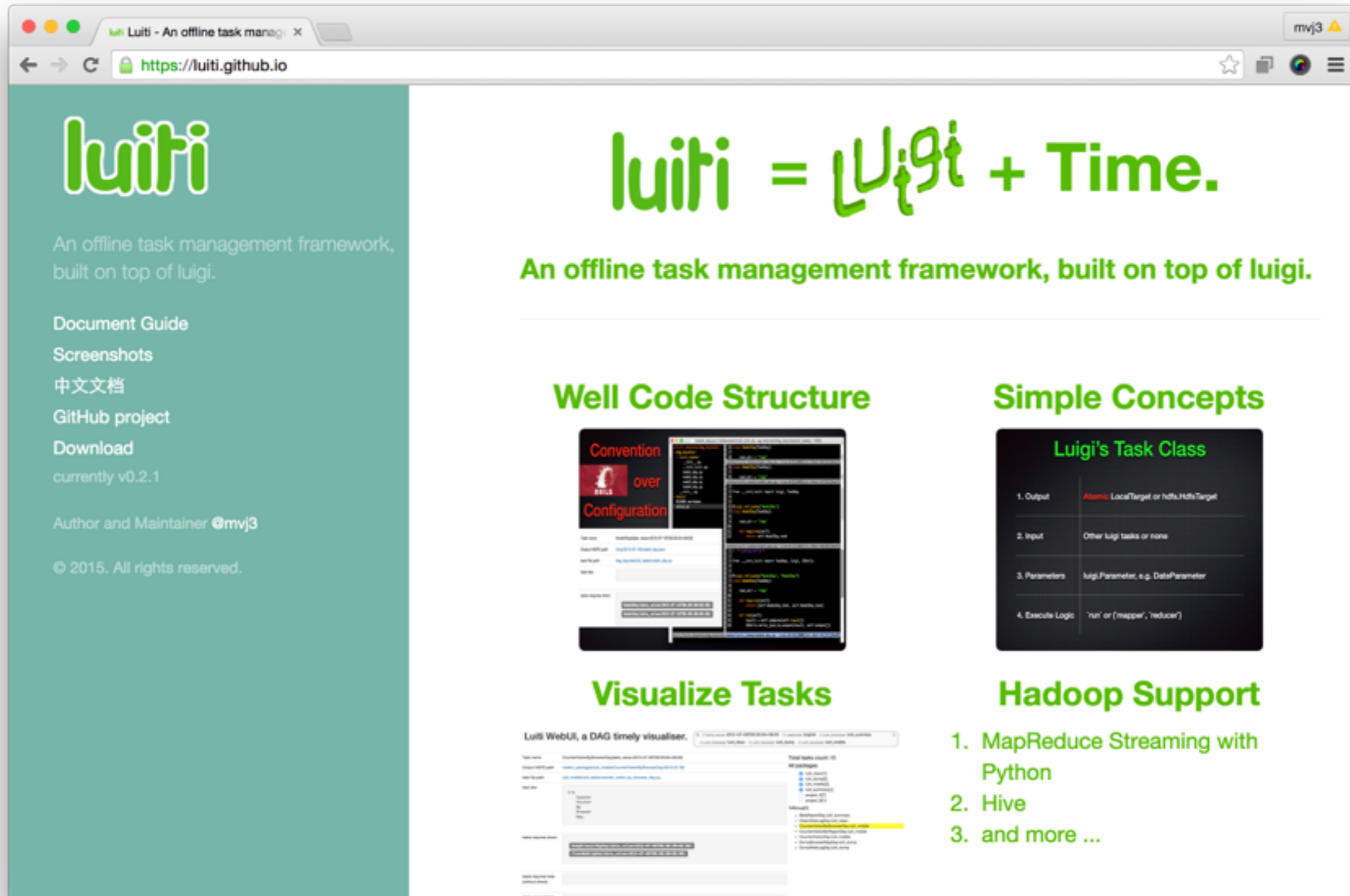


Abstract Machine



Luiti Demo!

http://luiti.github.io



The screenshot shows the homepage of luiti, an offline task management framework. The page features a teal sidebar on the left with navigation links: Document Guide, Screenshots, 中文文档, GitHub project, Download (currently v0.2.1), and Author and Maintainer @mvj3. The main content area has a large heading 'luiti = Luigi + Time.' and a sub-heading 'An offline task management framework, built on top of luigi.' Below this are four key features: 'Well Code Structure' (with a code editor screenshot), 'Simple Concepts' (with a table of Luigi's Task Class), 'Visualize Tasks' (with a DAG visualizer screenshot), and 'Hadoop Support' (with a list of supported technologies).

luiti

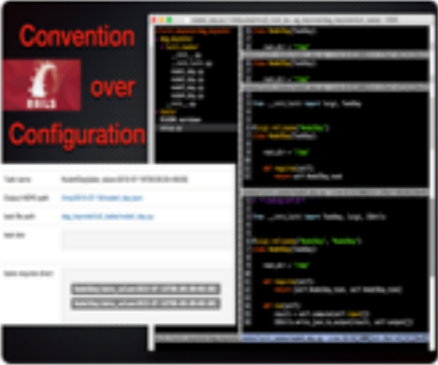
An offline task management framework, built on top of luigi.

- Document Guide
- Screenshots
- 中文文档
- GitHub project
- Download
currently v0.2.1
- Author and Maintainer @mvj3
- © 2015. All rights reserved.

luiti = Luigi + Time.

An offline task management framework, built on top of luigi.


Well Code Structure



Simple Concepts

Luigi's Task Class	
1. Output	Atomic, LocalTarget or hdfs.HdfsTarget
2. Input	Other luigi tasks or none
3. Parameters	luigi.Parameter, e.g. DataParameter
4. Execute Logic	'run' or ('mapper', 'reducer')

Visualize Tasks



Hadoop Support

1. MapReduce Streaming with Python
2. Hive
3. and more ...

Questions?
