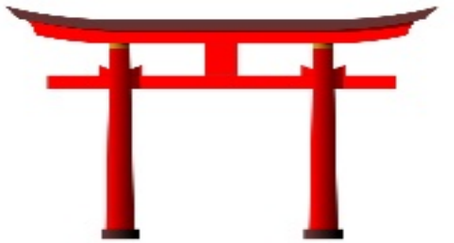
# Apache Toree:
## A Jupyter Kernel for Spark

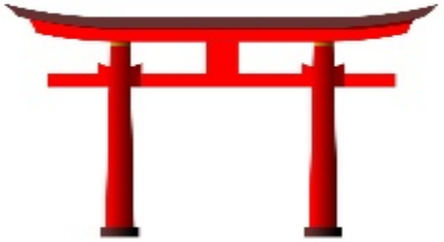
Marius van Niekerk
Apache Toree Contributor

# Apache Toree



A **Jupyter** kernel to connect to **Spark** and create **interactive applications**

# Ecosystem

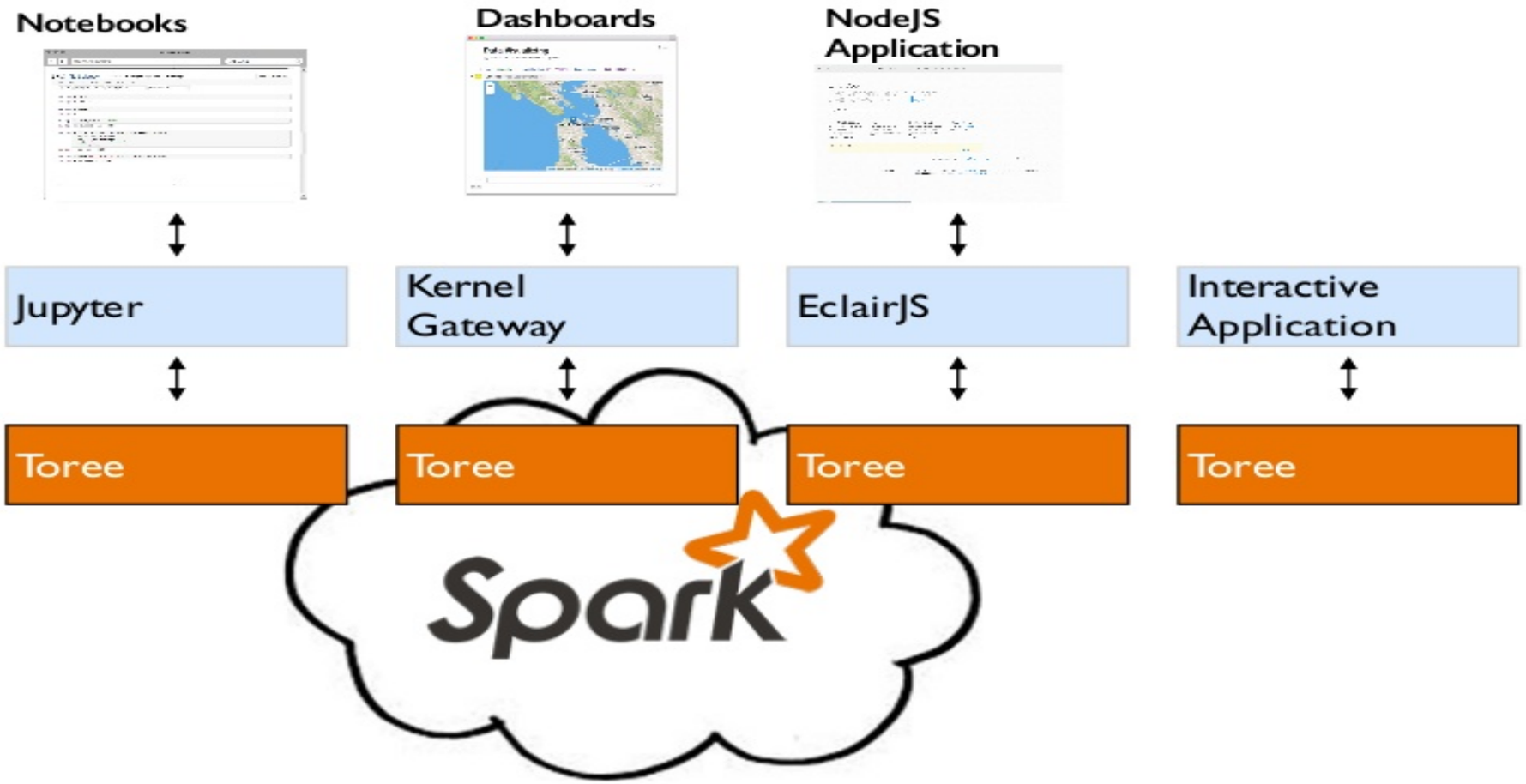


# Jupyter

- Open source, interactive data science and scientific computing across over 40 programming languages.
- Toree is an implementation of the Jupyter Kernel Protocol

# Apache Toree History

- Started as ibm/spark-kernel in 2013
- Started Apache incubation in late 2015
- First Apache Release is coming soon.


Notebooks
Dashboards
NodeJS Application
Jupyter
Kernel Gateway
EclairJS
Interactive Application
Toree
Toree
Toree
Toree
Spark

# Compatibility

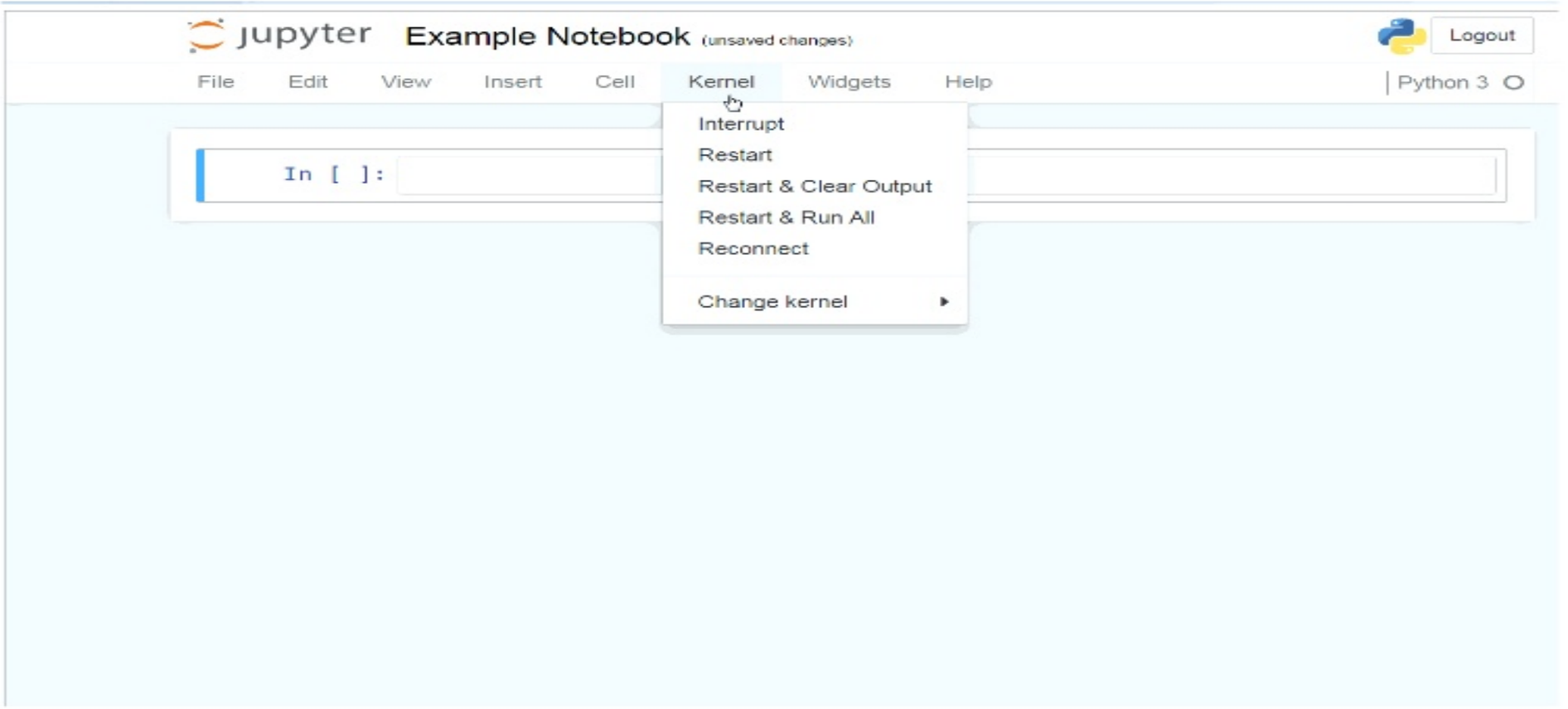
- Toree 0.1.x supports Spark 1.6.x
- Toree 0.2.x supports Spark 2.x

# Features

- Kernel languages: Scala, Python, R
- Magics
- Tab completion
- Plugin system

# Try it out

- Welcome to Spark with Scala:
  - https://tmpnb.org
- Docker
  - docker run -it --rm -p 8888:8888 \
    jupyter/all-spark-notebook
  - This is a very large container ~4gb


jupyter
Example Notebook
(unsaved changes)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Python 3
In [ ]:
Interrupt
Restart
Restart & Clear Output
Restart & Run All
Reconnect
Change kernel

# Examples

- github.com/Lull3rSkat3r/apache-toree-demos
- github.com/apache/incubator-toree/tree/master/etc/examples/notebooks

# Extending on top of Toree

- github.com/Brunel-Visualization/Brunel
- github.com/jupyter/declarativewidgets

# Other Alternatives

- sparkmagic (livy)
- jupyter-scala
- Apache Zeppelin
- spylon-kernel
- databricks

# Help out

Looking for contributors

- Web - toree.apache.org
- Mailing - dev@toree.incubator.apache.org
- Chat - gitter.im/apache/toree

# Thank You.

Github: @mariusvniekerk

Twitter: @__mvn__