

Apache Toree

Corey Stubbs

Follow Along

Slides and Code can be found here:

<https://github.com/Lull3rSkat3r/apache-toree-demos>

Try Apache Toree

Welcome To Spark with Scala:

<https://try.jupyter.org/>

IBM Data Science Experience:

<http://datascience.ibm.com/>

Apache Toree

Started as Spark-Kernel in 2013

Started Apache incubation in late 2015

First Apache Release is coming soon

Apache Toree

A Jupyter kernel to connect to Spark and create interactive applications

Ecosystem



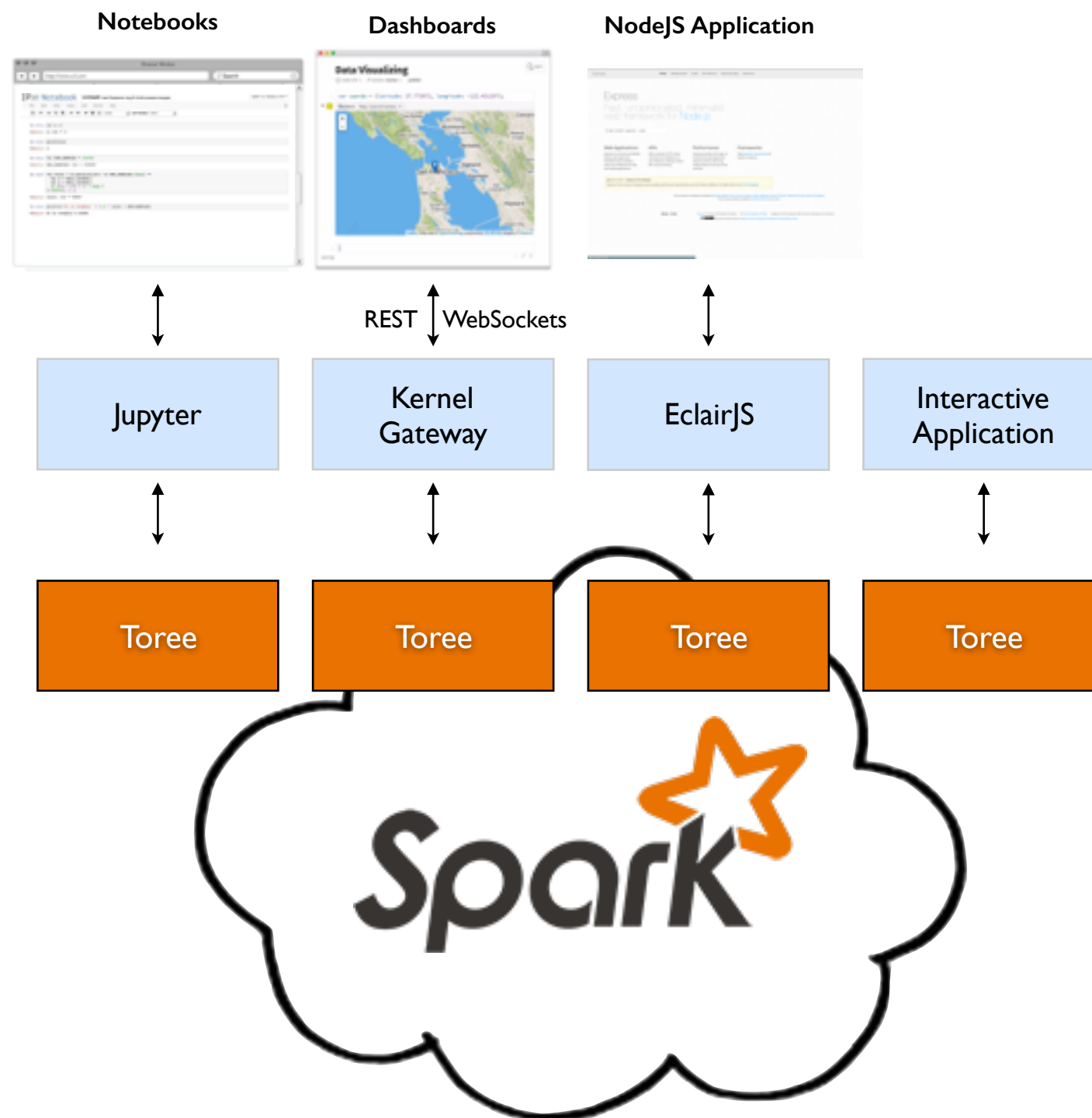
Apache Spark

Apache Spark™ is a fast and general engine for large-scale data processing.

<http://spark.apache.org/>

Apache Spark

Toree creates programmatic interfaces to
Apache Spark



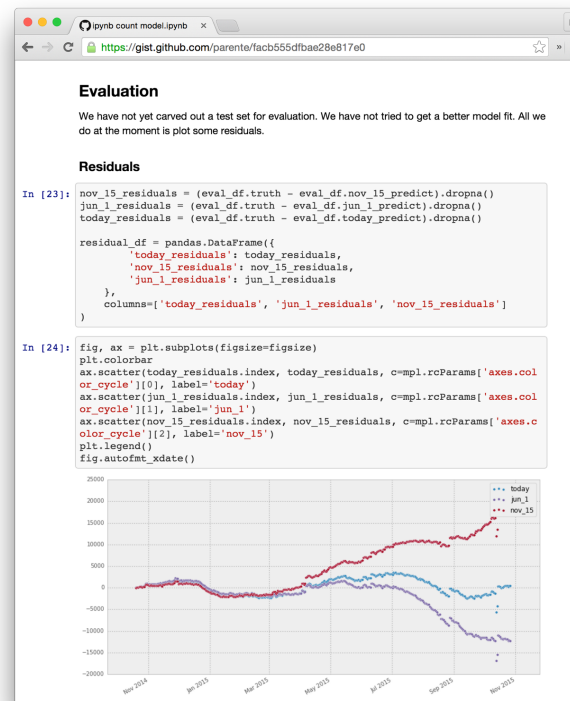
Jupyter

Open source, interactive data science and scientific computing across over 40 programming languages.

<http://jupyter.org/>

Jupyter

Toree is an implementation of the Jupyter Kernel Protocol



← Application

Code

Output

Toree →



Toree Community

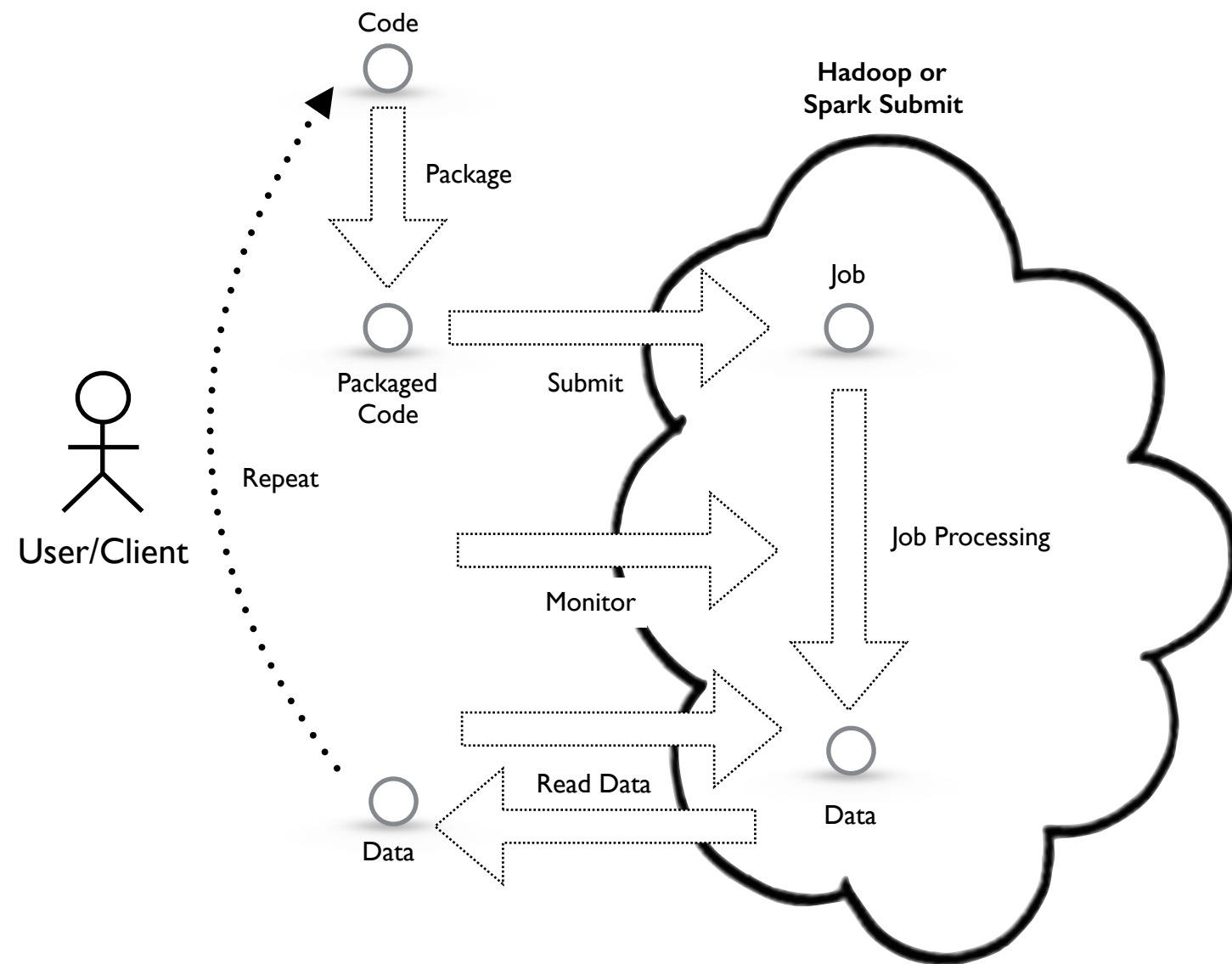
- Spark 2.0 and Scala 2.11 Support
- Hydrogen Plugin For Evaluating Spark Code
- EclairJS: JavaScript API for Spark

Ecosystem

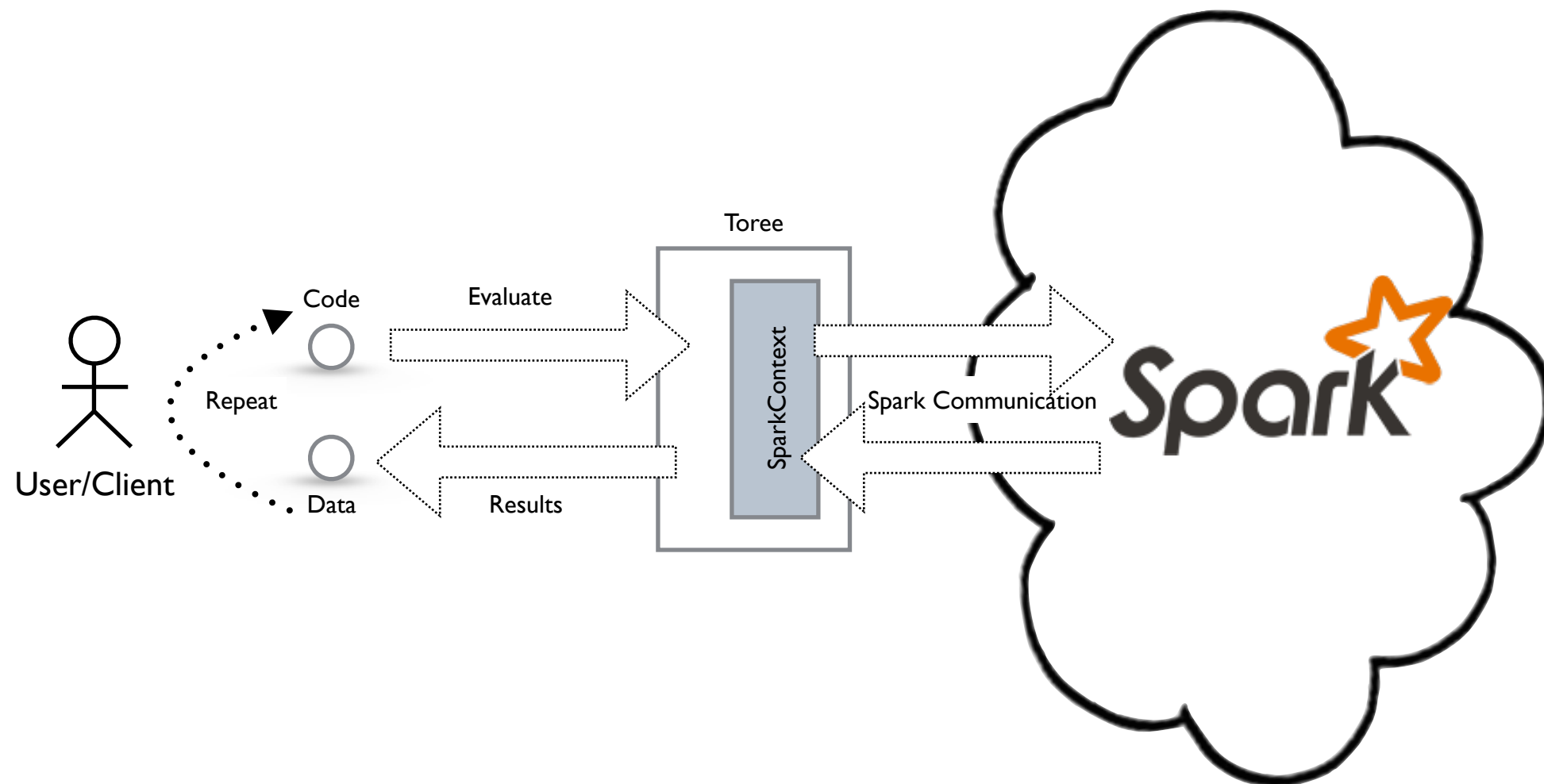


Toree Intro

Batch Applications



Interactive Applications



Application Demos

Basic Application Flow

Creating A RESTful Service

Creating A Dashboard

EclairJS Demo

Recap

Get Involved

Project site:

<http://toree.incubator.apache.org/>

Code repository:

<https://github.com/apache/incubator-toree>

Mailing List:

dev@toree.incubator.apache.org

Links

- <https://spark.apache.org/>
- <http://jupyter.org/>
- <https://github.com/jupyter-incubator/declarativewidgets>
- https://github.com/jupyter/kernel_gateway
- <https://github.com/jupyter-incubator/dashboards>
- <http://eclairjs.github.io/>



@Lull3rSkat3r

