

# ECLAIRJS = NODE.JS + APACHE SPARK

David Fallside

IBM



SPARK SUMMIT 2016  
DATA SCIENCE AND ENGINEERING AT SCALE

# Why EclairJS?

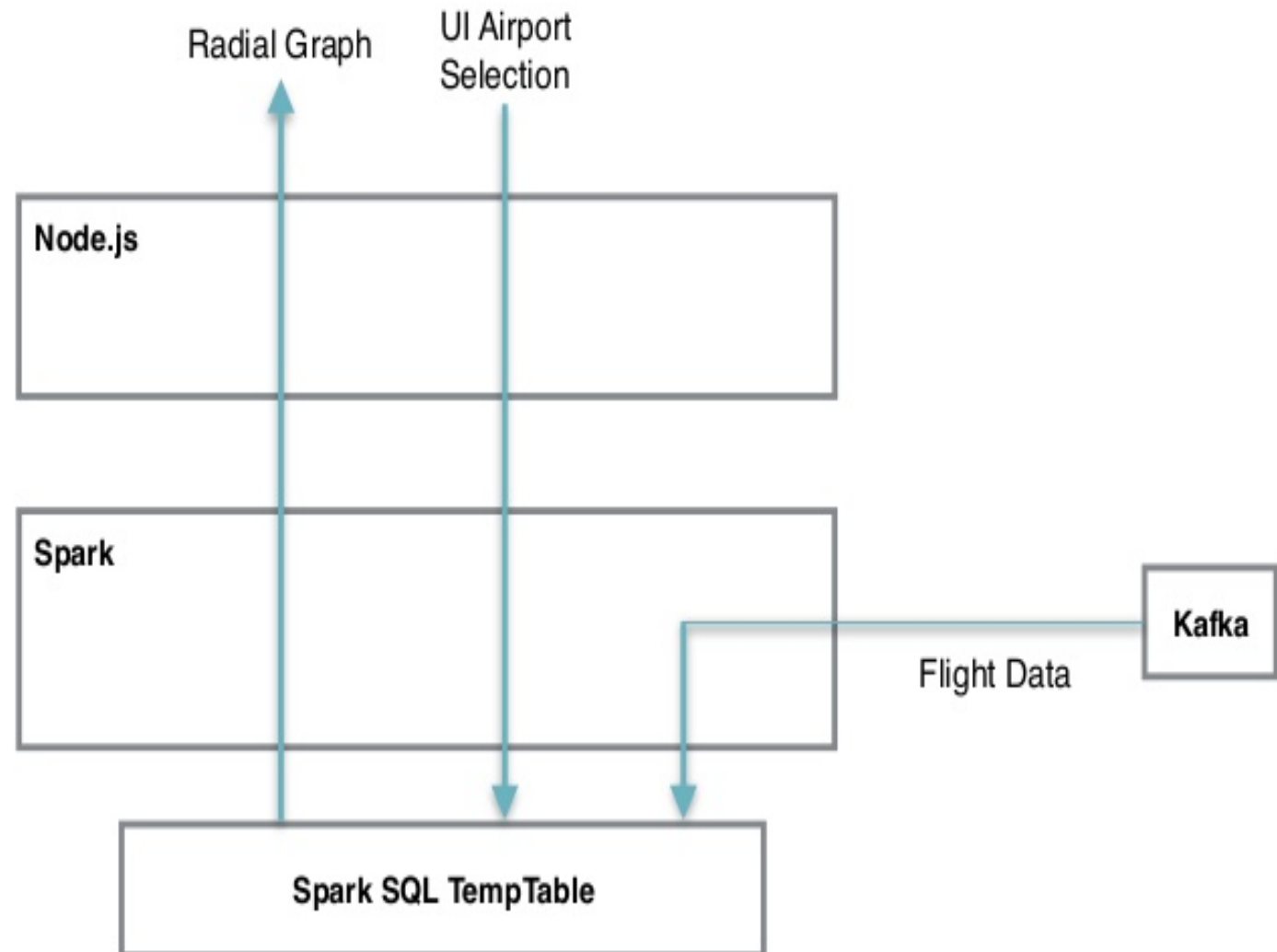
- Digital business looking to improve customer interactions, capture “perishable” insights (Forrester), use new data sources
- Today’s interactive & user-facing applications often developed in JavaScript using Node.js
  - npm provides largest (Node.js) package repo ([www.modulecounts.com](http://www.modulecounts.com))
  - Handles very large numbers of simultaneous requests
  - Compute-intensive workloads handed off to back-end engines
- Apache Spark as a back-end engine
  - Scalable, static & streaming data, Spark SQL, ML analytics, graph engine
  - But no Spark API for Node.js/JavaScript, hence EclairJS
  - So let’s look at an EclairJS application ...



demo



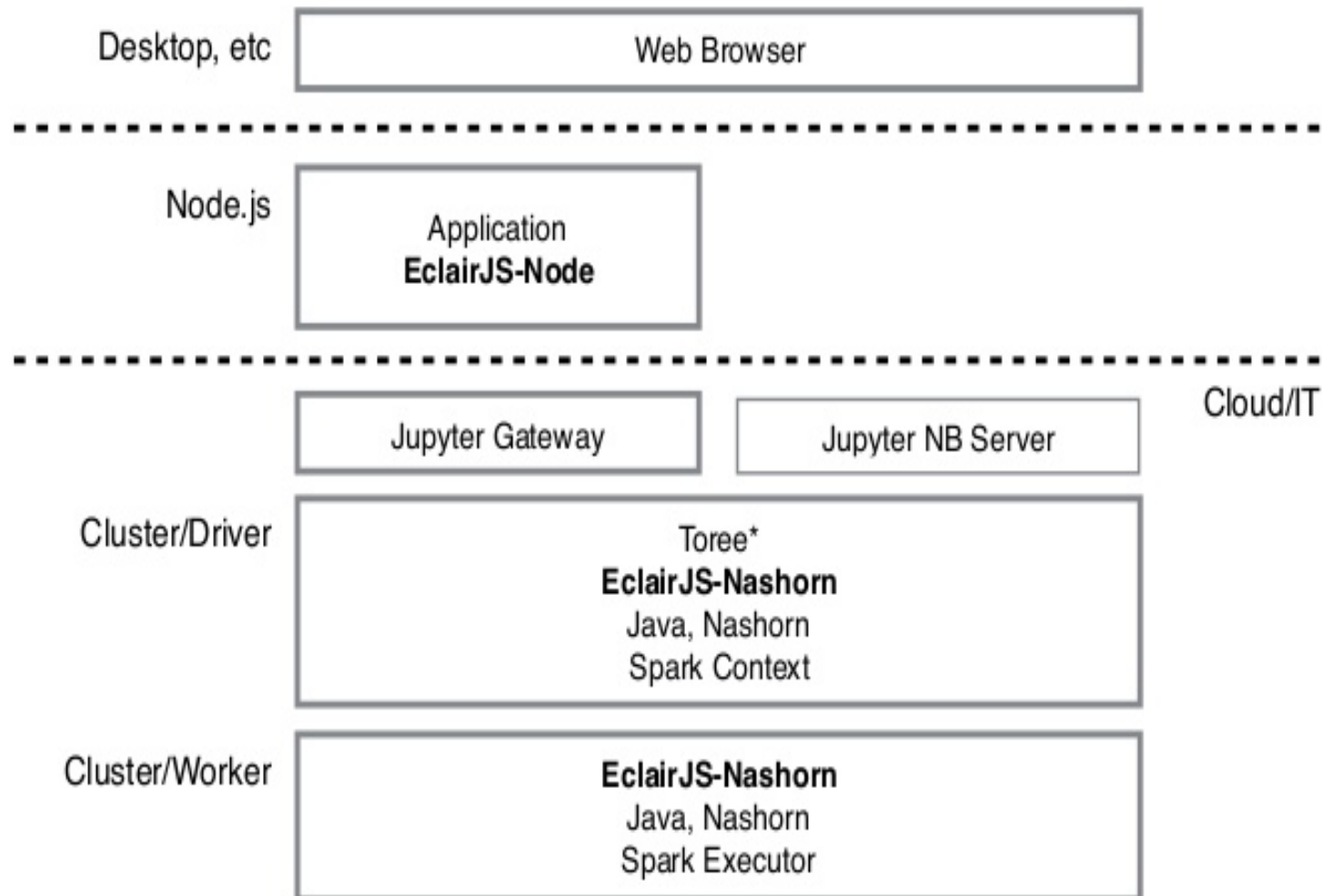
# Program Flow



# Word Count

```
var spark = require('eclairjs');  
var sc = new spark.SparkContext("local[*]", "foo");  
var file = __dirname + '/dream.txt';  
var rdd = sc.textFile(file);  
var rdd2 = rdd.flatMap(function(sentence) {  
    return sentence.split(" ");  
});  
var rdd3 = rdd2.filter(function(word) {  
    return word.trim().length > 0;  
});  
var rdd4 = rdd3.mapToPair(function(word, Tuple) {  
    return new Tuple(word.toLowerCase(), 1);  
}, [spark.Tuple]);  
var rdd5 = rdd4.reduceByKey(function(value1, value2) {  
    return value1 + value2;  
});  
var rdd6 = rdd5.mapToPair(function(tuple, Tuple) {  
    return new Tuple(tuple[1], tuple[0]);  
}, [spark.Tuple]);  
var rdd7 = rdd6.sortByKey(false);  
rdd7.take(10).then(function(val) {  
    console.log("Success:", val);  
});
```

# EclairJS Stack



\*Toree in Apache Incubator



# Notebooks

- Notebooks designed for (data) scientists, widely used for data cleaning and transformation, numerical simulation, statistical modeling, etc
  - Appear in browser as cells, may contain live code, visualizations, formatted text, widgets, etc
  - Jupyter notebooks have pluggable kernel architecture to enable different languages ([jupyter.org](https://jupyter.org))
- EclairJS provides JavaScript kernel so data engineers and web developers can try-out code and work with data in notebooks



movie recommending', and two code blocks. The first code block is titled 'Create the spark context and globals' and the second is titled 'load the small rating dataset'. Both code blocks start with 'In [1]:' and 'In [2]:' respectively." data-bbox="240 80 932 1000"/>

movie\_recommender

localhost:8888/notebooks/movie\_recommender.ipynb

Most Visited Getting Started

jupyter movie\_recommender Last Checkpoint: 2 hours ago (unsaved changes)

File Edit View Insert Cell Kernel Help Spark 1.6.0 (EclairJS)

Markdown Cell Toolbar: None

## Movie ranking example

this example is a EclairJS (JavaScript) implementation of [movie recommending](#)

### Create the spark context and globals

```
In [1]: var SparkContext = require('eclairjs/SparkContext');
var sc = new SparkContext("local[*]", "myapp");

var Tuple = require('eclairjs/Tuple');
var List = require('eclairjs/List');
var ALS = require('eclairjs/mllib/recommendation/ALS');
var Rating = require('eclairjs/mllib/recommendation/Rating');
```

### load the small rating dataset

```
In [2]: var small_ratings_raw_data = sc.textFile('../data/mllib/ml-latest-small/ratings.csv');
var small_ratings_raw_data_header = small_ratings_raw_data.take(1)[0];
var small_ratings_data = small_ratings_raw_data.filter(function(line, small_ratings_raw_data_header) {
  // filters out the header
  return line !== small_ratings_raw_data_header;
}, [small_ratings_raw_data_header])
.map(function(line) {
  return line.split(",");
})
.map(function(tokens, Rating) {
  // tokens are userId,movieId,rating,timestamp
  return new Rating(tokens[0],tokens[1],tokens[2]);
}, [Rating]).cache()
JSON.stringify(small_ratings_data.take(3));
```



# In Closing

- EclairJS for web application development in Node.js and JavaScript
  - For Data Engineers with JavaScript in Notebooks
- Project under active development in Github, [eclairjs.org](https://eclairjs.org)
  - Examples, documentation, getting-started, etc
  - EclairJS Node and EclairJS Nashorn
  - Open source, Apache v2 license
- Looking for collaborators!



# THANK YOU.

[eclairjs.org](http://eclairjs.org)

fallside at us.ibm.com



SPARK SUMMIT 2016  
DATA SCIENCE AND ENGINEERING AT SCALE