Who we are?
What is Apache Zeppelin?

A web-based notebook that enables interactive data analytics. You can make beautiful data-driven, interactive and collaborative documents with SQL, Scala and more.
When do you need Zeppelin?

When multiple tools and different peoples are involved in your data pipeline

Engineers

Data Scientists

Business user

Visualizations
Access control
Report

JDBC

Zeppelin

Apache Spark

Python

ZEPL
Notebook

Zeppelin allows use multiple interpreter (language backend) at the same time

Welcome to Zeppelin.

%md
## Welcome to Zeppelin.

Welcome to Zeppelin.

Took 3 sec. Last updated by admin at May 16 2017, 9:27:58 AM.

%spark
sc.version
println("Hello spark")

res2: String = 2.1.0
Hello spark

Took 2 sec. Last updated by admin at May 16 2017, 9:27:18 AM.

%python
print("Hello Python")

Hello Python

Took 0 sec. Last updated by admin at May 16 2017, 9:27:18 AM.

Built-in visualization instantly visualize your data

%sql
select age, count(1) value from bank where age < 30 group by age
order by age

![Bar chart showing age distribution](chart.png)

Took 1 sec. Last updated by admin at May 14 2017, 3:26:13 PM. (outdated)
Built-in Visualization

6 Basic visualizations are built in
(Table, Bar chart, Pie chart, Area chart, Scatter chart, Line chart)
Problems: Built-in Visualizations

Why isn’t it easy to add new visualizations?
- Dependent on Zeppelin release cycle
- Restricted License (e.g. commercial chart)

“What if I want to display things differently?”
Problems: Built-in Visualizations

Why isn’t it easy to add new visualizations?
- Restricted License (e.g. highchart)
- Dependent on Zeppelin release cycle

Solution?

Let’s add **pluggable** visualizations
- external add-ons
- can update frequently
- highly customizable
- shared via online registry
Apache Zeppelin: Helium
VISUALIZATION
DEMO
Helium Visualization Examples
Apache Zeppelin: Helium

SPELL
Problems: Backend Interpreter

It is not easy to add a new interpreter
- interpreters written in Java
- sometimes need to handle HTML dynamically

Can’t be combined with other interpreters
- e.g. Spark + Markdown
- Interpreter != Display System

“Can we write interpreters easily?”
Backgrounds: **Display System**

**customize interpreted outputs in** **frontend**
- can be combined with interpreters
- %html, %table, %angular, ...

```python
name = "lambda"
print("\n" + name)
print("%html <h3 style='color: #56568e'>" + name + "</h3>"

lambda

1ambda
```

1ambda

Took 0 sec. Last updated by anonymous at May 17 2017, 1:03:02 PM.
Problems: Backend Interpreter

```
%spark

println(s""""%markdown
_PV_: **${pv}**
_UV_: **${uv}**
"""")
```

“Frontend interpreter can be a display system”
Pluggable Frontend Interpreter

Easy to create and handle HTML
- written in JS
- can utilize many existing JS libraries
  (flowchart, sigmajs, vega, papaparse, ...)

(WIP) Can be a display system like %html, %table
- e.g. Spark Interpreter + Markdown Display
- allows to customize output %myGraph
Apache Zeppelin: Helium
SPELL
DEMO
Helium Spell Examples

---

**ApacheCon Demo: SPELL**

**Emojis**

Classic markup: 😊 crush: 😞 tear: 😍

Shortcuts (emoticons): 😊😊😊😊

see how to change output with twemoji.

---

**ApacheCon Demo: SPELL**

```r
%counter http://localhost:3000/api/v1/stat/dau
```

Daily Active Users

4021777
How can we share VISUALIZATION or SPELL packages?
ONLINE REGISTRIES

A community index of third-party packages for Apache Spark.

Showing packages 1 - 50 out of 340

spark-als
Another, hopefully better, implementation of ALS on Spark (already merged into MLlib)
@meir / Latest release: 0.1.0 (2014-11-27) / BSD 3-Clause / ★★★★★ (4.3)

mlib-grid-search
An example project for doing grid search in MLlib
@spark-mt / Latest release: 0.0.1 (2016-11-27) / BSD 3-Clause / ★★★★★ (4.3)

spark-avro
Integration utilities for using Spark with Apache Avro data
@datadiscrete / Latest release: 3.2.0-k.2.11 (2017-02-08) / Apache-2.0 / ★★★★★★★ (9.0)

spark-redshift
Redshift Data Source for Apache Spark
@datadiscrete / Latest release: 3.0.0 preview (2016-11-01) / Apache-2.0 / ★★★★★★ (7.0)
ONLINE REGISTRIES

npm is the package manager for gulp

481,746 total packages
391,763,488 downloads in the last day
2,223,968,876 downloads in the last week
8,945,897,538 downloads in the last month

packages people ‘npm install’ a lot

browserify
Grunt package manager

express
Fast, unopinionated, minimalist web framework

npm
A package manager for JavaScript

gulp
Caching build system

cordova
Cordova command line interface tool

forever
A simple CLI tool for ensuring that a given process stays running

less
Lesser CSS

getting started

installing npm

npm command-line tool is bundled with Node.js. If you have installed Node.js already, you already have npm installed.

screencasts and docs

We’ve got a new docs site featuring videos and tutorials to help you make your JavaScript dreams come true.

get hired

BEWAREOFBIZARREPEOPLE, micropasts, PaperG, and lots of other companies are hiring JavaScript
Online Registry for Helium packages?
BUT ONE ISSUE IS

*Apache Zeppelin* is an *Open Source* project
BUT ONE ISSUE IS

*Apache Zeppelin* is an *Open Source* project
SO, WE SHOULD CONSIDER

- Who can build up the *infra* and operate the *service*?
- **Apache Zeppelin** and the logo are *trademarks* of the ASF
SO, WE SHOULD CONSIDER

- Who can build up the *infra* and operate the *service*?
- Apache Zeppelin and the logo are *trademarks* of the ASF

So, the registry should be completely separated from

**Apache® Zeppelin™**
Too complicated..

Is there any other SIMPLE way to solve this?
A SOLUTION ALREADY EXISTED
A SOLUTION ALREADY EXISTED

Helium VISUALIZATION & SPELL packages are package
A SOLUTION ALREADY EXISTED

Helium VISUALIZATION & SPELL packages are

The package information can be saved in npm registry!
Then, is there any way to fetch **ONLY** Helium package information from npm registry? 🤔
HOW?

Create a Helium package
HOW?

Create a Helium package → Publish to npm Registry

http://registry.npmjs.org/
HOW?

Create a Helium package → Publish to Registry → Filter Helium pkg & fetch only necessary metadata
HOW?

Create a Helium package

Publish to Registry

Filter Helium pkg & fetch only necessary metadata

Does the package has **zeppelin-vis** or **zeppelin-spell** as its dependency?
HOW?

Create a Helium package

Publish to Registry

Filter Helium pkg & fetch only necessary metadata

If so, take only necessary metadata:

- name
- description
- version
- license
- ...
HOW?

Create a Helium package

Publish to Registry

Filter Helium pkg & fetch only necessary metadata

Integrate whole data and create *helium.json*
HOW?

Create a Helium package

Publish to Registry

Filter Helium pkg & fetch only necessary metadata

Integrate whole data and create helium.json

Save the file in ZEPL
HOW?

1. Create a Helium package
2. Publish to Registry
3. Trigger Lambda function every 1 hour using CloudWatch
4. Filter Helium pkg & fetch only necessary metadata
5. Integrate whole data and create `helium.json`
6. Save the file in S3
HOW?

Create a Helium package

Publish to Registry

Save the file in helium.json

Read helium.json

Filter Helium pkg & fetch only necessary metadata

Integrate whole data and create helium.json

Trigger Lambda function *every 1 hour* using

Create a Helium package

Filter Helium pkg & fetch only necessary metadata

Integrate whole data and create helium.json

Trigger Lambda function *every 1 hour* using

Create a Helium package
Helium Packages

From Zeppelin-0.7, you can load/unload a pluggable Apache Zeppelin package on runtime through Helium framework in Zeppelin. Here are the lists of Helium packages registered in the NPM Registry and Maven Central Repository. If you need more information about how you can use the below packages in Zeppelin, see What is Apache Zeppelin Visualization and What is Apache Zeppelin Spell.

List by

- Recently published
- Type: VISUALIZATION
- Type: SPELL (only available in development version 0.8.0-SNAPSHOT)
- Type: INTERPRETER

14 Package(s) Registered

- **volume-leaflet** VISUALIZATION
  by Tom Grant
  Geospatial visualization using the Leaflet map library.
  v 1.0.1

- **ultimate-heatmap-chart** VISUALIZATION
  by ZEPL
  The Ultimate Heatmap Chart for Apache Zeppelin using <a href="http://www.highcharts.com/">highcharts</a>
  v 0.0.1

- **ultimate-range-chart** VISUALIZATION
  by ZEPL
  The Ultimate Range Chart for Apache Zeppelin using <a href="http://www.highcharts.com/">highcharts</a>
Helium Packages

From Zeppelin-0.7, you can load/unload Helium packages. Here are the lists of Helium.

What is Apache Zeppelin Spell.

LIST BY
- Recently published
- Type: VISUALIZATION
- Type: SPELL only available in development
- Type: INTERPRETER

14 PACKAGE(S) REGISTERED

- **volume-leaflet** VISUALIZATION
  - volume-leaflet@1.0.1
  - Geospatial visualization using the Leaflet map library.

- **ultimate-heatmap-chart** VISUALIZATION
  - ultimate-heatmap-chart@0.0.1
  - The Ultimate Heatmap Chart for Apache Zeppelin using highcharts

- **ultimate-range-chart** VISUALIZATION
  - ultimate-range-chart@0.0.1
  - The Ultimate Range Chart for Apache Zeppelin using highcharts

- **ultimate-scatter-chart** VISUALIZATION
  - ultimate-scatter-chart@0.0.1
  - The Ultimate Scatter Chart for Apache Zeppelin using highcharts

- **ultimate-column-chart** VISUALIZATION
  - ultimate-column-chart@0.0.1
  - The Ultimate Column Chart for Apache Zeppelin using highcharts

- **ultimate-pie-chart** VISUALIZATION
  - ultimate-pie-chart@0.0.1
  - The Ultimate Pie Chart for Apache Zeppelin using highcharts

- **ultimate-area-chart** VISUALIZATION
  - ultimate-area-chart@0.0.1
  - The Ultimate Area Chart for Apache Zeppelin using highcharts
JIRA ISSUES

- **ZEPPELIN-1973**: List all available Helium packages in Zeppelin website
- **ZEPPELIN-2004**: List helium packages in Zeppelin GUI by reading file
Helium extends Zeppelin eco-system

Visualizations
- Bubble
- Map
- Heatmap
- Range
- Spline

Spell
- Sigma
- D3
- Flowchart
- Markdown
- Translator

Interpreters
- Spark
- Python
- Geode
- Cassandra
- Kylin
- Groovy
- JDBC
- Flink

Users / Developers
- 2833
- 211 contributors

3rd parties
- Amazon Web Services
- Microsoft Azure
- Oracle Data Cloud
- Quabele
- Mesosphere
- ZEPL
Useful service

Zeppelin notebook online viewer

https://www.zeppelinhub.com/viewer
Future Roadmap

0.7.2
- Maintenance release.

0.8.0
- Helium online registry.
- Interpreter Cluster mode.

1.0
- Finest, the most stable release

2Q /2017

3Q /2017

Hopefully

3-4Q /2017