



Apache Zeppelin

Helium makes Zeppelin Fly

Moon soo Lee (moon@zepl.com)

Hoon Park (1ambda@zepl.com)

Ahyoung Ryu (ahyoungryu@zepl.com)

@ZEPL

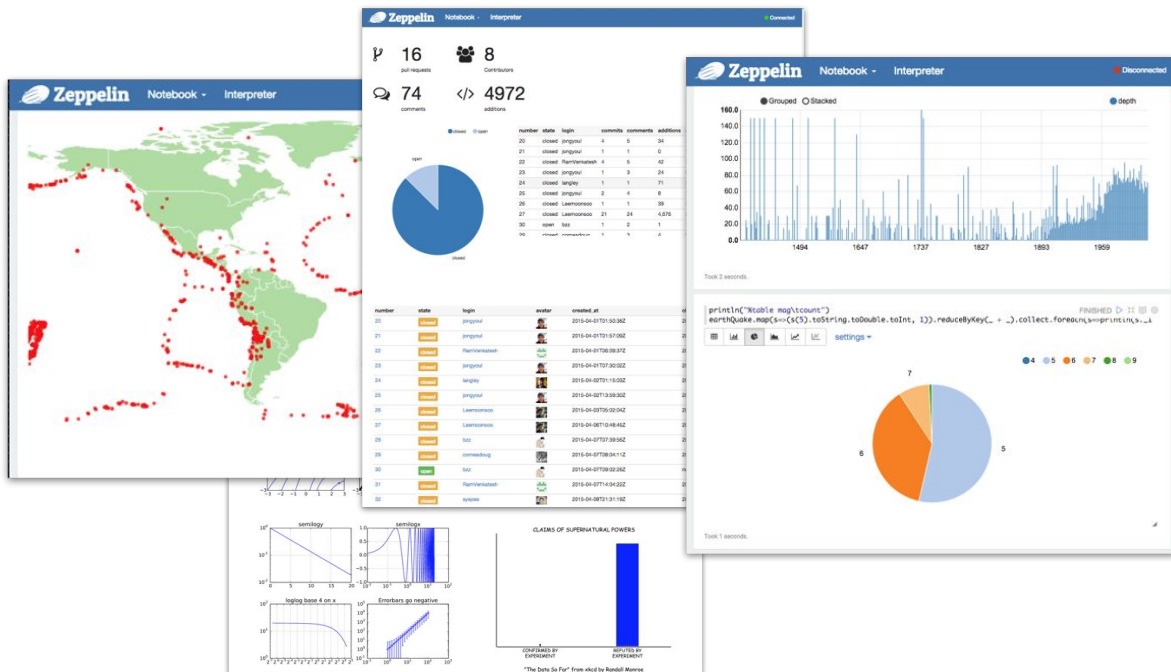


Who we are?



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.



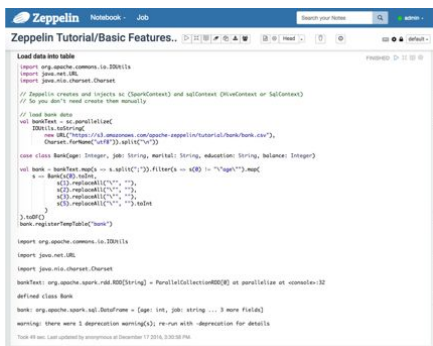
When do you need Zeppelin?

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

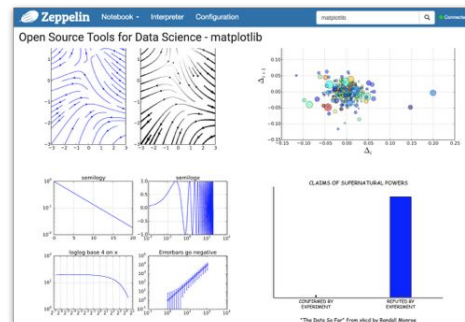
Engineers

Data Scientists

Business user



```
Zeppelin Tutorial/Basic Features...  
Load data into table  
import org.apache.commons.io.IOUtils  
import java.net.URL  
import java.nio.charset.Charset  
  
// Zeppelin creates and injects its Context (ExecutionContext or ExecutionContext)  
// So you don't need create them manually  
  
// Load bank data  
val bankText = io.github.danielmiller.  
  ZeppelinTutorial.  
  readURL("https://raw.githubusercontent.com/danielmiller/zeppelin-tutorial/master/src/main/resources/bank.txt").  
  toString(Charset.forName("UTF-8")).split("\n")  
  
case class BankEntry(idemp: Int, job: String, work: Int, Service: String, education: String, balance: Integer)  
  
val bank = bankText.map(_.split(",")).filter(_._1 != "0").map {  
  s => BankEntry(s._1.toInt, s._2, s._3.toInt, s._4, s._5, s._6.toInt)  
}.toSeq  
  
bank.registerTable("bank")  
  
import org.apache.commons.io.IOUtils  
import java.net.URL  
import java.nio.charset.Charset  
  
bankText = io.github.danielmiller.  
  ZeppelinTutorial.  
  readURL("https://raw.githubusercontent.com/danielmiller/zeppelin-tutorial/master/src/main/resources/bank.txt").  
  toString(Charset.forName("UTF-8")).split("\n")  
  
defined class BankEntry  
bank = org.apache.spark.sql.DataFrame = [idemp: int, job: string ... 3 more fields]  
warning: there were 3 deprecation warnings (3), re-run with -deprecation for details  
Total 48 lines. Last updated by anonymous at December 17 2016, 3:00:00 PM
```



JDBC

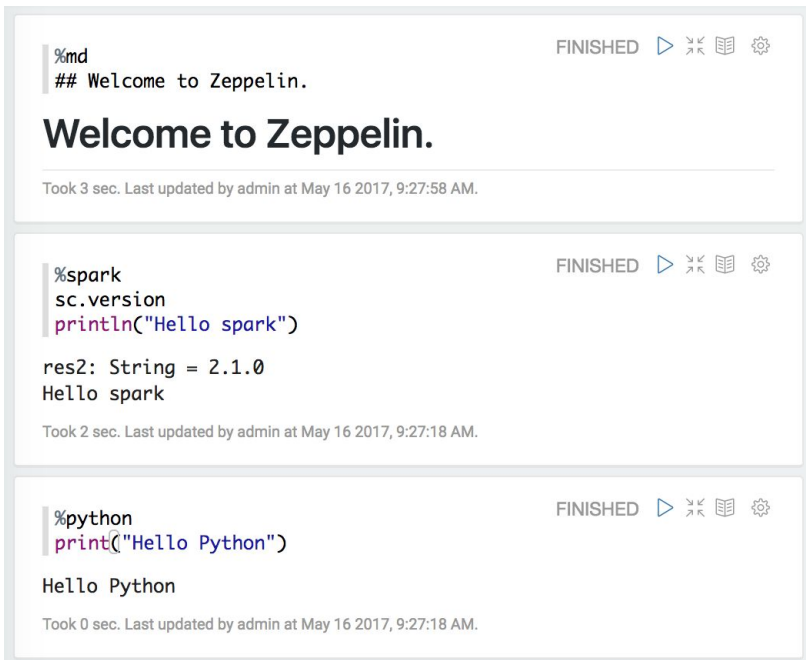


Visualizations
Access control
Report



Notebook

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



The screenshot displays a Zeppelin notebook interface with three code blocks. Each block has a status bar at the top right indicating 'FINISHED' and icons for execution, undo, redo, and settings.

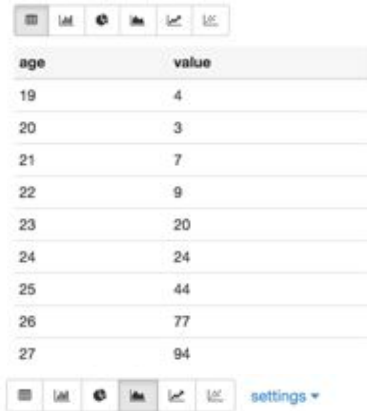
- Block 1 (Markdown):** Starts with `%md` and contains the text `## Welcome to Zeppelin.`. The output is **Welcome to Zeppelin.** in a large, bold font. The status bar shows 'Took 3 sec. Last updated by admin at May 16 2017, 9:27:58 AM.'
- Block 2 (Spark):** Starts with `%spark` and contains `sc.version` and `println("Hello spark")`. The output shows `res2: String = 2.1.0` and `Hello spark`. The status bar shows 'Took 2 sec. Last updated by admin at May 16 2017, 9:27:18 AM.'
- Block 3 (Python):** Starts with `%python` and contains `print("Hello Python")`. The output is `Hello Python`. The status bar shows 'Took 0 sec. Last updated by admin at May 16 2017, 9:27:18 AM.'

Zeppelin supports multi-user environment

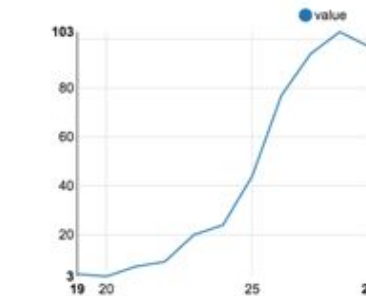
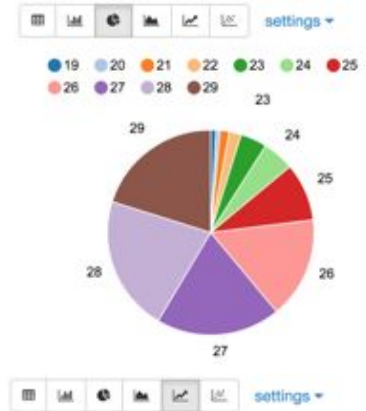
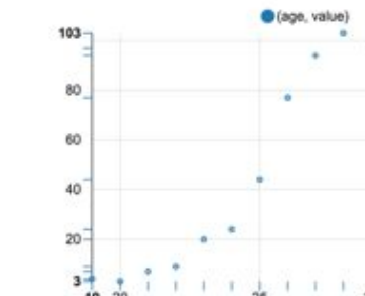
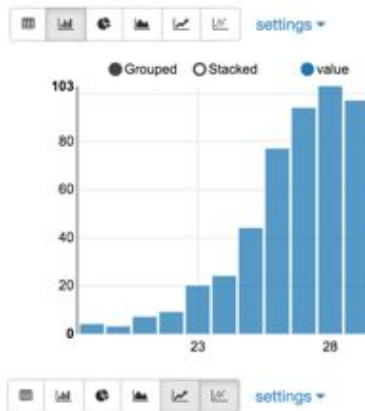
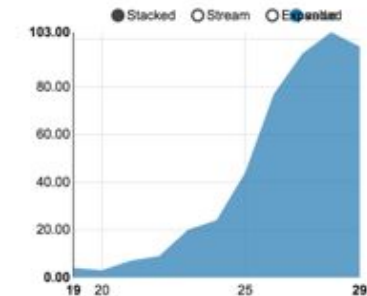
- Real-time collaboration
- Authentication
- Notebook ACL
- Interpreter ACL
- Fine grained interpreter session - notebook/user mapping
- Built-in scheduler
- Pluggable notebook storage
- Pluggable interpreter

Built-in Visualization

6 Basic visualizations are built-in, (in addition to matplotlib, ggplot integration)
(Table, Bar chart, Pie chart, Area chart, Scatter chart, Line chart)



age	value
19	4
20	3
21	7
22	9
23	20
24	24
25	44
26	77
27	94





Hoon Park @ ZEPL

[linkedin.com/in/1ambda](https://www.linkedin.com/in/1ambda)



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?  lium

Let's add **pluggable** visualizations

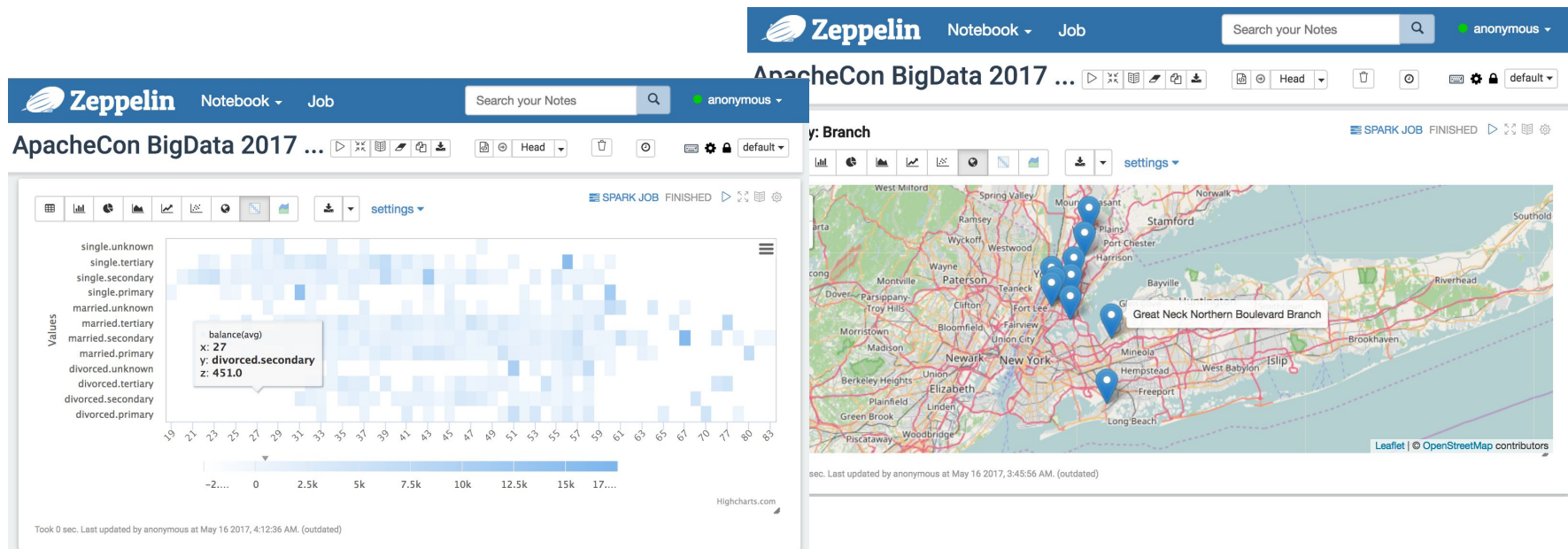
- external add-ons
- can update frequently
- highly customizable
- shared via online registry

Apache Zeppelin: Helium

VISUALIZATION

DEMO

Helium Visualization Examples



“What about extending Interpreters?”

“Can we leverage Helium framework for interpreters?”

Backgrounds: Interpreter

Execute paragraph (code) and return output

- separated JVM process
- `%spark`, `%jdbc`, `%python`, ...

```
%spark
```

```
import org.apache.spark.sql.SQLContext
```

```
val sqlContext = new SQLContext(sc)
```

```
def readCsvAndRegisterTable(url: String, table: String, sep: String = ",") {  
  val df = sqlContext.read  
    .format("com.databricks.spark.csv")
```

FINISHED ▶ ⌵ 📖 ⚙️

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 = "1lambda"
print("\n" + name)
print("%html <h3 style='color: #56568e'>" + name + "</h3>")
```

1lambda

1lambda

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

Problems: Backend Interpreter

```
%spark
```

READY ▶ ⌕ 📖 ⚙️

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

“Frontend interpreter can be a display system”

SPELL: Frontend Interpreter

Easy to create and handle HTML

- written in JS
- can utilize many existing JS libraries
([flowchart](#), [sigma.js](#), [vega](#), [papaparse](#), ...)

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

Zeppelin Notebook Job Search your Notes anonymous

ApacheCon Demo: SPELL

Finished

Emojies

Classic markup: 😊:crush: 😢:tear: 😊😢

Shortcuts (emoticons): 😊😢😎😢

see [how to change output](#) with twemoji.

Zeppelin Notebook Job Search your Notes anonymous

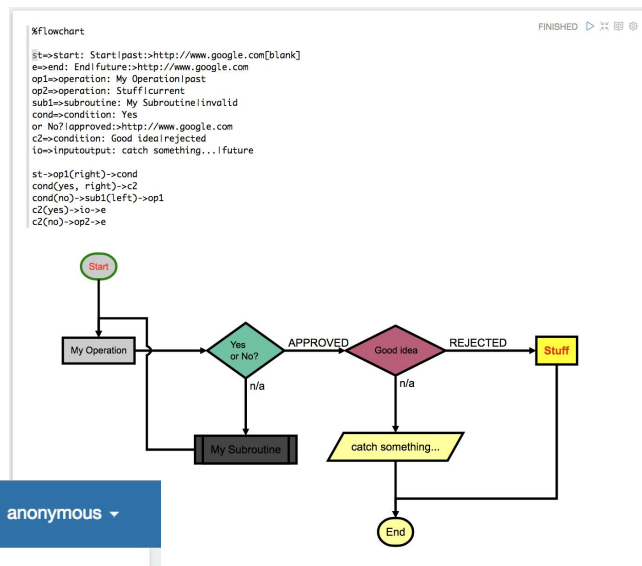
ApacheCon Demo: SPELL

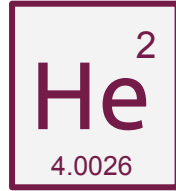
Finished

%counter http://localhost:3000/api/v1/stat/dau Daily Active Users

Daily Active Users

4021777






Helium Online Registry



Ahyoung Ryu @ ZEPL
[linkedin.com/in/AhyoungRyu](https://www.linkedin.com/in/AhyoungRyu)

How can we share
VISUALIZATION *or*
SPELL
packages?

ONLINE REGISTRIES

SparkPackages[Feedback](#)[Register a package](#)[Login](#)[Find a package](#)

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

Showing packages 1 - 50 out of 340 [Next >](#)

[All \(340\)](#) [Core \(13\)](#) [Data Sources \(45\)](#) [Machine Learning \(71\)](#) [Streaming \(49\)](#) [Graph \(17\)](#) [PySpark \(10\)](#) [Applications \(11\)](#) [Deployment \(12\)](#) [Examples \(24\)](#) [Tools \(28\)](#)

spark-als

Another, hopefully better, implementation of ALS on Spark (already merged into MLlib)

[@mengxr](#) / Latest release: 0.1.0 (2014-11-27) / BSD 3-Clause / ★★★★★ (1)

[2 ml](#) [1 mllib](#) [1 recommendation](#)

mllib-grid-search

An example project for doing grid search in MLlib

[@spark-ml](#) / Latest release: 0.0.1 (2014-11-27) / BSD 3-Clause / ★★★★★ (1)

[1 ml](#) [1 example](#) [1 examples](#)

spark-avro

Integration utilities for using Spark with Apache Avro data

[@databricks](#) / Latest release: 3.2.0-s_2.11 (2017-02-08) / Apache-2.0 / ★★★★★ (11)

[6 sql](#) [4 input](#) [4 avro](#)

spark-redshift

Redshift Data Source for Apache Spark

[@databricks](#) / Latest release: 3.0.0-preview1 (2016-11-01) / Apache-2.0 / ★★★★★ (3)

Spark Packages is a community site hosting modules that are not part of Apache Spark. Your use of and access to this site is subject to the terms of use. Apache Spark and the Spark logo are trademarks of the Apache Software Foundation. This site is maintained as a community service by Databricks.

ONLINE REGISTRIES

SparkPackages[Feedback](#)[Register a package](#)[Login](#)[Find a package](#)

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

Showing packages 1 - 50 out of 340

[All \(340\)](#) [Core \(13\)](#) [Data Sources \(45\)](#) [Machine Learning \(71\)](#) [Streaming \(49\)](#) [Graph \(17\)](#) [PySpark \(10\)](#) [Applications \(11\)](#) [Deployment \(12\)](#) [Examples \(24\)](#) [Tools \(10\)](#)

spark-als

Another, hopefully better, implementation of ALS on Spark (already merged into MLlib)

@mengxr / Latest release: 0.1.0 (2014-11-27) / BSD 3-Clause / ★★★★★ (1)

2 ml 1 millib 1 recommendation

mllib-grid-search

An example project for doing grid search in MLlib

@spark-ml / Latest release: 0.0.1 (2014-11-27) / BSD 3-Clause / ★★★★★ (1)

1 ml 1 example 1 examples

spark-avro

Integration utilities for using Spark with Apache Avro data

@databricks / Latest release: 3.2.0-s_2.11 (2017-02-08) / Apache-2.0 / ★★★★★ (11)


6 sql 4 input 4 avro

spark-redshift

Redshift Data Source for Apache Spark

@databricks / Latest release: 3.0.0-preview1 (2016-11-01) / Apache-2.0 / ★★★★★ (13)

Spark Packages is a community site hosting modules that are not part of Apache Spark. Your use of and access to this site is subject to the terms of use. Apache Spark and the Spark logo are trademarks of the Apache Software Foundation. This site is maintained as a community service by Databricks.

 **PLUGINS** [IDEs](#) [TEAMCITY](#) [DEV GUIDE](#) [UPLOAD PLUGIN](#) [SIGN IN](#)

IntelliJ IDEA Plugins

Search 2234 plugins

New or Updated

WordPress Support

In Framework integration

WordPress Support.

16.05.2017 29 780

Performance Testing

In Misc

Plugin for automated execution of test scripts, capturing performance snapshots and gathering performance statistics.

16.05.2017 11 780

Drupal Support

In Framework integration

Drupal Support.

16.05.2017 17 114

Command Line Tool Support

In Tools integration

Supports command execution.

16.05.2017 152 587 4.5

Google App Engine Support for PHP

In Framework integration

Google App Engine Support for PHP.

16.05.2017 23 156 5

Upsource Integration

In Tools integration

Integration with JetBrains Upsource.

16.05.2017 115 347 5

IdeaVimExtension

In Editor

为IdeaVim插件增加自动切换为英文输入法的功能。

16.05.2017 195

Stepik Union

In Code tools

Stepik plugin for code challenges on Python, Java, asm, C, C++, Clojure, Haskell, JavaScript, Mono C#, Octave, R, Ruby, Rust, Shell Script.



ONLINE REGISTRIES

SparkPackages

Feedback Register a package Login Find a package

A community index of third-party Spark packages

All (340) Core (13) Data Sources (45) Machine Learning (71) Streaming (49)

spark-als

Another, hopefully better, implementation of ALS on Spark (already in [Spark](#))

@mengxr / Latest release: 0.1.0 (2014-11-27) / BSD 3-Clause / ★★★★★ (1)

2 ml 1 millib 1 recommendation

mllib-grid-search

An example project for doing grid search in MLlib

@spark-ml / Latest release: 0.0.1 (2014-11-27) / BSD 3-Clause / ★★★★★ (1)

1 ml 1 example 1 examples

spark-avro

Integration utilities for using Spark with Apache Avro data

@databricks / Latest release: 3.2.0-s_2.11 (2017-02-08) / Apache-2.0 / ★★★★★

6 sql 4 input 4 avro

spark-redshift

Redshift Data Source for Apache Spark

@databricks / Latest release: 3.0.0-preview1 (2016-11-01) / Apache-2.0 / ★★★★★

Spark Packages is a community site hosting modules that are not part of the official Spark distribution. Apache Spark and the Spark logo are trademarks of the Apache Software Foundation.

IntelliJ IDEA Plugins

IDEs TEAMCITY DEV GUIDE

UPLOAD PLUGIN SIGN IN

npm is the package manager for gulp

481,746 total packages 391,763,408 downloads in the last day 2,223,946,876 downloads in the last week 8,945,897,538 downloads in the last month

packages people 'npm install' a lot

browserify browser-side require() the node way 14.3.0 published a month ago by feross	express Fast, unopinionated, minimalist web framework. 4.15.2 published 2 months ago by dougmlton	pm2 Production process manager for Node.js applications. 2.4.6 published 3 weeks ago by thekai
grunt-cli The grunt command line interface 1.2.0 published a year ago by vladikoff	npm a package manager for JavaScript 4.6.1 published 4 weeks ago by iarna	karma Spectacular Test Runner for JavaScript. 1.7.0 published a week ago by dignifiedquire
bower The browser package manager 1.8.0 published 6 months ago by sheerun	cordova Cordova command line interface tool 7.0.1 published 4 days ago by stevegil	coffee-script Unfancy JavaScript 1.12.6 published yesterday by geoffreybooth
gulp The streaming build system 3.9.1 published a year ago by phated	forever A simple CLI tool for ensuring that a given... 0.15.3 published 6 months ago by indexzero	statsd Network daemon for the collection and aggregation of metrics. 0.8.0 published a year ago by pkhzrd
grunt The JavaScript Task Runner 1.0.1 published a year ago by shama	less Leamer CSS 2.7.2 published 4 months ago by matthew-de...	yo CLI tool for running Yeoman generators 1.8.5 published 8 months ago by sboudrias

getting started

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

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
WEAREAREASONABLEPEOPLE, microapps, PaperG and lots of other companies are hiring JavaScript developers.

Testing

Execution of test performance snapshots and performance statistics.

Drupal Support

In Framework integration

Drupal Support.
16.05.2017 17 114

Engine PHP

Support for PHP.
16.05.2017 115 347 5

Upsource Integration

In Tools integration

Integration with JetBrains Upsource.

16.05.2017 115 347 5

challenges on Python, Clojure, Haskell, R, Octave, R, Ruby, Rust,

Online Registry for Helium packages?

THEN WE SHOULD CONSIDER

1. **Who/ How** can we build up the *infra* and operate the **service**?

- Need to set up the authentication system
- Need user/ package DB
- Versioning/ building / packaging
- ...

2. External library **Licenses**

Too complicated..

Is there any other **SIMPLE** way to solve this?

A SOLUTION ALREADY EXISTED

A SOLUTION ALREADY EXISTED

Helium **VISUALIZATION** & **SPELL** packages are




A SOLUTION ALREADY EXISTED

Helium **VISUALIZATION** & **SPELL** packages are



The package information can be saved in **npm** registry!

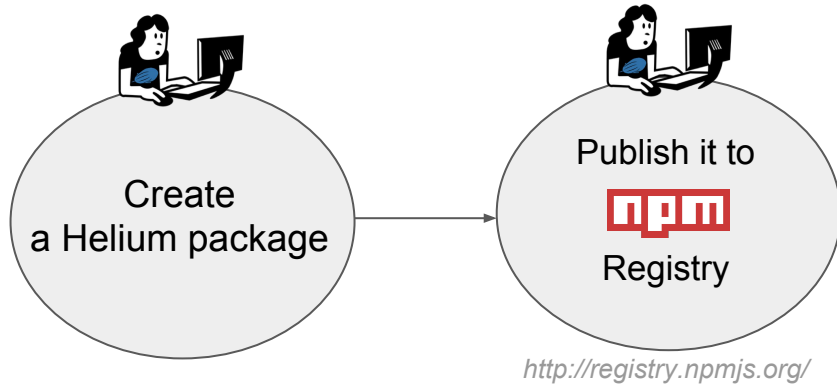
Then, is there any way to fetch
ONLY Helium package information
from  registry? 🤔

HOW?

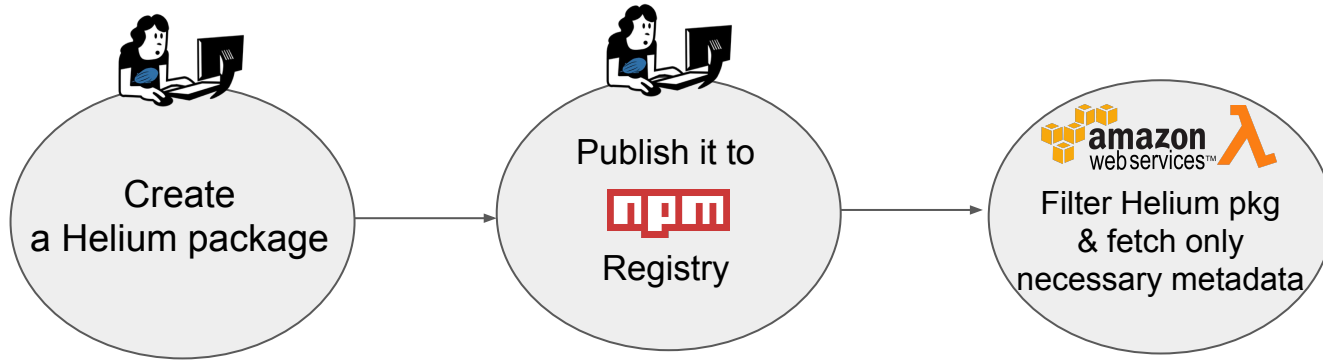


Create
a Helium package

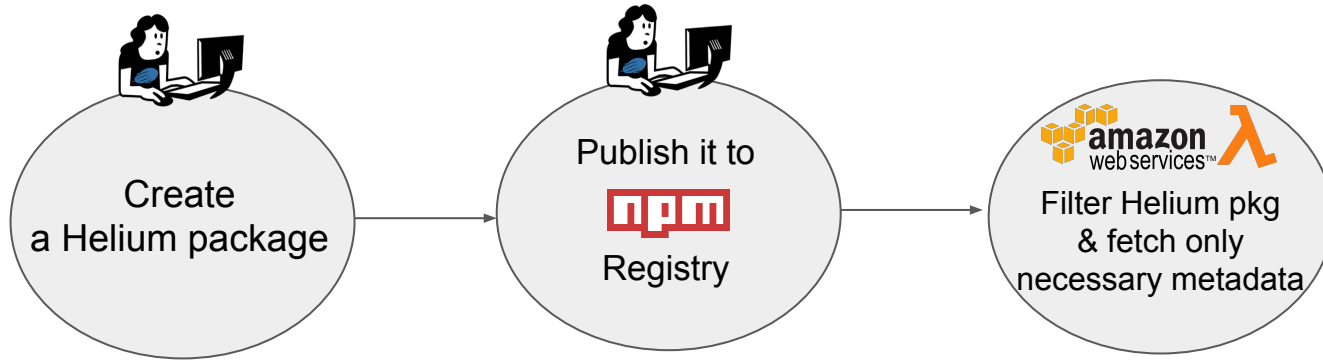
HOW?



HOW?

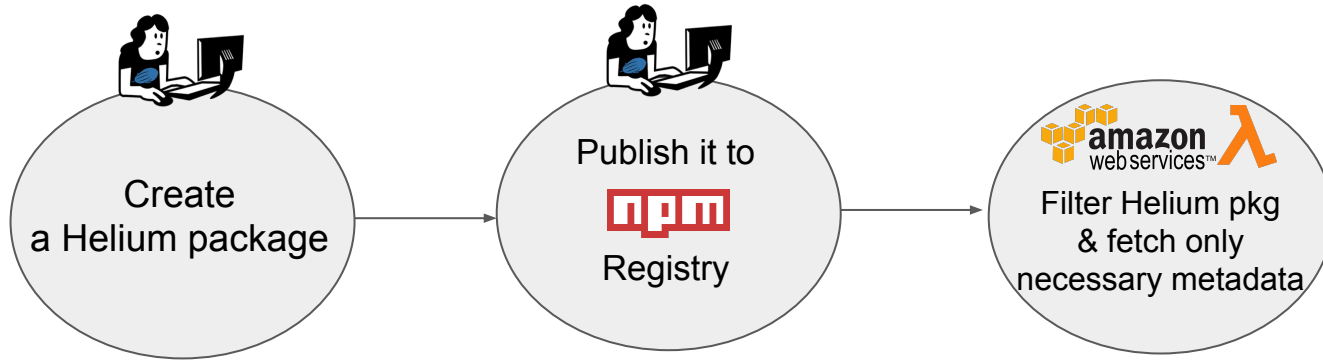


HOW?



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

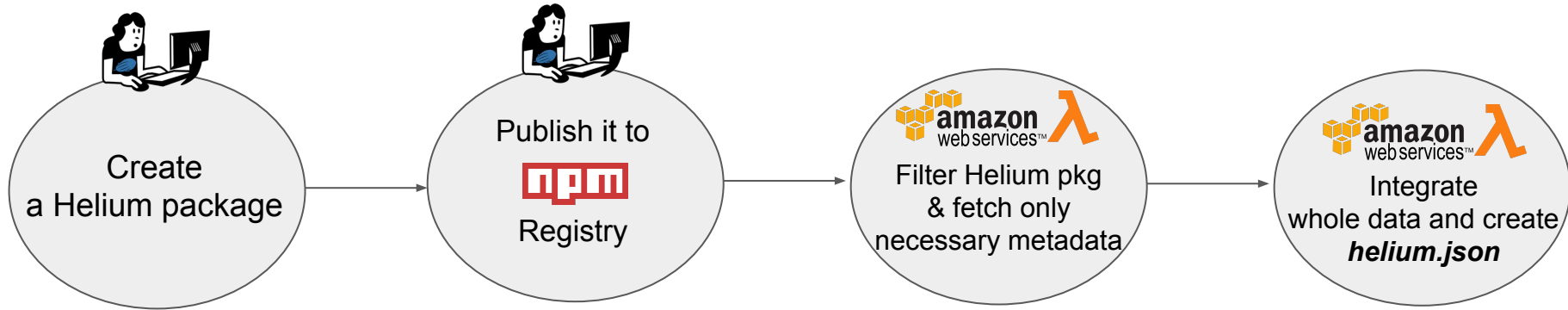
HOW?



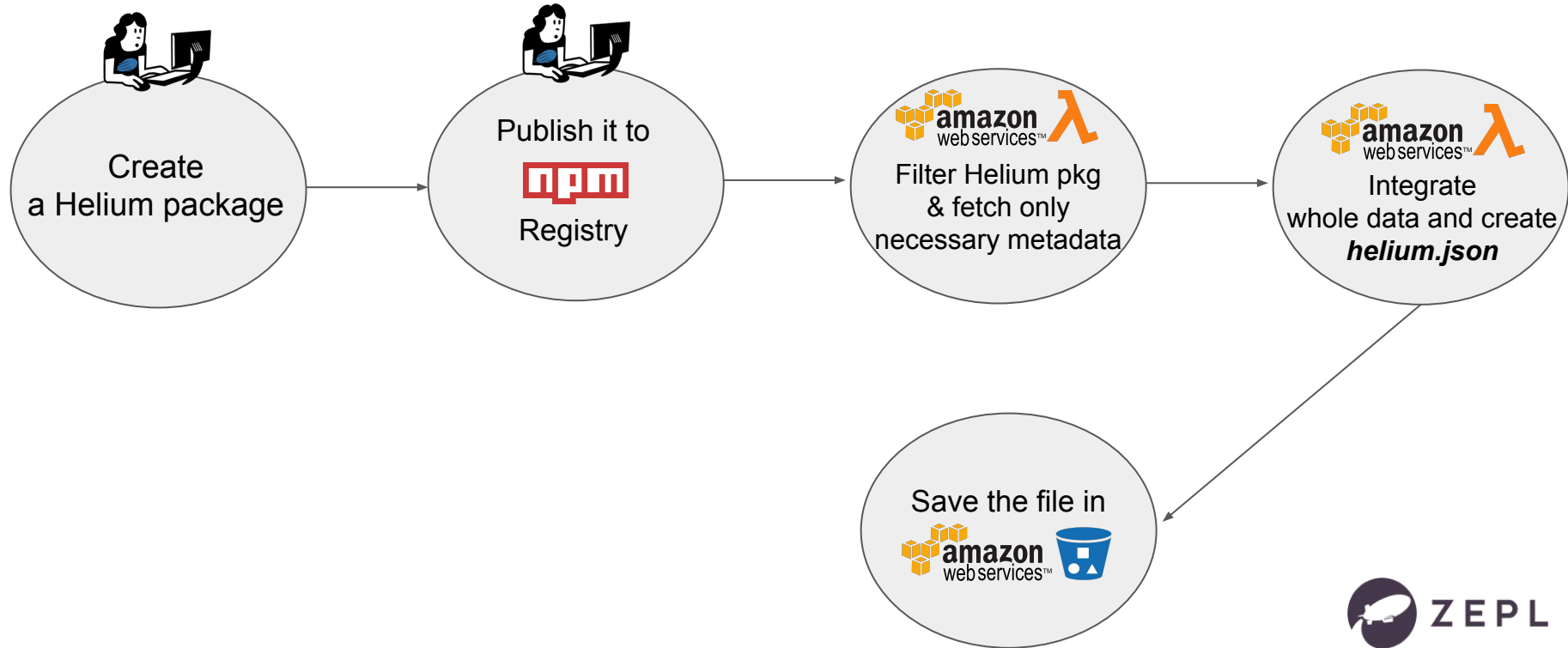
If so, take only necessary metadata:

- name
- description
- version
- license
- ...

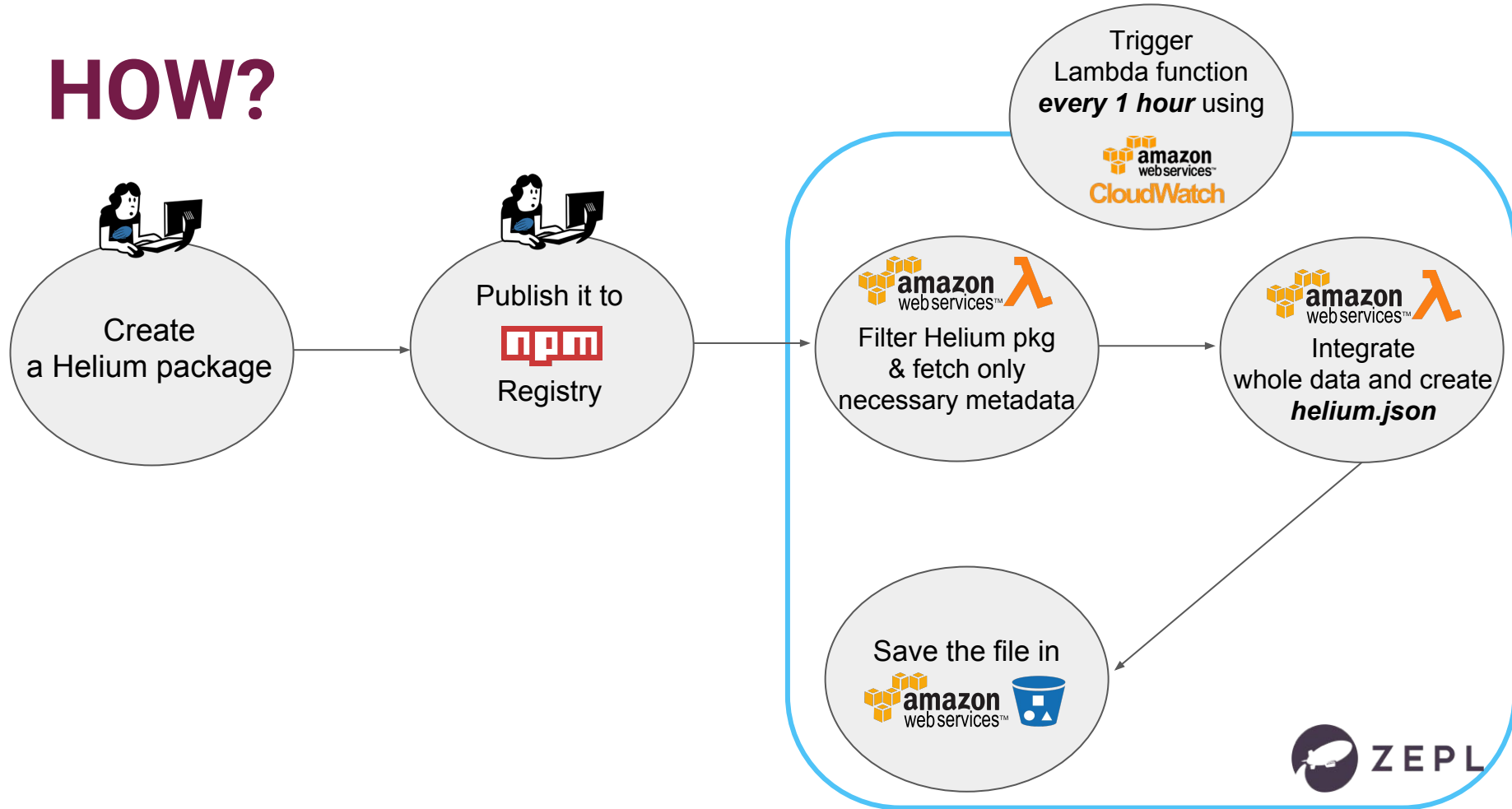
HOW?



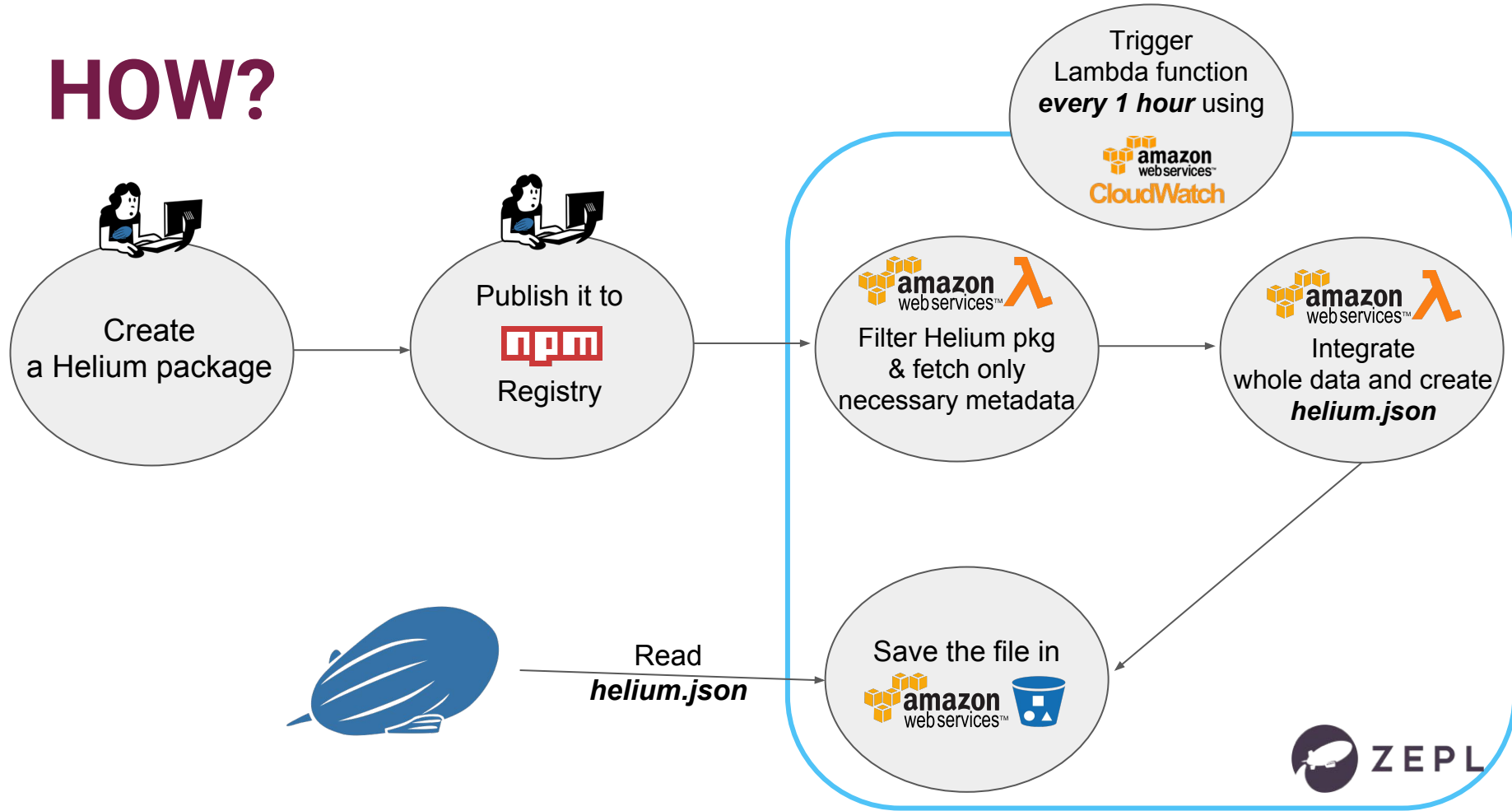
HOW?



HOW?



HOW?



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 <http://www.highcharts.com/> highcharts

v 0.0.1

ultimate-range-chart VISUALIZATION

by ZEPL

The Ultimate Range Chart for Apache Zeppelin using <http://www.highcharts.com/> highcharts




Helium Packages

From Zeppelin-0.7, you can load/unload Helium packages in Zeppelin. Here are the lists of Helium packages. For more information about how you can use Helium packages, see [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
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 [highcharts](#).
v 0.0.1

 **ultimate-range-chart** VISUALIZATION
by ZEPL
The Ultimate Range Chart for Apache Zeppelin using [highcharts](#).

Helium


 VISUALIZATION

 SPELL

 INTERPRETER

** Local registry package*

 **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](#)

Enable

 **ultimate-range-chart** VISUALIZATION

ultimate-range-chart@0.0.1

The Ultimate Range Chart for Apache Zeppelin using [highcharts](#)

Enable

 **ultimate-scatter-chart** VISUALIZATION

ultimate-scatter-chart@0.0.1

The Ultimate Scatter Chart for Apache Zeppelin using [highcharts](#)

Enable

 **ultimate-column-chart** VISUALIZATION

ultimate-column-chart@0.0.1

The Ultimate Column Chart for Apache Zeppelin using [highcharts](#)

Enable

 **ultimate-pie-chart** VISUALIZATION

ultimate-pie-chart@0.0.1

The Ultimate Pie Chart for Apache Zeppelin using [highcharts](#)

Enable

 **ultimate-area-chart** VISUALIZATION

Enable

About Zeppelin

Interpreter

Notebook Repos

Credential

Helium

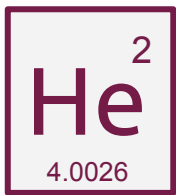
Configuration

JIRA ISSUES

- [ZEPPELIN-1973](#) : List all available Helium packages in Zeppelin website
- [ZEPPELIN-2004](#) : List helium packages in Zeppelin GUI by reading file

WHEN CAN I USE THIS?

- Not included in Zeppelin latest version **0.7.X**
- Will be available in ***Zeppelin 0.8.0***
- Release plan ?



Helium extends Zeppelin eco-system

Visualizations

Bubble Map Heatmap
Range Spline

Spell

Sigma D3 Flowchart
Markdown Translator



Interpreters

Spark Python Geode
Cassandra JDBC Groovy Kylin
Flink

Users / Developers



2833



211 contributors

3rd parties



Useful service



Zeppelin notebook online viewer

<https://www.zeppelinhub.com/viewer>

Notebook sharing and collaboration

<https://www.zeppelinhub.com>



Future Roadmap



2Q
/2017

0.7.2

- Maintenance release.

3Q
/2017

0.8.0

- Helium online registry.
- Interpreter Cluster mode.

3-4Q
/2017

1.0

- Finest, the most stable release

Hopefully



Thanks

Q&A