

APACHE ZEPPELIN TRAINING

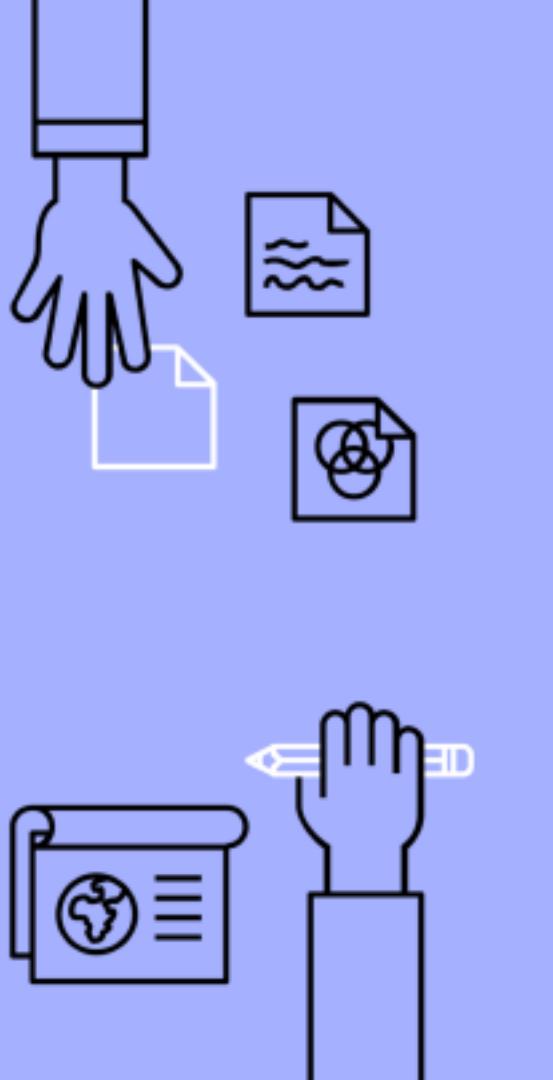


Session 1. Apache Zeppelin Overview

<https://bit.ly/2r5l7Fq>

INDEX

1. 저는요..!
2. 강의 각 회차 별 계획 소개
3. Apache Zeppelin 이 도대체 뭐죠?
4. Apache Zeppelin 설치
5. Hands-on Sessions
 - a. Notebook Versioning
 - b. Change notebook repo to AWS S3
 - c. Turn Shiro authentication on
6. (알쓸신잡) Apache Zeppelin community 이야기



류아영

AHYOUNG RYU

- A software development engineer [@ZEPL](#)
- A PMC(Project Management Committee) of [Apache Zeppelin](#)
- Github [@AhyoungRyu](#)
- Linkedin [@AhyoungRyu](#)
- ahyoungryu@zepl.com



저는 ZEPL 에서 왔어요

- [Apache Zeppelin](#) 을 기반으로하고 Enterprise 환경에서 Team 간 Sharing & Collaboration 에 좀더 집중한 SaaS service [ZEPL](#) 이 저희의 주 프로젝트랍니다.

저는



- Apache Zeppelin 을

간 Sharing & Collabora-

이 저희의 주 프로젝트

The screenshot shows the Zeplin website homepage. At the top right are 'LOG IN' and 'SIGN UP' buttons. Below them are navigation links: 'Products', 'About', 'FAQ', 'Plans & Pricing', 'Explore', 'Blog', and 'Resources'. The main header features the Zeplin logo and the tagline 'Do more with data, faster.' Below this is a sub-tagline '-Collaborative data science in a powerful, shared workspace.' Three callout boxes highlight features: 'A collaborative, cloud-hosted environment', 'A unified interface to all your data', and 'A catalyst for generating insights'. A paragraph below states: 'ZEPL helps businesses and organizations make smart, data-driven decisions by bringing together data science teams to work in a powerful, collaborative workspace.' To the right, a section titled 'Faster time to insight' discusses rapid deployment and tool integration. A large circular graphic at the bottom left contains the text 'Bring together your data and tools in a powerful, unified interface'. A note at the bottom right says 'Your secure, cloud-hosted workspace syncs live data with all your platforms. Get up and running in minutes, with no need to install or learn anything.'

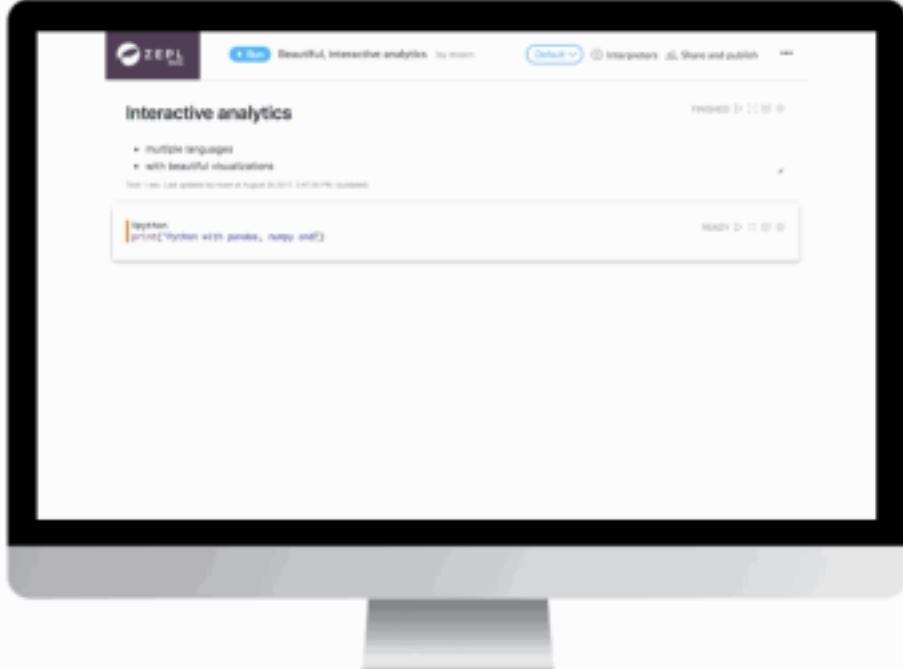
<https://zepl.com>

저는

- Apache Z

간 Sharir

이 저희으



The image shows a computer monitor displaying the Zeplin website. The website has a dark header with the Zeplin logo and navigation links for 'Beautiful, interactive analytics', 'Interpreters', and 'Share and publish'. Below the header, there's a section titled 'Interactive analytics' with a brief description: 'multiple languages', 'beautiful visualizations', and a note about being last updated on August 20, 2017, at 2:47:36 PM (UTC). A code editor window shows a single line of Python code: 'print("Hello world")'. To the right of the monitor, there's a promotional text block with a title 'Do more with data', a descriptive paragraph about Zeplin's capabilities, and a 'EXPLORE FEATURES >' button.

Do more with data

Take your team from discovery to visualization to analytics in a single notebook.

Run live code – Python, R, Scala, SQL, and more – against multiple data sources in a secure, multi-tenant environment.

[EXPLORE FEATURES >](#)

저는

- Apache Zeppelin

간 Sharir

이 저희으

The screenshot shows the Zeppelin organization page for 'ZEPL'. The left sidebar has 'Organization ZEPL' at the top, followed by 'Settings', 'Teams', and 'Members' (which is highlighted). Below the sidebar, a message says 'You are a member of this organization' and 'can see list of teams and members'. The main content area is titled 'Members' with a search bar. It lists ten members with their profile icons and names: lambda, lambda, shyoung, anthonycorbacho, khalid, linko, khald, khaldhusaynov, leemoo, and marc.

Member	Role
lambda	member
lambda	member
shyoung	member
anthonycorbacho	manager
khalid	member
linko	member
khald	member
khaldhusaynov	member
leemoo	member
marc	member

저는 ZEPL 에서 왔어요

- Apache Zeppelin 을 기반으로하고 Enterprise 환경에서 Team 간 Sharing & Collaboration 에 좀더 집중한 SaaS service ZEPL 이 저희의 주 프로젝트랍니다.
- Apache Zeppelin의 아버지, 이문수 CTO 님은 ZEPL 의 co-founder이자 Apache Zeppelin 프로젝트의 PMC Chair이며 이 프로젝트를 처음으로 만드셨어요.



Lee moon soo
Leemoonsoo

Unfollow

Block or report user

ZEPL

Organizations



저는 ZEPL 에서 왔어요

- [Apache Zeppelin](#) 을 기반으로하고 Enterprise 환경에서 Team 간 Sharing & Collaboration 에 좀더 집중한 SaaS service [ZEPL](#) 이 저희의 주 프로젝트랍니다.
- Apache Zeppelin의 아버지, 이문수 CTO 님은 ZEPL 의 co-founder이자 Apache Zeppelin 프로젝트의 PMC Chair이며 이 프로젝트를 처음으로 만드셨어요.
- 멋진 Open Source 프로젝트, Apache Zeppelin이 태어난 곳인 만큼 아주 cool 하고 open 된 문화를 가진 회사랍니다 :)



각 회차별 (간략한) 주제 소개

1회차 (2018. 04. 28) - Apache Zeppelin Overview (1)

- Apache Zeppelin의 Data Science Life Cycle에서의 위치 및 역할
- Apache Zeppelin A to Z
 - Interpreter Settings
 - Job Management
 - Notebook Repositories
 - Credential Settings
 - Helium
 - Configurations
- Hands on Sessions
 - Notebook Versioning
 - Change notebook repository to AWS S3
 - Shiro Authentication
- Open source project 인 Apache Zeppelin, 어떻게 개발되고 운영될까요 ?

1회차 (2018. 04. 28) - Apache Zeppelin Overview (1)

- Apache Zeppelin의 Data Science Life Cycle에서의 위치 및 역할
- Apache Zeppelin A to Z
 - Interpreter Settings
 - Job Management
 - Notebook Repositories
 - Credential Settings
 - Helium
 - Configurations
- Hands on Sessions
 - Notebook Versioning
 - Change notebook repository to AWS S3
 - Shiro Authentication
- Open source project 인 Apache Zeppelin, 어떻게 개발되고 운영될까요 ?

2회차 (2018. 05. 12) - Apache Zeppelin Overview (2)

- Multi user environment (2)
 - Notebook ACL / Interpreter ACL
 - Personalized mode
- Helium : pluggable system in Zeppelin
 - 기본 NVD3라이브러리 차트 말고 더 다양하고 화려한 차트 라이브러리를 원한다면? Helium!
 - Helium online package registry 에 있는 package runtime 에 설치 및 실행
 - Frontend interpreter: Spell

2회부터 각 리스트별 자세한 내용은 매회 강의때 다시 알려드릴 예정입니다

3회차 (2018.05.19) - Apache Zeppelin Advanced (1)

- Zeppelin & Exploratory Data Analysis (EDA)
 - Histogram / Scatter plot / Heatmap (?) / Box plot
 - Scikit-learn의 boston dataset
- Zeppelin as Business Intelligence (BI)
 - Dynamic forms (text / select / checkbox)
 - Using form templates (scope: notebook / paragraph)
 - Custom form using Angular Bootstrap

2회부터 각 리스트 별 자세한 내용은 매회 강의 때 다시 알려드릴 예정입니다

3회차 (2018.05.19) - Apache Zeppelin Advanced (1)

- Zeppelin & Exploratory Data Analysis (EDA)
 - Histogram / Scatter plot / Heatmap (?) / Box plot
 - Scikit-learn의 boston dataset
- Zeppelin as Business Intelligence (BI)
 - Dynamic forms (text / select / checkbox)
 - Using form templates (scope: notebook / paragraph)
 - Custom form using Angular Bootstrap

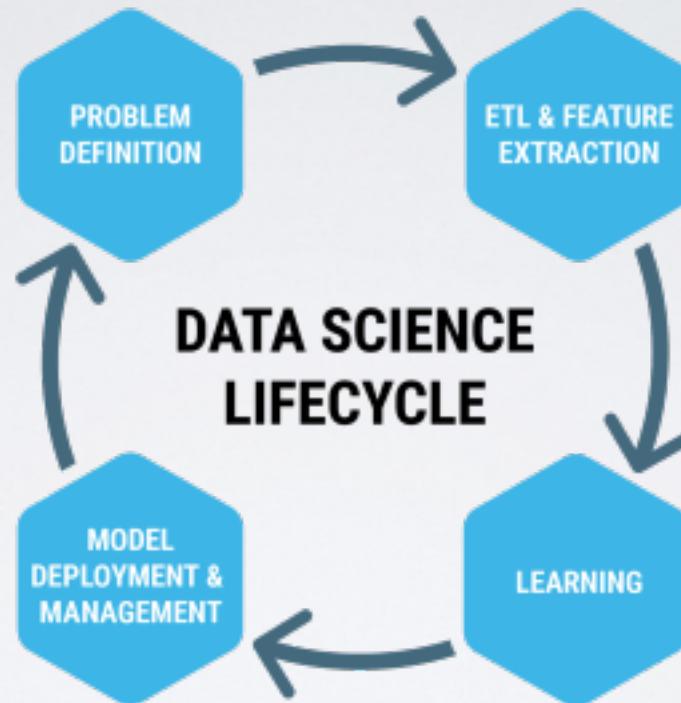
4회차 (2018.05.26) - Apache Zeppelin Advanced (2)

- Building my own BI interface
 - Building my own BI interface
 - CRON을 이용하여 노트 주기적으로 실행하기
 - REST API를 이용하여 비동기적으로 노트 실행하기
- Zeppelin 과 Machine Learning (ML)
 - Logistic regression & decision tree training result visualization
 - Text mining
 - Spark를 위한 한국어 형태소 분석기 “Spark NKP” 소개
 - Word2Vec visualization
- 추가 유용한 팁들
 - How to set a note as your Zeppelin homepage

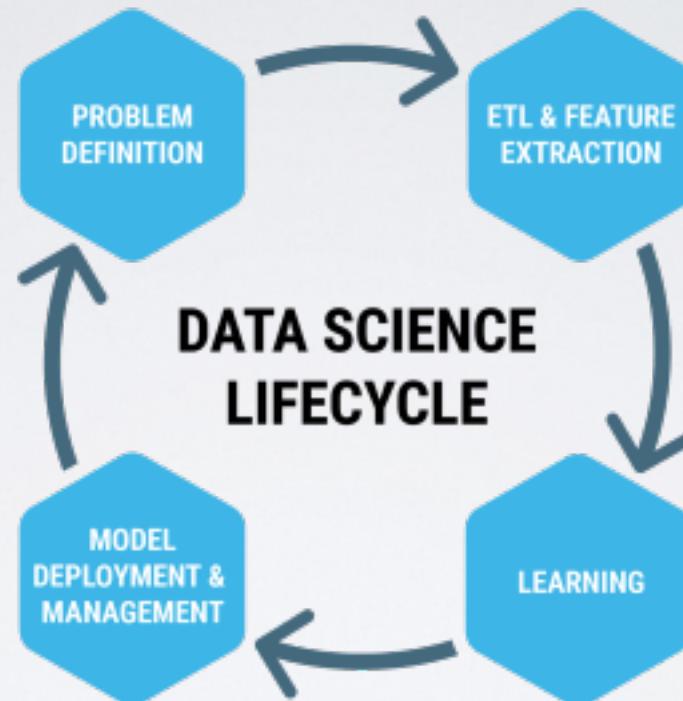
2회부터 각 리스트 별 자세한 내용은 매회 강의 때 다시 알려드릴 예정입니다

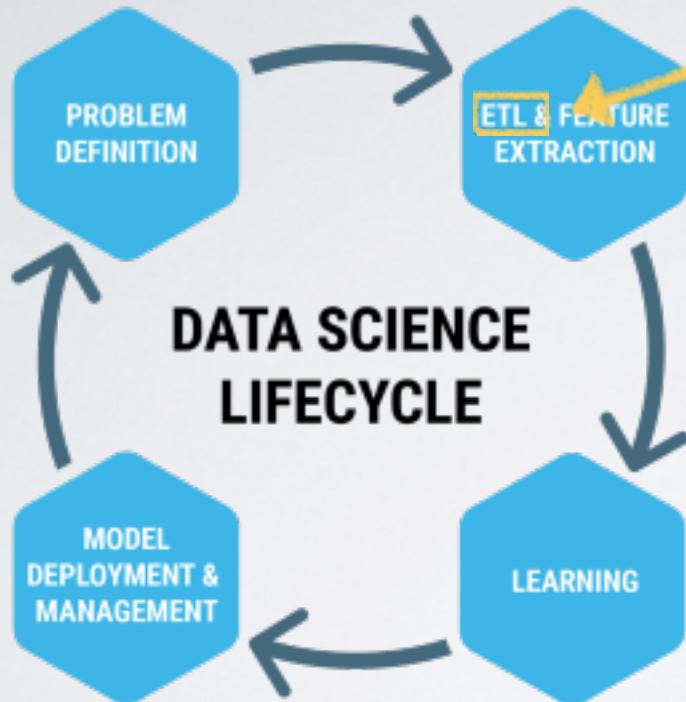
Apache Zeppelin ?

이런 다이어그램이 또 빠질수 없죠



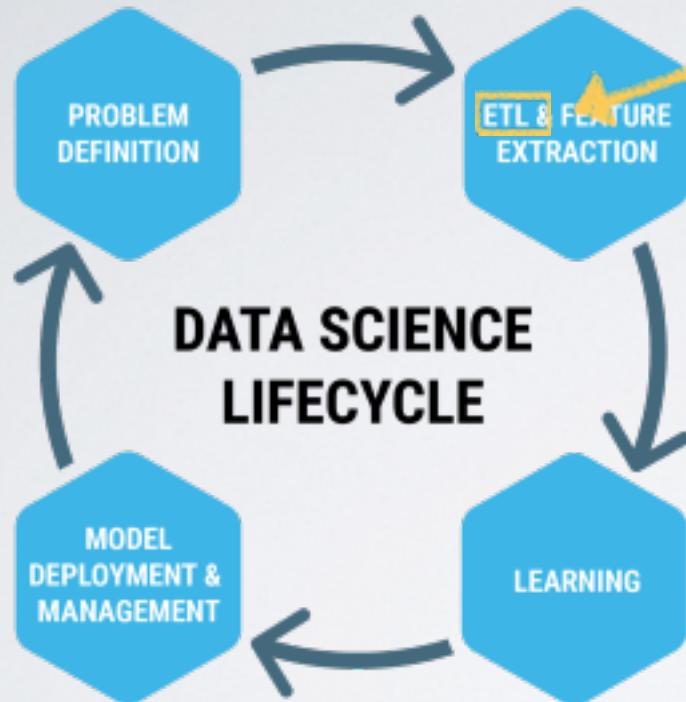
Apache Zeppelin은 어느 단계에 속할까요?





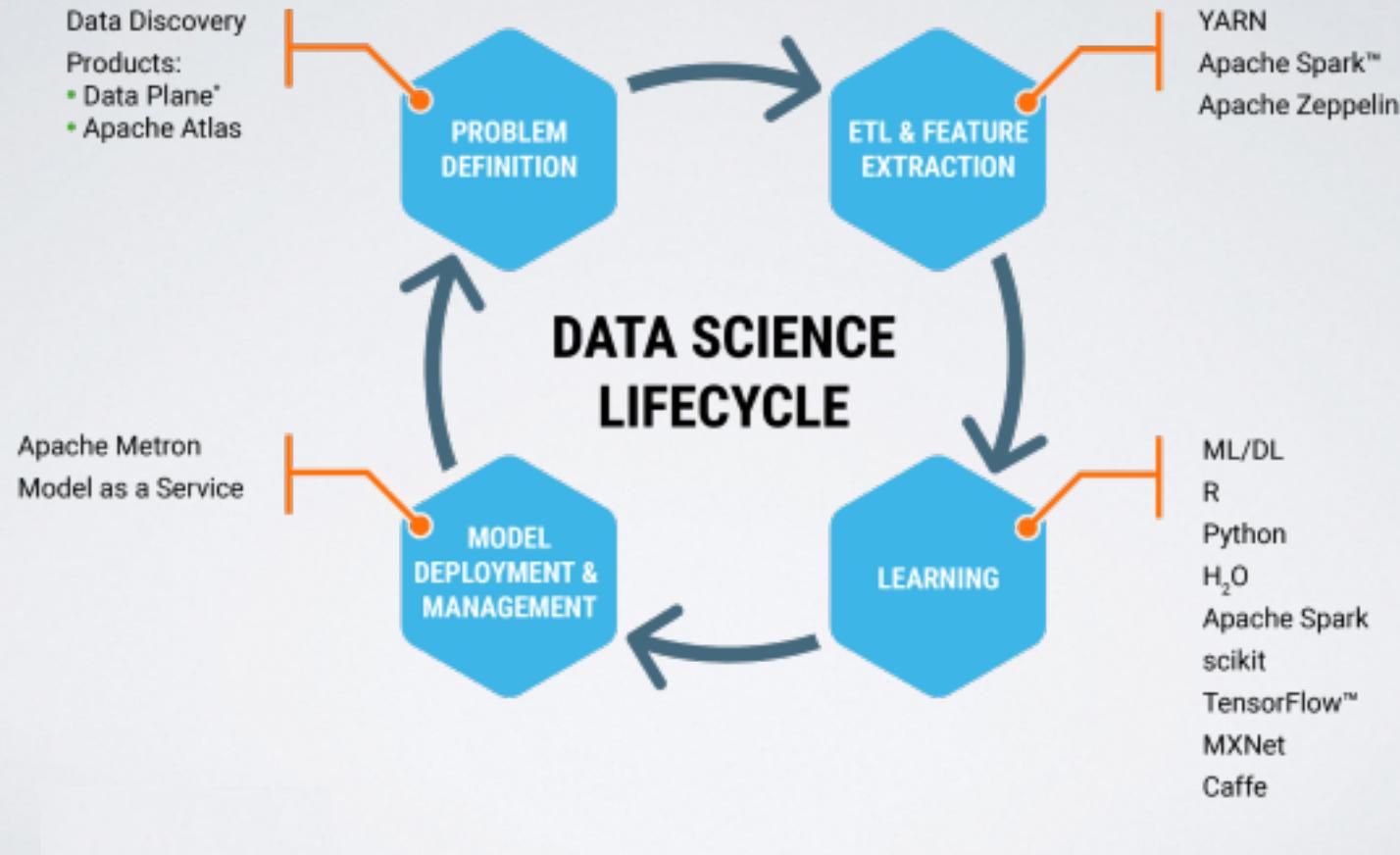
일반적으로 기업 인프라를 구성하는 시스템에서는 매우 다양한 로그 파일과 데이터를 생성해내죠.

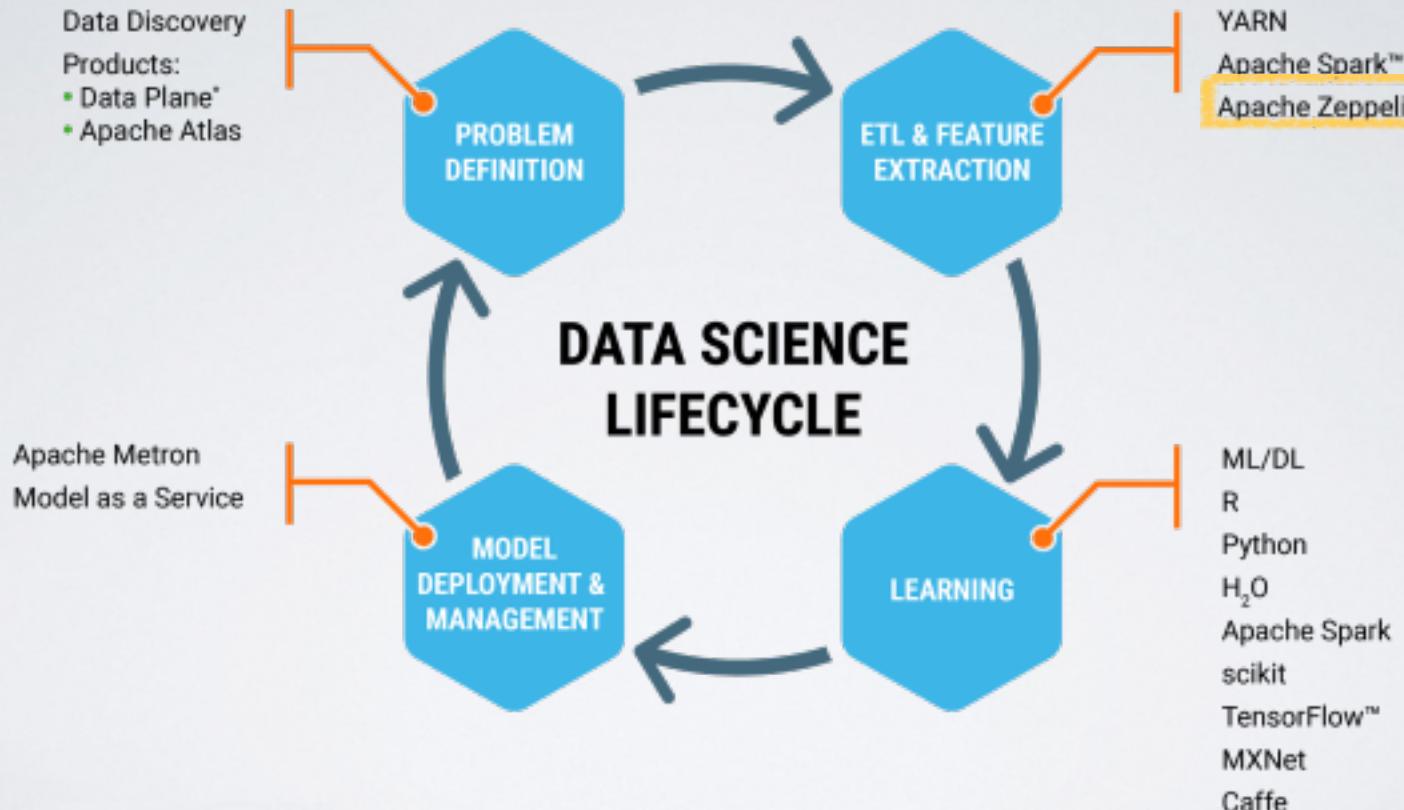
이러한 로그 파일을 처리하려면 적절하게 데이터를 추출하고, 사용하기 쉬운 형태로 가공해야 하며, 데이터베이스 같은데 이터 스토어에 로딩해야 합니다. 이런 일련의 작업을 지원하는 시스템을 ETL(Extract, Transform, Load) 시스템이라고 해요.

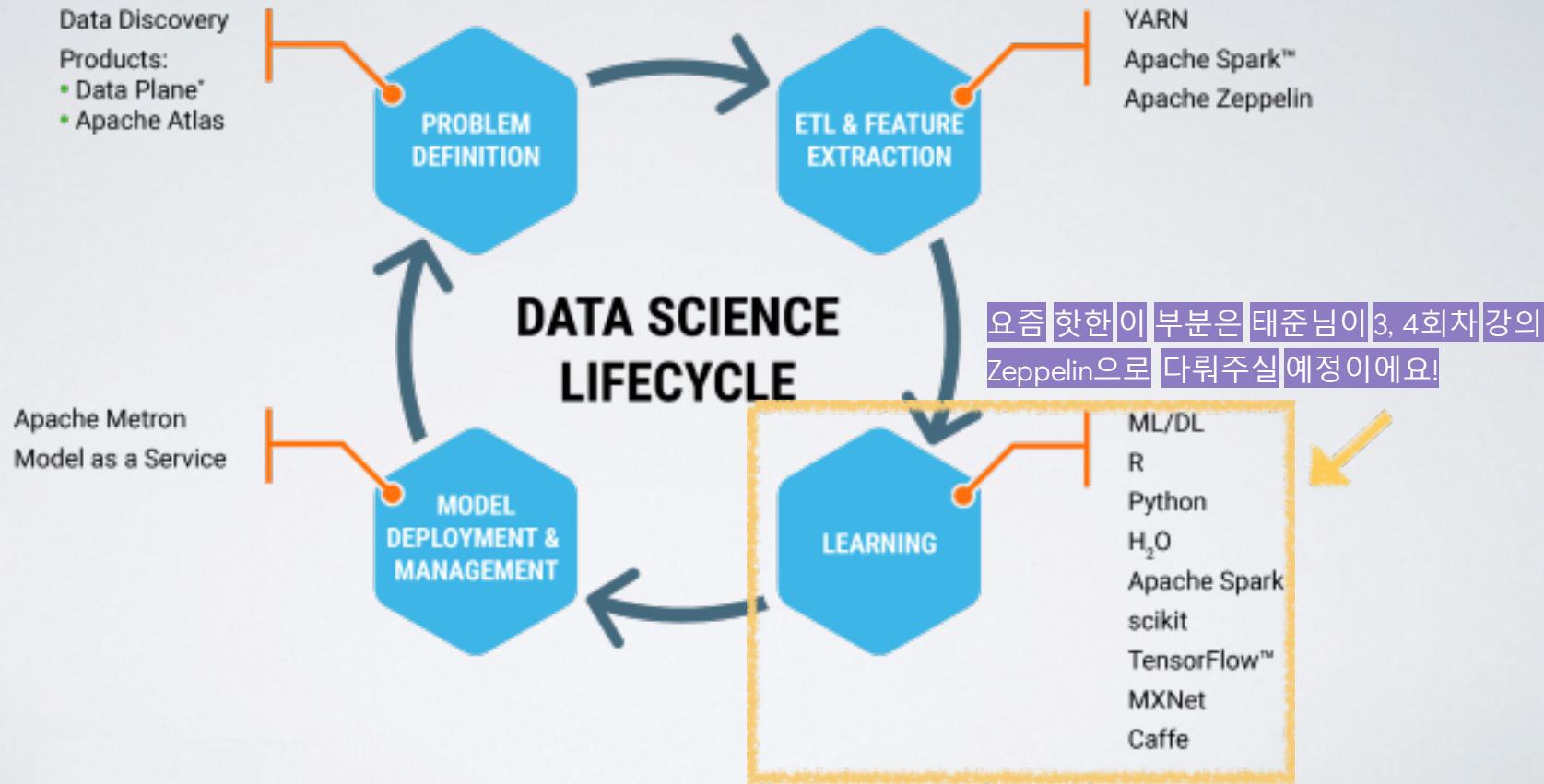


ETL 시스템의 목적 한 가지는

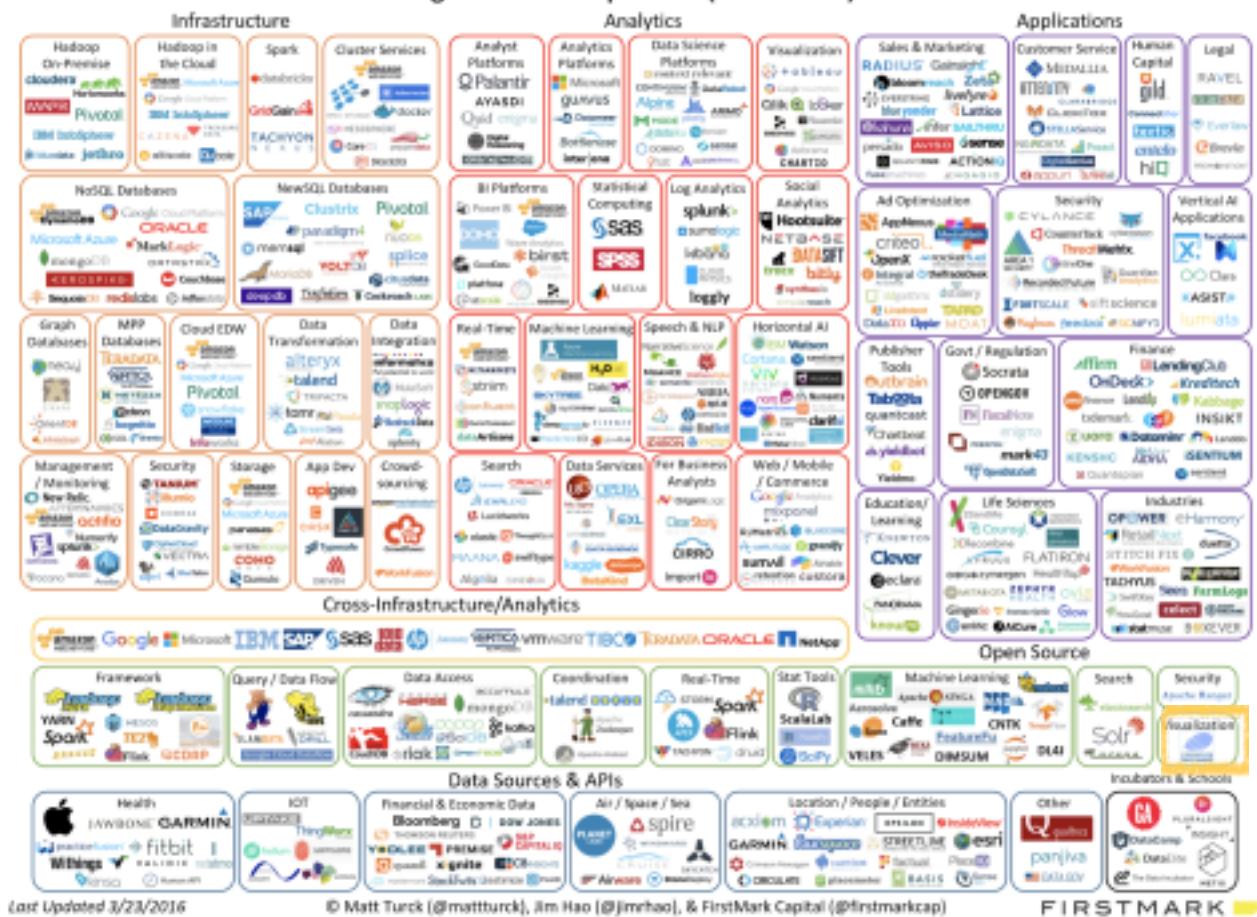
빠르고 손쉽게 데이터를 가공하는 것으로 손쉽게 구현하기 위한 GUI를 제공하고 이 GUI로 데이터의 흐름, 그리고 각 단계에서 데이터 형식을 변경하거나 처리하는 모듈을 추가하는 것을 지원합니다.





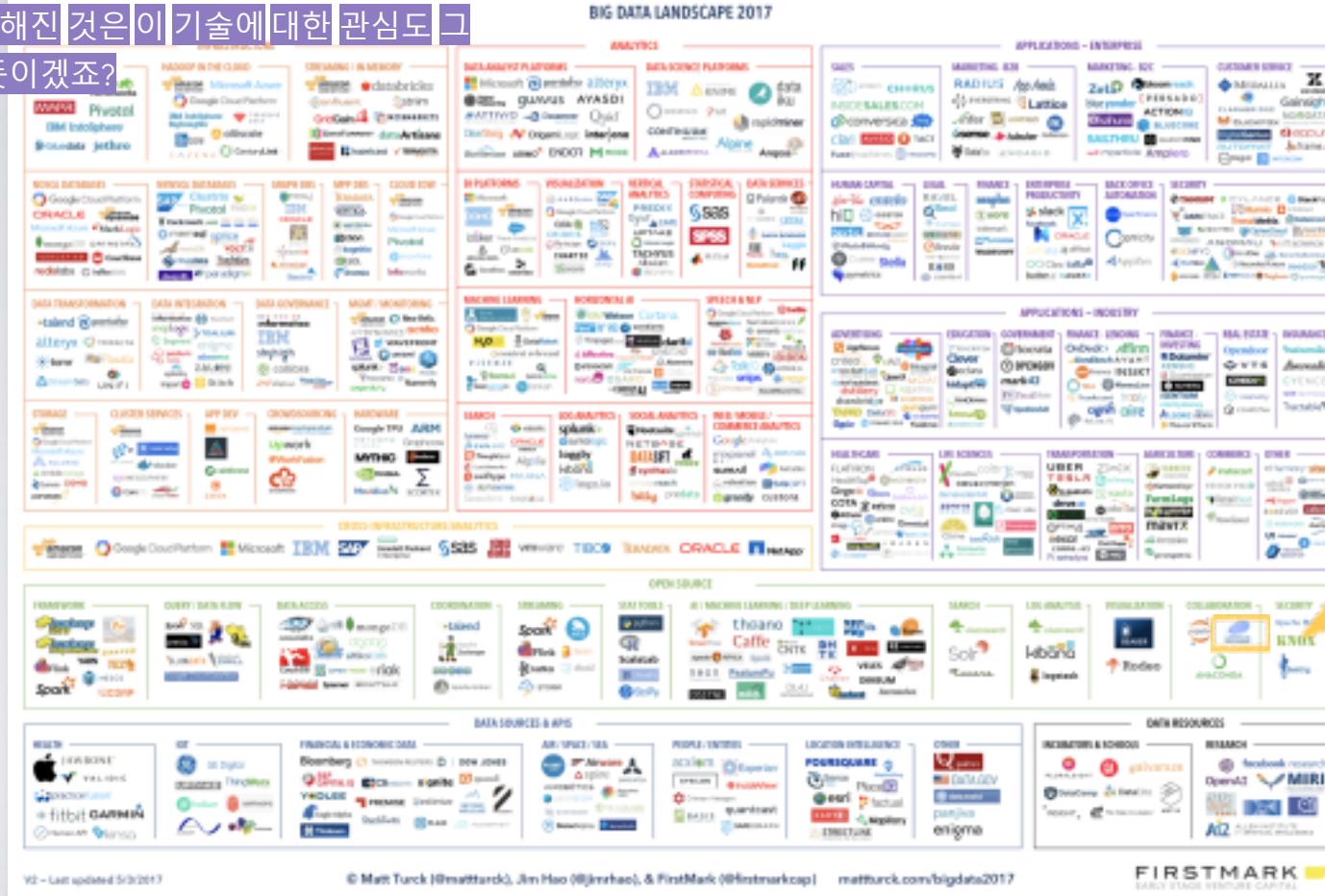


Big Data Landscape 2016 (Version 3.0)



“Visualization”
에 자리잡기 시
작하여
(Zeppelin 초창기
로고...)

1년 사이 더 거대해진 것은 이 기술에 대한 관심도 그 만큼 커졌다라는 뜻이겠죠?



“Collaboration” 으 로

Bossie Awards 2015, 2016 그리고 2017

The screenshot shows a news article from InfoWorld titled "Bossie Awards 2015: The best open source big data tools". The article discusses InfoWorld's top picks in distributed data processing, streaming analytics, machine learning, and other corners of large-scale data analytics. It features a sidebar for sharing the article and navigation controls. A large image of the Apache Zeppelin incubating project is displayed, showing its interface and capabilities. Below the image, there is a brief description of Zeppelin.

2015년엔 incubating project였음에도 불구하고

Bossie Awards 2015: The best open source big data tools

InfoWorld's top picks in distributed data processing, streaming analytics, machine learning, and other corners of large-scale data analytics

By InfoWorld staff, InfoWorld | Sep 16, 2015

Share

All Pages

Slide 21 of 22

Apache Zeppelin (incubating)

A web-based notebook that enables interactive data analysis. You can make beautiful data visual, interactive and collaborative documents.

Watch the video Get Started

Multi-purpose Notebook

The predecessor to the project is now called Zeppelin.

- Data Ingestion
- Data Discovery
- Data Analytics
- Data Visualization & Collaboration

Multiple language Backend

Apache Zeppelin is an open-source big data processing system developed by Apache. Current development is focused on support for various data sources, processing systems (Spark, Flink, Storm, etc.) and visualization libraries.

Zeppelin

Zeppelin

While still in incubation, **Apache Zeppelin** is nevertheless stirring the data analytics and visualization pot. The Web-based notebook enables users to ingest, discover, analyze, and visualize their data. The notebook also allows you to collaborate with others to

[Bossie Awards 2015: The best open source big data tools](#)

Bossie Awards 2015, 2016 그리고 2017

Software Development • Cloud Computing • Machine Learning • Analytics • Insider Exclusives • Resources & White Papers • Job Search

InfoWorld

Home > Open-Source

Bossie Awards 2016: The best open source big data tools

InfoWorld's top picks in large-scale search, SQL on Hadoop, streaming analytics, and other flavors of distributed data processing

By Andrew C. Oliver and Ian Pointer, InfoWorld | Sep 21, 2016

share All Slides Slide 13 of 13



Zepplin

Whether you're a developer who simply wants a pretty graph out of Hive or a data scientist who wants a notebook, **Zeppelin** might be for you. It uses the now-familiar notebook concept made popular by IPython, allowing you to write markup, embed

[Bossie Awards 2016: The best open source big data tools](#)

Apache Spark, Kafka, Jupyter, Tensorflow 등등

Bossie Awards 2015, 2016 그리고 2017

The screenshot shows a news article from InfoWorld titled "Bossie Awards 2017: The best databases and analytics tools". The article discusses the best open source software for large-scale search, SQL, NoSQL, and streaming analytics. It was written by Martin Heller, Steven Muey, and Andrew C. Oliver on Sep 27, 2017. The page includes a sidebar with navigation links like Home, Analytics, and a search bar. Below the article, there's a slide show featuring the Apache Zeppelin interface and the BOSSIE 2017 AWARDS logo.

Bossie Awards 2017: The best databases and analytics tools

InfoWorld picks the best open source software for large-scale search, SQL, NoSQL, and streaming analytics

By [Martin Heller](#), [Steven Muey](#) and [Andrew C. Oliver](#), InfoWorld | Sep 27, 2017

Share All Slides Slide 10 of 15

Welcome to Zeppelin

Apache Zeppelin

Technologies

spark SQL python

InfoWorld BOSSIE 2017 AWARDS

Apache Zeppelin

The Apache Software Foundation

Apache Zeppelin

Apache Zeppelin is a sort of Rosetta Stone for developers, analysts, and data scientists alike. A [view of interpreters](#) allow you to pull from different data stores and analyze in real-time environments. This means you can [pull data from your favorite NoSQL](#).

3년 연속으로 Bossie Awards에 선정된
open source project는 Apache Spark /
Kafka/ Apache Zeppelin 이렇게 세가지가
눈에 띠는군요!(일단 제 눈에는..)

[Bossie Awards 2017: The best databases and analytics tools](#)



Apache Zeppelin

공식 문서에 의하면...



Apache Zeppelin

공식 문서에 의하면...



- Data exploration and discovery
- Visualization
- Collaboration and publishing



Apache Zeppelin

공식 문서에 의하면...



- Data exploration and discovery
 - Visualization
 - Collaboration and publishing

“Modern Data Science Studio”



Apache Zeppelin



Star 갯수로는 ASF 프로젝트들 중 15위 이내

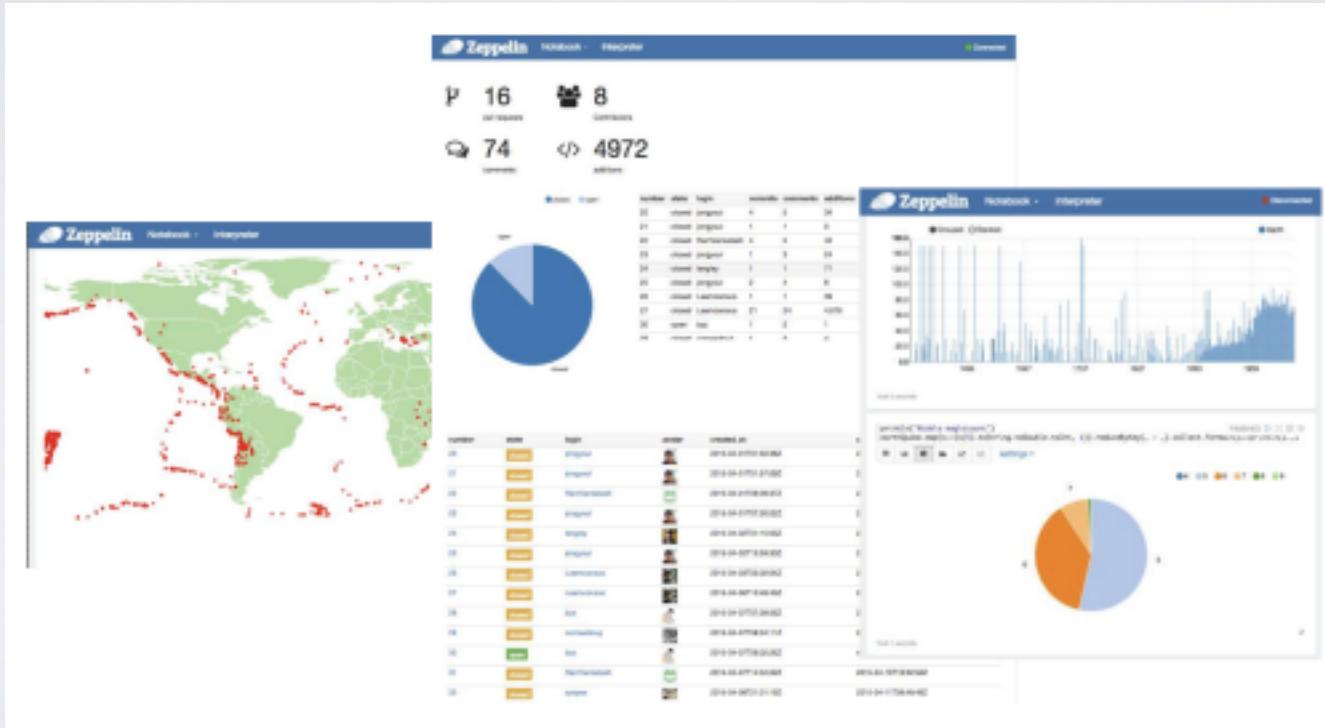
- Apache Top Level Project
- <https://github.com/apache/zeppelin> (3500+ stars)
- Built-in Apache Spark 지원

Zeppelin 내에서 Spark 을 사용하기 위해 따로

Spark 을 설치하지 않아도 되어요

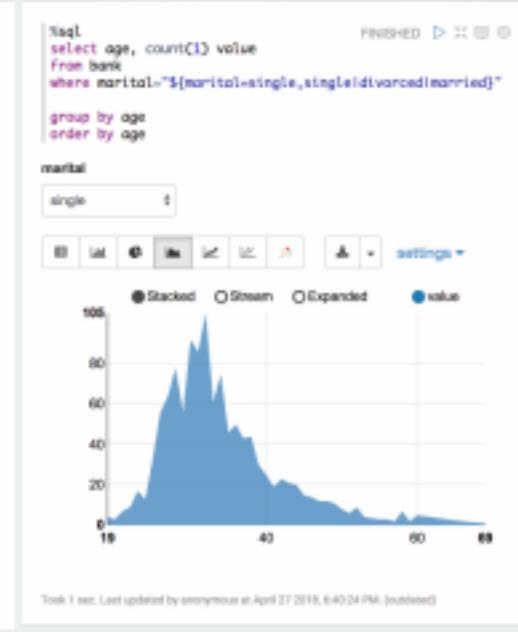
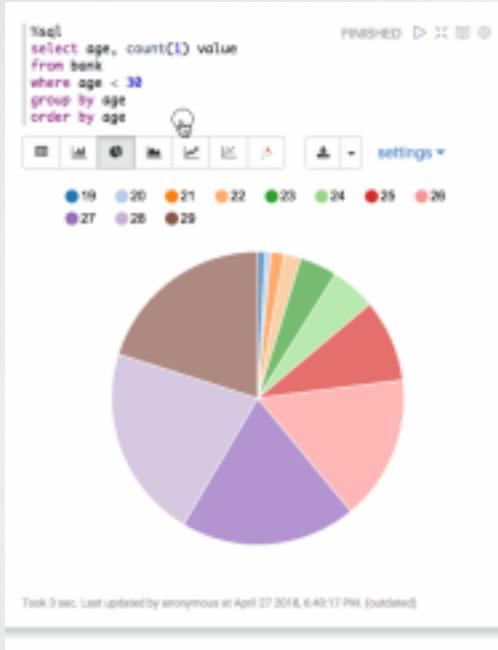
(하지만 이 때문에 덩치가 크죠ㅠㅠ)

Interactive Notebooks



Interactive Visualization

Zeppelin 의 Dynamic Form 기능은 visualization을 더
욱 interactive 하고 빠른 결과 확인을 가능하게 하죠



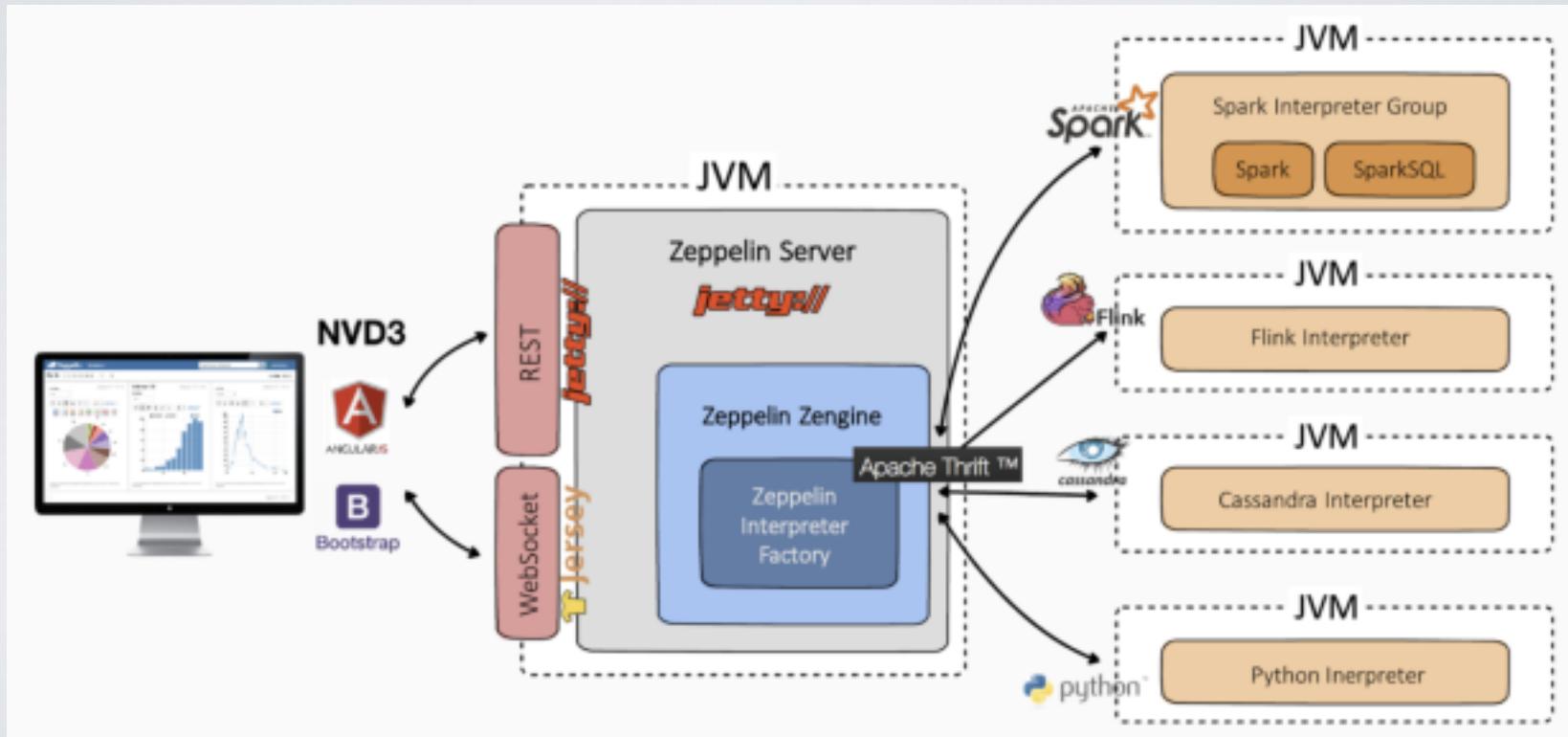
Multiple Backend



현재 20개가 넘는 Backend interpreter 를 지원하고 있답니다



한눈에 보는 Zeppelin Structure





Apache Zeppelin 이럴때 쓰면 좋아요

- 팀 Dash board 가 지금 당장 필요한데 개발할 시간이 없을 때
- 데이터가 이곳 저곳에 파편화 되어 있고 한군데서 보길 원할 때 (RDB, S3, ElasticSearch, HDFS, Local CSV, ...)
- 매번 parameter만 다른 데이터를 뽑아주기 귀찮을 때 => Dynamic Form!
- 데이터를 디자이너나 비즈니스팀에게 바로 쓸 수 있게 공유하길 원할 때

서론이 많이 길었죠?

Zeppelin 설치부터 같이 해볼까요 ?

이 수업에 사용 될 ZEPPELIN의 2가지 VERSION



0.7.3 Ver.

이 수업에 사용 될 ZEPPELIN의 2가지 VERSION



0.7.3 Ver.

- 2017.09.21 에 릴리즈된 Official latest & stable version
- 2회차 강의 “Apache Zeppelin & Helium” 수업을 제외하고 모든 수업에 사용될 버전입니다
- <https://zeppelin.apache.org/docs/0.7.3/>

이 수업에 사용 될 ZEPPELIN의 2가지 VERSION



0.7.3 Ver.

마지막 릴리즈가 벌써 7개월전이군요..

- 2017.09.21 에 릴리즈된 Official latest & stable version
- 2회차 강의 “Apache Zeppelin & Helium” 수업을 제외하고 모든 수업에 사용될 버전입니다
- <https://zeppelin.apache.org/docs/0.7.3/>

이 수업에 사용 될 ZEPPELIN의 2가지 VERSION



0.7.3 Ver.

- 2017.09.21 에 릴리즈된 Official latest & stable version
- 2회차 강의 “Apache Zeppelin & Helium” 수업을 제외하고 모든 수업에 사용될 버전입니다
- <https://zeppelin.apache.org/docs/0.7.3/>



0.8.0-SNAPSHOT Ver.

이 수업에 사용 될 ZEPPELIN의 2가지 VERSION



0.7.3 Ver.

- 2017.09.21 에 릴리즈된 Official latest & stable version
- 2회차 강의 “Apache Zeppelin & Helium” 수업을 제외하고 모든 수업에 사용될 버전입니다
- <https://zeppelin.apache.org/docs/0.7.3/>



0.8.0-SNAPSHOT Ver.

The "SNAPSHOT" term means that the build is a snapshot of your code at a given time.

- 최신 master branch 의 source code 를 의미
- 아직 공식 릴리즈 되지 않아 직접 Source Code 를 빌드해서 사용해야 해요
- 다음 2회차 강의에선 시간 단축을 위해 제가 만든 Docker image 를 이용할 예정이랍니다 □
- <http://zeppelin.apache.org/docs/0.8.0-SNAPSHOT/>

이 수업에 사용 될 ZEPPELIN의 2가지 VERSION



0.7.3 Ver.

- 2017.09.21 에 릴리즈된 Official latest & stable version
- 2회차 강의 “Apache Zeppelin & Helium” 수업을 제외하고 모든 수업에 사용될 버전입니다
- <https://zeppelin.apache.org/docs/0.7.3/>



0.8.0-SNAPSHOT Ver.

- 최신 master branch 의 source code 를 의미
- ~~• 아직 공식 릴리즈 되지 않아 직접 Source Code 를 빌드해서 사용해야 해요~~
- ~~• 다음 2회차 강의 예선 시간 단축을 위해 제가 만든 Docker image 를 이용할 예정이랍니다 □~~
- <http://zeppelin.apache.org/docs/0.8.0-SNAPSHOT/>

Starting to Prepare 0.8.0 RC1

Inbox x Apache Zeppelin/Users x

 Jeff Zhang <jifdu@gmail.com>
to dev, users ▾

10:12 AM (4 hours ago) ⭐ ↗ ↘

Hi Folks,

All the block issues are fixed, I am starting to prepare 0.8.0 RC1

바로 어제! 저희가 Zepl project 개발에 매진하느냐고
 계속 하지 못하고 있던 0.8.0 릴리즈를 Zeppelin의 또 다른
 PMC member 인 Jeff 가 릴리즈 프로세스를 자처하여 시작
 했어요

 Xiaohui Liu <herobh@gmail.com>
to users, dev ▾

10:13 AM (4 hours ago) ⭐ ↗ ↘

Fantastic!

...

 Jongyoul Lee <jongyoul@gmail.com>
to users, dev ▾

10:15 AM (4 hours ago) ⭐ ↗ ↘

Sounds great!

...

 Jun Kim
to users, dev ▾

12:17 PM (2 hours ago) ⭐ ↗ ↘

Awesome!

2018년 4월 27일 (금) 오전 10:15, Jongyoul Lee <jongyoul@gmail.com>님이 작성:

...

 Mina Lee <minallee@apache.org>
to users, dev ▾

12:38 PM (1 hour ago) ⭐ ↗ ↘

Nice! Let me know if you need any help

...

- <https://zeppelin.apache.org/docs/0.7.3/>

→ 디렉토리에서 실행되는 시나리오를 위해 새 가상 Docker
 image 를 이용할 예정이랍니다 □
 • <http://zeppelin.apache.org/docs/0.8.0-SNAPSHOT/>

이 수업에 사용 될 ZEPPELIN의 2가지 VERSION



0.7.3 Ver.

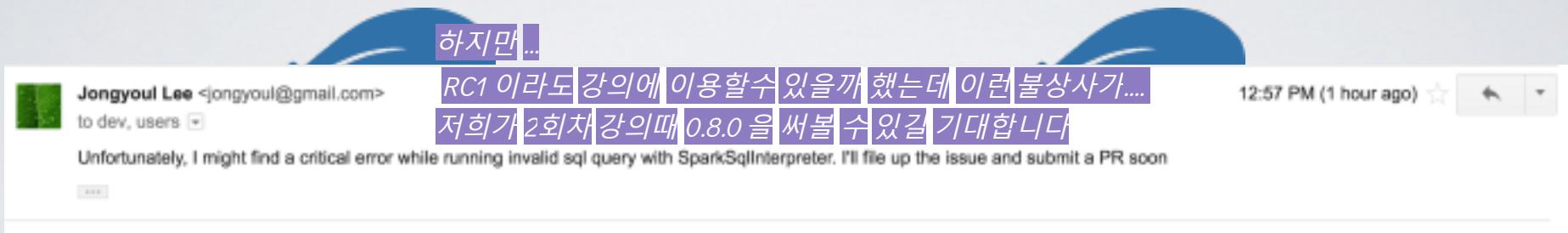
- 2017.09.21에 릴리즈된 Official latest & stable version
- 2회차 강의 “Apache Zeppelin & Helium” 수업을 제외하고 모든 수업에 사용될 버전입니다
- <https://zeppelin.apache.org/docs/0.7.3/>



0.8.0-SNAPSHOT Ver.

- 최신 master branch 의 source code 를 의미
- 다행히 다음 0.8.0 버전이 곧 나올것 같아요 2회차 강의땐 다운로드 가능한 packaging 된 binary package 를 사용할 수 있을것 같아요□
- <http://zeppelin.apache.org/docs/0.8.0-SNAPSHOT/>

이 수업에 사용 될 ZEPPELIN의 2가지 VERSION



0.7.3 ver.

- 2017.09.21에 릴리즈된 Official latest & stable version
- 2회차 강의 “Apache Zeppelin & Helium” 수업을 제외하고 모든 수업에 사용될 버전입니다
- <https://zeppelin.apache.org/docs/0.7.3/>

0.8.0-SNAPSHOT ver.

- 최신 master branch 의 source code 를 의미
- 다행히 다음 0.8.0 버전이 곧 나올것 같아요 2회차 강의땐 다운로드 가능한 packaging 된 binary package 를 사용할 수 있을것 같아요
- <http://zeppelin.apache.org/docs/0.8.0-SNAPSHOT/>

Zeppelin 설치를 위한 환경 설정 - 방법 1

Linux 혹은 OSX 환경의 노트북 맞죠..?

1. JDK8 를 설치하고 JAVA_HOME 과 PATH 를 세팅해 주세요

- (Ubuntu Linux) https://zetawiki.com/wiki/우분투_Java_8_설치
- (OSX) <http://recipes4dev.tistory.com/52>

```
ahyoungryu at MacBook-Pro-6 in ~/Dev/zeppelin-bin/zeppelin-0.7.3-bin-all
$ javac -version
javac 1.8.0_152
ahyoungryu at MacBook-Pro-6 in ~/Dev/zeppelin-bin/zeppelin-0.7.3-bin-all
$ echo $JAVA_HOME
/Library/Java/JavaVirtualMachines/jdk1.8.0_152.jdk/Contents/Home
```

꼭 위의 버전과 같지 않아도 됩니다. 환경 세팅 여부만 확인되면 됩니다

2. 이제 Zeppelin 0.7.3 version을 다운받아 설치해 주세요

Zeppelin-0.7.X-bin-all vs. Zeppelin-0.7.x-bin-netinst ?

- 차이점 : “all” package엔 모든 interpreter들이 설치되어있고, “netinst” package엔 Apache Spark interpreter만 포함되어있고, 나머지 interpreter들은 직접 Script를 통해 설치해줘야 해요. 이를 위한 interpreter-installation script가 들어 있답니다.
- Why ? 20+ 개의 interpreter들이 모두 필요하지 않은 사용자의 경우 거의 800MB의 binary package 를 다운받아 사용하는 것이 불필요할 수 있기 때문이죠
- 0.7.3 버전 설치 가이드: <http://zeppelin.apache.org/docs/0.7.3/install/install.html#installation>
- “Netinst” pkg 를 위한 interpreter installation 가이드:
<http://zeppelin.apache.org/docs/0.7.3/manual/interpreterinstallation.html>

Zeppelin 설치를 위한 환경 설정 - 방법 2

Apache Zeppelin은 [Official Docker Image](#) 를 지원합니다

1. Docker (가 없다면) [설치](#)
2. Terminal에서 `docker run -p 8080:8080 apache/zeppelin:0.7.3` 실행
3. Zeppelin 설치 끝! 간단하죠?

자 이제 **Zeppelin** 화면을 띄워봅시다!

- Zeppelin 설치 디렉토리 아래에서 `./bin/zeppelin-daemon start`
- 브라우저에서 <http://localhost:8080> 열기
- 만약 실행 도중 문제가 있다면 Zeppelin 설치 디렉토리 안에 있는 `logs/` 아래에 있는 log 파일들을 보면 어느부분이 잘못되었는지 힌트를 얻을 수 있어요!
- <http://zeppelin.apache.org/docs/0.7.3/install/install.html#starting-apache-zeppelin>

이 화면 보이시나요?

 **Zeppelin** Notebook • Job Search your Notes  anonymous •

Welcome to Zeppelin!

Zeppelin is web-based notebook that enables interactive data analytics.
You can make beautiful data-driven, interactive, collaborative document with SQL, code and even more!

Notebook 

-  Import note
-  Create new note



-  Fast Campus
-  Viz example using Helium pkg
-  Zeppelin Tutorial

Help 

Get started with [Zeppelin documentation](#)

Community 

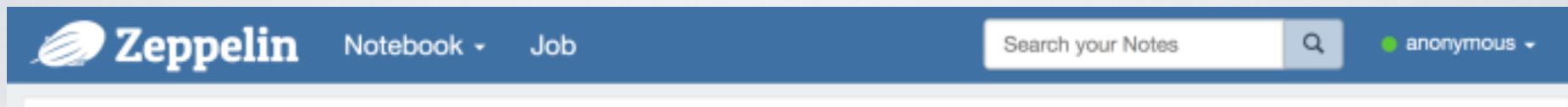
Please feel free to help us to improve Zeppelin,
Any contribution are welcome!

 Mailing list
 Issues tracking
 Github



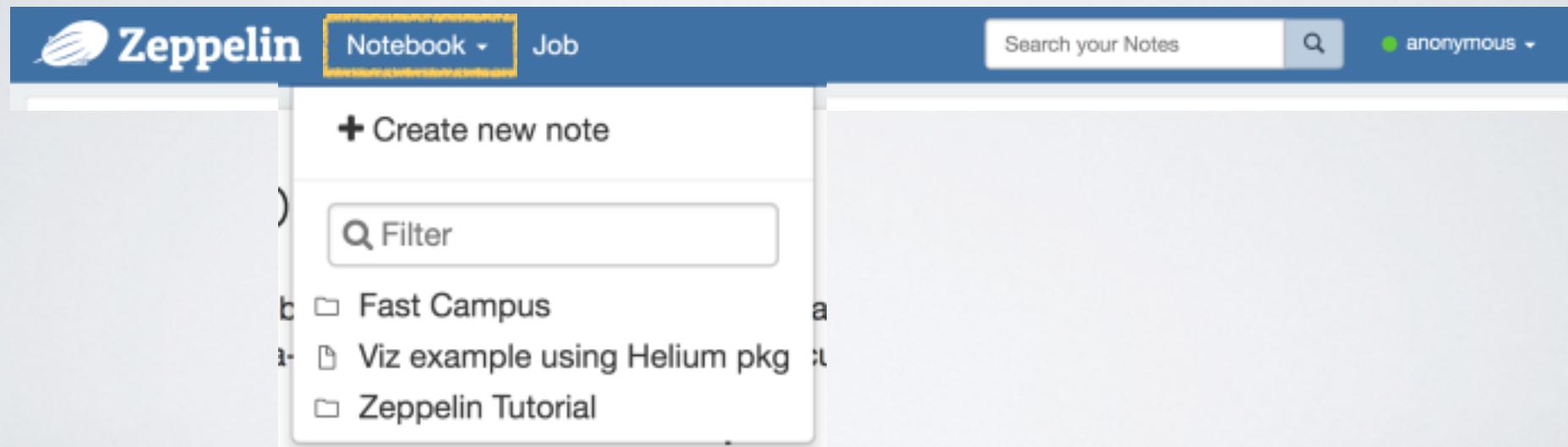
Apache Zeppelin 의 기능들, 속속들이 모두 알아보아요

Main Navbar menu



A screenshot of the Zeppelin user interface showing the main navigation bar. The bar is blue and contains the following elements from left to right: the Zeppelin logo (a stylized white airplane icon), the word "Zeppelin" in white, a dropdown menu labeled "Notebook", another dropdown menu labeled "Job", a search bar with the placeholder "Search your Notes" and a magnifying glass icon, and a user account section showing "anonymous" with a green dot next to it.

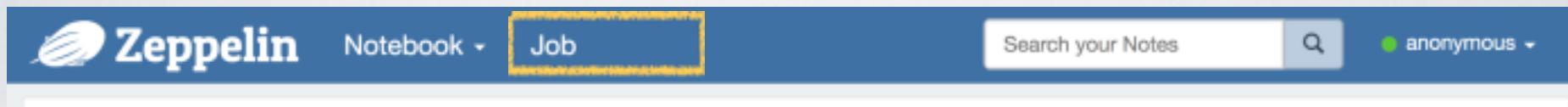
Main Navbar menu



The screenshot shows the Zeppelin Notebook interface. At the top is a blue header bar with the Zeppelin logo on the left, followed by two dropdown menus: "Notebook" and "Job". To the right of these are search and user authentication fields. A large white sidebar on the left contains a "Create new note" button, a "Filter" input field, and a list of notes. The notes listed are:

- Fast Campus
- Viz example using Helium pkg
- Zeppelin Tutorial

Main Navbar menu



Job Manager ?

localhost:3333/notebook/0A94MSU12

Zeppelin Notebook + Job

Zeppelin Tutorial/Basic Features (S...)

```
jupyter()
bank.registerTempTable("bank")
import org.apache.commons.io.IOUtils
import java.net.URL
import java.util.charset.Charset
bankFrom: org.apache.spark.sql.DataFrame = PervilleCollection[Bank] at parallize at <console>:38
bank: org.apache.spark.sql.BetaDataFrame = Dago: int, job string ... 3 more fields
warning: there was one deprecation warning; re-run with -deprecation for details
bank: org.apache.spark.sql.BetaDataFrame = Dago: int, job string ... 3 more fields
warning: there was one deprecation warning; re-run with -deprecation for details

```

Task Tree: Last updated by anonymous at April 27 2016, 11:07:58 PM.

```
spark
select age, count() value
from bank
where age > 18
group by age
order by age
```

value

Age	Count
19	~1
20	~10
21	~20
22	~30
23	~40
24	~50
25	~60
26	~70
27	~80
28	~90
29	~100

Task Tree: Last updated by anonymous at April 27 2016, 11:07:58 PM. (success)

```
spark
select age, count() value
from bank
where age > 18
group by age
order by age
```

value

Marital Status	Count
single	~45
divorced	~20
married	~15
widowed	~10
other	~5

Task Tree: Last updated by anonymous at April 27 2016, 11:07:58 PM. (success)

localhost:3333/jobmanager

Zeppelin Notebook + Job

Job

You can monitor the status of notebook and navigate to note or paragraph.

All Zeppelin Tutorial/Basic Features (S...)

READY FINISHED ABORT ERROR PENDING RUNNING

B: Zeppelin Tutorial/Basic Features (Spark) - spark 3 years ago READY D...

● ● ● ● ● ○

localhost:3333/notebook/0A94MSU12

Zeppelin Notebook + Job

Zeppelin Tutorial/Basic Features (S...)

```
jps()
bank.registerTempTable("bank")
import org.apache.commons.io.IOUtils
import java.net.URL
import java.util.Properties
bankFrom: org.apache.spark.rdd.RDD[String] = ParallelCollectionRDD at parallelize at <console>:38
bank: org.apache.spark.sql.Dataset[Bank] = Dataset[Bank] at <console>:38
warning: there was one deprecation warning; re-run with -deprecation for details
bank: org.apache.spark.sql.BetaDataFrame = Datasource: int, job, string ... 3 more fields
```

Task View. Last updated by anonymous at April 27 2018, 10:58 PM.

```
sql
select age, count() value
from bank
where age > 18
group by age
order by age
```

value

Task View. Last updated by anonymous at April 27 2018, 10:47 PM. (success)

```
sql
select age, count() value
from bank
where age > 18
group by age
order by age
```

value

Task View. Last updated by anonymous at April 27 2018, 10:47 PM. (success)

```
sql
select age, count() value
from bank
where marital='Married' and single, single and divorced married
group by age
order by age
```

localhost:3333/jobmanager

Zeppelin Notebook + Job

Job

You can monitor the status of notebook and navigate to note or paragraph.

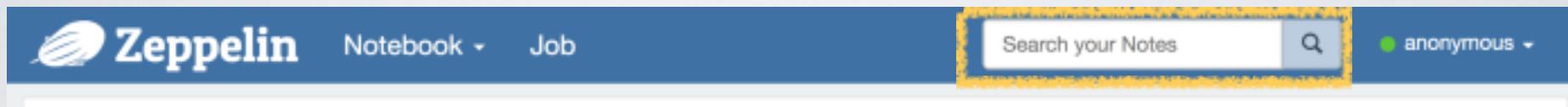
All Zeppelin Tutorial/Basic Features (S...)

READY FINISHED ABORT ERROR PENDING RUNNING

B: Zeppelin Tutorial/Basic Features (Spark) - spark 3 years ago READY

무조건 3 years ago 라고 나오는 bug 를 발견했어요...

Main Navbar menu

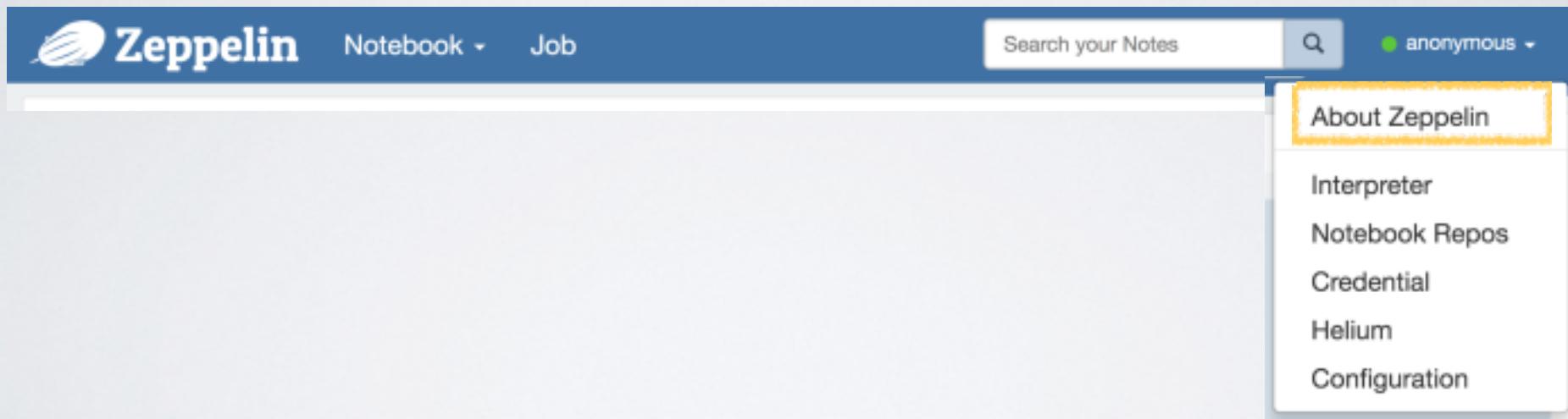


Main Navbar menu

“anonymous” 는 Zeppelin에서 Shiro authentication 을 활성화 시키지 않았을 때의 user name 이에요

The screenshot shows the Zeppelin web interface. At the top, there is a blue header bar with the Zeppelin logo on the left, followed by "Notebook" and "Job" buttons. To the right of the header is a search bar containing "Search your Notes" and a magnifying glass icon. Further to the right is a user authentication dropdown labeled "anonymous". A yellow box highlights this "anonymous" button. A white dropdown menu is open from the "anonymous" button, listing several options: "About Zeppelin", "Interpreter", "Notebook Repos", "Credential", "Helium", and "Configuration".

Main Navbar menu



The screenshot shows the Zeppelin web interface. At the top is a blue header bar with the Zeppelin logo on the left, followed by navigation links "Notebook" and "Job". To the right is a search bar containing "Search your Notes" with a magnifying glass icon, and a user status indicator "anonymous". A dropdown menu is open on the right side of the header, listing several options: "About Zeppelin" (which is highlighted with a yellow border), "Interpreter", "Notebook Repos", "Credential", "Helium", and "Configuration".

Zeppelin

Notebook Job

Search your Notes

anonymous

- About Zeppelin
- Interpreter
- Notebook Repos
- Credential
- Helium
- Configuration

About Zeppelin

The screenshot shows the Zeppelin web interface. A modal dialog box titled "About Zeppelin" is displayed in the center. The dialog contains the Apache logo, the text "Apache Zeppelin", "Version 0.7.3", "Get involved!", and "Licensed under the Apache License, Version 2.0". A yellow arrow points from the text "Version 0.7.3" to the text "Version 0.7.3" in the dialog. The background shows the Zeppelin dashboard with various navigation links and a search bar.

Welcome to Zeppelin

Zeppelin is web-based notebook that allows you to make beautiful data-driven, interactive documents.

Notebook

- Import note
- Create new note
- Filter
- Fast Campus
- Viz example using Helium pkg
- Zeppelin Tutorial

Mailing list

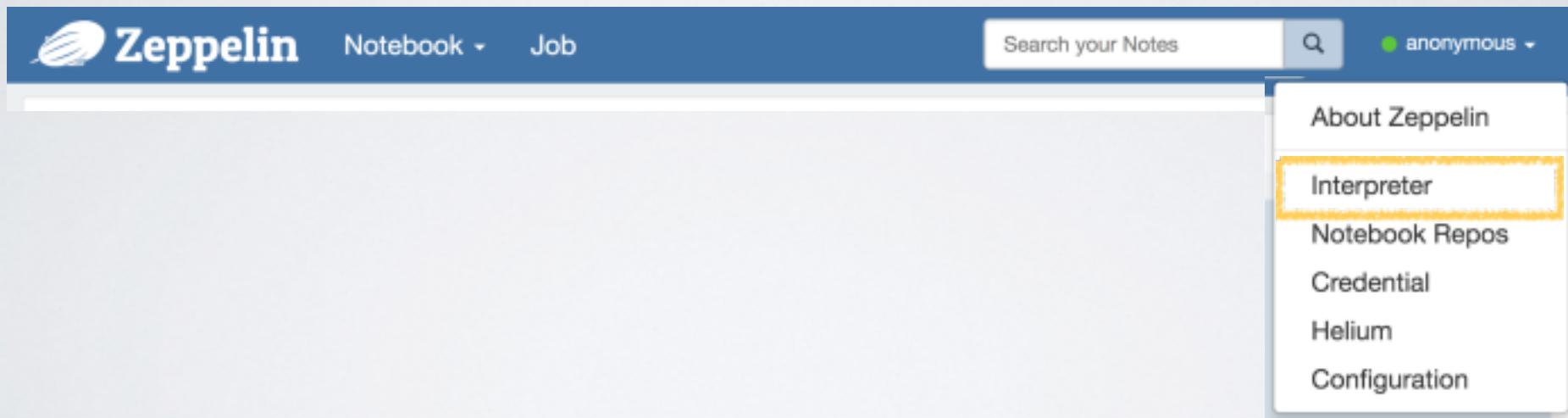
Issues tracking

Github

Search your Notes

anonymous

Main Navbar menu



The screenshot shows the Zeppelin web interface. At the top is a blue header bar with the Zeppelin logo on the left, followed by navigation links "Notebook" and "Job". To the right is a search bar containing "Search your Notes" and a magnifying glass icon. Further right is a user status indicator showing "anonymous". A dropdown menu is open on the far right, listing several options: "About Zeppelin", "Interpreter" (which is highlighted with a yellow border), "Notebook Repos", "Credential", "Helium", and "Configuration".

Zeppelin

Notebook Job

Search your Notes

anonymous

About Zeppelin

Interpreter

Notebook Repos

Credential

Helium

Configuration

Interpreter Settings (1)

The screenshot shows the Zeppelin web interface with the following details:

- Header:** Zeppelin Notebook Job Search your Notes anonymous Repository Create
- Section:** Interpreters
- Text:** Manage interpreters settings. You can create / edit / remove settings. Note can bind / unbind these interpreter settings.
- Search Bar:** Search Interpreters
- Interpreters List:**
 - alluxio** (Hadoop) [edit] [restart] [remove]
 - Option:** The interpreter will be instantiated Globally In shared process.
 - Connect to existing process
 - Set permission
 - Properties:**

name	value
alluxio.master.hostname	localhost
alluxio.master.port	19990
 - angular** (Angular) [edit] [restart] [remove]
 - Option:** The interpreter will be instantiated Globally In shared process.

- Interpreter 만들기 / 설정 변경 / 삭제
- Interpreter 검색

Interpreter Settings (2)

The screenshot shows the Zeppelin web interface under the 'Interpreters' tab. At the top, there's a search bar for notes and a user status indicator ('anonymous'). Below the header, a sub-header reads 'Manage interpreters settings. You can create / edit / remove settings. Note can bind / unbind these interpreter settings.' A search bar for interpreters is present. The main area displays two interpreters: 'alluxio' (selected) and 'angular'. The 'alluxio' row includes an 'edit' button. The 'Properties' section lists 'name' and 'value' for 'alluxio.master.hostname' (localhost) and 'alluxio.master.port' (19990). The 'angular' row also includes an 'edit' button.

Interpreter Creation

- + Create 버튼 클릭
- 기본 property 값 확인 후 “Save”
- Dependency(jars) 도 필요하다면 추가로 지정 가능 e.g JDBC interpreter에서 Mysql connector를 로딩할 때

<https://zeppelin.apache.org/docs/0.7.3/manual/interpreters.html>

Interpreter Settings (3)

The screenshot shows the Zeppelin web interface under the 'Interpreters' tab. There are two entries listed:

- alluxio**:
 - Option**: The interpreter will be instantiated **globally** - **In thread** - **process**.
Checkboxes for "Connect to existing process" and "Set permission" are present.
 - Properties**:

name	value
alluxio.mesher.hostname	localhost
alluxio.mesher.port	19998
- angular**:
 - Option**: The interpreter will be instantiated **globally** - **In thread** - **process**.
Checkboxes for "Connect to existing process" and "Set permission" are present.
 - Properties**:

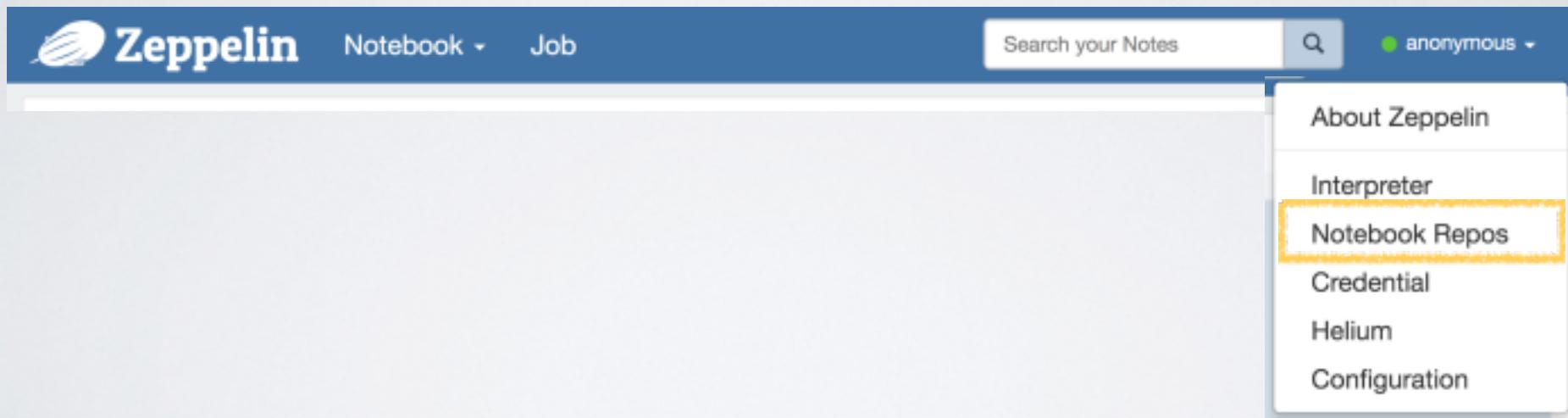
Currently there are no properties and dependencies set for this interpreter.

Repository 설정

- Central - <http://repo1.maven.org/maven2/>
- Local - `~/.m2/repository`
- Interpreter Dependency loading을 위한 maven repository 설정 등

[Add repository for dependency resolving](#)

Main Navbar menu



The screenshot shows the Zeppelin web interface. At the top is a blue header bar with the Zeppelin logo on the left, followed by navigation links "Notebook" and "Job". To the right is a search bar containing "Search your Notes" and a magnifying glass icon. Further right is a user status indicator "anonymous". A dropdown menu is open on the far right, listing several options: "About Zeppelin", "Interpreter", "Notebook Repos" (which is highlighted with a yellow border), "Credential", "Helium", and "Configuration".

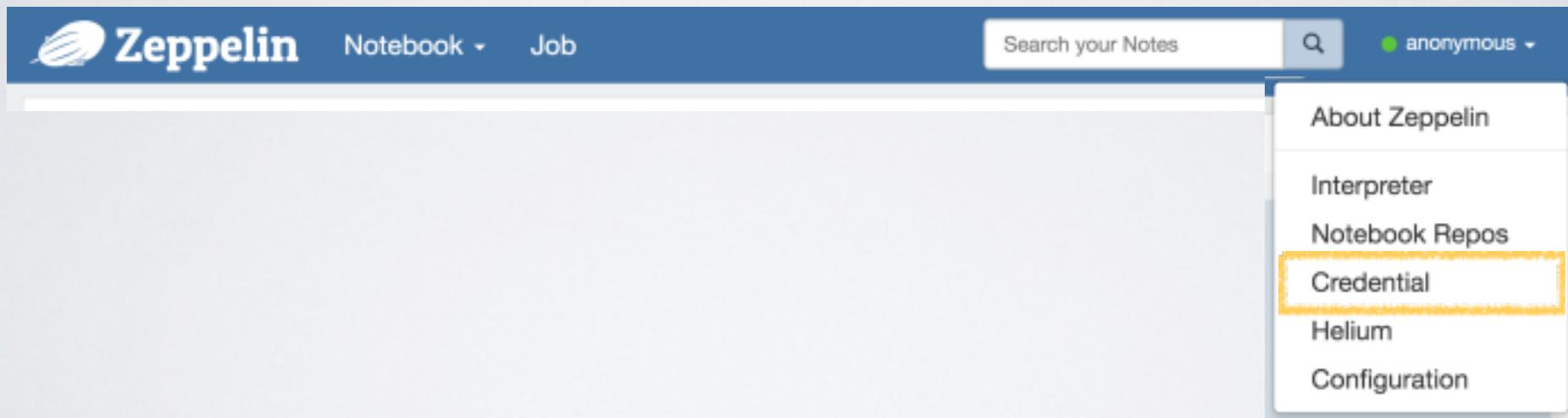
- About Zeppelin
- Interpreter
- Notebook Repos
- Credential
- Helium
- Configuration

Notebook Repository

The screenshot shows the Zeppelin interface for managing notebook repositories. At the top, there's a navigation bar with the Zeppelin logo, 'Notebook', 'Job', a search bar, and a user dropdown. Below the header, the title 'Notebook Repos' is displayed, followed by the sub-instruction 'Manage your Notebook Repositories' settings. A specific section for 'GitNotebookRepo' is highlighted with a yellow box. It contains a 'Settings' table with one row: 'name' (Notebook Path) and 'value' (/Users/ahyoungryu/Dev/zeppelin-bin/zeppelin-0.7.3-bin-all/notebook/). There's also an 'edit' button next to the table.

- Default notebook repository 는 **GitNotebookRepo**(local “o” remote “x”)로 설정되어 있어요
-> Version Control 이 가능해요
- Note들이 저장 될 위치 설정은 **conf/**아래 **zeppelin-env.sh** 혹은 **zeppelin-site.xml** 파일을 통해 변경 가능합니다
- Local Git Repo/ AWS S3/ Azure/ ZeppelinHub (a.k.a ZEPL) repository type 지원
- <https://zeppelin.apache.org/docs/0.7.3/storage/storage.html>

Main Navbar menu



The screenshot shows the Zeppelin web interface. At the top is a blue header bar with the Zeppelin logo on the left, followed by navigation links "Notebook" and "Job". To the right is a search bar containing "Search your Notes" and a magnifying glass icon. Further right is a user status indicator showing "anonymous" next to a green dot. A dropdown menu is open on the far right, listing several options: "About Zeppelin", "Interpreter", "Notebook Repos", "Credential" (which is highlighted with a yellow border), "Helium", and "Configuration".

Zeppelin

Notebook Job

Search your Notes

anonymous

- About Zeppelin
- Interpreter
- Notebook Repos
- Credential
- Helium
- Configuration

Credential Settings

The screenshot shows the Apache Zeppelin web interface with the title "Credential Settings". The top navigation bar includes links for "Notebook" and "Job", a search bar "Search your Notes", and a user dropdown "user1". Below the header, a sub-header "Credentials" is displayed with the sub-instruction "Manage your credentials. You can add new credential information." A "Add" button is located in the top right corner of the main content area.

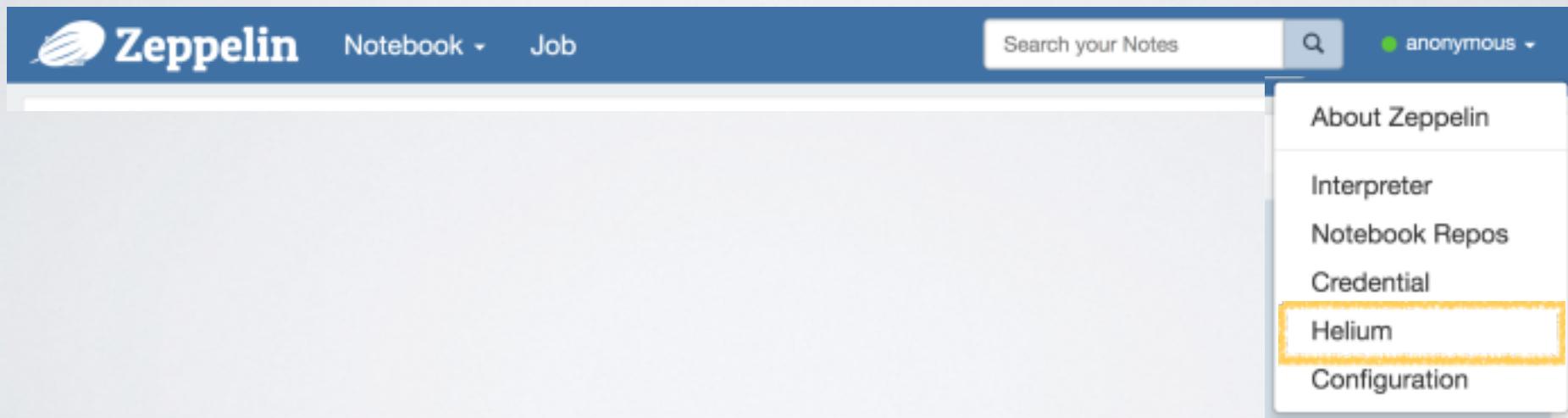
The main section is titled "Add new credential" and contains three input fields: "Entity" (set to "jdbo.mysql"), "Username" (set to "mysql_username"), and "Password" (represented by a series of asterisks). Below these fields are "Save" and "Cancel" buttons.

Below this form, a table lists existing credentials:

Entity	Username	Password	Action
jdbo.jdbc	user	*****	

- [Data Source Authorization in Apache Zeppelin](#)

Main Navbar menu

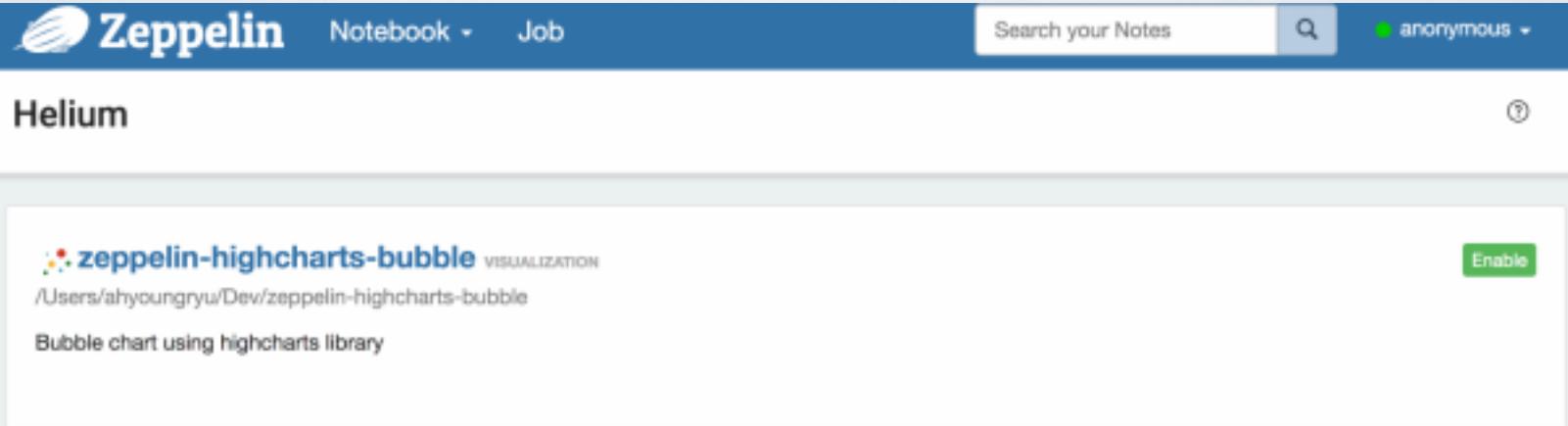


The screenshot shows the Zeppelin web interface. At the top is a blue header bar with the Zeppelin logo on the left, followed by navigation links "Notebook" and "Job". To the right is a search bar containing "Search your Notes" and a magnifying glass icon. Further right is a user status indicator showing "anonymous". A dropdown menu is open from the "Job" link, listing several options: "About Zeppelin", "Interpreter", "Notebook Repos", "Credential", "Helium" (which is highlighted with a yellow border), and "Configuration".

- About Zeppelin
- Interpreter
- Notebook Repos
- Credential
- Helium
- Configuration

Helium - Plugin System

Zeppelin 0.7.3 Version : Loading Helium packages from local Helium package registry



The screenshot shows the Zeppelin web interface. At the top, there is a blue header bar with the Zeppelin logo, a search bar labeled "Search your Notes", and a user status indicator "anonymous". Below the header, the main content area has a title "Helium". In the center, there is a card for a plugin named "zeppelin-highcharts-bubble". The card includes the plugin's name, a small icon, a "VISUALIZATION" label, the path "/Users/ahyoungryu/Dev/zeppelin-highcharts-bubble", a description "Bubble chart using highcharts library", and a green "Enable" button.

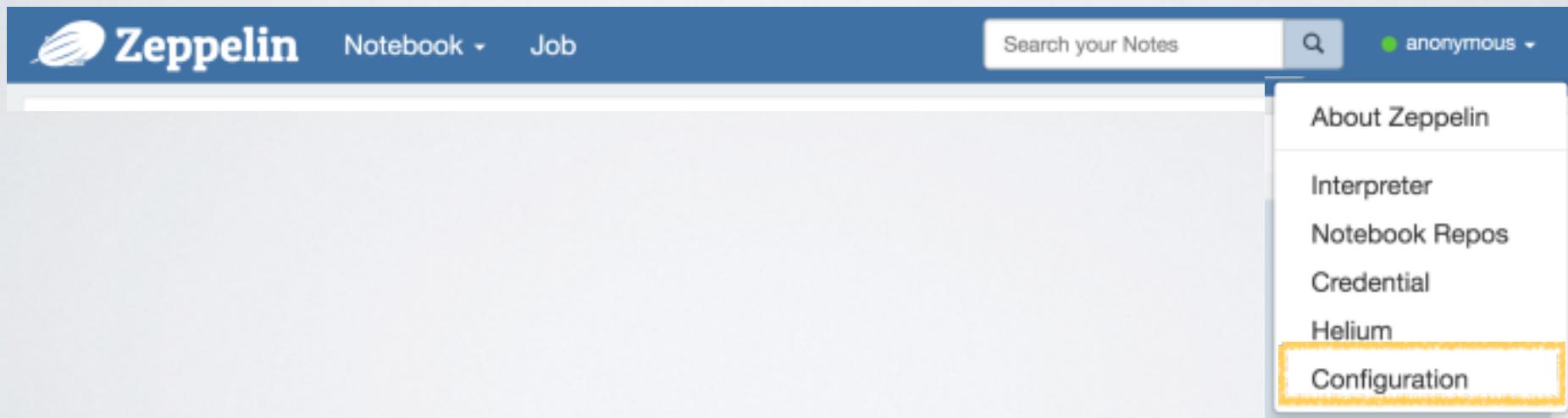
Helium - Plugin System

Zeppelin 0.8.0 Version : Loading Helium packages from online Helium package registry

The screenshot shows the Apache Zeppelin web interface with a blue header bar. The header includes the Zeppelin logo, a 'Notebook' dropdown, a search bar, and a 'anonymous' user indicator. Below the header, the page title is 'Helium'. A sub-header says 'Bundle package display order (drag and drop to reorder)'. There are two small icons: a grey one and an orange one. A note at the top right says 'Local registry package name is gray colored.' Below this, there are four listed packages:

- zeppelin-aggrid** VISUALIZATION
zeppelin-aggrid@0.1.1
Data visualization with Ag-Grid for Apache Zeppelin Enable
- zeppelin-highmaps** VISUALIZATION
zeppelin-highmaps@1.0.0
Zeppelin plugin to visualize data using Highmaps Disable
- ultimate-heatmap-chart** VISUALIZATION
ultimate-heatmap-chart@0.0.2
The Ultimate Heatmap Chart for Apache Zeppelin using highcharts Enable
- ultimate-scatter-chart** VISUALIZATION
ultimate-scatter-chart@0.0.2
The Ultimate Scatter Chart for Apache Zeppelin using highcharts Enable

Main Navbar menu



The screenshot shows the Zeppelin web interface. At the top is a blue header bar with the Zeppelin logo on the left, followed by navigation links "Notebook" and "Job". To the right is a search bar containing "Search your Notes" with a magnifying glass icon, and a user status indicator "anonymous". A dropdown menu is open on the far right, listing several options: "About Zeppelin", "Interpreter", "Notebook Repos", "Credential", "Helium", and "Configuration". The "Configuration" option is highlighted with a yellow border.

Zeppelin

Notebook Job

Search your Notes

anonymous

- About Zeppelin
- Interpreter
- Notebook Repos
- Credential
- Helium
- Configuration

Configuration

```
ahyoungryu at MacBook-Pro-6 in ~/Dev/zeppelin-bin/zeppelin-0.7.3-bin-all
$ ll
total 56056
-rw-r--r--@ 1 ahyoungryu staff 59K Sep 18 2017 LICENSE
-rw-r--r--@ 1 ahyoungryu staff 5.5K Sep 18 2017 NOTICE
-rw-r--r--@ 1 ahyoungryu staff 1.3K Sep 18 2017 README.md
drwxr-xr-x@ 12 ahyoungryu staff 4888 Sep 18 2017 bin
drwxr-xr-x@ 14 ahyoungryu staff 4768 Apr 22 23:41 conf
drwxr-xr-x  5 ahyoungryu staff 1708 Apr 23 01:06 data
-rw-r--r--  1 ahyoungryu staff 8258 Apr 23 08:08 derby.log
drwxr-xr-x  3 ahyoungryu staff 1028 Oct 13 2017 helium
drwxr-xr-x@ 23 ahyoungryu staff 7828 Sep 18 2017 interpreter
drwxr-xr-x@ 131 ahyoungryu staff 4.3K Sep 18 2017 lib
drwxr-xr-x@ 61 ahyoungryu staff 2.0K Sep 18 2017 licenses
drwxr-xr-x  3 ahyoungryu staff 1028 Oct 13 2017 local-repo
drwxr-xr-x  14 ahyoungryu staff 4768 Apr 23 01:09 logs
drwxr-xr-x  9 ahyoungryu staff 3868 Apr 23 08:08 metastore_db
drwxr-xr-x@ 16 ahyoungryu staff 5448 Apr 23 01:15 notebook
drwxr-xr-x  3 ahyoungryu staff 1028 Apr 23 03:29 run
drwxr-xr-x  3 ahyoungryu staff 1028 Apr 23 03:29 webapps
-rw-r--r--@ 1 ahyoungryu staff 27M Sep 18 2017 zeppelin-web-0.7.3.war
```

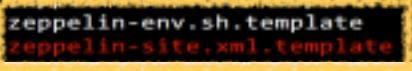


Zeppelin 설치 디렉토리 안에 있는 **conf/** 디렉토리 안에 .xml 혹은 .sh 확장자의 설정파일들이 들어있어요

Configuration

conf/ 디렉토리 아래에서

```
ahyoungryu at MacBook-Pro-6 in ~/Dev/zeppelin-bin/zeppelin-0.7.3-bin-all/conf
$ ll
total 168
-rw-r--r--@ 1 ahyoungryu  staff   1.3K Sep 18  2017 configuration.xlsx
-rw-r--r--  1 ahyoungryu  staff  343B Oct 13  2017 helium.json
-rw-r--r--@ 1 ahyoungryu  staff   2.7K Sep 18  2017 interpreter-list
-rw-r----- 1 ahyoungryu  staff   26K Apr 23 03:42 interpreter.json
-rw-r--r--@ 1 ahyoungryu  staff   1.3K Sep 18  2017 log4j.properties
-rw-r--r--@ 1 ahyoungryu  staff   20B Apr 23 03:42 notebook-authorization.json
-rw-r--r--@ 1 ahyoungryu  staff   3.8K Sep 18  2017 shiro.ini.template
-rw-r--r--@ 1 ahyoungryu  staff   5.2K Sep 18  2017 zeppelin-env.cmd.template
-rw-r--r--@ 1 ahyoungryu  staff   6.2K Sep 18  2017 zeppelin-env.sh.template
-rwxr-xr-x@ 1 ahyoungryu  staff   14K Sep 18  2017 zeppelin-site.xml.template
```

A yellow arrow points from the bottom right towards the last two files in the terminal output: "zeppelin-env.sh.template" and "zeppelin-site.xml.template". Both of these files are highlighted with a red rectangular box.

설정 파일 활성화하기

- cp zeppelin-env.sh.template zeppelin-env.sh
- cp zeppelin-site.xml.template zeppelin-site.xml
- <http://zeppelin.apache.org/docs/0.7.3/install/configuration.html>

Configuration

설정값 변경 후 Zeppelin을 재시작하면 여기 반영된 것을 확인할 수 있어요

The screenshot shows the Zeppelin web interface with the title "Configurations". It displays the current configurations for the Zeppelin Server. A note states that for security reasons, some key/value pairs including passwords would not be shown. The table lists the following configurations:

name	value
zeppelin.anonymous.allowed	true
zeppelin.conf.dir	/Users/shyoungryu/Dew/zeppelin-bin/zeppelin-0.7.3-bin-all/conf
zeppelin.credentials.persist	true
zeppelin.dep.localrepo	local-repo
zeppelin.encoding	UTF-8
zeppelin.helium.localregistry.default	helium
zeppelin.helium.npm.registry	http://registry.npmjs.org/
zeppelin.home	/Users/shyoungryu/Dew/zeppelin-bin/zeppelin-0.7.3-bin-all
zeppelin.interpreter.connect.timeout	30000
zeppelin.interpreter.dep.mavenRepo	http://repo1.maven.org/maven2/

Note Creation

 **Zeppelin** Notebook • Job Search your Notes  anonymous •

Welcome to Zeppelin!

Zeppelin is web-based notebook that enables interactive data analytics.
You can make beautiful data-driven, interactive, collaborative document with SQL, code and even more!

Notebook 

-  Import note
-  Create new note
-
-  Mahout Tutorial
-  R Tutorial
-  Zeppelin Tutorial
-  Zeppelin Tutorial: Python - matplotlib basic

Help 

Get started with [Zeppelin documentation](#)

Community

Please feel free to help us to improve Zeppelin,
Any contribution are welcome!

-  Mailing list
-  Issues tracking
-  Github

Zeppelin Home에 있는 이곳에서도 만들 수 있고



Note Creation

The screenshot shows the Zeppelin web interface. At the top, there is a blue header bar with the Zeppelin logo on the left, followed by 'Notebook - Job' and a search bar 'Search your Notes' with a magnifying glass icon. On the far right, it says 'anonymous' with a dropdown arrow. A yellow arrow points from the text above to the 'Notebook' button in the header.

Navbar 의 "Notebook" 메뉴에서도 바로 만들 수 있어요

Welcome to Zeppelin!

Zeppelin is web-based notebook. You can make beautiful data visualization and reports.

Notebook

- Import note
- Create new note

Filter

+ Create new note

Filter

- Mahout Tutorial
- R Tutorial
- Zeppelin Tutorial
- Zeppelin Tutorial: Python - matplotlib basic

SQL, code and even more!

[View documentation](#)

Community

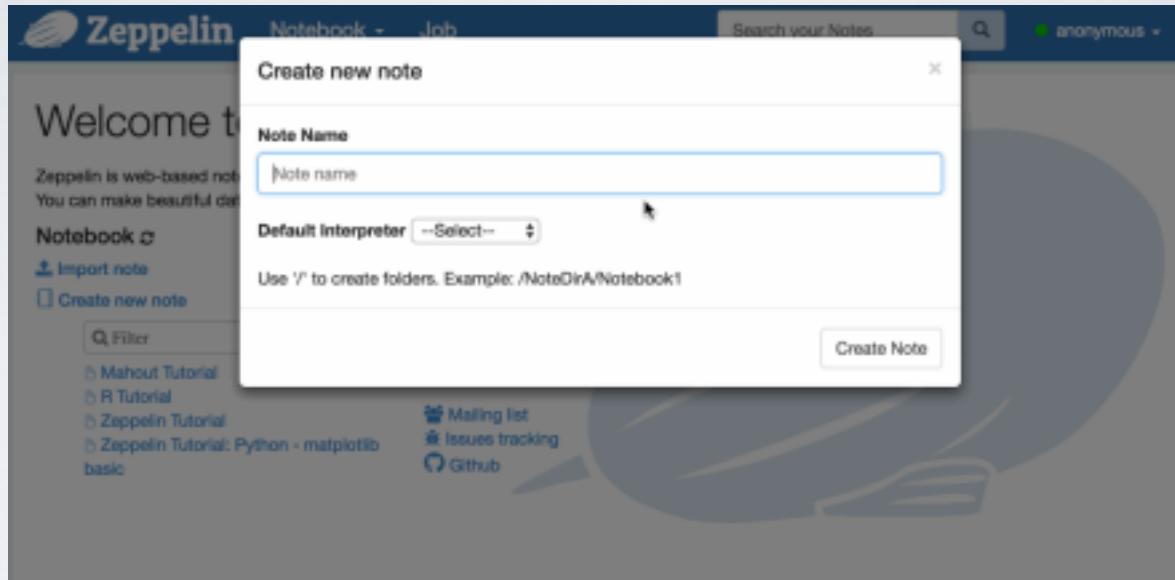
Please feel free to help us to improve Zeppelin. Any contribution are welcome!

Mailing list

Issues tracking

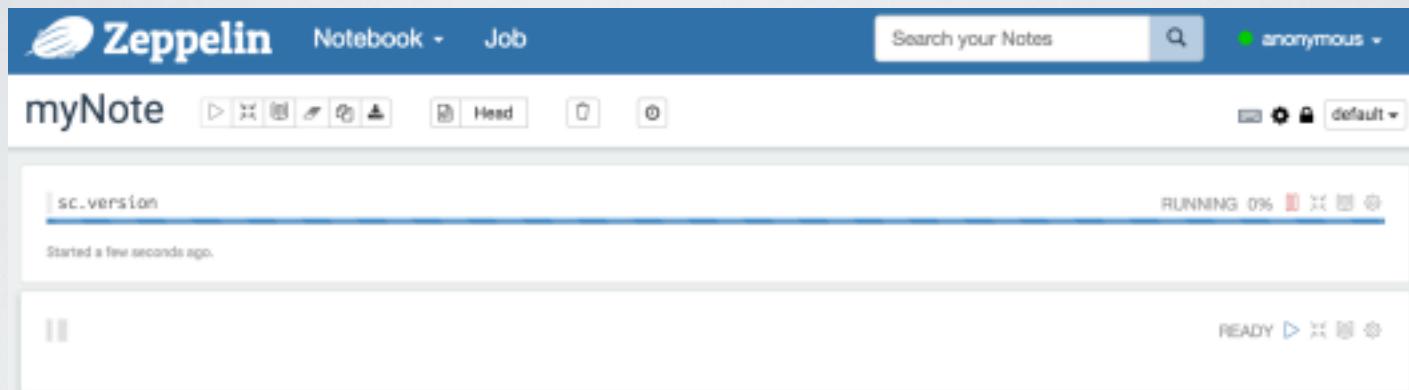
Github

Note Creation



- Default interpreter 설정 "%interpreter_name" 을 paragraph editor에 쓰지 않아도 되요
- `/NoteDirName/NotebookName` 형식으로 디렉토리 구조도 쉽게 생성 가능합니다

Paragraph



- Note 하나는 여러 개의 Paragraph들로 이루어져 있어요
- Paragraph의 editor 안에 code 를 쓰고 Shift + Enter 혹은 Play 버튼을 눌러 paragraph 실행
- 실행되는 도중에 Progress bar를 통해 얼마나 실행되었는지 확인할 수 있어요

Paragraph

The screenshot shows a Zeppelin Notebook interface. At the top, there's a blue header bar with the Zeppelin logo, a search bar labeled "Search your Notes", and a status indicator "anonymous". Below the header, the notebook title is "myNote". The main content area displays a paragraph of text in Korean, which translates to: "Paragraph 실행이 끝나면 결과값과 함께 경과시간/실행시킨 user name/ 최근 update 시간 등을 확인할 수 있어요". Above this text, there are two code snippets: "sc.version" and "res0: String = 2.0.1". At the bottom left, it says "Took 19 sec. Last updated by anonymous at November 12 2016, 10:43:26 PM." The bottom right shows a "READY" status with navigation icons.

Paragraph

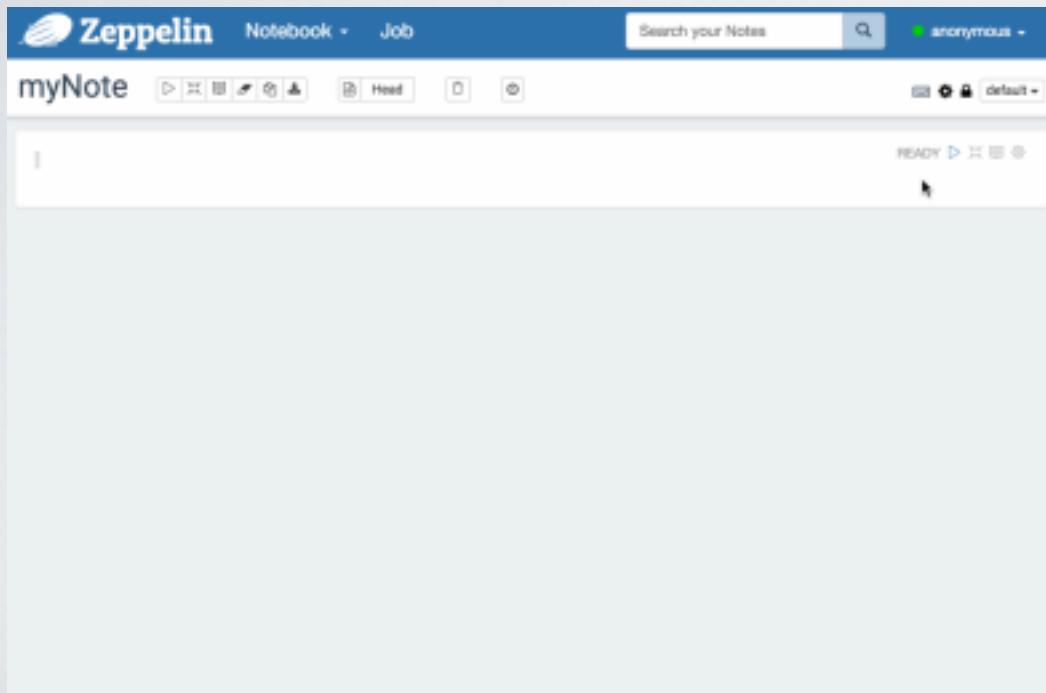
The screenshot shows the Zeppelin web interface. At the top, there's a navigation bar with the Zeppelin logo, 'Notebook - Job', a search bar 'Search your Notes', and a user status 'anonymous'. Below the header, the title 'myNote' is displayed, along with various toolbar icons for document operations like copy, paste, bold, etc. The main content area contains a code cell:

```
| sc.version  
res0: String = 2.0.1
```

Below the code, a note says 'Took 19 sec. Last updated by anonymous at November 12 2016, 10:43:26 PM.' To the right of the code cell, it says 'FINISHED' with icons for refresh, copy, and delete. The bottom section is a blank paragraph editor with a cursor, labeled 'READY' with similar icons.

Default interpreter로 설정된 interpreter 이외에 다른 interpreter를 사용하고
싶다면 `%interpreter_name`을 paragraph editor 최상단에 적어주면 됩니다

Interpreter Binding



- Default interpreter 변경
- Activate / Deactivate
- 실행중이던 interpreter process 를 kill하고 restart 가능

Interpreter Binding

The screenshot shows a Jupyter Notebook interface with three code cells:

- spark:**

```
// load data  
// create bank dataframe  
// ...  
bank.registerTempTable("bank")
```
- spark.sql:**

```
select *  
from bank  
limit 1000
```
- spark.pyspark:**

```
# visualize bank table using python library (ggplot, ...)
```

Each cell has a yellow header bar. The first cell is labeled "READY", the second "FINISHED", and the third "READY". Below the cells, a status bar indicates "Task 0 sec. Last updated by anonymous at October 21 2017, 8:08 PM. [unlisted]".

Interpreter Group

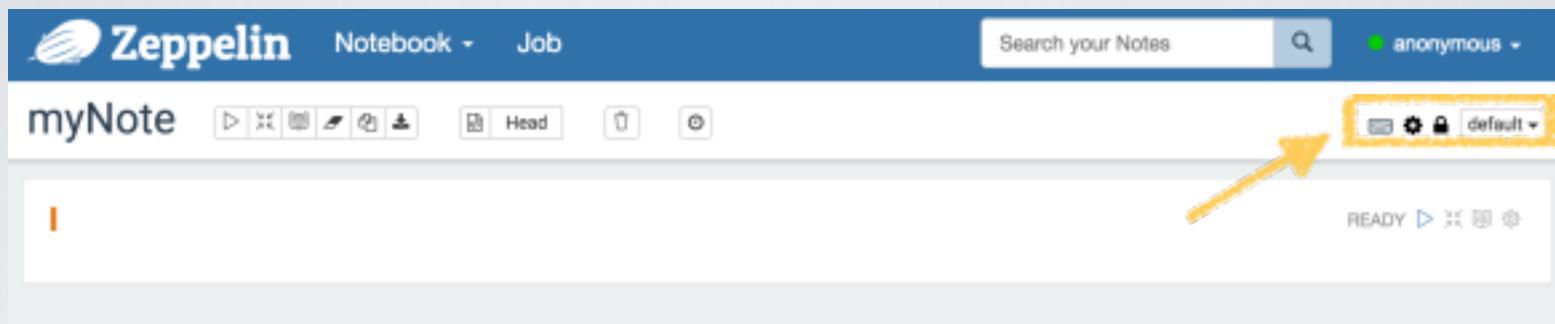
- Interpreter Group 은 관련된 Interpreter 여러개를 포함할 수 있어요
- E.g. Spark Interpreter Group 은 Spark, PySpark, SparkR, SparkSQL 를 포함
- E.g. Python Interpreter Group 은 Python, Ipython, Pandas, Docker, Conda 를 포함

Note Control Menu

The screenshot shows the Zeppelin web interface. At the top, there's a blue header bar with the Zeppelin logo, a 'Notebook' dropdown, a 'Job' link, a search bar ('Search your Notes'), and a user status ('anonymous'). Below the header is a toolbar with several icons: a magnifying glass, a trash can, a file folder, a plus sign, a 'Head' icon, and three other icons. A yellow arrow points to the third icon from the left in this toolbar. The main area is titled 'myNote' and contains a single paragraph with the letter 'I'. At the bottom right of this area, there are status indicators: 'READY', a blue triangle, and some small icons.

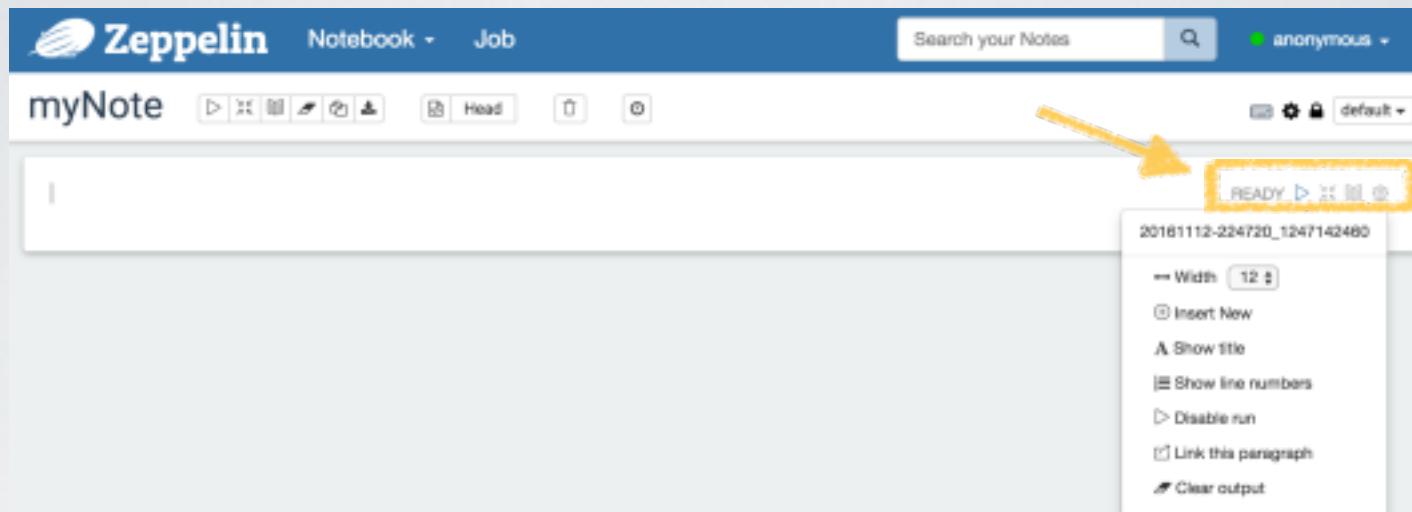
- 전체 Paragraph 실행
- 전체 Paragraph code editor 숨기기/보이기
- 전체 Paragraph result output 숨기기/보이기
- 전체 Paragraph result clear
- Clone
- Export
- Version Control
- Delete
- Cron expression 으로 Note auto-running 스케줄 설정

Note Control Menu



- Keyboard shortcut
- Interpreter binding
- Note 권한 설정
- Note LookNFeel(default/ simple/ report 모드) 설정

Paragraph Control Menu



- Paragraph 실행 결과 확인 (READY/ FINISHED/ ERROR/ PENDING/ RUNNING)
- Run ▶ Stop ▶
- Code 숨기기 / 보이기
- Output 숨기기/ 보이기
- Paragraph settings

Hands-on 1.

Notebook Version Control

Via local Git repository setting

Hands-on 1.

Notebook Version Control

Via local Git repository setting

만약 같이 일하는 팀원이 실수로 내가 만들여 만든 paragraph 하나를 날려버렸다면?

Hands-on 1.

Notebook Version Control

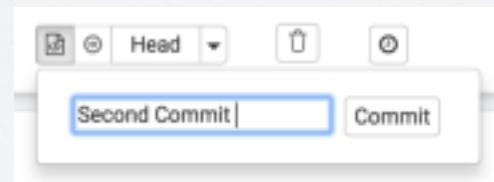
Via local Git repository setting

만약 같이 일하는 팀원이 실수로 내가 공들여 만든 paragraph 하나를 날려버렸다면?
어떻게 하면 이런 불상사를 막을 수 있을까요? 각 Note에 Version Control 을 할 수 있다면?

Notebook Version Control

Via local Git repository setting

“Version Control” click-> commit msg 를 적은 후 현재 note 상태를 history 로 남길 수 있어
요



Notebook Version Control

Via local Git repository setting

“Commit history 리스트중 원하는 commit 을 선택하여 preview 를 볼 수 있어요. 즉 preview mode 이기 때문에, 이 상태에선 아무것도 할 수 없답니다. Preview 로 본 상태가 내가 돌아가고 싶은 상태라면 “Set revision” 으로 해당 commit 을 Head 로 만들어 history 를 rollback 할 수 있어요.



Hands-on session 1.

직접 note history 를 만들어보세요

사용하기 불편한 점 없었나요?

어디에 이 정보가 저장되고 있는
거죠?

The screenshot shows the Zeppelin interface with the following elements:

- Header:** Zeppelin logo, Notebook >, Job, Search your Notes (with a magnifying glass icon), anonymous.
- Section:** Notebook Repos
- Description:** Manage your Notebook Repositories' settings.
- Table:** GitNotebookRepo settings

Settings	
name	value
Notebook Path	/Users/shyoungryu/Dew/zeppelin-bin/zeppelin-0.7.3-bin-all/notebook/

- Buttons:** edit (next to the table)

“Notebook Repos” menu => 기본 notebook repository setting 값을 확인해 볼 수 있어요

```
ahyoungryu at MacBook-Pro-6 in ~/Dev/zeppelin-bin/zeppelin-0.7.3-bin-all/notebook on master*
$ ls -al
total 0
drwxr-xr-x@ 13 ahyoungryu staff 442 Apr 26 22:48 .
drwxr-xr-x@ 20 ahyoungryu staff 680 Apr 28 02:13 ..
drwxr-xr-x  12 ahyoungryu staff 408 Apr 28 02:15 .git
drwxr-xr-x@  3 ahyoungryu staff 102 Apr 27 19:09 2A94M5J1Z
drwxr-xr-x@  3 ahyoungryu staff 102 Sep 18 2017 2BWJFTXKJ
drwxr-xr-x@  3 ahyoungryu staff 102 Sep 18 2017 2BYEZ5EVK
```

해당 위치에 가보면 여타 git으로 version 관리를 하는 프로젝트와 같이 **.git** 디렉토리가 생긴것을 확인할 수 있어요.

Hands-on session 2.

Change Notebook Repository

AWS S3 를 notebook repo로 사용해보기

Hands-on session 2.

Change Notebook Repository

AWS S3 를 notebook repo로 사용해보기

팀원들하고 협업을 해야하는 상황에서 notebook을 공유해야 한다면?

Hands-on session 2.

Change Notebook Repository

AWS S3 를 notebook repo로 사용해보기

팀원들하고 협업을 해야하는 상황에서 notebook을 공유해야 한다면?

이런 경우 내 local 보단 remote repository 에 notebook을 저장해야 하는게 좋겠
죠?

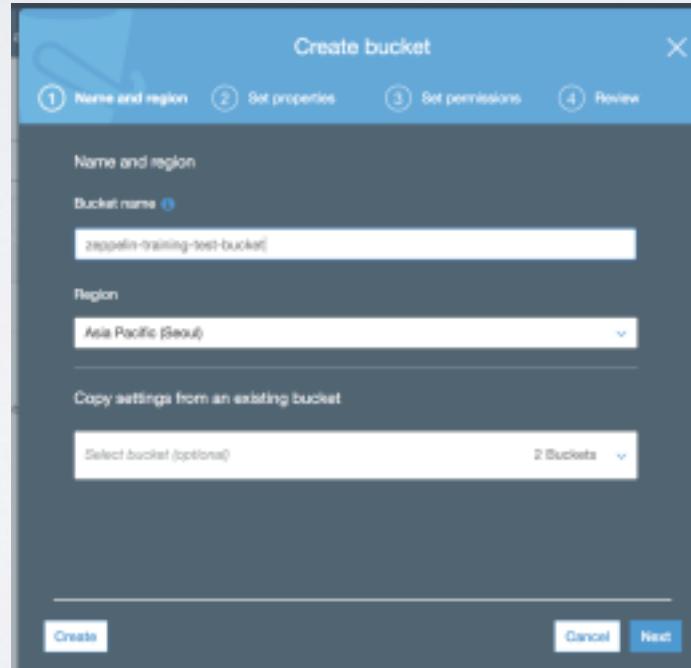
AWS S3 를 notebook repo로 사용해보기

- AWS 계정이 없다면 ?(시간 단축을 위해 제 계정을 알려드릴게요)
- 계정이 있다면 AIM에서 `AWS_ACCESS_KEY_ID` / `AWS_SECRET_ACCESS_KEY`를 생성 해주세요
- Zeppelin 위 credential 값을 알아내는데는 아래 순서를 따른답니다

- The `AWS_ACCESS_KEY_ID` and `AWS_SECRET_ACCESS_KEY` environment variables
- The `aws.accessKeyId` and `aws.secretKey` Java System properties
- Credential profiles file at the default location (`~/.aws/credentials`) used by the AWS CLI
- Instance profile credentials delivered through the Amazon EC2 metadata service

AWS S3 를 notebook repo로 사용해보기

AWS S3 service menu로 이동해서 bucket 을 하나 만들어 주세요



[ge/storage.html#notebook-storage-in-s3](#)

AWS S3 를 notebook repo로 사용해보기

- conf/zeppelin-site.xml 파일 수정
- 아래 부분을 찾아 uncomment 해주세요 => S3 로 새로운 NotebookRepo 간 설정
- 앞에서 만든 bucket 이름/ user name 을 각각 넣어 주세요

```
<!-- Amazon S3 notebook storage -->
<!-- Creates the following directory structure: s3://{bucket}/{username}/{notebook-id}/note.json -->
<property>
  <name>zeppelin.notebook.s3.user</name>
  <value>ahyung</value>
  <description>user name for s3 folder structure</description>
</property>

<property>
  <name>zeppelin.notebook.s3.bucket</name>
  <value>zeppelin-training-test</value>
  <description>bucket name for notebook storage</description>
</property>

<property>
  <name>zeppelin.notebook.s3.endpoint</name>
  <value>s3.amazonaws.com</value>
  <description>endpoint for s3 bucket</description>
</property>

<property>
  <name>zeppelin.notebook.storage</name>
  <value>org.apache.zeppelin.notebook.repo.S3NotebookRepo</value>
  <description>notebook persistence layer implementation</description>
</property>
```

<https://zeppelin.apache.org/docs/0.7.3/storage/storage.html#notebook-storage-in-s3>

AWS S3 를 notebook repo로 사용해보기

- conf/zeppelin-site.xml 파일 수정
- 아래 부분은 comment out 처리 해주세요 => 기존 Git 으로 설정되었던 NotebookRepo 값 비활성화

```
<!-- <property>
    <name>zeppelin.notebook.storage</name>
    <value>org.apache.zeppelin.notebook.repo.GitNotebookRepo</value>
    <description>versioned notebook persistence layer implementation</description>
</property> -->
```

AWS S3 를 notebook repo로 사용해보기

- `./bin/zeppelin-daemon.sh restart` 로 Zeppelin 재시작
- 설정값이 GitNotebookRepo -> S3NotebookRepo 로 제대로 바뀌었는지 확인 해보세요

The screenshot shows the Zeppelin web interface. At the top, there is a blue header bar with the Zeppelin logo, a search bar labeled "Search your Notes", and a user dropdown set to "anonymous". On the left, there is a sidebar with links: "About Zeppelin", "Interpreter", "Notebook Repos", "Credential", "Helium", and "Configuration". The main content area has a title "Notebook Repos" and a sub-section titled "S3NotebookRepo" which states "Currently there are no settings for this Notebook Repository".

AWS S3 를 notebook repo로 사용해보기

Note 를 새로 만든 후 AWS S3 bucket 으로 이동해서 해당 note 가 bucket 아래에 제대로 저장되고 있는지 확인!

The screenshot shows the AWS S3 console interface. At the top, there's a navigation bar with the AWS logo, 'Services' dropdown, 'Resource Groups' dropdown, a user icon for 'Ahyoung Ryu', and 'Global' and 'Support' links. Below the navigation bar, the URL path is displayed as 'Amazon S3 > zeppelin-training-test / ahyoung / notebook / 2DC39C0K2'. The main area is titled 'Overview' and contains a search bar with the placeholder 'Type a prefix and press Enter to search. Press ESC to clear.' Below the search bar are three buttons: 'Upload', '+ Create folder', and 'More'. To the right of these buttons, it shows the region as 'Asia Pacific (Seoul)' with a dropdown arrow. A table follows, showing one item: 'note.json'. The table has columns for 'Name' (with a checkbox), 'Last modified', 'Size', and 'Storage class'. The 'note.json' entry shows 'note.json' in the Name column, 'Apr 28, 2018 3:43:09 AM GMT+0900' in the Last modified column, '2.3 KB' in the Size column, and 'Standard' in the Storage class column. At the bottom of the table, it says 'Viewing 1 to 1'.

Name	Last modified	Size	Storage class
note.json	Apr 28, 2018 3:43:09 AM GMT+0900	2.3 KB	Standard

Activate Shiro Authentication

No more “anonymous”!

Shiro authentication in Zeppelin



- [Apache Shiro](#) is a powerful and easy-to-use Java security framework that performs authentication, authorization, cryptography, and session management.
- Zeppelin 내에서 notebook / interpreter ACL은 Shiro authentication 을 활성화 시켰을때 비로소 사용 가능해요

Shiro authentication in Zeppelin

- conf/ 디렉토리 아래에서 `cp shiro.ini.template shiro.ini` 를 실행 한 후
`shiro.ini`에서 login 에 필요한 user 정보 및 role 을 확인해보세요

```
[users]
# List of users with their password allowed to access Zeppelin.
# To use a different strategy (LDAP / Database / ...) check the shiro doc at http://shiro.apache.org/configuration.html#Configuration-INISections
admin = password1, admin
user1 = password2, role1, role2
user2 = password3, role3
user3 = password4, role2
```

```
[roles]
role1 = *
role2 = *
role3 = *
admin = *
```

Shiro authentication in Zeppelin

- 어떤 user 는 security 문제로 특정 페이지로의 접근을 막아야 하는 경우가 있어요.
- E.g. Configuration 혹은 Credentials 페이지

```
[urls]
# This section is used for url-based security.
# You can secure interpreter, configuration and credential information by urls. Comment or uncomment the below urls that you want to hide.
# anon means the access is anonymous.
# authc means Form based Auth Security
# To enforce security, comment the line below and uncomment the next one
/api/version = anon
#/api/interpreter/** = authc, roles[admin]      URL based security!
#/api/configurations/** = authc, roles[admin]
#/api/credential/** = authc, roles[admin]
#/** = anon
/** = authc
```

Shiro authentication in Zeppelin

- 어떤 user 는 security 문제로 특정 페이지로의 접근을 막아야 하는 경우가 있어요.
- E.g. Configuration 혹은 Credentials 페이지

```
[urls]
# This section is used for url-based security.
# You can secure interpreter, configuration and credential information by urls. Comment or uncomment the below urls that you want to hide.
# anon means the access is anonymous.
# authc means Form based Auth Security
# To enforce security, comment the line below and uncomment the next one
/api/version = anon
#/api/interpreter/** = authc, roles[admin]      URL based security!
#/api/configurations/** = authc, roles[admin]
#/api/credential/** = authc, roles[admin]
#/** = anon
/** = authc
```

이 URL 은 어디에서 오는거죠? 바로 Rest api
의 endpoint입니다.

Hands-on 3.

Admin 이외 다른 user 들의 아래 페이지로의 접근을 막아주세요

- Credential
- Notebook Repository

알쓸신잡 : Open Source 인 Apache Zeppelin은 어떻게 관리되고 있을까요?

질문 있나요 ?

3시간 동안 수고 많으셨습니다 :)