

Little Big Data

ZEPL[®]

Apache Zeppelin을
띠어넘자

2018.10.19
류아영

저는 ZEPL에서 왔어요



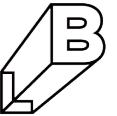
류아영

- ZEPL frontend developer
- Apache Zeppelin PMC member
- <https://www.linkedin.com/in/ahyoungryu/>

ZEPL

- Apache Zeppelin creator
- Seoul, Korea & Palo Alto
- Main service - ZEPL <http://www.zepl.com>

“Apache Zeppelin 을 등에 업고 ZEPL project (Zeppelin SaaS version)를 시작하게 되었지만, 아이러니하게도 Zeppelin의 기능을 뛰어 넘어야 하는 것이 가장 먼저 넘어야 할 산이 되었던 상황과 이를 위해 저희 ZEPL 팀에서 어떤 노력들을 했는지에 대한 경험을 공유하고자 합니다.”



Apache Zeppelin

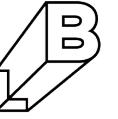
The screenshot shows the Apache Zeppelin Notebook interface with several cells:

- Bank**: Contains three charts: a pie chart of age distribution (e.g., 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19), a bar chart of age under 35 (e.g., 19, 22, 26, 30, 34), and a line chart of marital status over time.
- Sample Zeppelin note accessing Neo4j database**: Contains a sample note with a query and a histogram showing the distribution of actors per movie (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 12).

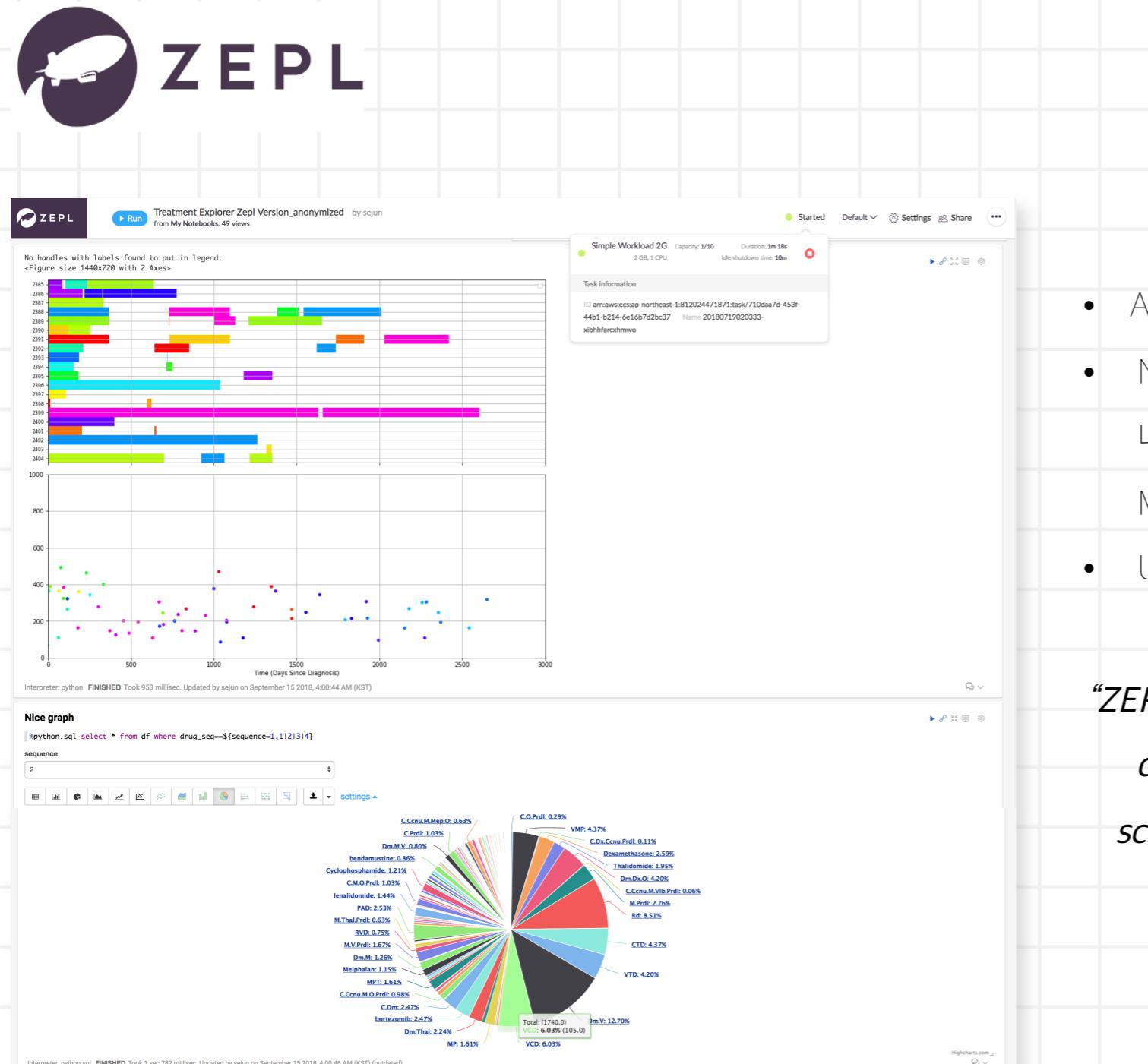
- 데이터 로딩, 정제, 분석, 시각화, 분석 결과 공유 등 모든 업무를 한 곳에서 할 수 있는 웹 기반 노트북 서비스
- Apache Software Foundation 의 top level project 중 하나
- Apache Spark 이 한창 인기를 얻기 시작할 그 초창기 무렵 Spark의 인기를 등에 업고 같이 빠르게 유명해짐
- Apache Spark / JDBC / Python / Google BigQuery / Shell / Apache Flink / Angular 등 20가지가 넘는 인터프리터들이 지원되고 있음

2018년 10월, 프로젝트의 개발 진행 상황

ZEPL 개발자들 (PMC 대부분이 ZEPL에 있(었)음)이 메인으로 리딩을 하고 있진 않지만 외부 PMC 들(Hortonworks / Microsoft 등)에 의해 릴리즈, 코드 리뷰가 진행되고 있음. 아직 ver 1.0.0도 나오기 전이지만 진행 속도가 느리긴 함..

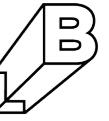


ZEPL, Apache Zeppelin을 뛰어넘자 - 오늘의 (진짜) 주인공 #2



- Apache Zeppelin 의 SaaS version
- Notebook 별 세분화된 유저 권한 설정 / Organization
내 team member 들의 role 설정을 통한 권한 제어 등
- Multi user environment 에 좀 더 적합한 환경
- UI 를 통한 notebook running 환경 제어 기능 제공

"ZEPL helps businesses and organizations make smart, data-driven decisions by bringing together data science teams to work in a powerful, collaborative workspace."



지피지기, 백전백승

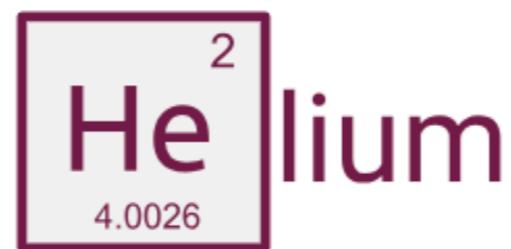
이래요

Apache Zeppelin의 장점부터 알아보죠

1. 20개 넘는 인터프리터 지원



2. Customizable Visualization (투박하고 한정된 nvd3 chart 말고)



3. 인기 오픈소스 프로젝트 (feat. 공짜)



지피지기, 백전백승

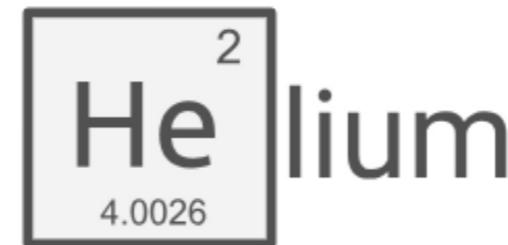
연 아래요

~~Apache Zeppelin의 장점부터 알아보죠..~~ ZEPL 에겐 뛰어넘어야 할 점이에요

1. 20개 넘는 인터프리터 지원



2. Customizable Visualization (투박하고 한정된 nvd3 chart 말고)



3. 인기 오픈소스 프로젝트 (feat. 공짜)



돈받고 서비스하는데 검증 안된

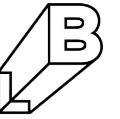
License 문제는 ?

인터프리터들까지

Helium Online Repository 에 있는
package 들 다 믿을만 할까?

모두다 지원 해야할까 ?

돈주고 쓸만한 기능들 ? 많습니다



지피지기, 백전백승 이래요

Zeppelin은 이런 한계점도 있긴하지만,
ZEPL에겐 오히려 보완해서 Selling Point로 만들 수 있는 점이기도 해요

1. Multi-user 환경에서의 한계



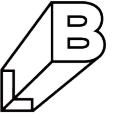
노트북별로 Read / Write / Run 권한 설정 가능하던데?

Zeppelin에 Apache Shiro 들어가 있던데 ?

2. 안정성 문제



Zeppelin 서버 또 죽었네 재시작 해야겠다.
뭐지 ? 아깐 됐는데 왜 지금은 안되는거지 ?
아까는 이랬는데 왜 지금은 반응이 다른거지 ?

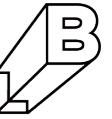


그래서 ZEPL 에서는...

#1. Multi-user Environment Support

The screenshot displays the Apache Zeppelin web interface. On the left, there's a 'Setup your organization - Start up guide' section with a 'Get Started with Team Plan' button and a 'What is Team Plan?' section listing benefits like creating teams and setting user roles. In the center, there's a 'Members' list for the 'ZEPL' organization, showing various team members. On the right, a 'Sharing and Permissions' dialog is open for a specific notebook, showing detailed permission settings for individual users like 'mina'.

- Team workspace (a.k.a organization) & Collaboration 가능
- Notebook / Space / Resource에 user 별 ACL 설정 가능



그래서 ZEPL 에서는...

#2. 더더 안정화되고 직관적인 노트북 실행 환경

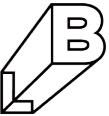
The screenshot shows the ZEPL interface with the following components:

- Top Left:** A dashboard notebook titled "Treatment Explorer Zepl Version_anonymized" by sejun. It displays two charts: a horizontal bar chart and a scatter plot. Below the charts is a terminal window showing Python code execution.
- Top Right:** A modal window titled "Simple Workload 2G" showing task information for a specific task ID.
- Middle Right:** A sidebar menu with tabs for "Resources", "Containers", and "Interpreters". The "Containers" tab is selected, displaying a table of running containers.
- Bottom Right:** A table of "Past Containers" showing completed tasks.

Containers	Resource	Start Time	Usage Duration	Idle Shutdown Time
Untitled Note 4 started	Simple Workload 2G	Oct 19th 2018, 4:39:39 am	57s	10m
Untitled Note 5 started	Simple Workload 2G	Oct 19th 2018, 4:39:59 am	37s	10m

Past Containers	Resource	Start Time	End Time	Usage Duration
Untitled Note 5 stopped	Simple Workload 2G	Oct 19th 2018, 4:38:46 am	Oct 19th 2018, 4:39:23 am	37s
Untitled Note 4 stopped	Simple Workload 2G	Oct 19th 2018, 4:38:54 am	Oct 19th 2018, 4:39:06 am	12s

- 노트북 하나당 Docker container 하나를 띄우는 방식인 ZEPL 노트북
- Container status 를 사용자가 UI로 컨트롤 할 수 있도록 구성
- 사용자는 까만창 터미널 창과 씨름하지 않아도 됨
- 사용자에게 할당된 Resource CRUD 도 UI를 통해 가능



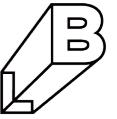
그래서 ZEPL 에서는...

#3. 더더 안정화된 Graph Visualization 환경 (feat. 안되겠다 갈아엎자)



- 사실상 Apache Zeppelin이나 ZEPL을 사용하는 사용자는 데이터 분석 후 “시각화” 까지 한번에 할 수 있다는 점인데, 이 “시각화” 부분에 Zeppelin은 아주 많이 buggy 함
- Frontend 팀은 이걸 하나하나 뜯어 고치다 지칠대로 지친상황 이었죠
- 그래서 Zeppelin에서 온 코드 갈아엎고 새로 다시 썼습니다

결론은 저희도 유저도 행복해졌습니다 😊

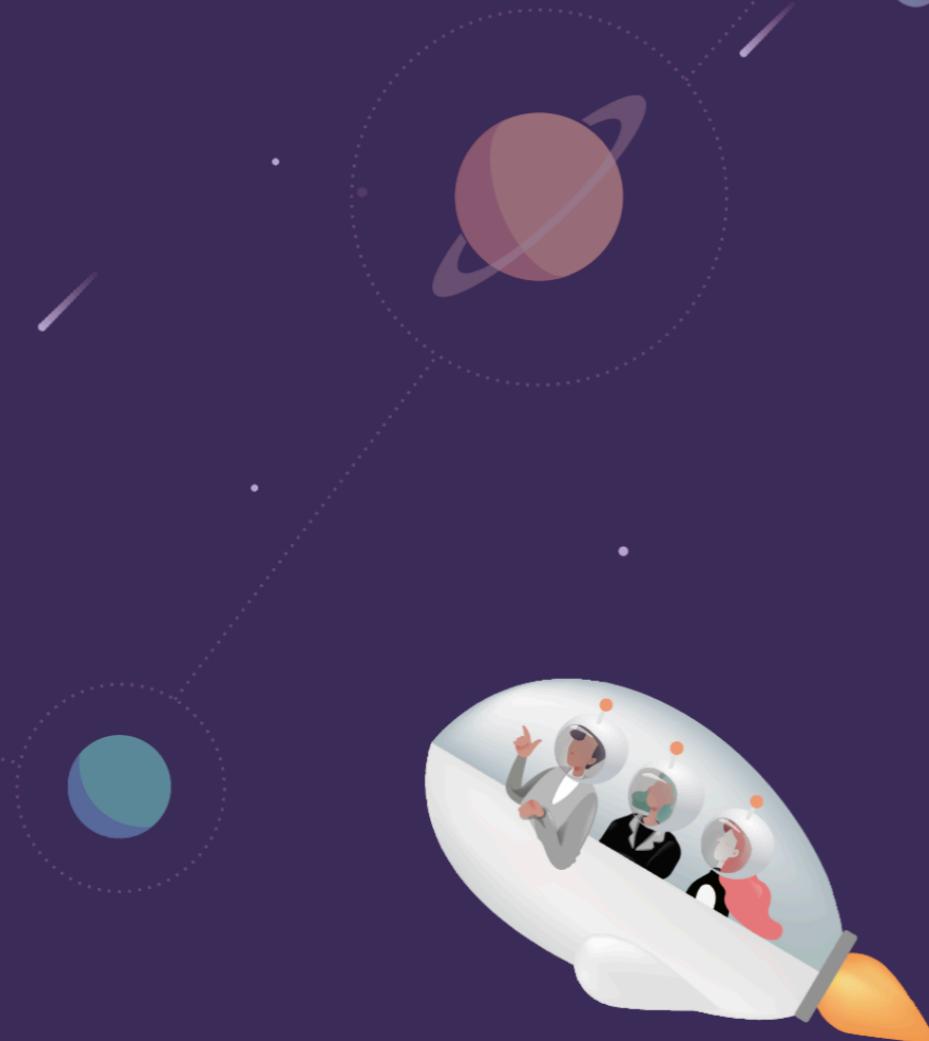


Apache Zeppelin 언급을 1도 하지
않아도 얘기가 통할 그날을 위해
열씨미 일해 보겠습니다



• ZEPL여권 케이스 끼고 ZEPL 미국 지사로 캠프 가자

ZEPL개발자에게 궁금한 점을 물어보세요.
채용 상담을 하시면 선착순으로 ZEPL 여권 케이스를 드립니다.



Apache Zeppelin의 회사
ZEPL에서
DevOps와
웹 개발자(Front/Backend)를
모십니다.



1. 웹 기반 데이터 분석 노트북을 만듭니다
2. 최고의 복지는 최고의 동료입니다. 전 세계에서 온 다국적 개발자와 함께 협업합니다.
3. 실리콘밸리와 서울을 거점으로 시장을 개척합니다.
4. 업무 장소와 시간은 개개인이 자율적으로 결정합니다
5. Slack, GitHub, Jira를 통해 영어로 커뮤니케이션합니다.
6. 오픈소스와 커뮤니티를 사랑합니다. Little Big Data 커뮤니티를 꾸리고 있습니다.

감사합니다

