

クラウド環境構築テンプレート - Jupyter Notebook によるノウハウ共有 -

政谷 好伸

国立情報学研究所

クラウド基盤研究開発センター／先端ICTセンター

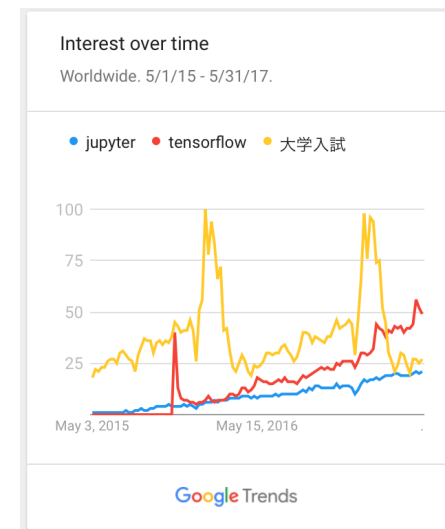
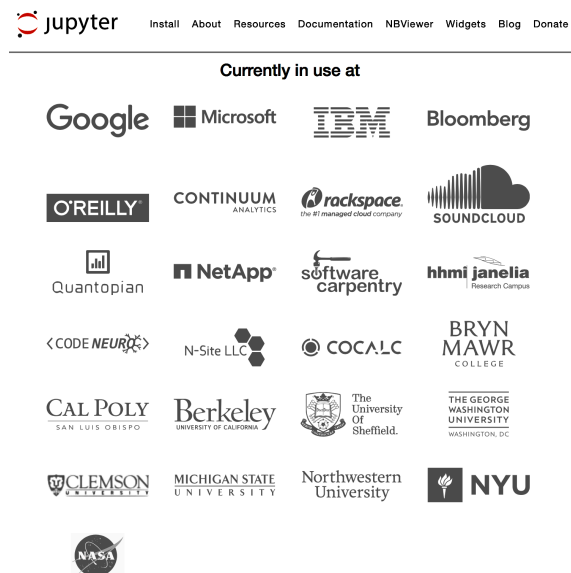
Jupyter Notebook によるノウハウ共有

- クラウド導入後... 研究・教育・業務でクラウドを活用する？
 - 再現性のある計算実験環境の構築ができないか？
 - オープンソースアプリケーションでは、そもそもインストールにつまずく
 - 構築した環境の維持管理業務が煩雑
- システム機能を提供するだけでなく、
使い方・利用ノウハウ “そのもの” の共有も推進したい

Jupyter Notebook ?

- Rich Web client な対話インターフェース
- 実行Code, 実行結果（計算の結果や, 結果を埋め込んだ図表）, 自然言語による手順や経緯の説明を ひとつにまとめて記述
- 数理系研究者のツールとして発展

アイデアの探索課程の記録, 共同開発, 再現可能な成果の公開, 教育



Notebook を用いた環境構築手順の記述

Contents [-] [n](#) [t](#)

- 運用への適用事例: Webサービス
- ログファイルの取得
- Nginxのログを解析する
- 3.1 Setup
- 3.2 Parsing the log
- 解析結果を可視化する
- 参考資料

1 運用への適用事例: Webサービスのログ解析

計算機インフラの御守では種々雑多なドキュメンテーションが不可欠です。日々の作業で証跡を残す、手順を整理して共有・再利用する、ユーザマニュアルや教材を整備する.. 国立情報学研究所 (NII) のクラウド運用担当では、これらをシームレスに記述・蓄積する方法を研究しています。

プロジェクトでは [Jupyter Notebook](#) を用いてドキュメンテーションを行うことで、運用作業の信頼性向上、手順やノウハウの蓄積・流通が容易になることを目指しています。

ここではログを解析する手順・ノウハウをNotebookで記述した例を紹介します。

2 ログファイルの取得

この事例ではみなさんがアクセスしているWebサービスのアクセスログを題材にします。サービスをホスティングしている環境 (JupyterHub) からログをダウンロードします。

● をクリックすると一連の手順をまとめて実行できます。表題左側の ▶ をクリックすると畳み込まれている手順を確認できます。

```
In [1]: master_node = '10.24.1.241' # JupyterHub のアドレス
```

ダウンロード先の一時ファイルをダウンロードするディレクトリを作成します。

```
In [2]: import tempfile
import os
work_dir = tempfile.mkdtemp() # テンポラリのディレクトリを作成
notebook_dir = os.getcwd()

print work_dir, notebook_dir

/tmp/tmpJkH18o /notebooks
```

ファイルをダウンロードします。

```
In [3]: !cd {work_dir} && wget -r http://{master_node}:8080/ && cd {notebook_dir}; \
ls -latr {work_dir}/{master_node}:8080/

2017-06-07 17:22:35 -- http://10.24.1.241:8080/
Connecting to 10.24.1.241:8080... connected.
HTTP request sent, awaiting response... 200 OK
Length: unspecified [text/html]
Saving to: '10.24.1.241:8080/index.html'

10.24.1.241:8080/in [ <=]
2017-06-07 17:22:35 (42.4 MB/s) - '10.24.1.241:8080/index.html' saved [380]

Loading robots.txt; please ignore errors.
--2017-06-07 17:22:35-- http://10.24.1.241:8080/robots.txt
Reusing existing connection to 10.24.1.241:8080.
HTTP request sent, awaiting response... 404 Not Found
2017-06-07 17:22:35 ERROR 404: Not Found.

--2017-06-07 17:22:35-- http://10.24.1.241:8080/access.log
Reusing existing connection to 10.24.1.241:8080.
HTTP request sent, awaiting response... 200 OK
Length: 2444384 (2.3M) [text/plain]
Saving to: '10.24.1.241:8080/access.log'

10.24.1.241:8080/ac 100%[=====] 2.33M --.-KB/s in 0.02s
2017-06-07 17:22:35 (111 MB/s) - '10.24.1.241:8080/access.log' saved [2444384/2444384]

--2017-06-07 17:22:35-- http://10.24.1.241:8080/error.log
Reusing existing connection to 10.24.1.241:8080.
HTTP request sent, awaiting response... 200 OK
Length: 41624 (41K) [text/plain]
```

Webサービスのログ解析

ログファイルの取得

ログの解析

解析結果の可視化

Notebookによるノウハウ共有

- 作業の信頼性向上・再現性を保証
- 手順の再利用
- ポータブルな手順
 - 環境に依存しない...
 - 環境の変更に追従できる...
 - Systems System への適応



- 計算機インフラの御守では種々雑多なドキュメンテーションが不可欠です。日々の作業で証跡を残す、手順を整理して共有・再利用する、ユーザマニュアルや教材を整備する.. 国立情報学研究所（NII）のクラウド運用担当では、これらをシームレスに記述・蓄積する方法を研究しています。
- プロジェクトでは Jupyter Notebook を用いてドキュメンテーションを行うことで、運用作業の信頼性向上、手順やノウハウの蓄積・流通が容易になることを目指しています。
- “機械的に再現できる、人が読み解ける手順”
- 共通認識の形成に寄与すること！ .. 現場でのコミュニケーションを促すもの

いろいろなHadoop環境(だけじゃないけど)を作ったり壊したり動かし続けたり

- 国立情報学研究所では、BIGCHA 2015、トップエスイーなどでHadoopを利用した集中講義を実施したり、研究目的でいろいろなクラスタを必要に応じて構築・運用したりしています。
- 学術機関の利用者ですと、**利用目的**により必要な構成・設定が異なってきます。
- 例えば、**セキュリティ**の設定を必要としていたり、**最新**のパッケージを使いたかったり、Hadoop**外のツール**が欲しかったり、付帯的なソフトウェアも含めたかったり、プロビジョニング**対象が異なったり**など。
- これらの、**少しずつ異なる環境**の構築・運用作業を再現可能かつ再利用しやすい形で実施していくことを追求..

畳みこみ表示・まとめ実行

- 詳細は気にせずNotebook形式の手順を気軽に利用したい
- 定型的な作業をまとめて実行したい
- 関心のある部分だけアレンジやカスタマイズができるように、利用者のスキルや活用シーンに合わせ見通し良く手順を記述する
- まとめて実行してエラーが発生した場合には、詳細を確認する

畳みこみ表示・まとめ実行

2 ログファイルの取得

この事例ではみなさんがアクセスしているWebサービスのアクセスログを題材にします。サービスをホスティングしている環境（JupyterHub）からログをダウンロードします。

まとめ実行機能
表題の左側にある ▶ は、詳細な手順が畳み込まれていることを示しています。▶ をクリックすると、▼ に変化し、畳み込まれている内容を表示することができます。

畳みこまれていた内容に「手順」が含まれていると、右のように表示されます。■ の数は畳み込まれている手順のステップ数を示しています。表示から ● を押すと、配下の手順をまとめて実行することができます。

すべてのステップが終了すると右のように表示が変化します。実行が完了したステップは薄緑 ■ に表示されます。また、完了したステップは、再度まとめ実行ボタン ● をクリックしても実行されないようロック 🔒 されます。

途中でエラーが発生した場合は当該のステップが薄紅 ■ に表示されるので、(▶) をクリックして内容を確認します。

● をクリックすると一連の手順をまとめて実行できます。表題左側の ▶ をクリックすると畳み込まれている手順を確認できます。

▶

3 Nginx のログを解析する

このWebサービスはNginxを用いています。先に取得したログを可視化してみましょう。以下の手順は [By Jess Johnson](#) (<http://grokcode.com>) を参考にしています。

以下の手順で、前節で読み込んだログファイルをパースして [PandasのDataFrame](#) に変換します。

● をクリックすると一連の手順をまとめて実行できます。表題左側の ▶ をクリックすると畳み込まれている手順を確認できます。

▶

Setup 🔒 🔒 🔒

Parsing the log 🔒 🔒 🔒 🔒 🔒 🔒 🔒 🔒 🔒

Webサービスのログ解析

ログファイルの取得

ログの解析

解析結果の可視化

ログ出力の要約

- 大量のログ出力を要約表示する
- 複数の実行ログを各々ファイルに保存し、後で全体を参照したり、結果を比較できるようにする

例えば、某かのパッケージ群をインストールしたりすると大量のログが出力されますが、JupyterのWeb UI上で大量の出力を扱うのはなにかと不便です。JupyterのCellの中でログの内容を検索したり、比較したりするのはまどろっこしく手間がかかります。

ログ出力の要約

```
In [11]: !!from time import sleep
with open("/var/log/bootstrap.log", "r") as f:
    count = 0
    limit = 100
    for line in f:
        count = count+1
        if count > limit: break
        print(line),
        sleep(0.05)
    print ("after", limit, "lines are ignored")
```

```
start time: 2017-06-02 20:18:54(JST)
end time: 2017-06-02 20:18:59(JST)
Output Size(byte): 5043, Lines: 113, Path: /notebooks/.log/20170602/20170602-201855-0506.log
13 keyword matched or stderr happened
```

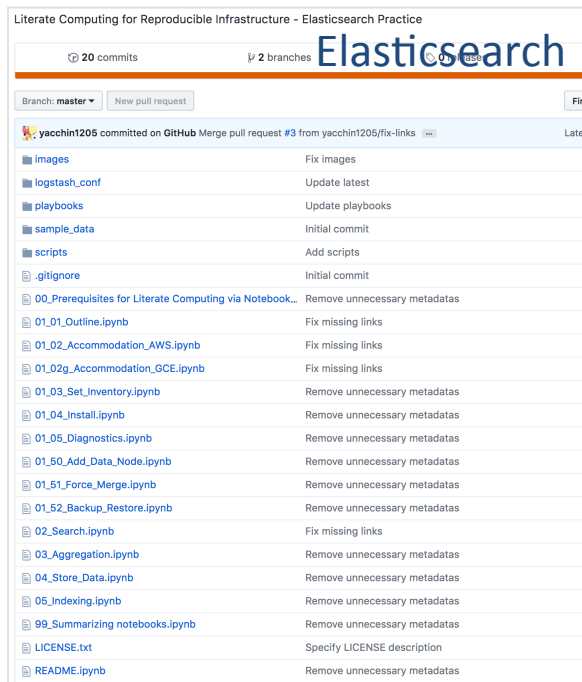
```
gpgv: Signature made Sat May 6 05:15:04 2017 PDT
gpgv: using RSA key 8B48AD6246925553
gpgv: Good signature from "Debian Archive Automatic Signing Key (7.0/w
gpgv: Signature made Sat May 6 05:15:04 2017 PDT
gpgv: using RSA key 7638D0442B90D010
gpgv: Good signature from "Debian Archive Automatic Signing Key (8/jes
gpgv: Signature made Sat May 6 05:28:49 2017 PDT
gpgv: using RSA key CBF8D6FD518E17E1
gpgv: Good signature from "Jessie Stable Release Key <debian-release@l
dpkg: warning: parsing file '/var/lib/dpkg/status' near line 5 package
...
dpkg: warning: parsing file '/var/lib/dpkg/status' near line 5 package
dpkg: warning: parsing file '/var/lib/dpkg/status' near line 5 package
dpkg: warning: parsing file '/var/lib/dpkg/status' near line 24 packag
dpkg: warning: parsing file '/var/lib/dpkg/status' near line 24 packag
dpkg: warning: ignoring pre-dependency problem!
dpkg: warning: parsing file '/var/lib/dpkg/status' near line 50 packag
dpkg: warning: parsing file '/var/lib/dpkg/status' near line 50 packag
dpkg: warning: ignoring pre-dependency problem!
dpkg: warning: ignoring pre-dependency problem!
dpkg: warning: ignoring pre-dependency problem!
dpkg: warning: ignoring pre-dependency problem!
dpkg: warning: ignoring pre-dependency problem!
dpkg: warning: ignoring pre-dependency problem!
...
```

```
Setting up dpkg (1.17.27) ...
Selecting previously unselected package libc6:amd64.
(Reading database ... 380 files and directories currently installed.)
Preparing to unpack .../libc6_2.19-18+deb8u9_amd64.deb ...
Unpacking libc6:amd64 (2.19-18+deb8u9) ...
dpkg: libc6:amd64: dependency problems, but configuring anyway as you requested:
libc6:amd64 depends on libgcc1; however:
('after', 100, 'lines are ignored')
```

- ✓ 要約されてた結果が出力されるようになります(例:最初10行と最後の10行)
- ✓ オリジナルの出力結果全体はファイルに保存されます。実行毎に各々のファイルに結果が保存されるので出力結果を比較することができます
- ✓ 要約中でも、エラーなど特定のパターン含む行を表示します(カスタマイズ可)

Notebook群の関係性を見える化

- 複雑なシステムの構築手順、利用手順の記述もまた複雑
- 関係性を整理して「見える化」



Literate Computing for Reproducible Infrastructure - Elasticsearch Practice

20 commits 2 branches

Branch: master New pull request

File	Commit Message
images	Fix images
logstash_conf	Update latest
playbooks	Update playbooks
sample_data	Initial commit
scripts	Add scripts
.gitignore	Initial commit
00_Prerequisites for Literate Computing via Notebook...	Remove unnecessary metadatas
01_01_Outline.ipynb	Fix missing links
01_02_Accommodation_AWS.ipynb	Fix missing links
01_02g_Accommodation_GCE.ipynb	Fix missing links
01_03_Set_Inventory.ipynb	Remove unnecessary metadatas
01_04_Install.ipynb	Remove unnecessary metadatas
01_05_Diagnostics.ipynb	Remove unnecessary metadatas
01_50_Add_Data_Node.ipynb	Remove unnecessary metadatas
01_51_Force_Merge.ipynb	Remove unnecessary metadatas
01_52_Backup_Restore.ipynb	Remove unnecessary metadatas
02_Search.ipynb	Fix missing links
03_Aggregation.ipynb	Remove unnecessary metadatas
04_Store_Data.ipynb	Remove unnecessary metadatas
05_Indexing.ipynb	Remove unnecessary metadatas
99_Summarizing notebooks.ipynb	Remove unnecessary metadatas
LICENSE.txt	Specify LICENSE description
README.ipynb	Remove unnecessary metadatas



Literate Computing for Reproducible Infrastructure - Hadoop Practice

36 commits 1 branch

Branch: master New pull request

File	Commit Message
group_vars	Add notebooks for Hadoop
images	Convert SVG to PNG to display on GitHub
playbooks	Merge pull request #13 from yacchin1205/spark-single
scripts	Add notebooks for Hadoop
D00_Prerequisites for Literate Comput...	Add notebooks for Hadoop
D01_GCE - Set! Go! (Google Compute ...	Add notebooks for Hadoop
D03_KVM - Ready! on CentOS.ipynb	Add notebooks for Hadoop
D03b_KVM - Set! CentOS6.ipynb	Add notebooks for Hadoop
D03c_KVM - Go! VM.ipynb	Add notebooks for Hadoop
D10_Hadoop - Set! Inventory.ipynb	Add notebooks for Hadoop
D11_Hadoop Prerequisites - Ready! on...	Add notebooks for Hadoop
D12_Hadoop - Ready! on CentOS6 - Z...	Add notebooks for Hadoop
D13a_Hadoop Swimlanes - Ready! on ...	Add notebooks for Hadoop
D13b_Hivemall - Ready! on Hive.ipynb	Add notebooks for Hadoop
D90_Postscript - Operational Policy Se...	Add notebooks for Hadoop
LICENSE.txt	Specify LICENSE description
O03_GCE - Destroy VM (Google Comp...	Add notebooks for Hadoop
O03_KVM - Destroy VM on KVM.ipynb	Add notebooks for Hadoop
O12a_Hadoop - Start the services - ZK...	Add notebooks for Hadoop
O12b_Hadoop - Stop the services - Sp...	Add notebooks for Hadoop
O12c_Hadoop - Decommission DataNo...	Add notebooks for Hadoop
O12d_Hadoop - Restore a Slave Node...	Add notice about error counters
README.ipynb	Specify LICENSE description
README.md	Specify LICENSE description
/M - Confirm KVM is healthy .ip...	Add notebooks for Hadoop
/M - Status Report of running V...	Add notebooks for Hadoop
adoop - Confirm the services a...	Add notebooks for Hadoop
adoop - Simple YARN job for Te...	Add notebooks for Hadoop
adoop - Simple HBase query fo...	Add notebooks for Hadoop
adoop - Simple Spark script for...	Add notebooks for Hadoop
adoop - Simple Hivemall query...	Add notebooks for Hadoop
sv	Add notebooks for Hadoop

Elasticsearch : 14 Notebooks & 17 Playbooks
Hadoop : 25 Notebooks & 57 Playbooks

Notebook群の関係性を見える化

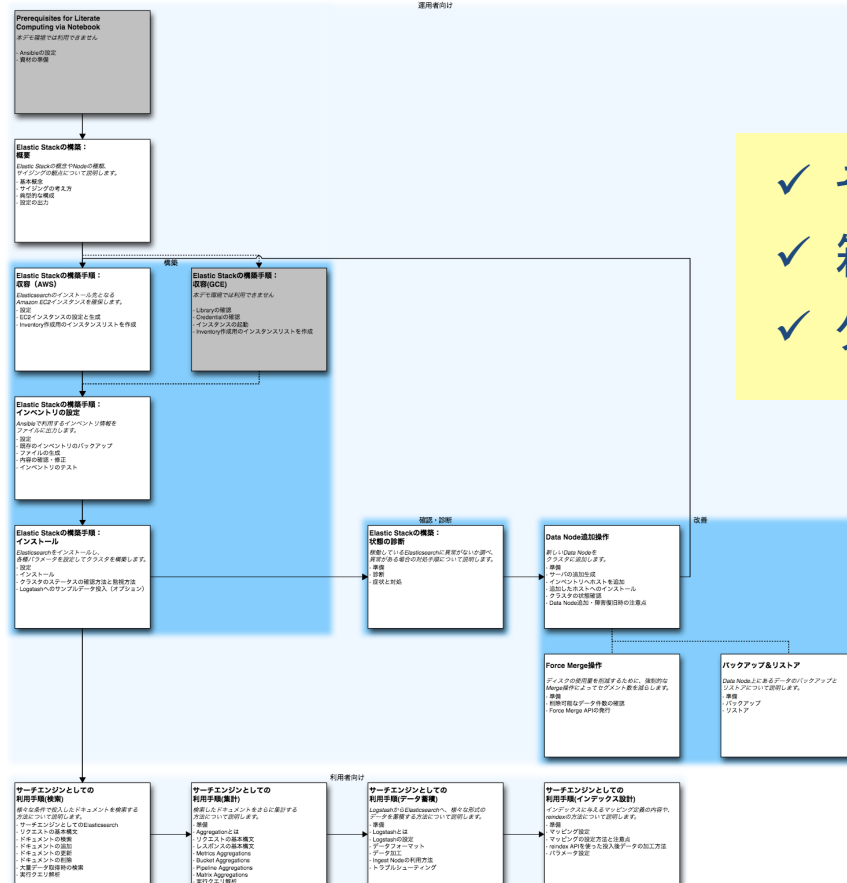
▼ 3.1 お手本Notebookの一覧

お手本Notebookとそれぞれの利用局面の関係については、以下のフローを参考にしてください。

それぞれの箱はNotebookを表しており、**タイトル部分をクリックすることでNotebookを開くことができます。** グレーアウトされたNotebookはこのデモ環境ではお試しいただくことはできません。

```
In [1]: from IPython.display import SVG
SVG(filename='images/notebooks-filled.svg')
```

Out[1]:



- ✓ それぞれの箱は Notebook を表す
- ✓ 箱の中に Notebook の TOC を表示
- ✓ クリックすることで Notebook を開く

お試しデモ環境

NII 国立情報学研究所
学術情報基盤
オープンフォーラム2017
2017年 6/7(水) 9(金)
学術総合センター
21st Century Academic Information Infrastructure for Advancing Open Science
共に考え共に創る学術情報基盤

Day 2 6/8 13:00-14:30 2F 小会議室(定員60名)
クラウドトラック

「オンデマンドクラウド構築サービスとその活用事例」
平成30(2018)年度より、SINETとクラウドを活用した「オンデマンドクラウド構築サービス」の運用開始を予定しています。本セッションでは、サービスの概要と、Jupyter Notebookを使った本サービスの活用事例について紹介します。

B14 インフラの運用って大変で分かりにくいよね、スマートに見えませんか？
運用作業の証跡も、再利用する手順の記述も、教材作成もNotebookでやってみる
クラウド基盤研究開発センター

NIIでは研究者向けクラウドサービスを提供するに際し、Literate Computing for Reproducible Infrastructure という考え方を提案しています。日々の運用作業の証跡を記録する、そこから手順を整理して再利用する、マニュアルや教材を整備するなど複数局面での計算機利用を、同じような様式で見える化・蓄積することを目的としてJupyter Notebookを活用している実態を示します。

「Jupyter Notebookによるオンデマンド環境構築」をテーマに展示・デモしている環境を、以下のURLにアクセスすることで、体験頂けます。

アクセスURL

<https://nbdemo.ecloud.nii.ac.jp/>



最初の一步

上記URLにアクセスし、提供されたアカウント情報でログインした後、「01_Literate Computingデモ環境について.ipynb」をご覧ください。

ご注意

- 本体験環境は、興味を持って頂いた方に、機能評価を目的として提供するものです。提供されたNotebook内の手順や、Jupyter Notebookの機能を理解するための評価の範囲を外れた操作は、禁止とさせていただきます。
- 体験にアカウントは必要ありません。
- 万が一、先まで本体験



NII Cloud Operation Team
Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan

Repositories 9 People 9 Teams 3 Projects 0 Settings

Search repositories... Type: All Language: All Customize pinned repositories New

Literate-computing-Elasticsearch
Literate Computing for Reproducible Infrastructure - Elasticsearch Practice
Jupyter Notebook ★10 Updated 9 days ago

Jupyter-LC_wrapper
Python Updated 12 days ago

Jupyter-multi_outputs
Jupyter notebook extension - accumulate multiple outputs from a code cell into tabs
JavaScript Updated 22 days ago

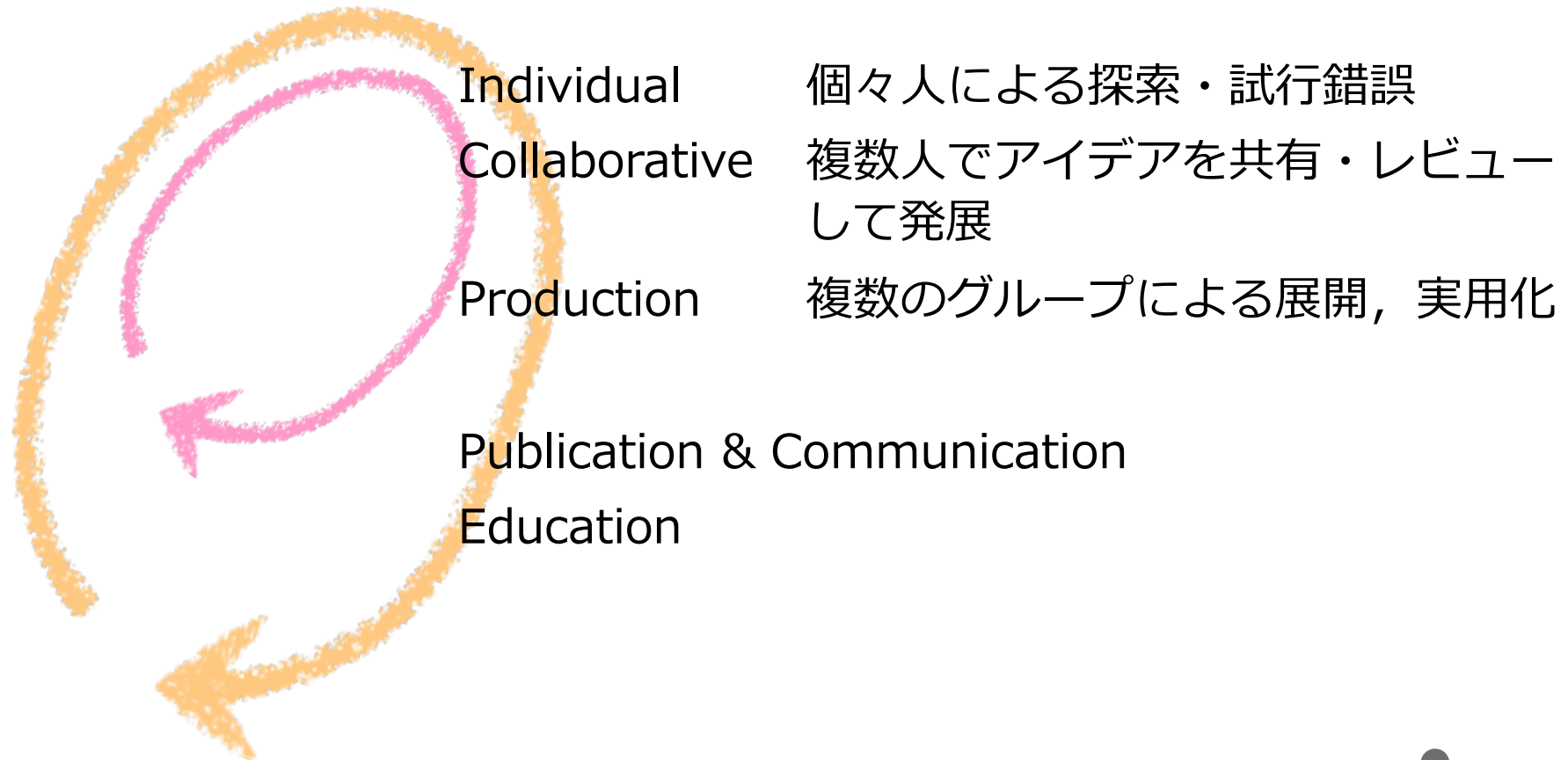
Jupyter-LC_run_through
Jupyter Notebook extension to run cells contained in a section with one click
JavaScript ★1 Updated on Apr 27

Top languages
JavaScript Jupyter Notebook Python

People 9
Invite someone

<https://github.com/NII-cloud-operation>

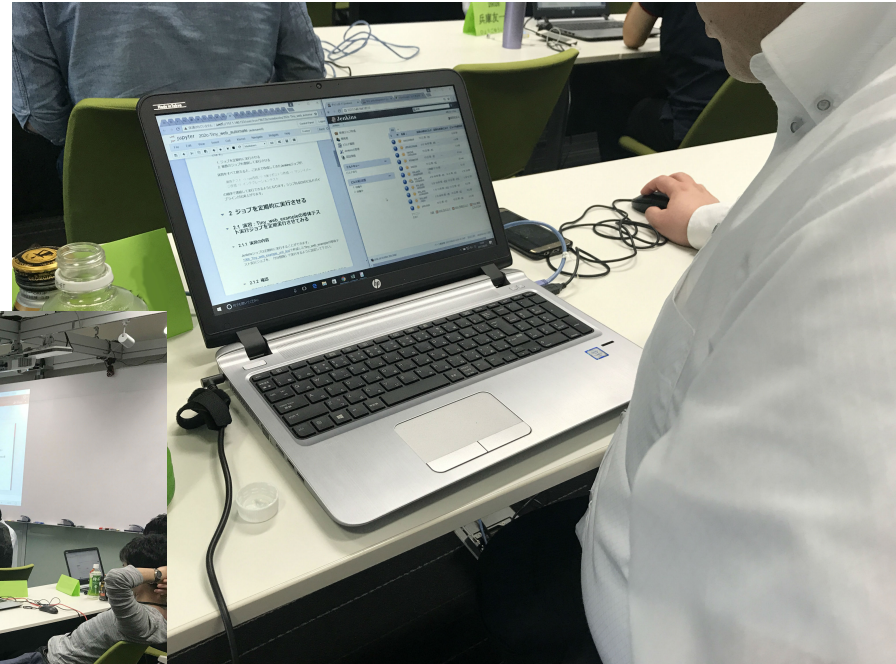
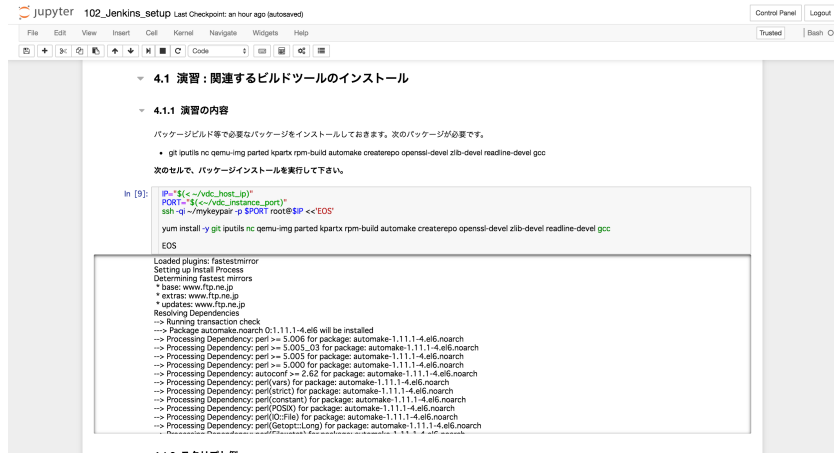
Lifecycle of a Scientific Idea ≡ Engineering Practices



Literate Computing
for Reproducible Infrastructure



TopSE講義でも Jupyter を活用



クラウド実践演習

可能

实践的

担当講師： 山崎 泰宏

応用

講義
演習

1單位

クラウド入門受講で得られた知見を元に、実際にクラウド環境を用いた開発をグループ単位で実施する。プレ修了制作として活用することもできる。

設計

構築

エンジニアリング・マネジメント

品質