

CODE2250BCN

vmware® EXPLORE

Productionizing Jupyter Notebooks

Duygu Hasan(She/her)
Member Of Technical Staff

#vmwareexplore #CODE2250BCN



Jupyter

From Data Exploration to Production

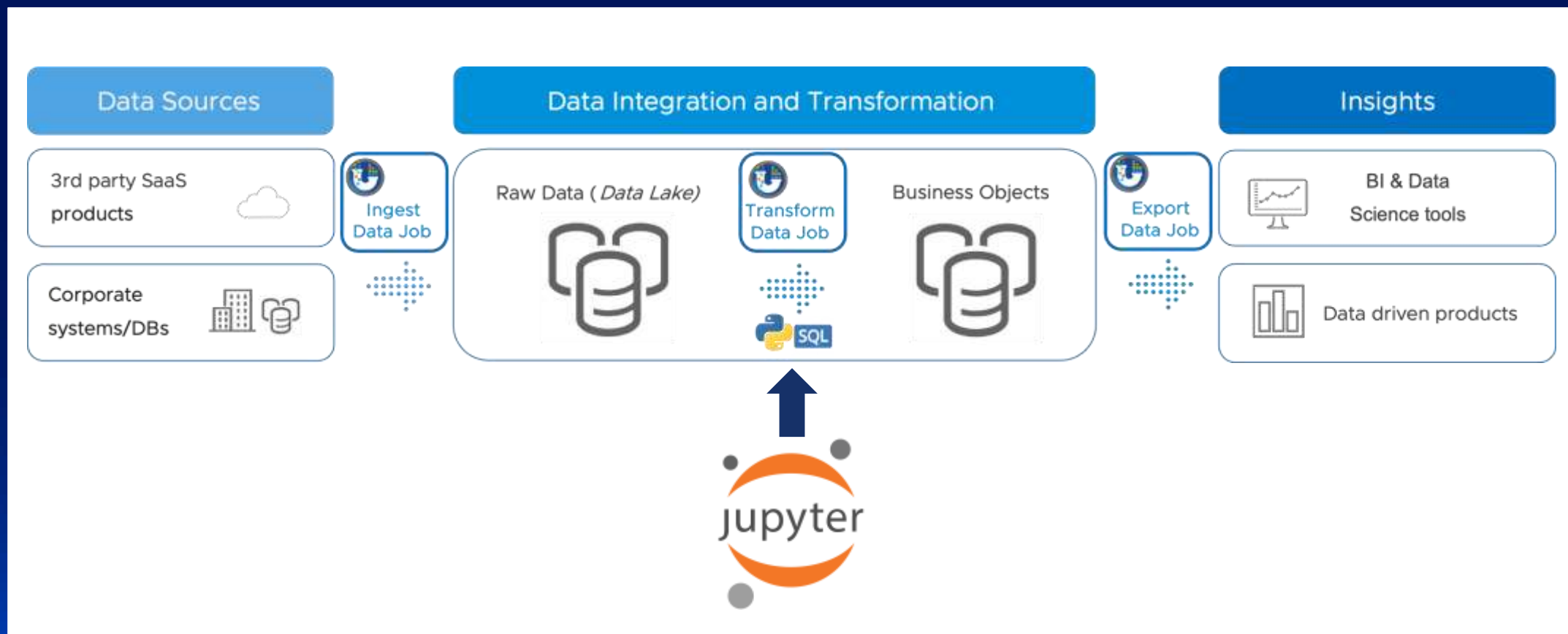
vmware® EXPLORE



Versatile Data Kit



vmware® EXPLORE



Demo

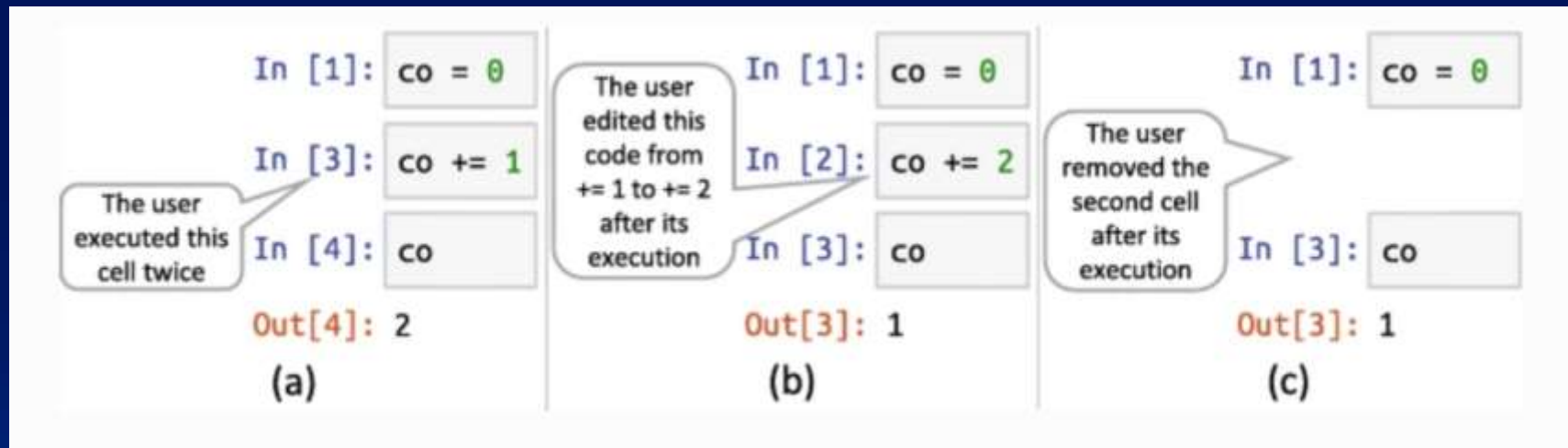
You can access the integration here:

<https://mybinder.org/v2/gh/duyguHsnHsn/vmware-explore/HEAD>



Reproducibility: Non-Linear Execution and Hidden State Risks

vmware® EXPLORE



Irrelevant code for production

vmware® EXPLORE

Irrelevant



```
[ ]: import pandas as pd  
[ ]: url = "some-url"  
     df = pd.read_csv(url)  
[ ]: visualise(df)
```



Relevant

Testing

vmware® EXPLORE





How to Test Jupyter Notebooks with Pytest and Nbmake

Dec 14, 2021 — This tutorial describes how you can use the nbmake, a pytest plugin, to automate end-to-end testing of notebooks. jupyter notebook A Jupyter ...
ig Notebooks Locally · Write Executed Notebooks...

Unit testing for notebooks | Databricks on AWS

How to call these functions from Python, R, Scala, and SQL notebooks. How to write unit tests in Python, R, and Scala by using the popular test

interact/testbook:   Unit test your Jupyter Notebooks ...

Previous attempts at unit testing notebooks involved writing the tests in the notebook itself. However, testbook will allow for unit tests to be run against ...

Version Control

vmware® EXPLORE

Without VDK

```
350 {
351   "cell_type": "code",
352   - "execution_count": null,
353   + "execution_count": 4,
354   "id": "cc05260f-2457-4174-9788-f185b24dd821",
355   "metadata": {
356     "tags": [
357       "vdk"
358     ],
359   - "outputs": [],
360   + "outputs": [
361   +   {
362   +     "ename": "NameError",
363   +     "evalue": "name 'job_input' is not defined",
364   +     "output_type": "error",
365   +     "traceback": [
366   +       "\u001b[0;31m-----\u001b[0m",
367   +       "\u001b[0;31mNameError\u001b[0m      Traceback (most recent call last)",
368   +       "Cell \u001b[0;32mIn [4], line 1\u001b[0m\n\u001b[0;32m----> 1\u001b[0m run(\u001b[43mjob_input\u001b[49m)\n",
369   +       "\u001b[0;31mNameError\u001b[0m: name 'job_input' is not defined"
370   +     ]
371   +   },
372   "source": [
373     "run(job_input)"
374   ]
375 }
```

With VDK

```
350 {
351   "cell_type": "code",
352   "execution_count": null,
353   "id": "cc05260f-2457-4174-9788-f185b24dd821",
354   "metadata": {
355     "tags": [
356       "vdk"
357     ],
358   },
359   "outputs": [],
360   "source": [
361     "run(job_input)"
362   ]
363 }
364 ],
```


Learn more about Versatile Data Kit with these sessions

- Streamlining Dataset Creation and Debugging in AI Models with Versatile Data Kit[CO DE2223BCN]

Monday, Nov 6, 4:30 PM - 4:50 PM CET Hall 6.0, VMware {code} Theater

- Centralized Control and Flexibility in Machine Learning Operations [VMTN2224BCN]

Thursday, Nov 9, 1:00 PM - 1:15 PM CET Hall 6.0, VMTN Theater, VMware Communities

Stay Connected

Let's continue the conversation

Follow us @<https://twitter.com/VDKProject>

Visit us @<https://github.com/vmware/versatile-data-kit>

Linkedin: <https://www.linkedin.com/in/duygu-hasan/>

Thank You

Please take
the survey.

Productionizing Jupyter
Notebooks

