

# ***M2 Writeback Extension Use Case***

## **Enterprise Architectures of Big Data - Student Project**

Vahe Shelunts (1850563)

Quynh Tran (569677)

Elias Brummund (519999)

July 2021



alteryx



Exasol



# *M2 Writeback Extension Use Case*



Hochschule für  
Wirtschaft und Recht Berlin  
Berlin School of Economics and Law



1

**Human-in-the-loop Machine Learning**

2

**Machine Learning Pipeline**

3

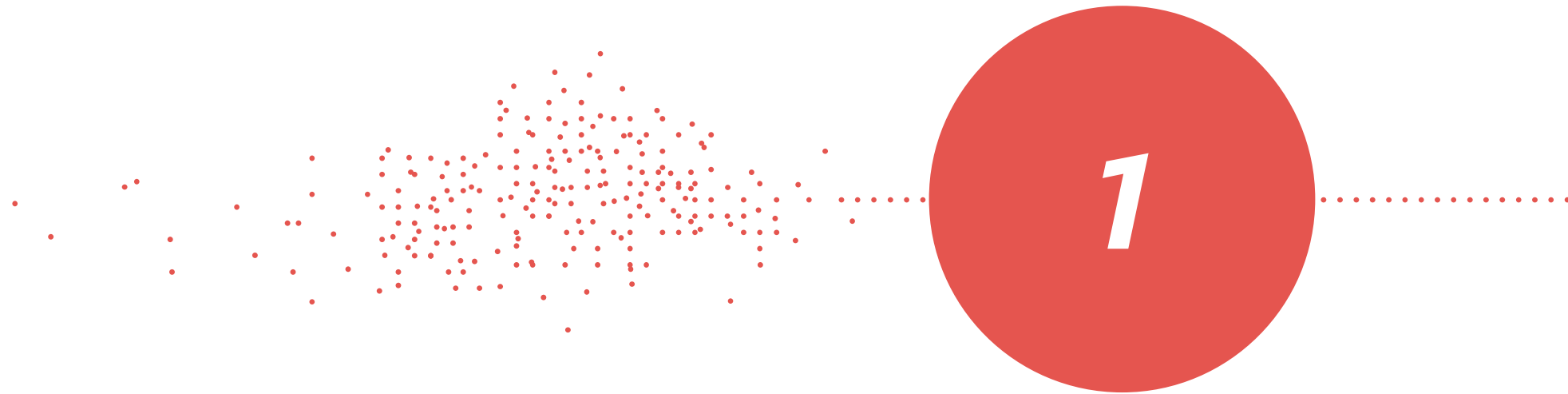
**Server Deployment & Automation**

4

**Writeback Extension**

5

**Outlook**



*Human-in-the-loop Machine Learning*

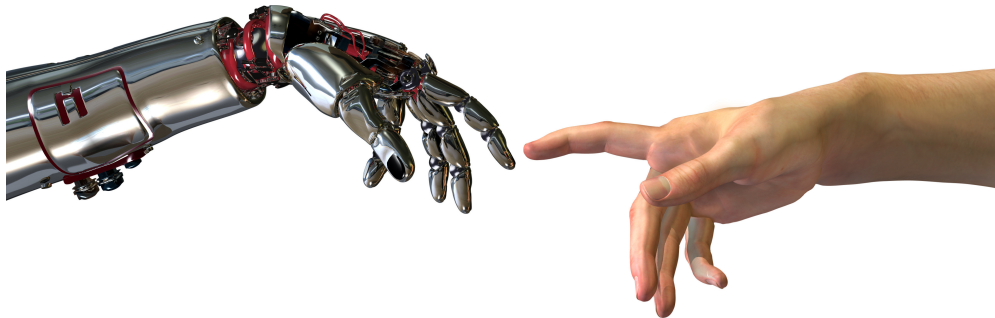
**Better with every Feedback**

# Human-in-the-loop Machine Learning

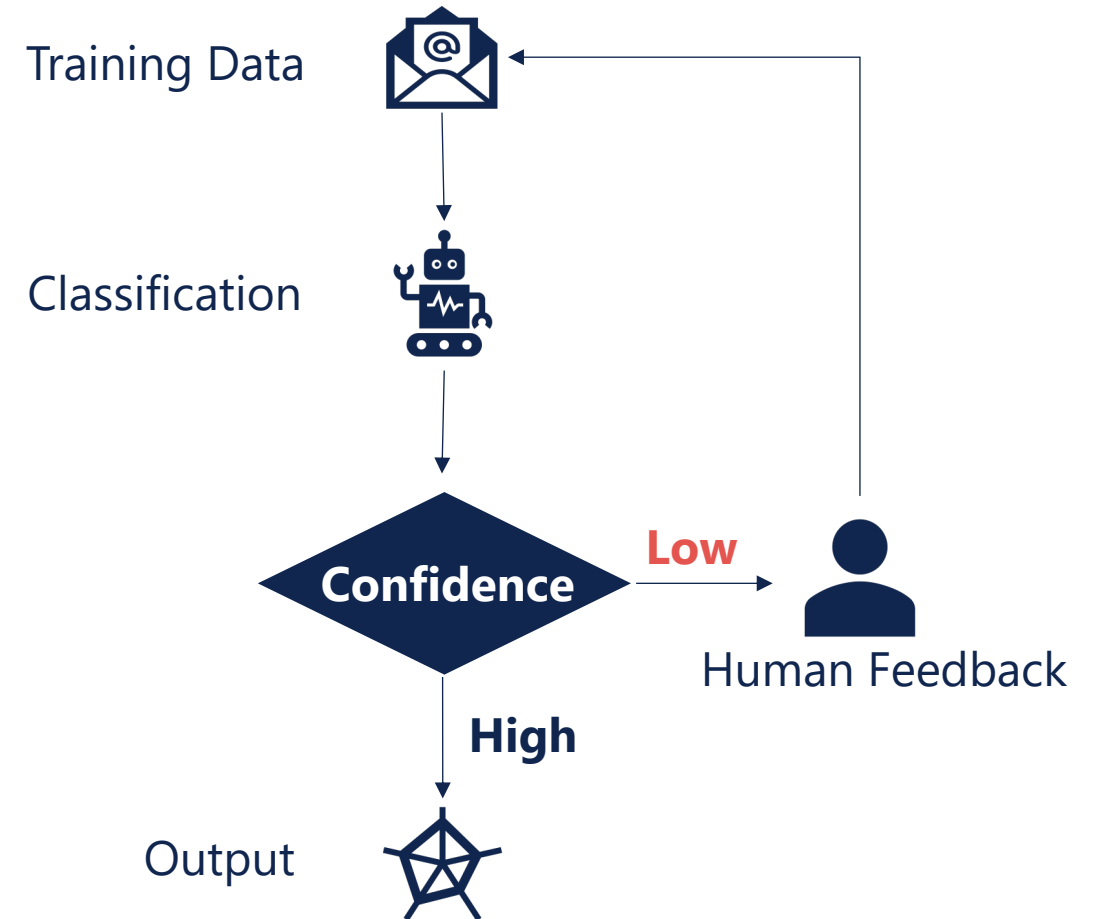
Better with every Feedback

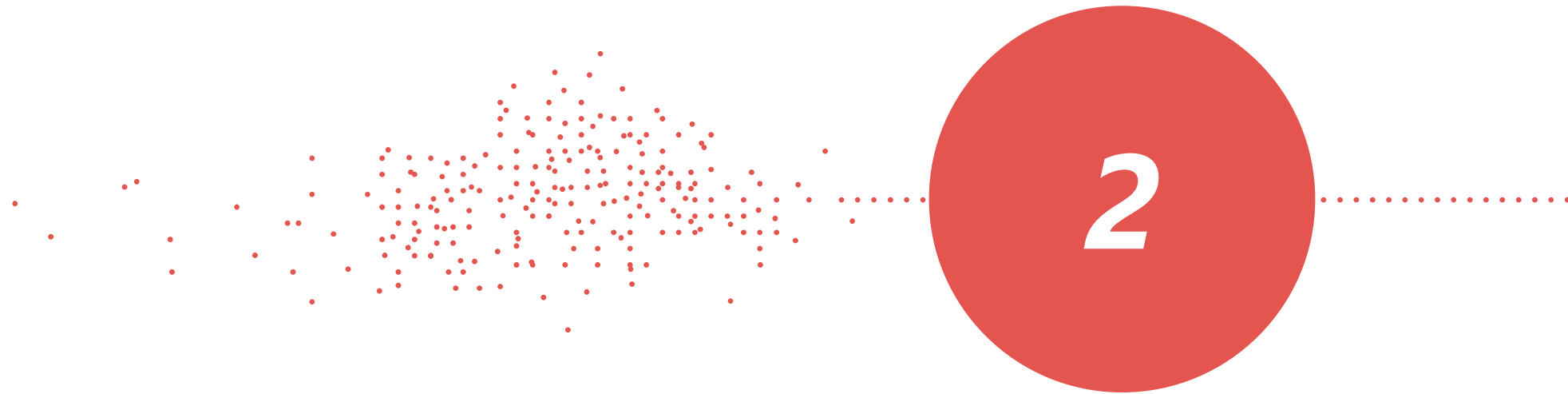


Hochschule für  
Wirtschaft und Recht Berlin  
Berlin School of Economics and Law



Source: <https://images.app.goo.gl/DKNtHEPQ11FYtefr9>

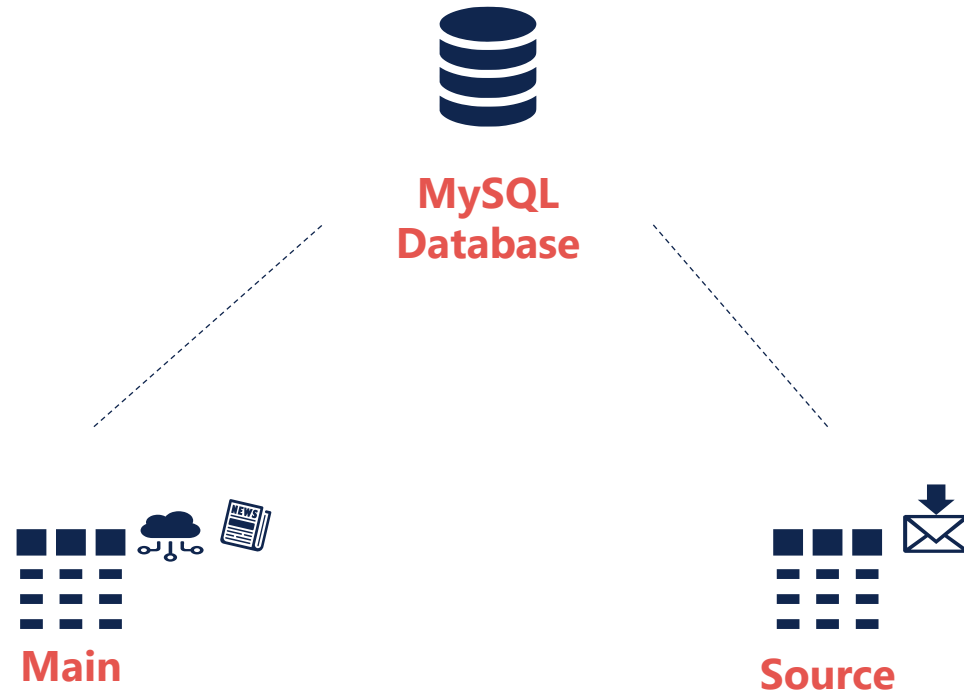




*Data and Classification Algorithm*

**Machine Learning Pipeline**

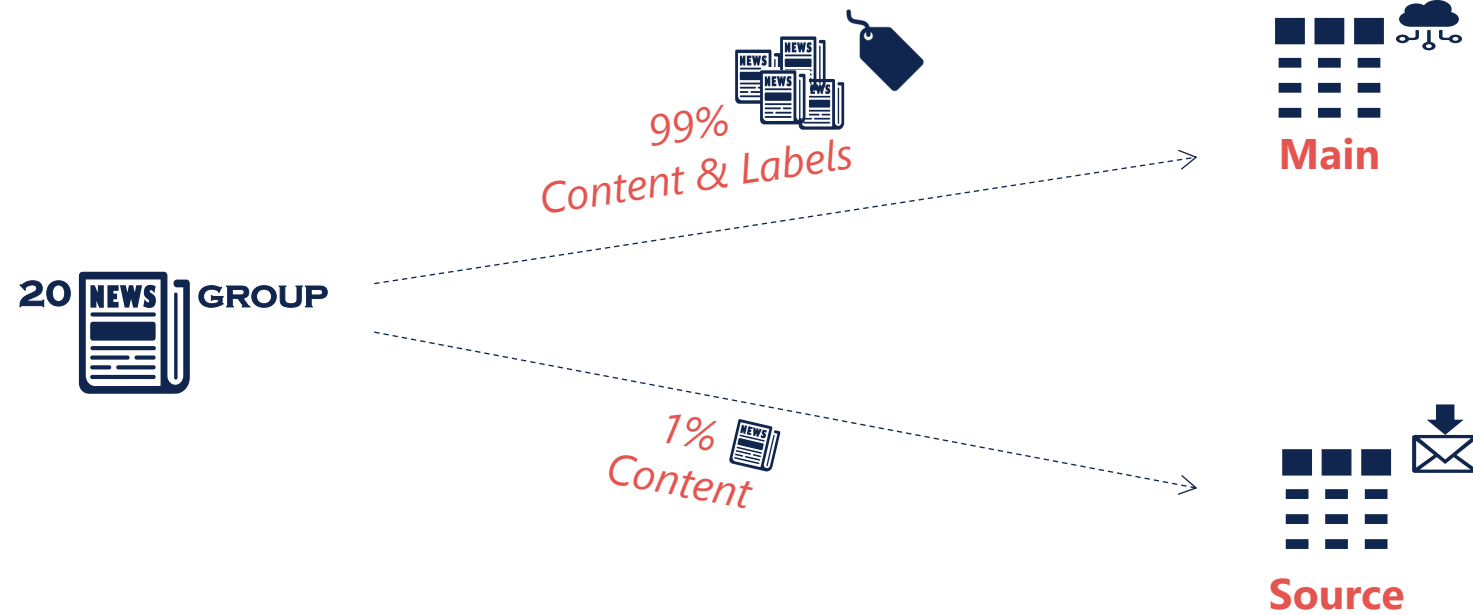
# Data Science Pipeline



# Data Science Pipeline

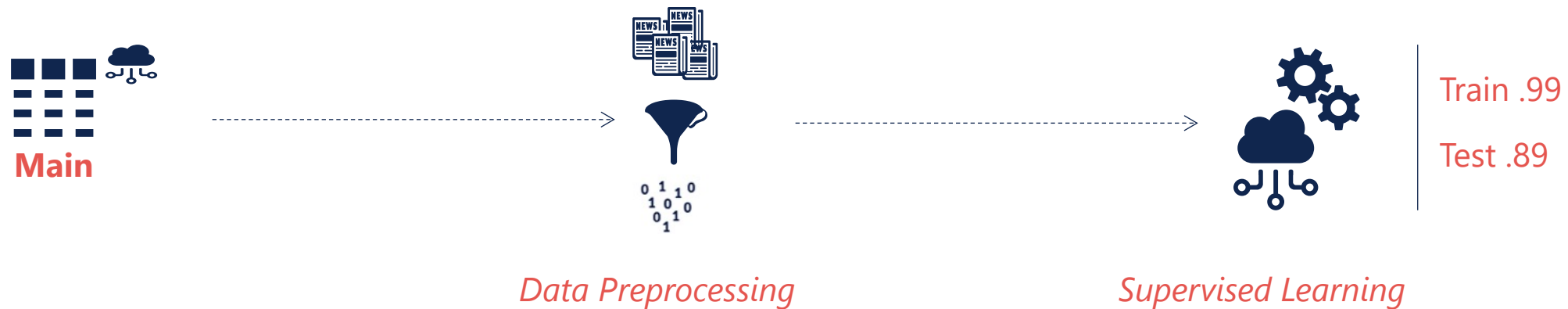


Hochschule für  
Wirtschaft und Recht Berlin  
Berlin School of Economics and Law



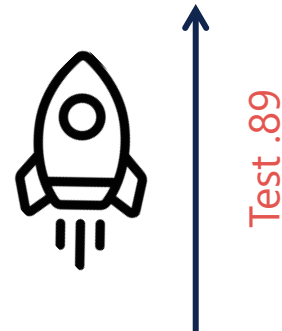


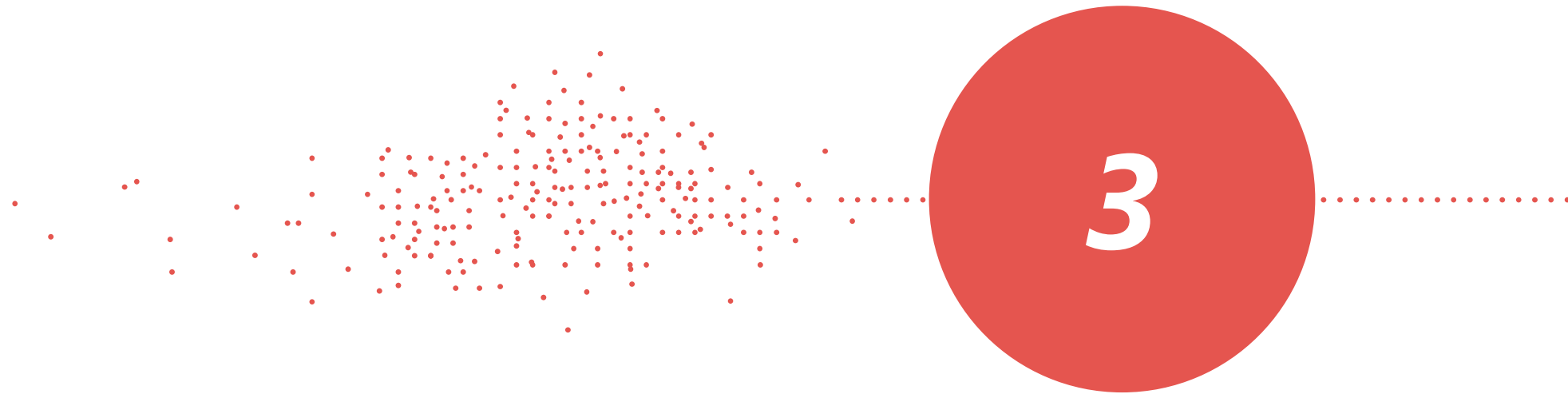
# Data Science Pipeline





# *Data Science Pipeline*





## *Server Deployment & Automation*

**Automatically classify new entries to the database**



Vahe

Quynh

Elias



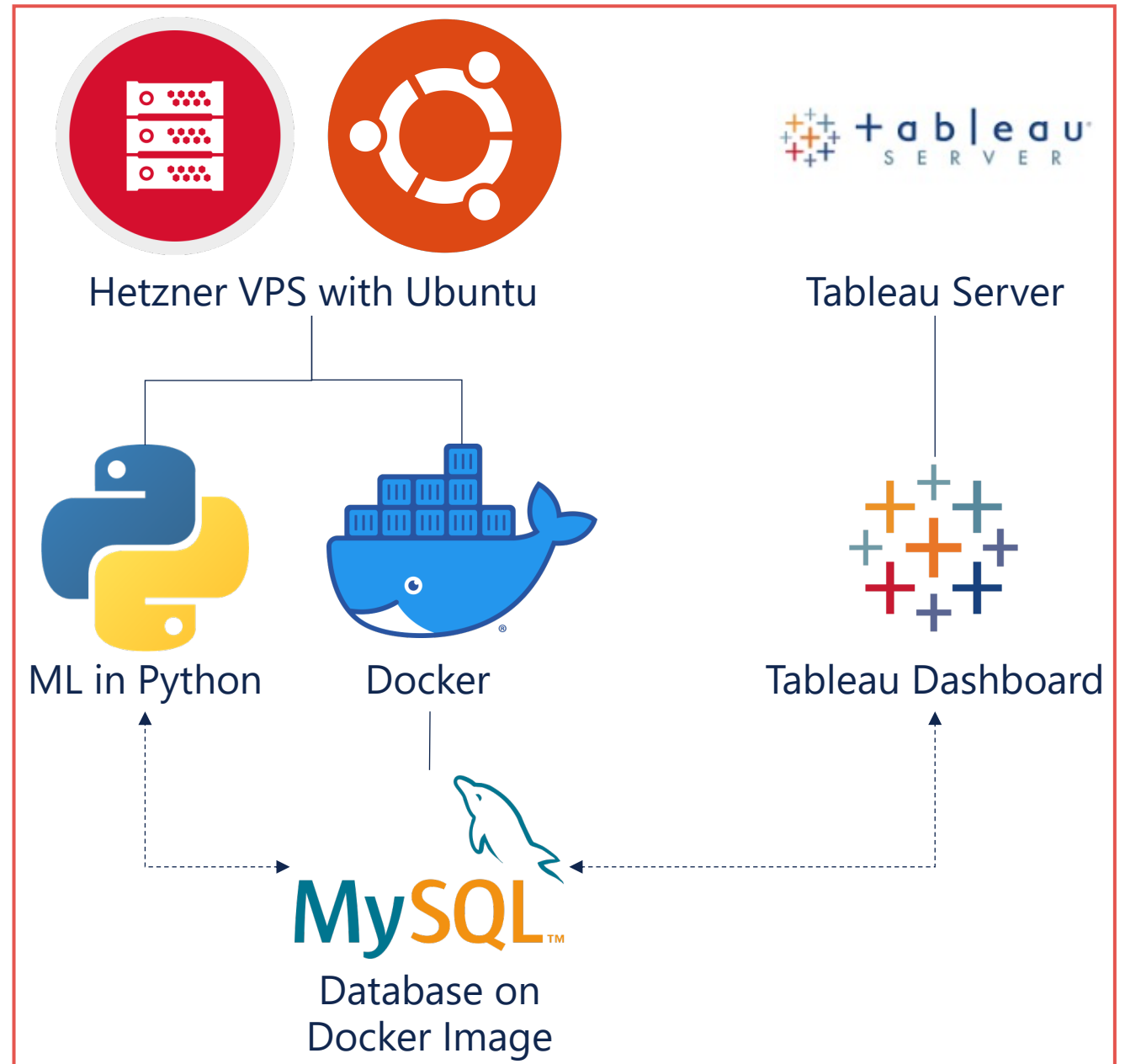
Tableau Desktop



Jupyter Lab

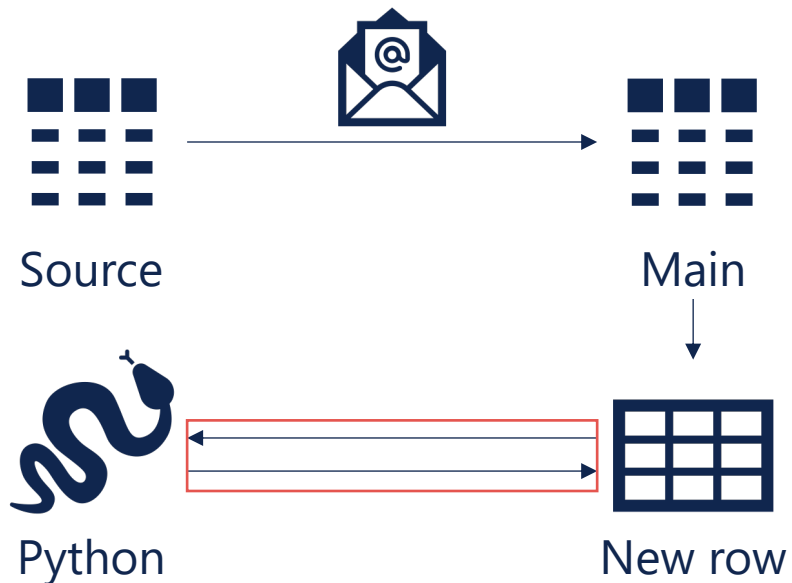


PyCharm



# Server Deployment & Automation

## Crontab



1. Check for new rows
2. Classify new rows
3. Write classification to DB

```
elias — root@ml-writeback-sql-server: ~ — ssh root@162.55.171.160 — 87x34
GNU nano 4.8 /tmp/crontab.ZDgfsX/crontab Modified
# Edit this file to introduce tasks to be run by cron.
#
# Each task to run has to be defined through a single line
# indicating with different fields when the task will be run
# and what command to run for the task
#
# To define the time you can provide concrete values for
# minute (m), hour (h), day of month (dom), month (mon),
# and day of week (dow) or use '*' in these fields (for 'any').
#
# Notice that tasks will be started based on the cron's system
# daemon's notion of time and timezones.
#
# Output of the crontab jobs (including errors) is sent through
# email to the user the crontab file belongs to (unless redirected).
#
# For example, you can run a backup of all your user accounts
# at 5 a.m every week with:
# 0 5 * * 1 tar -zcf /var/backups/home.tgz /home/
#
# For more information see the manual pages of crontab(5) and cron(8)
#
# m h dom mon dow command
* * * * * (cd /home/run/ml_writeback_db ; python3 classification.py)
* * * * * (sleep 10 ; cd /home/run/ml_writeback_db ; python3 classification.py)
* * * * * (sleep 20 ; cd /home/run/ml_writeback_db ; python3 classification.py)
* * * * * (sleep 30 ; cd /home/run/ml_writeback_db ; python3 classification.py)
* * * * * (sleep 40 ; cd /home/run/ml_writeback_db ; python3 classification.py)
* * * * * (sleep 50 ; cd /home/run/ml_writeback_db ; python3 classification.py)
* 4 * * 7 (cd /home/run/ml_writeback_db ; python3 train.py)

^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify ^C Cur Pos
^X Exit ^R Read File ^\ Replace ^U Paste Text ^T To Spell ^_ Go To Line
```



4

***Writeback Extension***



5

## *Outlook*

**Things to consider in a real environment**

## Outlook

### Things to keep in mind when applying the solution to a real-world environment

- How to input emails from webserver to database?
- Design triggers for new emails independent of size of database
- Monitoring model performance over time to validate human-in-the-loop machine learning.
- Writeback Extension can also be used to verify unsupervised learning /classification.



Source: <http://www.quickmeme.com/meme/3gsytz>





***Thank you!***



***Vahe Shelunts***

***Quynh Tran***

***Elias Brummund***

M2. technology & project consulting GmbH  
Milastr. 2 · 10437 Berlin  
+49 30 2089 870 10  
mail@m2dot.com · m2dot.com