

# Where can I run this? An Introduction to **Wikimedia** **Cloud Services**



# What is Wikimedia Cloud Services?

- Platforms and services to help Wikimedia technical contributors
- Free and open for the Wikimedia movement!



# What is Cloud Services?

A stable and efficient hosting platform.

Services that support the creation, operation, and maintenance of tools.

Technical and community support for users of our products.

**30.4%**

Percentage of total edits  
made to Wikimedia wikis by  
Wikimedia Cloud Services  
hosted bots and tools.  
May of 2020



**51.4%**

Percentage of total edits  
made to Wikidata by  
Wikimedia Cloud Services  
hosted bots and tools.  
May of 2020

**161 Projects**  
**800 Instances**

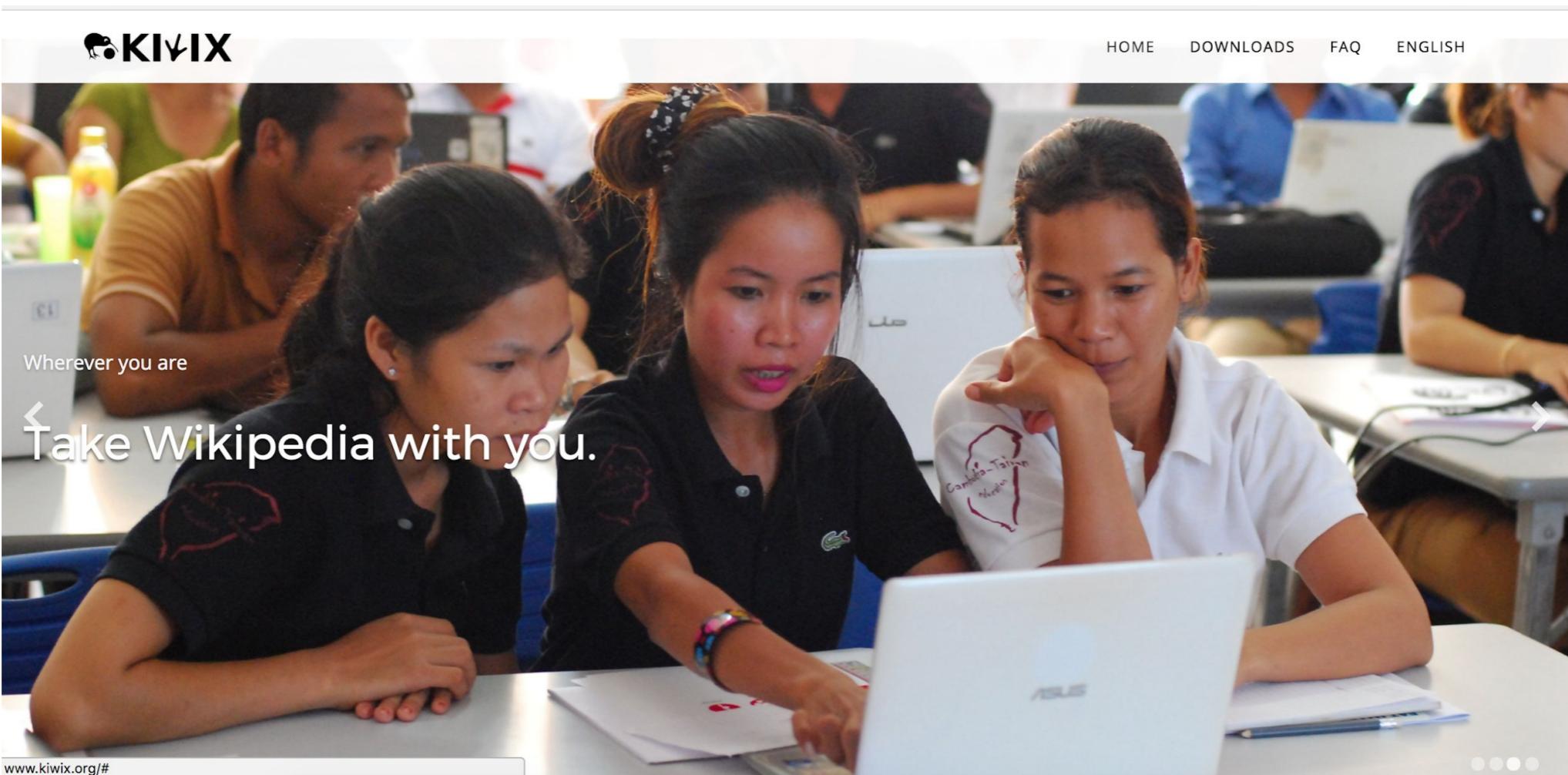
2021-05-20

**2,934 Tools**  
**2,146 Maintainers**

2021-05-20

# Some things hosted on cloud Services...





# Kiwix - Wikipedia Offline

<http://kiwix.org>

[https://meta.wikimedia.org/wiki/Kiwix\\_-\\_Wikipedia\\_Offline](https://meta.wikimedia.org/wiki/Kiwix_-_Wikipedia_Offline)

# Export tool of Wikisource books in many file formats.

## Export a file

Language code	<input type="text" value="en"/>
Language code of Wikisource domain, like en or fr	
Title of the page	<input type="text"/>
Name of the mainpage of the book in Wikisource	
File format	<input type="text" value="epub 3"/> <span style="float: right;">▼</span>
Include fonts	<input type="text" value="None"/> <span style="float: right;">▼</span>
Options	<input type="checkbox"/> Do not include images
<input type="button" value="Export"/>	

Wsexport is a tool under [GNU GPLv2+ licence](#).

The source code is available [here](#) and a small documentation [here](#).

Some [statistics](#) are also provided.

Please report issues [here](#).

# WikiSource Export

<https://tools.wmflabs.org/wsexport/tool/book.php>

Wikidata - The Distribute X +



https://wikidata-game.toolforge.org/distributed/#



The Distributed Game

Log in

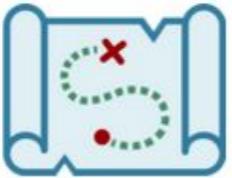


## Available games

This is a list of available games. Games are loaded from different sources, and you can [add your own!](#)

## Reference hunt!

[Reference hunt](#) suggestions to add references in items based on structured data in the web. Source code can be found in [here](#)



## Category Facts

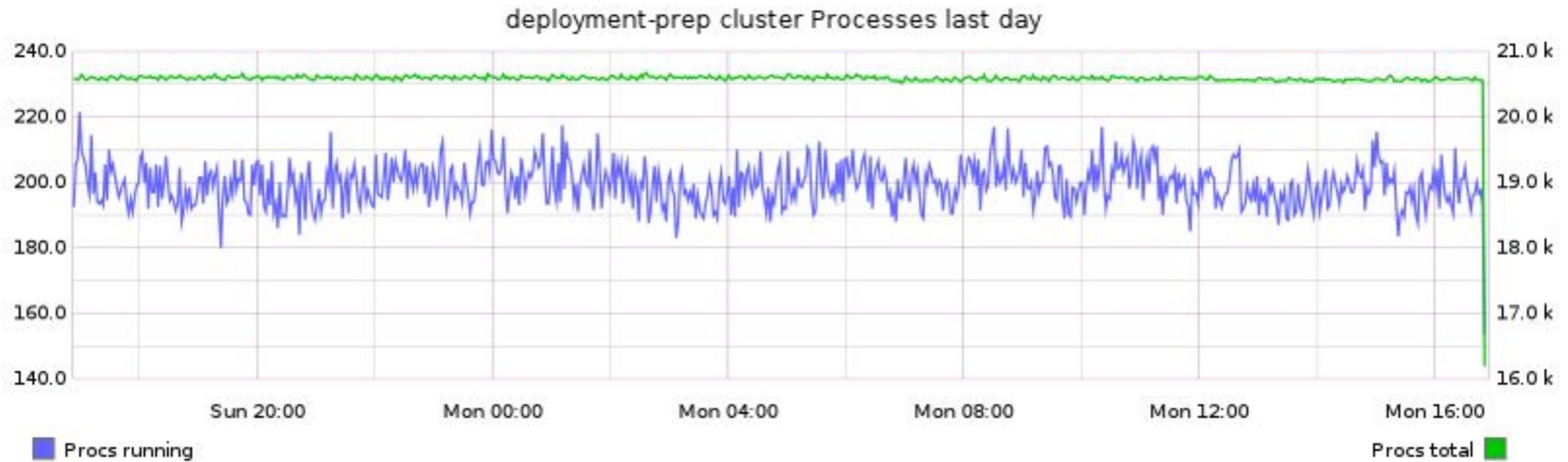
Explore hidden facts from category.



# Wikidata Game

<https://tools.wmflabs.org/wikidata-game>

# Deployment Prep



[https://wikitech.wikimedia.org/wiki/Nova\\_Resource:Deployment-prep](https://wikitech.wikimedia.org/wiki/Nova_Resource:Deployment-prep)

# Service and platform offerings: Code



# Wikimedia Cloud Services products

## Cloud VPS

OpenStack powered  
public cloud hosting

## Toolforge

Managed hosting for  
community tools,  
**Grid Engine**  
batch-queuing system,  
Kubernetes cluster

## Quarry

Web-based SQL  
query interface

## PAWS

JupyterHub service  
with custom integrations  
for accessing APIs  
and wiki replicas

## Wiki replicas

Real-time replicas of  
Wikimedia production  
databases

## ToolsDB

Shared access database  
hosting for Toolforge  
hosted projects

## Dumps

Wikimedia database  
XML dump files,  
Curated datasets

Orphan talk pages - Quarry - Mozilla Firefox

Orphan talk pages - Qua X + https://quarry.wmflabs.org/query/47487

Quarry Beta Home New Query Recent Queries Discuss Database tables Login

Orphan talk pages Fork

This query is marked as a draft by Pppery.

SQL Toggle Highlighting

```
set @namespace = 100;
select * from page where page_namespace=@namespace+1
and not exists (select 1 from page page2 where page2.page_namespace=@namespace and page2.page_title=page.page_title)
and page_title not like "%/Archive_%"
limit 10000;
```

All SQL code is licensed under [CC0 License](#).

Query status: **complete**

Executed in 2.36 seconds as of Thu Oct 15 2020.

Resultset (670 rows) Download data ▾

Show 100 entries Search:

page_id	page_namespace	page_title	page_restrictions	page_is_redirect
27676136	101	African_American/to_do	0	

A screenshot of the Quarry web interface. The title bar says "Orphan talk pages - Quarry - Mozilla Firefox". The address bar shows "https://quarry.wmflabs.org/query/47487". The main header has tabs for "Home", "New Query", "Recent Queries", "Discuss", "Database tables", and "Login". Below the header, the title "Orphan talk pages" is displayed with a "Fork" button. A note says "This query is marked as a draft by Pppery.". The "SQL" section contains the query code. A note at the bottom says "All SQL code is licensed under CC0 License.". Below the SQL section, it says "Query status: complete". It shows the execution time as "Executed in 2.36 seconds as of Thu Oct 15 2020". The "Resultset" section shows "670 rows" with a "Download data" button. There are filters for "Show 100 entries" and a search bar. The results table has columns for page\_id, page\_namespace, page\_title, page\_restrictions, and page\_is\_redirect. One row is shown: page\_id 27676136, page\_namespace 101, page\_title African\_American/to\_do, page\_restrictions 0, and page\_is\_redirect empty.

# DB queries with your Wikimedia account

<https://quarry.wmcloud.org/>

```
In [1]: import os
import pymysql
from pprint import pprint

host = os.environ['MYSQL_HOST']
user = os.environ['MYSQL_USERNAME']
password = os.environ['MYSQL_PASSWORD']

all_names = set()
free_names = {}

def iter_states():
    conn = pymysql.connect(
        host=host,
        user=user,
        password=password
    )
    query = """
SELECT page_title
FROM page
JOIN categorylinks ON cl_from = page_id
WHERE
    cl_to = "States_of_the_United_States"
    AND cl_type = "page"
    AND cl_sortkey_prefix = ""
"""
    with conn.cursor() as cur:
        cur.execute("USE enwiki_p;")
        cur.execute(query)
```

# Run code with your Wikimedia account

<https://paws.wmcloud.org/>

```
[andrewtesttool@tools-sgebastion-07:~$ echo "export PYTHONPATH=/data/project/shared/pywikibot/core:/shared/pywikibot/core/scripts" > .bash_profile
[andrewtesttool@tools-sgebastion-07:~$ mkdir .pywikibot
[andrewtesttool@tools-sgebastion-07:~$ python /data/project/shared/pywikibot/core/generate_user_files.py
WARNING: /mnt/nfs/labstore-secondary-tools-project/pywikibot/public_html/core/pywikibot/__init__.py:124: FutureWarning:
Python 2.7.13 will be dropped in 2020.
It is recommended to use Python 3.5 or above.
See T213287 for further information.

FutureWarning)

You can abort at any time by pressing ctrl-c

Your default user directory is "/mnt/nfs/labstore-secondary-tools-project/andrewtesttool"
user-config.py already exists in the target directory "/mnt/nfs/labstore-secondary-tools-project/andrewtesttool".
Would you like to change the directory? ([Y]es, [n]o): ]
```

# Shell access with a toolforge account

```
andrew — andrew@tools-sgebastion-07: /usr/lib/python3.5 — ssh tools-login.wmflabs.org — 130x23
#!/usr/bin/python3

import pywikibot
import pymysql
from pywikibot import pagegenerators

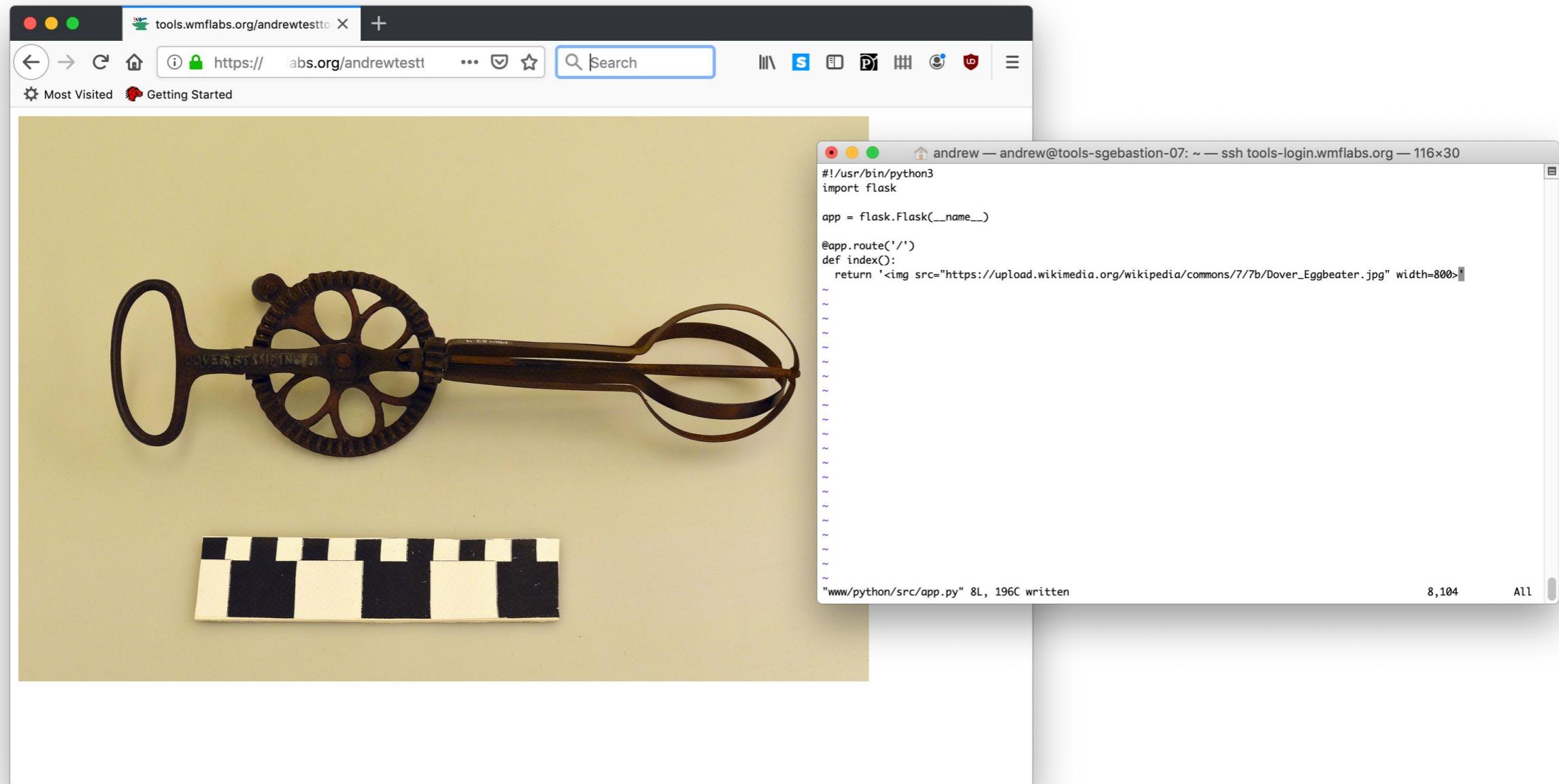
db = pymysql.connect(host="tools.db.svc.eqiad.wmflabs", database="s53605__andrewtesttool", read_default_file="~/replica.my.cnf")
cursor = db.cursor()
query = "SELECT title FROM egg_pages"
cursor.execute(query)
results = cursor.fetchall()
dbpages=[row[0] for row in results]

site = pywikibot.Site('commons', 'commons')
cat = pywikibot.Category(site, 'Category:Hand-powered_beaters')
gen = pagegenerators.CategorizedPageGenerator(cat)

for page in gen:
    text = page.text
    if "egg" in page.text.lower():
        print(page.title())
        if page.title() not in dbpages:
            print("Found new page: " + page.title())
            print("Text: " + text)
```

22,1      Top

# Run a bot with Toolforge



# Make a web service with ToolForge

Project / Compute / Instances

## Instances

Instances

Images API Access Displaying 6 items

	Instance Name	Image Name	IP Address	Flavor	Key Pair	Status	Availability Zone	Task	Power State	Time since created	Actions
Network	<input type="checkbox"/> video-redi	debian-8.2-jessie (deprecated 2016-02-16)	172.16.1.102	m1.small	-	Active	nova	None	Running	8 months, 3 weeks	<button>Associate Floating IP</button>
DNS	<input type="checkbox"/> encoding0	debian-8.5-jessie (deprecated 2016-08-01)	172.16.1.13	m1.gigantic	-	Active	nova	None	Running	8 months, 3 weeks	<button>Associate Floating IP</button>
Puppet	<input type="checkbox"/> encoding0	debian-8.5-jessie (deprecated 2016-08-01)	172.16.1.201	m1.gigantic	-	Active	nova	None	Running	8 months, 3 weeks	<button>Associate Floating IP</button>
Access	<input type="checkbox"/> gfg01	debian-8.5-jessie (deprecated 2016-10-24)	172.16.1.183	m1.xlarge	-	Active	nova	None	Running	8 months, 3 weeks	<button>Associate Floating IP</button>
Identity											

# Run your own virtual server, or a cluster of servers

# Service and platform offerings: Data



# Wikimedia Cloud Services products

## Cloud VPS

OpenStack powered  
public cloud hosting

## Toolforge

Managed hosting for  
community tools,  
**Grid Engine**  
batch-queuing system,  
Kubernetes cluster

## Quarry

Web-based SQL  
query interface

## PAWS

JupyterHub service  
with custom integrations  
for accessing APIs  
and wiki replicas

## Wiki replicas

Real-time replicas of  
Wikimedia production  
databases

## ToolsDB

Shared access database  
hosting for Toolforge  
hosted projects

## Dumps

Wikimedia database  
XML dump files,  
Curated datasets

# wiki replicas

**“Real-time” sanitized queryable  
data from production MediaWiki  
databases**



# ToolsDB

**Database hosting utilized for tool  
and user specific data needs**



# Dumps

## content and metadata of Wikipedia and all other Wikimedia projects

**Front page**

<https://dumps.wikimedia.org/>

**General information**

<https://meta.wikimedia.org/wiki/Data.dumps>



# Working with us



# How to get an account?

## Toolforge

- Purpose built for hosting web based services and tools
- Accounts are granted based on demonstrated productive use and reviewed by an admin before granting
- Create a request  
<https://toolsadmin.wikimedia.org/register/>

## PAWS and Quarry

- If you have a wiki account, you can start using PAWS and Quarry immediately.

## Cloud VPS

- Create a project request  
<https://phabricator.wikimedia.org/project/view/2875/>
  - See guidelines  
[https://wikitech.wikimedia.org/wiki/Help:Cloud\\_VPS\\_project#Guidelines\\_for\\_project\\_requests](https://wikitech.wikimedia.org/wiki/Help:Cloud_VPS_project#Guidelines_for_project_requests)
- Requests are reviewed every few days by the WMCS team
- Unused or abandoned projects are purged yearly

# How to contact us

## Mailing Lists

- Internal mailing list
  - [cloudservices@wikimedia.org](mailto:cloudservices@wikimedia.org)
- Community Lists
  - [cloud-announce@lists.wikimedia.org](mailto:cloud-announce@lists.wikimedia.org)
  - [cloud@lists.wikimedia.org](mailto:cloud@lists.wikimedia.org)
  - [cloud-admin@lists.wikimedia.org](mailto:cloud-admin@lists.wikimedia.org)

## IRC

- **#wikimedia-cloud** on libera.chat for help and issues
- **#wikimedia-cloud-admin** on libera.chat for backend infrastructure chatter

## Newsfeed

- <https://phabricator.wikimedia.org/phame/blog/view/5/>

## Documentation

- [https://wikitech.wikimedia.org/wiki/Main\\_Page](https://wikitech.wikimedia.org/wiki/Main_Page)

## Bugs & Feature Requests

- <https://phabricator.wikimedia.org/tag/cloud-services/>

# Q&A time!



# Credits

- [Wmf logo vert K.svg](#) By Logo and trademark of the Wikimedia foundation, designed by Wikipedia user [Neolux](#), (SVG version created by [DarkEvil](#), revised by [Philip Ronan](#) and optimized by [Zscout370](#) and [Artem Karimov](#)) Later revised by Wikimedia Foundation, [CC BY-SA 3.0](#)
- Various icons By [Outstandy](#) - Graphic designs for Wikipedia 15 from Mule Design, [CC0](#)
- [What is Cloud Services poster.svg](#) By BDavis (WMF) - Own work , includes elements from [Wikimedia labs logo.svg](#), [Wikimedia Foundation logo - horizontal.svg](#) & [Creative Commons yy-sa small.svg](#), [CC BY-SA 4.0](#)
- Various screenshots By [Andrew Bogott](#), [CC BY-SA 4.0](#)
- [WMCS components.svg](#) By Bryan Davis - Own work, based on diagram initially created by [Andrew Bogott](#), [CC BY-SA 4.0](#)

Copyright © 2020-2021, [Bryan Davis](#), [Andrew Bogott](#), [Arturo Borrero](#), [Brooke Storm](#), and the [Wikimedia Foundation](#).

This work is licensed under a [Creative Commons Attribution-Share Alike 4.0 International](#) license.