

Serverless Approach to Working with Data

Jakub Nowacki

whoami

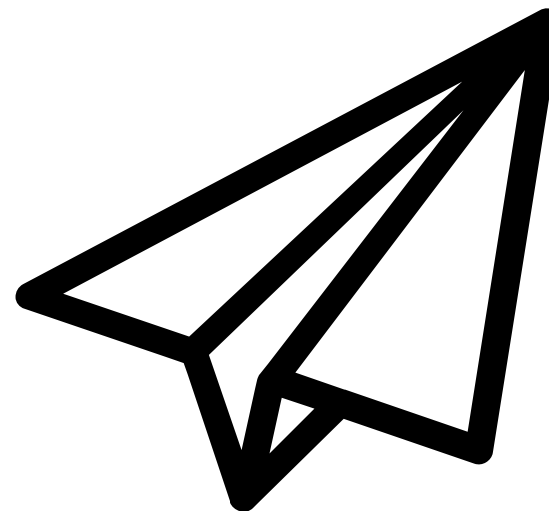
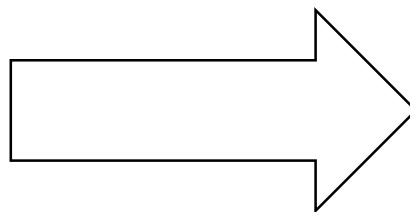
Lead Machine Learning Engineer @ Sotrender (sotrender.com)

Trainer @ Sages (sages.com.pl)

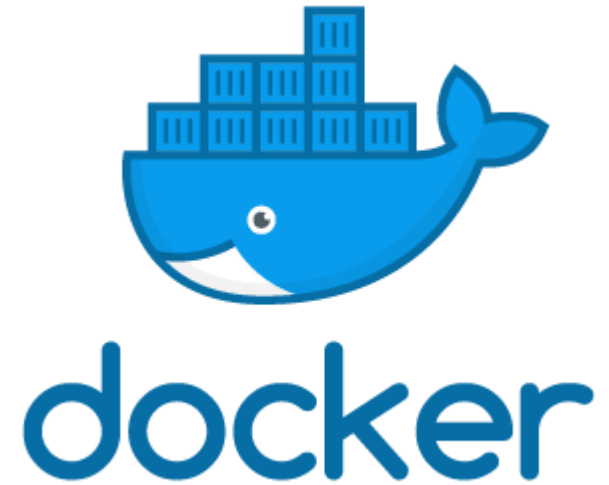
I can code, I do maths

@jsnowacki

So you've got a function...



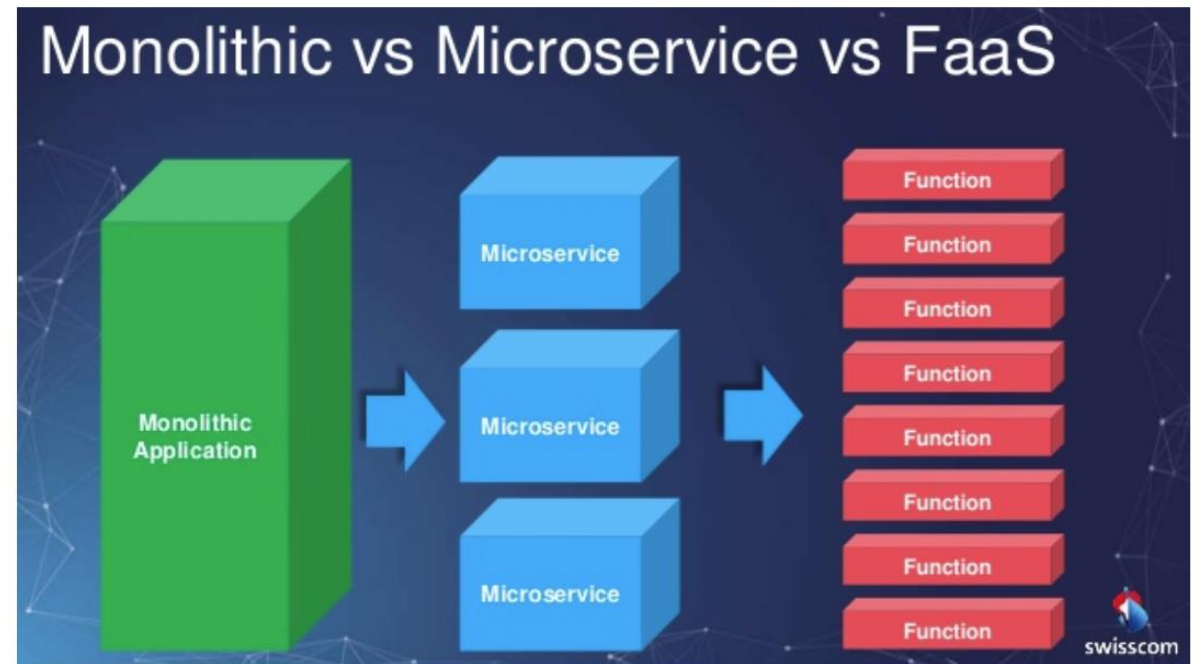
Your options



Function as a Service (FaaS)

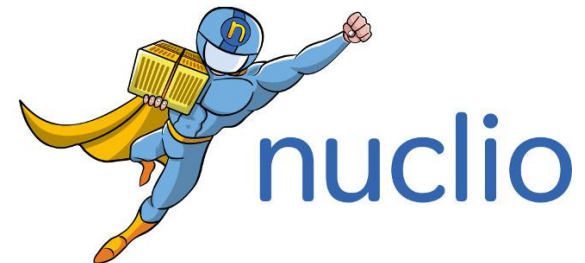


NORDICAPIS.COM



Source: <https://blog.algorithmia.com/serverless-computing/>

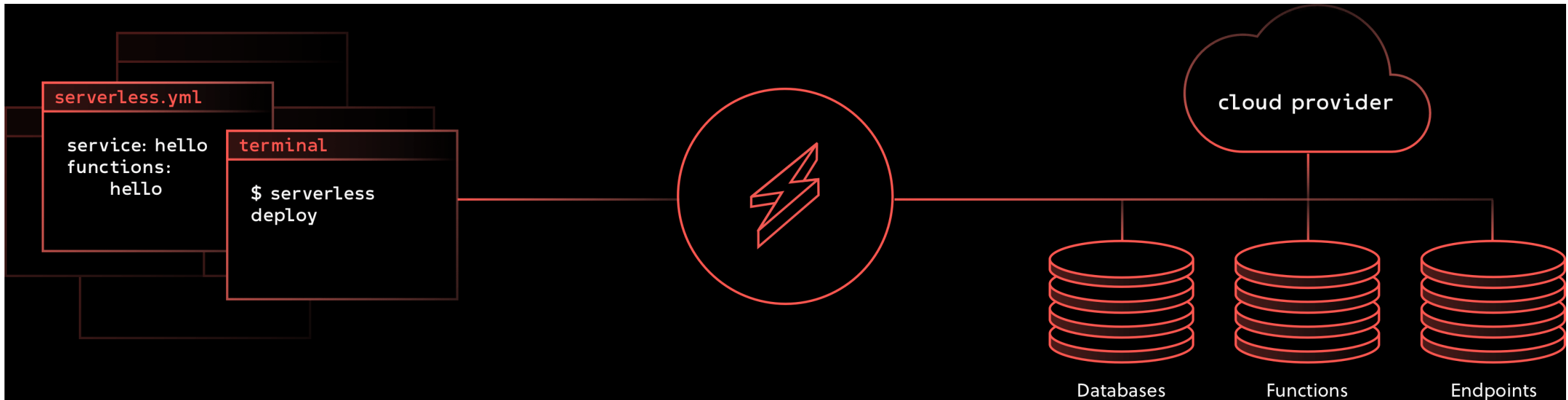
Where can I run FaaS?



Poliglot(ish)



Serverless



Source: <https://serverless.com/framework>

Providers



[AWS Docs](#)



Microsoft Azure

[Azure Functions Docs](#)



IBM OpenWhisk

[OpenWhisk Docs](#)



Google Cloud
Platform

[Cloud Functions Docs](#)



Kubeless

[Kubeless Docs](#)



spotinst

[Spotinst Docs](#)



fn

[Fn Docs](#)

Events



⚡ Events

- API Gateway
- Streams
- S3
- Schedule
- SNS
- SQS
- Alexa Skill
- Alexa Smart Home
- IoT
- CloudWatch Event
- CloudWatch Log
- Cognito User Pool



⚡ Events

- HTTP
- Timer
- Queue Storage
- Service Bus
- Event Hubs
- Blob Storage
- Other Bindings



⚡ Events

- HTTP
- Event

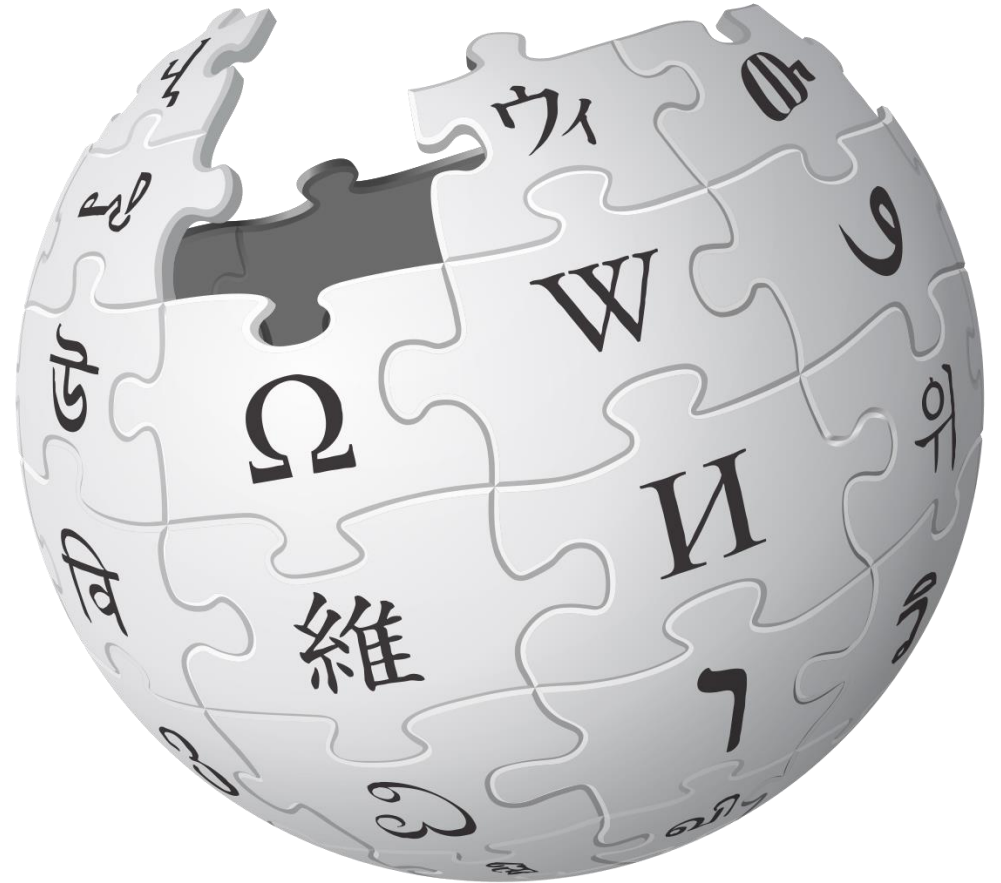
Getting data from the Internets



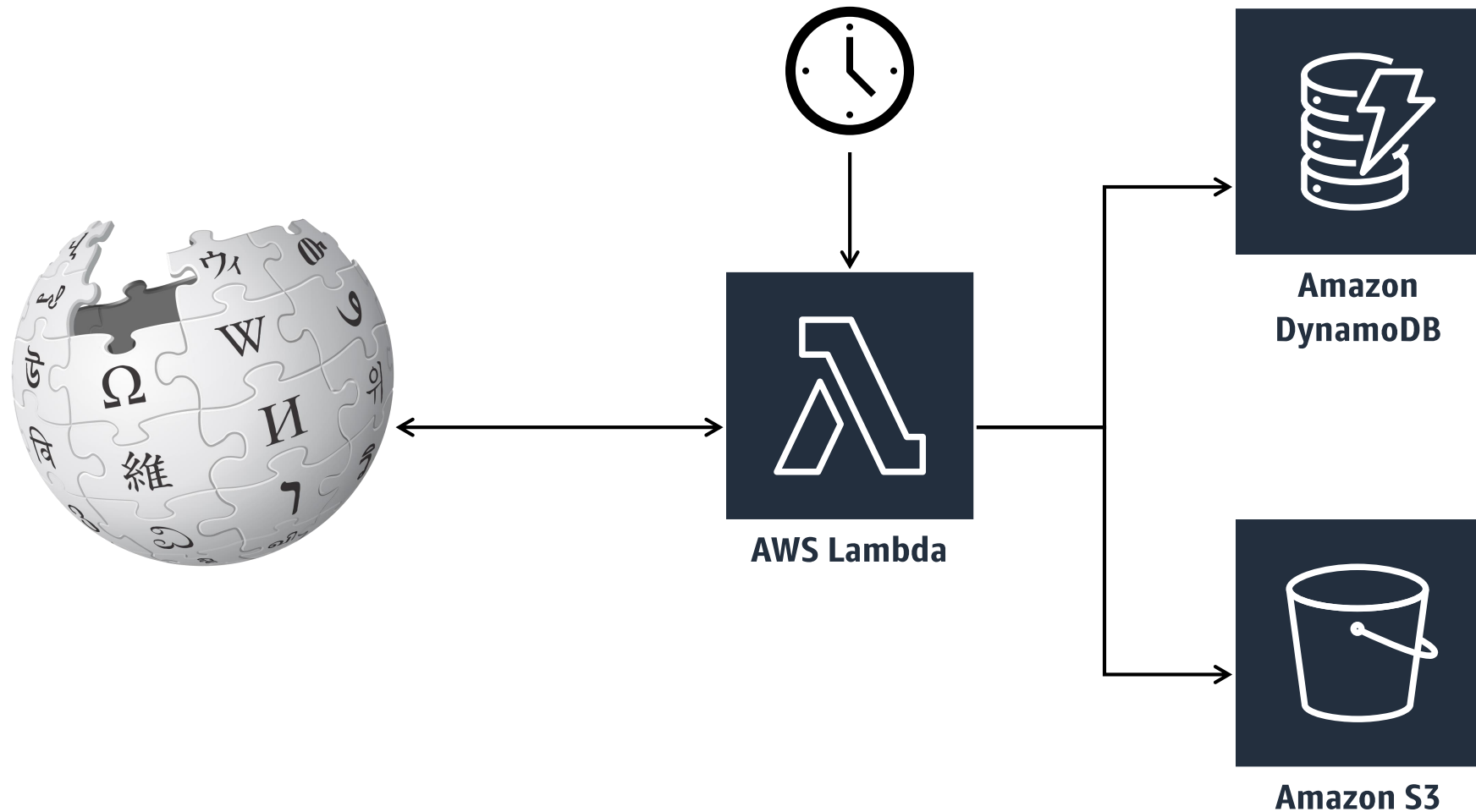
The task

Get event data from Wikipedias main page, in particular news.

Save extracted data to JSON file and a record in a database. Data should be pulled daily.



The task - architecture



Let's do this!



GitHub: <https://github.com/jsnowacki/serverless-wiki-crawler>

Thank you!

Questions?