

DevOps for Data Science (DataOps)

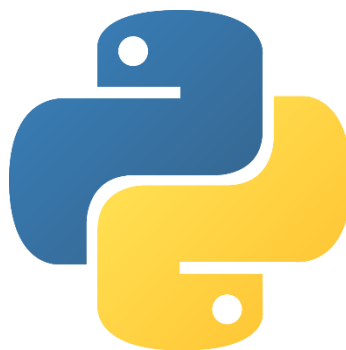
With Azure DevOps, Docker and Kubernetes

Terry McCann | @SQLShark



“Productionisation of models is the ***TOUGHEST*** problem in data science”

But why?



Data Scientist



Highly Academic

Statistically minded

Particularly strong in ML languages

Not from a software background

Thinks testing is hard

Doesn't use source control

Thinks deployments are the
responsibility of another team

Models are not in
production

How do we fix this
problem?

Developer



I write code in Python

I write new features

I fix old code & Bugs

I wait weeks to deploy my code

Time to market for my products is suffering

My customers are **not** happy!

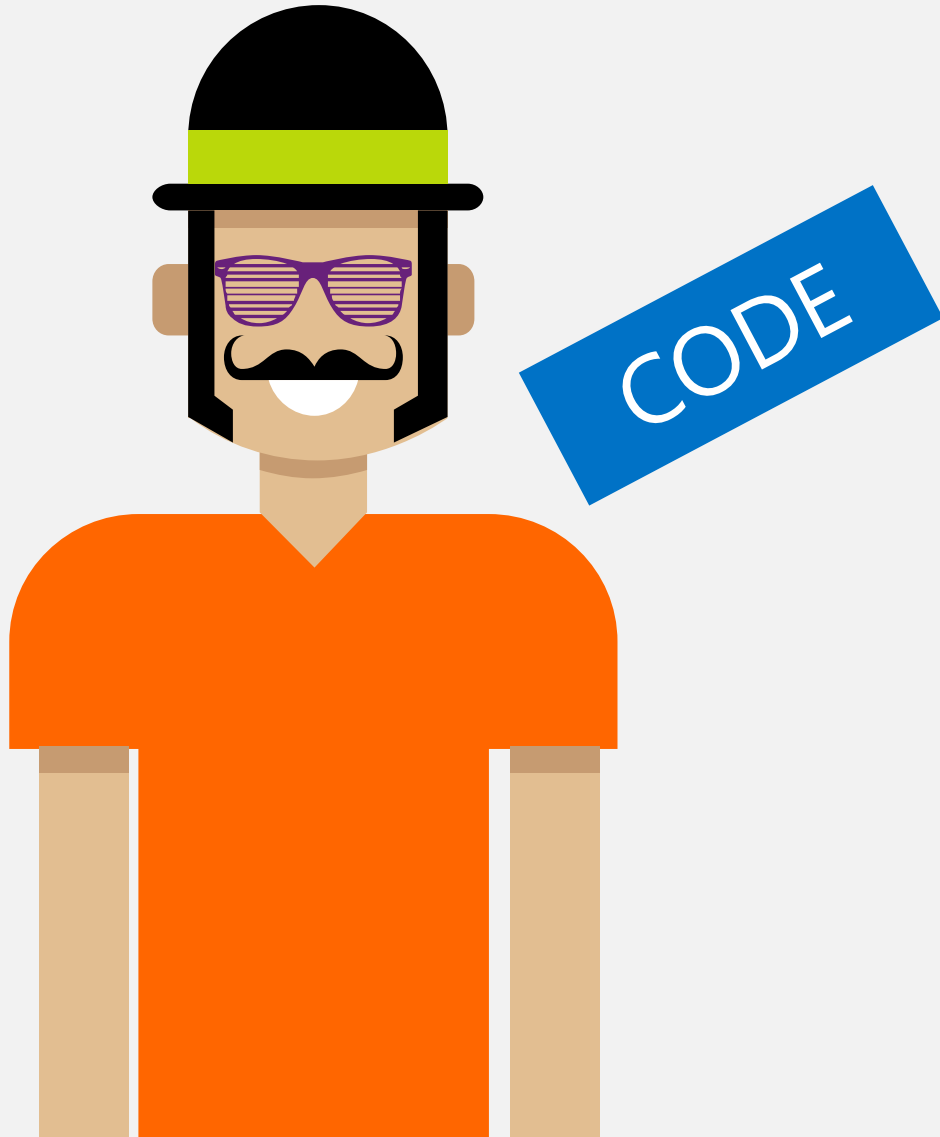
Operations are too slow!

Operations

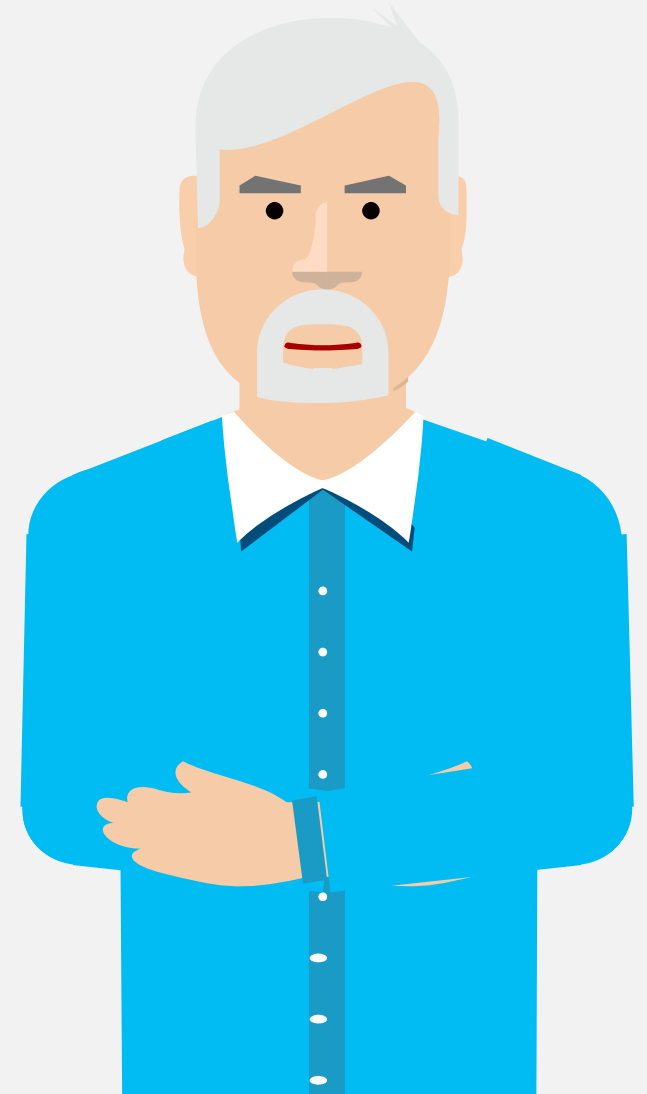
I manage Production code in Java
Our environments != the same
I am responsible for 99.9% uptime
I manage hundreds of servers
Our company is growing fast and
so is my responsibility
I own the bugs from deployments
Developers don't care about uptime!



Developer



Operations



Dev Ops



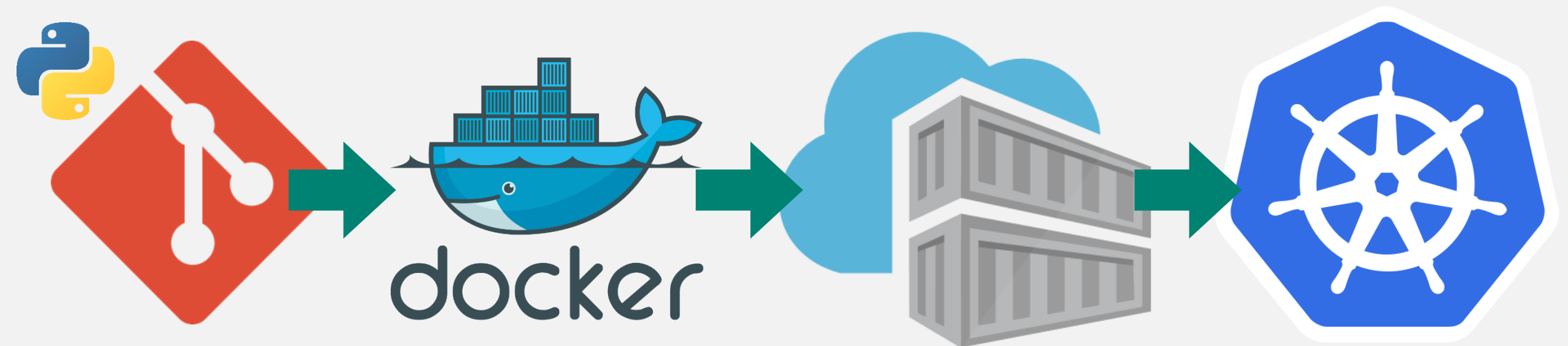
1. Culture
2. Source Control
3. Continuous integration/testing
4. Continuous deployment
5. Infrastructure as code
6. Configuration as code
7. Automation
8. Operational Monitoring / Feedback

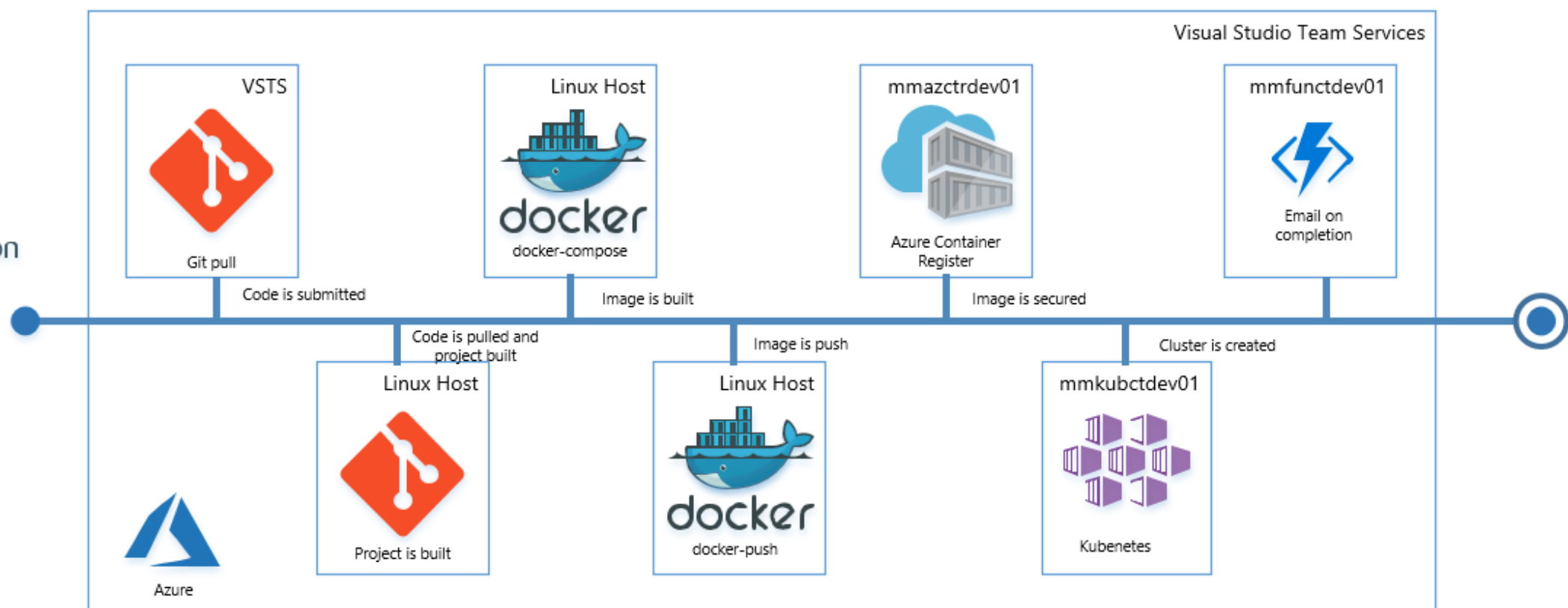
Dev Ops



Lets automate the
deployment of data
science with DevOps

Demo







MICROSOFT MVP, MASTER OF DATA SCIENCE & PRINCIPAL
CONSULTANT AT ADATIS

ABOUT

APPLYING DEVOPS TO DATA SCIENCE

PRESENTATION TIPS

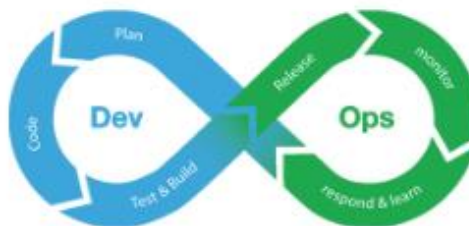
PRESENTATIONS

SOFTWARE DEVELOPMENT

DATA PLATFORM MVP



APPLYING DEVOPS TO DATA SCIENCE



Hi everyone,

Some of you might know that for the last 2 years I was studying a Master's degree in data science from the University of Dundee. This was a 2 years part-time course delivered by Andy Cobley and Mark Whitehorn. This course was fantastic and I recommend it – If you want to know more about the course, please give me a shout. The course is comprised of multiple modules. The final module is a research project, which you need to start thinking about towards the end of the first year of study. I selected my topic very early on, however being indecisive, I changed my idea 3 times (each time having written a good chunk of the project).

Why did I do this? I simply was not passionate about the subject of those projects. They were good ideas, but I just was not researching or building anything new. The outcome of my dissertation might have been a working project, however it would have felt hollow to me. I needed a topic which I was passionate about, and I found it. I chose DevOps because of that, that I have spent much of my career working either with or developing automation tools to accelerate and simplify my development processes. Having attended a lot of conferences, I became familiar with DevOps and how it accelerated the software industry. DevOps allows software developers to ship code faster. I have been applying the core principles of DevOps to all my recent projects, with great success.

The course covered a lot of the techniques required for data science, however it did not cover how to deploy a model in to production. I started researching deployment techniques. I read a lot of books which described the development process, each stopping at

<http://www.hyperbi.co.uk/applying-devops-to-data-science/>