

Hands-on with the AWS ML platform: putting the power of the AWS AI&ML tools in the hands of EVERY researcher, no matter how large the data and training job

Kevin Jorissen¹ and Ben Thurgood²

¹ Amazon Web Services, Seattle, USA, jorissen@amazon.com

² Amazon Web Services, Sydney, Australia, btgood@amazon.com

GENERAL INFORMATION

- Is this workshop half-day or full-day? *Half-day*
- Does the workshop include a hands-on component? *Yes*
- Are there any constraints on the number of attendees? *No, we can accommodate up to hundreds of participants if there is a lot of interest. Amazon will provide the necessary cloud computing resources.*
- Are there any special seating or table requirements (e.g. for breakouts, teams)? *No, except that participants will need to work on a laptop.*
- Are there any technical requirements beyond AV and access to wireless network? *No, but power strips would be nice so people can plug in their laptops.*

DESCRIPTION

1. The AWS stack for ML/AI. Through the lens of a scientist working with research data we give a very brief and conceptual overview of the different ML services in the AWS platform and when you would use them, as well as where you would store your datasets and how you would access them.

(e.g.) 20 minutes

2. Amazon SageMaker: putting ML in the hands of every scientist. SageMaker is an AWS platform where a researcher can execute all steps of the ML cycle: data selection and preparation; model training; and model deployment. It has two remarkable properties: you can access the power of the cloud (there's always space for another PB of data, always another 100 GPUs if you need huge training jobs in order to get accurate, publishable models) – and you can do it all from the familiar comfort of a Jupyter notebook. From building a toy model on your laptop it's a massive increase in power without much of a learning curve.

25 minutes

3. Basic tutorials: From the web browser on your laptop you'll log into an AWS account, open a Jupyter notebook on an AWS notebook server, and step through the notebook. In the first exercises we do simple things like deploy a TensorFlow notebook, or scale up a training job from a single server to a cluster for speedup.

60 minutes

4. Scientific tutorials: We'll work through a genomics notebook where you pull a slice of genomics data from a massive, real-world data lake available on AWS (the "1000 genomes" public dataset), prep the data, then kick off a training job

and analyze the results. Additional tutorials will be available for fast learners. AWS support will be available in the classroom to help participants. You can also bounce your ideas off them for applications in your own lab or classroom.

60 minutes

WHO SHOULD ATTEND

The skills we teach will be useful to a wide range of practitioners: researchers, educators, research/IT support staff, research software engineers, data scientists, and more.

WHAT TO BRING

Attendees need to bring their own laptop of any kind. Workshop activities will take place on cloud computing resources accessed in a web browser. Attendees should be familiar with the basic functionality of a Jupiter notebook. Some basic familiarity with python will be helpful but is not required. The workshop content will be available in English only.