Data Scientists spend 80% of their time gathering, formatting, and processing their data and only 20% of their time innovating. It should be the other way around.

**The Problem**

Modern data science uses a ton of computing power, and with AWS and Azure it’s cheaper than ever to rent processing power.

The problem is that it’s not easy to use. Right now, data scientists waste too much time dealing with logistics when they should spend most of their time solving problems.

**Project Goals**

**Goal:** Design an application that builds and deploys Azure Cloud resources automatically.

**Requirements:**
- Automatically create Azure Cloud resources and deploy jobs to them.
- Store results and log them in a database.
- Expose an API so that developers can easily create new jobs.
- Build a User Interface so non-developers can create new jobs and monitor existing jobs.
- Deploy a demo application using “One-Click” Infrastructure as Code

**Design**

- The application is built in Go
- The application *sits idle* until jobs arrive to save costs.
- Once a job arrives, VMs are automatically allocated to hold as many processing tasks as possible.
- It can deploy Docker Containers from any repository - users just need to provide the location.

**Conclusions**

- We have a working demo
- The demo has been tested with sample workloads and meets the requirements

**Next Steps:**
- Work with an industry contact to test large-scale workloads! (we can’t afford to do this on our own)

**Demo**

- It constantly monitors running jobs and saves all of their log messages.
- The program is *reentrant* - even if it crashes, it can reconnect to running jobs and see them to completion.
- The project includes both a REST API and a web app built with React
- The demo was deployed with our Infrastructure as Code and is running well!!

**References:**