data iku × snowflake

The ultimate shortcut to data science products

Lea Fleckner (Dataiku)

Daniel Mannino (Snowflake)

Today's speakers

Lea
FLECKNER

Sales Engineer
Dataiku

Daniel
MANNINO

Principal Sales Engineer
Snowflake

Imagine…

… you have a new project to get started on.

Your dataset

250 TB + 30TB

Your machine

You

What your business teams are expecting as an end result:

Python on your local machine

Sounds great because…

- You only need your laptop!
- You can install whatever packages you want (ideally)
- Use the IDE you love most

… however

- Data access
- Memory limited and compute single threaded
- Data must be copied and ingestion is slow
- IT security
- Documentation and collaboration setup
- Doesn't scale beyond quick ad hoc use cases
- No operationalisation

Python in the cloud

Sounds great because…

You can watch a movie!

… however

- Data access + **cloud setup**
- Memory is **still** limited and compute is **still** single threaded
- Data must be copied but **better** ingestion speed
- IT security + **cloud security** topics
- Documentation and collaboration
- Potentially operationalizable
- **Frontend?**

Spark cluster

Sounds great because…

- It scales!

… however

- Data access
- Can’t work in Python anymore
- Data must be copied
- Data distribution to avoid shuffles
- Complex configuration
- IT security
- Documentation and collaboration
- Model serving and monitoring
- **Frontend?!**

By now months have gone by and your project is still not live.

So what's the point?

Three pieces of advice…

Don't try to solve everything in one language | machine.

Depending on the source of your data, the size of it and your intended operation you should be selecting the engine & language to go with that.

Work in an environment that can be a vehicle for operationalisation.

By accessing source data directly and modelling for it you will be able to move into production much faster.

Push the computation to the data, not vice versa.

Copying data around means incurring costs, wasting time, and taking risks.

Thank You

dataiku.com | contact@dataiku.com