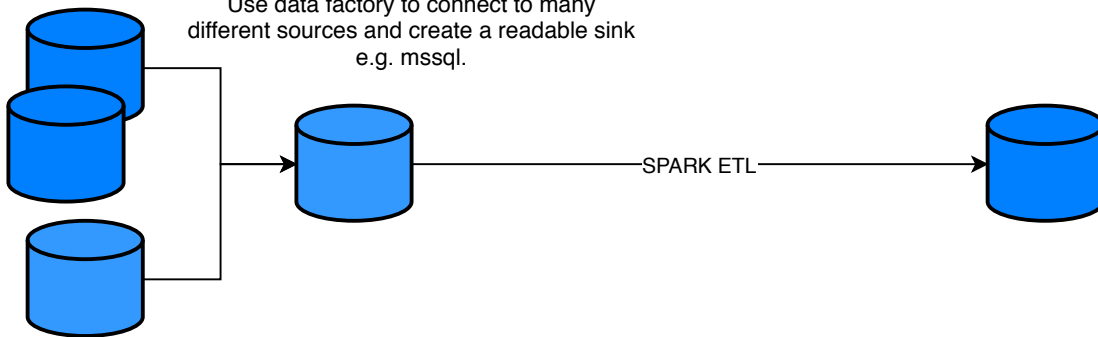


Use data factory to connect to many different sources and create a readable sink e.g. mssql.



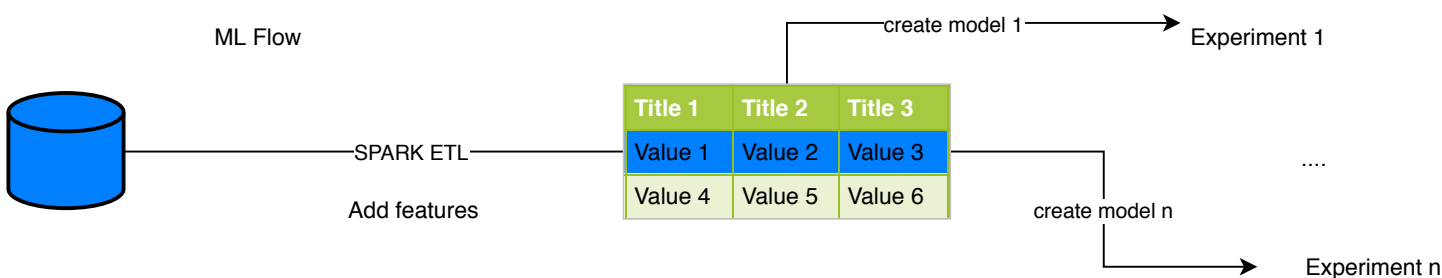
DataBricks Delta is ideal for the daily loads.

Downside is that fewer tools can connect to the table as it is no longer .parquet

If the sources are all easily connectable we could skip data factory



ML Flow

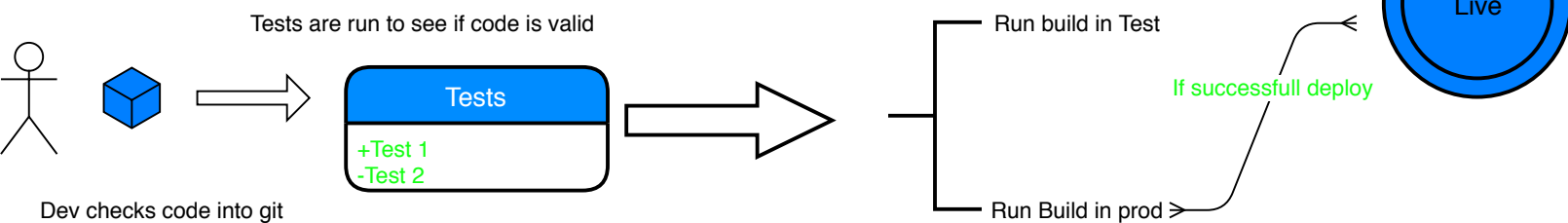


Add features

Use ML flow to note the metrics from different runs of your model.

Azure dev ops:

- Git
- Hooks for tests
- Deployment
- connect to or act as scheduling agent
- Has build in trello



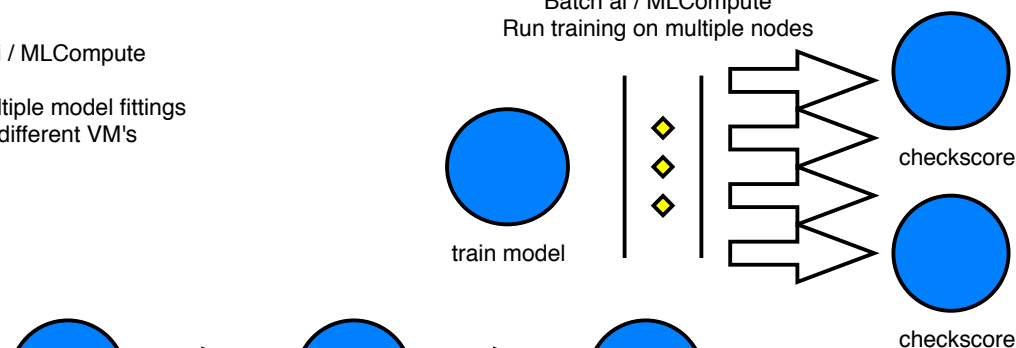
Tests are run to see if code is valid

Dev checks code into git

Batch ai / MLCompute

Run multiple model fittings on different VM's

Batch ai / MLCompute
Run training on multiple nodes



Easy to mirror this with python + onprem cluster.

Sparkcontexts might also be created in Databricks

