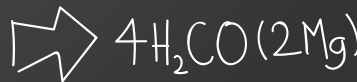
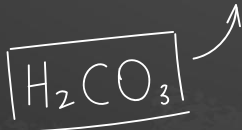
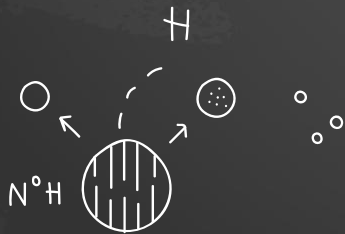
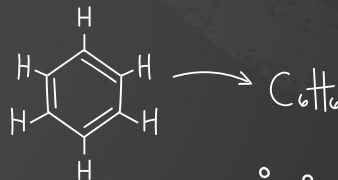
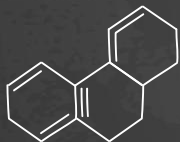


Day 1



# Landschap en tools

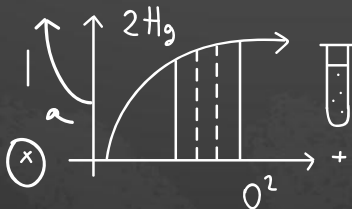
Hackaton





## Tools gebruikt in deze hackaton

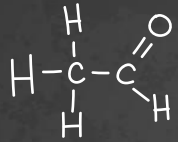
- Microsoft Azure
- Databricks
- MLflow
- R-Studio



databricks®

mlflow

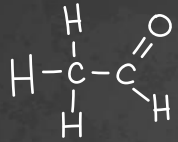




Microsoft Azure is een **platform**  
waarbinnen een reeks aan  
**cloudservices** wordt geboden.

Denk hierbij aan rekenkracht,  
analytics, opslag en netwerken.

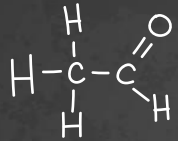




Binnen Azure kun je **applicaties**  
deployen met een theoretisch  
**ongelimiteerd** vermogen aan  
**rekenkracht.**

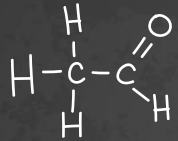
*En dat is wat we nodig hebben bij ML*





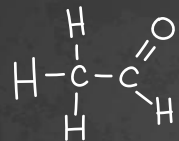
Azure neemt een hoop uit handen  
zodat een bedrijf en/of ontwikkelaar  
zich kan **focussen** op het  
**beschikbaar** stellen van een  
**applicatie**.





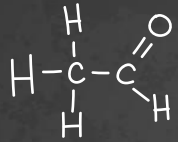
Het is relatief **eenvoudig** om  
machines te starten waar  
**automatische** updates plaats  
vinden, basis **security** is ingeregeld  
en zich bevindt binnen een eigen  
**afgeschermd** (private) omgeving.





Infrastructure &  
Platform  
**As A Service**





**" Ok, leuk dat Eezjeur....  
Maar wat moeten wij er  
dan mee in de praktijk? "**



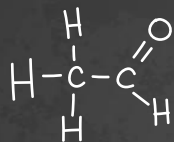


# Azure, voor onze doeleinden

*Wij zouden er het volgende mee kunnen:*

- Experiment tracking
- Sneller trainen in de cloud (compute)
- Model store
- Feature store
- Data warehousing / virtualisatie
- Automatisering van deployments
- API's (publiceren van endpoints)
- Logging
- Dashboards / monitoring
- Samenwerkings tools
- Etc.

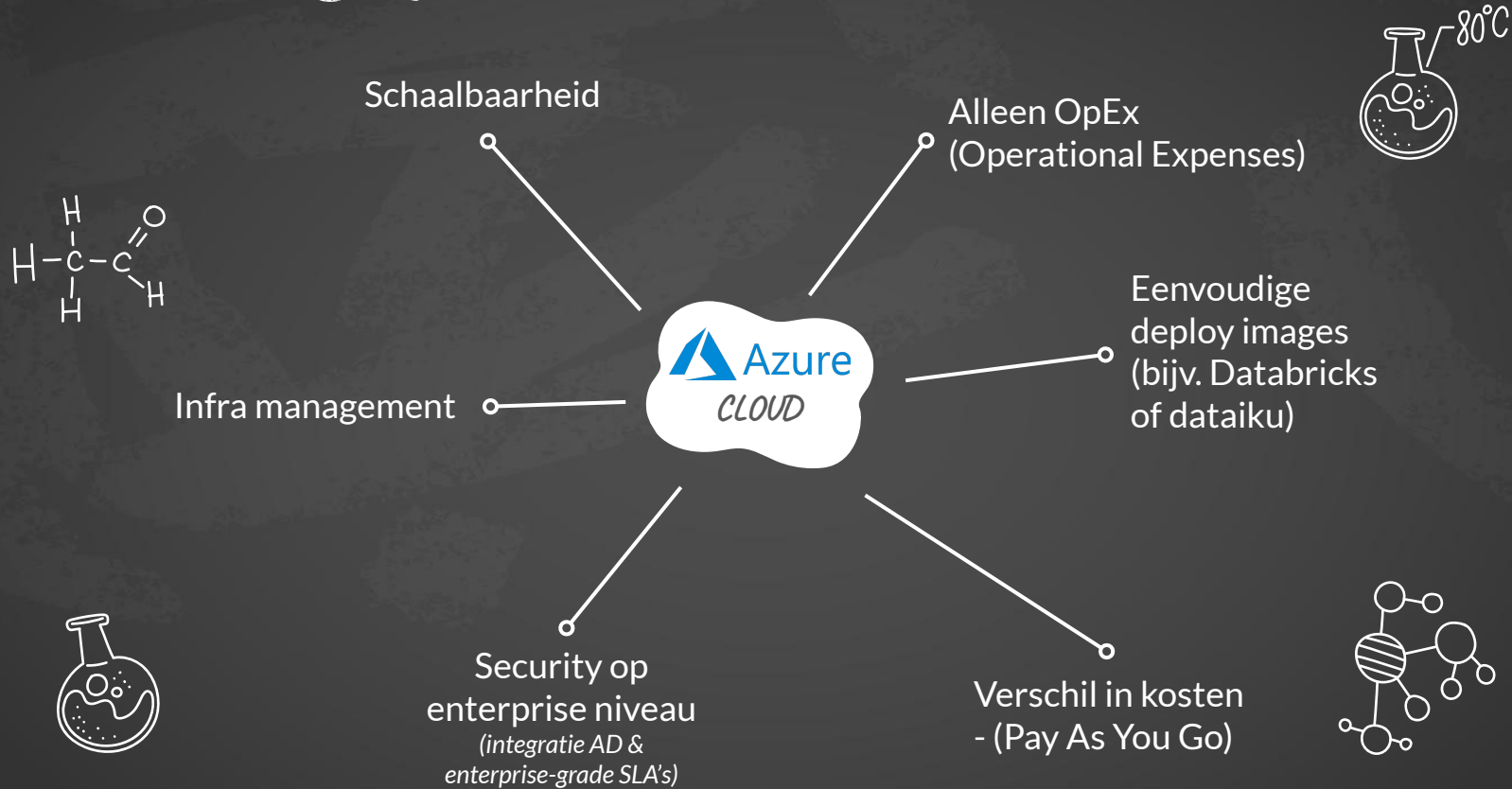
» en.... deployen Databricks MLFlow



**Waarom niet gewoon  
on-prem (op de zaak)**



# Belangrijkste voordelen van de cloud



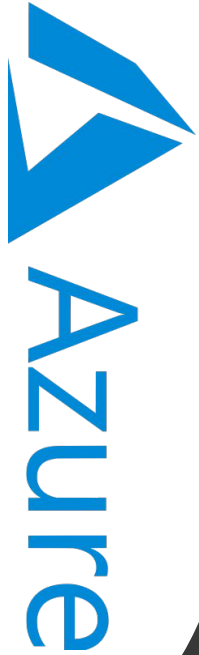
# wat is Databricks?



Databricks is een open Source platform waarop data uit alle verschillende bronnen worden samengebracht, zodat data-engineers en data-analisten deze kunnen bewerken en er modellen en algoritmes op los kunnen laten.



# Databricks tech...

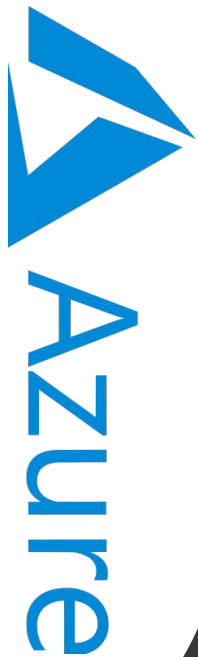


Databricks maakt gebruik van Apache Spark. Spark is een hele snelle unified analytics engine voor big data en machine learning. Spark is ontwikkeld om grote hoeveelheden data gedistribueerd (dus verdeeld over meerdere machines) te verwerken. Databricks is de workspace waarin je vervolgens deze data bijvoorbeeld kunt bewerken, door middel van coderen.



databricks®

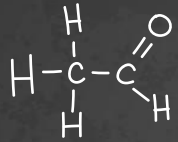
# Azure & Databricks



Met Azure Databricks kun je heel snel die workspace inrichten om met big data aan de slag te gaan, kun je gemakkelijk op- of afschalen en werk je in een veilige omgeving. Je omgeving is bovendien gemakkelijk te koppelen aan Azure Datalake en bijvoorbeeld aan Power BI.



databricks®

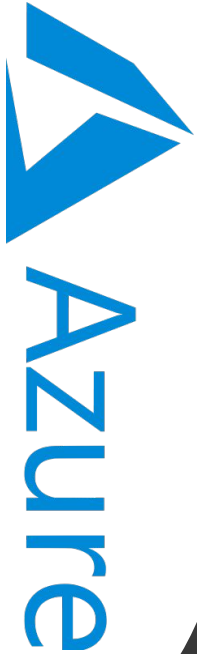


**" En wat moeten wij dan  
vandaag met  
Databricks? "**



# MLFlow

Binnen Databricks draait MLFlow. Dit is een open-Source library voor het **tracken, beheren en deployen van Machine Learning modellen**. Omdat het op Azure draait hebben we de mogelijkheid tot een enterprise-grade architectuur.





# mlflow

A diagram with the mlflow logo at the top, connected by lines to five feature boxes below. The boxes are arranged horizontally and contain the following text: 'High performance trainen in de cloud', 'In de cloud tracking van experimenten, runs, metrics en configuratie', 'Run reproduceerbaarheid', 'Model Store', and 'Eenvoudige deploy van algoritmes'.

High performance  
trainen in de  
cloud

In de cloud  
tracking van  
experimenten,  
runs, metrics en  
configuratie

Run  
reproduceerbaar-  
heid

Model Store

Eenvoudige  
deploy van  
algoritmes