

Data Analytics End-2-End

Datenorientierte Lösungen Abteilungsweit Entwickeln

Twitter: @kein_mark

Github: <https://github.com/crazzle>

Vagrant

Erlaubt einfache und reproduzierbare Erstellung von
Entwicklungsumgebungen

Provisionierung auf Basis von Virtual Box, VMWare, AWS ...

Nutzung von Tools wie Chef, Puppet, Shell-Skripte

Docker

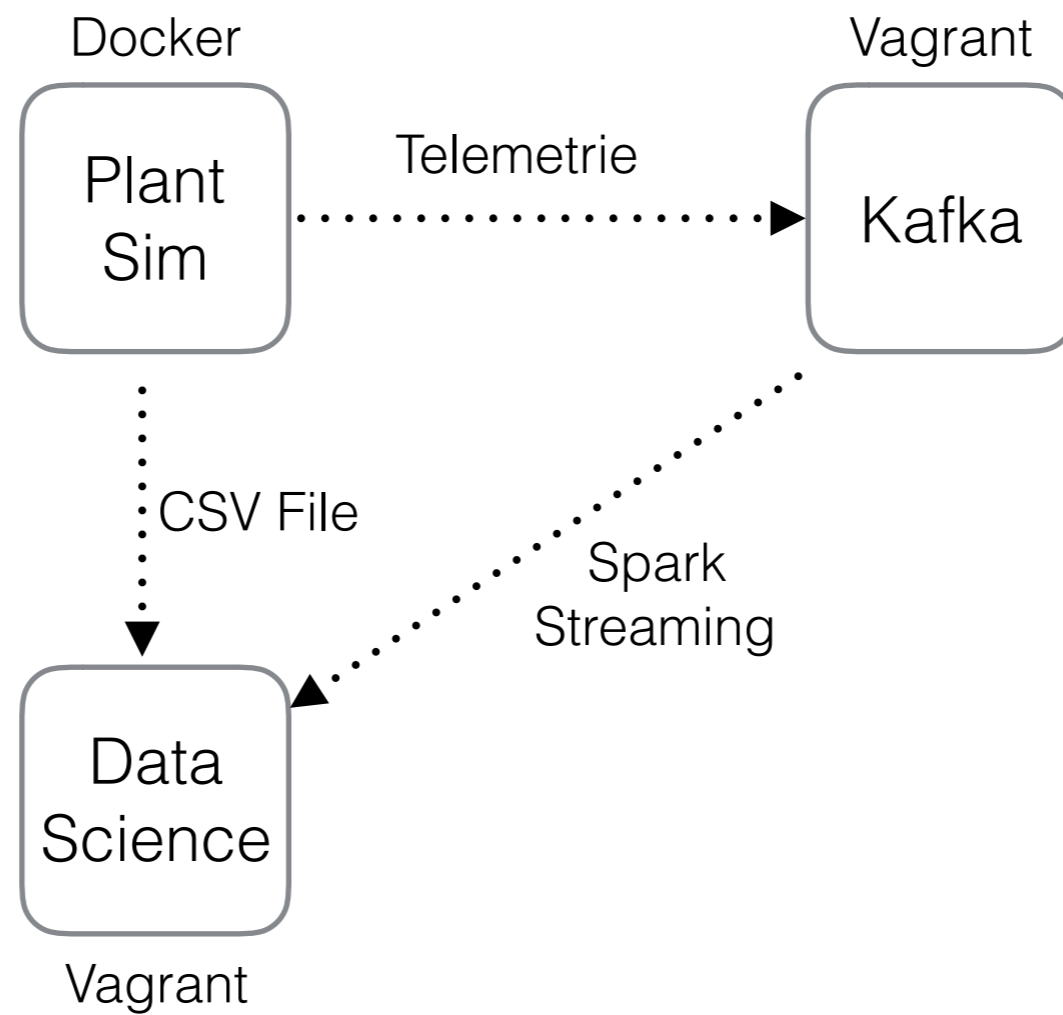
Docker ist eine Plattform zur Virtualisierung in Containern

Container sind Software auf einem Filesystem

Nutzen Kernel des Hosts

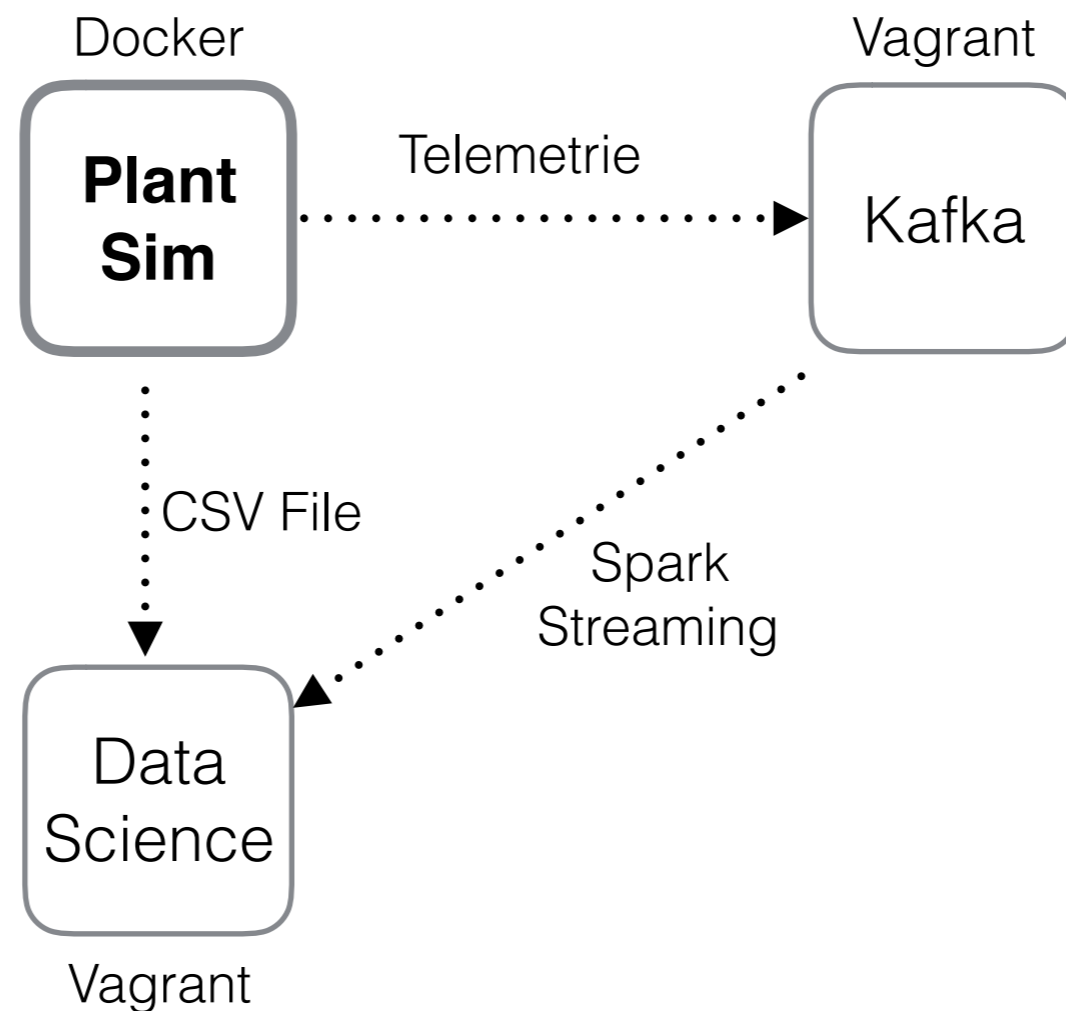
Sind Prozesse auf dem Host

Unser Setup

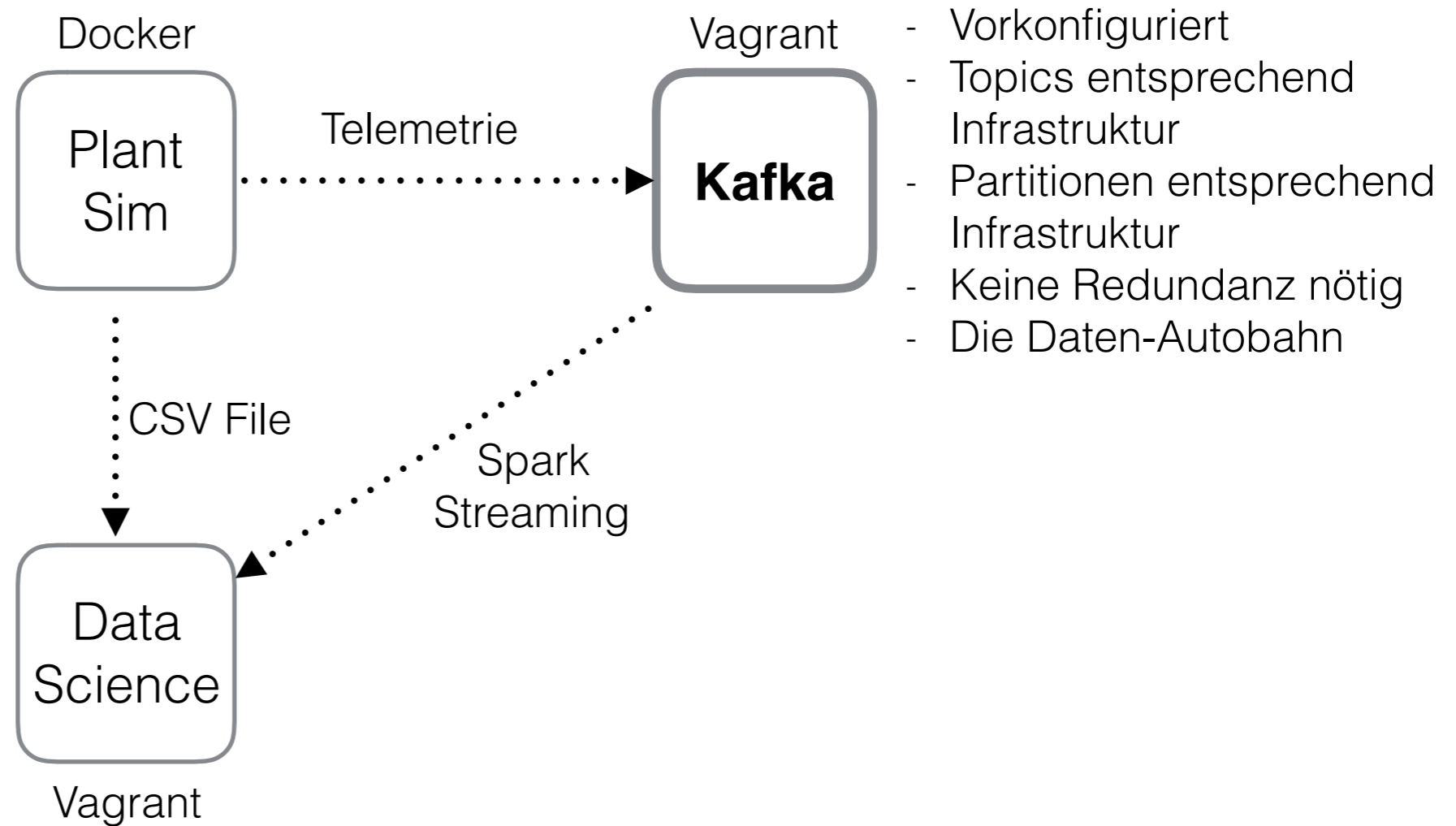


Kraftwerke simulieren

- Python für "einfaches Hacken"
- Von außen steuerbar
- Plattformunabhängig
- In Entwicklungsprozess integriert
- Der Daten-Generator

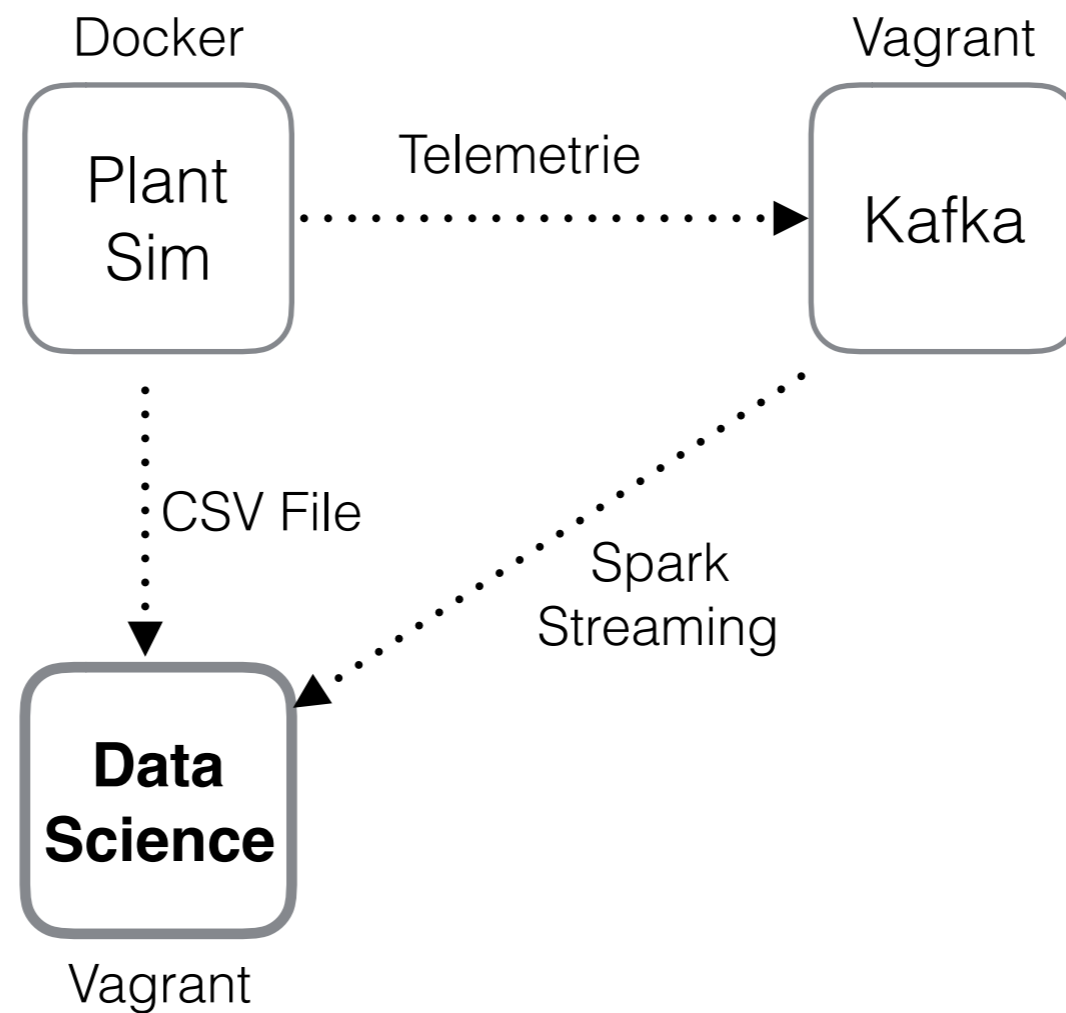


Datenströme aufbauen



Daten Nutzen

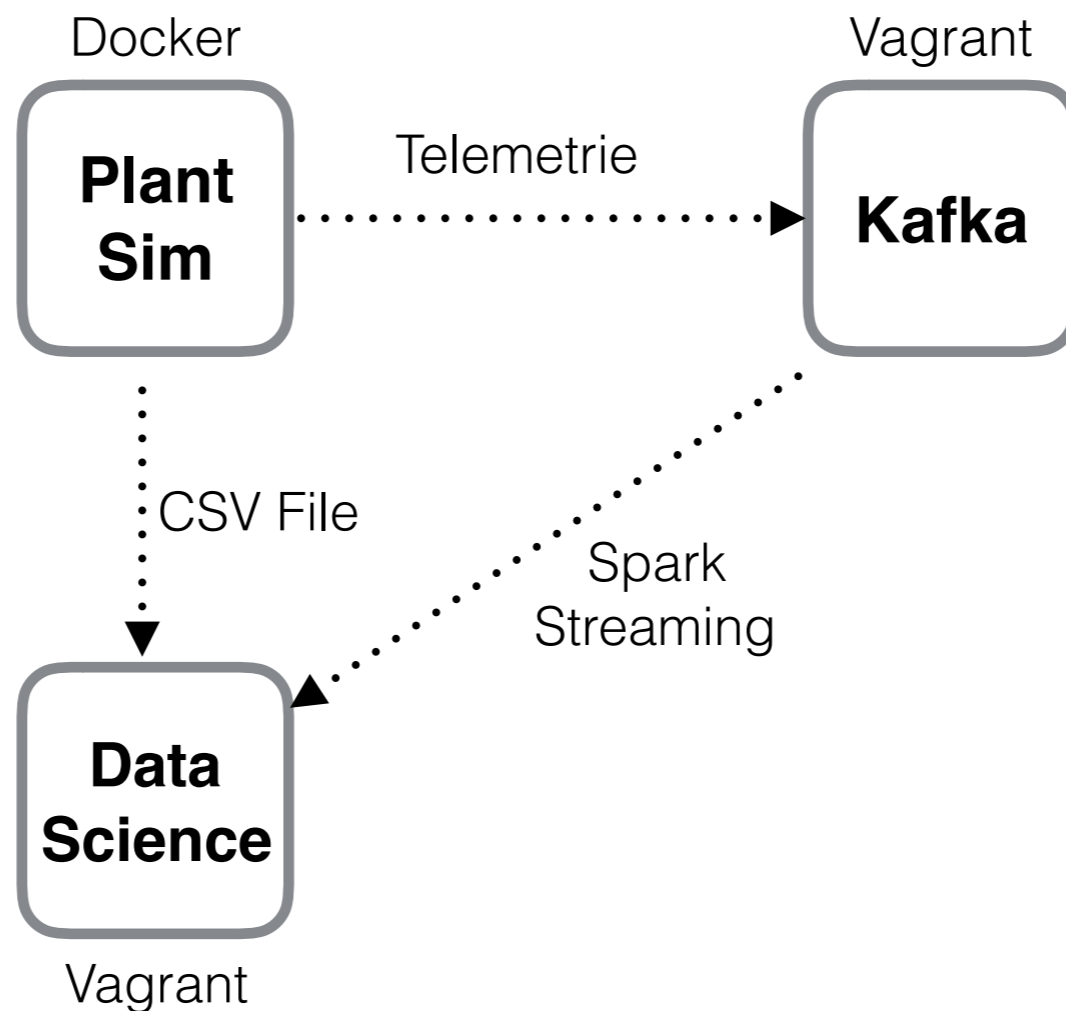
- Großes Repertoire an Tools
- Apache Spark
entsprechend Infrastruktur
- Jupyter Notebooks
- In Entwicklungsprozess integriert
- Der Daten-Transformator



Das große Ganze

- Ein wenig Development für den Fachbereich
- Immer auf dem neuesten Stand
- Direktes Feedback in Form von Code
- Business "nah" am Code

- Analysten werden zu Data-Scientists
- Vorgefertigter "Glue Code"
- Jupyter Notebooks können von allen genutzt werden
- Direktes Feedback in Form von Code



- Kann Blackbox sein
- Insights für Developer
- Entspricht einer Integrationsumgebung