

DataOps as a Service

<https://github.com/vmware/versatile-data-kit>

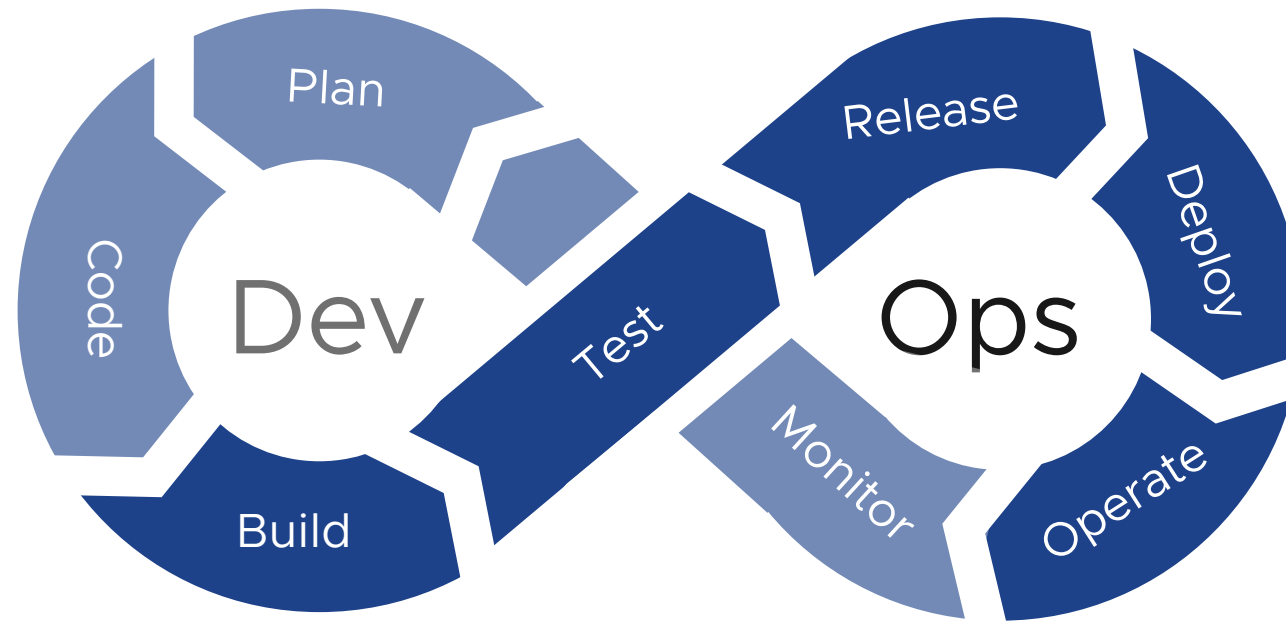
Antoni Ivanov

DataOps Promise

Single Goal: ensure the most efficient creation of business value from big data

DevOps

Single Goal: ensure the most efficient software development of an idea to reality



Why do we need DataOps ?

Infra/Ops Team and Data Teams tension

DevOps & Infrastructure knowledge

Maintain infrastructure

Optimizes reliability, availability and security



Infra &
Operations
Team

Why do we need DataOps ?

Infra/Ops Team and Data Teams tension

Domain knowledge

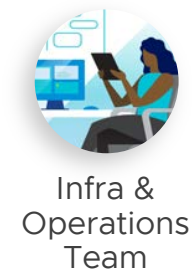
Implement business logic

Optimizes for agility and speed

DevOps & Infrastructure knowledge

Maintain infrastructure

Optimizes reliability, availability and security



Why do we need DataOps ?

Infra/Ops Team and Data Teams tension

Inefficient Operations Stalled development.

Domain knowledge

Implement business logic

Optimizes for agility and speed



Data Team



Data Team



Data Team



Data Team



Data Team



DevOps & Infrastructure knowledge

Maintain infrastructure

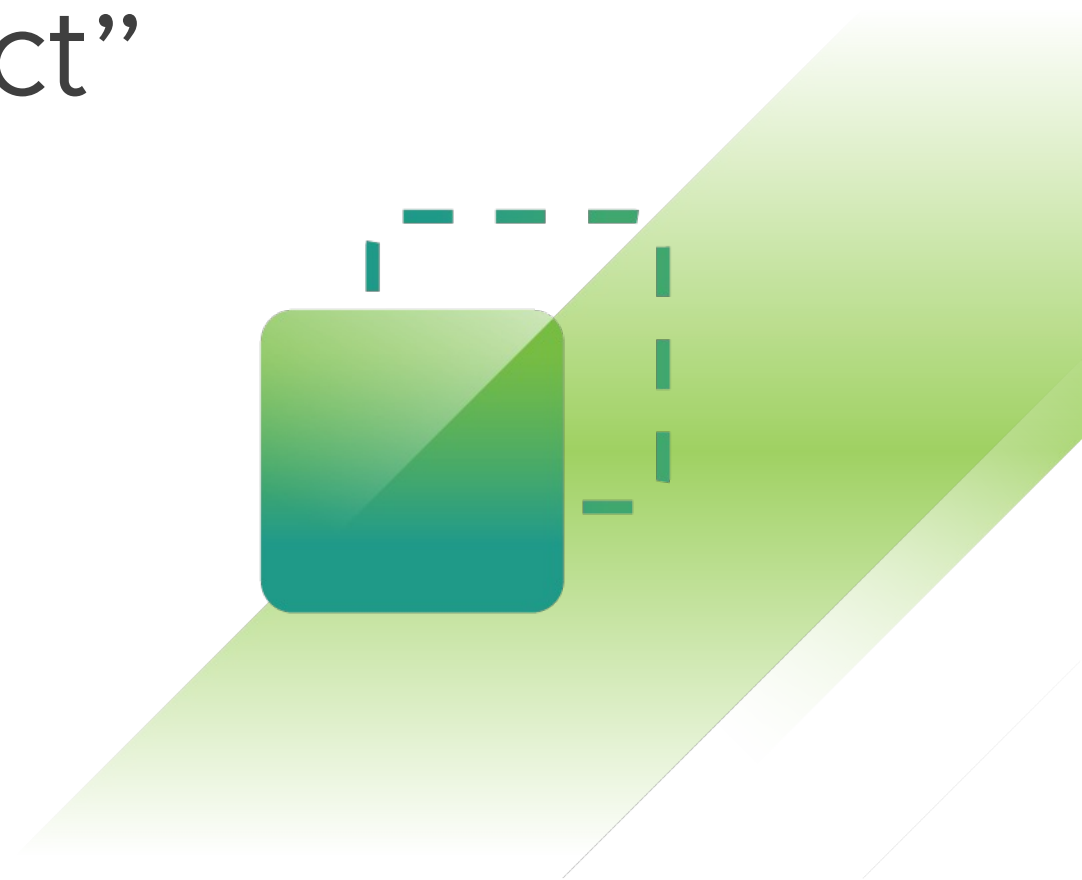
Optimizes reliability, availability and security



Infra &
Operations
Team

Blurred lines of responsibility

”Treat data as a product”

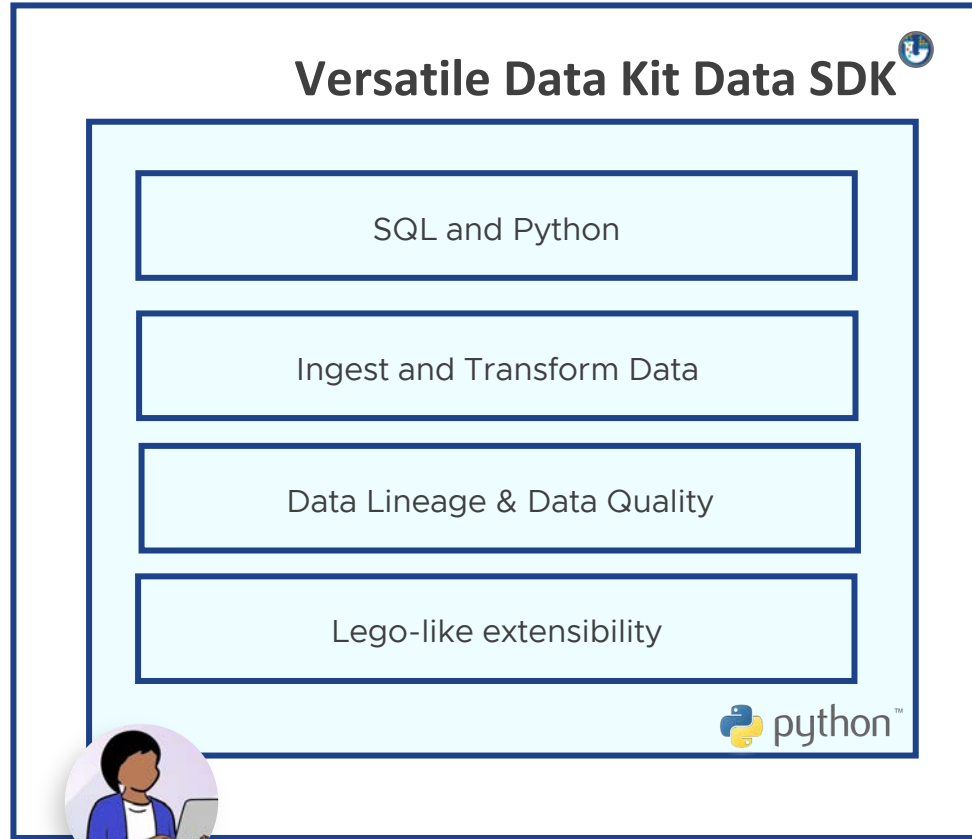




Versatile Data Kit

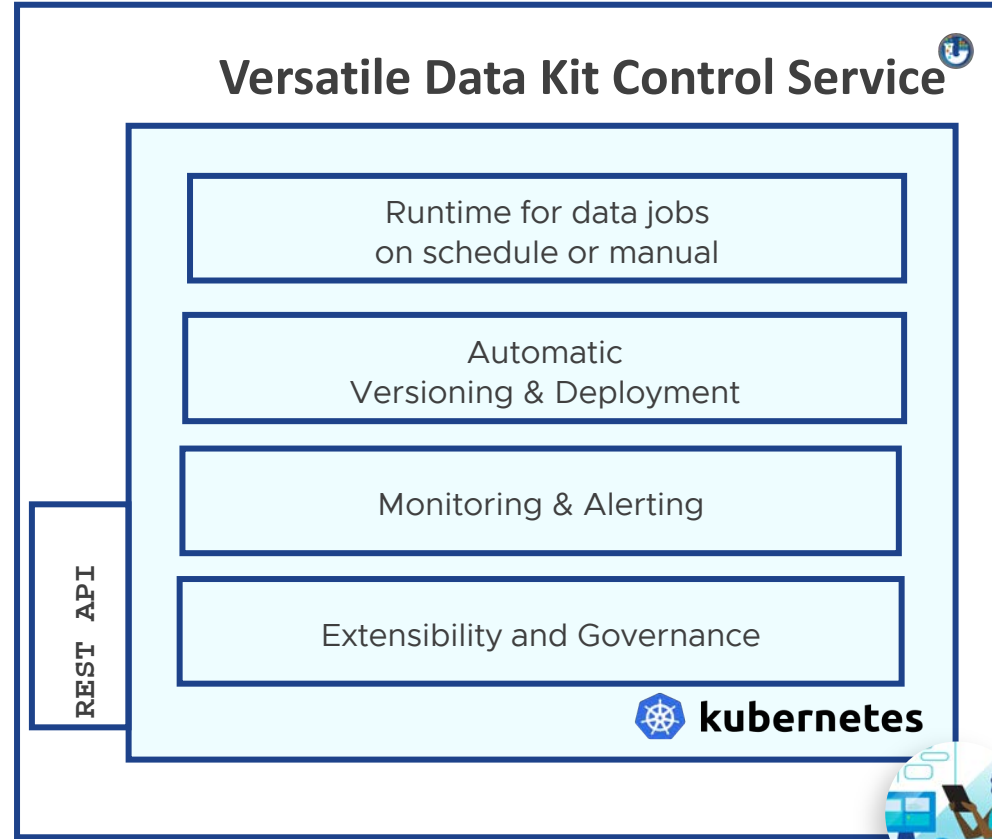
What does it take to run Versatile Data Kit and start deploying data jobs?

Automate and abstract the Data Journey



Data Teams

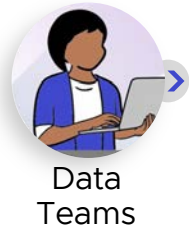
Automate and abstract the DataOps Cycle



Infra & Operations Team

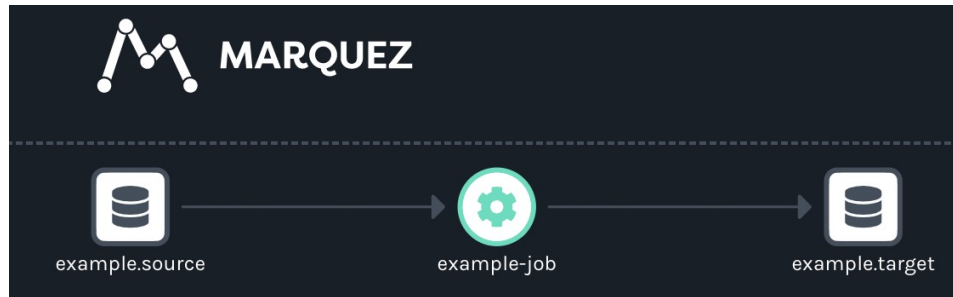
Automate and abstract data journey example

 `vdk query -q "select..." job_input.execute_query(...) vdk run sql_file.sql`

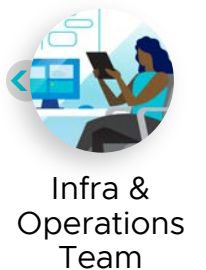


```
insert into example.target  
select ...  
from example.source
```

Managed SQL

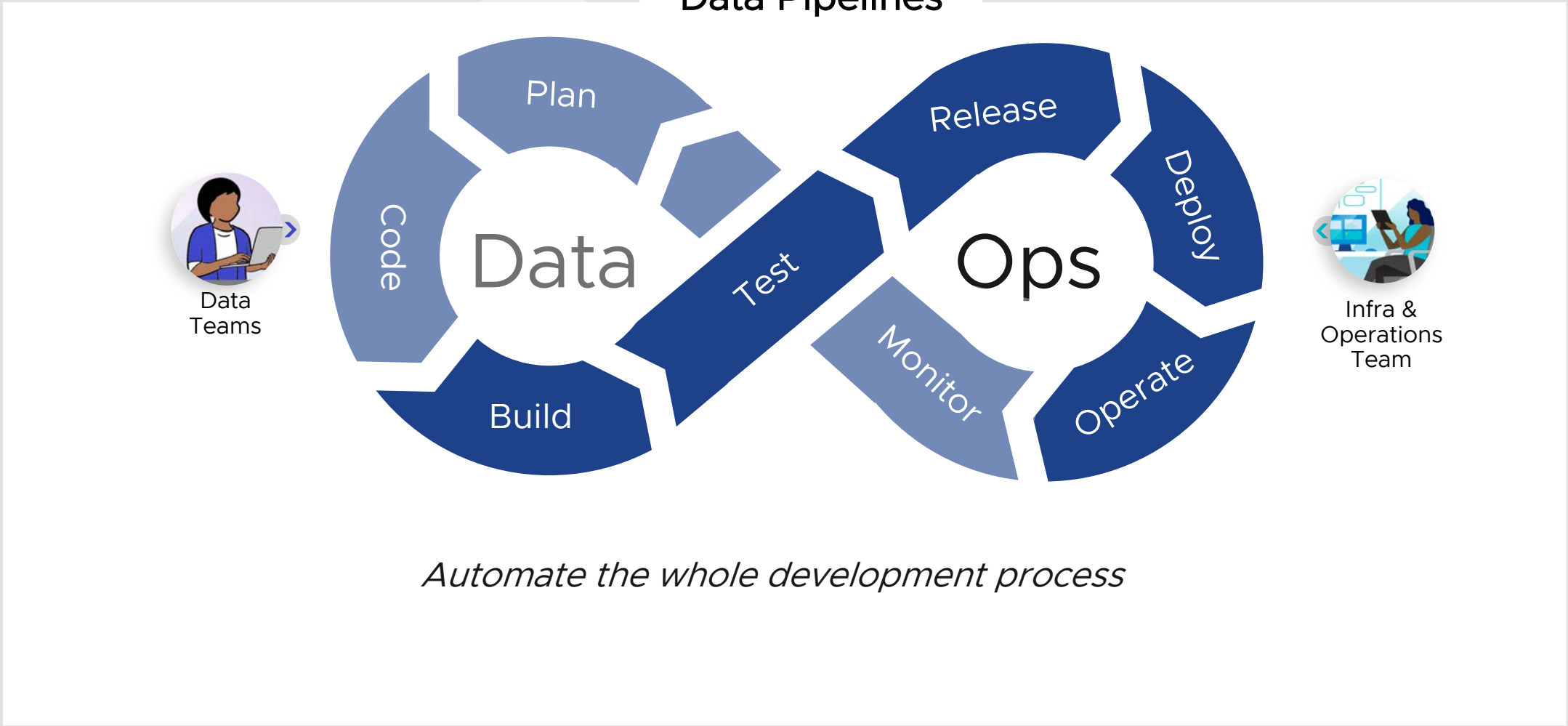


Open Lineage



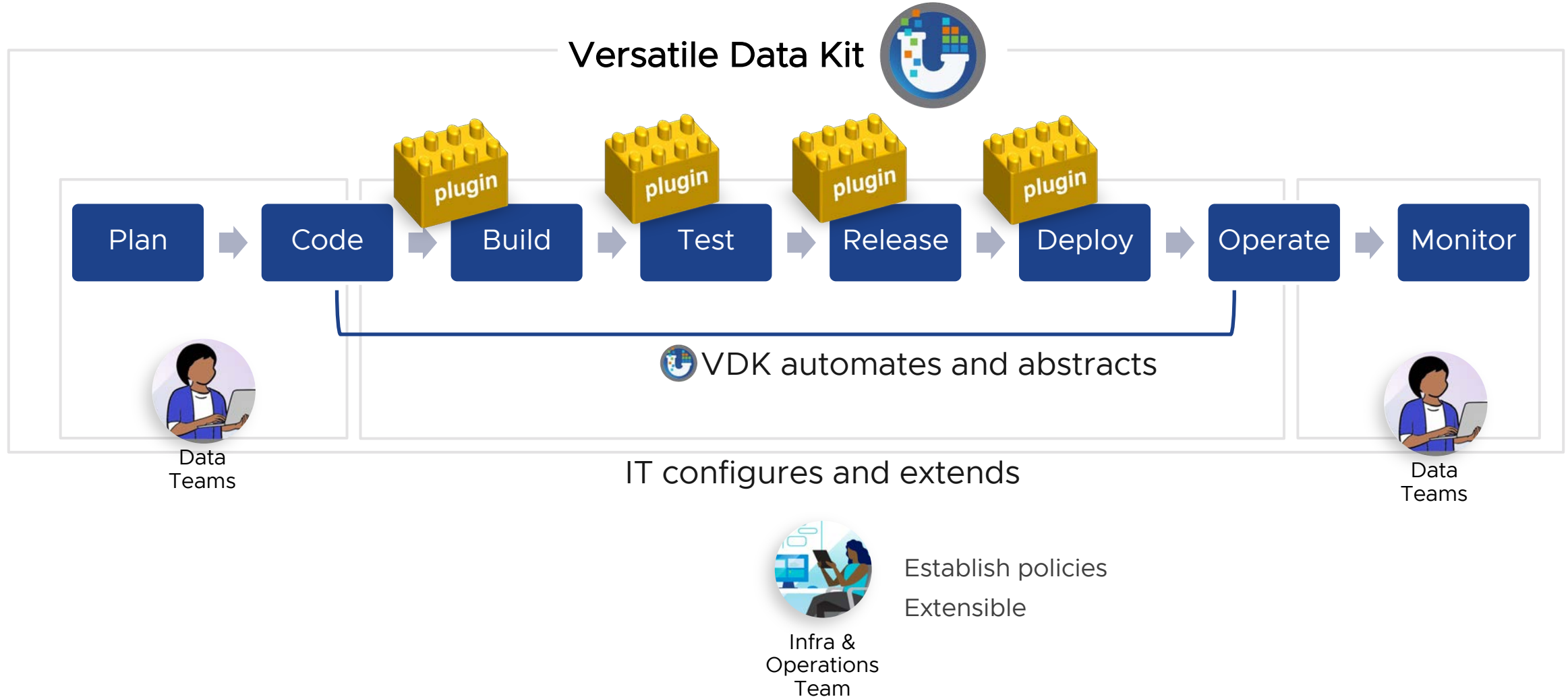
DevOps Cycle

Data Pipelines



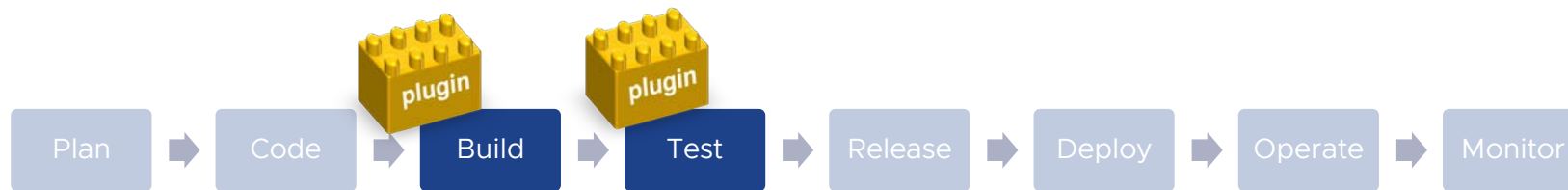
Automate and Abstract the Development Process

Give power to Operators to establish best dev practices; Ease data job development



Quick example: DevOps Plugin

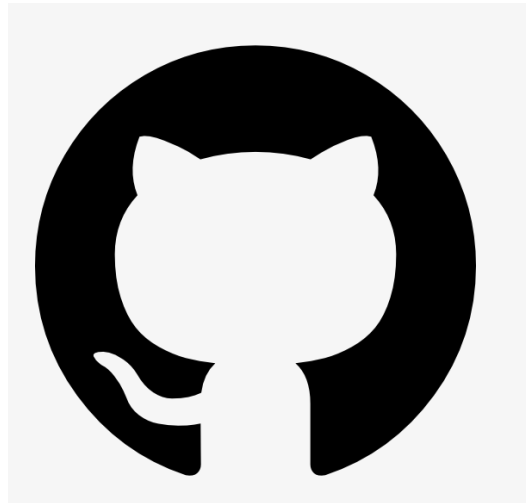
Establish standard system tests and security hardening



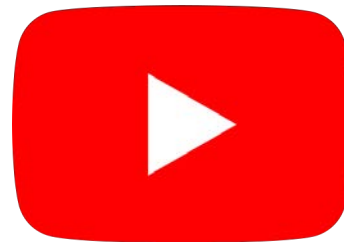
```
helm install --set job-builder=my-job-builder-image
```

```
2 >> FROM versatiledatakit/job-builder
3
4 # Run system test before accepting the new job code
5 RUN pytest system_test.py || die 'Failed system test'
6
7 # Remove execution privileges from files during container build
8 RUN chmod -R -x $job_name/
9
```

Search & Test Versatile Data Kit



Raise an issue in
GitHub if you have any
questions!



Search & contact us

<https://github.com/vmware/versatile-data-kit/#contacts>

Thank you!



If you liked the talk

follow us on GitHub