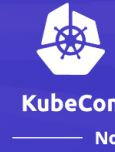


Cloud-native machine learning systems at Day Two and beyond



Sophie Watson • @sophwats
William Benton • @willb

Forecast



Virtual

North America 2020

The basics of machine learning systems

Special challenges of reproducibility

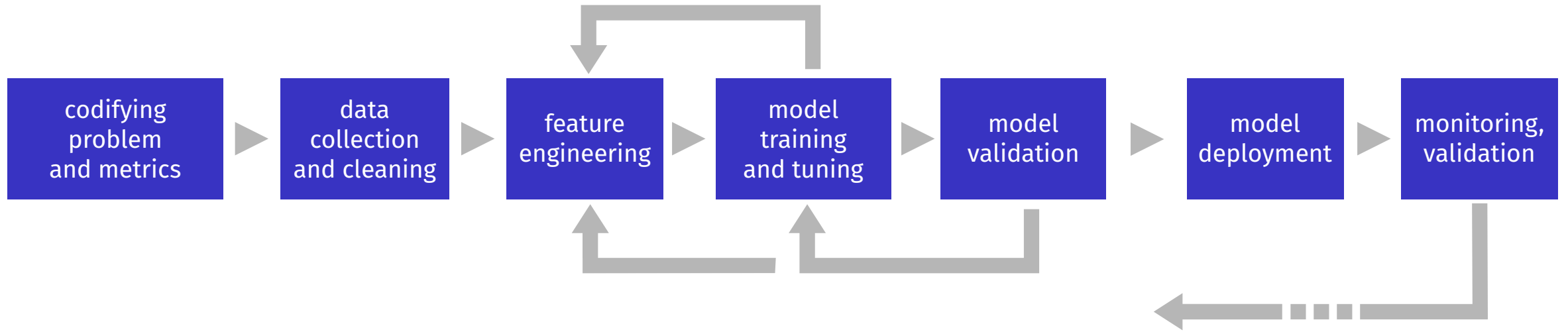
Supporting high-velocity data science

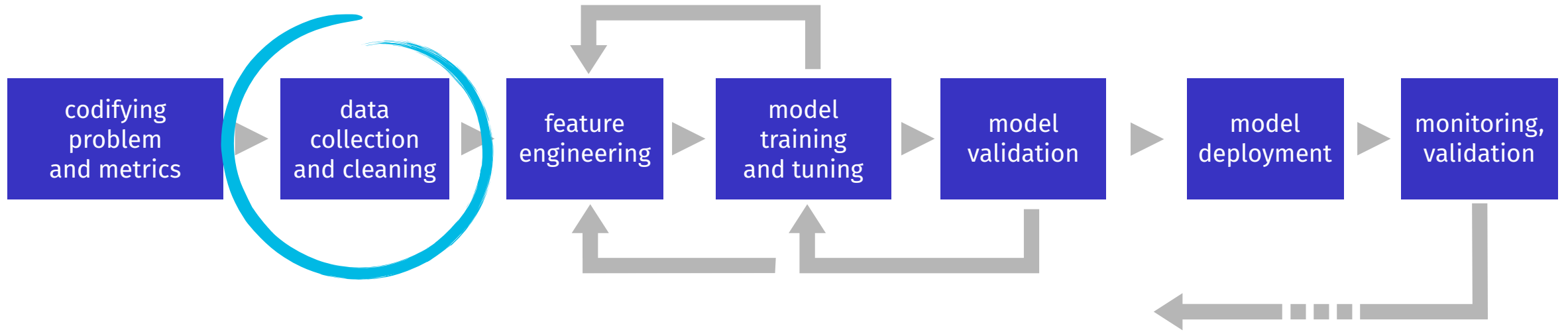
Monitoring models and pipelines in production

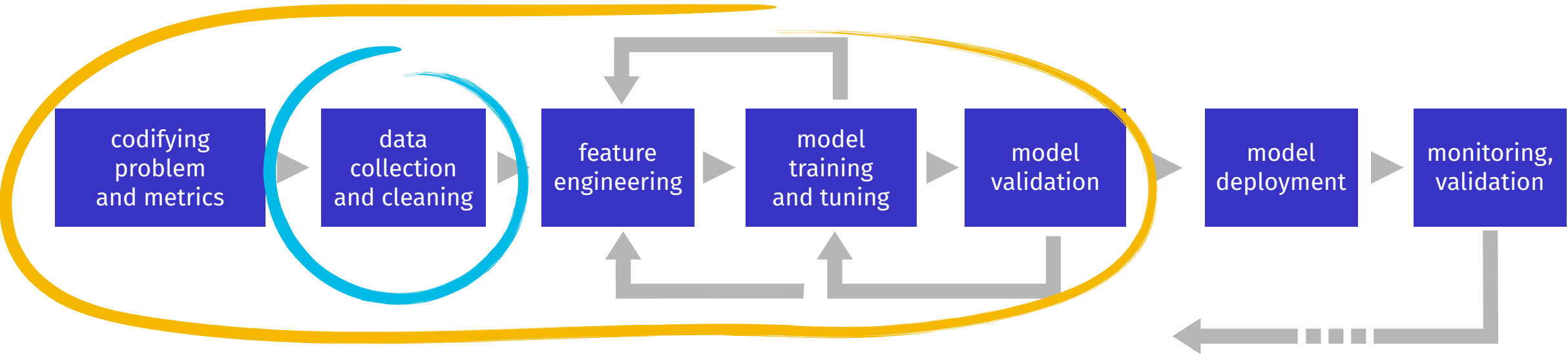
The big picture

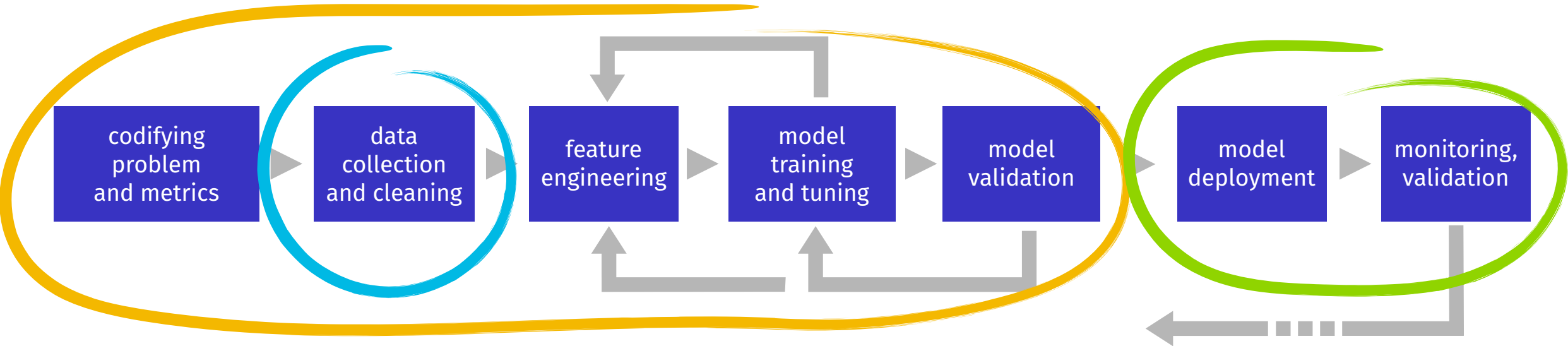


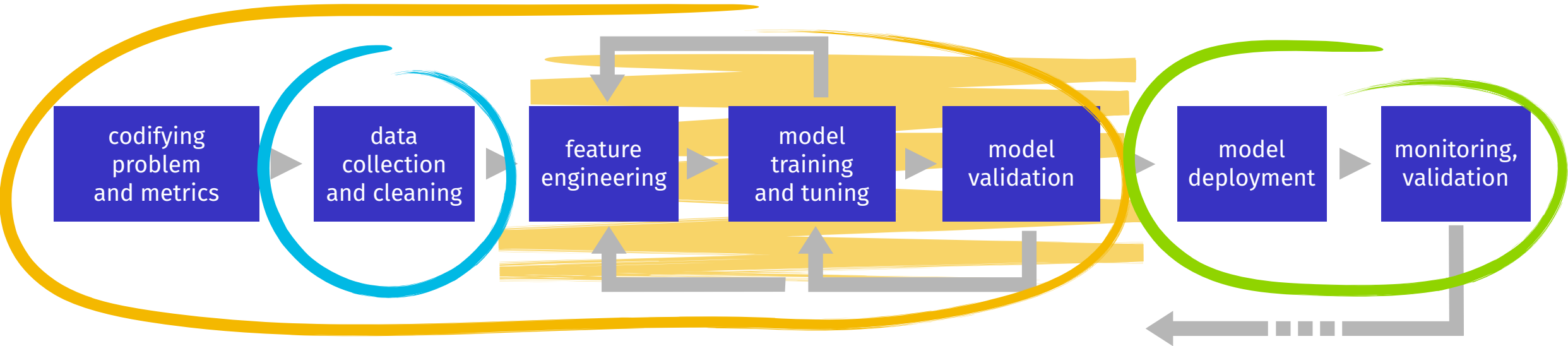
Machine learning systems

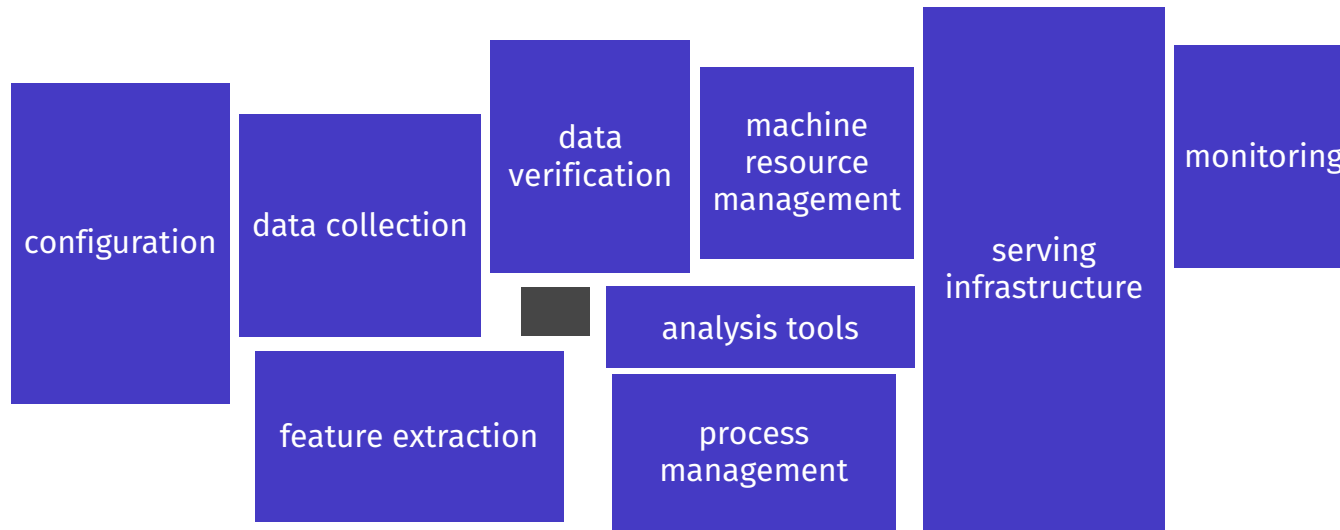












Sculley et al., “Hidden Technical Debt in Machine Learning Systems.” *NIPS 2015*



Ensuring reproducibility

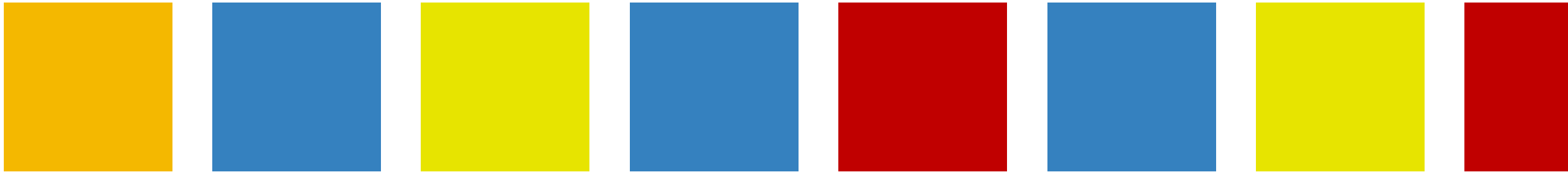


00

01

01

10



01

10

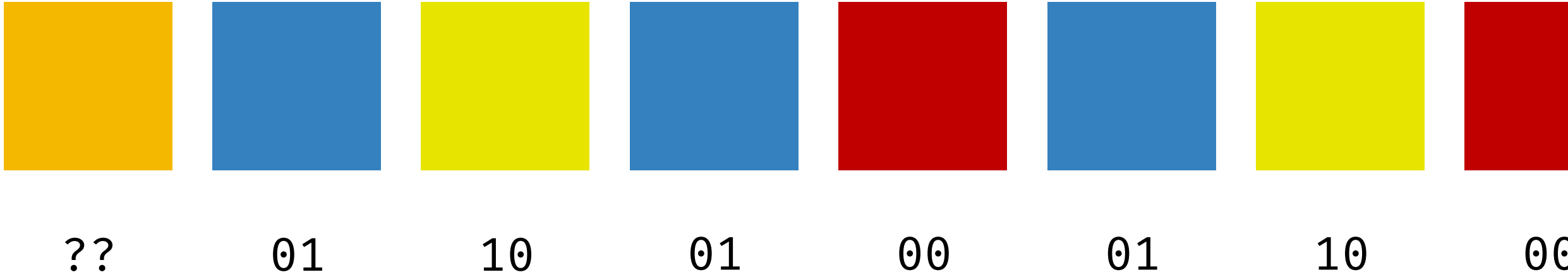
01

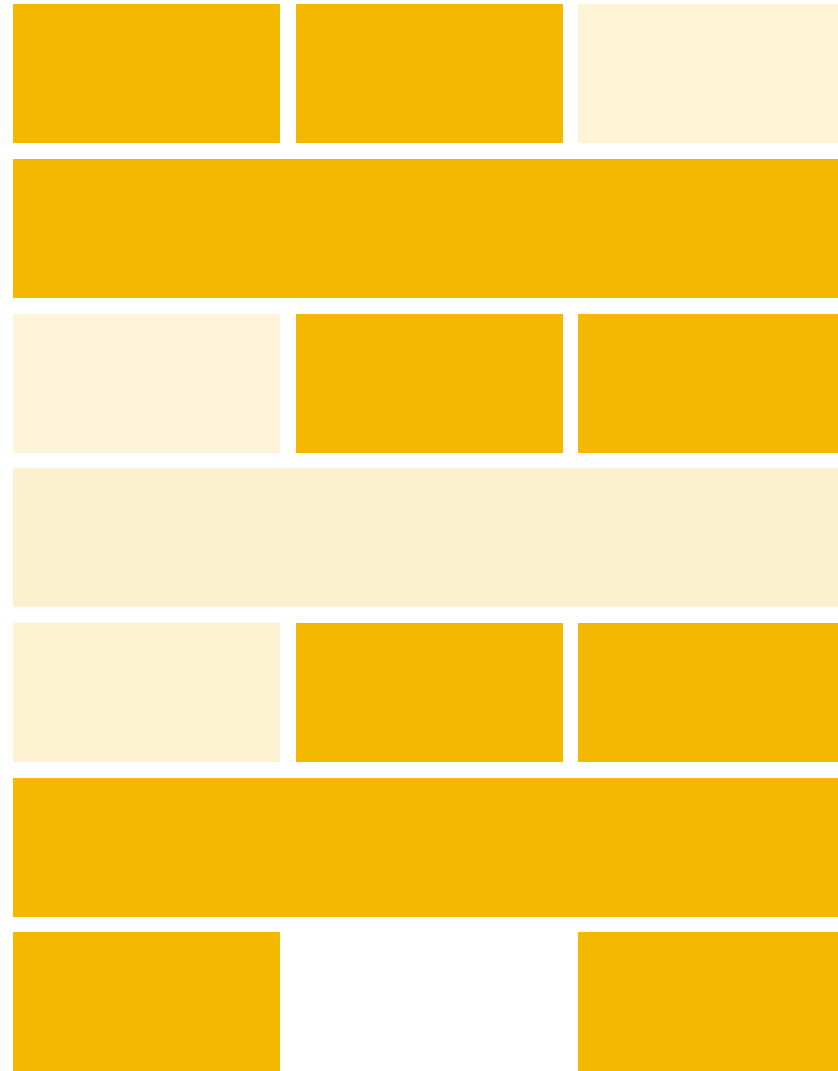
00

01

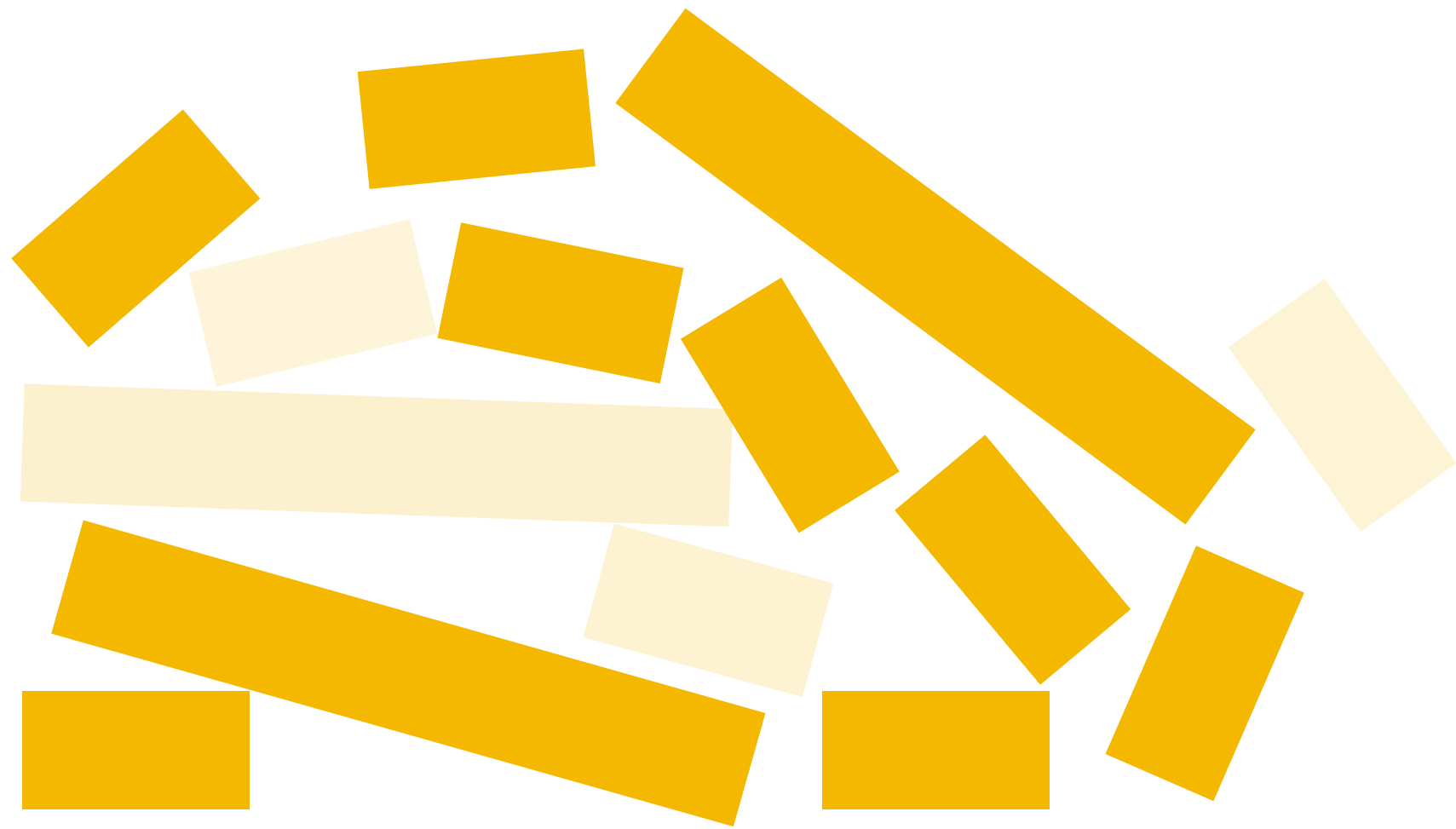
10

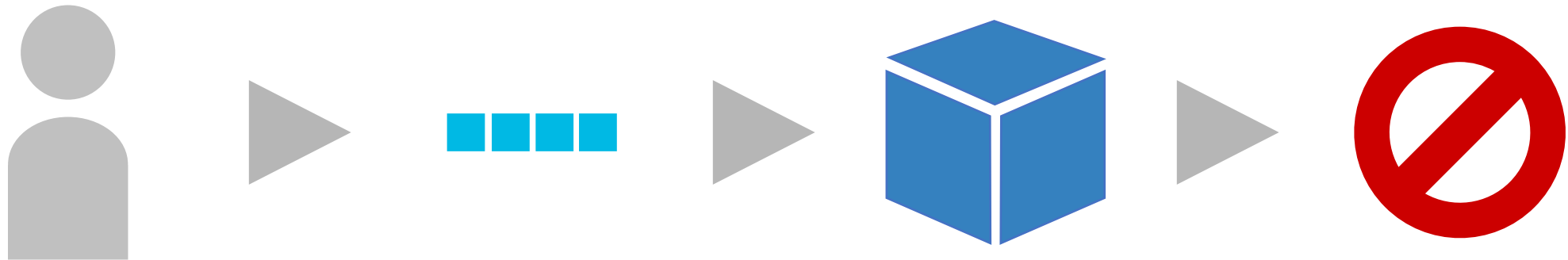
00

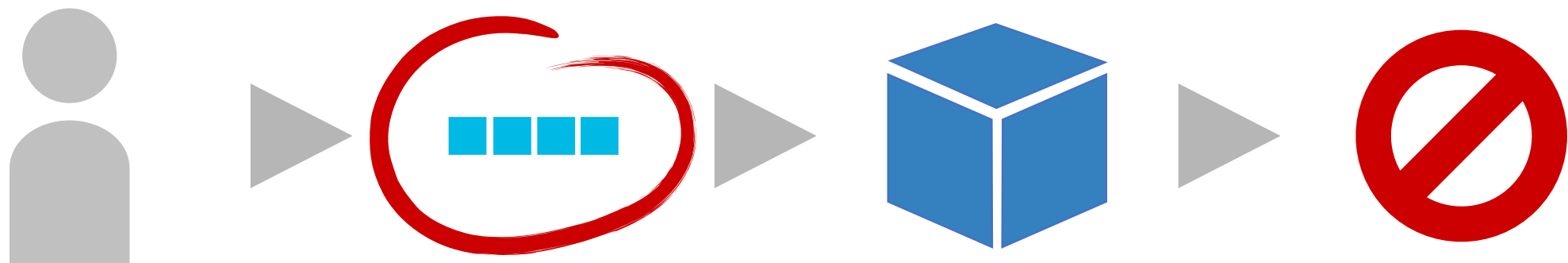


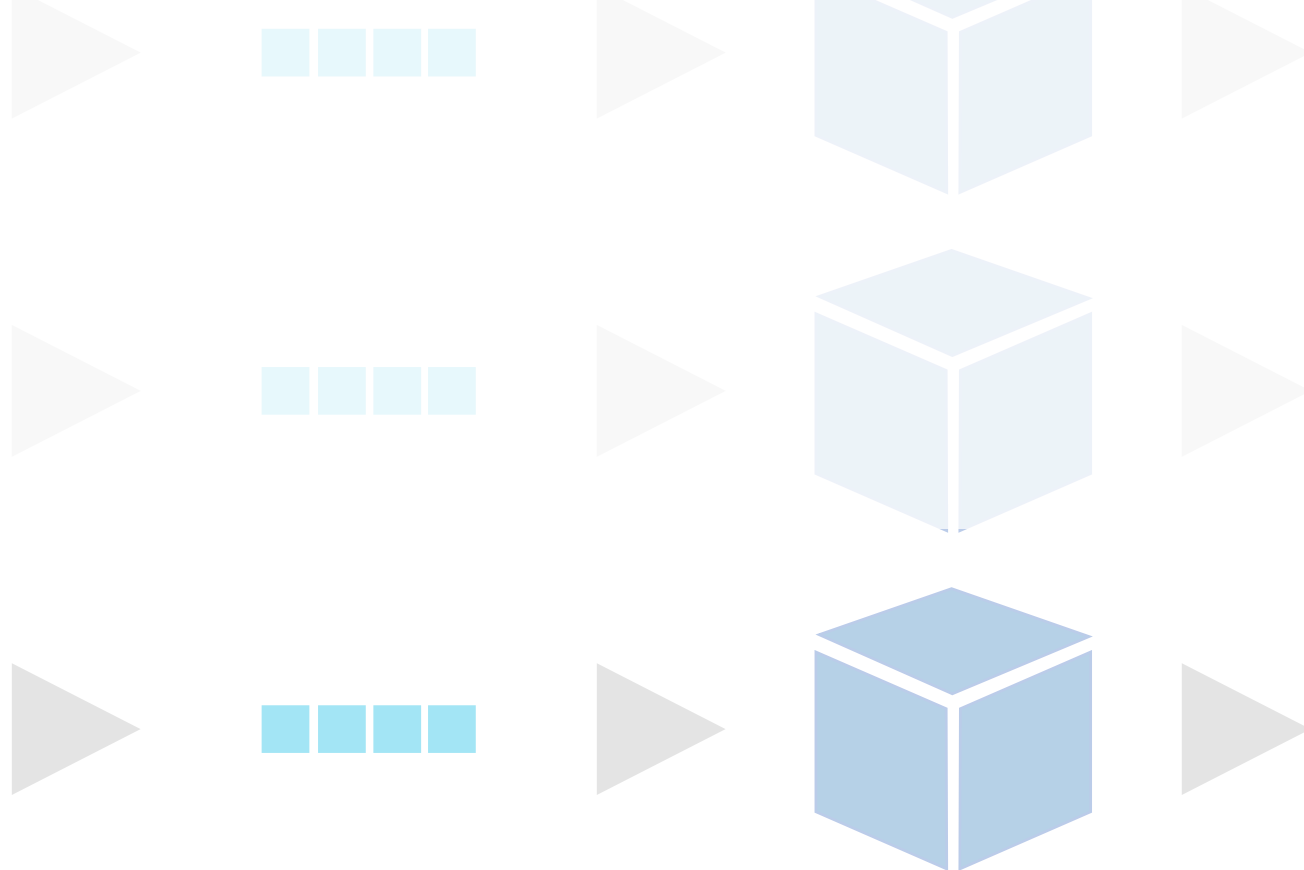
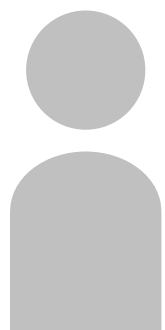


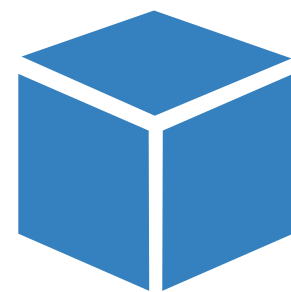
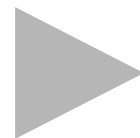
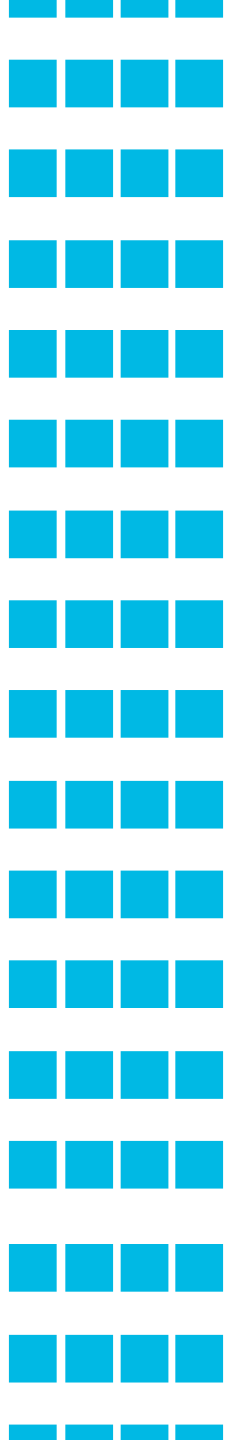
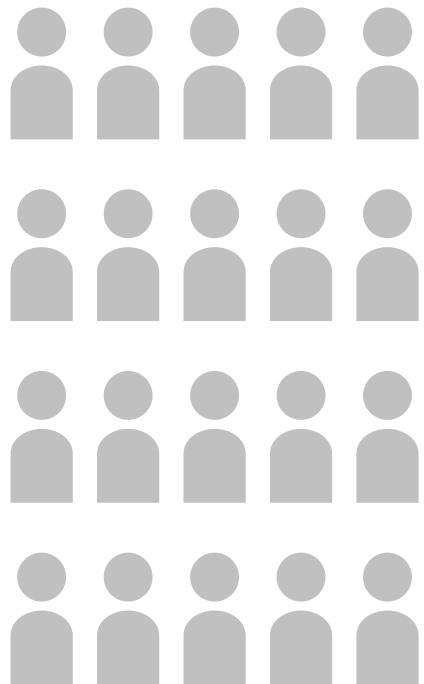


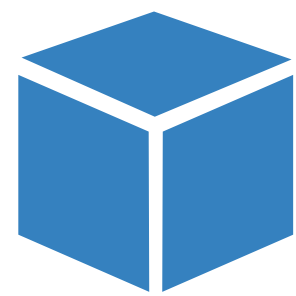
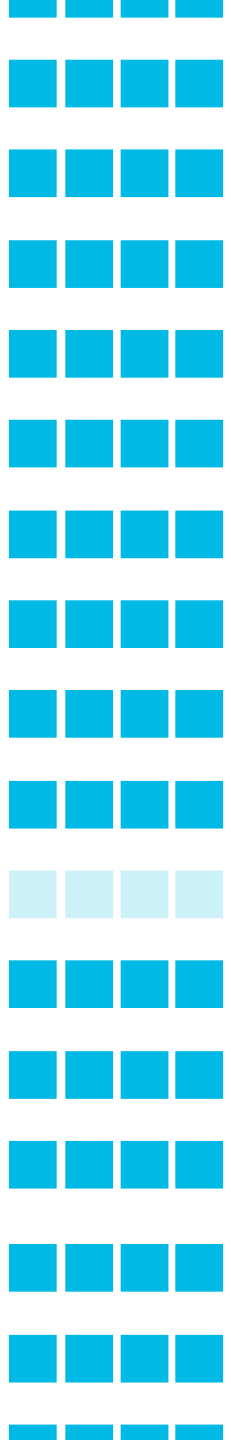
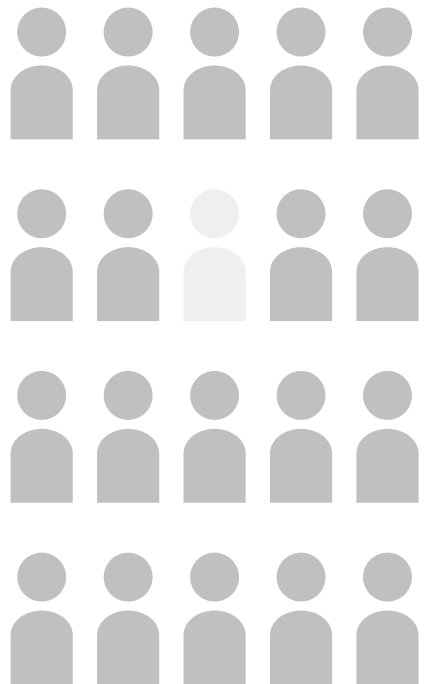


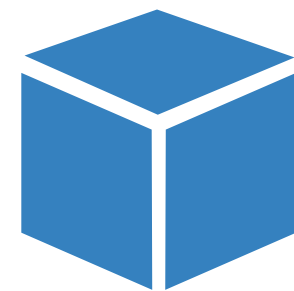
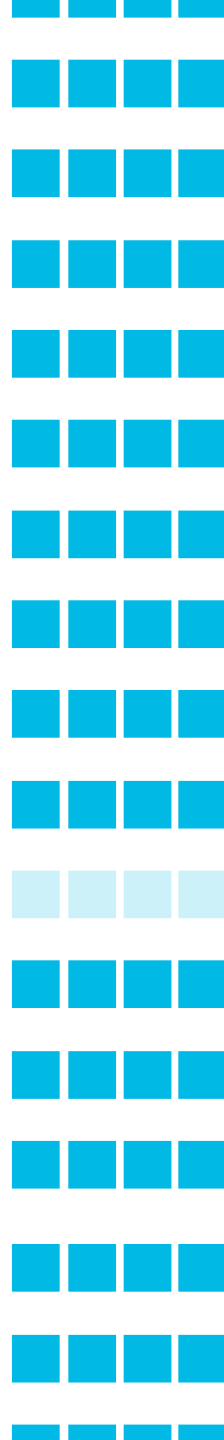
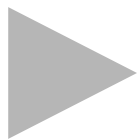
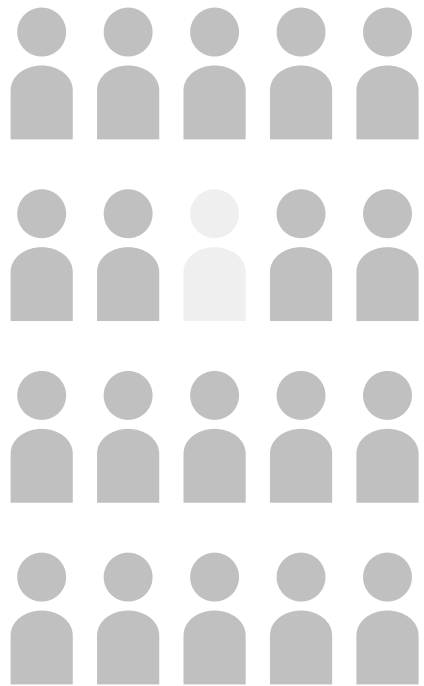








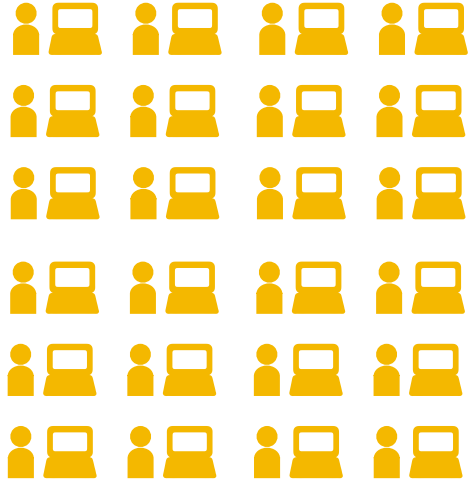




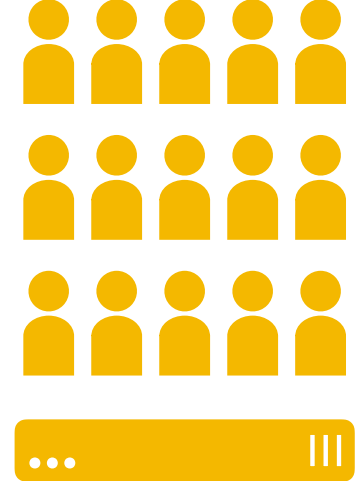


High-velocity data science

A continuum of approaches

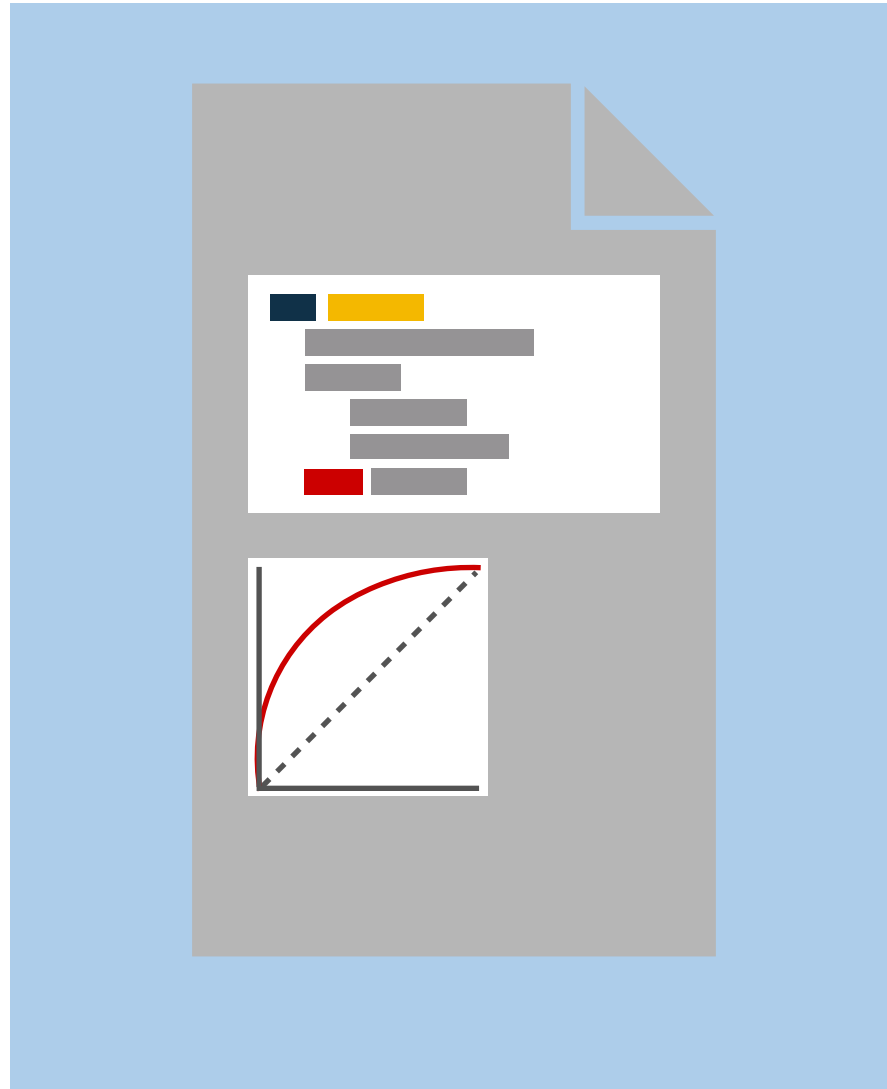


**maximal flexibility,
minimal reproducibility
and governance**

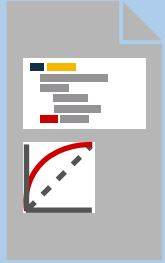


**maximal governance,
minimal flexibility**

notebook environment



notebook environment



notebook environment



altair==3.2.0	jedi==0.17.2	packaging==20.4	qtconsole==4.7.7
appnope==0.1.0	Jinja2==2.11.2	pandas==1.0.0	QtPy==1.9.0
argon2-cffi==20.1.0	joblib==0.16.0	pandocfilters==1.4.2	scikit-learn==0.23.2
async-generator==1.10	jsonschema==3.2.0	parso==0.7.1	scipy==1.5.2
attrs==20.2.0	jupyter==1.0.0	pexpect==4.8.0	Send2Trash==1.5.0
backcall==0.2.0	jupyter-client==6.1.7	pickleshare==0.7.5	six==1.15.0
bleach==3.1.5	jupyter-console==6.2.0	prometheus-client==0.8.0	terminado==0.8.3
cffi==1.14.2	jupyter-core==4.6.3	prompt-toolkit==3.0.7	testpath==0.4.4
cloudpickle==1.6.0	jupyterlab-pygments==0.1.1	ptyprocess==0.6.0	threadpoolctl==2.1.0
decorator==4.4.2	MarkupSafe==1.1.1	pyarrow==1.0.1	toolz==0.10.0
defusedxml==0.6.0	mistune==0.8.4	pyparser==2.20	tornado==6.0.4
entrypoints==0.3	nbclient==0.5.0	Pygments==2.7.0	traitlets==4.3.3
importlib-metadata==1.7.0	nbconvert==6.0.2	pyparsing==2.4.7	vega==2.5.0
ipykernel==5.3.4	nbformat==5.0.7	pyrsistent==0.17.3	wcwidth==0.2.5
ipython==7.16.1	nest-asyncio==1.4.0	python-dateutil==2.8.1	webencodings==0.5.1
ipython-genutils==0.2.0	notebook==6.1.4	pytz==2020.1	widetsnbextension==3.5.1
ipywidgets==7.5.1	numpy==1.19.2	ppyzmq==19.0.2	xgboost==1.2.0

notebook environment



altair==3.2.0	jedi==0.17.2	packaging==20.4	qtconsole==4.7.7
appnope==0.1.0	Jinja2==2.11.2	pandas==1.0.0	QtPy==1.9.0
argon2-cffi==20.1.0	joblib==0.16.0	pandocfilters==1.4.2	scikit-learn==0.23.2
async-generator==1.10	jsonschema==3.2.0	parso==0.7.1	scipy==1.5.2
attrs==20.2.0	jupyter==1.0.0	pexpect==4.8.0	Send2Trash==1.5.0
backcall==0.2.0	jupyter-client==6.1.7	pickleshare==0.7.5	six==1.15.0
bleach==3.1.5	jupyter-console==6.2.0	prometheus-client==0.8.0	terminado==0.8.3
cffi==1.14.2	jupyter-core==4.6.3	prompt-toolkit==3.0.7	testpath==0.4.4
cloudpickle==1.6.0	jupyterlab-pygments==0.1.1	ptyprocess==0.6.0	threadpoolctl==2.1.0
decorator==4.4.2	MarkupSafe==1.1.1	pyarrow==1.0.1	toolz==0.10.0
defusedxml==0.6.0	mistune==0.8.4	pyparser==2.20	tornado==6.0.4
entrypoints==0.3	nbclient==0.5.0	Pygments==2.7.0	traitlets==4.3.3
importlib-metadata==1.7.0	nbconvert==6.0.2	pyparsing==2.4.7	vega==2.5.0
ipykernel==5.3.4	nbformat==5.0.7	pyrsistent==0.17.3	wcwidth==0.2.5
ipython==7.16.1	nest-asyncio==1.4.0	python-dateutil==2.8.1	webencodings==0.5.1
ipython-genutils==0.2.0	notebook==6.1.4	pytz==2020.1	widetsnbextension==3.5.1
ipywidgets==7.5.1	numpy==1.19.2	ppyzmq==19.0.2	xgboost==1.2.0

notebook environment



```
altair==3.2.0  
cloudpickle==1.6.0  
jupyter==1.0.0  
numpy==1.19.2  
pandas==1.0.0  
pyarrow==1.0.1  
scikit-learn==0.23.2  
vega==2.5.0  
xgboost==1.2.0
```

notebook environment



```
altair==3.2.0  
cloudpickle==1.6.0  
jupyter==1.0.0  
numpy==1.19.2  
pandas==1.0.0  
pyarrow==1.0.1  
scikit-learn==0.23.2  
vega==2.5.0  
xgboost==1.2.0
```



pandas== 1.0.0



Tags

v1.0.0 ...

🕒 on Jan 29

🔗 fd9ceb9

📄 zip

📄 tar.gz

📄 Notes

⬇ Downloads

v1.0.0rc0 ...

🕒 on Jan 9

🔗 d3f0856

📄 zip

📄 tar.gz

📄 Notes

⬇ Downloads

pandas== 1.0.0



Tags

v1.1.3 ...

🕒 16 days ago 🔑 db08276 📄 zip 📄 tar.gz 📄 Notes
⬇ Downloads

v1.1.2 ...

🕒 on Sep 8 🔑 2a7d332 📄 zip 📄 tar.gz 📄 Notes
⬇ Downloads

v1.1.1 ...

🕒 on Aug 20 🔑 f2ca0a2 📄 zip 📄 tar.gz 📄 Notes
⬇ Downloads

pandas== 1.1.3



Tags

v1.1.3 ...

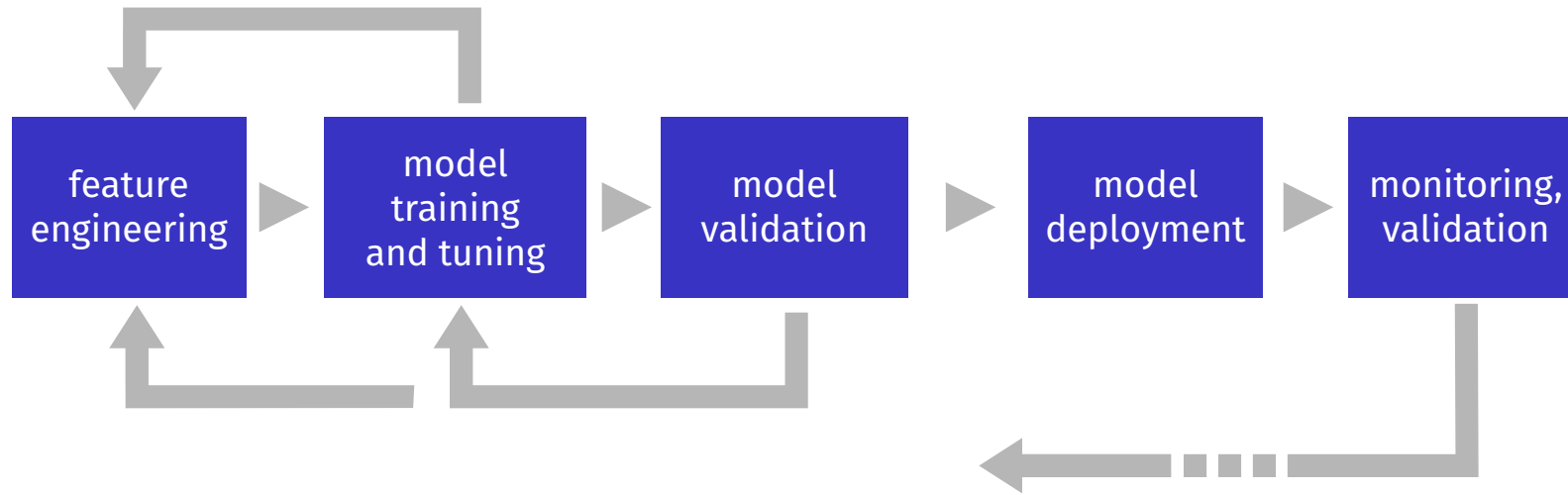
🕒 16 days ago 🔑 db08276 📄 zip 📄 tar.gz 📄 Notes
⬇ Downloads

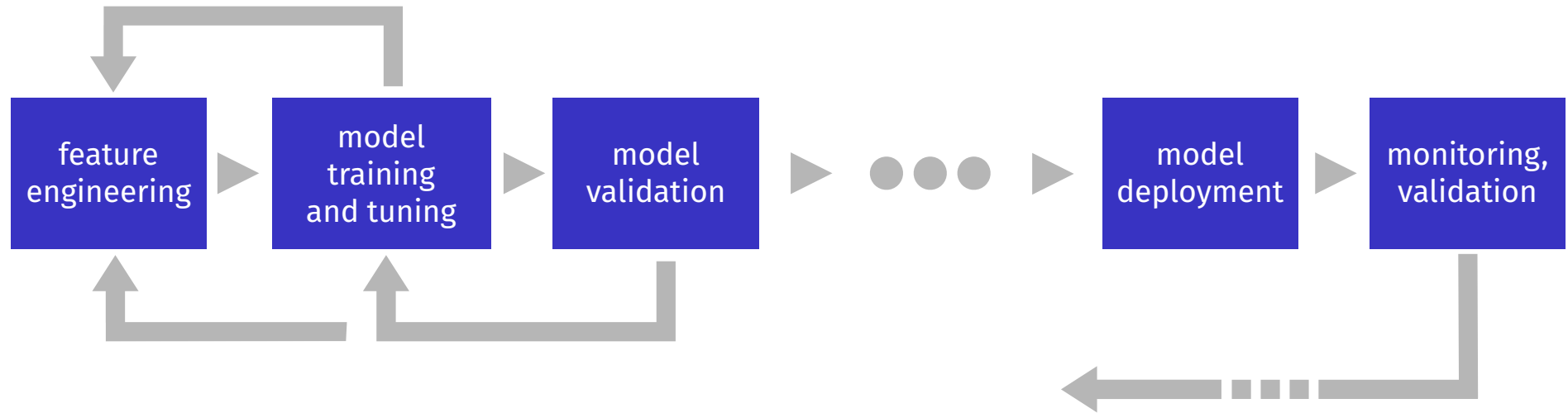
v1.1.2 ...

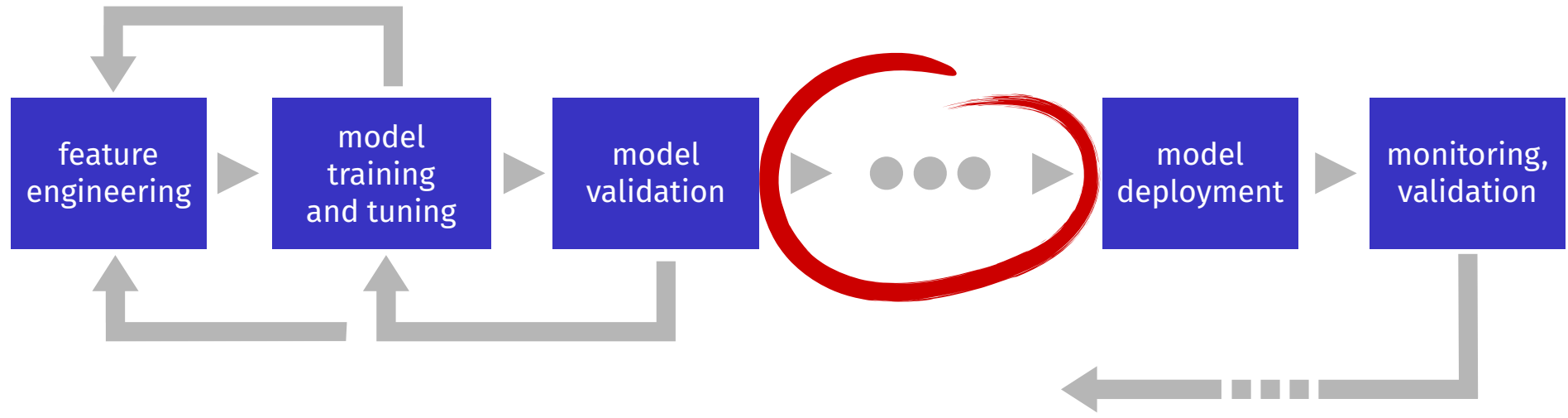
🕒 on Sep 8 🔑 2a7d332 📄 zip 📄 tar.gz 📄 Notes
⬇ Downloads

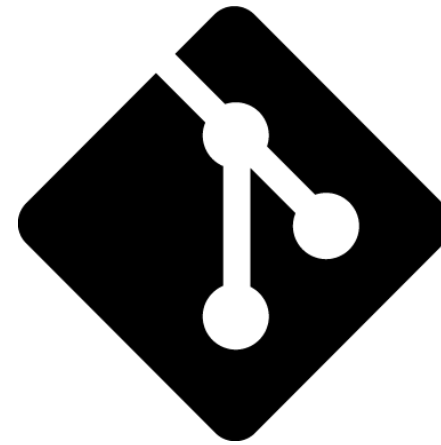
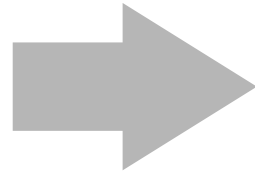
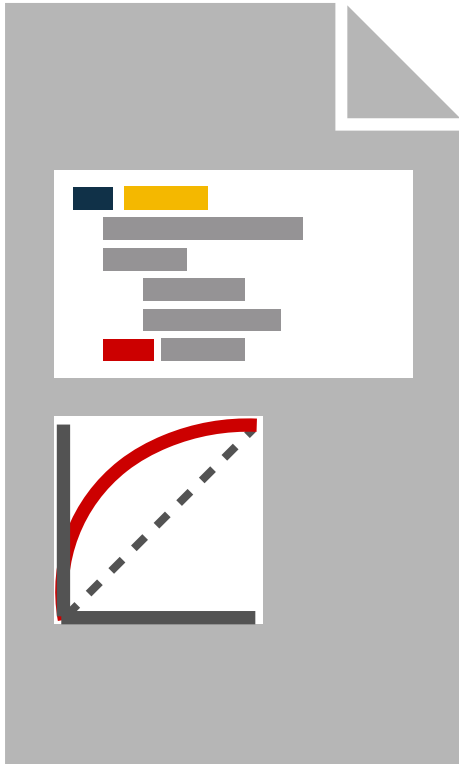
v1.1.1 ...

🕒 on Aug 20 🔑 f2ca0a2 📄 zip 📄 tar.gz 📄 Notes
⬇ Downloads

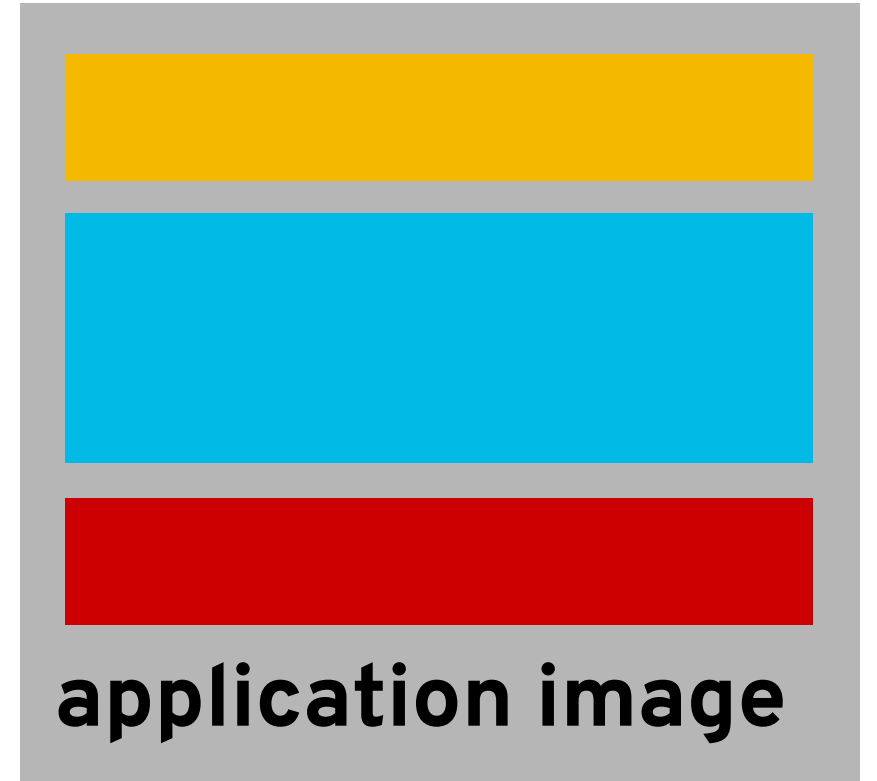
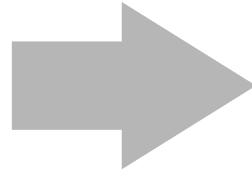
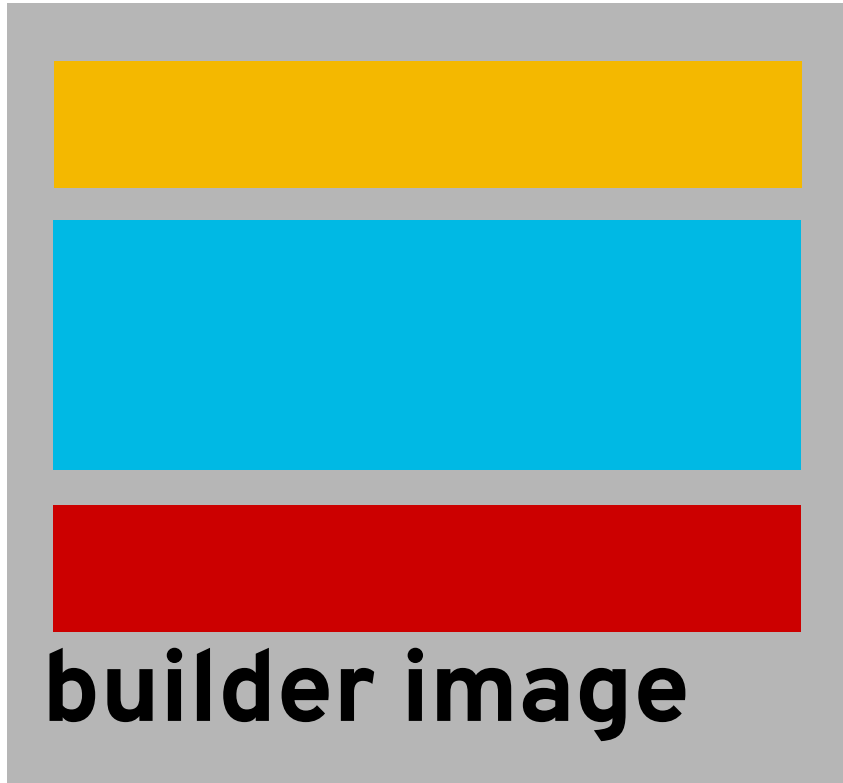




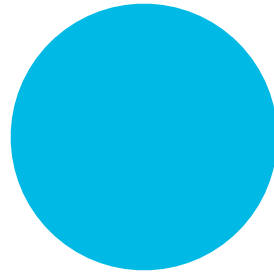
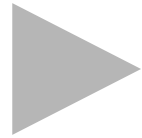




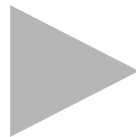
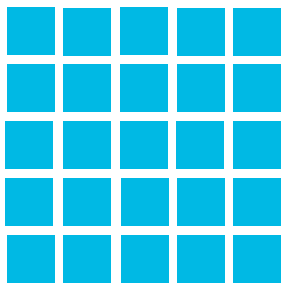
git

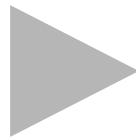
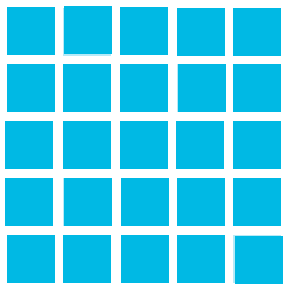


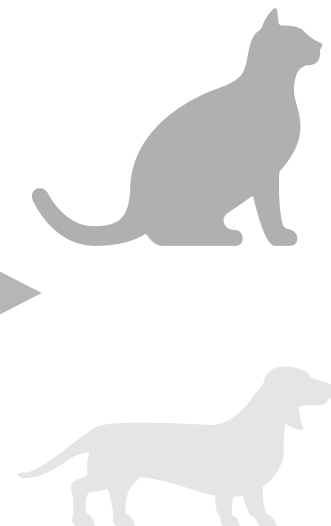
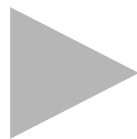
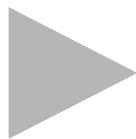
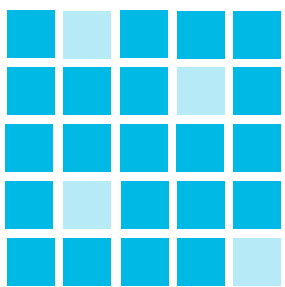
```
{"Employee ID": 1234,  
"Name": "Jane Doe",  
"Request":  
["2020-10-22", "2020-10-23"]}
```

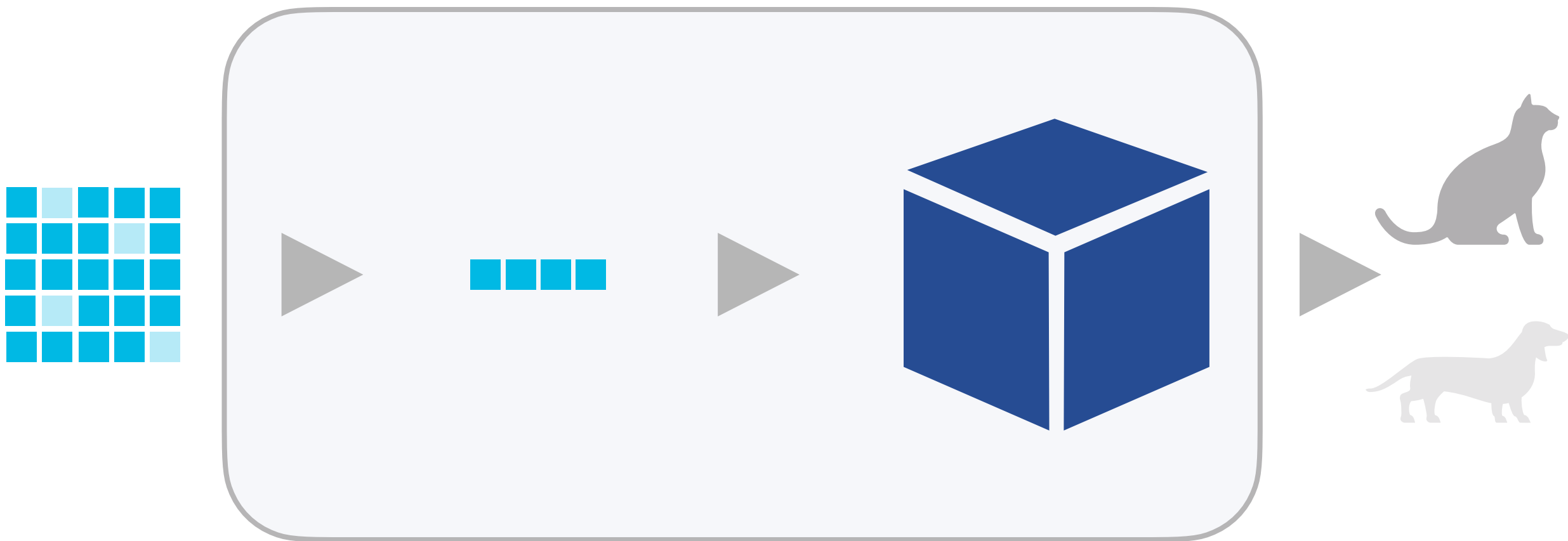


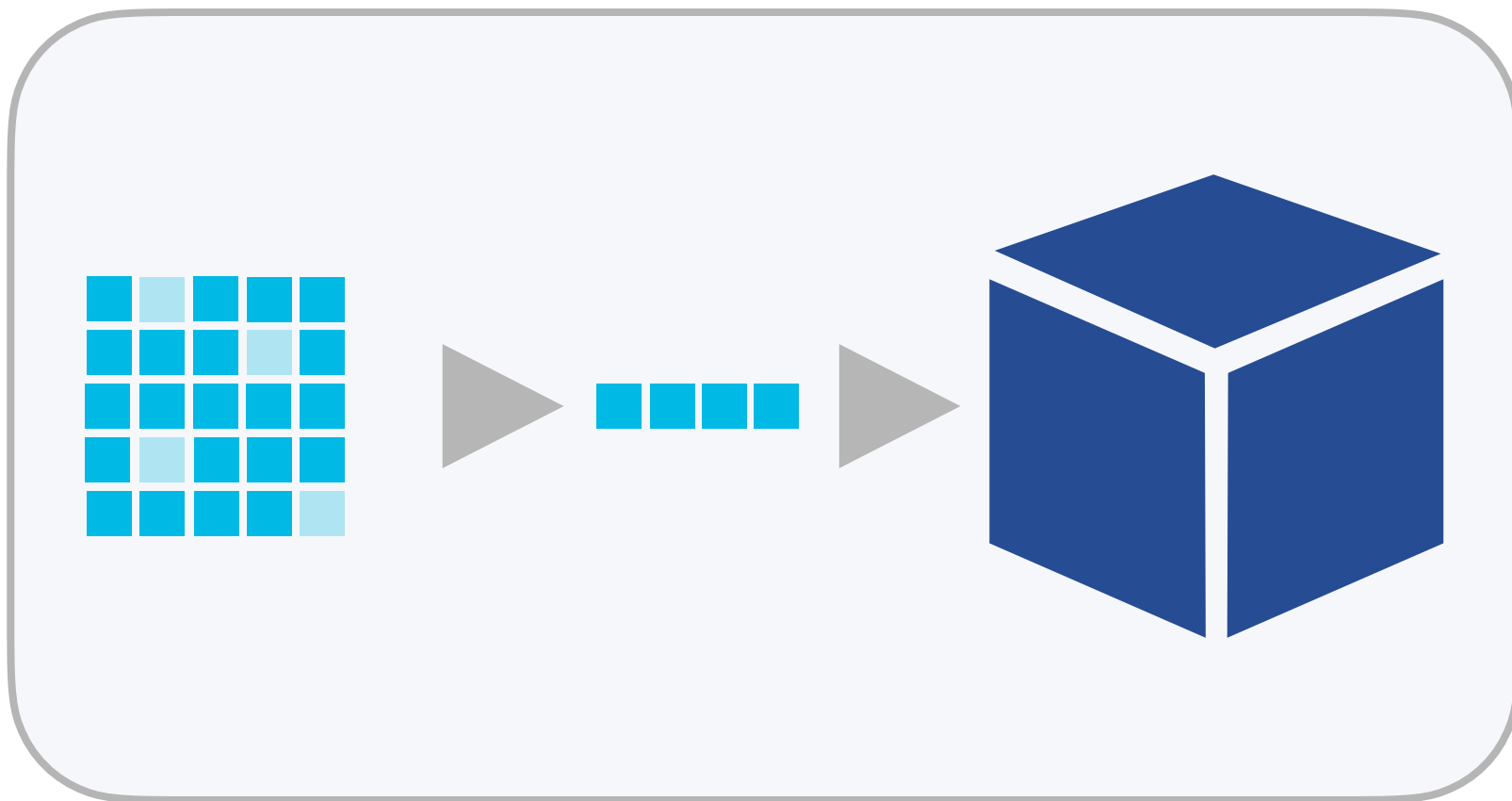
```
{"Status": "OK"}
```

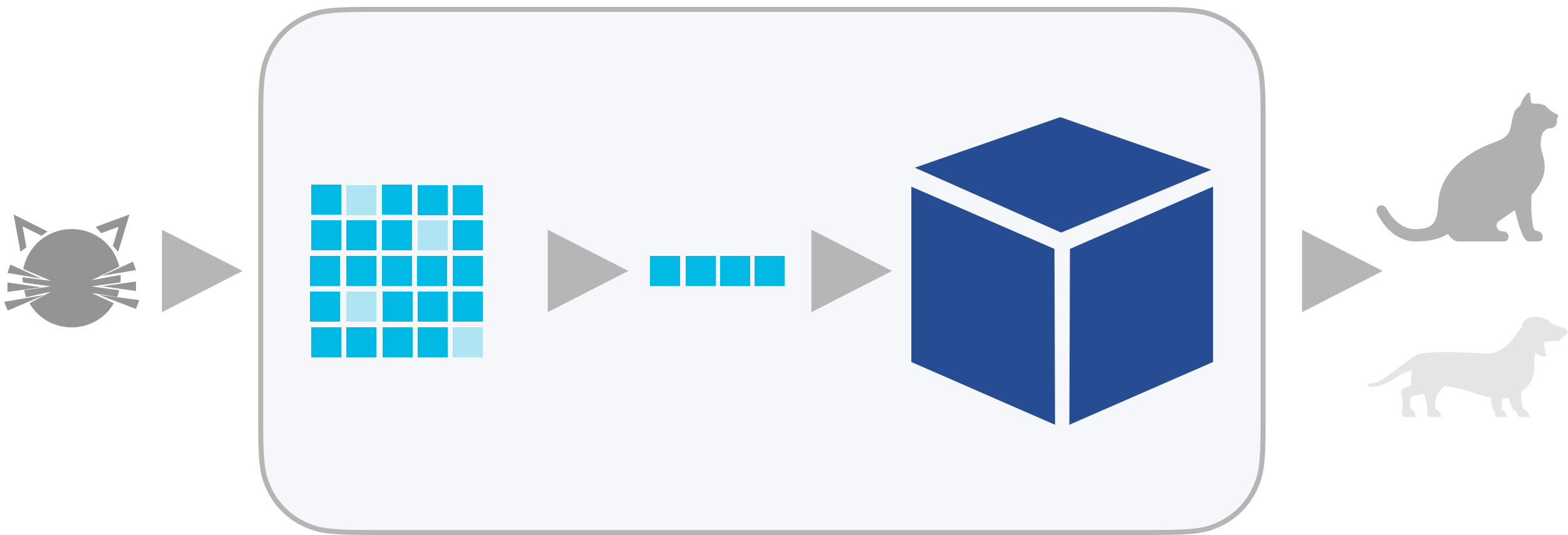


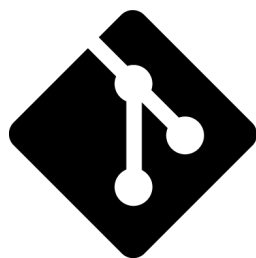
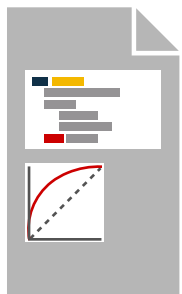




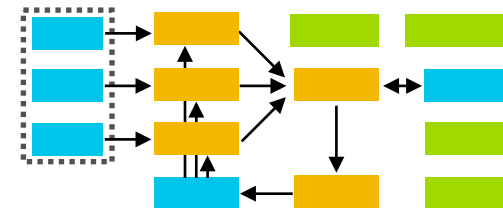
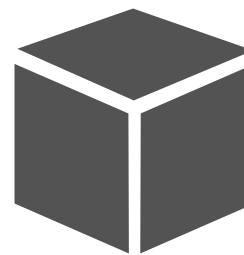
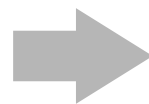






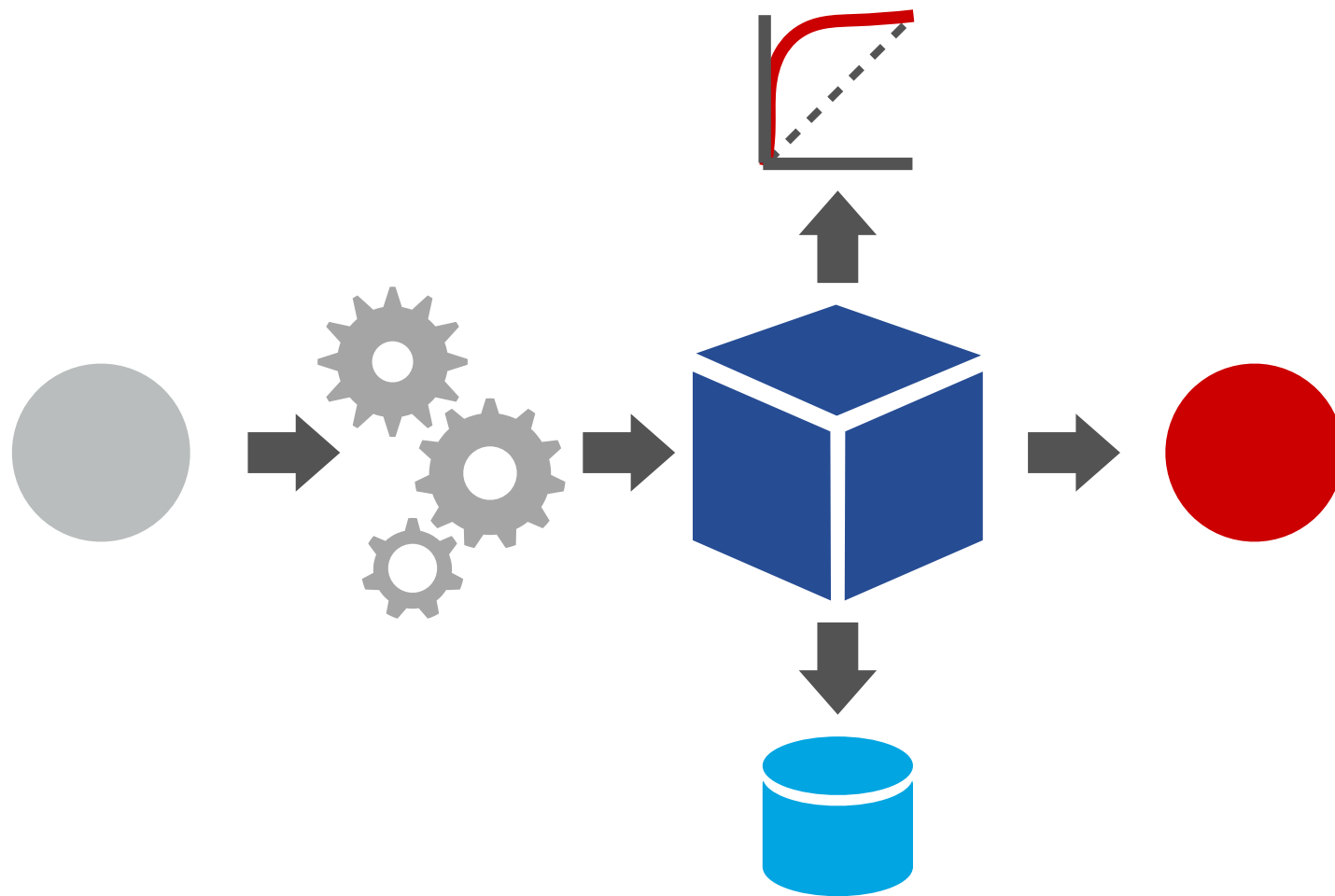


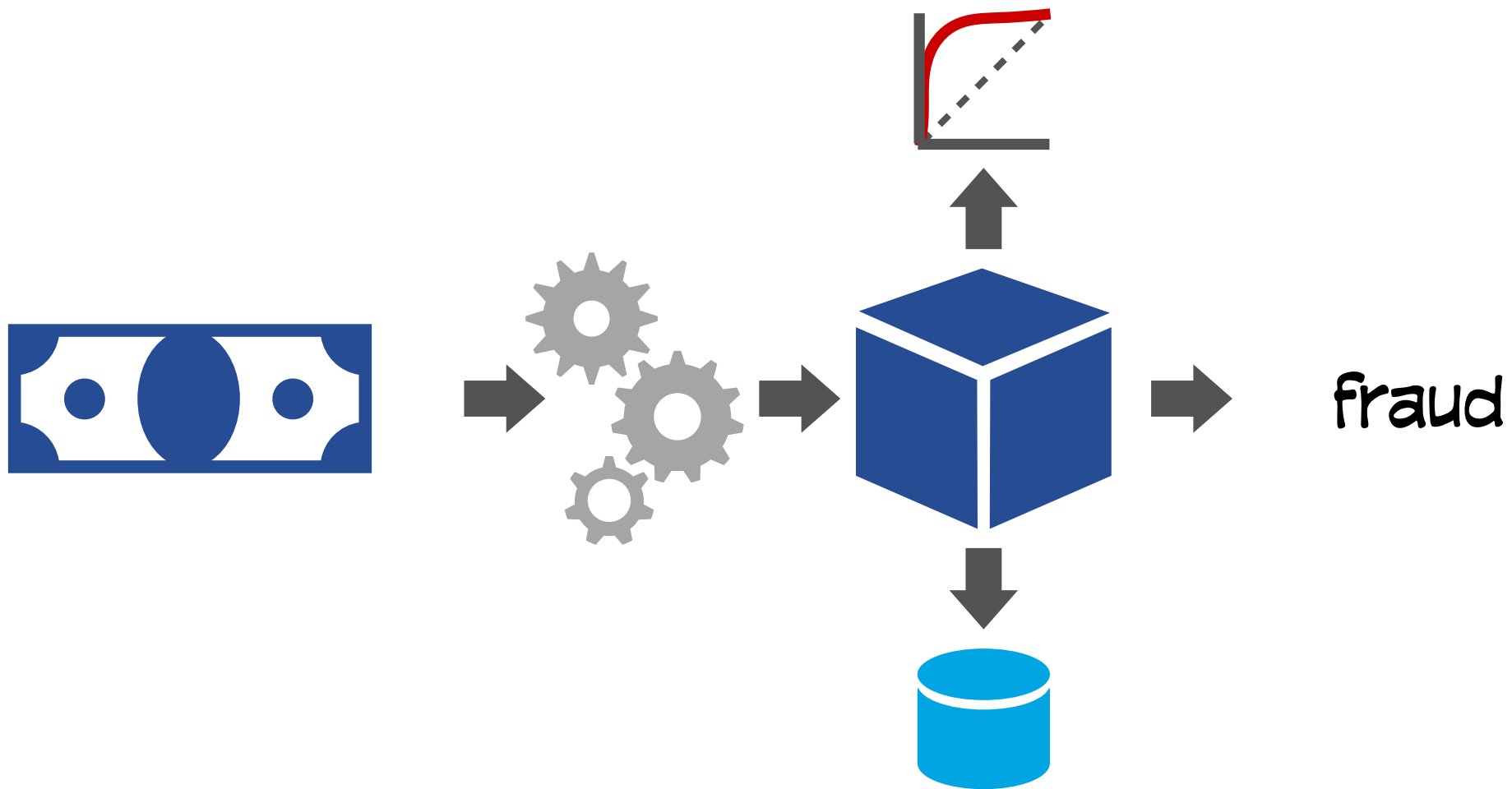
git

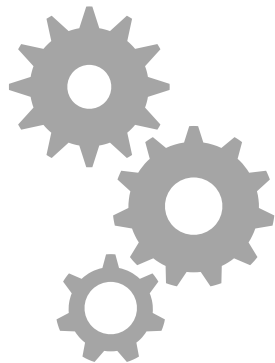


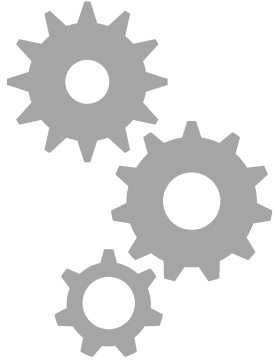


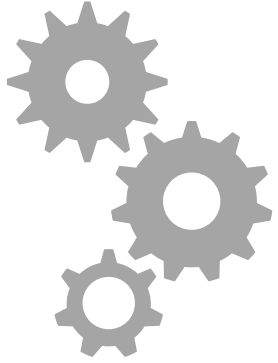
Observability for ML systems

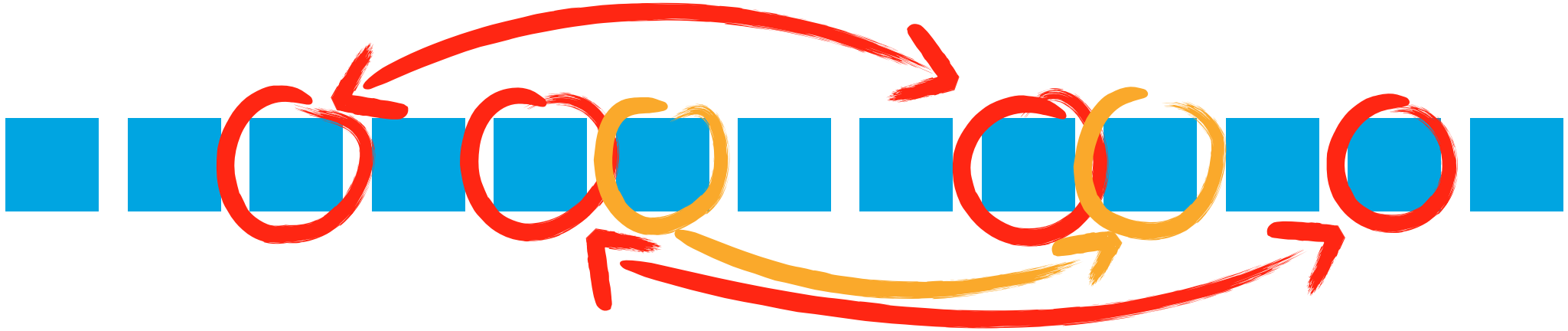
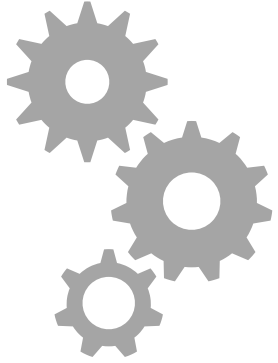


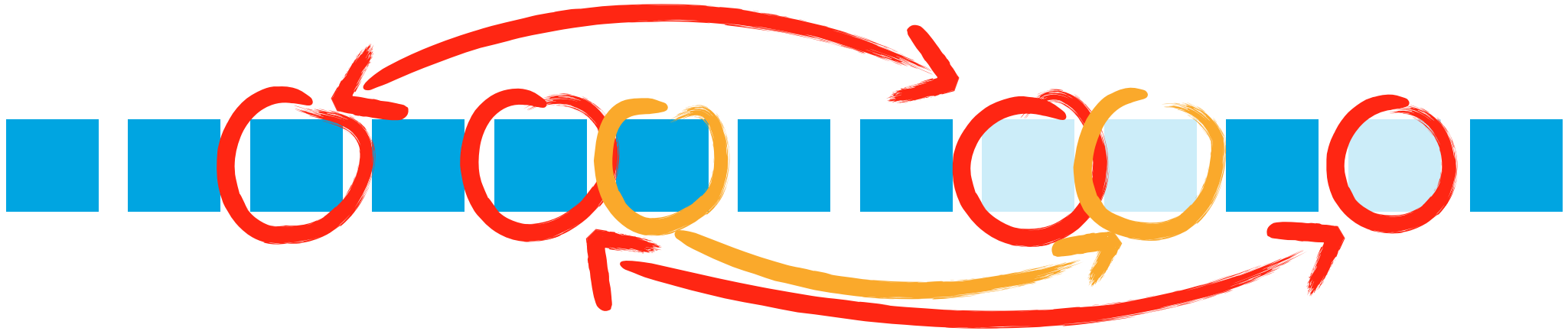
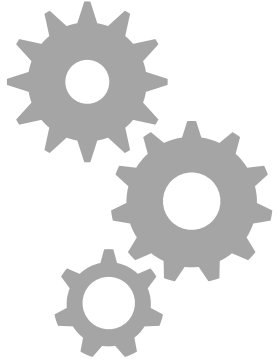


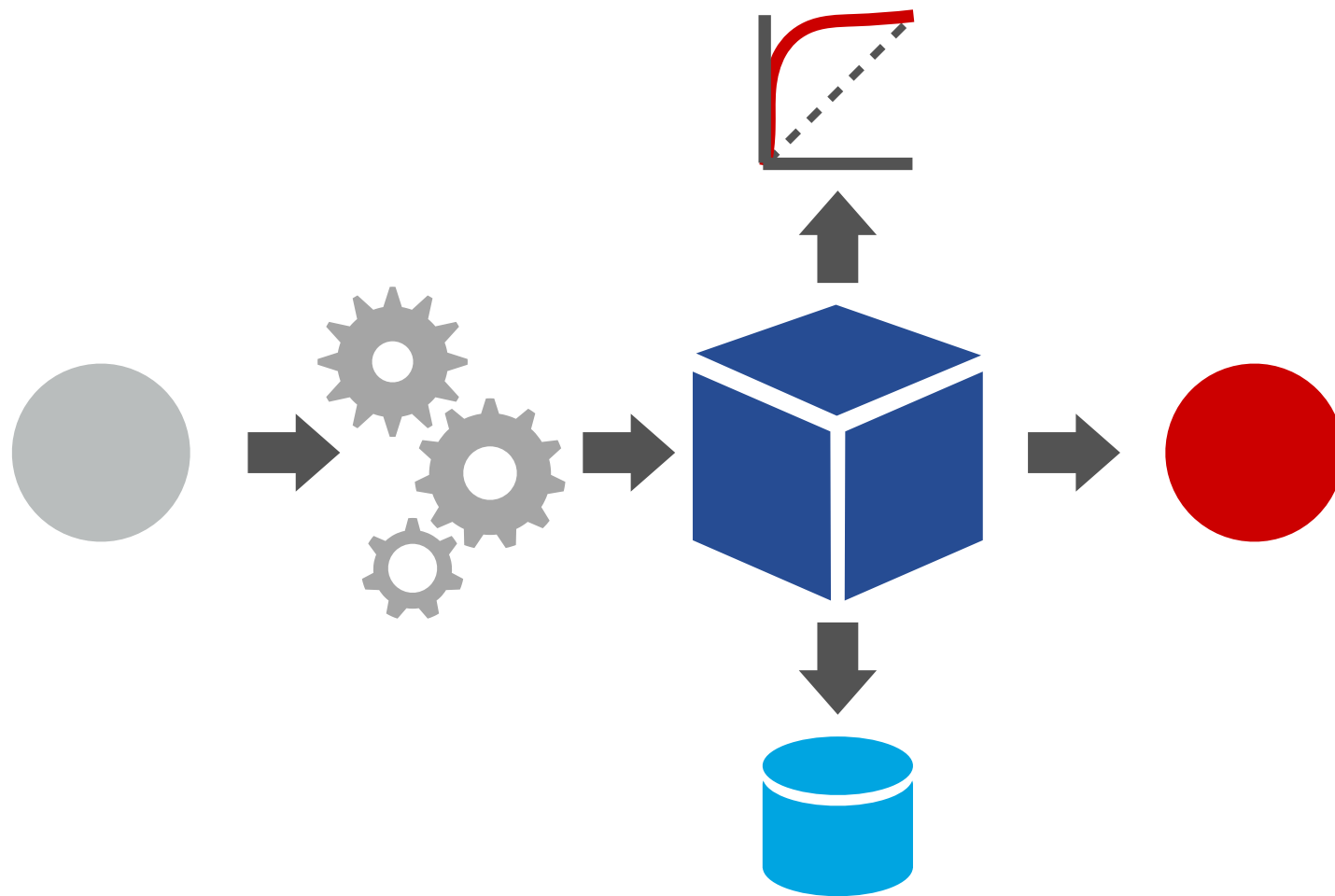














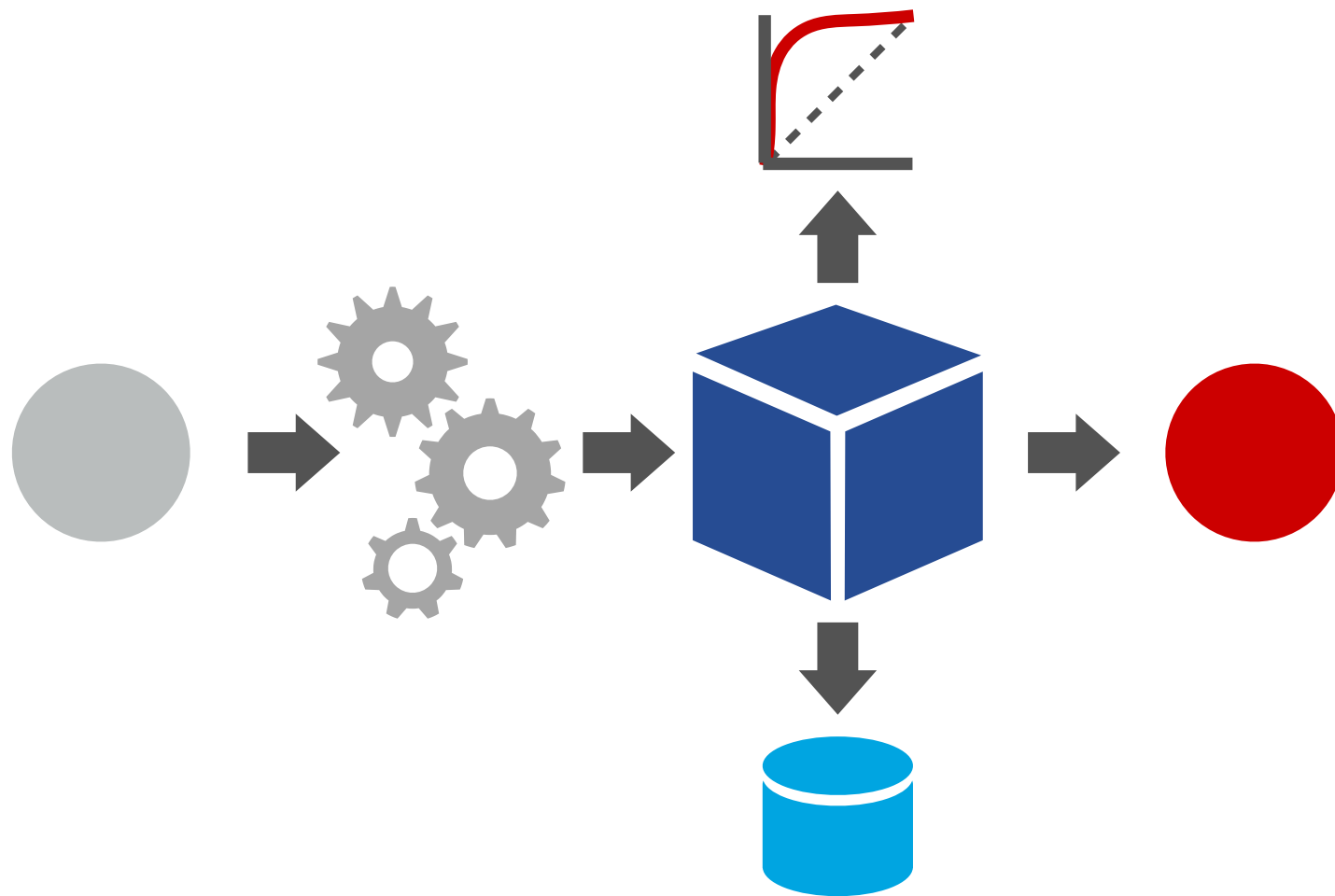
January 2020

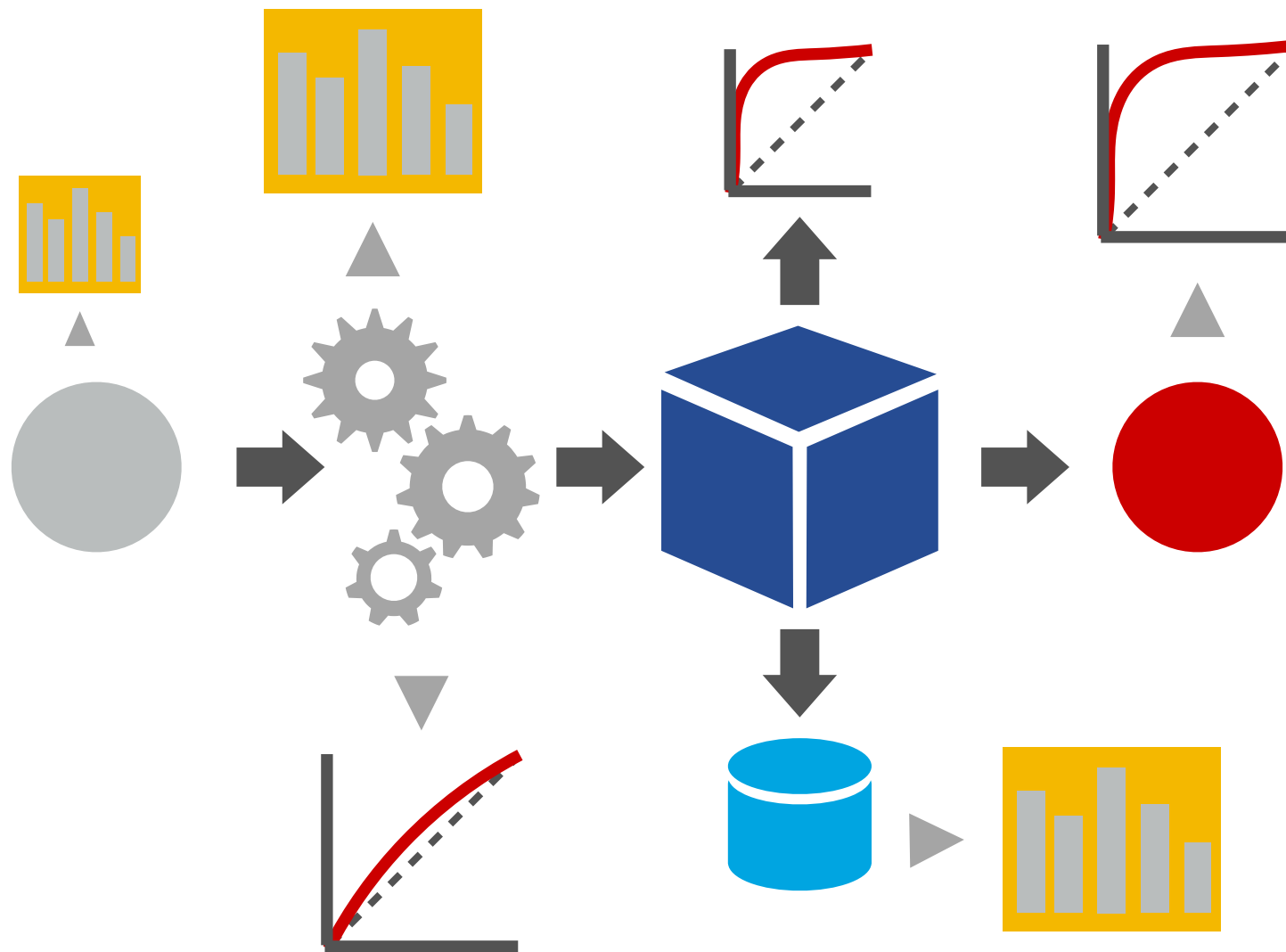
flu lung diabetes cirrhosis
alzheimer's
dna trachea arthritis pulmonary
chronic protein dementia
dengue patients tuberculosis

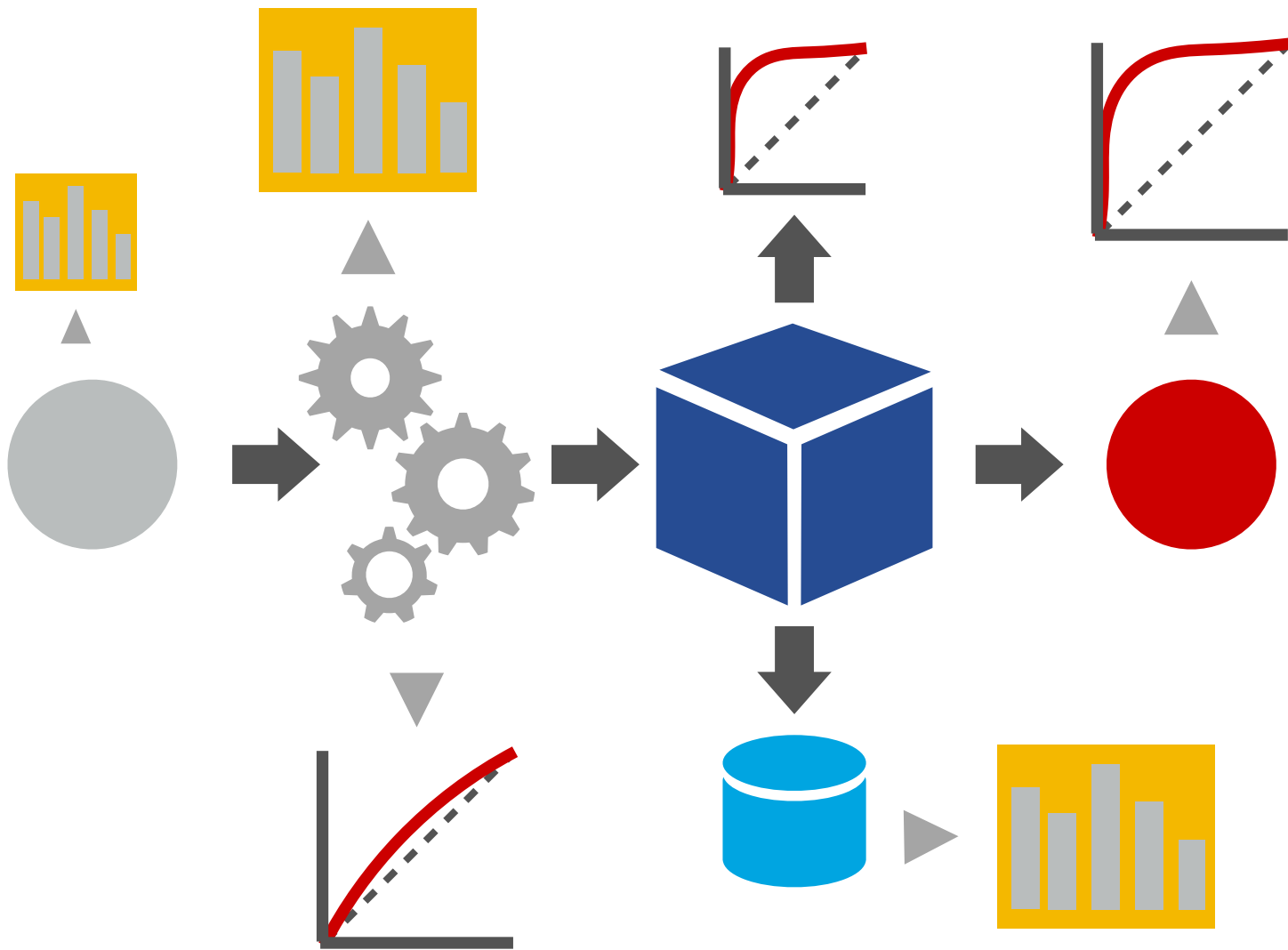


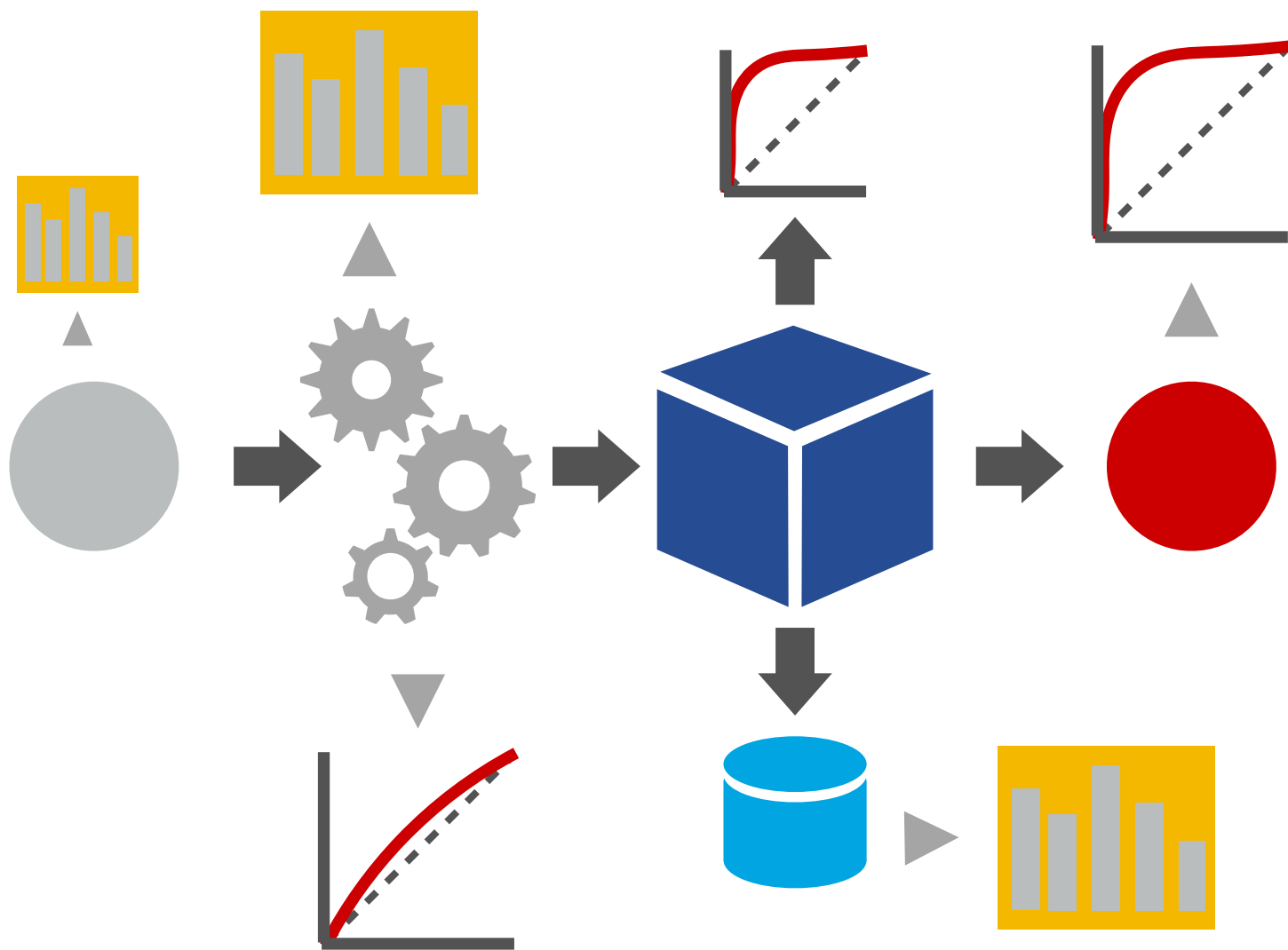
July 2020

flu
alzheimer's
trachea
dna
chronic
pulmonary
protein
cirrhosis
tuberculosis
coronavirus
dengue
patients
arthritis
lung
diabetes
dementia







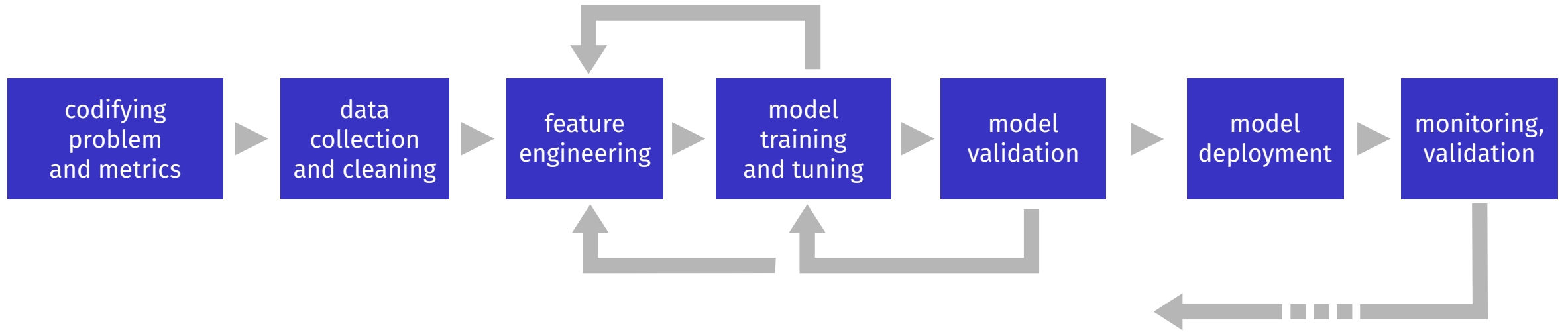


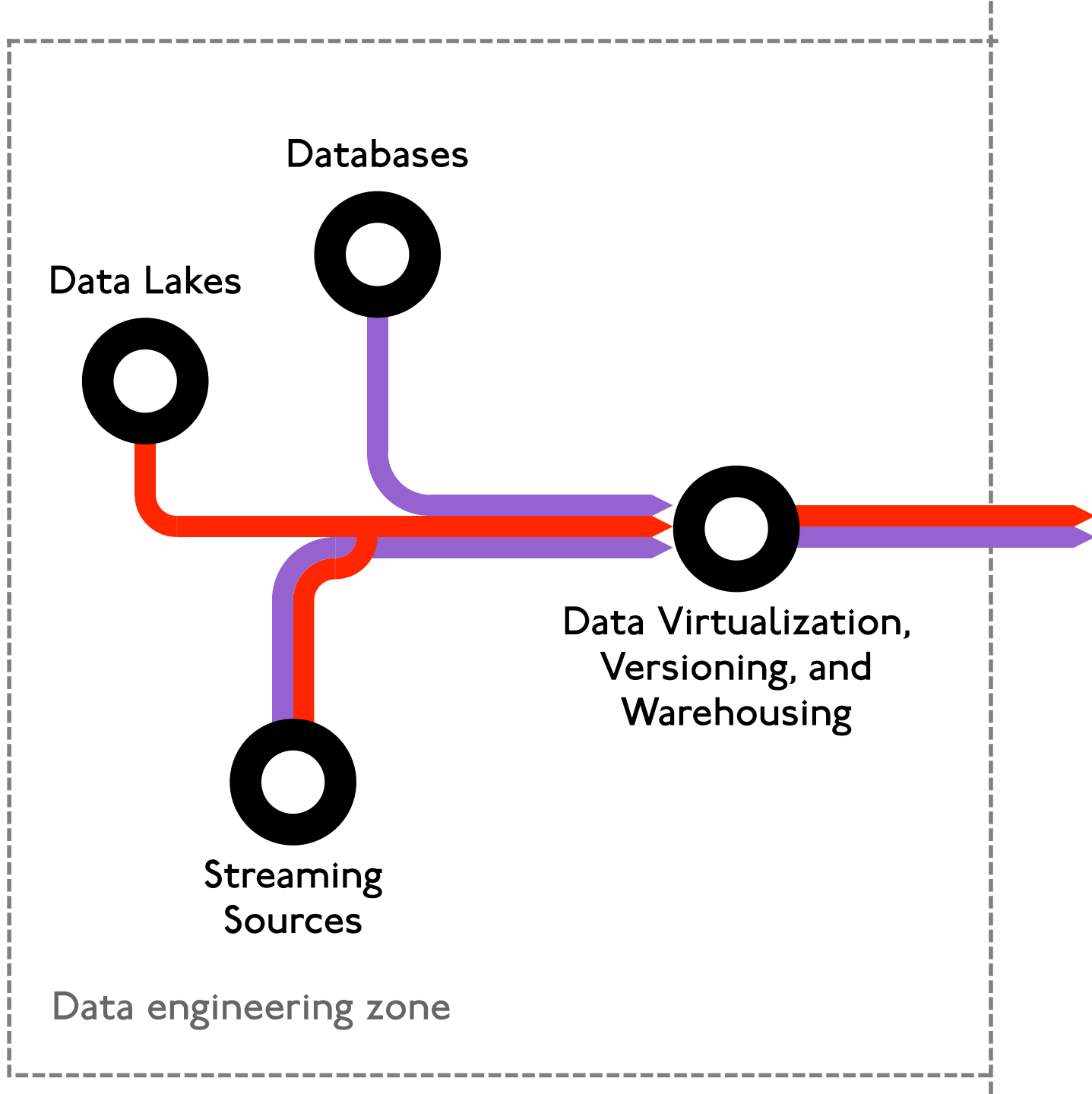
 **ALIBI**
EXPLAIN

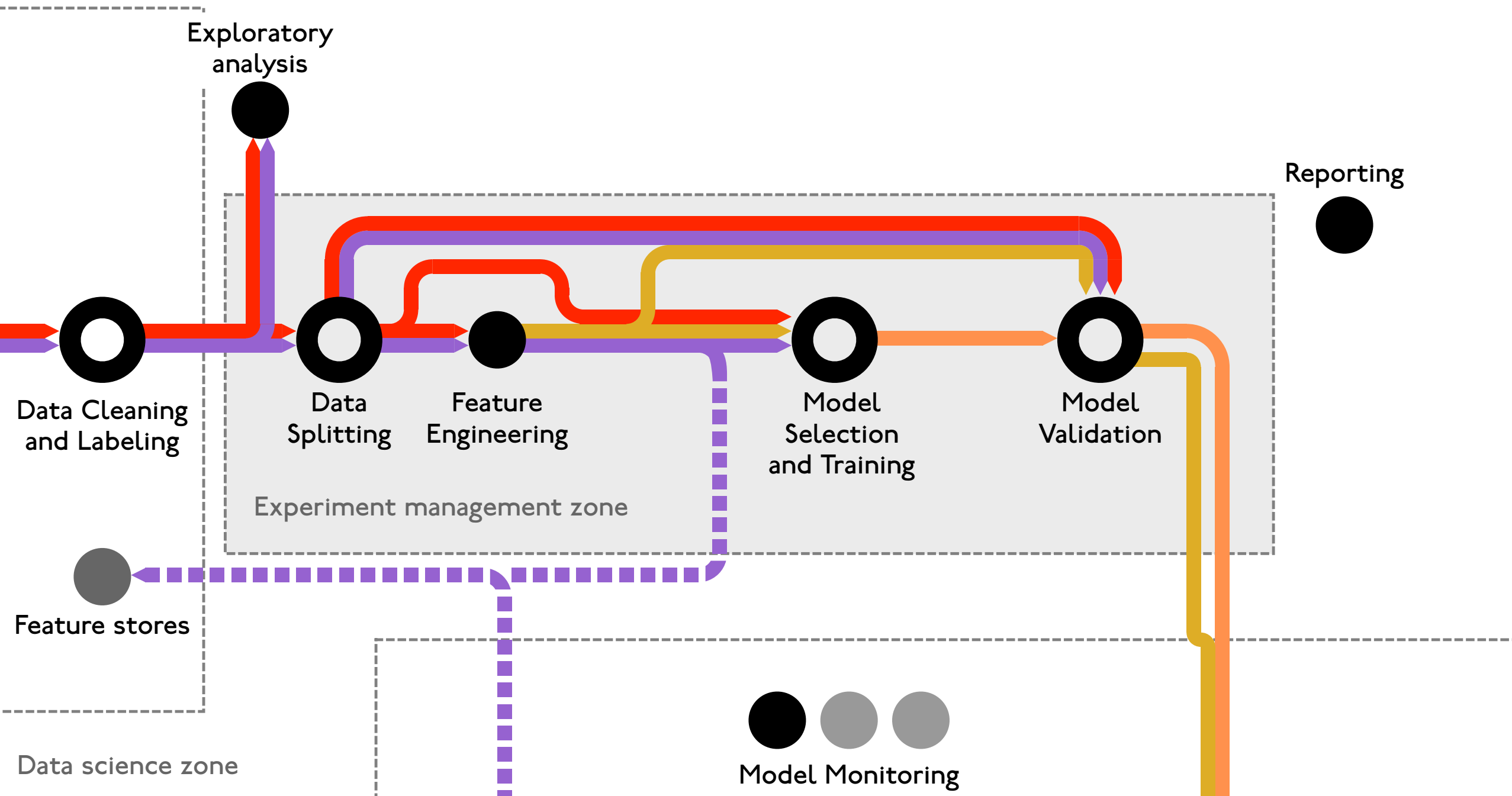


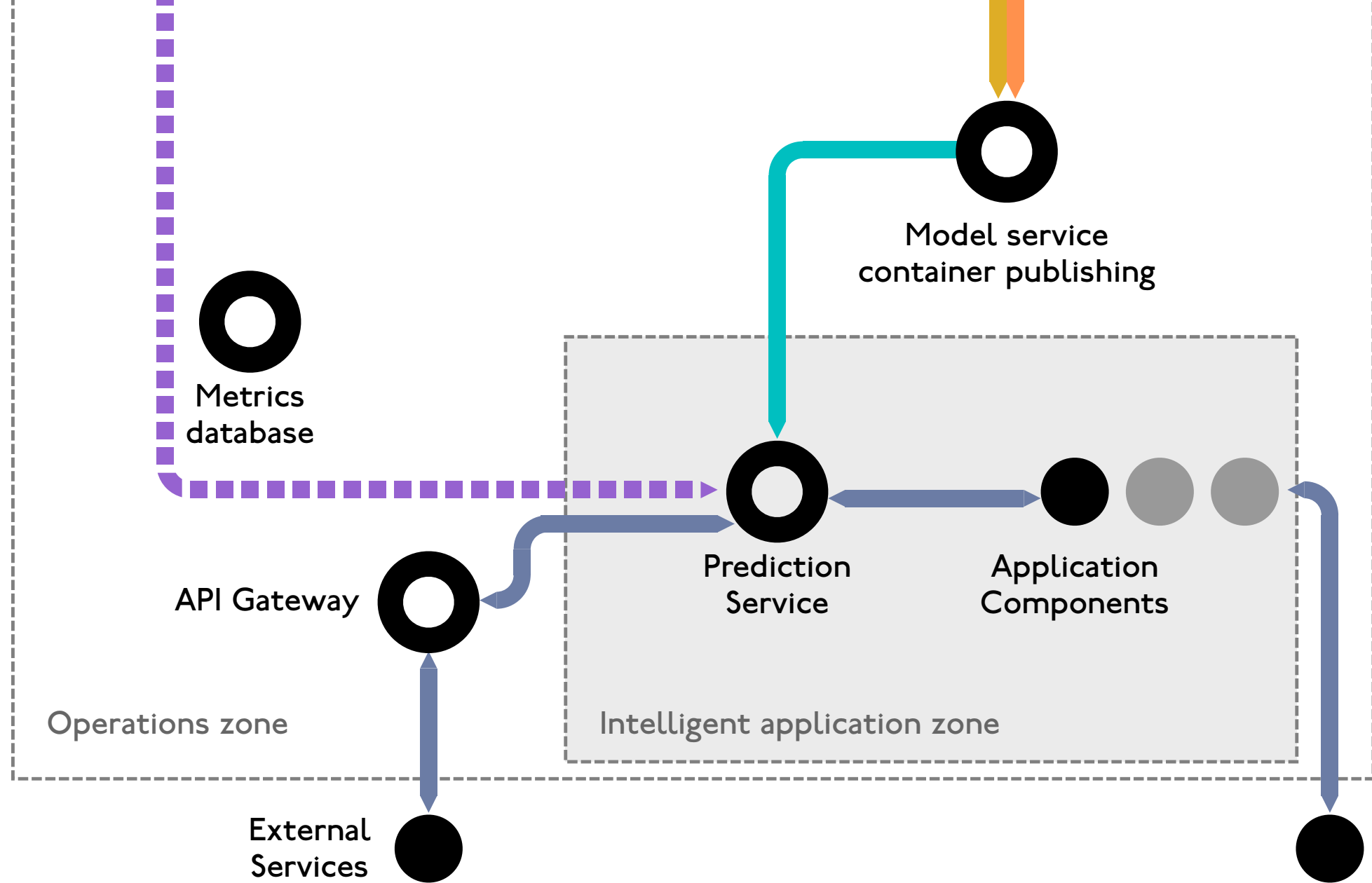


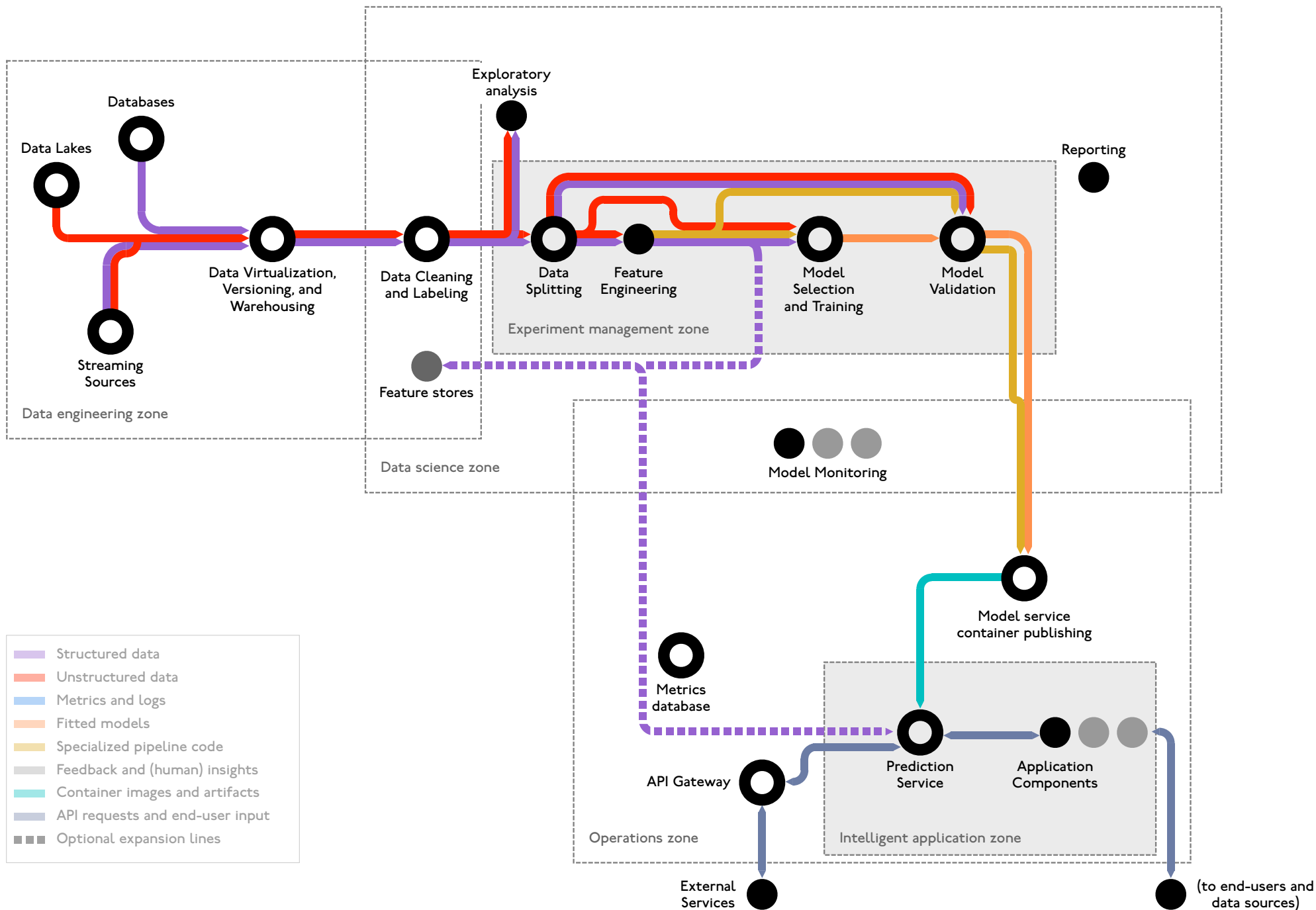
The big picture

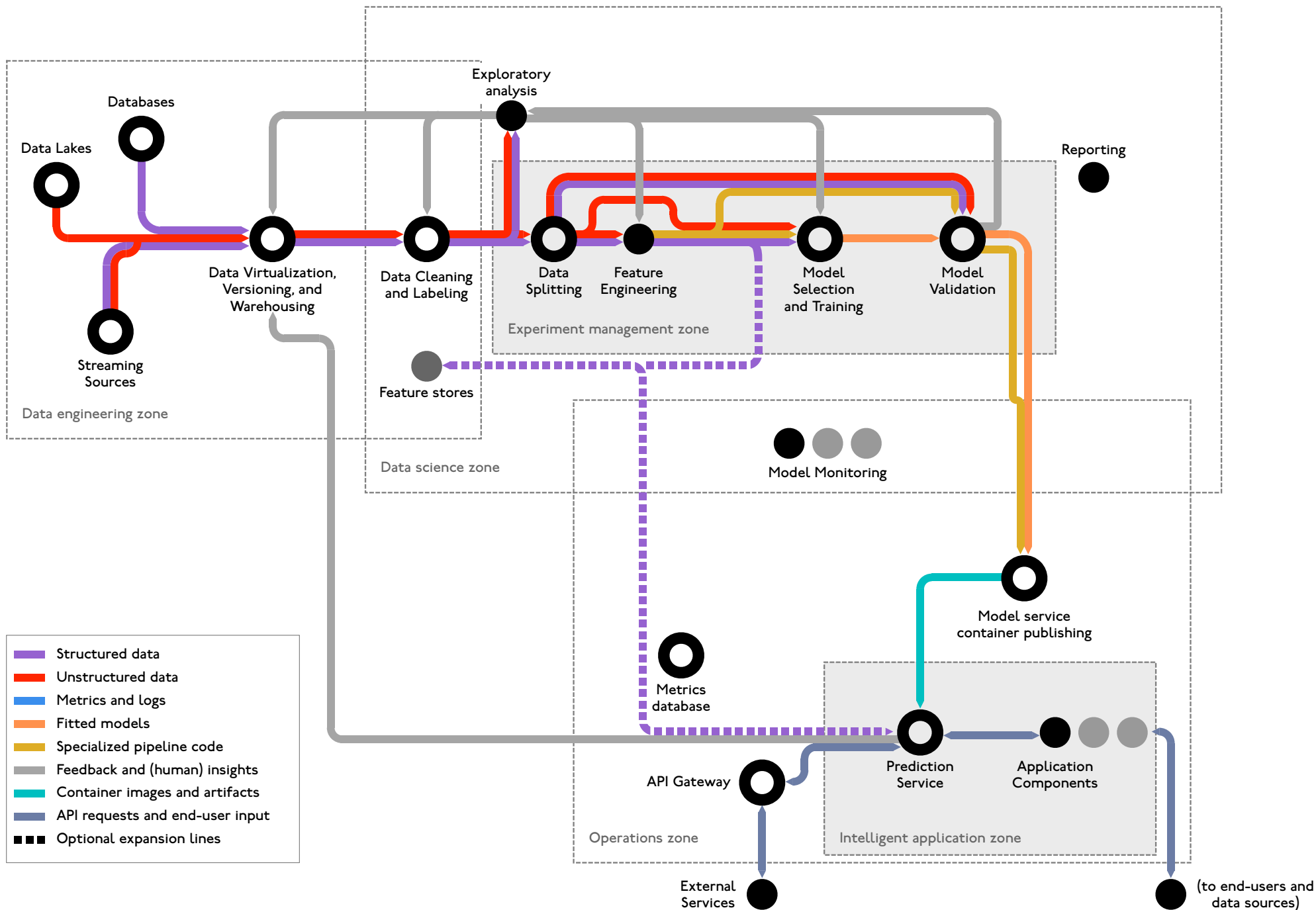


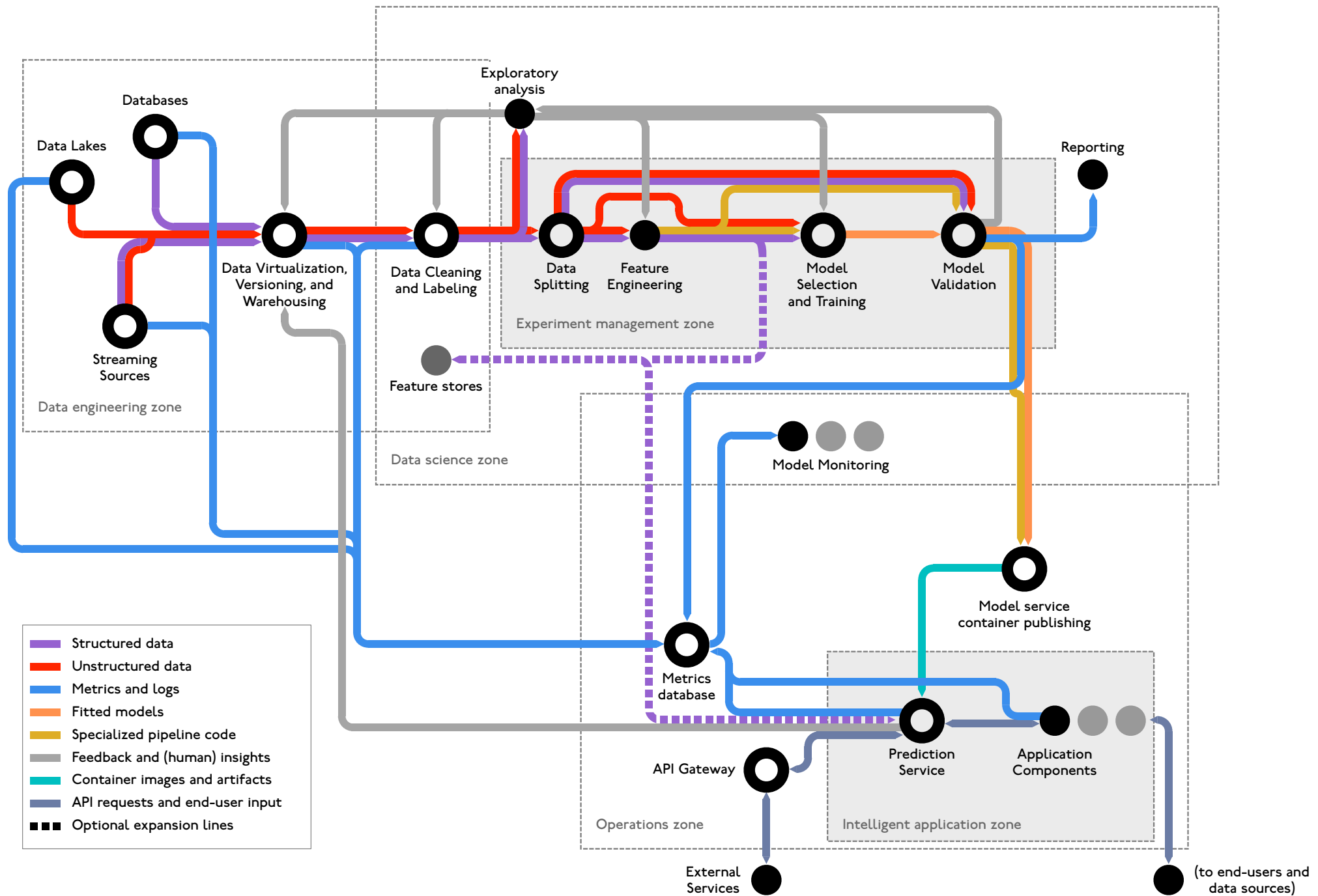


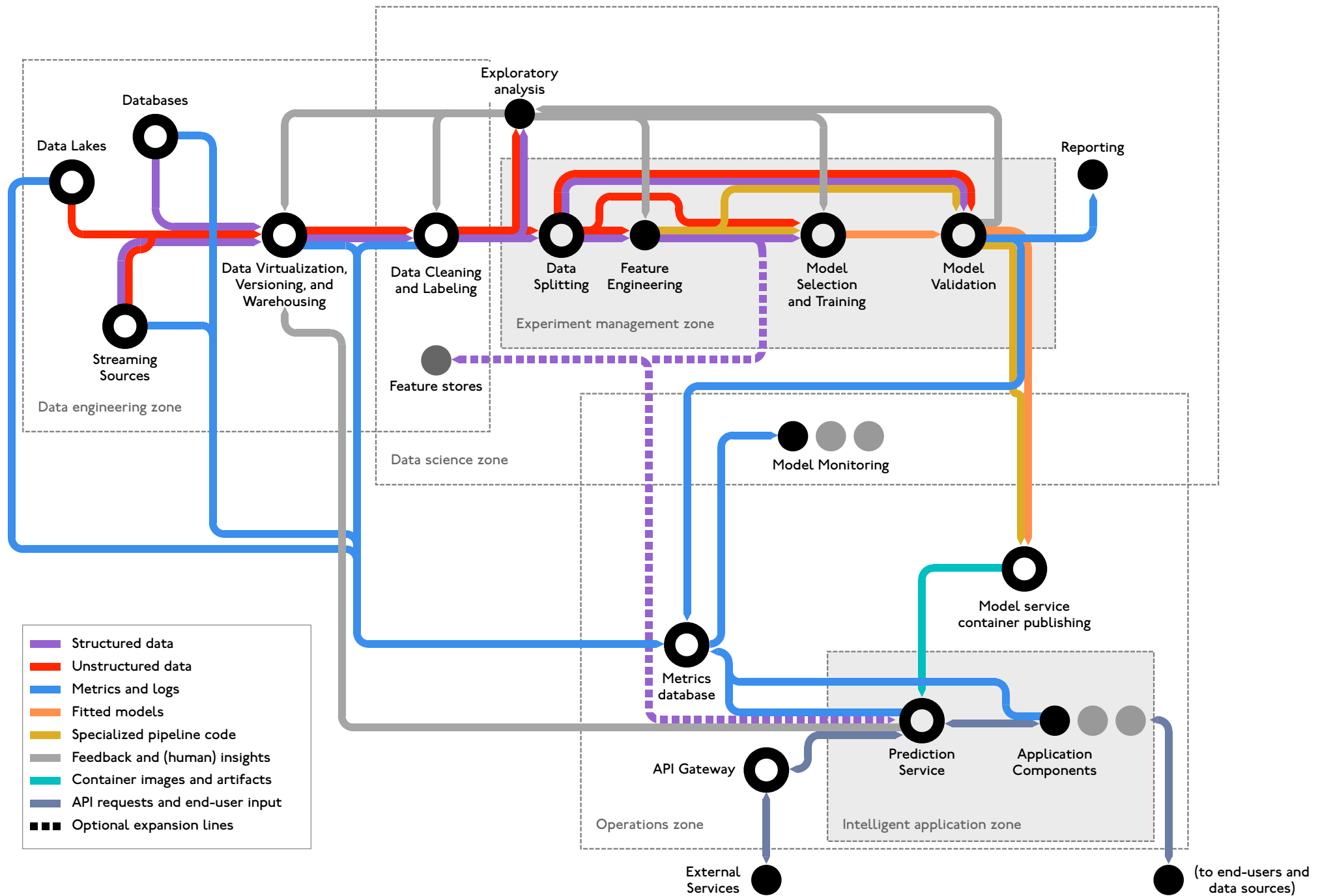


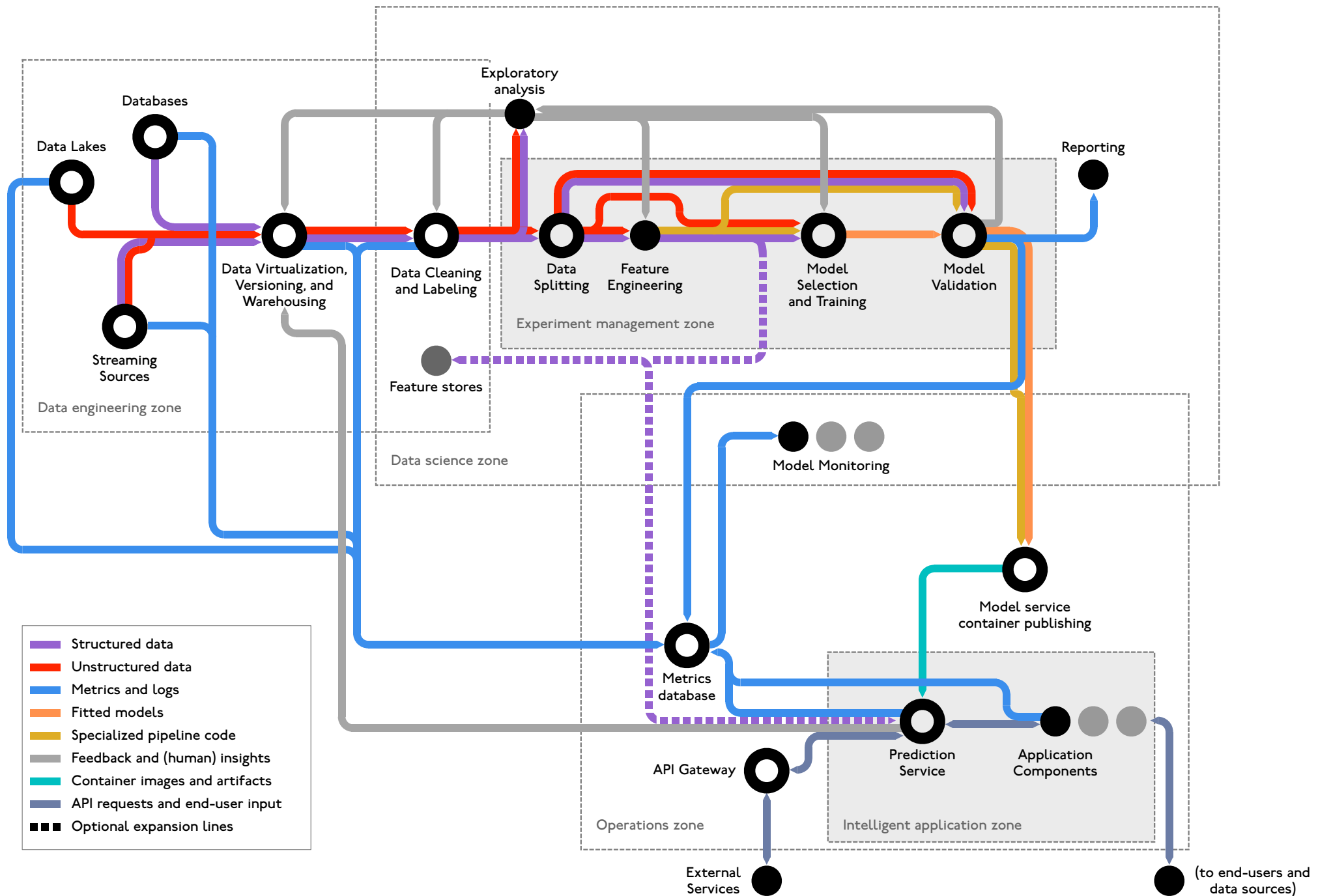


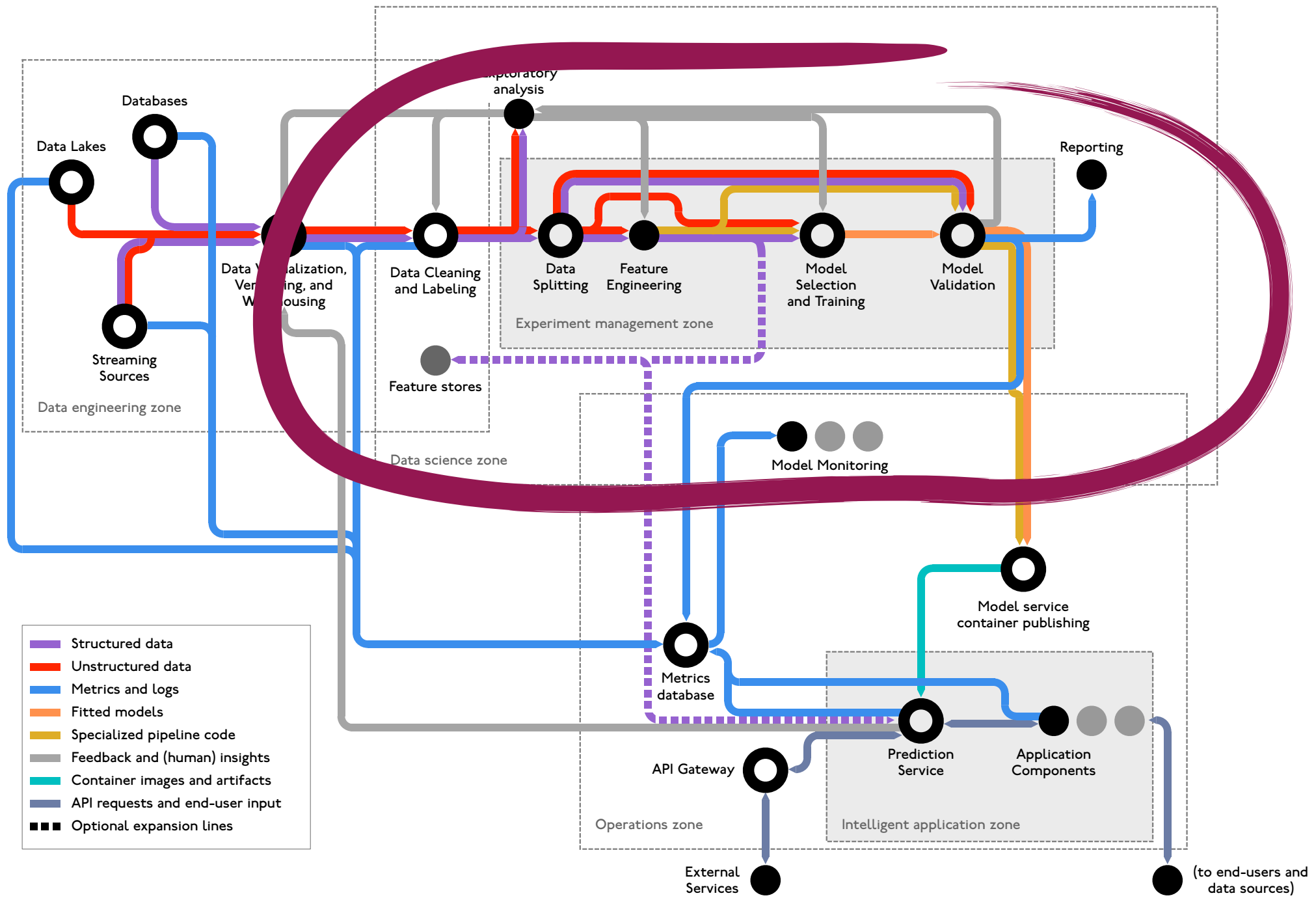


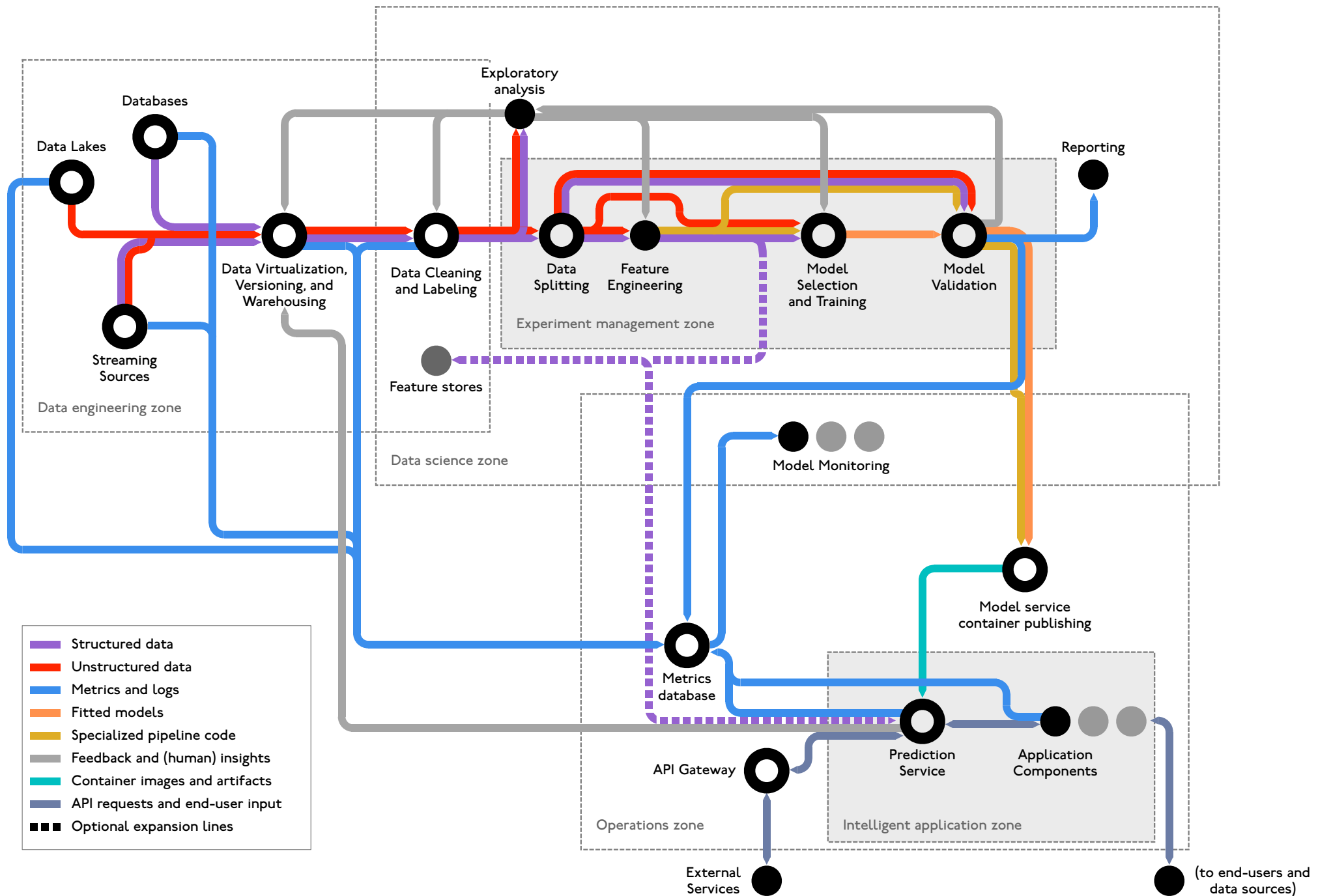


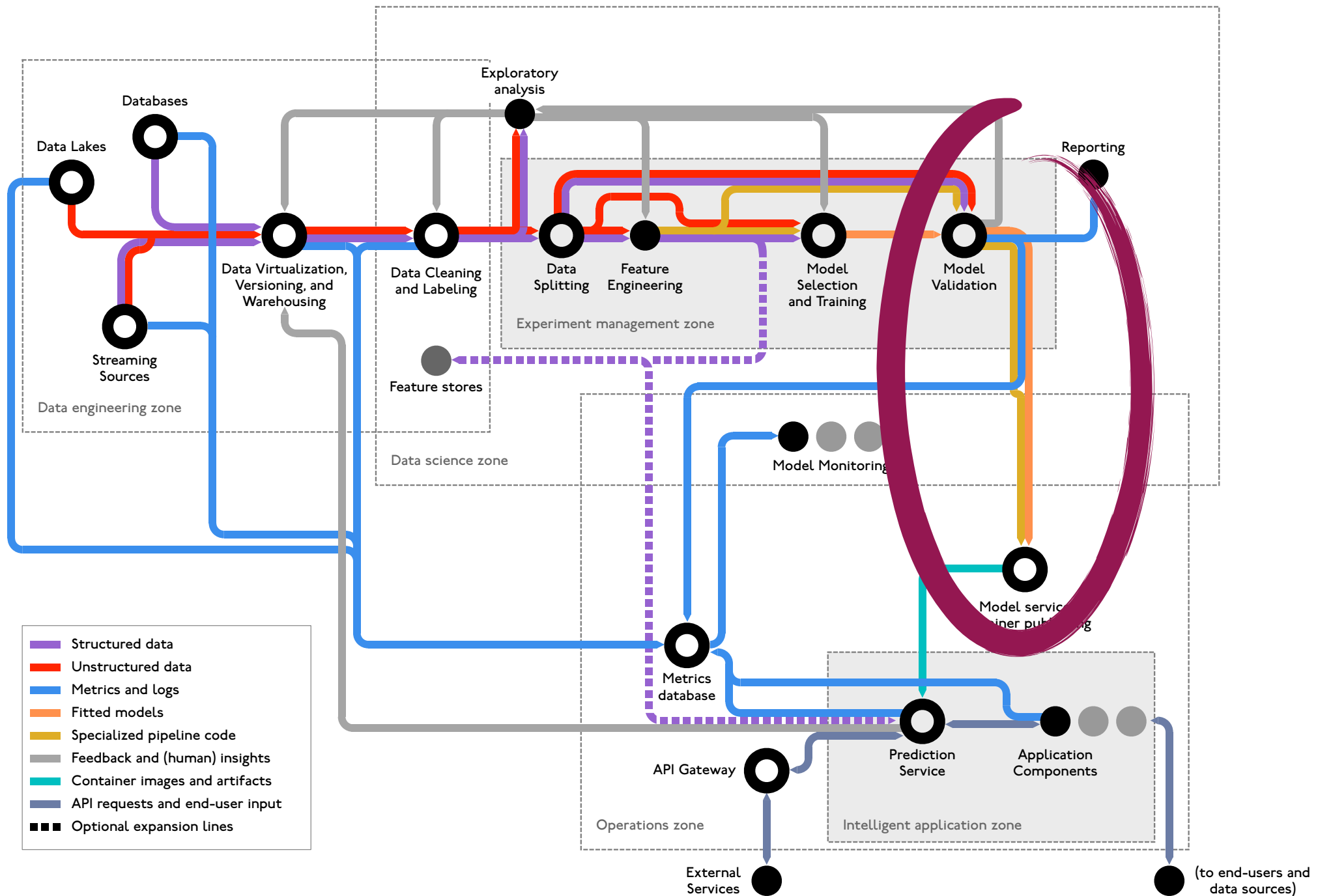


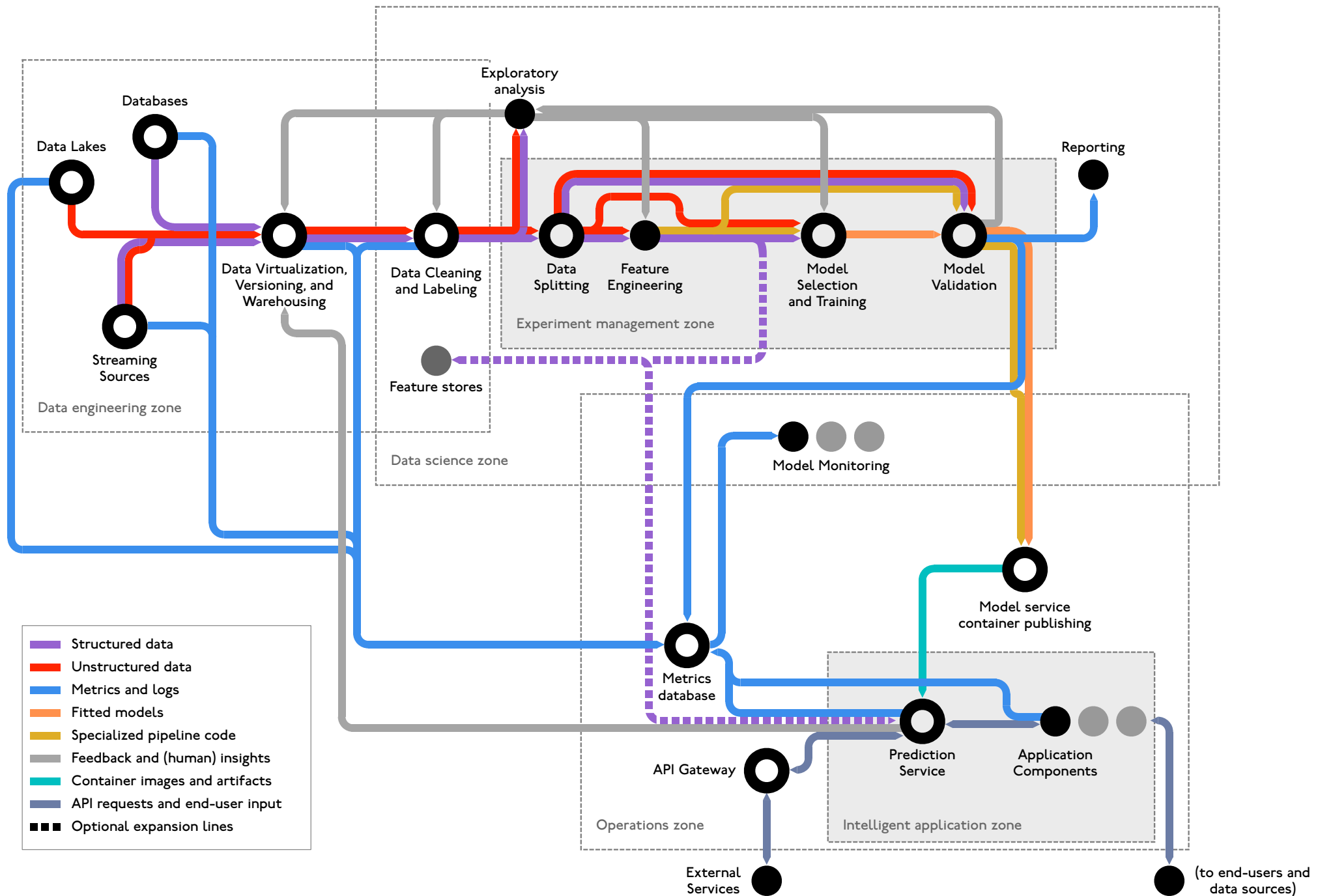


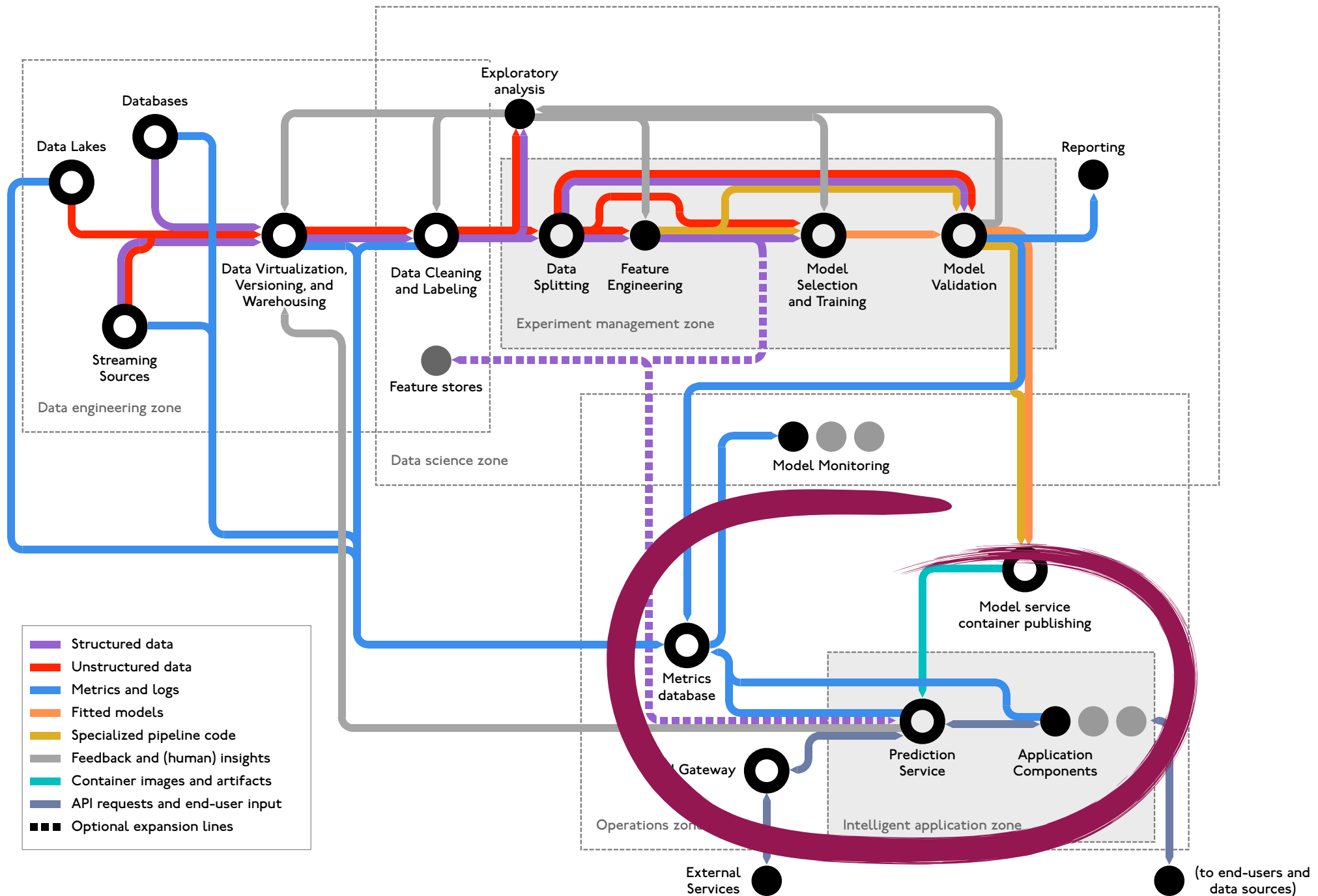


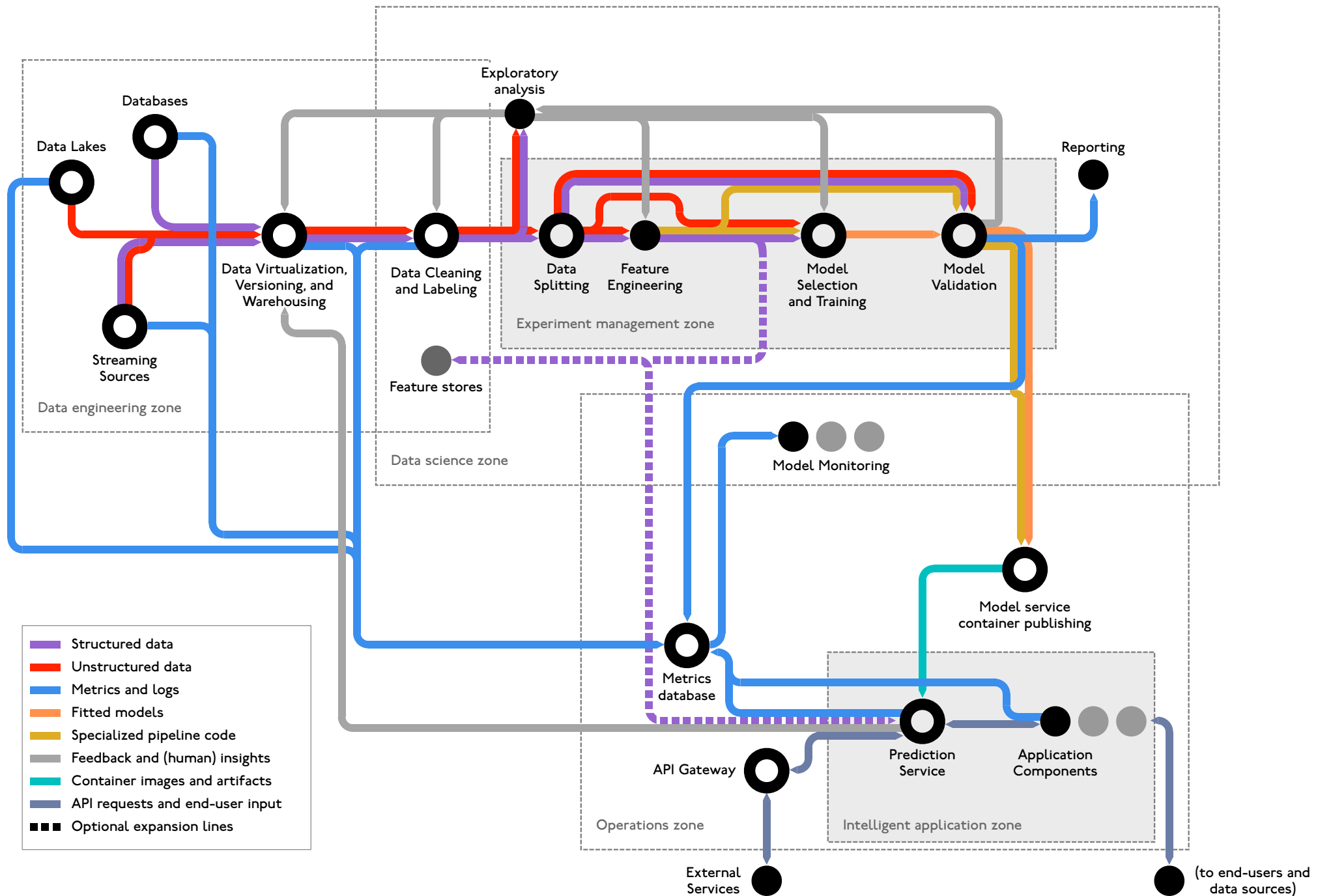


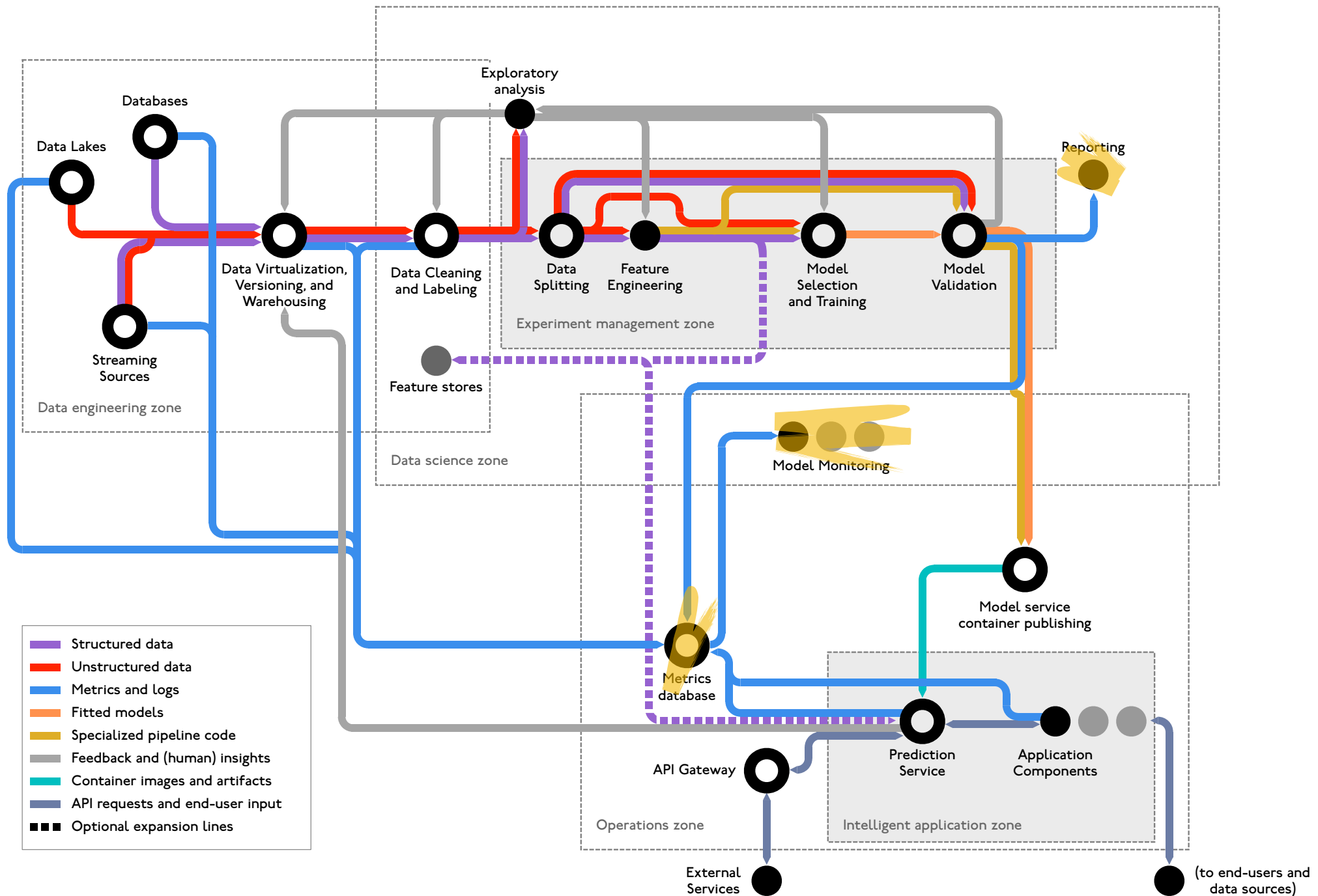






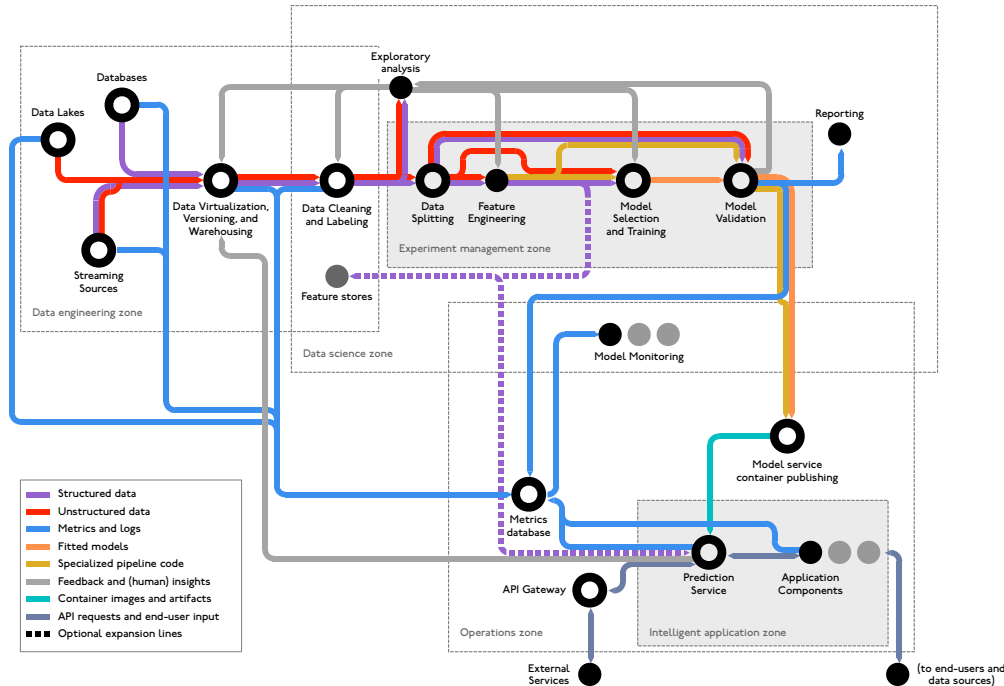






THANKS!

Kubernetes for ML:
Productivity over primitives:
<http://bit.ly/oscon-19>



Machine learning systems
and intelligent applications.
Benton, IEEE Software
<http://bit.ly/willb-ieee-iapps>

@sophwats • @willb

