

kokchun giang

# what is the **data** **engineering lifecycle?**



# what is **data engineering**?



*Data engineering* is the development, implementation and maintenance of systems and processes that take in raw data and produce high-quality, consistent information that supports downstream use cases, such as analysis and machine learning.

*Data engineering* is the intersection of security, data management, DataOps, data architecture, orchestration and software engineering.

A *data engineer* manages the data engineering lifecycle, beginning with getting the data from source systems and ending with serving data for use cases, such as analysis and machine learning.



# what is **data engineering**?



*Data engineering* is the development, implementation and maintenance of systems and processes that take in **raw data** and produce high-quality, consistent information that **supports downstream use cases**, such as analysis and machine learning.

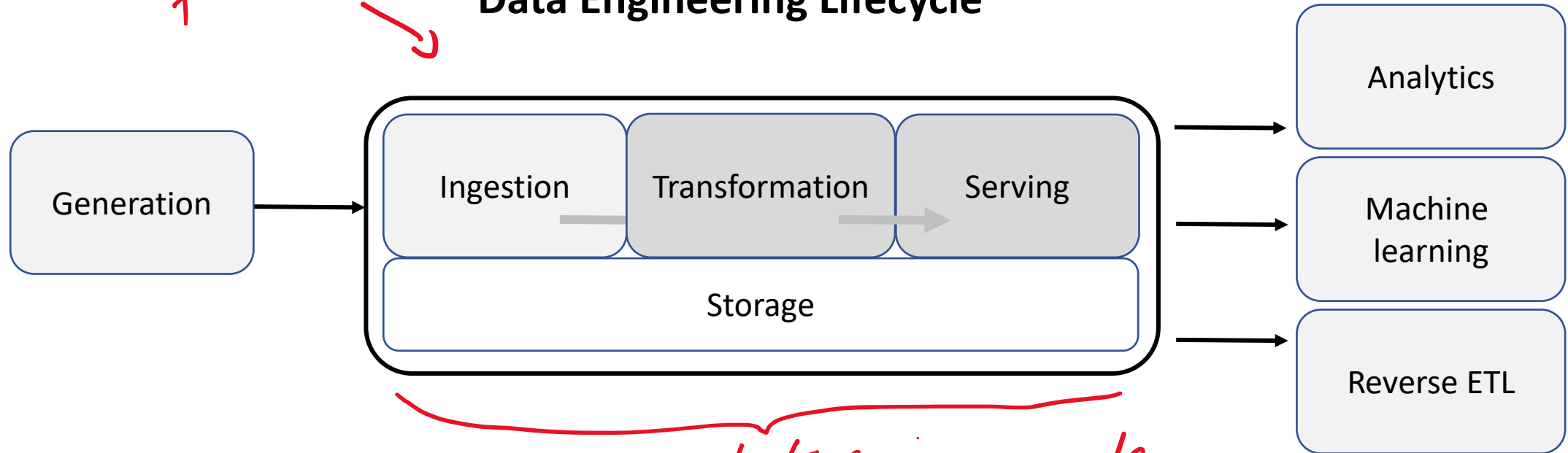
*Data engineering* is the intersection of **security, data management, DataOps, data architecture, orchestration and software engineering.**

A *data engineer* manages the data engineering lifecycle, beginning with getting the **data from source systems and ending with serving data for use cases, such as analysis and machine learning.**



*fun today*

## Data Engineering Lifecycle



*data engineer role*

### Undercurrents:

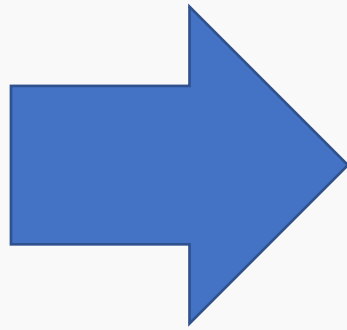
cut across the data engineering lifecycle



**ingesting data** from source system into pipeline



source systems



target/destination

*∃ different strategies  
for ingesting  
data*

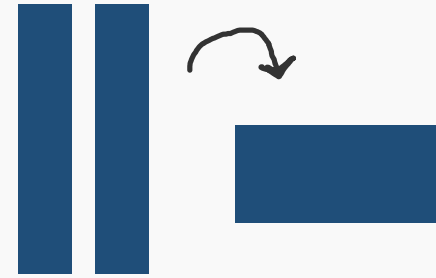
# transform data to fit your needs



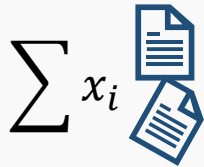
data integration



clean data



data transformation



data aggregation



data filtering

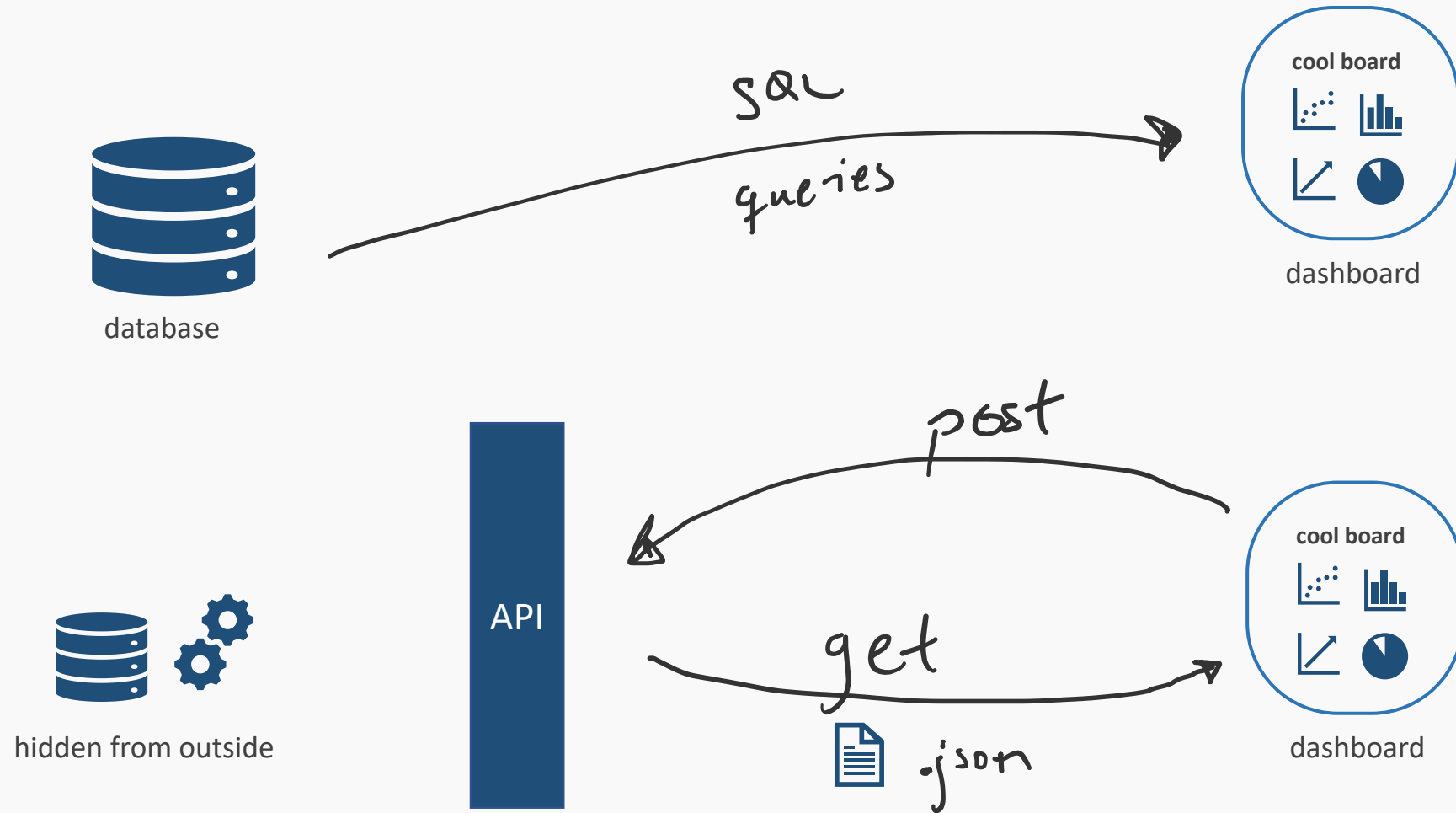


data enrichment

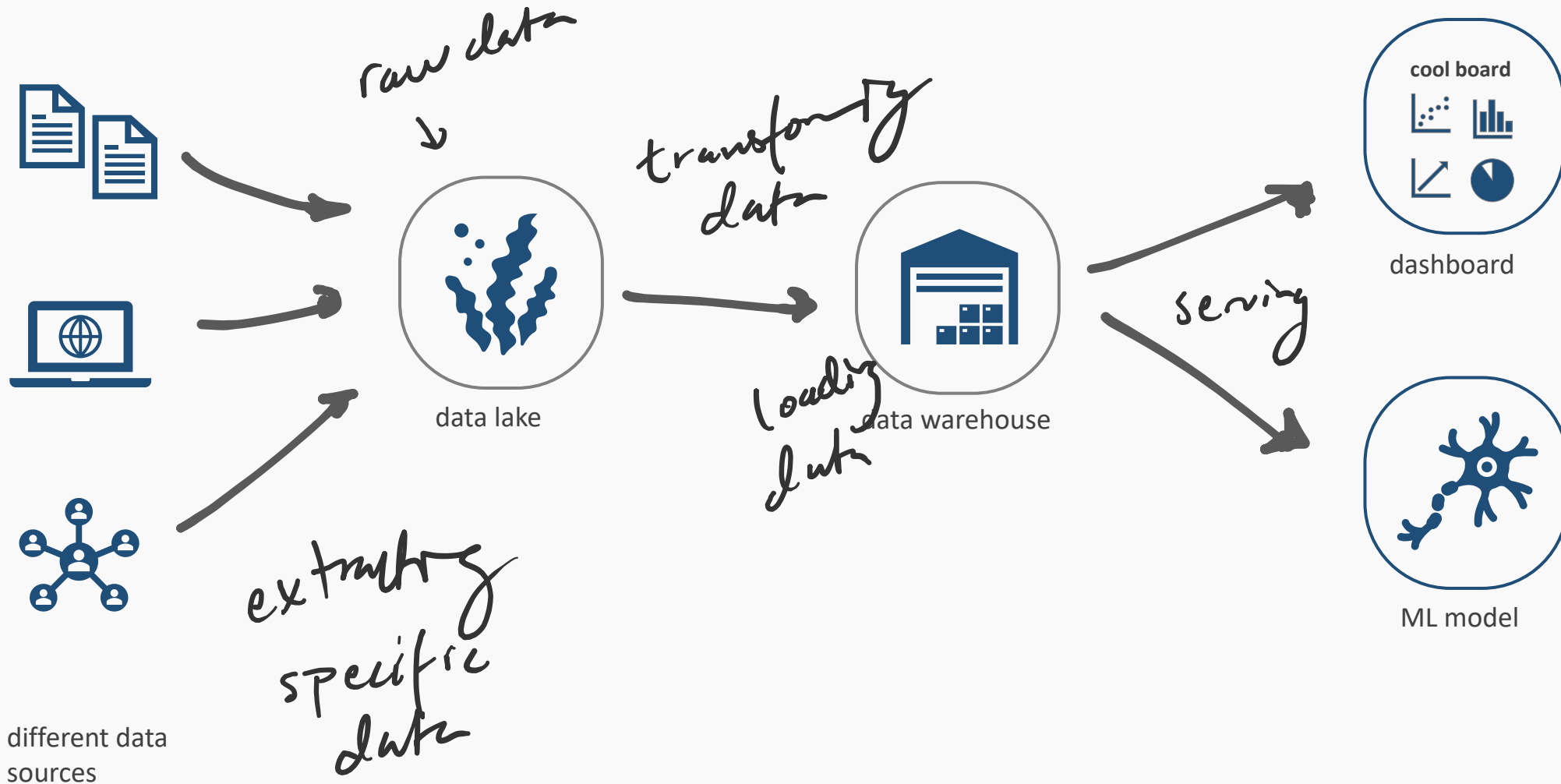


data masking

# serve data to downstream

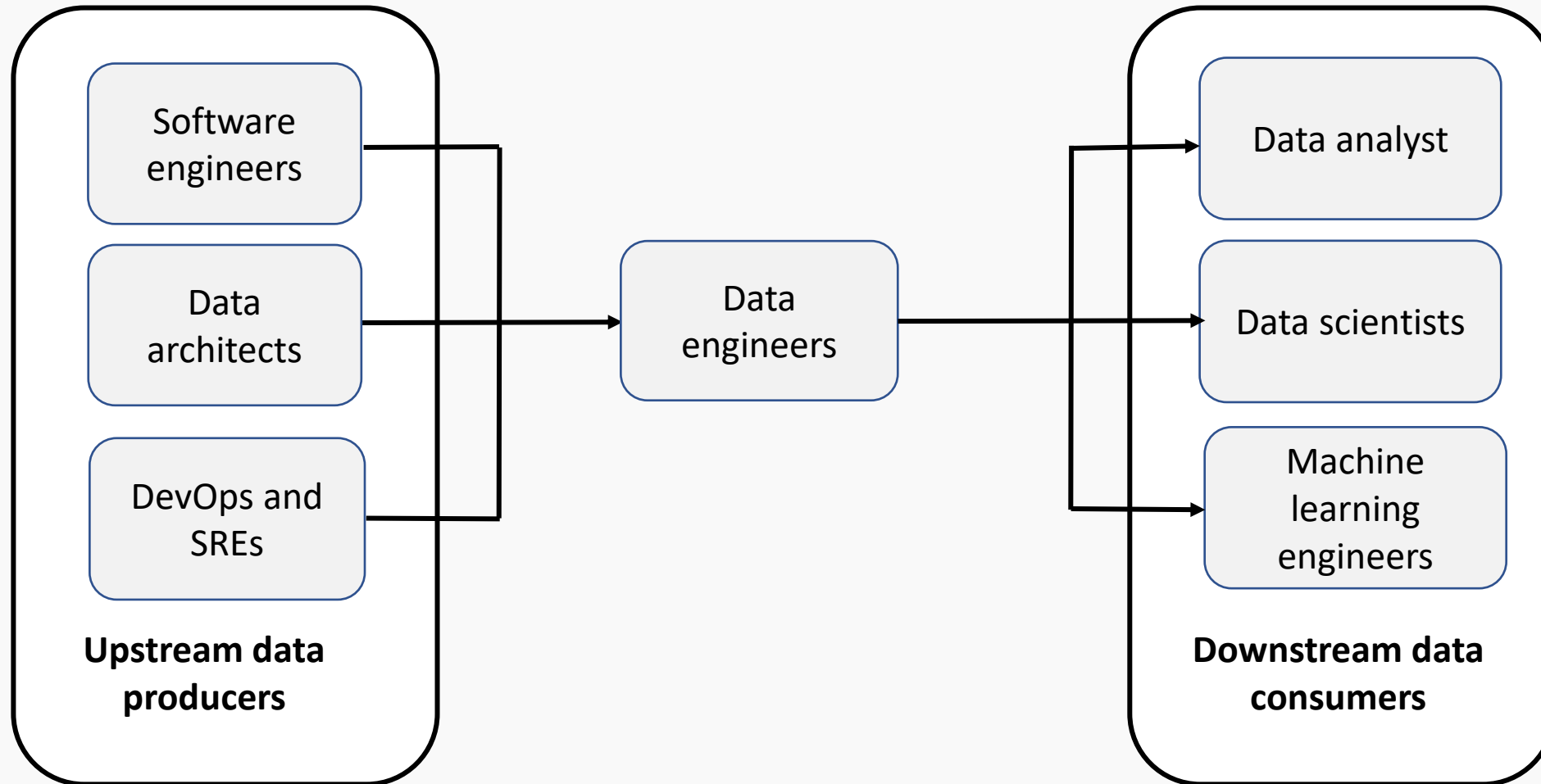


# example **ETL** data pipeline

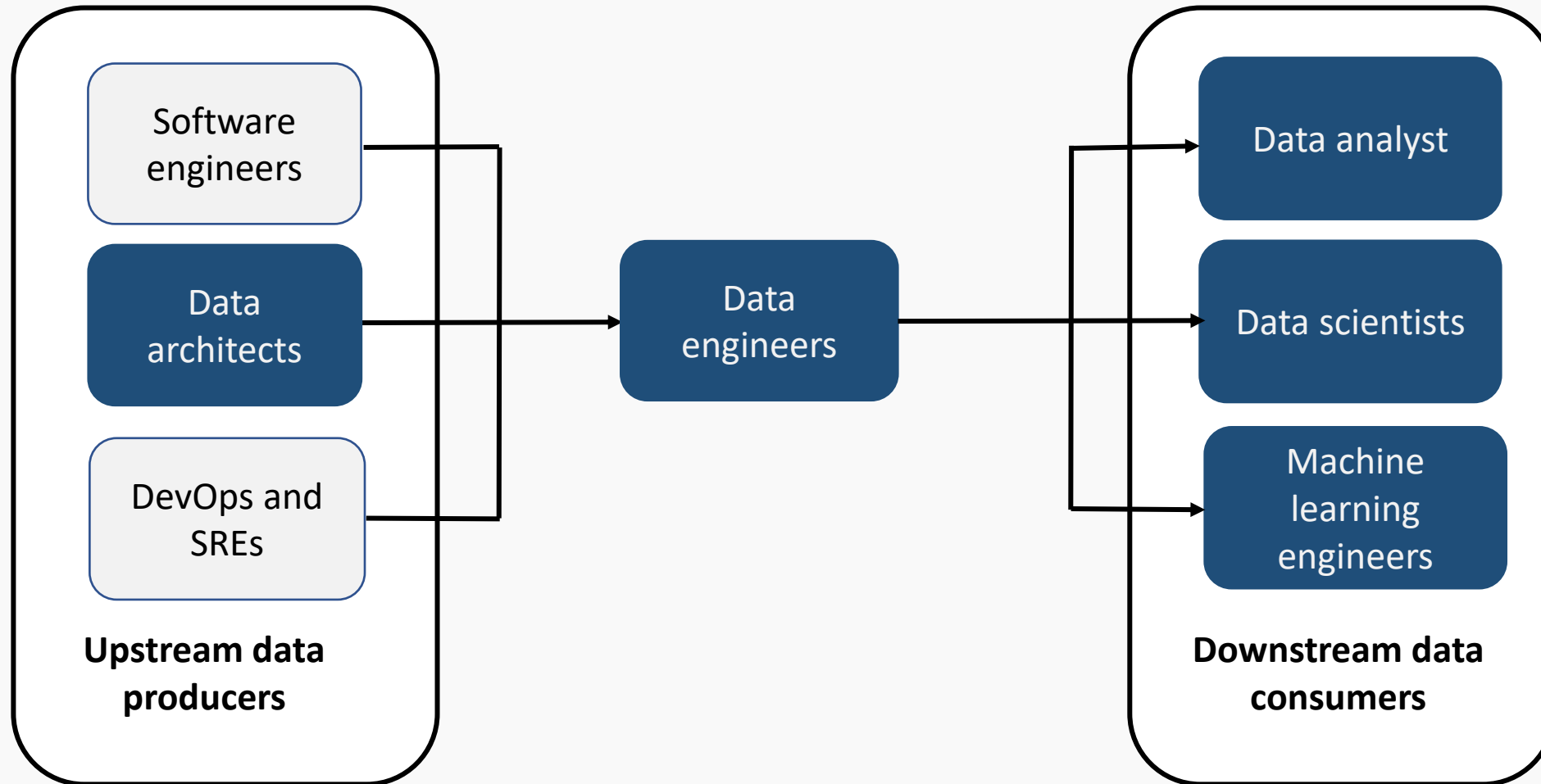




## different data **roles**



## different data **roles**

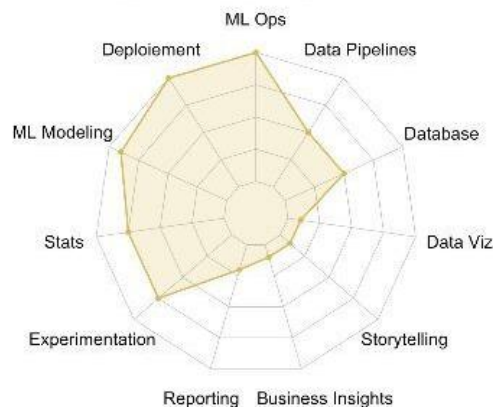


# different data **roles** and their **skills**

## Data Engineer



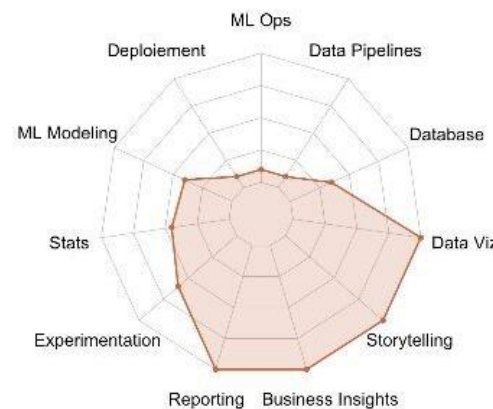
## ML Engineer



## Data Scientist



## Data Analyst



note that the skillsets and actual work vary a lot throughout different organizations

for example if you are the only data scientist in a smaller organization, you might have all these roles

Granted permission from: Kevin Rosamont Prombo on linkedin

business req.  
→ data architecture

company's  
data vision

data architect 


extract  
insights

predictive  
models

mathematical  
models

data scientist 

build & deploy  
ML models

machine learning  
engineer 

bridge gap  
between data  
science &  
software eng.

data roles

data engineer 

data analyst 

explore,  
clean data

visualize  
transform  
storytel

pipelines

data infrastructure

integrate  
to product  
systems