

Machine learning and data pyramid

MACHINE LEARNING FOR BUSINESS



Karolis Urbonas

Head of Machine Learning & Science,
Amazon

Machine Learning applications

ML is applying statistical or computer science methods on data to:

1. Draw causal insights

"What is causing our customers to cancel their subscription to our services?"*

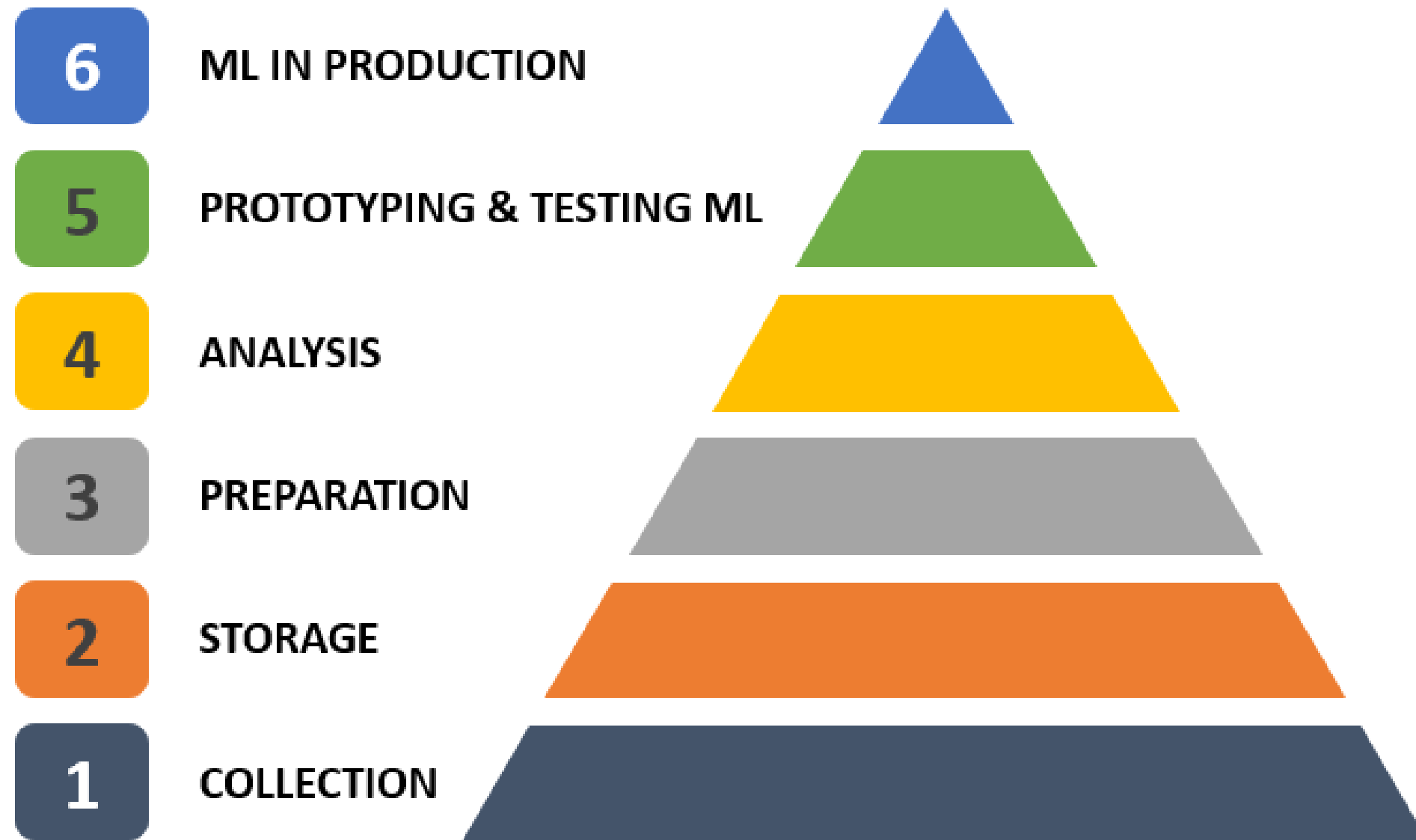
2. Predict future events

"Which customers are likely to cancel their subscription next month?"*

3. Understand patterns in data

"Are there groups of customers who are similar and use our services in a similar way?"*

Data hierarchy of needs



Collection

1

COLLECTION



Extract data from source systems

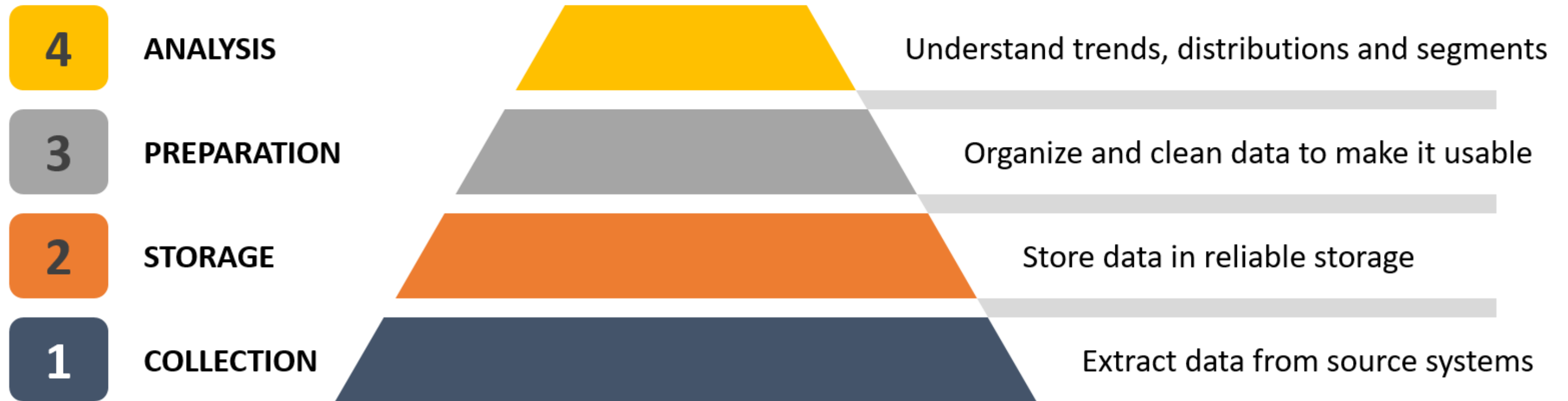
Storage



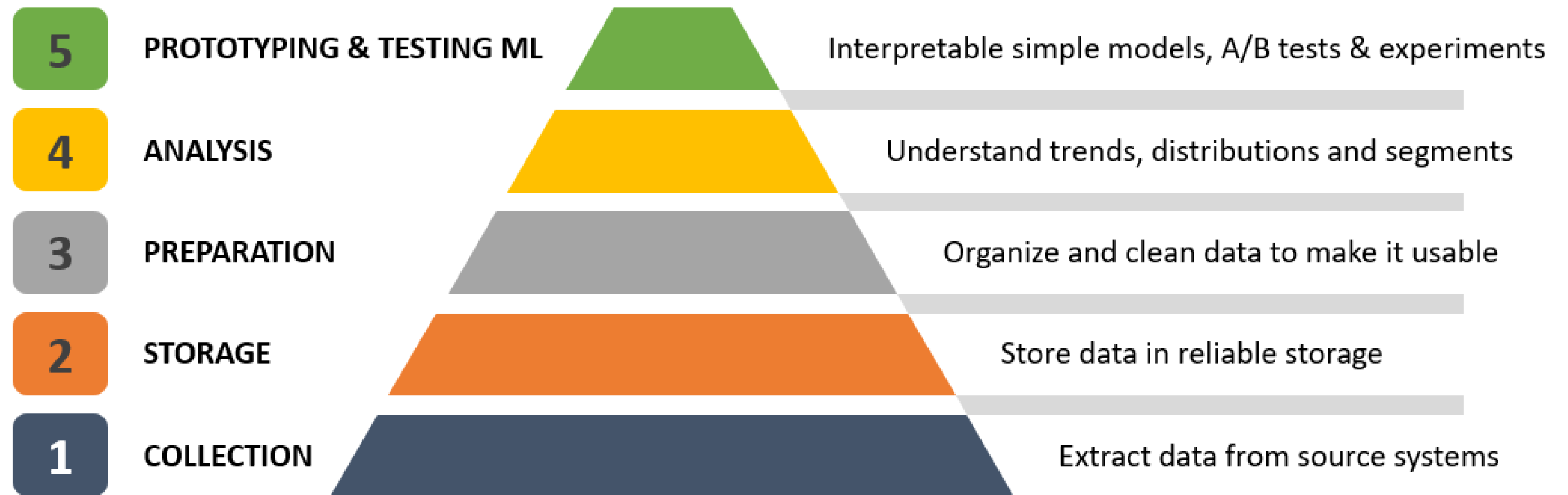
Preparation



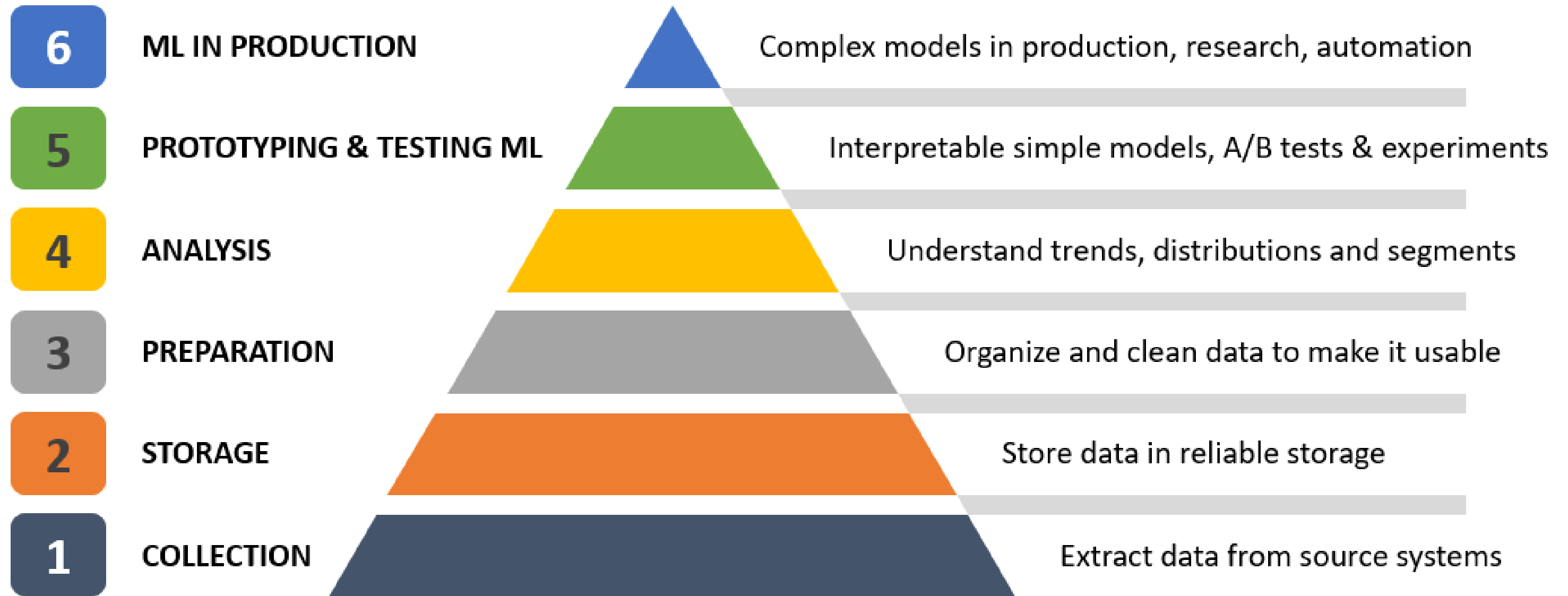
Analysis



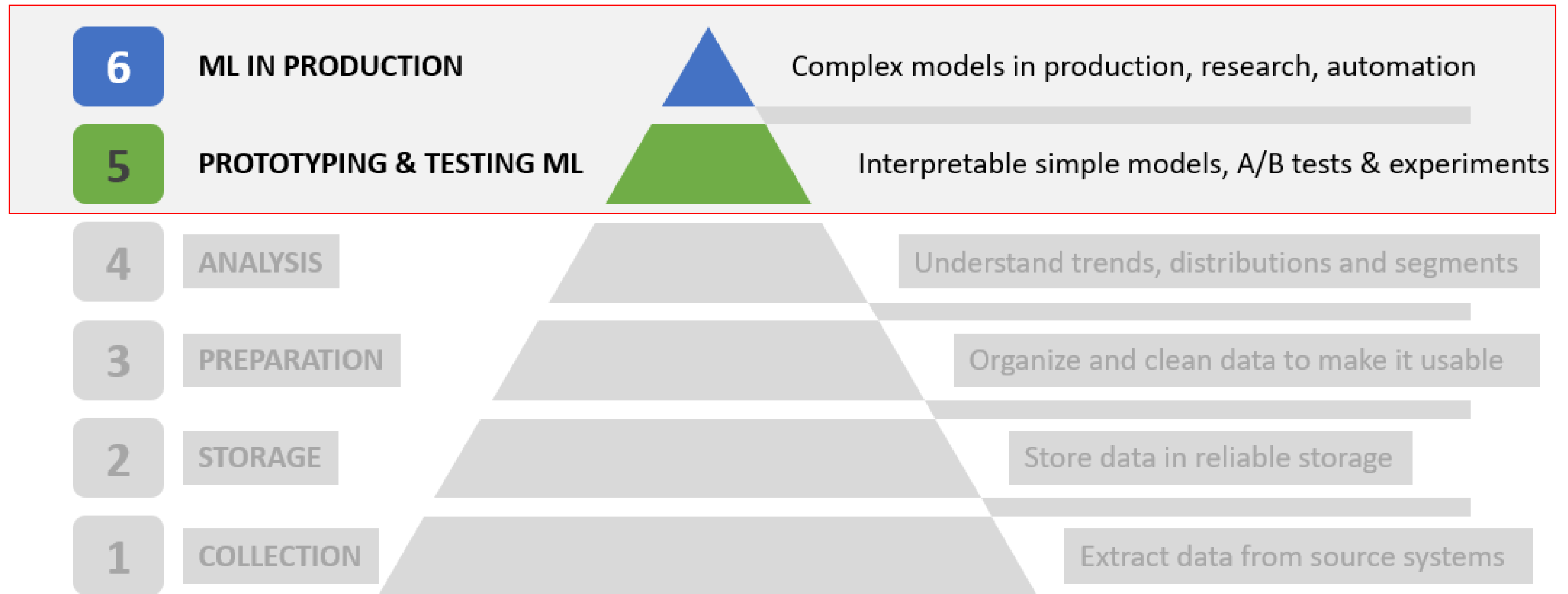
Model prototyping and testing



ML in production



Focus



Let's practice!

MACHINE LEARNING FOR BUSINESS

Machine learning principles

MACHINE LEARNING FOR BUSINESS



Karolis Urbonas

Head of Machine Learning & Science,
Amazon

Machine learning types

Machine learning - applying statistical or computer science methods on data to:

1. Draw causal insights

"What is causing our customers to cancel their subscription to our services?"

2. Predict future events

"Which customers are likely to cancel their subscription next month?"

3. Understand patterns in data

"Are there groups of customers who are similar and use our services in a similar way?"

Supervised vs. unsupervised ML

Machine learning - applying statistical or computer science methods on data to:

1. Draw causal insights

SUPERVISED Machine Learning

2. Predict future events

SUPERVISED Machine Learning

3. Understand patterns in data

UNSUPERVISED Machine Learning

Supervised ML data structure

	Transaction data A	Transaction data B	Transaction data C	Transaction data D		Fraud probability
Transaction 1						
Transaction 2						
Transaction 3						
Transaction ...						
Transaction N						

Target variable

Transaction 1
Transaction 2
Transaction 3
Transaction ...
Transaction N

Transaction data A	Transaction data B	Transaction data C	Transaction data D

Target variable

Fraud probability

Input features

Data about transactions that the business collected
(input features)

	Transaction data A	Transaction data B	Transaction data C	Transaction data D
Transaction 1				
Transaction 2				
Transaction 3				
Transaction ...				
Transaction N				

Target variable

Fraud probability

Example input features

	Past fraud count	Time of transaction	Declined in T-30 days	Amount
Transaction 1	20	3 am	Yes	5.25 USD
Transaction 2	1	9 pm	Yes	19.5 USD
Transaction 3	0	9.30 am	No	500 USD
Transaction ...				
Transaction N				

Fraud
Yes
Yes
No

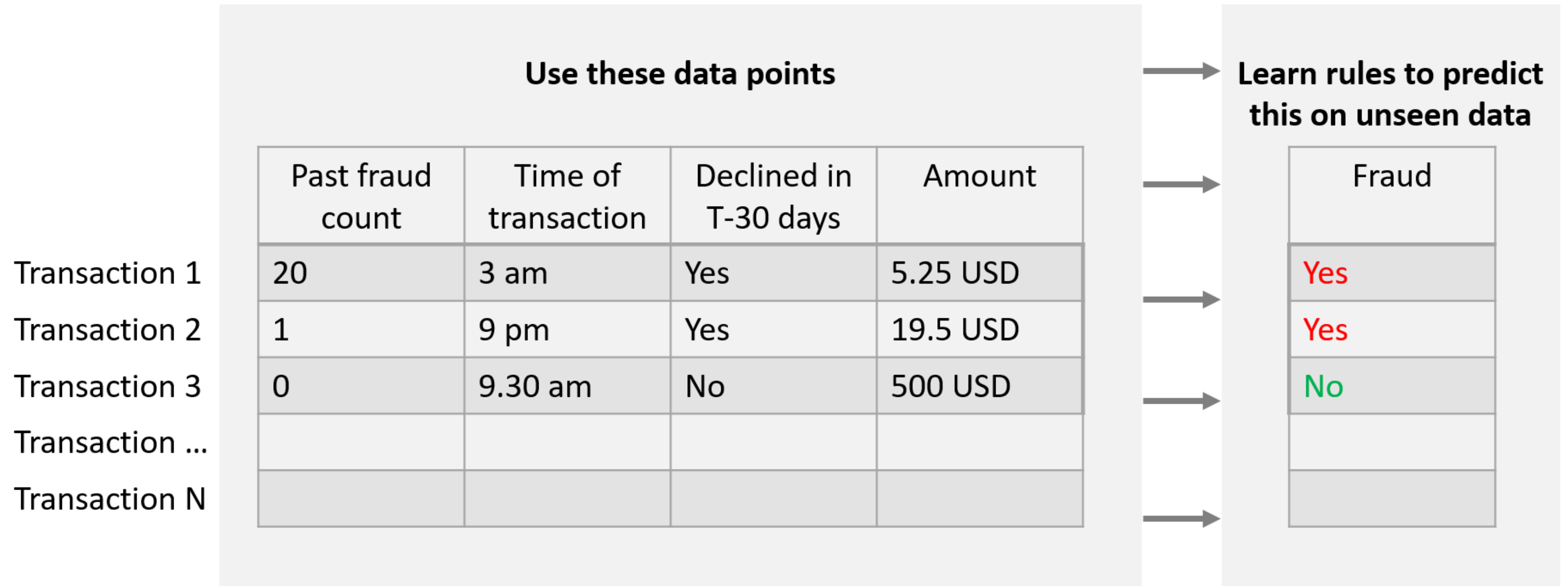
Using input features

Use these data points

	Past fraud count	Time of transaction	Declined in T-30 days	Amount
Transaction 1	20	3 am	Yes	5.25 USD
Transaction 2	1	9 pm	Yes	19.5 USD
Transaction 3	0	9.30 am	No	500 USD
Transaction ...				
Transaction N				

Fraud
Yes
Yes
No

Predicting target variable



Unsupervised ML data structure

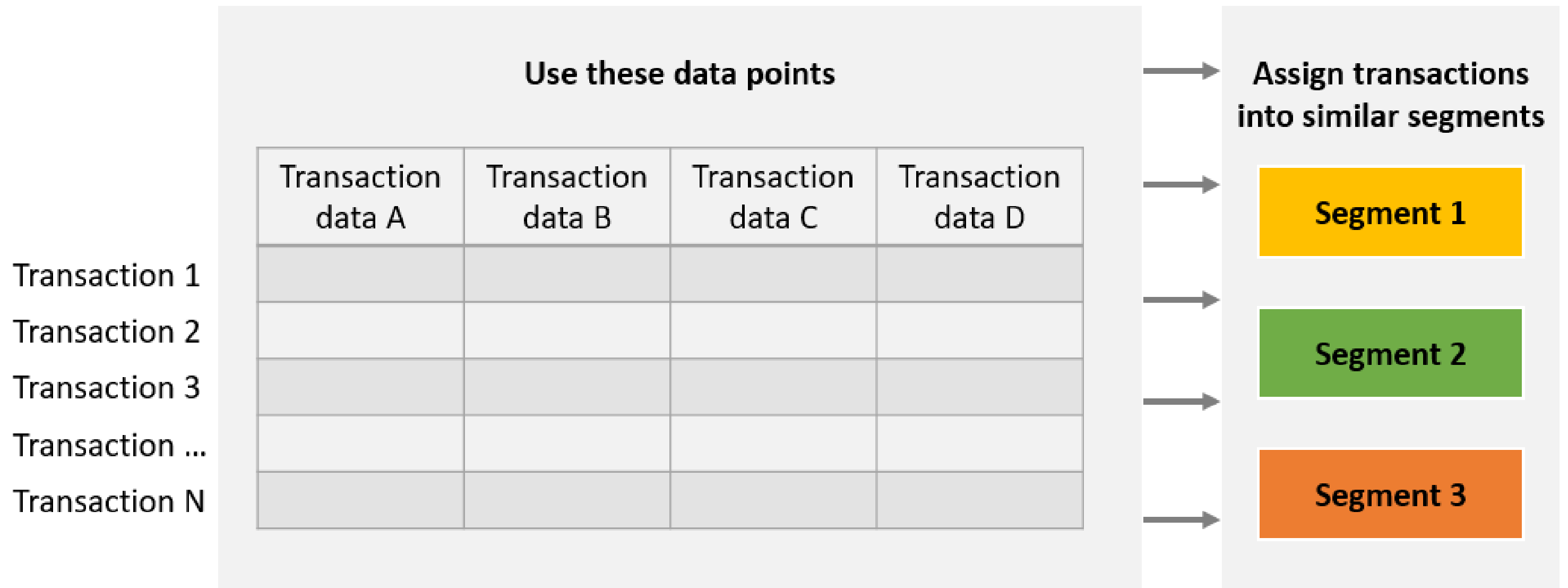
	Transaction data A	Transaction data B	Transaction data C	Transaction data D
Transaction 1				
Transaction 2				
Transaction 3				
Transaction ...				
Transaction N				

Unsupervised input features

Use these data points

	Transaction data A	Transaction data B	Transaction data C	Transaction data D
Transaction 1				
Transaction 2				
Transaction 3				
Transaction ...				
Transaction N				

Unsupervised ML results



ML examples - Marketing

SUPERVISED Machine Learning:

- Predict which customers are likely to purchase next month
- Predict each customer's expected lifetime value

UNSUPERVISED Machine Learning:

- Group customers into segments based on their past purchases

ML examples - Finance

SUPERVISED Machine Learning:

- Identify key transaction attributes that indicate a potential fraud
- Predict which customers will default on their mortgage payments

UNSUPERVISED Machine Learning:

- Group transactions into segments based on their attributes to understand which segments are the most profitable

ML examples - Manufacturing

SUPERVISED Machine Learning:

- Predict which items in production are likely faulty and should be manually inspected
- Predict which machines are likely to break and need maintenance

UNSUPERVISED Machine Learning:

- Group readings from machine sensors and identify anomalies for potential manufacturing malfunctions

ML examples - Transportation

SUPERVISED Machine Learning:

- Predict the expected delivery of the parcel
- Identify the fastest route for driving
- Predict product demand to prepare enough stock, rent/buy vehicles and hire workers

Let's practice!

MACHINE LEARNING FOR BUSINESS

Job roles, tools and technologies

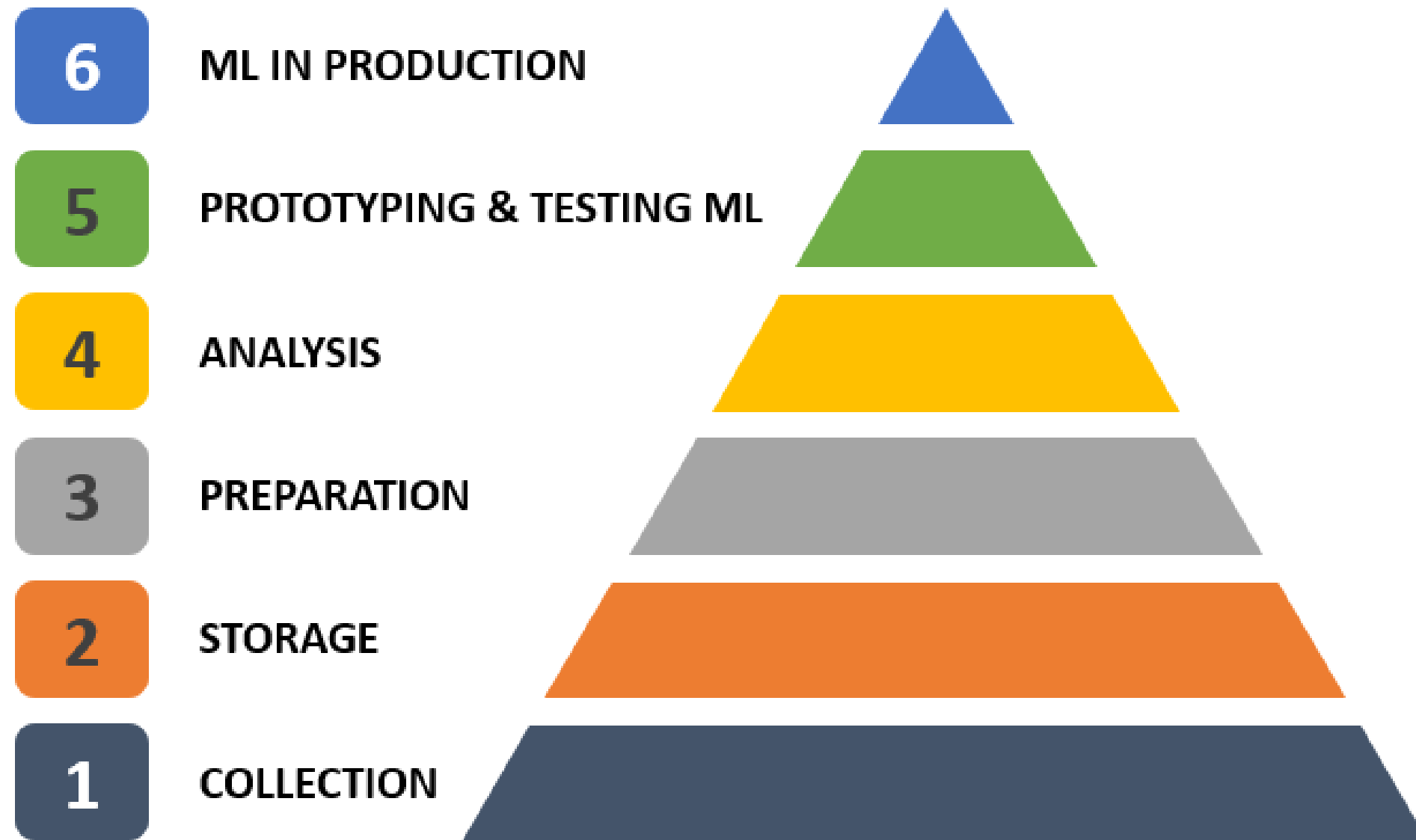
MACHINE LEARNING FOR BUSINESS



Karolis Urbonas

Head of Machine Learning & Science,
Amazon

Data pyramid and roles



Infrastructure owner

1

COLLECTION



Infrastructure owners

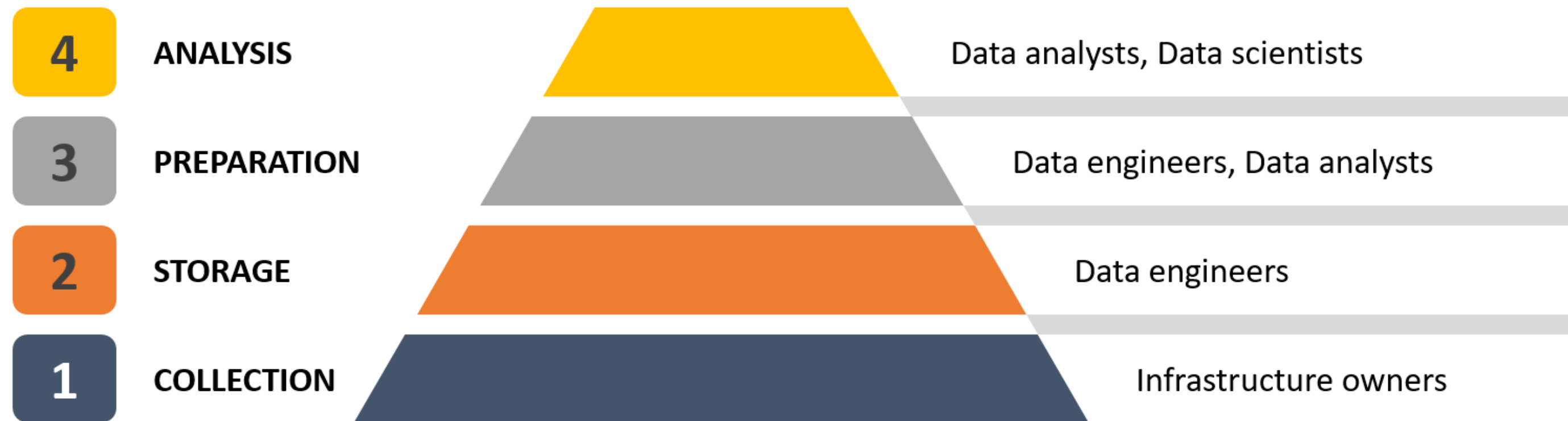
Data Engineer



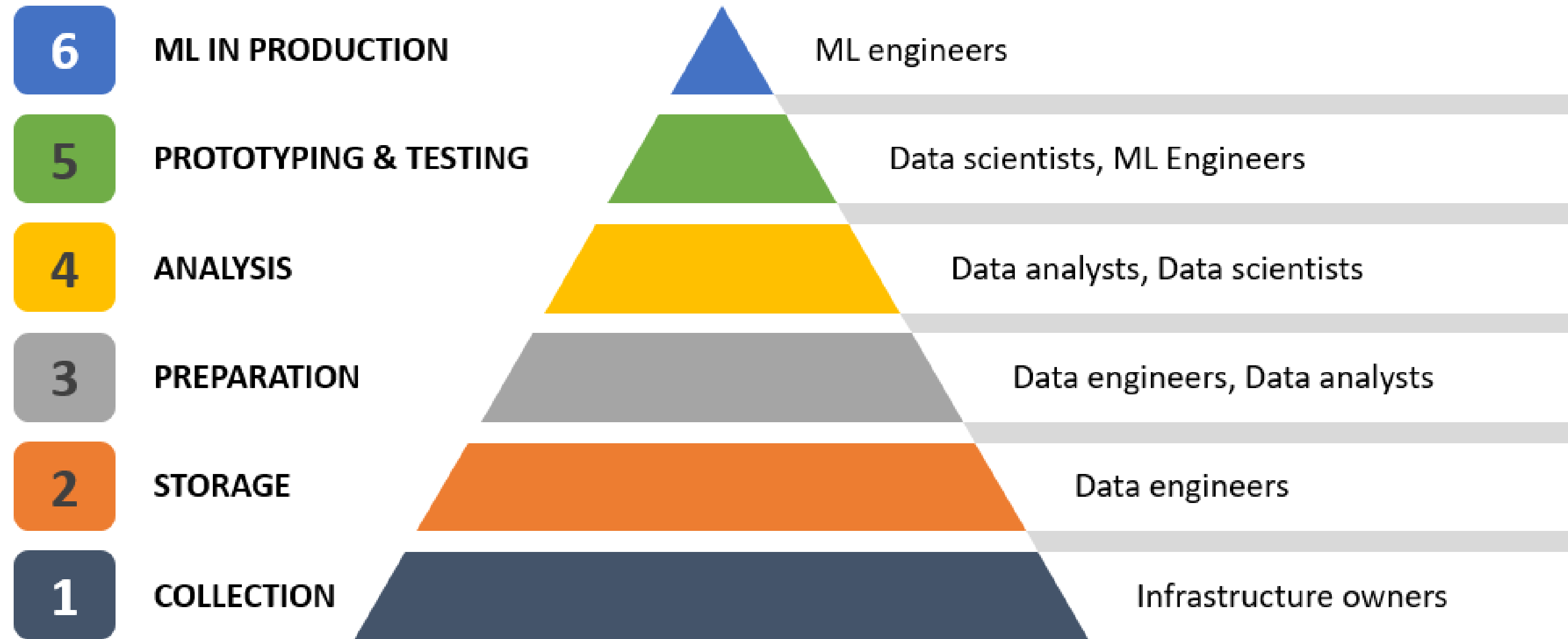
Data Analyst



Data Scientist



Machine Learning Engineer



Team structure

1. Centralized
2. Decentralized (or embedded)
3. Hybrid

Team structure comparison

1. **Centralized** - all data functions in one central team. Works well for small companies, startups, new organizations. Gets slow once business matures and requires focus
2. **Decentralized** - each business unit, geography or department have their own data functions. Works well for larger companies. Introduces issues with data governance, differences in definitions, redundancies, and added complexity
3. **Hybrid** - infrastructure, definitions, methods and tooling are **centralized**, while application and prototyping **decentralized**

Let's practice!

MACHINE LEARNING FOR BUSINESS