



kokchun giang

# building a modern data platform for **streaming data**



example robotic arm for welding needs **temperature monitoring**



AI generated with DALL-E-3

requires temperatures within certain ranges

temperature measured with temperature sensor is sent to a data platform

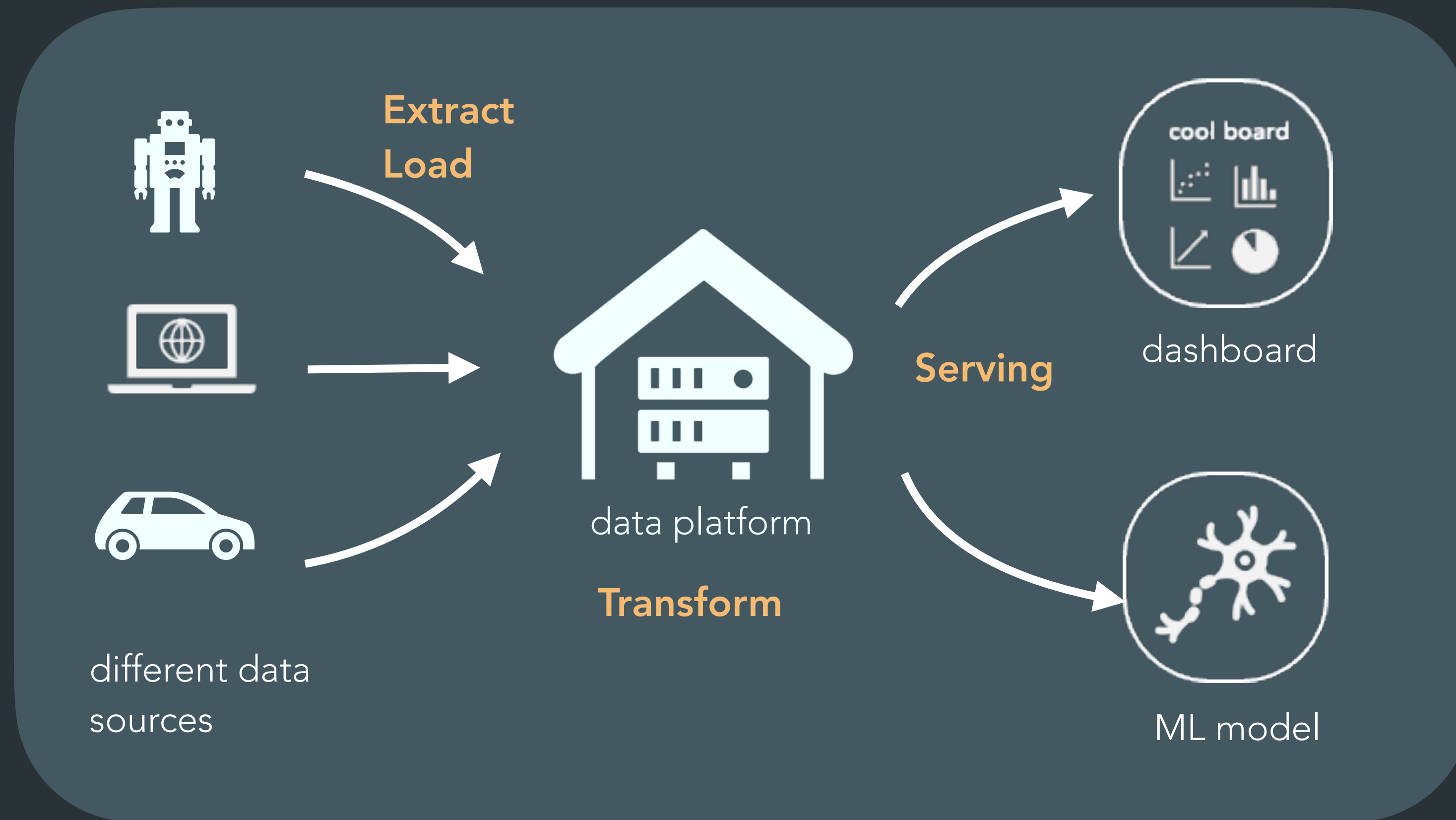
this is considered streaming data

# streaming data vs batch data processing

		STREAMING	BATCH
	DEFINITION	real-time or near real-time processing	data stored and processed in batches at scheduled intervals
	LATENCY	low latency - milliseconds to seconds	higher latency - minutes to days, weeks
	DATA SIZE	smaller chunks but high frequency	large volumes of data accumulated over time
	USE CASE	real-time dashboards, fraud detection, IoT, anomaly detection	sales report, data warehousing, BI dashboards, periodic backups
	COMPLEXITY	higher due to real-time	lower complexity
	COST	higher	lower

which to choose depends on the **business requirements**

source data is streamed and processed in  
the **data platform**



the whole **pipeline**  
could also be  
considered the data  
platform



