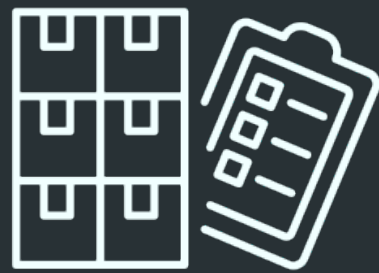kokchun giang

the role of a **data warehouse** in a modern data stack

analyzing ice cream sales may require
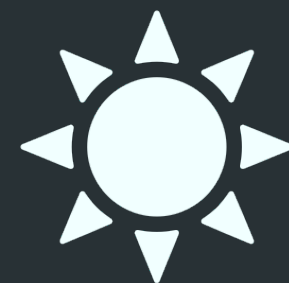# **disparate sources** of data

point-of-sale
(POS) system

inventory
management
system

supply chain
and logistics

weather API

marketing
and ads
platforms

social media
platforms

demographic
data

this is what some of the **data** could look like in the disparate sources

SOCIAL MEDIA

MentionID	Date	Platform	CustomerID	MentionText	Sentiment	Likes
5001	2023-06-24	Twitter	3001	Love The New Vanilla Cup!	Positive	15
5002	2023-06-24	Instagram	3002	Not A Fan Of The Chocolate	Negative	8
5003	2023-06-25	Facebook	3003	Strawberry Bar Is Amazing!	Positive	20

SUPPLY CHAIN

ID	ProductName	Category	StockLevel	ReorderPoint	ReorderQuantity	LastRestocked
2001	Vanilla Cup	Cups	50	20	100	2023-06-20
2002	Chocolate Cone	Cones	30	10	150	2023-06-22
2003	Strawberry Bar	Bars	20	5	50	

WEATHER

Date	Location	Temperature	Humidity	WeatherCondition
2023-06-25	New York	85°F	60 %	Sunny
2023-06-25	California	75°F	55 %	Cloudy
2023-06-25	Texas	90°F	70 %	Sunny

POINT OF SALE

SaleID	Date	Time	ProductID	Quantity	UnitPrice	TotalPrice	PaymentMethod	StoreLocation
1001	2023-06-25	14:35	2001	2	3.50	7.00	Credit Card	Store A
1002	2023-06-25	15:00	2002	1	4.00	4.00	Cash	Store B
1003	2023-06-25	15:30	2003	3	2.50	7.50	Debit Card	Store A

DEMOGRAPHICS

City	Population	MedianAge	AverageIncome	HouseholdSize
Stockholm	975,551	39	377,000 Sek	2.20
Gothenburg	583,056	38	336,000 Sek	2.10
Malmö	347,949	36	300,000 Sek	2.30

the **formats** of the data can vary

structured semi-structured data

csv
xslx
parquet
json

different databases

PostgreSQL
MySQL
SQL server
MongoDB

unstructured text data

docx
pdf

other unstructured data

mp3
video
images

the data can be scattered in **different locations**

on premise
servers

different peoples
computers

file storage systems

cloud vendors

different APIs

data gets **updated** with different frequencies

important that the data
visualizations, reports and
predictions to stakeholders
reflect reality to properly support
decision making

meet the **data warehouse**

data warehouse lifecycle

business requirements → data modeling → ETL/ELT development → testing → deployment → maintenance