Global Terrorism Data Warehouse

By Max Lasky, Ben Orr, Devon Smith

AGENDA

About the Data

01

Background
Type of Data
Data Preparation
System Design

Star Schema

02

Visual
Fact Table
Dimension Tables

OLAP Queries

03

Query Examples

Demo

04

Video
Questions

BACKGROUND

- We are using a subset of the Global Terrorism Database, which is "an open-source database including information on terrorist events around the world." (University of Maryland)
- The "National Consortium for the Study of Terrorism and Responses to Terrorism" (also known as START) at the University of Maryland maintains this database.
- The database is funded by several US Federal Agencies, as well as federal agencies from the United Kingdom and Germany.
- The database has a corresponding guidebook called the 'Global Terrorism Database Guidebook'. It provides extensive knowledge on contextual information and variables the database provides.

Source

LaFree, Gary, et al. "Global Terrorism Database (GTD)." Global Terrorism Database (GTD) | START.Umd.Edu, A Department of Homeland Security Emeritus Center of Excellence led by the University of Maryland, www.start.umd.edu/data-tools/GTD. Accessed 15 Apr. 2025.

TYPE OF DATA

- Subset of Database contains information from January 2021 to July 2021.
- The subset contains 135 different variables, and 4961 rows.
- The database contains most data types, including Ints, Floats, Strings, Booleans, Dates, Times, Arrays, and Strings.
- There are several categorical variables that require reference to the Global Terrorism Guidebook.
- For example, Criteria 1 (shown as variable crit1) from the GTD Guidebook states: "The violent act must be aimed at attaining a political, economic, religious, or social goal".

Source

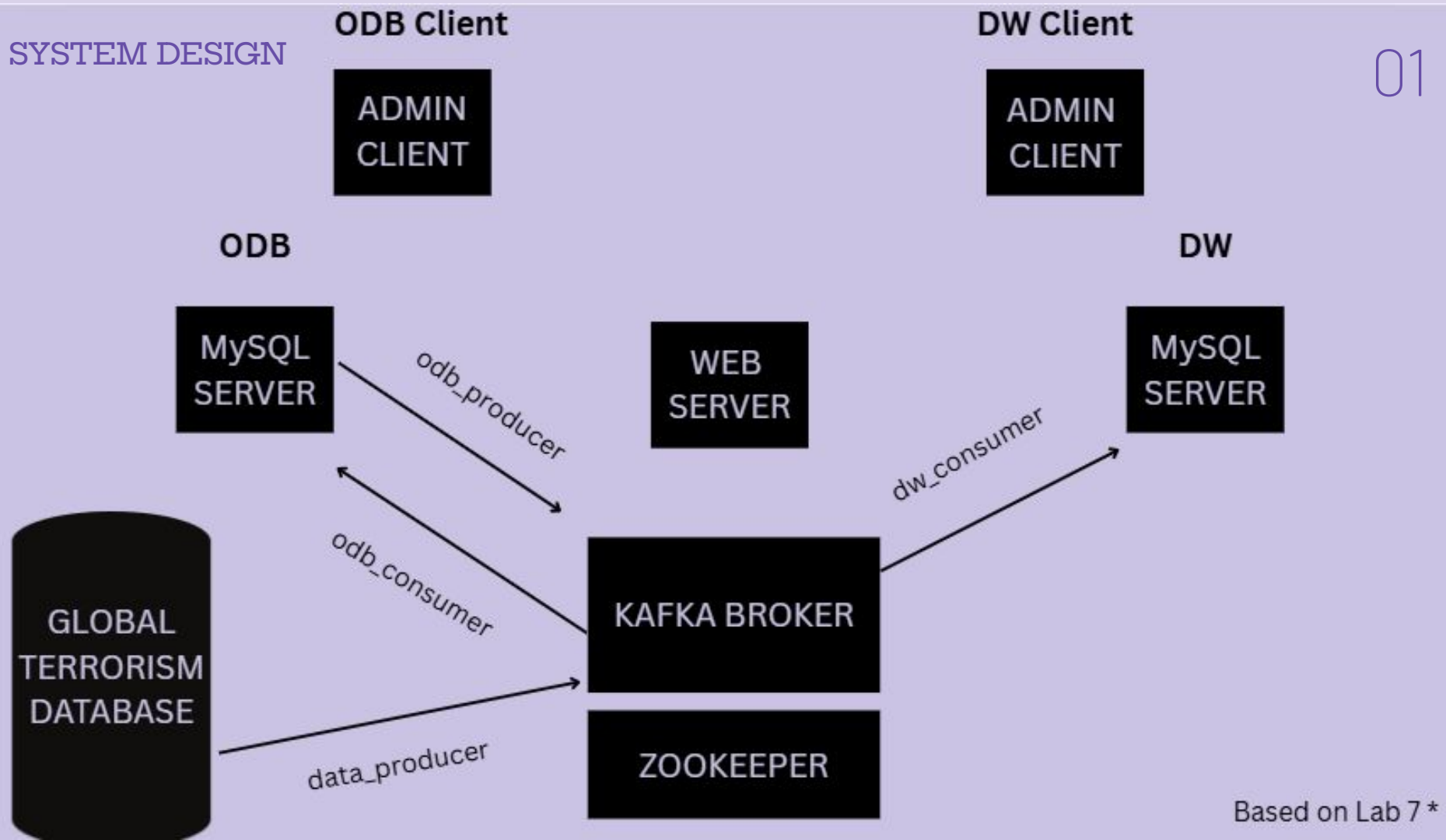
LaFree, Gary, et al. "Global Terrorism Database (GTD)." Global Terrorism Database (GTD) | START.Umd.Edu, A Department of Homeland Security Emeritus Center of Excellence led by the University of Maryland, www.start.umd.edu/data-tools/GTD. Accessed 15 Apr. 2025.

DATA PREPARATION

- Converted missing and “-99” values to NULL

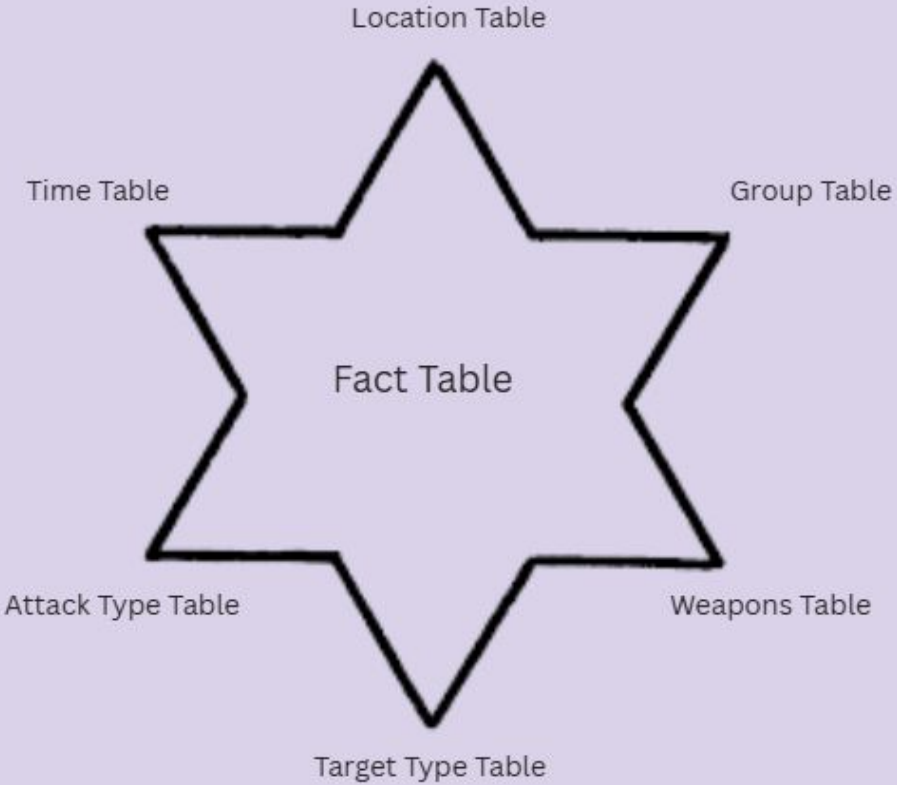
- Normalized field types such as converting iyear, imonth, and iday from String to INT

- Some fields exceeded our standard of 100 characters such as weapdetail



Based on Lab 7 *

STAR SCHEMA


Location Table
Time Table
Group Table
Fact Table
Attack Type Table
Weapons Table
Target Type Table

Time Table

The time table contains all descriptive time information in the database, including year, month, day, and the approximate time of occurence.

timeid	iyear	imonth	iday	approxdate

Location Table

The location table includes all descriptive variables regarding where an attack took place. It goes from the country all the way down to specific coordinates.

locationid	country_txt	region_txt	provstate	city	longitude	latitude

Group Table

The Group Table contains all information about groups that claim responsibility over the attack. It includes the groups name, subgroup involved, and the likelihood of that group claiming the attack being true.

groupid	gname	gsubname	guncertain1

Attack Type Table

The Attack Type table contains all information regarding the purpose of the attack. It includes different attack types, as well as criteria used to distinguish the purpose of an attack.

attackid	attacktype1_txt	attacktype2_txt	attacktype3_txt	crit1	crit2	crit3

Target Type Table

The Target Type table contains all information as to who was attacked, and information regarding the victim. It includes relationships to larger entities such as businesses / governments, and the nationality of the victim.

targetid	targtype1_txt	targsubtype1_txt	corp1	natlty1_txt

Weapon Type Table

The Weapon Type Table contains all information regarding the types of weapon used in an attack. It contains the overall genre of weapon, subgroup of weapon, and further details about the weapon used.

weaponid	weaptype1_txt	weapsubtype1_txt	weapdetail

Fact Table

The fact table contains all the ids for the dimension tables, as well as the total amount of attacks in the database.

factid	timeid	locationid	groupid	attackid	targetid	weaponid	total_attacks

***Research Question 1**: Which countries are most targeted for terrorism attacks?*

***Research Question 1**: Which countries are most targeted for terrorism attacks?*

Explanation

- This query is helpful in understanding which countries are most susceptible to terrorist attacks.
- Information from this query helps influence legislation and policy in specific countries to counter their risk of terrorism.

Completing Query

```
SELECT
    l.country_txt AS country,
    COUNT(*) AS total_attacks
FROM fact f
JOIN location l ON f.locationid
= l.locationid
GROUP BY l.country_txt
ORDER BY total_attacks DESC
LIMIT 10;
```

OLAP QUERIES

Showing rows 0 - 9 (10 total, Query took 0.0052 seconds.)

```
SELECT l.country_txt AS country, COUNT(*) AS total_attacks FROM fact f JOIN location l ON f.locationid = l.locationid GROUP BY l.country_txt ORDER BY total_attacks DESC LIMIT 10
```

Profiling [Edit inline] [Edit] [Explain SQL] [Create PHP cod

+ Options

country	total_attacks ▾ 1
Afghanistan	1453
Iraq	547
Nigeria	448
Yemen	290
India	235
Myanmar	216
Pakistan	150
Syria	144
Democratic Republic of the Congo	142
Somalia	125

***Research Question 2**: Which weapons are most commonly used in terroristic attacks?*

***Research Question 2**: Which weapons are most commonly used in terroristic attacks?*

Explanation

This query hopes to recognize common weapons that are used in terroristic attacks.

- Knowledge about weapon usage could influence policy about how specific weapons are handled legislatively and how policies against them are enforced.

Completing Query

```
SELECT
    w.weaptype1_txt AS weapon_type,
    t.iyear AS year,
    COUNT(*) AS total_attacks
FROM fact f
JOIN weapon w ON f.weaponid =
w.weaponid
JOIN time t ON f.timeid = t.timeid
GROUP BY w.weaptype1_txt, t.iyear
ORDER BY t.iyear, total_attacks DESC;
```

OLAP QUERIES

Showing rows 0 - 8 (9 total, Query took 0.0096 seconds.)

```
SELECT w.weaptype1_txt AS weapon_type, t.iyear AS year, COUNT(*) AS total_attacks FROM fact f JOIN weapon w ON f.weaponid = w.weaponid JOIN time t ON f.timeid = t.timeid GROUP BY
w.weaptype1_txt, t.iyear ORDER BY t.iyear, total_attacks DESC
```

Profiling [Edit inline] [Edit] [Explain SQL] [Create PHP code] [Refresh]

Show all | Number of rows: 25 Filter rows: Search this table

+ Options

weapon_type	year	total_attacks ▾ 2
Explosives	2021	2086
Firearms	2021	1715
Unknown	2021	775
Incendiary	2021	259
Melee	2021	97
Vehicle (not to include vehicle-borne explosives, ...	2021	9
Sabotage Equipment	2021	7
Chemical	2021	4
Other	2021	2

Show all | Number of rows: 25 Filter rows: Search this table

Reflection on the project

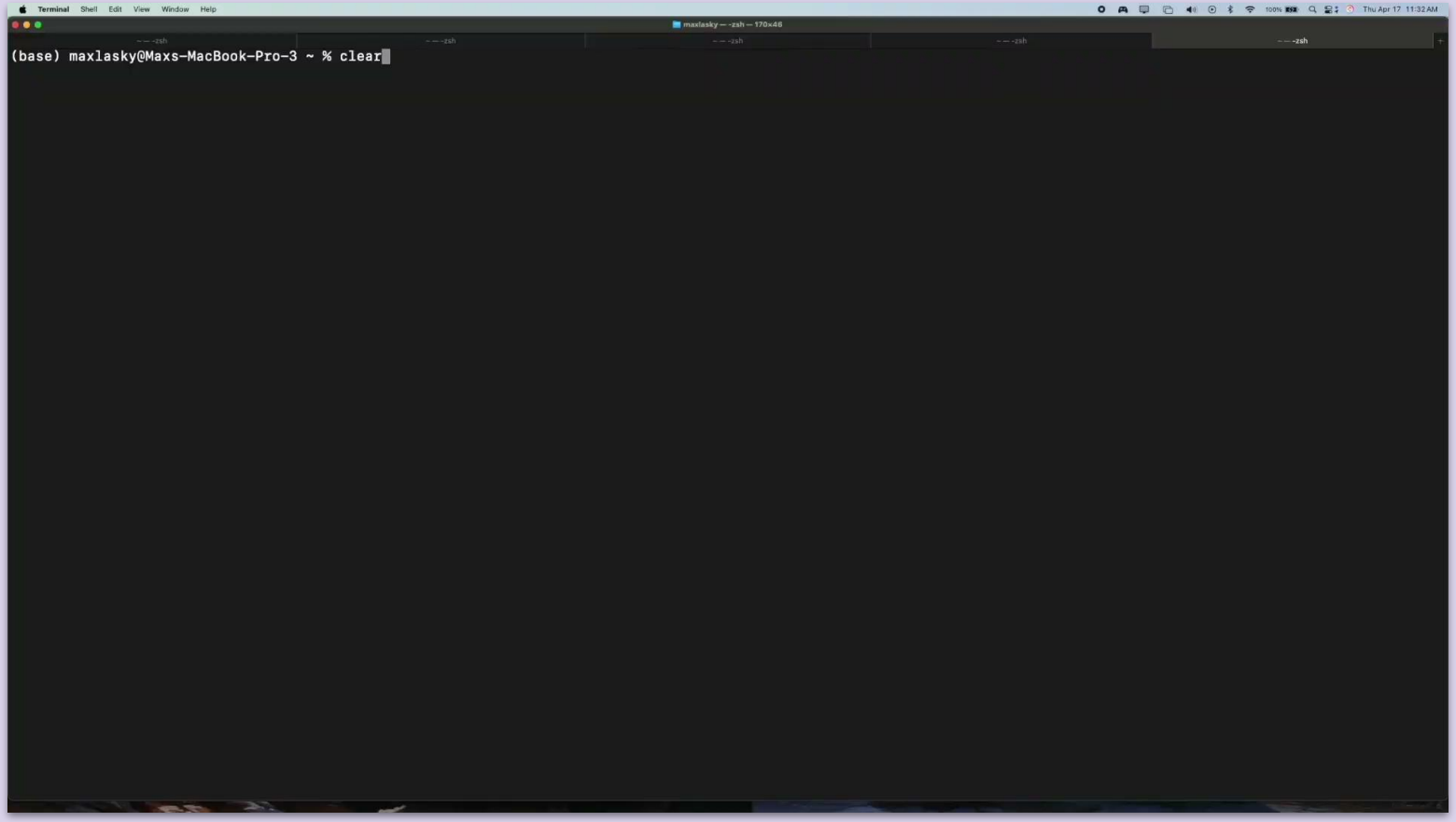
Challenges

Our limited knowledge of Python made it difficult to modify containers to correctly process the data we were using. This resulted in a lot of trial and error throughout the process of getting the data stream to populate the data warehouse.. Another challenge we encountered was missing or incorrectly formatted data, which prevented our SQL queries from correctly processing the data.

Pleasures

Finally getting the data to appear in the DW and ODB was a huge relief, confirming that our process was correct.

DEMO


Terminal Shell Edit View Window Help
Thu Apr 17 11:32 AM
maxlasky — -zsh — 170×46
(base) maxlasky@Maxs-MacBook-Pro-3 ~ % clear

Q & A

THANK YOU