NOVA IMS
Information Management School

Anastasia Nica (m20210516)
Beatriz Fernandes (m20210566)
Frederico Ferreira (m20210387)
Inês Rodrigues (m20210557)
Mª Inês Falcato (m20211079)
GROUP 19

PROFESSORS

Miguel Neto
Bruno Jardim

GLOBAL *global* TERRORISM *terrorism*

BUSINESS INTELLIGENCE I

INDEX

"WITH GUNS YOU CAN KILL TERRORISTS, WITH EDUCATION YOU CAN KILL TERRORISM."

INTRODUCTION

This project was developed with the objective of building a Data Warehouse for the global organization GTD (Global Terrorism Database), consolidating in this way the knowledge acquired in Business Intelligence I.
First, we started by searching the database, which verified all the parameters required by the Teacher. Then, based on the needs we identified, we defined some critical questions that are of high importance regarding the issues of terrorist attacks. These questions are the reason why we need to build a Data Warehouse (DW) in the first place.

Then, we designed the DW based on these needs in order to be able to answer these questions. We used the Kimball Methodology, first in a snowflake schema and then condensing it into a star schema.
Finally, we will move on to the ETL process, where we will do data extraction, transformation and loading. To start, we will extract the data from the source (Extraction), and then we will guarantee data quality and accessibility (Transformation). Lastly, we will load the new transformed data into the DW (Load). These ETL processes will permit us to have a bigger consolidated data view that will lead us to better business decisions.

BUSINESS OVERVIEW

Description of GTD model

The Global Terrorism Database (GTD)™ is considered the most comprehensive database when it comes to classifying terrorist attacks worldwide.
The GTD is an open-source database, which provides information on domestic and international terrorist attacks around the world since 1970, and now includes more than 200 000 events. For each event, a wide range of information is available, including the date and location of the incident, the weapons used, nature of the target, the number of casualties, and – when identifiable – the group or individual responsible.
The fall in deaths was mirrored by a reduction in the impact of terrorism, with 103 countries recording an improvement on their GTI score, compared to 35 that recorded a deterioration. The full GTI score takes into account not only deaths, but also incidents, injuries, and property damage from terrorism, over a five-year period.

Characteristics of the GTD

- Contains information on over 200 000 terrorist attacks;
- Currently the most comprehensive unclassified database on terrorist attacks in the world;
- Includes information on more than 95 000 bombings, 20 000 assassinations, and 15 000 kidnappings and hostage events since 1970;
- Includes information on at least 45 variables for each case, with more recent incidents including information on more than 120 variables;
- More than 4 000 000 news articles and 25 000 news sources were reviewed to collect incident data from 1998 to 2019 alone.

BIGGEST TAKEAWAYS FROM THE GLOBAL TERRORISM

1. DEATHS FROM GLOBAL TERRORISM CONTINUE TO DECLINE

Globally deaths from terrorism fell for the fifth consecutive year in 2019 to 13 826, a 15% decrease from the year prior.

The total number of deaths from terrorism declined for the fifth consecutive year in 2019, falling by 15 per cent to 13 826 deaths. This represents a 59 per cent reduction since the peak in 2014 when 33 438 people were killed in terrorist attacks.

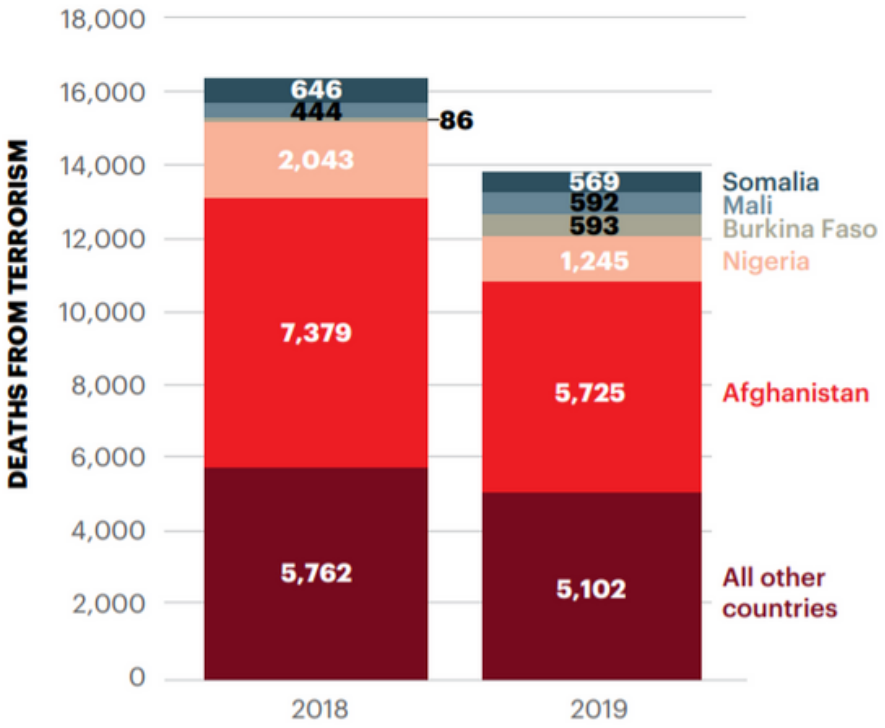
This chart shows the distribution of deaths in the countries with the largest number of terrorism deaths in 2019, compared to 2018.

The year-on-year fall in deaths mirrors a fall in the number of attacks, which dropped from 7 730 to 6 721, a 13 per cent decrease.

Since COVID-19 was declared a global pandemic by the World Health Organization (WHO) in March 2020, preliminary data suggests a decline in both incidents and deaths from terrorism across most regions in the world. However, the COVID-19 pandemic is likely to present new and distinct counter-terrorism challenges.

Total terrorism deaths by country, 2018–2019

Total deaths from terrorism fell 15.5 per cent from 2018 to 2019.



Source: START GTD, IEP calculations

BIGGEST TAKEAWAYS FROM THE GLOBAL TERRORISM

2. DESPITE A FALL IN THE GLOBAL IMPACT OF TERRORISM, IT REMAINS A THREAT IN MANY COUNTRIES

There were 63 countries in 2019 that recorded at least one death from a terrorist attack, and the largest increase in terrorism occurred in Burkina Faso – where deaths rose by 590%. The graph shows the countries that experienced the largest decreases in terrorism deaths in 2019. Afghanistan and Nigeria experienced the two largest falls in 2019.

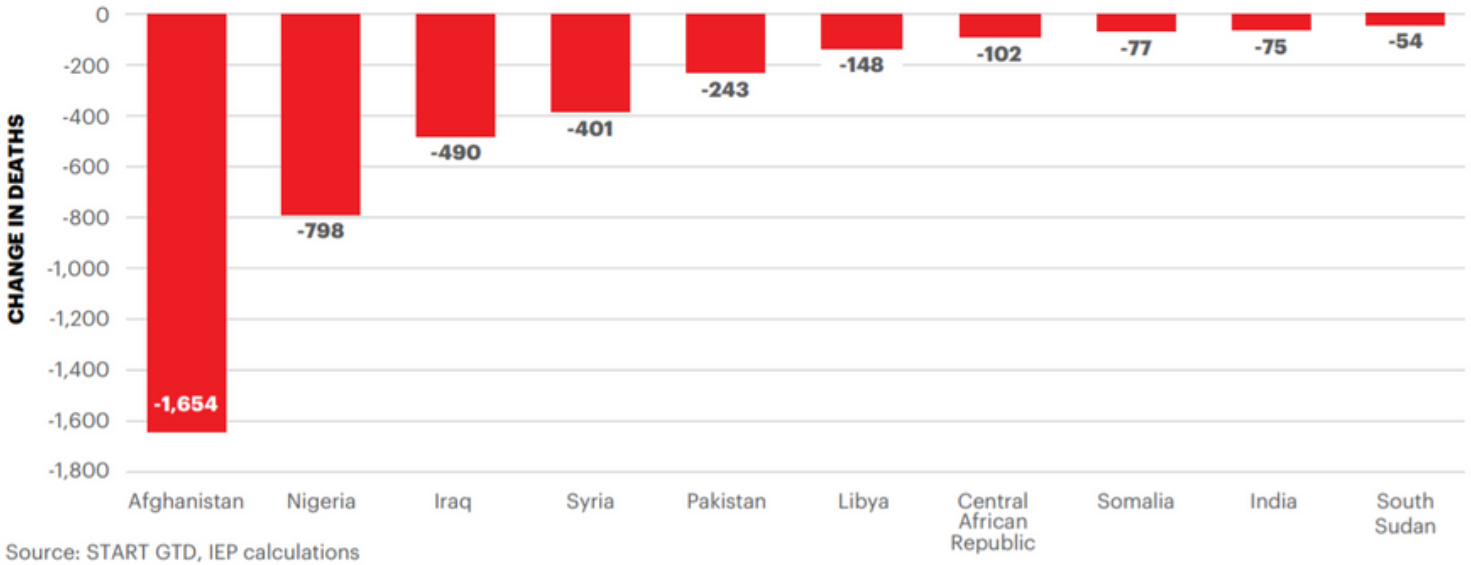
The fall in deaths in Afghanistan is particularly noticeable given its recent history. Since the peak of violence in 2018, deaths have fallen by just over 22 per cent in a year. This reduction was driven by a decline in terrorist deaths attributed to the Taliban.

Nigeria had the second largest fall in total deaths, owing largely to a 72 per cent reduction in fatalities attributed to Fulani extremists. Despite this decrease, the number of deaths attributed to Boko Haram increased by 25 per cent from 2018 to 2019. Renewed activity by Boko Haram in Nigeria and neighbouring countries, including Cameroon, Chad and Niger, remains a substantial threat to the region.

Iraq had the third largest total fall in deaths, with deaths from terrorism falling 46 per cent in a single year. This was the first year since 2003 that Iraq recorded less than a thousand deaths from terrorism. The fall in deaths in Iraq can be attributed to the near total defeat of ISIL in Iraq, which has decreased the level of internal conflict.

Largest decreases in deaths from terrorism, 2018–2019

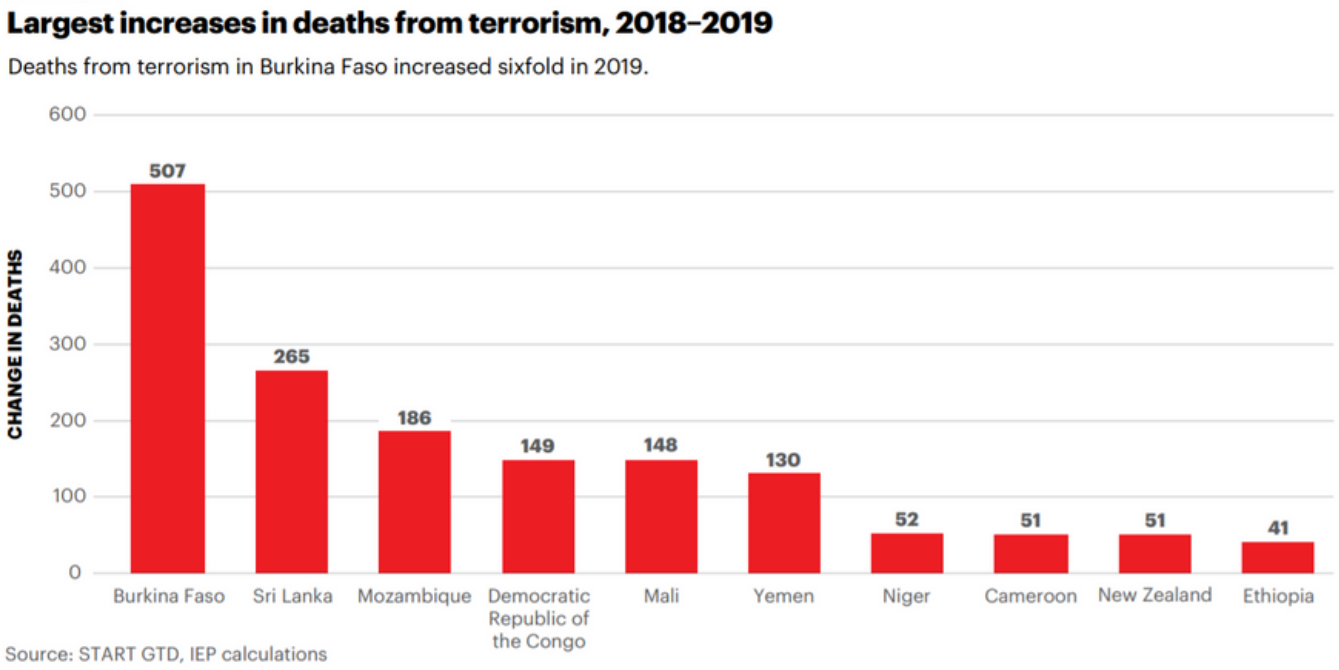
Afghanistan had the largest decrease in the number of deaths from terrorism, reversing a steady increase in terrorism deaths since 2001.



Source: START GTD, IEP calculations

BIGGEST TAKEAWAYS FROM THE GLOBAL TERRORISM

The chart below highlights the countries with the largest increases in deaths from terrorism in 2019. While the increases were offset by much more significant decreases elsewhere, there were a number of countries with worrying increases. Seven of the ten countries with the largest increases in deaths are in sub-Saharan Africa. The country with the largest total increase in deaths from terrorism was Burkina Faso, where the number of people killed rose from 86 in 2018 to 593 in 2019.
Sri Lanka recorded the second largest increase in 2019, with the Easter Sunday bombings accounting for the entirety of this increase. Sri Lanka recorded the deadliest attack of 2019 when eight coordinated suicide attacks across the country targeted churches and hotels on Easter Sunday, killing 266 people and injuring at least 500.

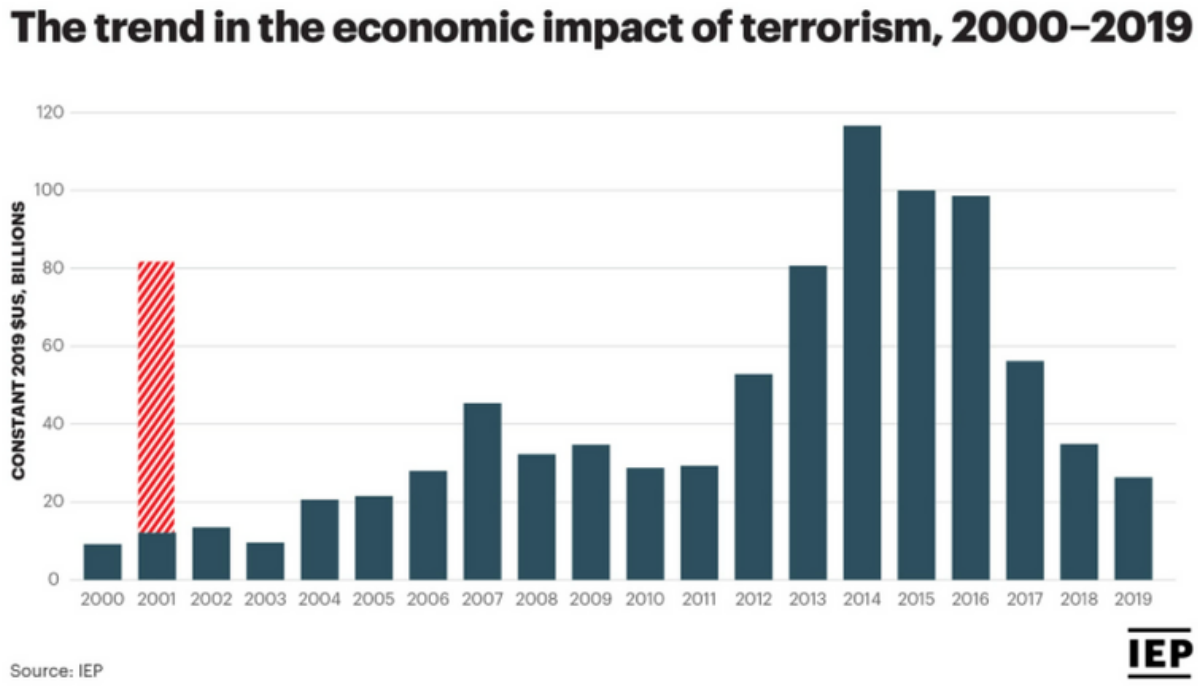


3. TERRORISM ENTAILS A SIGNIFICANT ECONOMIC TOLL

The global economic impact of terrorism was estimated to be US$26.4 billion in 2019; 25 per cent less than the prior year and the fifth consecutive year that it has declined.
This chart shows the trend in the economic impact of terrorism globally from 2000 to 2019. The impact of the September 11, 2001 terrorist attacks is highlighted separately.

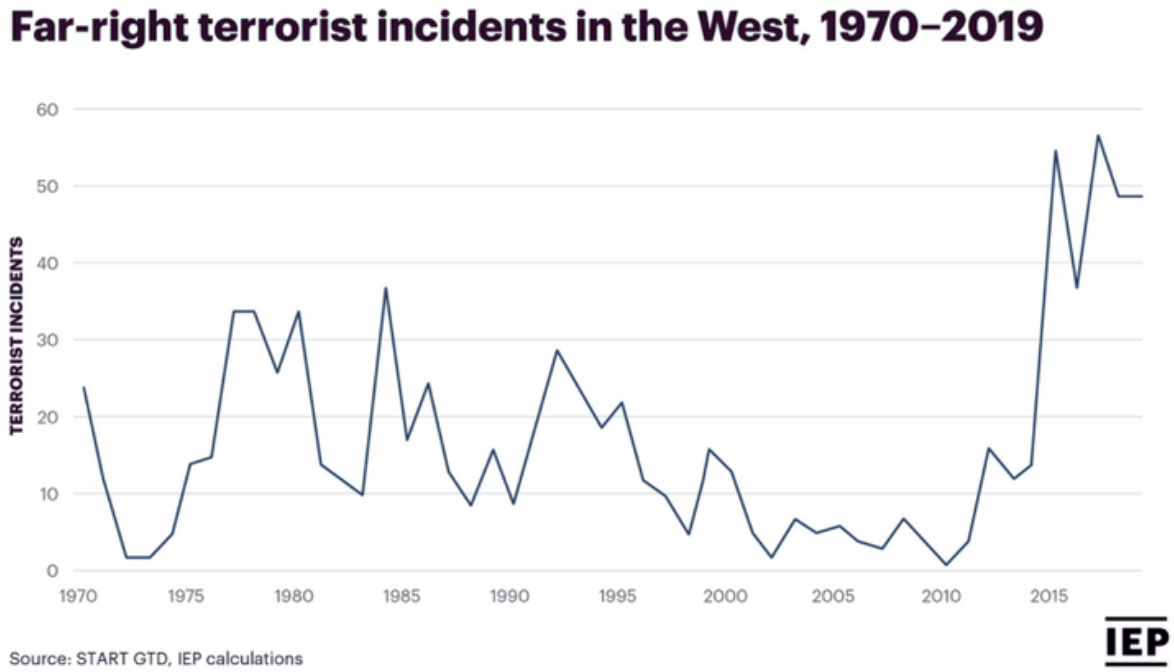
The improvement over the last four years is largely driven by the declining level of terrorism in Iraq, Nigeria, Pakistan and Syria.

BIGGEST TAKEAWAYS FROM THE GLOBAL TERRORISM



4. THERE HAS BEEN A SURGE IN FAR-RIGHT TERRORISM IN THE WEST

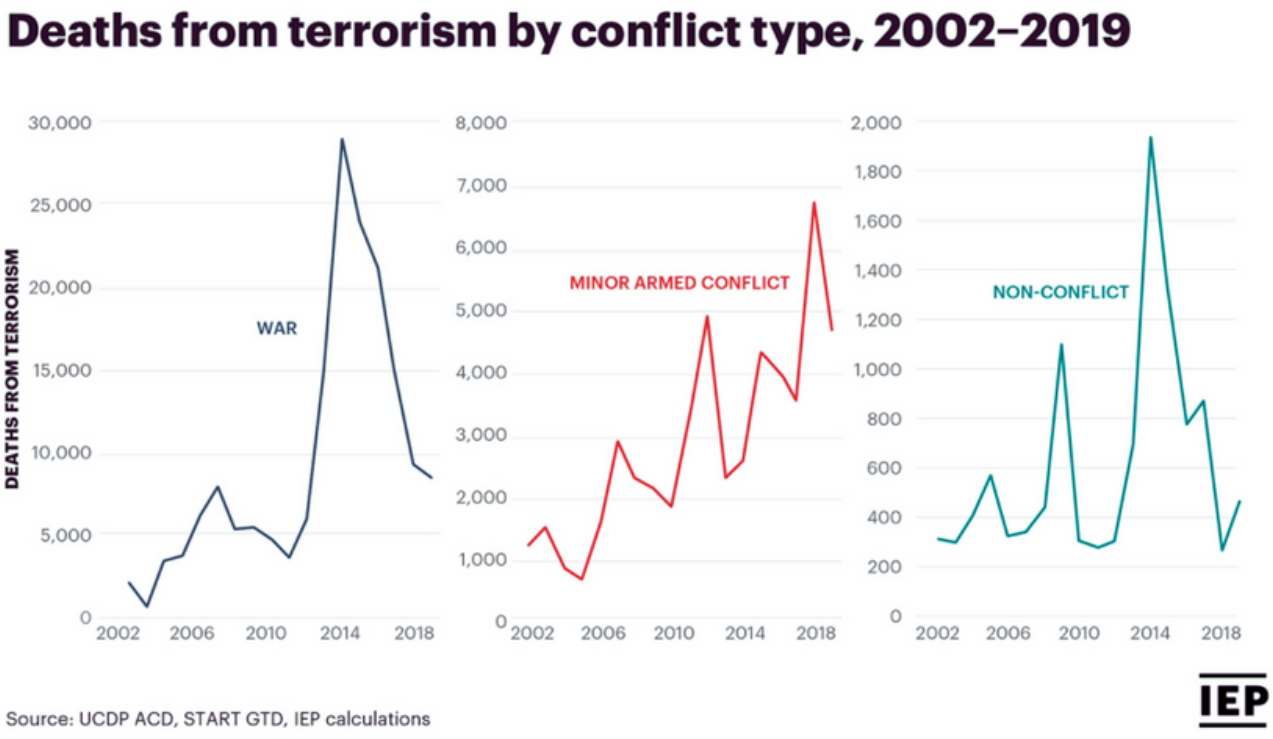
In North America, Western Europe and Oceania, far-right attacks have increased by 250 per cent since 2014 – they are higher now than at any time in the last 50 years.
This chart shows that there was just one recorded far-right terrorist attack in 2010, compared to 49 in 2019.



BIGGEST TAKEAWAYS FROM THE GLOBAL TERRORISM

5. CONFLICT REMAINS THE PRIMARY DRIVER OF TERRORISM

Conflict has been the primary driver of terrorism since 2002.
Every one of the ten countries most impacted by terrorism from 2002 to 2019 was involved in an armed conflict, meaning that they had at least one conflict that led to 25 or more battle-related deaths.
There were 236 422 deaths from terrorism between 2002 and 2019. Of these deaths, just under 95 per cent, or 224 582, occurred in countries involved in conflict.
This chart illustrates the trend in deaths from terrorism by conflict type.



It's crucial understand application of systems thinking to terrorism, using mainly statistical techniques and mathematical models to better understand the dynamics of terrorism and its subsequent impact on society.
Terrorist groups flourish when they can increase their influence. The major drivers of influence are media coverage, recruitment of sympathisers, and finances. All of these different facets need to be tackled together to break up terrorist organizations.

BUSINESS NEEDS

It is important to take security measures to protect people and organizations that could become the target of attacks. This reduces the chances of a terrorist attack.

It is not just the surrounding areas that are suffering from the attacks, but all regions and the country concerned, as well as means of transport, hotels, restaurants, and other sectors.

The impact of terrorism on a country's economy can bring unemployment, lack of coexistence with other cultures, encourage crime and other social problems.

For such catastrophes not to happen, the country has to impose several security measures to prevent everything.

Prevention means early prevention of terrorism, violent extremism and radicalisation leading to terrorism, with the participation of stakeholders from different segments of society. It assumes a developed system of protection which enables timely identification and reduces the potential causes of radicalisation and violent extremism leading to terrorism, while at the same time discouraging an individual or group to resort to an act of terrorism or to otherwise support terrorists. In defining the strategic objectives in this area, account was taken of causes found in the root of radicalisation, as well as indirect causes which accelerate the process of radicalization.

Protection from terrorism implies a built system with defined procedures, activities and measures of responsible stakeholders, whose implementation will reduce the threat from a terrorist attack, i.e. prevent a concrete terrorist attack through detection and removal of weaknesses in the system of protection and limiting the possibilities of an individual or a group to commit a terrorist attack.

Timely taken **preventive measures** are crucial in combating terrorism, taking into account that they affect the instigating forces that incite violent extremism and radicalization leading to terrorism, thereby reducing the number of individuals ready to commit terrorist acts or provide support to terrorists.

BUSINESS NEEDS

Improving the existing and **developing more advanced prevention programs** implies the creation of political, social, economic and other circumstances which are not conducive to the emergence of radicalization, spreading of extremist ideologies and recruitment of terrorists, along with strengthening the value system. It is particularly important to establish intensive and close cooperation and coordination between representatives of government authorities, civil society, the private sector and families, in order to provide support in case that radicalization is detected.

The capacity of the system to protect itself from terrorism implies an improved mechanism for **timely detection** of a threat from terrorism, by way of early identification of terrorist organizations and networks providing support to them, methods of terrorist operations, as well as by uncovering terrorist plots. This reduces the risk of a terrorist attack and increases the capacities for early detection of criminal offences of terrorism and other criminal offences linked to terrorist organizations and terrorist activities, particularly in relation to foreign terrorist fighters and financing of terrorism, including those criminal offences perpetrated through the social media.

Such improved system requires a necessary level of operative capabilities of the police and intelligence-security capacities, adequate protection of critical infrastructure, efficient integrated border management, improved system for combating the financing of terrorism, and the establishment of a system for deradicalization and reintegration of radicalized persons.

The response to a terrorist attack, is to **prepare institutions and systems** that provide vital services to citizens, and citizens themselves, to manage and mitigate the consequences of terrorist attacks, including the provision of assistance to victims of attacks.

BUSINESS NEEDS

Taking everything into consideration, the needs we propose as a group end up being all very correlated.
These needs end up being based on the safety and security of all ordinary citizens. That said, the following needs are worth mentioning:

01

INCREASE THE SECURITY LEVEL

As is well known, terrorism is something that covers the whole world. Through this project we aim to help increase security in the world. As mentioned before, this dataset focuses on global terrorism, which ends up giving us data information regarding terrorist attacks around the world. Through this data we aim to maximise existing security measures in order to create an environment of harmony between citizens and an environment of safety and well being.

Strategic Objective 1.1 – Enhanced level of operational capacity of the police and intelligence-security capacities

Appropriate operational capacities of the police and intelligence-security structures when it comes to human and material resources, including knowledge and skills based on experience and scientific methods, through specialization and other forms of professional training, will contribute to better protection of citizens and property from terrorist attacks. This objective will be achieved through the efforts to improve the normative and institutional framework, with adequate development of human and material resources of police and intelligence security capacities, so that they are capable of efficiently responding to threats of terrorism.

BUSINESS NEEDS

Strategic Objective 1.2 – Raised level of security in the field of transport, trade, exchange of goods and services

A higher level of security in the field of transport, trade, exchange of goods and services will lower the risks of forbidden transfer of means and materials which can be used to commit a terrorist attack, while at the same time improving the overall economic development and efficient and safe communication, both among different parts of the country and among different countries. This objective will be achieved through the efforts to improve the security of transport of passengers and goods.

Strategic Objective 1.3 – Enhanced system of managing the consequences of a terrorist attack

Developed national capacities, standards and procedures in situations caused by terrorist attacks, based on a previously conducted situation analysis, analyses of available capacities and operational preparedness, along with provision of the necessary preconditions for emergency response, rehabilitation of threatened values, minimizing the damage, and cost-efficiency of response, will enable an effective and coordinated response of society to a terrorist attack.

This objective will be achieved through the efforts to establish a system with defined competencies, forces and means, raise it's capacities to the level which enables adequate preparedness and ensure coordination and communication within the system and with stakeholders outside the system, including citizens and the public.

BUSINESS NEEDS

02

PROTECT THE MOST ATTACKED TARGETS

Through this dataset, we have access to data from around the globe, which is useful for analysing the areas and targets most affected by terrorism. With this, it is important to stress the importance of reinforcing prevention and security measures for these attacks, measures that can be improved with the knowledge of, for example, the type of weapons, the type of attack, the most common type of targets, among others, given that, as we have already mentioned, are present in the dataset. All this information will be useful to the competent authorities, who will thus be able to increase their security perimeters and be more alert of the places and targets most affected by terrorist attacks.

Strategic Objective 2.1- Full understanding of terrorism threats in through early identification of target groups and radical methods

Improved understanding of vulnerability to terrorism, including the identification of protection system weaknesses will, by way of strategic planning, coordination, cooperation and exchange of information, enable competent authorities to efficiently and proportionally distribute resources and undertake relevant measures and activities resulting in the elimination of threats and risks of terrorism. This objective will be achieved through the efforts to improve the existing methodologies on the collection, processing and estimate of intelligence, which will contribute to revealing of preparation of a terrorist attack and identification of the perpetrator in an early phase.

BUSINESS NEEDS

Strategic Objective 2.2 – Enhanced level of protection of critical infrastructure

The vulnerability of contemporary societies to terrorist attacks has increased as facilities and systems of special and vital importance for the functioning of a community may become a target of attacks, which can have devastating consequences to the lives and health of people, economy and the environment. This is one of the hallmarks of modern terrorism – creation of fear and disturbance of regular routines for a wide range of the population. Therefore, improving the system for the protection of critical infrastructure is one of the most important objectives of countering terrorism. This objective will be achieved through the efforts to raise awareness of the importance of critical infrastructure for the functioning of the community, develop a normative and institutional framework and enhance human and material capacities of bodies in charge of protection of critical infrastructure, which guarantees efficient readiness and ensures appropriate protection for persons, vital functions, systems and installations.

Strategic Objective 2.3 – Protecting public spaces

Recent terrorist attacks have focused on public spaces, including places of worship and transport hubs, exploiting their open and accessible nature. The rise of terrorism triggered by political or ideologically motivated extremism has made this threat even more acute. This calls for both stronger physical protection of such places and adequate detection systems, without undermining citizens' freedoms. Enhance public-private cooperation for the protection of public spaces, with funding, the exchange of experience and good practices, specific guidance and recommendations. Awareness raising, performance requirements and testing of detection equipment and enhancing background checks to address insider threats will also be part of the approach. An important aspect to reflect is the fact that minorities and vulnerable individuals can be disproportionately affected including persons targeted because of their religion or gender, and therefore require particular attention. Regional and local public authorities have an important role to improve security of public spaces

BUSINESS NEEDS

03

WHAT CAN LEAD TO AN ATTACK

When we talk about reasons why there can be terrorist attacks, these are innumerable: they can be for political, economic, coercion, intimidation, ideological, social, religious or racial reasons.
One never knows for sure the reasons why there are attacks but with this project we intend to set some standards so that these reasons can be better explained, understood and hopefully erased.

Strategic Objective 3.1 – Early identification of causes and factors conducive to spreading of violent extremism and radicalization leading to terrorism

Early identification of causes which motivate individuals and groups to terrorism will enable the undertaking of planned, coordinated and targeted activities of the society aimed at creating a political, social and economic environment not conducive to radicalization, spreading of extremist ideologies and recruitment of terrorists along with strengthening the basic value system that this strategy rests upon.
This objective will be achieved with the efforts – through coordinated engagement of all factors of society involved in prevention – to first adequately identify the key instigators of radicalization, violent extremism and terrorism, and thereafter to act in a planned way with the aim of mitigation and elimination, by undertaking the activities at the economic and social level, by strengthening the value system and recognizing special needs of vulnerable groups.

BUSINESS NEEDS

04 PAY ATTENTION TO CERTAIN TERRORIST GROUPS

When we talk about terrorist groups it should be noted that there are several and spread all over the world, with Al-Qaeda standing out, which we all know not for the best reasons but maybe for the worst ones. The aim of most of these terrorist groups is to draw the public's attention to major problems, such as economic, social and, in most cases, religious crises.

Terrorism is used by these groups as a political and not military strategy, which ends up influencing with the lives of all ordinary citizens like us.

It should be noted the urgency to understand the strategies of the main terrorist groups and the places where they incur their evil acts, so that people and local security can be prevented.

Strategic Objective 4.1 – Target key terrorist and terrorists groups

Using both military and non-military capabilities, we will target the terrorists and terrorist groups who pose the greatest threat to citizens and interests. This will include terrorist leaders, operational planners, and individuals deploying their expertise in areas such as WMD (weapon mass destruction, explosives, cyber operations, and propaganda). We try to understand which are the most influential terrorist groups to try to dismantle these terrorist networks.

Apply persistent pressure and partner intelligence, law enforcement, economic and financial measures, and military action to disrupt, degrade, and prevent the reconstitution of terrorist networks.

ORIGINAL DATA SOURCES

Information about the GTD' dataset

The GTD (Global Terrorism Database), that is the original source, is in a CSV file that has an unique table identified by one key (*eventid*). The variables of this database are divided in nine parts: GTD ID and Date, Incident Information, Incident Location, Attack Information, Weapon Information, Target/Victim Information, Perpetrator Information, Casualties and Consequences and, Additional Information and Sources, as we can see in this document.

We decided to transform the CSV file to an Excel file, that has each table in one sheet, and to change the division of the variables mentioned before to have more specific tables. The names will also change. The changes made were:

- GTD ID Date → Date;
- Incident Information → Incident;
- Incident Location → Location;
- Attack Information → Attack Details;
- Weapon Information → Weapon;
- Target/Victim Information → Target;
- Perpetrator Information → Perpetrator + Claim;
- Casualties and Consequences → Wounded + Deaths + Hostages + Ransom + Property Damage;
- Additional Information and Sources → Sources and Info + International.

For those new tables, we created an ID for each one of them (*incidentid, localid, attackid, weaponid, targetid, perpetratorid, claimid, woundedid, deathsid, hostagesid, ransomid, propertyid, sourcesid, internationalid*), except for the Date table that will have the *eventid* has an ID.

We also created another Excel file that has a table that joins all the tables mentioned before (Attack).

ORIGINAL DATA SOURCES

Information about the GTD' dataset

We felt that it would be useful to the business needs, to have a variable that has information about the brand of the general type of weapon used in the attack (*weaptype1*), therefore we made an Excel file with that variable (*weapbrand*), that was filled randomly with 17 brands (Smith & Wesson, Remington Outdoor, Sturm, Ruger & Co, SIG Sauer, Heckler and Koch, Mossberg, Colt Defense, Beretta, Springfield Armony, Inc, Savage Arms, Barret Firearms, FN Herstal, Taurus International, Browning Arms Company, Winchester Repeating Arms, Glock Ges.m.b.H, Benelli) with exception to the weapons which type is "Biological", "Chemical", "Radiological", "Nuclear", "Fake Weapon", "Incendiary", "Vehicle", "Other" or "Unknown". This variable also adds another level of depth to the hierarchy.

We changed the value of the variables that received 0, 1 and -99/-9 to "No", "Yes" and "Unknown", respectively (*compclaim, property, ishostkid, ransom, int_log, int_ideo, int_misc, int_any, claimed, vicinity, doubtterr*).

We also wanted more variables regarding the date so we also made an Excel file which countains the variables that we wanted. Those variables were built through Excel functions (proper date, number of the weekday, name of the weekday, shortened name of the weekday, type of the weekday, name of the month, shortened name of the month, number of the quarter, name of the quarter, shortened name of the quarter, number of the semester, name of the semester, shortened name of the semester). If the date is not fully known (e.g. 00/07/2015), then we will consider the data to be 01/07/2015, e.g. This table can also help with the business needs and adds more depth levels to the hierarchies.

Both Excel Files are available in the following formats: 2003 and 2016.

We deleted some transactions randomly, since the database is quite big and takes up a lot of space, but the original database can be found in this link.

ORIGINAL DATA SOURCES

Description of each division/table

DATE TABLE

VARIABLE	DESCRIPTION
EVENTID	ID of the date
IYEAR	Number of the year
IMONTH	Number of the month
IDAY	Number of the day
APPROXDATE	Approximate date (used when the exact date is unknown)
EXTENDED	Indicates if it extended for more/less than 24 hours
RESOLUTION	Only used when *extended* = 1 Date of the resolution
PROPERDATE	Date in MM/DD/YY format
WEEKDAYNAME	Name of the day
DAYNAMESHORT	Shortened name of the day
WEEKDAYTYPE	Type of the weekday
MONTHNAME	Name of the month
MONTHNAMESHORT	Shortened name of the month
QUARTERNUMBER	Number of the quarter
QUARTERNAME	Name of the quarter
QUARTERNAMESHORT	Shortened name of the quarter
SEMESTERNUMBER	Number of the semester
SEMESTERNAME	Name of the semester
SEMESTERNAMESHORT	Shortened name of the semester

INCIDENT

VARIABLE	DESCRIPTION
INCIDENTID	ID of the incident
SUMMARY	Brief narrative summary
CRIT1	Indicates if it meets the criterion 1
CRIT2	Indicates if it meets the criterion 2
CRIT3	Indicates if it meets the criterion 3
DOUBTTERR	Indicates is there is doubt that it is an act of terrorism
ALTERNATIVE	Only used when *doubtterr* = 1 Categorization of it other than terrorism
ALTERNATIVE_TXT	
MULTIPLE	Indicates if it is part of a multiple incident
RELATED	Only used when *multiple* = 1 Has the other *eventids*

ORIGINAL DATA SOURCES

Description of each division/table

LOCATION

VARIABLE		DESCRIPTION
NUMERICAL	LOCALID	ID of the location
NUMERICAL	COUNTRY	Country of the incident
TEXT	COUNTRY_TXT	
NUMERICAL	REGION	Region of the incident
TEXT	REGION_TXT	
TEXT	PROVSTATE	Name of the 1st order subnational administrative region
TEXT	CITY	City of the incident
TEXT	VICINITY	Indicates if the incident occurred in the immediate vicinity of the city
TEXT	LOCATION	Additional information about the location
NUMERICAL	LATITUDE	Latitude of the *city*
NUMERICAL	LONGITUDE	Longitude of the *city*
NUMERICAL	SPECIFICITY	Identifies the geospatial resolution of the latitude and longitude fields

ATTACK DETAILS

VARIABLE		DESCRIPTION
NUMERICAL	ATTACKID	ID of the attack
NUMERICAL	ATTACKTYPE1	Captures the general method of attack
TEXT	ATTACKTYPE1_TXT	
NUMERICAL	ATTACKTYPE2	Method of the other attacks Used when the attack is comprised of a sequence of events
TEXT	ATTACKTYPE2_TXT	
NUMERICAL	ATTACKTYPE3	
TEXT	ATTACKTYPE3_TXT	
NUMERICAL	SUCCESS	Indicates if the attack was successful
NUMERICAL	SUICIDE	Indicates if it was a suicide attack

ORIGINAL DATA SOURCES

Description of each division/table

WEAPON

VARIABLE	DESCRIPTION
NUMERICAL WEAPONID	ID of the weapon
NUMERICAL WEAPONTYPE1	General type of weapon used
TEXT WEAPONTYPE1_TXT	
NUMERICAL WEAPSUBTYPE1	More specific type for the general weapon
TEXT WEAPSUBTYPE1_TXT	
NUMERICAL WEAPONTYPE2	Type of the 2nd most used weapon
TEXT WEAPONTYPE2_TXT	
NUMERICAL WEAPSUBTYPE2	More specific type for the 2nd most used weapon
TEXT WEAPSUBTYPE2_TXT	
NUMERICAL WEAPONTYPE3	Type of the 3rd most used weapon
TEXT WEAPONTYPE3_TXT	
NUMERICAL WEAPSUBTYPE3	More specific type for the 3rd most used weapon
TEXT WEAPSUBTYPE3_TXT	
NUMERICAL WEAPONTYPE4	Type of the 4th most used weapon
TEXT WEAPONTYPE4_TXT	
NUMERICAL WEAPSUBTYPE4	More specific type for the 4th most used weapon
TEXT WEAPSUBTYPE4_TXT	
TEXT WEAPDETAIL	Pertinent informations on the type of weapon(s) used
TEXT WEAPBRAND	Brand of the general weapon used

TARGET

VARIABLE	DESCRIPTION
NUMERICAL TARGETID	ID of the target
NUMERICAL TARGTYPE1	General type of target/victim
TEXT TARGTYPE1_TXT	
NUMERICAL TARGSUBTYPE1	More specific category for the main target, next level of designation
TEXT TARGSUBTYPE1_TXT	
TEXT CORP1	Name of the main corporate entity or government agency that was targeted
TEXT TARGET1	Main specific person, building, installation, etc., that was targeted
NUMERICAL NATLTY1	Nationality of the main target that was attacked
TEXT NATLTY1_TXT	
NUMERICAL TARGTYPE2	Type of the 2nd target/victim
TEXT TARGTYPE2_TXT	
NUMERICAL TARGSUBTYPE2	More specific category for the 2nd target, next level of designation
TEXT TARGSUBTYPE2_TXT	
TEXT CORP2	Name of the 2nd corporate entity or government agency that was targeted
TEXT TARGET2	2nd specific person, building, installation, etc., that was targeted
NUMERICAL NATLTY2	Nationality of the 2nd target that was attacked
TEXT NATLTY2_TXT	
NUMERICAL TARGTYPE3	Type of the 3rd target/victim
TEXT TARGTYPE3_TXT	
NUMERICAL TARGSUBTYPE3	More specific category for the 3rd target, next level of designation
TEXT TARGSUBTYPE3_TXT	
TEXT CORP3	Name of the 3rd corporate entity or government agency that was targeted
TEXT TARGET3	3rd specific person, building, installation, etc., that was targeted
NUMERICAL NATLTY3	Nationality of the 3rd target that was attacked
TEXT NATLTY3_TXT	

ORIGINAL DATA SOURCES

Description of each division/table

PERPETRATOR

	VARIABLE	DESCRIPTION
NUMERICAL	PERPETRATORPID	ID of the perpetrator
TEXT	GNAME	Name of the group that carried out the attack
TEXT	GSUBNAME	Additional qualifiers or details about the name of the main group
TEXT	GNAME2	Name of the 2nd group
TEXT	GSUBNAME2	Additional qualifiers or details about the 2nd group
TEXT	GNAME3	Name of the 3rd group
TEXT	GSUBNAME3	Additional qualifiers or details about the 3rd group
NUMERICAL	GUNCERTAIN1	Indicates if the main group attribution(s) for the attack are suspected
NUMERICAL	GUNCERTAIN2	Indicates if the 2nd group attribution(s) for the attack are suspected
NUMERICAL	GUNCERTAIN3	Indicates if the 3rd group attribution(s) for the attack are suspected
NUMERICAL	INDIVIDUAL	Indicates if the attack was carried out by an individual(s) not known to be affiliated with a group
NUMERICAL	NPERPS	N° of terrorists participating
NUMERICAL	NPERPCAP	N° of perpetrators taken into custody
TEXT	MOTIVE	Specific motive for the attack

DEATHS

	VARIABLE	DESCRIPTION
NUMERICAL	DEATHID	ID of the deaths
NUMERICAL	NKILL	N° of total fatalities for the incident
NUMERICAL	NKILLUS	N° of U.S. citizens who died as a result of the incident
NUMERICAL	NKILLTER	N° of perpetrator fatalities

WOUNDED

	VARIABLE	DESCRIPTION
NUMERICAL	WOUNDEDID	ID of wounded
NUMERICAL	NWOUND	N° of non-fatal injuries to both perpetrators and victims
NUMERICAL	NWOUNDUS	N° of non-fatal injuries to U.S. citizens, both perpetrators and victims
NUMERICAL	NWOUNDTE	N° of non-fatal injuries of the perpetrators

CLAIM

	VARIABLE	DESCRIPTION
NUMERICAL	CLAIMID	ID of the claim
TEXT	CLAIMED	Indicates if the main group claimed responsibility for the attack
NUMERICAL	CLAIMMODE	Records the mode used by the main group to claim responsibility
TEXT	CLAIMMODE_TXT	
TEXT	COMPCLAIM	Indicate if more than one group claimed separate responsibility
NUMERICAL	CLAIM2	Indicates if the 2nd group claimed responsibility for the attack
NUMERICAL	CLAIMMODE2	Records the mode used by the 2nd group to claim responsibility
TEXT	CLAIMMODE2_TXT	
NUMERICAL	CLAIM3	Indicates if the 3rd group claimed responsibility for the attack
NUMERICAL	CLAIMMODE3	Records the mode used by the 3rd group to claim responsibility
TEXT	CLAIMMODE3_TXT	

PROPERTY DAMAGE

	VARIABLE	DESCRIPTION
NUMERICAL	PROPERTYID	ID of the property
TEXT	PROPERTY	Indicates if the incident resulted in property damage
NUMERICAL	PROPEXTENT	Only used when *property* = 1 Describes the extent of the property damage
TEXT	PROPEXTENT_TXT	
NUMERICAL	PROPVALUE	Only used when *property* = 1 Exact U.S. dollar amount of total damages
TEXT	PROPCOMMENT	Only used when *property* = 1 Non-monetary or imprecise measures of damage

ORIGINAL DATA SOURCES

Description of each division/table

HOSTAGES

VARIABLE	DESCRIPTION
HOSTAGESID (NUMERICAL)	ID of the hostages
ISHOSTKID (TEXT)	Indicates if victims were taken hostage or kidnapped
NHOSTKID (NUMERICAL)	Nº of hostages or kidnapping victims
NHOSTKIDUS (NUMERICAL)	Nº of U.S. citizens that were taken hostage or kidnapped
NHOURS (NUMERICAL)	If the kidnapping/hostage incident lasted for less than 24h, this field records the approximate nº of hours
NDAYS (NUMERICAL)	If the kidnapping/hostage incident lasted more than 24h, this field records the approximate nº of days
DIVERT (TEXT)	Country that hijackers diverted a vehicle to, or the country that the kidnap victims were moved to and held
KIDHIJCOUNTRY (TEXT)	Country in which the kidnapping/hostage incident was resolved or ended
HOSTKIDOUTCOME (NUMERICAL)	Indicates the eventual fate of hostages and kidnap victims
HOSTKIDOUTCOME_TXT (TEXT)	
NRELEASED (NUMERICAL)	Nº of hostages who survived the incident

RANSOM

VARIABLE	DESCRIPTION
RANSOMID (NUMERICAL)	ID of the ransom
RANSOM (TEXT)	Indicates if the incident involved a demand of monetary ransom
RANSOMAMT (NUMERICAL)	Only used when *ransom* = 1 Amount (in U.S. dollars)
RANSOMAMTUS (NUMERICAL)	Amount (in U.S. dollars) if a ransom was demanded from U.S. sources
RANSOMPAID (NUMERICAL)	Amount (in U.S. dollars) if a ransom amount was paid
RANSOMPAIDUS (NUMERICAL)	Amount (in U.S dollars) if a ransom amount was paid by U.S. sources
RANSOMNOTE (TEXT)	Specific details relating to a ransom

INTERNATIONAL

VARIABLE	DESCRIPTION
INTERNATIONALID (NUMERICAL)	ID of the international
INT_LOG (NUMERICAL)	Indicates if a group crossed a border to carry out an attack
INT_IDEO (NUMERICAL)	Indicates if a group attacked a target of a different nationality
INT_MISC (NUMERICAL)	Indicates if the location of the attack differs from the nationality of the target
INT_ANY (NUMERICAL)	Indicates if the attack was international on any of the dimensions described above

ATTACK

VARIABLE	DESCRIPTION
EVENTID (NUMERICAL)	ID of the date
LOCALID (NUMERICAL)	ID of the location
INCIDENTID (NUMERICAL)	ID of the incident
ATTACKID (NUMERICAL)	ID of the attack
TARGETID (NUMERICAL)	ID of the target
PERPETRATORID (NUMERICAL)	ID of the perpetrator
WEAPONID (NUMERICAL)	ID of the weapon
CLAIMID (NUMERICAL)	ID of the claim
WOUNDEDID (NUMERICAL)	ID of the wounded
DEATHSID (NUMERICAL)	ID of the deaths
HOSTAGESID (NUMERICAL)	ID of the hostages
RANSOMID (NUMERICAL)	ID of the ransom
PROPERTYID (NUMERICAL)	ID of the property
SOURCESID (NUMERICAL)	ID of the source and info
INTERNATIONALID (NUMERICAL)	ID of the international

ORIGINAL DATA SOURCES

Description of each division/table

SOURCES AND INFO

VARIABLE		DESCRIPTION
NUMERICAL	SOURCESID	ID of sources and info
NUMERICAL	ADDINFOID	ID of the additional information and sources
TEXT	ADDNOTES	Additional relevant details about the attack
TEXT	SCITE1	1st source that was used to compile information on the specific incident
TEXT	SCITE2	2nd source that was used to compile information on the specific incident
TEXT	SCITE3	3rd source that was used to compile information on the specific incident
TEXT	DBSOURCE	Original data collection effort in which each event was recorded

A more detailed explanation about the variables can be found in this document.

The attack types, the weapon types and subtypes, the target types and subtypes and the claim modes, are also available in the previous document.

The variables *claimmode2_txt* and *claimmode3_txt* are not in the document but are in the database.
Claimed, *doubtterr* and *vicinity* can also have -9 has a value, looking at the database.
If a variable has -99/-9 then the value is unknown.

APPENDIX

- Criteria 1 - Political, economic, religious, or social goal

The violent act must be aimed at attaining a political, economic, religious, or social goal. This criterion is not satisfied in those cases where the perpetrator(s) acted out of a pure profit motive or from an idiosyncratic personal motive unconnected with broader societal change.

- Criteria 2 - Intention to coerce, intimidate or publicize to larger audience(s)

To satisfy this criterion there must be evidence of an intention to coerce, intimidate, or convey some other message to a larger audience (or audiences) than the immediate victims. Such evidence can include (but is not limited to) the following: pre or post attack statements by the perpetrator(s), past behavior by the perpetrators, or the particular nature of the target/victim, weapon, or attack type.

- Criteria 3 - Outside international humanitarian law

The action is outside the context of legitimate warfare activities, insofar as it targets non-combatants.

DATA WAREHOUSE

The Process Used in the Data Warehouse Design

In order to achieve a well-designed warehouse, we will follow the Kimball methodology which consists in the steps presented next.

We strongly believe that this is the most suitable methodology to add value to the business and matches our bussiness requirements and needs.

Considering this, our methodology will respect the following steps:

01

SELECT THE BUSINESS PROCESS

Identify the business process. This will be the source of the metrics and measurements.

02

DECLARE THE GRAIN

Determine the meaning behing each fact table row.

03

IDENTIFY THE DIMENSIONS

Determine the tables that support the construction and constitution of fact tables.

04

IDENTIFY THE FACTS

The measures, facts or metrics are identified. They should confirm the grain defined in the second step.

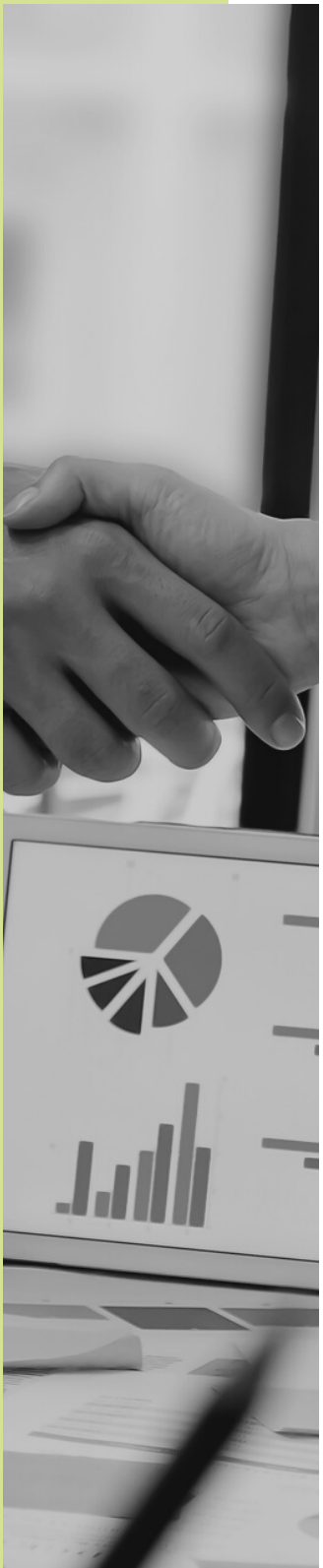
05

CREATE SCHEMA

In the final step, we connect the fact tables and the dimension tables that are directly connected in the Relational.

DATA WAREHOUSE

1. *Select the business process*



All that are business related processes, more specifically, the operational activities, are always performed by a given organization.

The above is advocated by the Kimball group.
Having said that, we emphasize that they describe not only transactions but also generate metrics for GTD. All this means that these processes allow us to do several things:

1. Define the grain;
2. Define dimensions;
3. Define the facts.

So we started by defining GTD's business processes, based on the information in our database, figures provided to us by the National Consortium for the Study of Terrorism and Responses to Terrorism (START) at the University of Maryland.

It is an objective of our project to analyse the response to the terrorist attack, so that is our business process. Identifying our business processes is important for building the Data Warehouse.

For greater understanding, we supplement each previously defined business need with the questions that will help us clarify the needs of our citizens.

DATA WAREHOUSE

1. *Select the business process*

1.INCREASE THE SECURITY LEVEL

The following **questions** were defined:

- What are the top 3 weapon type/subtype, by attack type?
- What is the most used brand of weapon by attack type?
- What is the most common type of attack by country/region/state/city, per semester/quarter/month/week/day?
- What is the number of dead/injured by type of attack?
- What is the percentage of successful attacks by country?
- Which are the countries that have the most attacks that were successful and the perpetrators had to cross a border to carry out the attack?
- Which are the countries that had the most number of successful attacks in which the targets were not from this country?

2.PROTECT THE MOST ATTACKED TARGETS

The following **questions** were defined:

- In which country and in which month did the incidents that covered a larger audience, by criteria 2, took place?
- Which city and year has the most attacks, by criteria 1/2/3?
- What is the most common type/subtype of target, by 1/2/3 criteria?
- Which corporate entity/target is the most attacked, by criteria 1/2/3?
- Which nationalities are the most attacked, by criteria 1/2/3?
- Which are the nationalities of the targets that were victims in the most attacks that were successful and the perpetrators were from a different nationality?

DATA WAREHOUSE

1. *Select the business process*

3.WHAT CAN LEAD TO AN ATTACK

The following **questions** were defined:

- Which target is the most attacked, by criteria 1/2/3?
- Which nationality is the most attacked, by criteria 1/2/3?
- What is the most common motives, per attack type?
- What is the total number of ransoms, per type of attack?

4.PAY ATTENTION TO CERTAIN TERRORIST GROUPS

The following **questions** were defined:

- What is the year by year percentage of perpetrators captured, per group?
- What is the most common claim mode for each attack type, by group?
- How many attacks were suicide attacks, by group?
- What is the total number of successful attacks, per group?
- What is the most common mode of claim that passes the 1/2/3 criteria, per group?
- What is the total number of wounded/dead people by attack, per group?
- What is the ransom paid by type of attack, per group?
- What is the total number of hostages by attack, per group?
- What is the principal country in which the kidnapping/hostage incident has started and resolved/ended by attack, per group?
- What is the total amount of property damage by attack type, per group?
- Which is the percentage of hostages free, per group?

DATA WAREHOUSE

2. Declare The Grain

For step of the process defined above, we shall consider:

- Different measures required for our fact table - **black color**
- Identify the context that supplied data for each dimension - **green color**

1.INCREASE THE SECURITY LEVEL

- What are the top 3 **weapon type/subtype**, by **attack type**?
- What is the most used **brand of weapon** by **attack type**?
- What is the most common **type of attack** by **country/region/state/city**, per **semester/quarter/month/day/week**?
- What is the number of **dead/injured** by **type of attack**?
- What is the percentage of **successful** attacks by **country**?
- Which are the **countries** that have the most attacks that were **successful** and the **perpetrators had to cross a border to carry out the attack**?
- Which are the **countries** that had the most number of **successful** attacks in which the **targets were not from this country**?

2.PROTECT THE MOST ATTACKED TARGETS

- In which **country** and in which **month** did the incidents that covered a **larger audience, by criteria 2**, took place?
- Which **city** and **year** has the most attacks, by **criteria 1/2/3**?
- What is the most common **type/subtype of target**, by **1/2/3 criteria**?
- Which **corporate entity/target** is the most attacked, by **criteria 1/2/3**?
- Which **nationalities** are the most attacked, by **criteria 1/2/3**?
- Which are the **nationalities** of the targets that were **victims** in the most attacks that were **successful** and **the perpetrators were from a different nationality**?

DATA WAREHOUSE

2. Declare The Grain

For step of the process defined above, we shall consider:

- Different measures required for our fact table - **black color**
- Identify the context that supplied data for each dimension - **green color**

3.WHAT CAN LEAD TO AN ATTACK

- Which **target** is the most attacked, by **criteria 1/2/3?**
- Which **nationality** is the most attacked, **by criteria 1/2/3?**
- What is the most common **motives**, per **attack type?**
- What is the total **number of ransoms**, per **type of attack?**

4.PAY ATTENTION TO CERTAIN TERRORISTS GROUPS

- What is the **year by year percentage of perpetrators captured**, per **group**?
- What is the most common **claim mode** for each **attack type,** by **group**?
- How many attacks were **suicide attacks**, by **group**?
- What is the total **number of successful attacks**, per **group**?
- What is the most common **mode of claim** that passes the **1/2/3 criteria**, per **group**?
- What is the total number of **wounded/dead** people by **attack**, per **group**?
- What is the **ransom** paid by **type of attack**, per **group**?
- What is the total **number of hostages** by **attack**, per **group**?
- What is the principal **country** in which the **kidnapping/hostage** incident has started and resolved/ended by **attack**, per **group**?
- What is the total **amount of property damage** by **attack type**, per **group**?
- Which is the **percentage of hostages free**, per **group**?

DATA WAREHOUSE

3. Identify the dimensions

1.INCREASE THE SECURITY LEVEL

What are the top 3 **weapon type/subtype**, by attack type?	**Weapon**
What is the most used **brand of weapon** by attack type?	**Weapon**
What is the most common type of attack by **country/region/state/ city**, per **semester/quarter/month/day/week**?	**Location** **Date**
What is the number of **dead**/**injured** by type of attack?	**Deaths** **Wounded**
What is the percentage of successful attacks by **country**?	**Location**
Which are the **countries** that have the most attacks that were successful and the **perpetrators had to cross a border to carry out the attack**?	**Location** **International**
Which are the **countries** that had the most number of successful attacks in which the **targets were not from this country**?	**Location** **International**

2.PROTECT THE MOST ATTACKED TARGETS

In which **country** and in which **month** did the incidents that covered a larger audience, by criteria 2, took place?	**Location** **Date**
Which **city** and **year** has the most attacks, by criteria 1/2/3?	**Location** **Date**
What is the most common **type/subtype of target**, by 1/2/3 criteria?	**Target**
Which **corporate entity/target** is the most attacked, by criteria 1/2/3?	**Target**
Which **nationalities** are the most attacked, by criteria 1/2/3?	**Target**
Which are the **nationalities** of the targets that were victims in the most attacks that were successful and the **perpetrators were from a different nationality**?	**Target** **International**

DATA WAREHOUSE

3. Identify the dimensions

3. WHAT CAN LEAD TO AN ATTACK

Which **target** is the most attacked, by criteria 1/2/3?	**Target**
Which **nationality** is the most attacked, by criteria 1/2/3?	**Target**
What is the most common **motives**, per attack type?	**Perpetrator**
What is the total number of **ransoms**, per type of attack?	**Ransom**

4. PAY ATTENTION TO CERTAIN TERRORISTS GROUPS

What is the **year by year** percentage of perpetrators captured, per **group**?	**Date** Perpetrator
What is the most common **claim mode** for each attack type, by **group**?	**Claim** Perpetrator
How many attacks were suicide attacks, by **group**?	**Perpetrator**
What is the total number of successful attacks, per **group**?	**Perpetrator**
What is the most common **mode of claim** that passes the 1/2/3 criteria, per **group**?	**Claim** Perpetrator
What is the total number of **wounded**/**dead** people by attack, per **group**?	**Wounded** Deaths Perpetrator
What is the **ransom** paid by type of attack, per **group**?	**Ransom** Perpetrator
What is the total **number of hostages** by attack, per **group**?	**Hostages** Perpetrator
What is the principal **country** in which the **kidnapping/hostage** incident has started and resolved/ended by attack, per **group**?	**Location** Hostages Perpetrator
What is the total **amount of property damage** by attack type, per **group**?	**Property Damage** Perpetrator
Which is the percentage of hostages free, per **group**?	**Perpetrator**

DATA WAREHOUSE

3. Identify the dimensions

DIMENSIONS	PARENT DIMENSION
Date	None
Location	None
Perpetrator	None
Claim	Perpetrator
Weapon	None
Deaths	None
Target	None
Property Damage	None
International	None
Wounded	None
Ransom	Hostages
Hostages	None

DATA WAREHOUSE

4. Identify the facts

1.INCREASE THE SECURITY LEVEL

Attack Type	What are the top 3 weapon type/subtype, by **attack type**?
Attack Type	What is the most used brand of weapon by **attack type**?
Attack Type	What is the most common **type of attack** by country/region/state/city, per semester/quarter/month/day/week?
Attack Type	What is the number of dead/injured by **type of attack**?
Successful Attack	What is the percentage of **successful** attacks by country?
Successful Attack	Which are the countries that have the most attacks that were **successful** and the perpetrators had to cross a border to carry out the attack?
Successful Attack	Which are the countries that had the most number of **successful** attacks in which the targets were not from this country?

2.PROTECT THE MOST ATTACKED TARGETS

Criteria 2	In which country and in which month did the incidents that **covered a larger audience, by criteria 2**, took place?
Criteria 1/2/3	Which city and year has the most attacks, by **criteria 1/2/3**?
Criteria 1/2/3	What is the most common type/subtype of target, by **criteria 1/2/3**?
Criteria 1/2/3	Which corporate entity/target is the most attacked, by **criteria 1/2/3**?
Criteria 1/2/3	Which nationalities are the most attacked, by **criteria 1/2/3**?
Successful Attack	Which are the nationalities of the targets that were victims in the most attacks that were **successful** and the perpetrators were from a different nationality?

DATA WAREHOUSE

4. Identify the facts

3.WHAT CAN LEAD TO AN ATTACK

Criteria 1/2/3	Which target is the most attacked, by **criteria 1/2/3**?
Criteria 1/2/3	Which nationality is the most attacked, by **criteria 1/2/3**?
Attack Type	What is the most common motives, per **attack type**?
Attack Type	What is the total number of ransoms, per **type of attack**?

4.PAY ATTENTION TO CERTAIN TERRORISTS GROUPS

Percentage of perpetrators captured	
Attack Type	What is the most common claim mode for each **attack type**, by group?
Suicide Attack	How many attacks were **suicide attacks**, by group?
Successful Attack	What is the total number of **successful attacks**, per group?
Criteria 1/2/3	What is the most common mode of claim that passes the **criteria 1/2/3**, per group?
Attack Type	What is the total number of wounded/dead people by **attack**, per group?
Attack Type	What is the ransom paid by **type of attack**, per group?
Attack Type	What is the total number of hostages by **attack**, per group?
Attack Type	What is the principal country in which the kidnapping/hostage incident has started and resolved/ended by **attack**, per group?
Attack Type	What is the total amount of property damage by **attack type**, per group?
Percentage of hostages free	Which is the **percentage of hostages free**, per group?

DATA WAREHOUSE

4. Identify the facts

MEASURE	DATA TYPE	FORMULA
Criteria 1	int	-
Criteria 2	int	-
Criteria 3	int	-
Attack Type Text	nvarchar(50)	-
Attack Type	int	-
Success	int	-
Suicide	int	-
Percentage of hostages free	numeric(29,3)	(Number of hostages released / Number of hostages) * 100
Percentage of perpetrators captured	numeric(29,3)	(Number of perpetrators captured / Number of perpetrators) * 100

Comments: If the nº of perpetrators/hostages is 0 or -99 or if the nº perpetratores captured/hostages released is -99, then the formula should be Null.

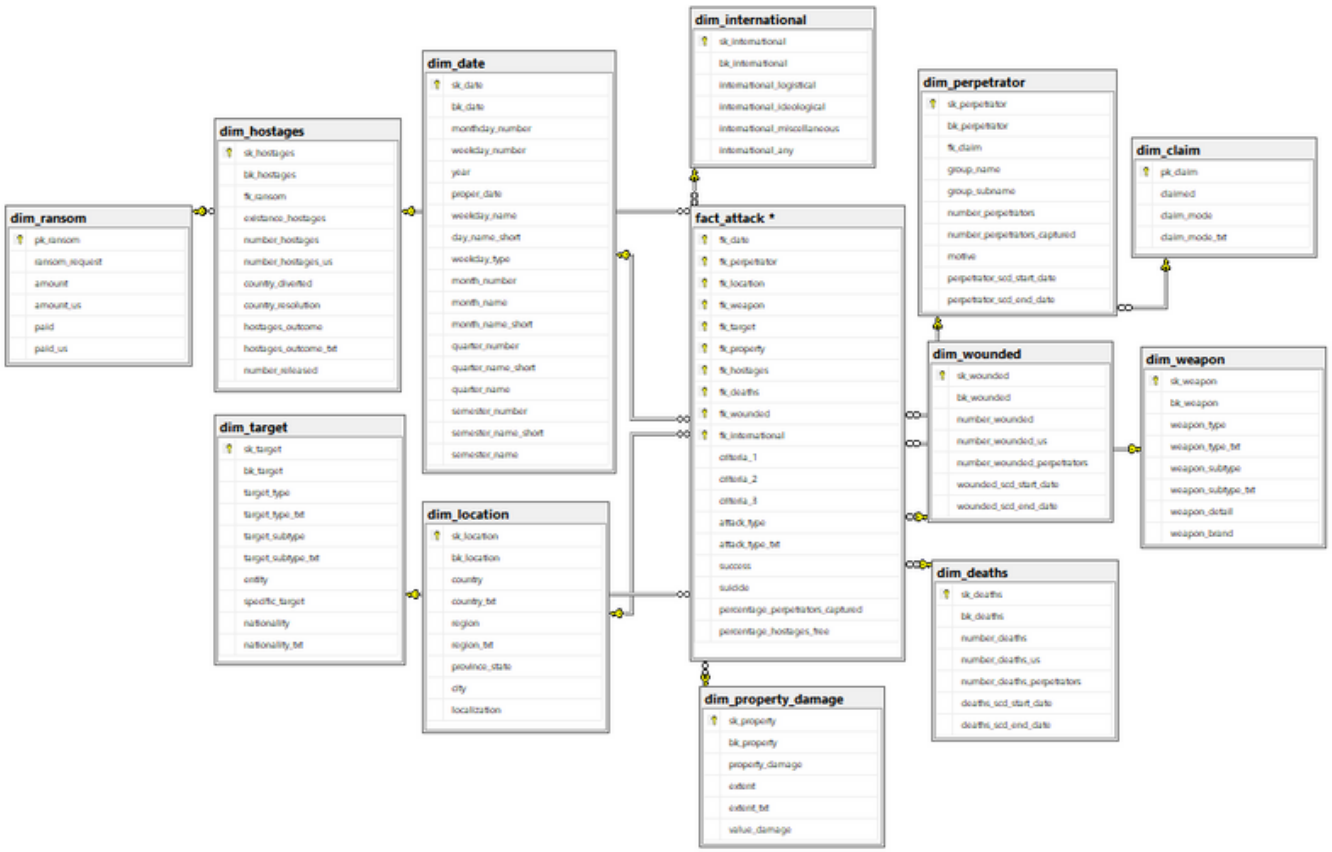
DATA WAREHOUSE

5. Create Schema

After making the connection of the dimension tables to the fact table, we began the assembly of the scheme.
We started by making a snowflake schema, which ended up allowing us a better quality of the data and, consequently, a better structure of the data. For these reasons mentioned above, we also improved the problems related to data integrity.

1.SNOWFLAKE SCHEMA

Although the snowflake scheme uses less space to store dimension tables, it should be noted that it has a complex database design, which ultimately makes it more complex with respect to information arrival.
When performing the transitions in the snowflake scheme, it should be noted that we have the fact table and its associated dimensions, however, it is quite complicated to analyze them because it requires a greater complexity when it comes to queries, queries that with a high number of connections leads to a lower performance.

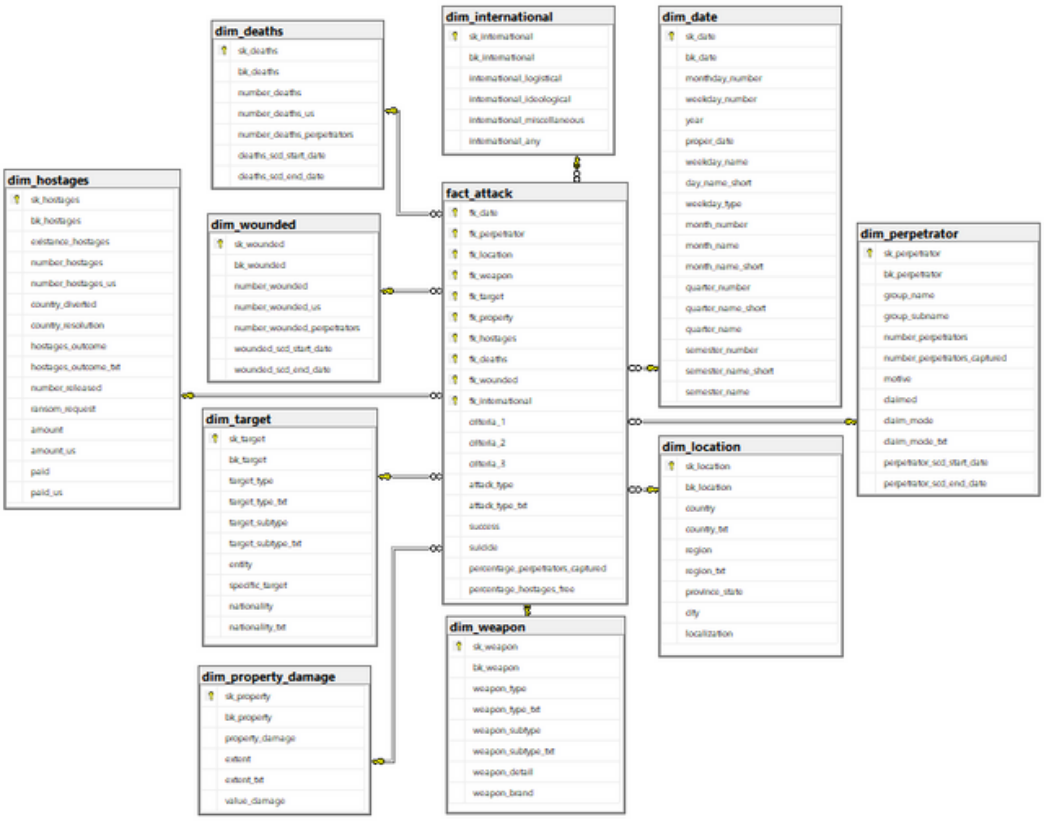


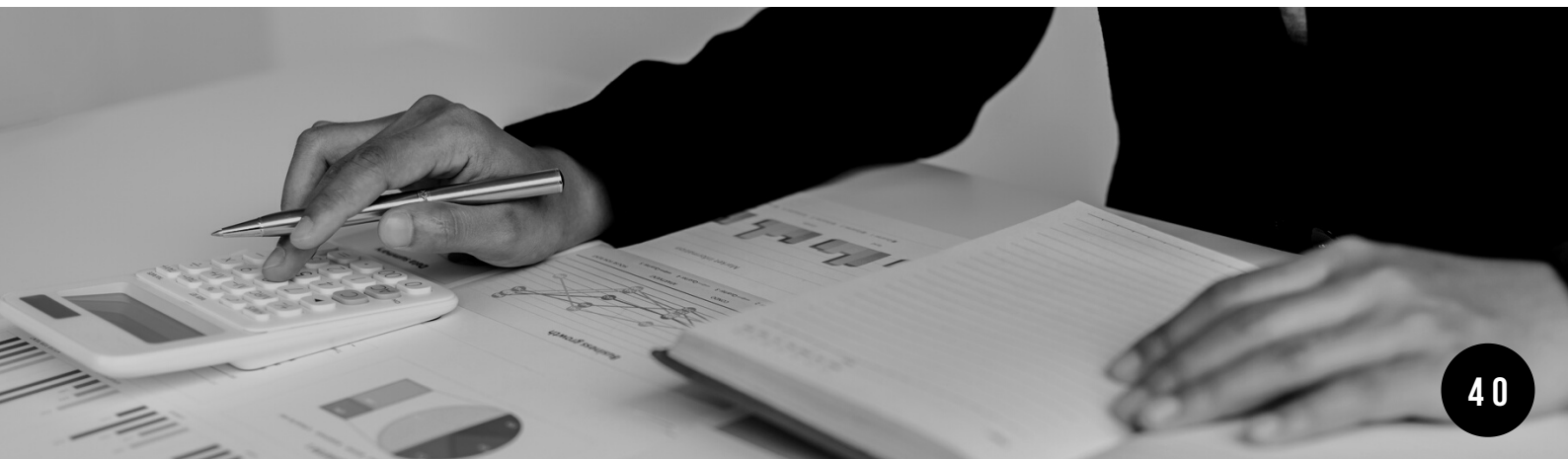
DATA WAREHOUSE

5. Create Schema

2.STAR SCHEMA

As for why we have changed from snowflake to star schema, we must emphasise the improvement in query performance, not forgetting that the analysis of each of the diagrams is better understood. As we know, the star schema joins the fact table to the respective dimension tables, which in turn leads to simpler and faster SQL queries.





DATA WAREHOUSE

5. Create Schema

3. FACT ATTACK

This table represents the facts of attacks that represent the characteristic derivatives of incidents, incidents that are carried out by terrorists.
Through these facts, we can better understand terrorist behaviour and, if possible, devise strategies to protect and secure civilians.
Once terrorist incidents are understood and analysed, we are able to draw certain conclusions that are beneficial to society.

ATRIBUTE NAME	IMPORTANCE	SOURCE
fk_date fk_perpertrator fk_location fk_weapon fk_target fk_property fk_death fk_wounded fk_international fk_hostages	Connectors between the dimension and the fact table	dim_date dim_perpetrator dim_location dim_weapon dim_target dim_property dim_death dim_wounded dim_international dim_hostages
criteria _1 criteria_2 criteria_3 attack_type attack_type_txt success suicide percentage_perpetrators_captured percentage_hosatges_free	Fields to respond the business needs	See the table in page 35 of the report

DATA WAREHOUSE

5. Create Schema

4. DIMENSIONS

SCD (SLOWLY CHANGING DIMENSIONS)

With regard to slowly changing dimensions, we think it is important to emphasise a few facts. With respect to the fundamental measures that we store in our data tables, it is important to note that these are time series, series that most of the fundamental measures that we store in our data tables are time series, which we write carefully with time stamps and foreign keys that in a sense link to calendar date dimensions. However it is important for us to know that time effects are not isolated acts when it comes to these timestamps, which in turn are based on activities.

Regarding all the other dimensions that are linked to the fact table, it should be noted that they are also affected by the passage of time, because as we know we are talking about attacks, in other words, the information that we receive one day may not be the same the next day regarding a certain attack, for example, the number of dead or wounded and the number of perpetrators (dimension wounded, perpetrator and deaths). As a solution to the above, we will use slowly changing dimensional attributes, which will allow us to pass on all the relevant information about an attack with precision. In order to deal with this situation, we will use variable attributes with the application of type 1 and 2 techniques, techniques that will allow us to keep with the expected precision of all kind of information, both historical and records, or just update the information, that is relevant for the fight against terrorism.

DATA WAREHOUSE

5. Create Schema

4. DIMENSIONS

DIM PERPERTRATOR

Dimension related to the individual or group that commited the crime which aids us in understanding the reason why they did it.
The specific motive for the attack, the nº of terrorists, the group, the mode used to claim responsibility, the motive and the atributtes are the variables inside this dimension. It a two level hierarchy: name of the group that made the attack and additional qualifiers or details about the name.

DIM WOUNDED

With regard to the wounded dimension, we can draw information about the number of injured, the number of injured in the United States and also the number of injured perpetrators. It has a two level hierarchy: number of non-fatal victims and perpetrators and number of non-fatal victims and perpetrators in US.

DIM LOCATION

Dimension related to the location that tell us where the attack happened. It has a five level hierarchy: country, region, province/state, city and specific location.

DIM TARGET

This part provides information about the target that perpetrator's reached.
Who or what was attacked, the general type of the target/victim, nationality, amongst other characteristics that led to the attack. It has a four level hierarchy: target type, target subtype, entity and the specific target.

DATA WAREHOUSE

5. Create Schema

4. DIMENSIONS

DIM WEAPON

From this dimension we get to see the specifics of the weapon: the type used, it's components and the brand. It has a four level hierarchy: weapon type, weapon subtype, weapon brand and weapon details.

DIM DATE

Dimension that allows us to analyze data in different aspects of date as efficiently and accurately as possible. It has a five level hierarchy: year, semester, quarter, month and day.

DIM PROPERTY DAMAGE

This dimension refers to the damaged property. From this we can derive information about whether or not the property was damaged, the extent of property damage, the value of the damage and specific details about the damage done.

DIM HOSTAGES

From this dimension we can conclude whether the attack involved hostages or not. It should be noted that kidnapping is a form of attack.
We also take the number of hostages and it is possible to identify the country in which they were held hostage. It has a two level hierarchy: number of hostages or kidnapping victims and number of US citizens that were kidnapped or taken hostage

DATA WAREHOUSE

5. Create Schema

4. DIMENSIONS

DIM DEATH

This dimension refers to the fatalities and therefore we can take the total number of deaths, the number of deaths for the United States and the number of deaths for perpetrators. It has a two level hierarchy: number of total fatalities and number of fatalities in US.

DIM INTERNATIONAL

Finally, there is the international dimension. This dimension retains logistical information, ideological information, miscellaneous information and also a mix of the three mentioned above.
We can say that it is based on the comparison of different nationalities.

5. COMMENTS

As can be seen, the Data Warehouse has less variables and tables comparing with the original data source. That is because we decided that some variables were not necessary taking into consideration our Business Needs. Another reason was that some of them did not have any value or had only a small percentage.

Just like we said in page 19, the name of the tables changed and we also changed the division of the tables. Another thing that we changed was the name of the variables because we thought that the original name was not very understandable.

ETL PROCESSES

INTRODUCTION TO ETL

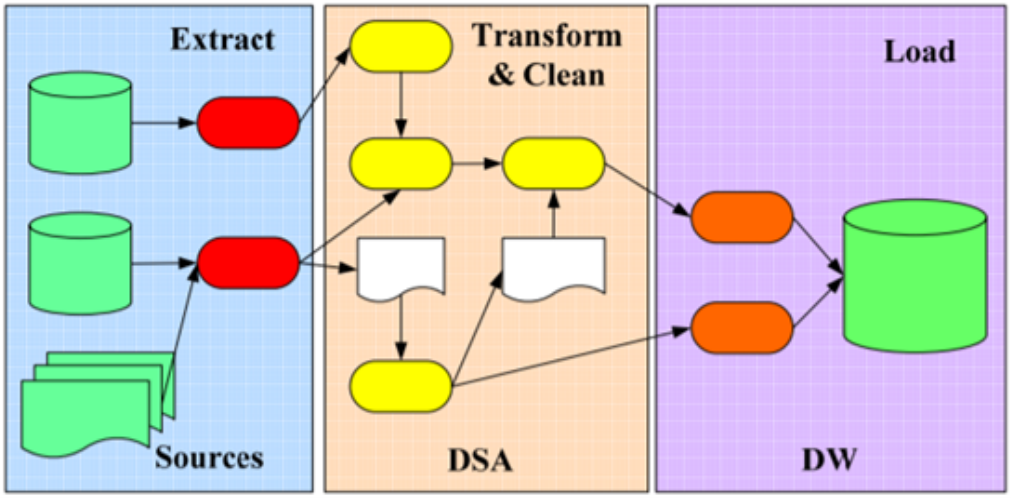
The ETL process, which means extraction, transformation and loading, responsible for the extraction of data from several sources, their cleansing, their customization, their transformation in order to fit business needs.

ETL is divided in three diferent stages with the main goal of synthesizing data. Basically, it consists of extracting the data from different source systems into the staging area, then transforming it, if needed, to provide important information and lastly, load the data into the Data Warehouse system.

The ETL process is used to make the process of extracting information easier, from the database and manages to take advantage of it's ability to reliably query data and obtain insights.

It should be highlighted that, as far as our project is concerned, we use two ETL processes. The first is to load the data from the dataset into the Staging Area, by loading the dimension tables and the fact table. The second is intended to load the first ETL process for the Data Warehouse.

We also made the ETL work suitable for a Full Load and also an Incremental Load, but we will mainly talk about the process of the second one.



ETL PROCESSES

IMPORTANCE OF INCREMENTAL LOAD

We took into consideration that ETL is a recurring activity, which can be daily, weekly or monthly, of a Data Warehouse system and wants agility, automation and good documentation. We aim to make ETL fast and easy to use.

Incremental loads are useful for large data sets because they run efficiently due to the speed and the less risk associated. At the same time, only the difference between the target and source data is loaded through the ETL process in Data Warehouse and, therefore, is much faster than full load.

Data that didn't change will be left alone, so we only use new and loaded data to the destination in a process of loading data incrementally.

Incremental loads can be challenging because you have to monitor the data in order not to have erros or anomalies and check if the process to ensure data in an ETL Data Warehouse is correct and consistent.

Incremental data loads have several advantages over full data load:

- They typically run considerably faster since they touch less data;
- Because they touch less data, the surface area of risk for any given load is reduced;
- Incremental load performance is usually steady over time;
- They can preserve historical data.

ETL PROCESSES

1.STAGING AREA

INTRODUCTION TO THE STAGING AREA

Staging area or landing zone works as an intermediate storage area used for data processing during the ETL process and, therefore, is mainly used to quickly extract data from its data sources and minimize the impact of the sources and, therefore, increase efficiency of ETL processes, ensuring data integrity and support data quality operations.

Typically, data enters a staging database, so that if something goes wrong, it is easier to identify problems and correct them.

> Staging area is where you hold the data and perform data cleansing and merging, before loadinging it into the Warehouse, (i.e., it's a place where you hold temporary tables).

IMPORTANCE OF THE STAGING AREA

The use of a Staging Area is extremely important and useful.

First of all, it should be pointed out that data in production systems undergo intensive processing, and not to mention that the insertion of incorrect values requiring transformation, or variables with missing values, would create interferences with sufficient impact on the systems.

Next, it is of equal relevance that the production systems must be ordered.

Remembering, and going back to why it is important to have a Staging Area, it should be mentioned that once we have some data that we are not sure whether or not it would be useful in the Data Warehouse environment, it is not disposable and may be useful for further analysis in the future, and is then kept in the Staging Area.

Last but not least, the tables in the DW tables are updated so that there is no drift of information.

ETL PROCESSES

1.STAGING AREA

LOG ETL AND LOG ERRORS TABLES

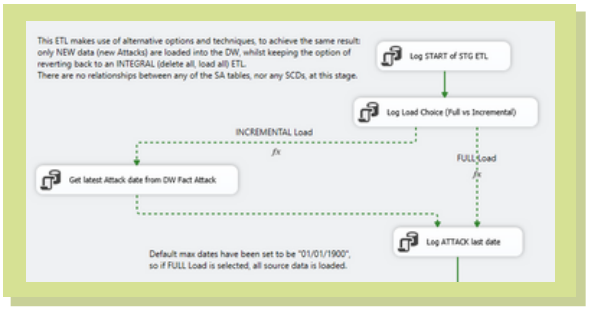
It's always important to record information when transforming the data, so that every task done through the process in SA is registered and, therefore, viewed when needed.

By creating, two tables: the first focus on the errors received during the execution of the package during the all process; and the second focus on registing the execution of the containers (deletion and loading of the execution, including time and targeted rows).

To report these steps and errors, as mentioned before, we resorted to creating two different tables, **log_stg_etl,** which records the execution (time and lines) of the containers (delete and upload) and another table that records all the errors, **log_stg_errors**, errors that were obtained when trying to execute the package during the whole process.

INCREMENTAL LOAD

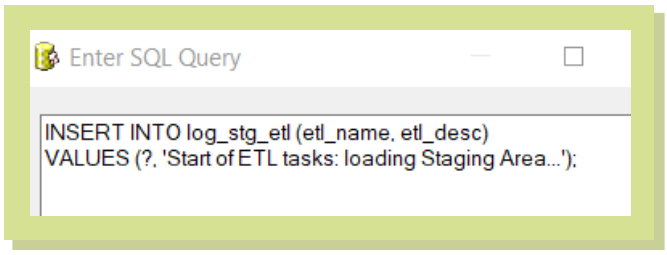
Regarding the incremental load underlying the Staging Area, it is important to mention that it needs more work than the full load, and there are some parameters that make it a better choice. These features are then its speed and efficiency since with a full load we have to delete all records and load them all again with the new records also, and the added capacity to store our historical information without it being lost, since with a full load all of these historical information will be deleted.



ETL PROCESSES

1.STAGING AREA

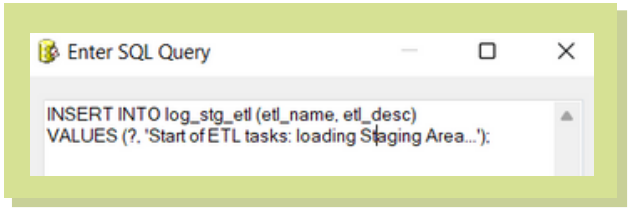
We start with the execution of the command **Log START of STG ETL**. The connection to the SA was defined in the Connection Manager, and a query to insert the information into the **log_stg_etl** table saying that the ETL tasks started was created.



In the Query shown in the previous page, the command refers to insert in the **log_stg_etl** table the following columns:

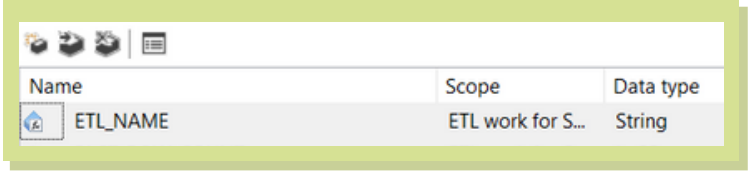
- **ETL_NAME**: String that consists in the date and time that the execution takes place. As the variable can take up many values, changing with time, we write a question mark in the following expression shown in the pictures below;
- **ETL_Desc**: Everytime we execute the Package, in the **log_stg_etl** in SQL the description in the table is going to be: "Start of ETL tasks: loading Staging Area...".

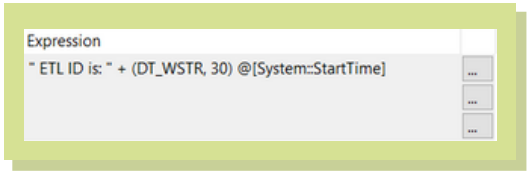
The expression mentioned previously concerning to the **log_stg_etl** can be noted in the picture on the right.



All of this is only possible by the addition of the variable **ETL_NAME**.

Since it is in a string format we need to add *(DT_WSTR, 30)* to convert. because it would not be read correctly in the SQL Server if it was left like a *timestamp*.





ETL PROCESSES

1.STAGING AREA

After the creation of the new variable, it is now possible to do a mapping between it and the "?" present in the expression of the command **Log START of STG ETL**, so the system substitues the date and time details properly.

Concerning the parameter mapping of the new variable, the parameter number is 0 since the variable is replacing the first "?" showing up. The parameter size is -1 (default size).

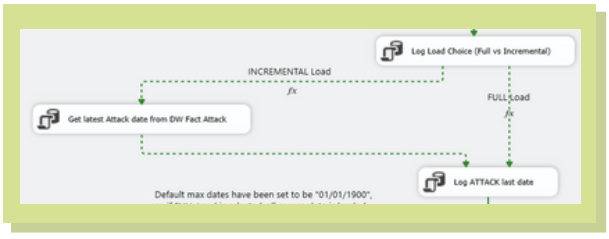
Variable Name	Direction	Data Ty...	Paramet...	Paramet...
User::ETL_NAME	Input	NVARC...	0	-1

As a result of the joint of all points described above, when we run the package, in the table **log_stg_etl**, in SQL Server Management Studio, the following will appear, showing a new row:

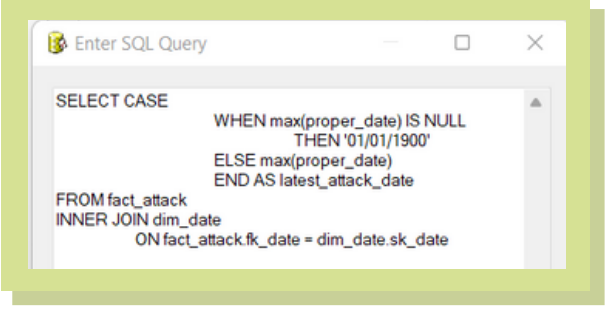
ETL ID is: 07/01/2022 16:08:09	Start of ETL tasks: loading Staging Area...

At this stage, there are no relationships between any of the Staging Area tables or any SCDs.

Although the incremental load is more efficient and quick, due to our database, we still need to use the full load. The full load consists in the deletion of all the information and data and reload it with the new records.



In our data there is this need regarding the date of attack. In this case we consider, in the incremental load, the last date available in the database of an attack and update this with the most recent one. Concerning the full load, we considered the maximum date as 1900. As all the dates are more recent, this way will register every attack.



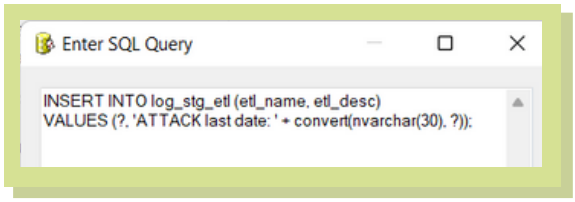
ETL PROCESSES

1.STAGING AREA

It is important to to highlight one important step still concerning the incremental load that is the creation of a variable that aims to save the details about the latest loaded attack date.

latest_loaded_attack_date	ETL work for STAGING AREA	DateTime	01/01/2022 19:57

A new command is created **Log ATTACK last date**. This command will be responsible to log the last attack date from one of the load methods available. The same thinking happens here with the "?" within the query of the command.



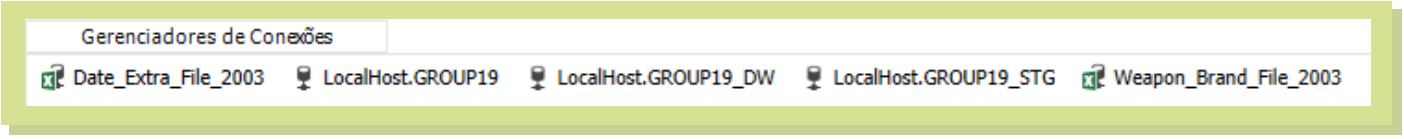
In the parameter mapping, we have present both created variables. However the Parameter Name of the variable **latest_loaded_attack_date** is 1 instead of 0 and the Parameter Size is -1.

Variable Name	Direction	Data Ty...	Parameter Name	Paramet...
User::ETL_NAME	Input	NVARC...	0	-1
User::latest_loaded_attack_date	Input	DATE	1	-1

ETL ID is: 07/01/2022 16:08:09	Start of ETL tasks: loading Staging Area...
ETL ID is: 07/01/2022 16:08:09	Is ETL using INCREMENTAL load? YES
ETL ID is: 07/01/2022 16:08:09	ATTACK last date: 1900-01-01 00:00:00

CONNECTION MANAGERS

Connection managers are used to configure a connection between SSIS and an external data source. In order to make the information flow we used five connectors: two of them in excel file (specifically in excel 2003) - to obtain the weapon brand and dates - a third connection manager to do the link to the Staging Area database in SQL Server Management Studio and the forth connection manager to do the link with Data Warehouse database in SQL Server Management Studio. Lastly, the last connection manager is concerning the main source (bacpac file that was imported to the SQL Server Management Studio).



ETL PROCESSES

1.STAGING AREA

DELETE THE DIMENSIONS AND TRUNCATE THE TABLES

It is important to note that before truncating the facts in the staging area, they were cleared from both the dimension tables and the fact table. To make this possible, both the dimension tables and the fact table were deleted and truncated, respectively.

As we are dealing with the Staging area, we can delete the fact table and the dimension tables all at once. This can happen because there are no relationships between the tables, at this stage.

To make the deletions of the dimensions tables and the truncate of the fact table, we created several SQL Execute Task - one per table - all contained in a sequence container. We decided to o with a sequence container as it is easier to group tasks together.

We still want to underline the fact that we are deleting all the values inside the tables of the dimensions but it is saved on the previously mentioned log in Management Studio. Regarding the Fact table, as it is not needed to have a lot of information we use the Truncate.

As we also created a SQL Execute Task named **Log Tables are Cleared**, everytime we run the package, in SQL Management Studio the following row will appear due to the expression inserted within the command query.

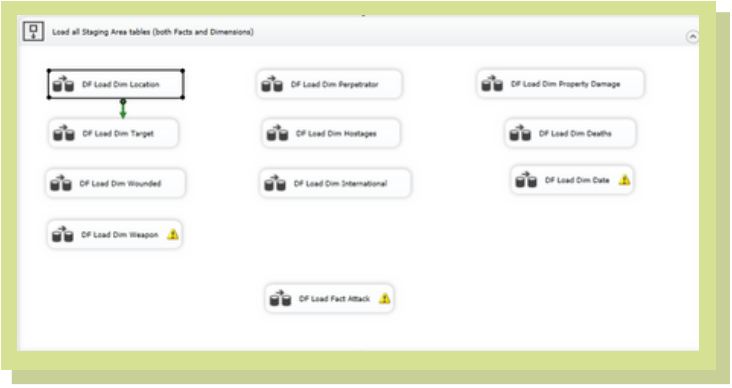
ETL ID is: 02/01/2022 17:45:42	All tables were deleted

ETL PROCESSES

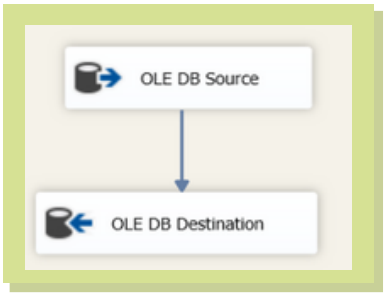
1.STAGING AREA

LOAD THE DIMENSIONS

In order to load all the data into the tables in SQL Management Studio we divided the process in two partes: first load the data of the Dimensions and then load the fact table. The order of this process does not really matter in the Staging Area once we did everything in one sequence cointainer.



The majority of the dimensions loaded came from the main source file (GROUP19) - bacpac.



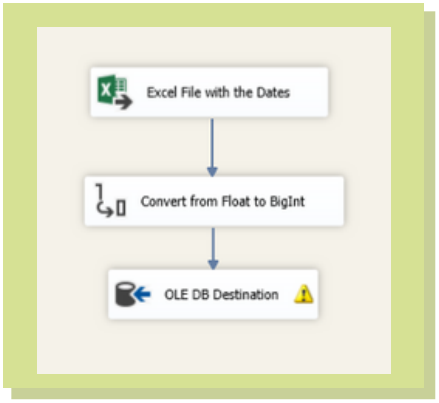
The only exceptions are the **Date** and **Weapon** Dimensions. The data from the **Date** Dimension came from the Excel file "Data_Extra_File_2003" and the data from the **Weapon** Dimension came from 2 different places: the Excel file "Weapon_Brand_File_2003" and the Bacpac "GROUP19".

We decided to change the Excel File that we used to the 2003 version. We also switched to "False" the *Run64BotRuntime* because without this the loading from the Excel would not function properly.

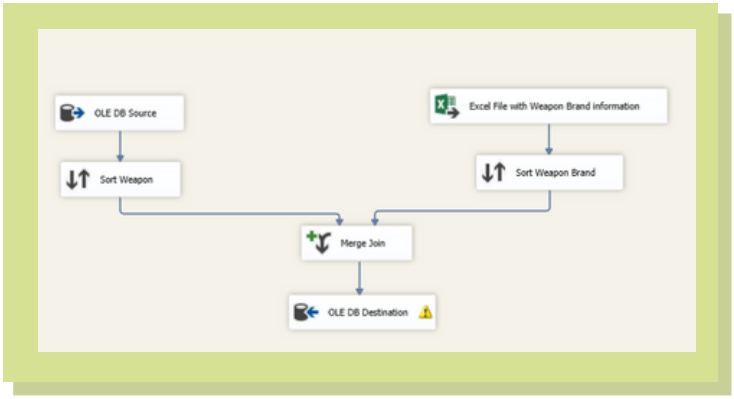
ETL PROCESSES

1.STAGING AREA

For the **Date** Dimension we also did a convertion of the variable *EventID - float* to *bigint* (eight-byte signed integer).



The data from the **Weapon** Dimension came from two different places so we did a merge join, but sorting in a ascendent way the information first. We use an inner join because we are sure that every single weapon has a match in the joining brand (inner join only returns matching rows).



Concerning the **Hostages** Dimension, in order to be able to get information about the hostages and the ransom (if it existed) we needed an SQL join statement across the tables of **Attack**, **Hostages** and **Random** to get the following variables: *HostagesID, Ishostkid, Nhostkid, Nhostkidus, Divert, Kidhijcountry, Hostkidoutcome, Hostkidoutcome_txt, Nreleased, Ransom, Ransomamt, Ransomamtus, Ransompaid* and *Ransompaidus.*

Regarding the **Perpetrator** Dimension, in order to be able to get information about each perpetrator and how we claimed (if he/they claimed) we needed an SQL join statement across the tables **Attack**, **Perpetrator** and **Claimed** to get the following variables: *PerpetratorID, Gname, Gsubname, Motive, Nperps, Nperpcap, Claimed, Claimmode* and *Claimmode_txt*.

ETL PROCESSES

1.STAGING AREA

LOAD THE FACTS

After we load the dimensions into the Data Warehouse, we load the fact table.

As we know, the fact table, when it comes to loading data, is actually more work than the dimension tables.

Firstly we use a SQL query to bring together all the variables that will be useful for the next steps - creation of measures and get the FK - (*EventID, InternationalID, HostagesID, PropertyID, WoundedID, DeathsID, WeaponID, PerpetratorID, TargetID, LocalID, Success, Suicide, AttackType1, AttackType1_txt, Crit1, Crit2, Crit3, Nperps, Nperpcap, Nhostkid* and *Nreleased*). To do that we use a SQL join statement across the following tables: **Attack**, **Attack Details**, **Perpetrator**, **Hostages** and **Incident**. This data comes from the main and principal source of information, the "GROUP19" bacpac.

At the same time we also import from the Excel 2003 File (Date_Extra_File_2003) the information about the dates. We will use it to obtain the *Proper Date* that will be useful for the incremental load.

After we sort both of the sources in a ascending way and join both of them, we calculate two measures - the calculation of the *percentage_perpetrators_captured* and the calculation of the *percentage_hostages_free*. The formulas of calculations are the following:

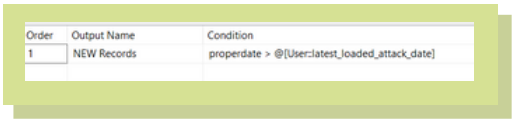
Derived Column Name	Derived Column	Expression	Data Type	Length	Precision	Scale	Code Page
percentage_perpetrators_captured	<add as new column>	(nperps == 0) || (nperps == -99) || (nperpcap == -99) ? NULL(DT_DECIMAL,3) : (nperpcap / nperps) * 100	numeric [DT_NUMERIC]		29	3	
percentage_hostages_free	<add as new column>	(nhostkid == 0) || (nhostkid == -99) || (nreleased == -99) ? NULL(DT_DECIMAL,3) : (nreleased / nhostkid) * 100	numeric [DT_NUMERIC]		29	3	

ETL PROCESSES

1.STAGING AREA

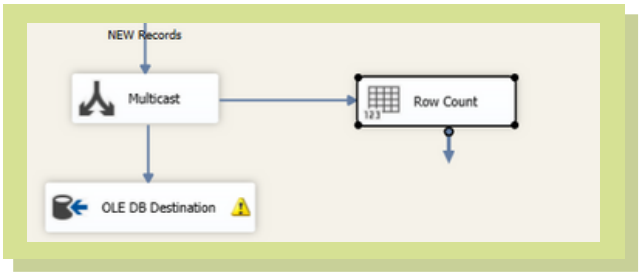
Then it is necessary to convert the variable *EventID - float* to *bigint* (eight-byte signed integer).

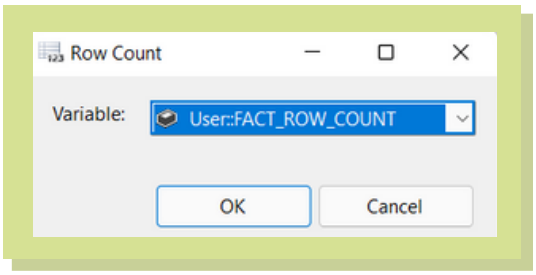
The following step is to do a conditional split. We will use the date of the last attack record that exists in the Staging Area to load newer records (ones that have more recent dates).



In order to get all the primary keys, we need to extract all the data from multiple tables in our source, sort and merge them at different stages of the process.

The last step is to do a multicast command that will load the data to 2 places: the final destination (to "Group19_STG") and to a row count that, based on the created variable " **Fact_Row_Count**, will count the number of rows that are created.





ETL PROCESSES

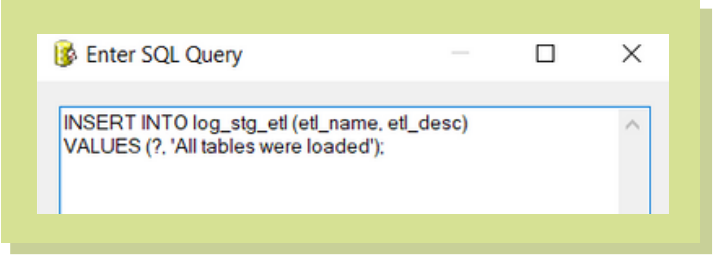
1.STAGING AREA

EXECUTE TASKS

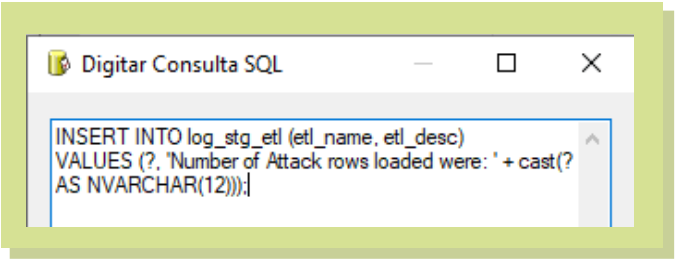
Coming now to the execution tasks, we made a copy of the SQL Execution Task titled **Load Tables are Cleared** which was then pasted below the **Load all Staging Area Tables** container being titled **Log Tables are Loaded**.

Finally, this last task will store in the **log_stg_etl** table a row with *All tables were loaded*.

At this point, we made some changes to the SQL command for the following changes below:



Having as an objective to know how many attack rows were loaded in the SQL Server Management Studio, another SQL Execute Task was made - **Log Count Attack Rows**. This task stores in the **log_stg_etl** table the number of rows loaded using the variable **FACT_ROW_COUNT.**



Nome da Variável	Direção	Tipo de Da...	Nome do ...	Tamanho d...
User::ETL_NAME	Input	NVARCHAR	0	-1
User::FACT_ROW_COUNT	Input	LONG	1	-1

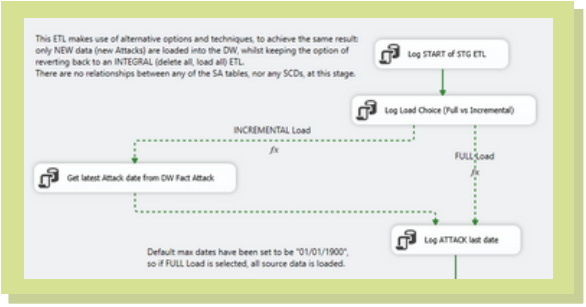
ETL ID is: 07/01/2022 16:08:09	All tables were loaded
ETL ID is: 07/01/2022 16:08:09	Number of Attack rows loaded were: 30644

ETL PROCESSES

1.STAGING AREA

FULL LOAD

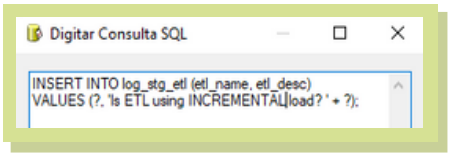
It is also possible to do a full load if the user wants to.



To do that we created a new param named **Is_Incremental_Load**. The value of it is setted to "YES", so it is going to be an incremental load. If the user wants a full load, he just has no change the value of this param to "NO".

Nome	Tipo de dados	Valor	Confidencial	Obrigatório	Descrição
Is_Incremental_Load	String	YES	False	True	Use "YES" for Incremental load, "NO" for Full load

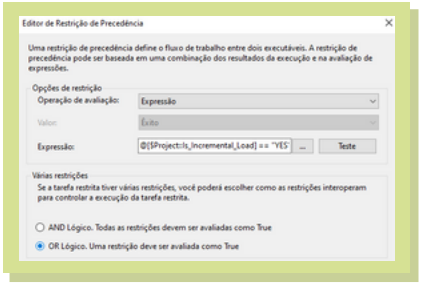
We also created a new SQL Execute Task named **Log Load Choice (Full vs Incremental)** that will store in the table **log_stg_etl** which was the choice (Incremental or Full). The same thinking that we used in the other logs, regarding the **ETL_NAME** and the **Is_Incremental_load**, and the query, is applied in these case.



Nome da Variável	Direção	Tipo de Da...	Nome do ...	Tamanho d...
User::ETL_NAME	Input	NVARCHAR	0	-1
$Project::Is_Incremental_Load	Input	NVARCHAR	1	-1

ETL ID is: 07/01/2022 16:08:09	Is ETL using INCREMENTAL load? YES

Then, depending on our choice (Full or Incremental), the path will change. This is possible with some changes regarding the precedence constraints - adding an expression ("YES" to incremental load and "NO" to full load) and changing from *AND* to *OR*.



The next steps regarding either a full or incremental load were already described.

ETL PROCESSES

2.DATA WAREHOUSE

With the end of the steps in the Staging Area, we transfer the ETL work from the Staging Area to the Data Warehouse, using the incremental load. For this, we started with the creation of the log start of ETL for DW by coping the **Log START of STG ETL** from the ETL work for STAGING AREA package.

After this, we created the variable **ETL_NAME** (just like we did for the **ETL_NAME** in the ETL work for STAGING AREA). This variable was used on the SQL Execute Task named **Log START of DW ETL**.

It is important to note that, for the incremental load there's no need to delete the tables first. We will only load the new rows from the Staging Area, so we created two sequence containers, one for loading the dimensions and one for the facts.

First we have to load the dimensions to after be able to load the facts.

In the case of a full load, we have to empty every table before we load the tables, so in this case we delete everything from the fact table first and then we delete from the dimensions.

LOAD THE DIMENSIONS

The first think we did was add the source and check the columns in all dimensions. We emphatize that Surrogate Key is automatically created as new data arrives in the system in the form of new rows.

This is justifyed by the fact that when we created the Data Warehouse we parameterized the Dimension tables saying that the Surrogate Key is an identity (1,1), in this way we order the server to create it mechanically.

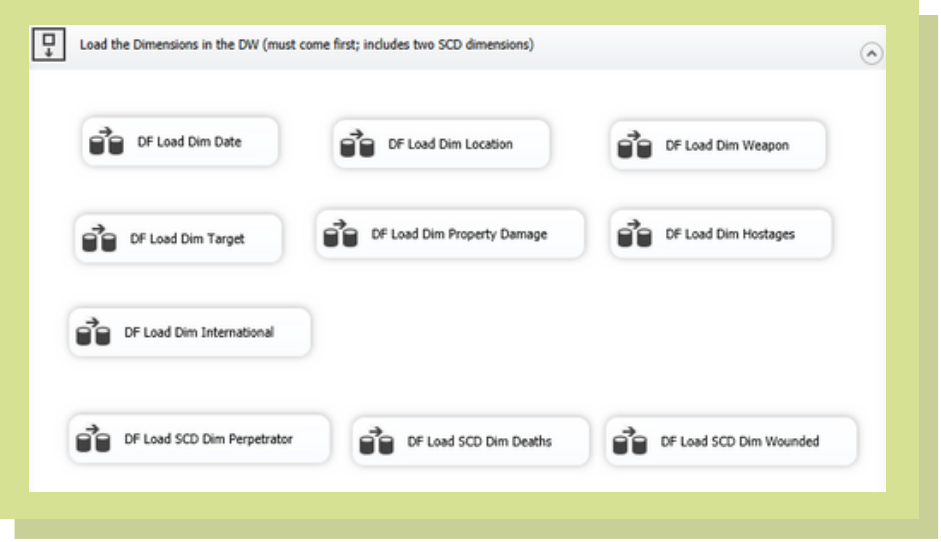
To complement what has been said, the code below has been added to each Surrogate key in the respective line of code in SQL:

```
sk_date INT IDENTITY(1,1) PRIMARY KEY,
```

ETL PROCESSES

2.DATA WAREHOUSE

When using incremental load we face two types of dimensions, some with SCD and others without it, so we have to apply Slowly Changing Dimension to both, but with different approaches.



LOAD THE DIMENSIONS WITH SGD

A key concept in data warehousing is tracking changes over time. In a multidimensional design we choose which frequently occurring events we want to capture, and then we store information about these in fact tables.

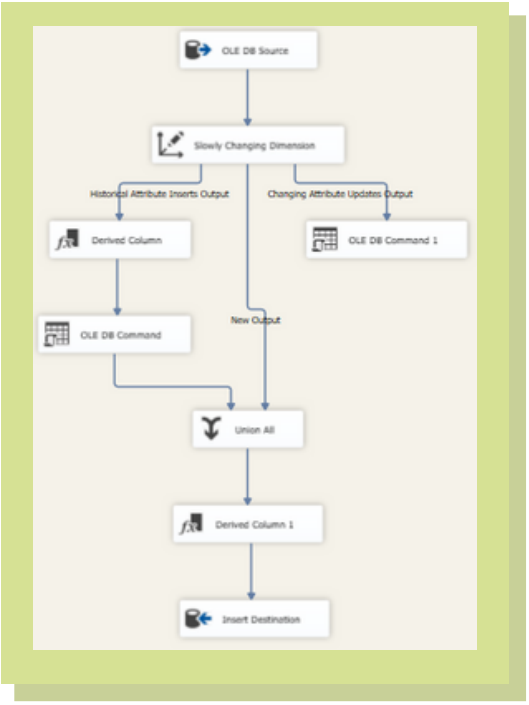
For this type of dimensions, after defining the source, we added a Slowly Changing Dimension task where we used the attributes of **Type 1 (Changing Attribute)** and **Type 2 (Historical Attribute)**.

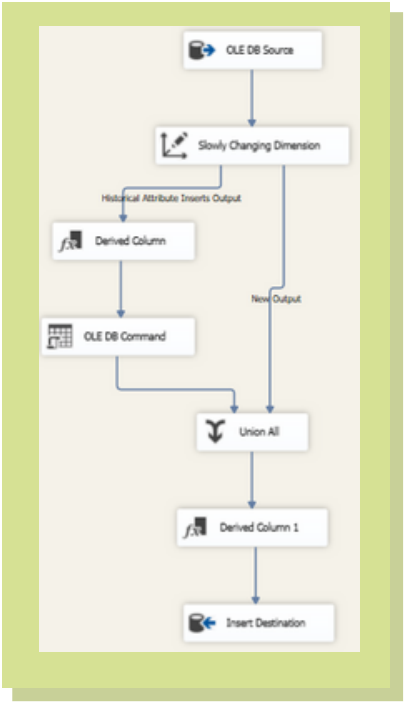
In **Historical Attribute (Type 2)**, an additional column is created for storing the historical attribute value, which means that every time a change occurs, a new row is created, making the method most useful for singular attribute changes. In this type all dimension rows are marked with a time interval in which the attribute values are valid and for each change a new row is inserted in the table. The *SCD_Start_Date* and *SCD_End_Date* will allow us to know the period of time the information is valid.

ETL PROCESSES

2.DATA WAREHOUSE

With a **Changing Attribute (Type 1)** we change the old records by overwriting them with the new ones, making it useful when old information is of no interest. Thus the existing data is lost as it is not stored anywhere else.





Dim Perpetrator

Changing Attribute: claim_mode, claim_mode_txt and claimed.
Historical Attribute: group_name, group_subname, motive, number_perpetrators and number_perpetrators_captured.

Dim Deaths

Historical Attribute: number_deaths, number_deaths_perpetrators and number_deaths_us.

Dim Wounded

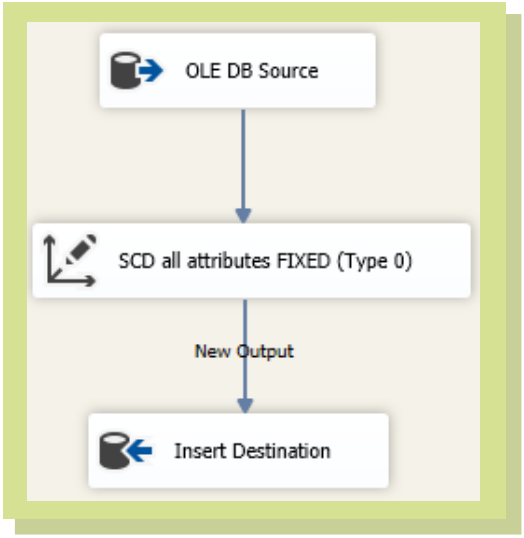
Historical Attribute: number_wounded, number_wounded_perpetrators, and number_wounded_us.

ETL PROCESSES

2.DATA WAREHOUSE

LOAD THE REMAINING DIMENSIONS

It's important to note that for the dimensions *Location*, *Target*, *Weapon*, *Date*, *Property_Damage*, *Hostages* and *International*, we will not change the values and because of that those are not really SCD and do not have the *SCD_Start_Date* and *SCD_End_Date.*



After we outline the source as the Staging Area, to proceed with the incremental load all attributes were defined as **Fixed Attributes (Type 0)**, we used the Slowly Changing Dimensions for that.

This procedure serves as a guarantee that when the data update is loaded everything in the Staging Area remains the same and warns us in error form if we try to load a line with an existing BK.

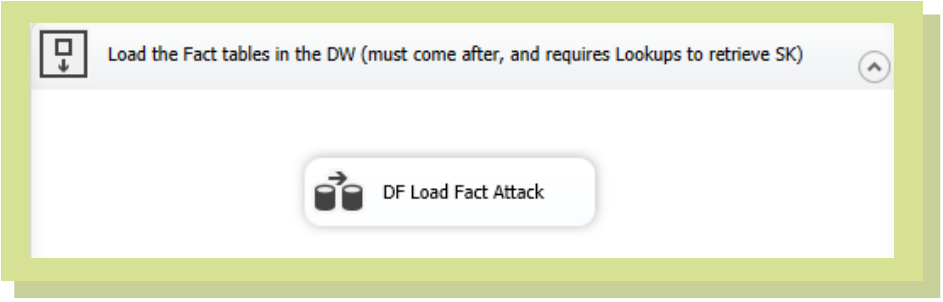
After all the dimensions have been loaded, we finish with a SQL Execute Task that was named **Log Loaded Dimensions**. This task records in the **log_stg_etl** table whenever the dimension tables are loaded.

ETL PROCESSES

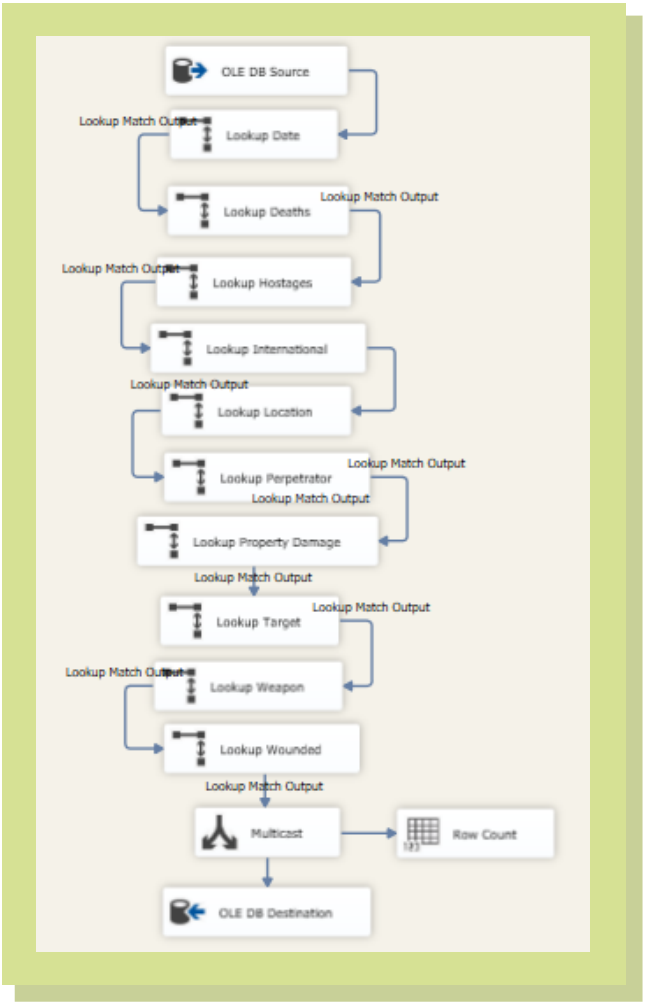
2.DATA WAREHOUSE

LOAD THE FACTS

Now that we loaded the dimensions, we will load the fact table.



We used the Lookup Task tool to load the fact table, as it was only through the dimensions that we got the Surrogate Key for each dimension. For that reason, we use the Lookup to make the link between the FK of the dimensions in the Fact Table and the BK of the dimensions in the Dimension Table and retrieve the SK of that Dimension Table to bring to the Fact Table.



Before adding the OLE DB Destination, we will use a Multicast to count the number of rows (Row Count was made just like in the ETL work for Staging Area) that were loaded into the fact table and to connect to the destination.

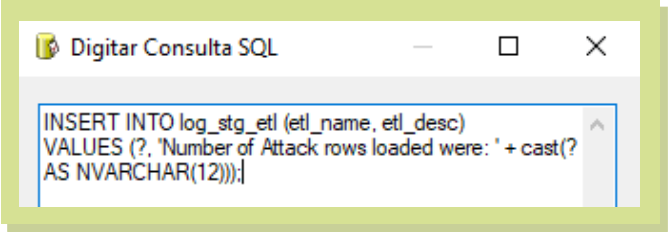
After completion of the entire process of loading the facts , we apply the SQL Execute Task named **Log Loaded Facts,** that records in the **log_stg_etl** table whenever the fact tables are loaded using the variable **ETL_NAME.**

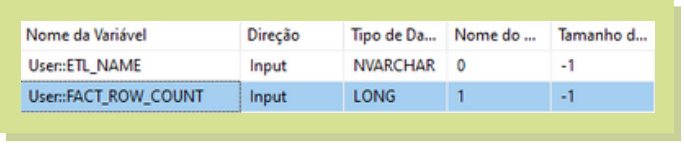
ETL PROCESSES

2.DATA WAREHOUSE

EXECUTE TASKS

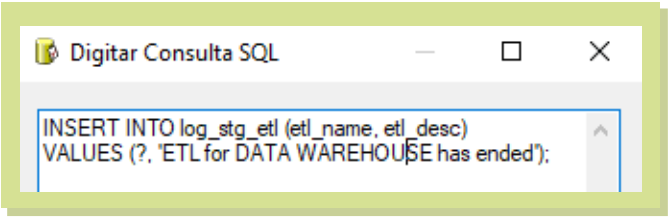
Having as an objective to know how many attack rows were loaded in the SQL Server Management Studio, another SQL Execute Task was made - **Log Count Attack Rows**. This task stores in the **log_stg_etl** table the number of rows loaded using the variable **FACT_ROW_COUNT**, just like we did for the ETL work for Staging Area.



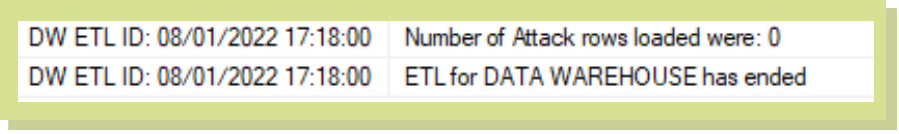


Nome da Variável	Direção	Tipo de Da...	Nome do ...	Tamanho d...
User::ETL_NAME	Input	NVARCHAR	0	-1
User::FACT_ROW_COUNT	Input	LONG	1	-1

To close the ETL work for Data Warehouse, we created the last SQL Execute Task - **Log END of DW ETL** - by copying the **Log START of DW ETL** and just changing the SQL query.



Nome da Variável	Direção	Tipo de Da...	Nome do ...	Tamanho d...
User::ETL_NAME	Input	NVARCHAR	0	-1



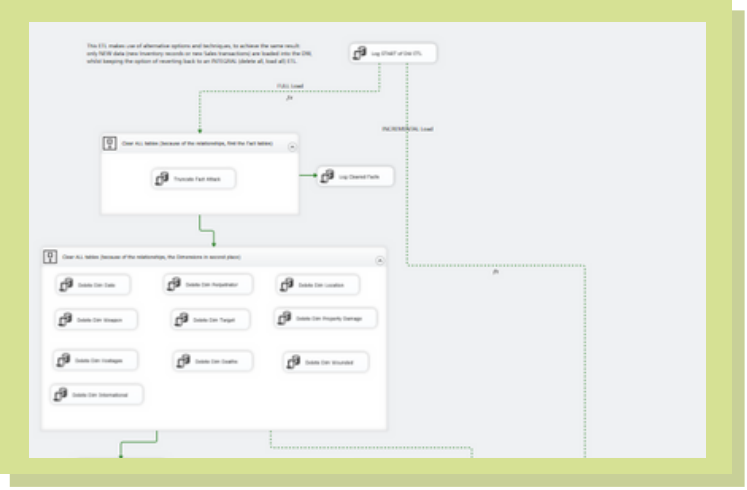
DW ETL ID: 08/01/2022 17:18:00	Number of Attack rows loaded were: 0
DW ETL ID: 08/01/2022 17:18:00	ETL for DATA WAREHOUSE has ended

ETL PROCESSES

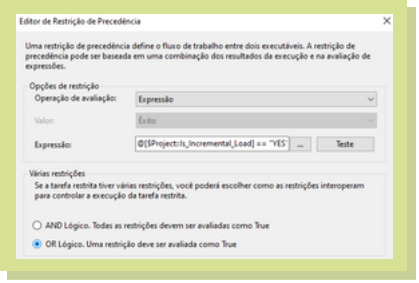
2.DATA WAREHOUSE

FULL LOAD

If the user chose a full load, then there are some changes in the process.



We have the same changes regarding the precedence constraints (dashed lines).



Then, just like we already said, if we choose a full load, we have to delete/truncate all the tables, in a certain order (fact table before dimensions) before we load them.

To delete/truncate the tables, we used the same process as in the ETL work for Staging Area. The only diference is that we have a container for the fact table and then a container for the dimensions.

We also have two new SQL Execute Tasks: **Log Cleared Facts** and **Log Cleared Dimensions**. Both of them work just like the log used in the ETL work for Staging Area, **Log Tables are Cleared**.

ETL PROCESSES

3. MASTER CONTROL

We created a new SSIS Package named MASTER Control of ALL ETL that calls the execution of all the other packages, in whatever order is necessary to ensure the correct and successful running of all tasks in ETL.

This Packages has two Execute Package Task, one for the Staging Area ETL and another one for the Data Warehouse ETL.

To make them work we select the Package that we want to run in the **PackageNameFromProjectReference**.

Pacote	
ReferenceType	**Referência do Projeto**
PackageNameFromProjectReference	**ETL work for DATA WAREHOUSE.dtsx**
Password	********
ExecuteOutOfProcess	**False**

Pacote	
ReferenceType	**Referência do Projeto**
PackageNameFromProjectReference	**ETL work for STAGING AREA.dtsx**
Password	********
ExecuteOutOfProcess	**False**

Then we just have to press *Start* and the Packages will run in the correct order.

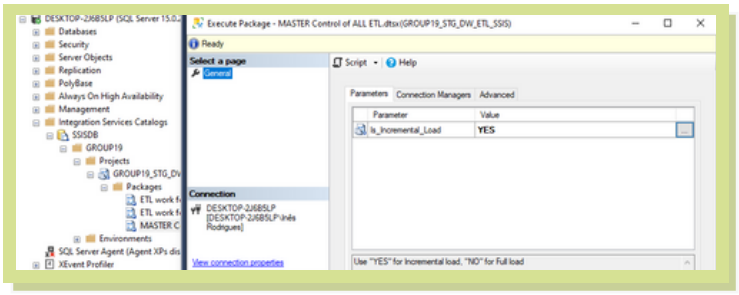
We also made it possible to execute the ETL work and see the information about this execution in the SQL Management Studio.

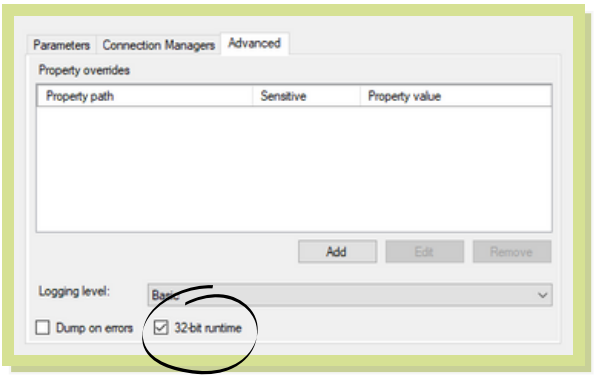
To do that we started by creating a new catalog in the Integration Services Catalogs on SQL named SSISDB. Next, in the Visual Studio, we deployed the GROUP19_STG_DW_ETL_SSIS SQL, in a folder name GROUP19 inside the SSISDB catolog.

ETL PROCESSES

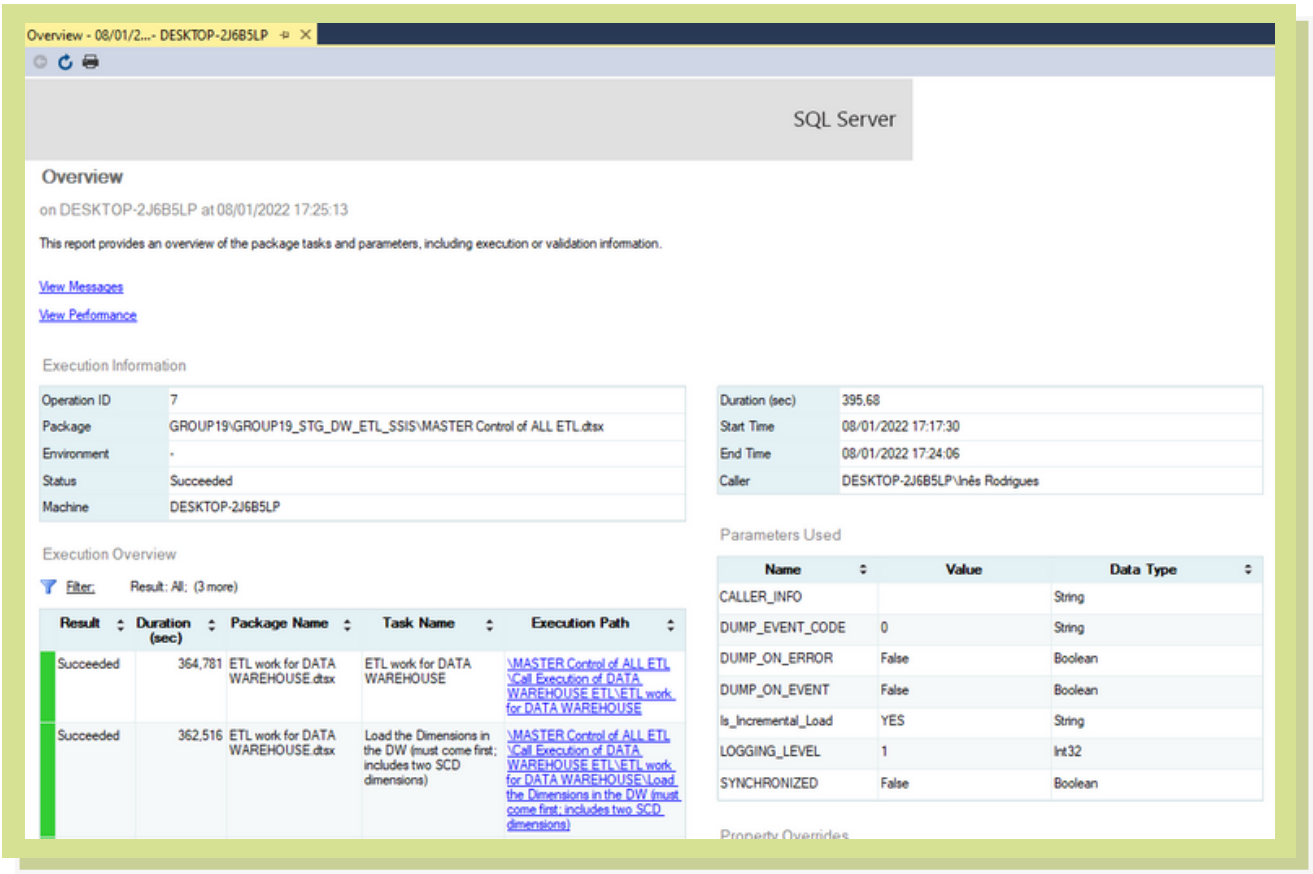
3. MASTER CONTROL

Now, to see the information about the execution of the ETL work in the SQL, we just need to go to the folder GROUP19 inside the SSIDB and execute the MASTER Control of ALL ETL. Then we choose if we want and incremental or full load and we also need to select the *32-bit runtime*.





Then the following page will appear, where we can see all the information about the execution of the packages.



ETL PROCESSES

4. IMPROVEMENTS

MAXCONCURRENTEXECUTABLES

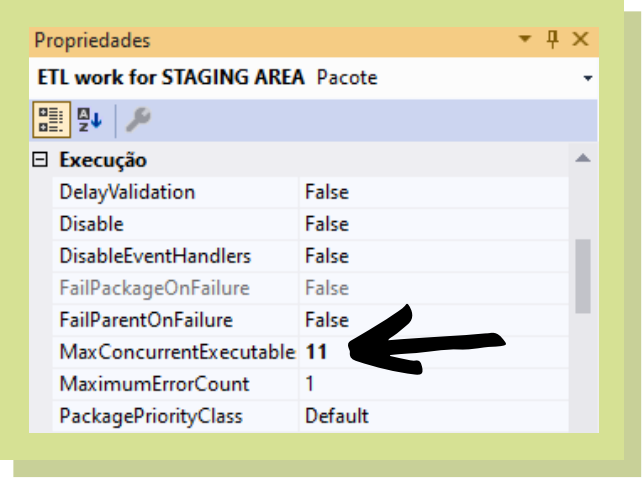
We thought about making the Visual Studio finish the ETL work faster so we configured the packages for parallel execution. This boosts the performance on computers that have various physical/logical processors.

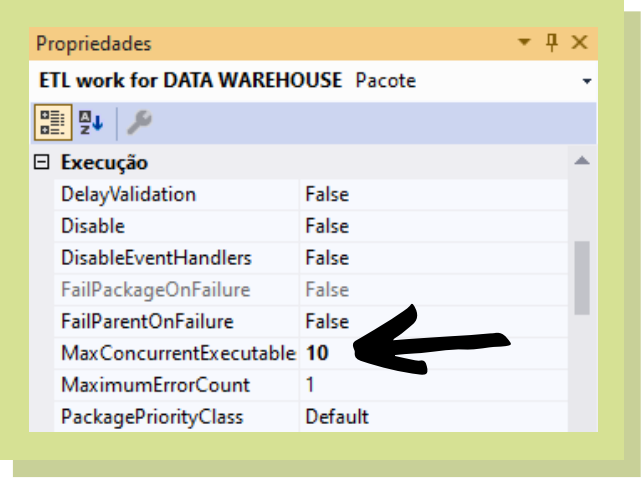
To configure the packages for parallel execution we changed the **MaxConcurrentExecutables**.

With the **MaxConcurrentExecutables** property we can establish how many tasks run at the same time. By default its value is -1, which means the number of physical/logical processors plus 2.

For example, if we have a package that has 3 Data Flows and if the **MaxConcurrentExecutables** is set to 3, then the 3 Data Flows will run at the same time.

Since the number of maximum Data Flows in a container in the ETL work for Staging Area is 11 and in the ETL work for Data Warehouse is 10, we changed the **MaxConcurrentExecutables** to 11 and 10, respectively.





CONCLUSION

The whole execution and organization of the project was in fact a help at various levels, with regard to the skills we possess, whether at problem solving level, or at the level of group work spirit, or at the level of personal learning.

Everything that involved the practical part of the project was extremely important for the complementary and individual understanding of the practice obtained in the classroom. We became in a certain way autonomous in what concerns the creation of Data Warehouses as well as in the ETL process, and combined they increased the efficiency of each of the group members in SQL and Visual Studio.

Speaking now of the first delivery, and what led us to choose this database. It is important to say that we tried to get a little out of our comfort zone and go beyond borders in terms of creativity, with this we tried to implement the knowledge learned in class and everything that the realization of the project required and apply all this to a problem that we find very relevant and that happens every day around the world: Terrorism. We tried to find as much information as possible about this subject and the database used also helped us a lot.

.

CONCLUSION

During the second part of the work and consequent second delivery, it was of general character the increase of responsibility as a group, and we all felt more familiar with the theme and with all the process that we had to do to conclude this project. In this second part we dedicated the realization of two ETL processes, one of them centred in the Staging Area and the other in the Data Warehouse, without forgetting that these same processes were very important for the correct realization of the Data Warehouse.

In the complete picture, we always had in mind our business needs and what we want to deliver as a benefit regarding the improvement and possible disappearance of Terrorism.

This way, and after the execution of our entire project, the competent authorities were able to access crucial information that could greatly facilitate the streamlining of processes when we talk about a possible terrorist attack, and it will also be easier to predict the ways in which these attacks could occur.

REFERENCES

Global Terrorism Database | Kaggle

GTD | Global Terrorism Database (umd.edu)

START (umd.edu)

Terrorism - Our World in Data

GTI-2020-web-2.pdf (visionofhumanity.org)

Global Terrorism Index [2020] » Vision of Humanity

NATIONAL STRATEGY FOR THE PREVENTION AND COUNTERING OF TERRORISM FOR 2017–2021

O impacto do terrorismo no mercado de capitais (uac.pt)

TERRORISMO INTERNACIONAL CONFERÊNCIA PROFERIDA NO INSTITUTO DA DEFESA NACIONAL, AO CURSO DE DEFESA NACIONAL EM FEVEREIRO DE 1982

Como impedir os ataques terroristas? As medidas da UE em síntese | Atualidade | Parlamento Europeu (europa.eu)

Automated presentation of slowly changing dimensions - Christer Boedeker

Slowly Changing Dimension Transformation | Microsoft

Data Flow Performance Features - SQL Server Integration Services (SSIS) | Microsoft Docs

Welcome To TechBrothersIT: What Is MaxConcurrentExecutables Property in SSIS Package?