Table of Contents

Chapter 1: Introduction 1

1.1. Introduction 1

1.2. Problem Statement 1

1.3. Objectives 2

1.4. Scope 2

Chapter 2: Literature Review 3

2.1. Data Warehousing 3

2.2. Online Analytical Processing (OLAP) 5

2.3. Snowflake Schema 7

2.4. MySQL Workbench 8

2.5. Visual Paradigm 8

Chapter 3: Methodology 9

3.1. Database Schemas and ERD 9

3.1.1. Original ERD for Operational Database 9

3.1.2. Constructed Data Warehouse Schema 11

3.2. Data Cube Representations 13

3.2.1. Information Package Diagram 13

3.3. Lattice of Cuboid 13

3.4. Hierarchical Structures 14

Chapter 4: Discussion and Analysis 15

4.1. Roll-up, or Drill-up 15

4.1.1. Every Continent 15

4.1.2. Cold War and Post-Cold War Era 17

4.2. Roll-down, or Drill-down 18

4.2.1. Every Country 18

4.2.2. Every Year after Cold War 20

4.3. Slicing 21

4.3.1. The United States 21

4.3.2. Pressurised Water Reactor 23

4.4. Dicing 24

4.4.1. Decommissioned Plants in Post-Soviet Nations 24

4.4.2. Operational Plants in China after 2000s 26

4.5. Drill-across 28

4.6. Pivoting 29

4.6.1. Status against Reactor Type 29
4.6.2. Reactor Type against Status 32
4.6.3. Decades against Top Nuclear Nations 35
4.6.4. Top Nuclear Nations against Decades 38
Chapter 5: Conclusion 40
References 41

List of Tables

Table 3.1: Information Package Diagram 13
Table 4.1: Rollup Query #1........ 15
Table 4.2: Rollup Query #2........ 17
Table 4.3: Rolldown Query #1 18
Table 4.4: Rolldown Query #2 20
Table 4.5: Slicing Query #1 21
Table 4.6: Slicing Query #1 23
Table 4.7: Dicing Query #1........ 24
Table 4.8: Dicing Query #2........ 26
Table 4.9: Pivoting #1 (1/2) 29
Table 4.10: Pivoting #1 (2/2) 32
Table 4.11: Pivoting #2 (1/2) 35
Table 4.12: Pivoting #2 (2/2) 38

List of Figures

Figure 2.1: OLAP Architecture 5
Figure 2.2: Snowflake Schema 7
Figure 2.3: MySQL Workbench 8
Figure 2.4: Visual Paradigm 8
Figure 3.1: Nuclear Power Plant Geodata Schema 9
Figure 3.2: Proposed Data Warehouse Schema 11
Figure 3.3: Lattice of Cuboid 13
Figure 4.1: Rollup Query #1 Output 16
Figure 4.2: Rollup Query #2 Output 18
Figure 4.3: Rolldown Query #1 Output 19
Figure 4.4: Rolldown Query #2 Output 21
Figure 4.5: Slicing Query #1 Output 22
Figure 4.6: Slicing Query #1 Output 24
Figure 4.7: Dicing Query #1 Output 26
Figure 4.8: Dicing Query #2 Output 27
Figure 4.9: Pivoting #1 Output (1/2) 31
Figure 4.10: Pivoting #1 Output (2/2) 34
Figure 4.11: Pivoting #2 Output (1/2) 37
Figure 4.12: Pivoting #2 Output (2/2) 39

List of Abbreviations

OLTP	Online Transaction Processing
OLAP	Online Analytical Processing
SQL	Structured Query Language
ERD	Entity Relationship Diagram
WNA	World Nuclear Association
IAEA	International Atomic Energy Agency

Chapter 1: Introduction

1.1. Introduction

According to IBM, OLAP (online analytical processing) is a software to perform multidimensional analysis at high speeds on large volumes of data from multiple sources such as a data warehouse, data mart, data lake, or any other data store. Most real-world data have multiple dimensions and categories of data which can be broken down for presentation, tracking, or analysis (IBM Cloud Education, 2020). For example, businesses analyse their sales data with several dimensions related to location (region, country, state, province, zip code, postcode, branch, store), time (year, quarter, month, week, day), product (category, brand, type), and many more.

In this report, the chosen domain is geographic data and information (geodata for short) for nuclear power plants. Geodata is basically data and information associated with a location on Earth. This report will provide a comprehensive review and analysis of all nuclear power plant locations in the world using OLAP. Geodata can benefit from OLAP because it can provide answers to many queries for who, what, where, when, and why. (Kramer & Nisbet, n.d.).

1.2. Problem Statement

1. **Difficult to interpret data in operational database**.

OLTP databases, or relational databases in general are not designed for analysis. As a result, it takes time and effort to get responses from these databases. Specialized databases called OLAP databases are made to extract this business intelligence data from the data. In this domain, there is one table linked to 3 lookup tables. This table stores data about nuclear power plants globally such as the reactor type, model, status, coordinates, etc. All this data is grouped together making them hard to interpret. By separating the data into several dimensions, useful patterns and correlations can be derived, and users can make faster, better decisions especially while doing nuclear-related research work.

2. **Lack of interesting patterns and correlations**.

Relational databases display data in a tabular form with only 2 dimensions (rows and columns). Though the data is organised and filtering can be performed too, it is difficult to interpret such data for common patterns and interesting knowledge. A data cube in

OLAP makes data easier to understand because it displays data in at least 3 dimensions. It is especially helpful when data and dimensions are represented together for decision making. OLAP can display information in a way such that users can see patterns, perform analysis, and facilitate decision-making process from different perspectives. In this domain, there are many factors that influence the evolution of nuclear power plants such as geopolitical tensions, economical factors, social changes, historical events, and geographical reasons. Hence, displaying the data based on locations, timelines, reactor types, etc allows for users to analyse correlations between real-world events and nuclear power plants.

1.3. Objectives

1. To develop a data warehouse based on the snowflake schema for nuclear power plants geodata.
2. To deliver a comprehensive report on process of converting the original datasets into data warehouse, and performing OLAP operations on the created data warehouse.
3. To discover interesting patterns and correlations between nuclear power plants and real-world events and timelines.

1.4. Scope

The target users are geography students and teachers, nuclear-related researchers, climate activists, or just about anyone who require data about global nuclear power to plants to be presented in a clear, intuitive manner. OLAP allows restructuring and reporting of nuclear power plants information in a more intuitive manner for them to perform analysis and construct useful charts.

This assignment will deliver a comprehensive report and an OLAP application consisting of the data warehouse schema, information package diagram, cube representations diagrams, Python scripts, and SQL codes. The SQL codes consist of the codes to create data warehouse, populate the tables with data from the original database, and run OLAP operations such as roll-up, drill-down, slicing, dicing, and pivoting. The operations utilise MySQL JOINS, aggregate functions such as COUNT and SUM, and specific functions such as TIMESTAMPDIFF, SUBSTRING, CONCAT, and YEAR.

Chapter 2: Literature Review

2.1. Data Warehousing

A data warehouse is a semantically consistent database that holds the data that an organisation requires to make strategic decisions. It acts as a physical implementation of a decision support data model. Data from several heterogeneous sources are combined to create a data warehouse, which is then used to support structured and/or ad hoc queries, analytical reporting, and decision-making (Han & Kamber, 2006). Essentially, a data warehouse is a collection of facts and dimensions. A value that represents a particular value is what is referred to as a fact. A dimension is a framework that includes the measures needed to give users the tools they need to solve queries (Nanda, Gupta, & Vijrania, 2019).

There are three types of data warehouse, namely enterprise warehouse, data mart, and virtual warehouse (Han & Kamber, 2006). The first type is enterprise warehouse It stores an enterprise's historical business data, that is all its previous operational data. These data usually originate from systems such as Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), and Online Transactional Processing (OLTP). Enterprise warehouse is a common data storage option today because it offers scalability based on the business and data requirements (Chhabra, 2022). The second type is data mart. A data mart is a basic type of data warehouse that is concentrated on one topic or area of business, such sales, finance, or marketing. Data mart offers a quick centralised source of information, provides better insights to decision making, and simple to implement (Oracle, 2022). The third and final type is virtual warehouse. A virtual data warehouse is a collection of independent databases that may be queried collectively, giving users access to all the data as if it were kept in a single data warehouse (Panoply, n.d.).

Data warehouse can connect its various sources in order to facilitate wise decision-making and perform analysis, reporting, and business intelligence. Data warehouse is used across different sectors and industries where good decisions are crucial. Some of the common applications of data warehouse include the financial and banking to maximise profits, education to track students' performance, healthcare to use resources optimally, manufacturing to manage costs, insurance to analyse existing and potential clients, and finally services to maintaining customer details, financial records, and resources (Data Channel, 2020).

According to (Watson, Goodhue, & Wixom, 2002), data warehouse applications have many tangible and intangible benefits. These include benefits that range from time savings to useful information to increased market share achieved by an organisation. The use of data warehousing solutions to rethink business processes and support strategic business goals results in the best advantages. These advantages entail more substantial adjustments to the organization's operations. A data warehouse may be a key facilitator for a significant strategic change within a business. This would entail reengineering a specific business process, but it would also most likely require a coordinated group of modified processes that all supported the new organisational strategy.

However, there are some challenges within data warehousing. One of which as stated by is security and performance issues. Data in a data warehouse is a vital asset for an organisation. There are numerous common encryption algorithms that can be used to secure data warehouses. However, as a result of extra computation work, the performance of the data warehouse system will suffer. The degree of anonymity that may be offered must therefore be balanced against the impact on performance and scalability needs. Due to this trade-off, new security algorithms/methods are needed in order to provide the desired level of security with the least amount of performance drop (Chandra & Gupta, 2017).

To conclude, data warehousing has become an important technology because it can efficiently store, manage and analyse huge amounts of data. Data warehousing is expanding in many applications. However, there are challenges in size, speed and distributed operation when implementing data warehouses. Therefore, organisations need to decide on a trade-off between efficiency and speed. There are still many areas in which data warehousing can be improved and that would depend on future research.

2.2. Online Analytical Processing (OLAP)

According to the OLAP council, OLAP is a class of software technology that allows analysts, managers, and executives to quickly, consistently, interactively access information in a wide range of information views that have been transformed from raw data to reflect the true dimensionality of the enterprise as understood by the user (Chandra & Gupta, 2017).

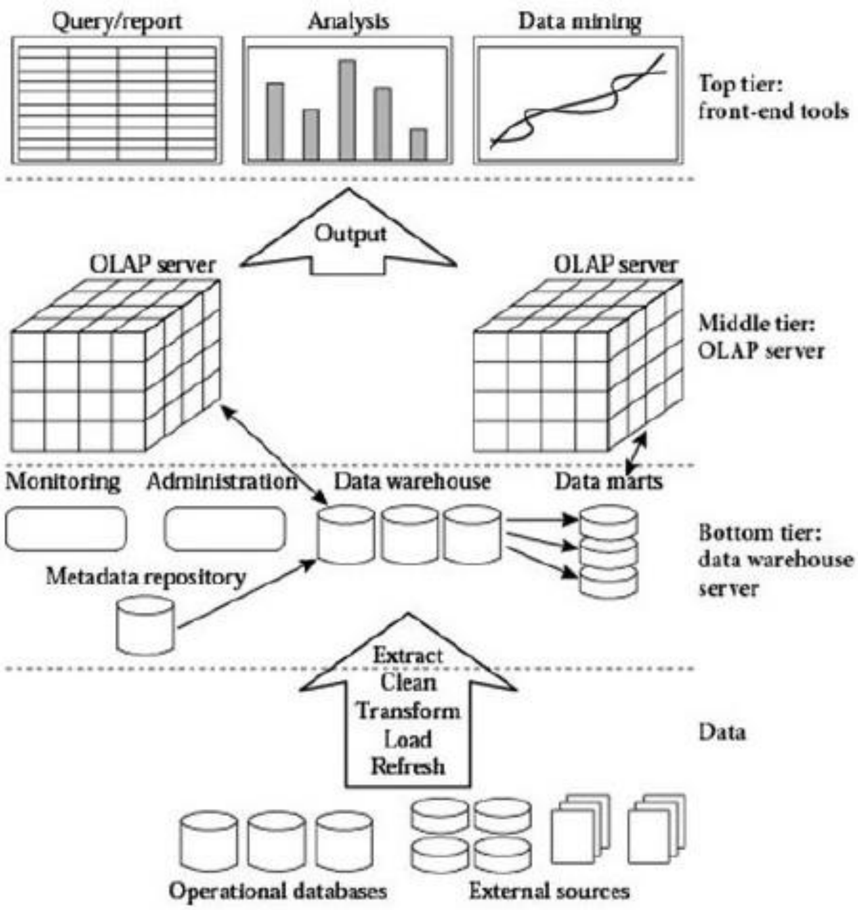

Figure 2.1: OLAP Architecture

Based on the OLAP architecture as shown in Figure 2.1, an OLAP consists of 4 stages (Nanda, Gupta, & Vijrania, 2019). The first stage is the process of obtaining data from various sources such as operational databases, relation tables, flat files, and other external sources. The data obtained is usually heterogenous in nature and structured differently such as structured, unstructured or semi-structured data. Thus, the data undergoes ETL which stands for extract, transform and load in order to organise and arrange the data for further processing.

The second stage, known as the Bottom Tier, comprises of data warehouse, data marts, and metadata. It also carries out administrative and monitoring tasks. Here, the data is set up for processing. Then, the third stage is Middle Tier which is the location of all transactions. It is made up of OLAP servers, which are in charge of carrying out different operations on the multidimensional data cubes. Finally, the last stage, Top Tier, is used for analysis, reporting, summarising, visualisation, data mining and other applications performed by users.

In this domain, OLAP helps to display geodata in multidimensional format. It helps to uncover patterns and correlations between events, locations, dates, and power plants. Since real-world events, geopolitical tensions, and economical factors all play a role in influencing the supply and demand of power generated by nuclear power plants, OLAP can display information in a way such that users can see patterns, perform analysis, and facilitate research decisions.

2.3. Snowflake Schema

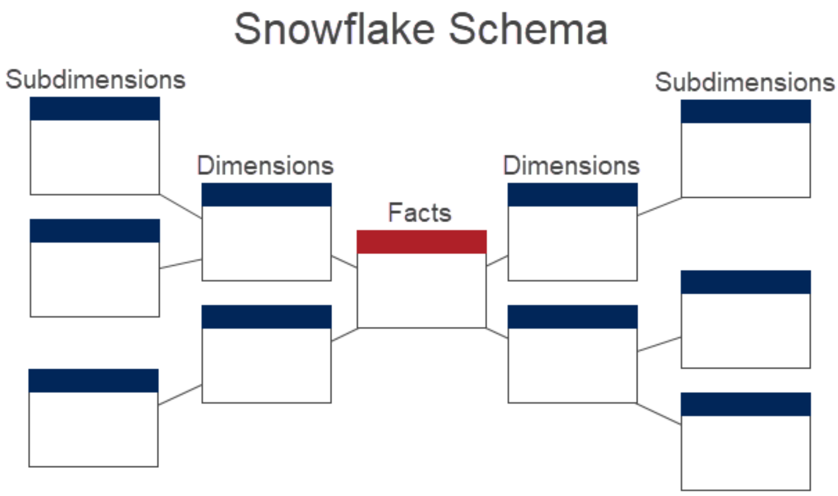

Figure 2.2: Snowflake Schema

Figure 2.2 above depicts a snowflake schema (Dancuk, 2021). It consists of three types of tables, a fact table in the middle, dimension tables, and subdimension tables. A fact table consists of several dimension tables. Sub dimension tables are derived normalizing dimension tables. Snowflake schema prevents data redundancy, and retain atomic values. As there is less redundancy, therefore less memory is required which leads to lower storage space and costs. However, there are several drawbacks of this schema. Due to its higher number of tables, querying data can be complex as it now requires linking from multiple dimension tables with multiple sub dimension tables (Karmani, et al., 2020).

Snowflake schema appropriate in this domain when there are many lookup tables. Lookup tables contain a fixed list of data. Such data do not change very often as they serve as reference for the operational tables. In order to prevent denormalisation which only increases data redundancy, snowflake schema allows lookup tables to retain their original form as subdimension tables. This however comes at a decrease in performance; hence a trade-off is made for saving storage and memory space at the expense of reduced performance.

2.4. MySQL Workbench

Figure 2.3: MySQL Workbench

Figure 2.3 shows MySQL Workbench which is a unified visual tool meant for database architects, developers, and database administrators. It provides data modelling, SQL creation, and extensive administrative tools for server configuration, user management, backup, and other database-related tasks. It is appropriate in this domain because it can be used to analyse the original database such as performing reverse engineering to obtain ERD diagram, and manage relationships between tables (Oracle, n.d.). In this context, it would be used to create the data warehouse including the fact table, dimension tables, and subdimension tables, and finally to write and test OLAP queries.

2.5. Visual Paradigm

Figure 2.4: Visual Paradigm

Figure 2.4 shows Visual Paradigm which is a software tool to perform UML Modelling, requirements management, business process modelling, data modelling ERD and Object Relational Mapping Diagrams (ORMD). Visual Paradigm serves as a guide for many stages of the software development lifecycle, including the compilation of reports during system analysis, software engineering, and programme development (Wikipedia, n.d.). It is appropriate in this assignment to design the schema for data warehouse.

Chapter 3: Methodology

3.1. Database Schemas and ERD

3.1.1. Original ERD for Operational Database

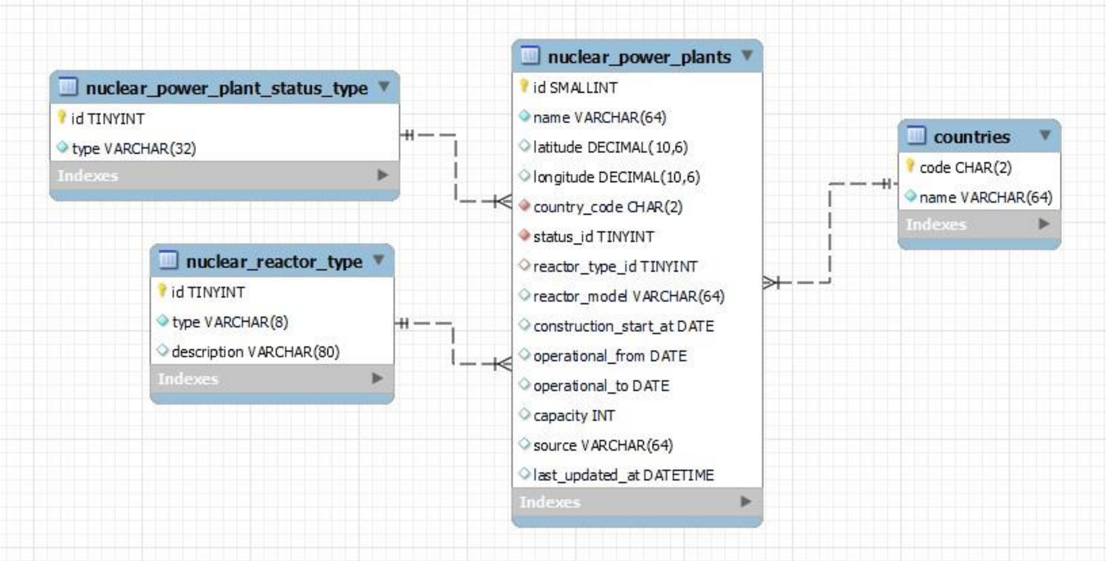

Figure 3.1: Nuclear Power Plant Geodata Schema

Figure 3.1 above shows the entity relationship diagram (ERD) for nuclear power plants geodata. The database is derived from datasets uploaded to Kaggle by Jonathan Bouchet in 2019. The datasets provide information about nuclear power plants globally. The main sources of data are from WNA/IAEA and Wikipedia.

There are 4 tables in the relational database. The main table is `nuclear_power_plants`. It has 14 columns to represent a nuclear power plant entity. The first column is `id` which acts as the primary key. The second column is name which tells the nuclear power plant name. The following columns, `latitude`, `longitude`, and `country_code`, provide information about the location of the nuclear power plant. `country_code` acts as a foreign key to `countries` table. `countries` table has 255 rows and each row provides the country code in ISO 3166-1 alpha-2 and the English name of the country.

The next column is status_id which also acts as a foreign key to nuclear_power_plant_status_type table and it is used to tell the status of the power plant, whether it is operational, shutdown, etc. There are 10 status types available in the table. The third and final foreign key is reactor_type_id that tells the reactor type of the power plant. The next column is reactor_model which describes the nuclear reactor model. There are 25 reactor types in the table.

Furthermore, construction_start_at is the date of which the nuclear power plant construction was started, operational_from is the date of which the nuclear power plant became operational (commercial operation date), and operational_to is the date of which the nuclear power plant was shut down (permanent shutdown date). The most important column is the capacity which tells the design net capacity in MWe. Design net capacity is the unit electrical output after deducting the self-consumption power to operate the power plant. The final two columns source and last_updated. They do not serve much purpose as they are meant for maintenance purpose only.

3.1.2. Constructed Data Warehouse Schema

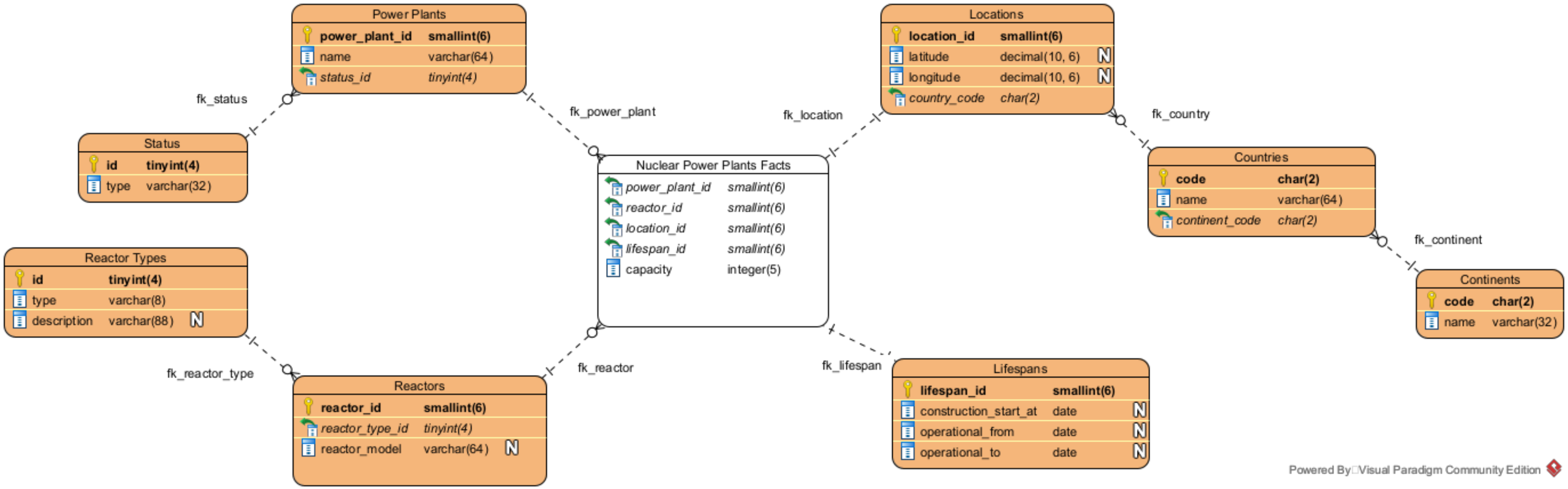

Figure 3.2: Proposed Data Warehouse Schema

Figure 3.2 depicts the proposed data warehouse schema that follows the snowflake schema. The fact table is `nuclear_power_plants_facts`. It has 4 foreign keys, each links to a dimension table. There is only measure which is capacity.

There are 4 dimension tables. The first dimension table is `power_plants` with `power_plant_id` as its primary key integer, `name` that tells the name of the power plant, and foreign key `status_id` to link to a subdimension table called `status`. `status` stores the 10 status types as in the original database. The second dimension table is `reactors` with `reactor_id` as its primary key integer, foreign key `reactor_type_id` to link to `reactor_types`, and `reactor_model` that tells the model name of the power plant. `reactor_types` store 25 reactor type information as in the original database. Next is locations with `location_id` as its integer primary key, `latitude` and `longitude` for the coordinates of the power plant, and `country_code` to link to countries table. `countries` contain the 255 country information as in the original database, and now it has a new foreign key `continent_code` which links to `continents` table. `continents` contain all 7 continents, each with a code and a name. Overall, there are 1 fact table, 4 dimension tables, and 4 subdimension tables in this schema.

3.2. Data Cube Representations

3.2.1. Information Package Diagram

Table 3.1: Information Package Diagram

Power Plant	Reactor	Location	Lifespan
power_plant_id	***reactor_id***	***location_id***	***lifespan_id***
status	reactor type	continent	era
name	reactor model	country	decade
		coordinates	year
			date
Metrics: Capacity			

Table 3.1 above outlines the hierarchal attributes for each dimension table. For `power_plant`, a power plant has a unique name and a status. Many power plants can be assigned the same status. For `reactor`, a power plant can be categories based on the type and model. For `locations`, a pair of coordinates belongs to a country which in turn belongs to a continent. For `lifespan`, an era can have several decades, a decade has 10 years, and a year has 365 or 366 dates.

3.3. Lattice of Cuboid

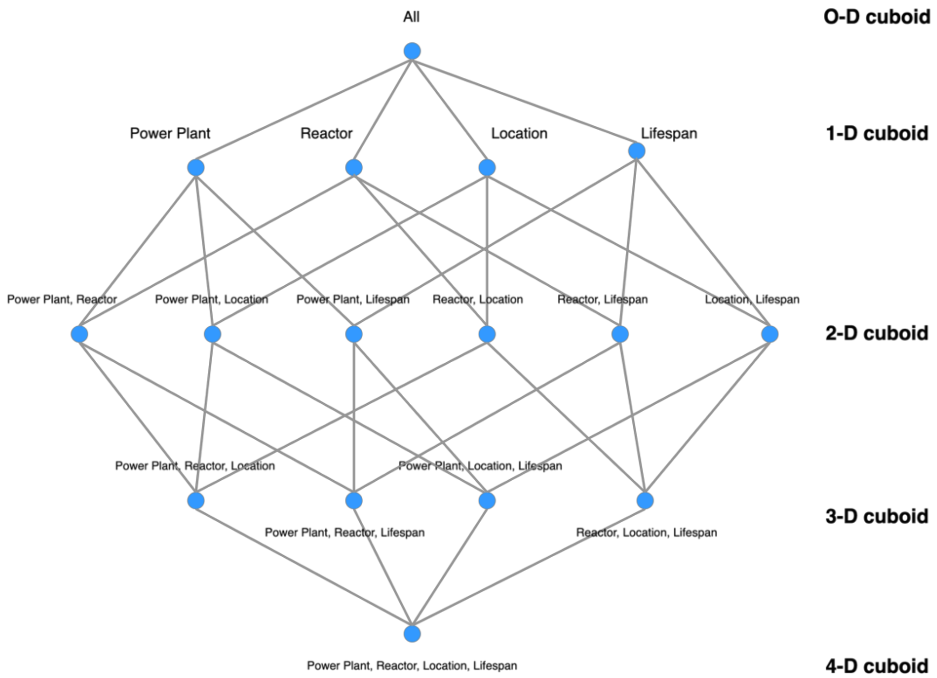

Figure 3.3: Lattice of Cuboid

3.4. Hierarchical Structures

For `power_plant`, power plants can be grouped based on their status such as operational, shutdown, planned, etc. Then, it can be drilled down into their respective name. Every power plant has a unique name, no two plants have the same name. Thus, name belongs to the bottom of the hierarchy.

For `reactor`, each nuclear power plant uses a single type of reactor. Therefore, multiple reactors can be grouped based on the same type of reactor. Though two reactors may have the same reactor type, they still have distinct models. A reactor is closely related to geographic data such as proximity to water sources, terrain, weather events, etc. Thus, geography influences the type of reactor being installed for a power plant.

For `locations`, continent encompasses a number of countries, and each pair of coordinates is located in a country. Countries have different capabilities and resources, as well as demands for electricity. Historically, nuclear energy is concentrated in wealthy countries in North America and Europe as they had the best infrastructure, technology and investments to build nuclear power plants. Developing nations in Asia and Africa start to pick up nuclear energy as their citizens and technology grow. Hence, continents and countries can serve as good parameters when analysing patterns over the years.

For `lifespan`, the attributes concern the dates of which a nuclear power plant began construction, started operation, and ended operation. These dates can be grouped accordingly. An era represents a period in time. One notable example is Cold War era from 12th May 1947 to 26th December 1991. This era saw an increase of nuclear power plants being constructed as the United States and the Soviet Union and their respective allies engaged themselves in the nuclear arms race. Then, an era may consist of several decades. Every decade undergoes different events and growths. And finally, a year can have 365/366 different dates, and every day is significant such that a new nuclear power plant can be planned or began construction.

Chapter 4: Discussion and Analysis

4.1. Roll-up, or Drill-up

4.1.1. Every Continent

Table 4.1: Rollup Query #1

Purpose	Display number of operational nuclear power plants and their capacity for every continent based on the year they were constructed.
Dimensions	Power Plants, Lifespans, and Locations
# of Columns	4
Conditions (if applicable)	Status is operational
Justifications	OLAP can find out how many power plants that are still operational today built by every continent in a given year.

SQL Command:

```sql
SELECT
con.`name` AS 'Continent',
YEAR(lif.`construction_start_at`) AS 'Construction Year',
COUNT(fact.`location_id`) AS '# Total',
SUM(fact.`capacity`) AS 'Total Capacity (in MWe)'
FROM `nuclear_power_plant_facts` fact
JOIN `locations` loc ON fact.`location_id` = loc.`location_id`
JOIN `countries` ctr ON loc.`country_code` = ctr.`code`
JOIN `continents` con ON ctr.`continent_code` = con.`code`
JOIN `power_plants` pow ON fact.`power_plant_id` = pow.`power_plant_id`
JOIN `status` sta ON pow.`status_id` = sta.`id`
JOIN `lifespans` lif ON fact.`lifespan_id` = lif.`lifespan_id`
WHERE sta.`type` = 'Operational'
GROUP BY con.`code`, YEAR(lif.`construction_start_at`) WITH ROLLUP;
```

Output:

Continent	Construction Year	# Total	Total Capacity (in MWe)
Africa	1976	2	1842
Africa	NULL	2	1842
Asia	1964	2	400
Asia	1965	1	207
Asia	1966	1	125
Asia	1968	1	207
Asia	1970	1	780
Asia	1971	2	982
Asia	1972	2	982
Asia	1973	2	1660

Figure 4.1: Rollup Query #1 Output

Explanation: 104 row(s) returned. From the output, we know that two nuclear power plants were constructed in Africa with a total capacity of 1842 MWe. Africa only constructed two plants over the years, hence the rollup results are the same as the previous values. The other continents totals are as followed: Asia built 156 plants with a total capacity of 127537 MWe, Europe built 182 plants with a total capacity of 158356 MWe, North America built 119 plants with a total capacity of 108906 MWe, and finally South America built 5 plants with a total capacity of 3482 MWe. The global total stands at 464 with a combined capacity of 400123 MWe.

4.1.2. Cold War and Post-Cold War Era

Table 4.2: Rollup Query #2

Purpose	Display number of nuclear power plants and their total capacity based on country and reactor type that began construction during and after Cold War era.
Dimensions	Reactors, Locations and Lifespans
# of Columns	5
Conditions (if applicable)	Construction began after the start of Cold War (12th March 1947)
Justifications	Cold War is a period in which the United States, Soviet Union, and their respective allies race to achieve dominance in the nuclear power. They engage in nuclear arms race for supremacy in nuclear warfare where they build bigger and more powerful nuclear bombs and weapons. Therefore, OLAP can be used to find the correlation between this event and the number of nuclear power plants being constructed by every country.

SQL Command:

```
SELECT
IF (
    lif.`construction_start_at` > '1991-12-26',
    'Post-Cold War',
    'Cold War'
) AS 'Era',
ctr.`name` AS 'Country',
rty.`description` AS 'Reactor Type',
COUNT(fact.`location_id`) AS '# Total',
SUM(fact.`capacity`) AS 'Total Capacity (in MWe)'
FROM `nuclear_power_plant_facts` fact
JOIN `reactors` rea ON fact.`reactor_id` = rea.`reactor_id`
JOIN `reactor_types` rty ON rea.`reactor_type_id` = rty.`id`
JOIN `locations` loc ON fact.`location_id` = loc.`location_id`
JOIN `countries` ctr ON loc.`country_code` = ctr.`code`
JOIN `lifespans` lif ON fact.`lifespan_id` = lif.`lifespan_id`
WHERE lif.`construction_start_at` > '1947-03-12'
GROUP BY rty.`type`, ctr.`code`,
ORDER BY lif.`construction_start_at`, ctr.`name`, rty.`description`;
```

Output:

	Era	Country	Reactor Type	# Total	Total Capacity (in MWe)
▸	Cold War	Russian Federation	Light Water Graphite Reactor	18	10472
	Cold War	United Kingdom of Great Britain and Northern Ireland	Fast Breeder Reactor	2	248
	Cold War	United States	Fast Breeder Reactor	1	60
	Cold War	United Kingdom of Great Britain and Northern Ireland	Gas-Cooled Reactor	41	13697
	Cold War	Belgium	Pressurised Water Reactor	8	5475
	Cold War	Sweden	Pressurised Heavy Water Reactor	1	9
	Cold War	Slovakia	Heavy Water Gas Cooled Reactor	1	110
	Cold War	Italy	Gas-Cooled Reactor	1	200
	Cold War	United States	Organic Cooled Reactor	2	86

Figure 4.2: Rollup Query #2 Output

Explanation: 383 row(s) returned. The output tells the number of power plants with their total capacity built by a country and grouped by a reactor type for a given era. For example, Russia built 18 plants of reactor type Light Water Graphite Reactor during Cold War Era.

4.2. Roll-down, or Drill-down

4.2.1. Every Country

Table 4.3: Rolldown Query #1

Purpose	Display the construction year, number of operational nuclear power plants and their total capacity for every country with at least a nuclear power plant based on the year they were constructed.
Dimensions	Power Plants, Reactors and Locations
# of Columns	4
Conditions (if applicable)	Status is operational
Justifications	OLAP can find out how many power plants that are still operational today built by every country in a given year.

SQL Command:

```sql
SELECT
ctr.`name` AS 'Country',
YEAR(lif.`construction_start_at`) AS 'Constructed in',
COUNT(fact.`location_id`) AS '# Total',
SUM(fact.`capacity`) AS 'Total Capacity (in MWe)'
FROM `nuclear_power_plant_facts` fact
JOIN `locations` loc ON fact.`location_id` = loc.`location_id`
JOIN `countries` ctr ON loc.`country_code` = ctr.`code`
JOIN `continents` con ON ctr.`continent_code` = con.`code`
JOIN `power_plants` pow ON fact.`power_plant_id` = pow.`power_plant_id`
JOIN `status` sta ON pow.`status_id` = sta.`id`
JOIN `lifespans` lif ON fact.`lifespan_id` = lif.`lifespan_id`
WHERE sta.`type` = 'Operational'
GROUP BY ctr.`code`, YEAR(lif.`construction_start_at`) WITH ROLLUP;
```

Output:

Country	Constructed in	# Total	Total Capacity (in MWe)
Armenia	1975	1	375
Armenia	NULL	1	375
Argentina	1968	1	319
Argentina	1974	1	600
Argentina	1981	1	692
Argentina	NULL	3	1611
Belgium	1969	1	392
Belgium	1970	1	870
Belgium	1971	1	392
Belgium	1975	1	890

Figure 4.3: Rolldown Query #1 Output

Explanation: 244 row(s) returned. The output tells the number of plants and total capacity built by a country in a given year, then it sums up the total over the years. For example, Argentina built a plant in 1968, 1974, and 1981 with total capacity of 319 MWe, 600, and 692 respectively. Its grand total comes at 3 plants and 1611 MWe combined capacity.

4.2.2. Every Year after Cold War

Table 4.4: Rolldown Query #2

Purpose	Display number of nuclear power plants and their total capacity based on country and reactor type for every year after Cold War.
Dimensions	Reactors, Locations and Lifespans
# of Columns	5
Conditions (if applicable)	Construction began after the start of Cold War (12th March 1947)
Justifications	There were significant events during the Cold War such as Korean War, Vietnam War, Cuban Missile Crisis, formation of NATO, fall of Berlin Wall, etc., as well as after Cold War ended when the Soviet Union collapsed. Therefore, OLAP drills down based on year to find out more patterns and correlations.

SQL Command:

```
SELECT
YEAR(lif.`construction_start_at`) AS 'Construction Year',
ctr.`name` AS 'Country',
rty.`description` AS 'Reactor Type',
COUNT(fact.`location_id`) AS '# Total',
SUM(fact.`capacity`) AS 'Total Capacity (in MWe)'
FROM `nuclear_power_plant_facts` fact
JOIN `reactors` rea ON fact.`reactor_id` = rea.`reactor_id`
JOIN `reactor_types` rty ON rea.`reactor_type_id` = rty.`id`
JOIN `locations` loc ON fact.`location_id` = loc.`location_id`
JOIN `countries` ctr ON loc.`country_code` = ctr.`code`
JOIN `lifespans` lif ON fact.`lifespan_id` = lif.`lifespan_id`
WHERE lif.`construction_start_at` > '1947-03-12'
GROUP BY YEAR(lif.`construction_start_at`), rty.`type`, ctr.`code`
ORDER BY lif.`construction_start_at`, rty.`description`, ctr.`name`;
```

Output:

Construction Year	Country	Reactor Type	# Total	Total Capacity (in MWe)
1951	Russian Federation	Light Water Graphite Reactor	1	5
1953	United Kingdom of Great Britain and Northern Ireland	Gas-Cooled Reactor	2	70
1954	United States	Pressurised Water Reactor	1	60
1955	United Kingdom of Great Britain and Northern Ireland	Fast Breeder Reactor	1	14
1955	France	Gas-Cooled Reactor	1	36
1955	United Kingdom of Great Britain and Northern Ireland	Gas-Cooled Reactor	6	210
1956	France	Gas-Cooled Reactor	1	36
1956	United States	Boiling Water Reactor	2	216
1956	United States	Pressurised Water Reactor	1	265
1956	United States	Fast Breeder Reactor	1	60

Figure 4.4: Rolldown Query #2 Output

Explanation: 383 row(s) returned. The output tells the number of power plants with their total capacity built by a country and grouped by a reactor type for a given year. For example, Russia built 1 plant of reactor type Light Water Graphite Reactor in 1951.

4.3. Slicing

4.3.1. The United States

Table 4.5: Slicing Query #1

Purpose	List all nuclear power plants in the United States.
Dimensions	Power Plants, Locations, and Lifespans
# of Columns	7
Conditions (if applicable)	Country is the United States
Justifications	The United States leads the world with more than 30% of all nuclear electricity generated worldwide. In 2019, the nation's nuclear reactors generated 843 billion kWh, or 19% of all electrical output (World Nuclear Association, 2022). OLAP can slice all records about plants built in the United States and display the useful information.

SQL Command:

```sql
SELECT
pow.`name` AS 'US Nuclear Power Plant',
fact.`capacity` AS 'Capacity (in MWe)',
sta.`type` AS 'Status',
IF(
    lif.`construction_start_at` IS NOT NULL,
    YEAR(lif.`construction_start_at`), '-'
) AS 'Constructed in',
CONCAT(
    IF(lif.`operational_from` IS NOT NULL, YEAR(lif.`operational_from`), 'n.d'),
    ' - ',
    IF(
        lif.`operational_to` IS NOT NULL,
        YEAR(lif.`operational_from`),
        IF(sta.`type` = 'Operational', 'Present', 'n.d')
    )
) AS 'Operation',
loc.`latitude` AS 'Latitude',
loc.`longitude` AS 'Longitude'
FROM `nuclear_power_plant_facts` fact
JOIN `locations` loc ON fact.`location_id` = loc.`location_id`
JOIN `countries` ctr ON loc.`country_code` = ctr.`code`
JOIN `power_plants` pow ON fact.`power_plant_id` = pow.`power_plant_id`
JOIN `status` sta ON pow.`status_id` = sta.`id`
JOIN `lifespans` lif ON fact.`lifespan_id` = lif.`lifespan_id`
WHERE ctr.`name` = 'United States'
ORDER BY sta.`type`;
```

Output:

US Nuclear Power Plant	Capacity (in MWe)	Status	Constructed in	Operation	Latitude	Longitude
Arkansas Nuclear One-1 (ANO-1)	850	Operational	1968	1974 - Present	35.310000	-93.230000
Arkansas Nuclear One-2 (ANO-2)	912	Operational	1968	1980 - Present	35.310000	-93.229000
Beaver Valley-1	835	Operational	1970	1976 - Present	40.624000	-80.432000
Beaver Valley-2	836	Operational	1974	1987 - Present	40.624000	-80.432000
Braidwood-1	1120	Operational	1975	1988 - Present	41.247000	-88.230000
Braidwood-2	1120	Operational	1975	1988 - Present	41.247000	-88.230000
Browns Ferry-1	1065	Operational	1967	1974 - Present	34.704000	-87.119000
Browns Ferry-2	1065	Operational	1967	1975 - Present	34.704000	-87.119000
Browns Ferry-3	1065	Operational	1968	1977 - Present	34.704000	-87.119000
Brunswick-1	821	Operational	1970	1977 - Present	33.958000	-78.007000

Figure 4.5: Slicing Query #1 Output

Explanation: 137 row(s) returned. The output lists the all the power plants in the US. It provides some details such as current status, the year it was built, the operation period, and the coordinates.

4.3.2. Pressurised Water Reactor

Table 4.6: Slicing Query #1

Purpose	Display all nuclear power plants with reactor type of Pressurised Water Reactor
Dimensions	Power Plants, Reactors and Locations
# of Columns	6
Conditions (if applicable)	Reactor type is Pressurised Water Reactor
Justifications	Pressurised water reactor (PWR) is the most common type of nuclear reactor with over 300 reactors currently operational to generate power, and several hundreds more used for naval propulsion (World Nuclear Association, 2022). OLAP can slice all the power plants of reactor type Pressurised Water Reactor and derive useful information such as which country favours it the most, and how many are still operational today and how much they contribute to the energy supply.

SQL Command:

```sql
SELECT
pow.`name` AS 'Nuclear Power Plant',
sta.`type` AS 'Status',
fact.`capacity` AS 'Capacity (in MWe)',
loc.`latitude` AS 'Latitude',
loc.`longitude` AS 'Longitude',
ctr.`name` AS 'Country'
FROM `nuclear_power_plant_facts` fact
JOIN `power_plants` pow ON fact.`power_plant_id` = pow.`power_plant_id`
JOIN `status` sta ON pow.`status_id` = sta.`id`
JOIN `reactors` rea ON fact.`reactor_id` = rea.`reactor_id`
JOIN `reactor_types` rty ON rea.`reactor_type_id` = rty.`id`
JOIN `locations` loc ON fact.`location_id` = loc.`location_id`
JOIN `countries` ctr ON loc.`country_code` = ctr.`code`
WHERE rty.`type` = 'PWR'
ORDER BY fact.`capacity`;
```

Output:

	Nuclear Power Plant	Status	Capacity (in MWe)	Latitude	Longitude	Country
▶	Saxton	Shutdown	3	40.226944	-78.241944	United States
	BR-3	Shutdown	11	51.217000	5.099000	Belgium
	CAREM25	Under Construction	25	-33.966667	-59.211111	Argentina
	Akademik Lomonosov-1 (Vilyuchinsk)	Under Construction	32	59.919000	30.261000	Russian Federation
	Akademik Lomonosov-2 (Vilyuchinsk)	Under Construction	32	59.919000	30.261000	Russian Federation
	Shippingport	Shutdown	60	40.630000	-80.417778	United States
	Rheinsberg	Shutdown	62	53.147000	12.991000	Germany
	Jose Cabrera-1 (Zorita)	Shutdown	153	40.362000	-2.818000	Spain
	Yankee Rowe	Shutdown	175	42.728000	-72.929000	United States
	Novovoronezh-1	Shutdown	197	51.275000	39.206000	Russian Federation

Figure 4.6: Slicing Query #1 Output

Explanation: 415 row(s) returned. The output lists down all nuclear power plants of reactor type PWR. For example, Saxton is a PWR plant built in the US. It is currently shutdown and used to generate 3 MWe of power. The table sorts the plants based on the capacity from the lowest to highest.

4.4. Dicing

4.4.1. Decommissioned Plants in Post-Soviet Nations

Table 4.7: Dicing Query #1

Purpose	List of all nuclear power plants that are either shutdown or suspended operation in post-Soviet nations.
Dimensions	Power Plants, Locations, and Lifespans
# of Columns	7
Conditions (if applicable)	1. Country is a post-Soviet nation (Russian Federation, Ukraine, Georgia, Belarus, Uzbekistan, Armenia, Azerbaijan, Kazakhstan, Kyrgyzstan, Moldova, Turkmenistan, Tajikistan, or Latvia), and 2. status is either shutdown or suspended operation

Justifications	After Soviet Union collapsed in 25th December 1991, 15 new republics were formed. This period saw changes in regimes, introduction of democracy, as well as many political, economic, and social changes. Many of these newly formed republics began to phase out Soviet-era plants in favour of other cheaper energy alternatives (Wikipedia, n.d.). Therefore, OLAP can find out the fate of these decommissioned plants and display information about them.

SQL Command:

```
SELECT
pow.`name` AS 'Nuclear Power Plant',
ctr.`name` AS 'Country',
sta.`type` AS 'Status',
fact.`capacity` AS 'Achieved Capacity (in MWe)',
lif.`operational_from` AS "Started Operation",
lif.`operational_to` AS "Shutdown / Suspended Operation",
TIMESTAMPDIFF(YEAR, lif.`operational_from`, lif.`operational_to`) AS
"Total Operational Years"
FROM `nuclear_power_plant_facts` fact
JOIN `locations` loc ON fact.`location_id` = loc.`location_id`
JOIN `countries` ctr ON loc.`country_code` = ctr.`code`
JOIN `power_plants` pow ON fact.`power_plant_id` = pow.`power_plant_id`
JOIN `status` sta ON pow.`status_id` = sta.`id`
JOIN `lifespans` lif ON fact.`lifespan_id` = lif.`lifespan_id`
WHERE ctr.`name` IN ('Russian Federation', 'Ukraine', 'Georgia',
'Belarus', 'Uzbekistan', 'Armenia', 'Azerbaijan', 'Kazakhstan',
'Kyrgyzstan', 'Moldova, Republic of', 'Turkmenistan', 'Tajikistan',
'Latvia')
AND (sta.`type` = 'Shutdown' OR sta.`type` = 'Suspended Operation')
ORDER BY ctr.`name`, sta.`type`;
```

Output:

Nuclear Power Plant	Country	Status	Achieved Capacity (in MWe)	Started Operation	Shutdown / Suspended Operation	Total Operational Years
Armenia-1 (Armenian-1 / Metsamor)	Armenia	Shutdown	376	1977-10-06	1989-02-25	11
Aktau (Shevchenko)	Kazakhstan	Shutdown	135	1973-07-16	1999-04-22	25
APS-1 Obninsk	Russian Federation	Shutdown	5	1954-12-01	2002-04-29	47
Beloyarsk-1	Russian Federation	Shutdown	102	1964-04-26	1983-01-01	18
Beloyarsk-2	Russian Federation	Shutdown	146	1969-12-01	1990-01-01	20
Bilibino-1	Russian Federation	Shutdown	11	1974-04-01	2019-01-14	44
Leningrad-1	Russian Federation	Shutdown	925	1974-11-01	2018-12-22	44
Novovoronezh-1	Russian Federation	Shutdown	197	1964-12-31	1988-02-16	23
Novovoronezh-2	Russian Federation	Shutdown	336	1970-04-14	1990-08-29	20
Novovoronezh-3	Russian Federation	Shutdown	385	1972-06-29	2016-12-25	44
Chernobyl-1	Ukraine	Shutdown	925	1978-05-27	1996-11-30	18
Chernobyl-2	Ukraine	Shutdown	925	1979-05-28	1991-10-11	12
Chernobyl-3	Ukraine	Shutdown	925	1982-06-08	2000-12-15	18
Chernobyl-4	Ukraine	Shutdown	925	1984-03-26	1986-04-26	2

Figure 4.7: Dicing Query #1 Output

Explanation: 14 row(s) returned. The output returns all 14 power plants that have been decommissioned in post-Soviet nations. The notable records include the final 4 rows which show the details of Chernobyl power plants. There were 4 power plants in Chernobyl before the world's worst nuclear accident happened in 26th April 1986.

4.4.2. Operational Plants in China after 2000s

Table 4.8: Dicing Query #2

Purpose	List of operational nuclear power plants that started operating after 2000 in People's Republic of China.
Dimensions	Power Plants, Locations, and Lifespans
# of Columns	6
Conditions (if applicable)	1. Status is operational, and 2. country is People's Republic of China, and 3. started operating after 15th December 1991
Justifications	People's Republic of China, or more commonly known as just China, independently built and developed its first nuclear power plant called Qinshan Nuclear Power Plant in 1984, and successfully connected to the power grid in 15th December 1991 (Shanghai Nuclear Office, n.d.). As China's economy boomed, the demand for electricity by its mammoth citizens grew. Therefore, OLAP can be used to analyse the nuclear power plants being built to match this growing demand.

SQL Command:

```sql
SELECT
pow.`name` AS 'Nuclear Power Plant',
rty.`description` AS 'Reactor Type',
loc.`latitude` AS 'Latitude',
loc.`longitude` AS 'Longitude',
fact.`capacity` AS 'Capacity (in MWe)',
lif.`operational_from` AS 'Started Operating'
FROM `nuclear_power_plant_facts` fact
JOIN `power_plants` pow ON fact.`power_plant_id` = pow.`power_plant_id`
JOIN `status` sta ON pow.`status_id` = sta.`id`
JOIN `locations` loc ON fact.`location_id` = loc.`location_id`
JOIN `countries` ctr ON loc.`country_code` = ctr.`code`
JOIN `lifespans` lif ON fact.`lifespan_id` = lif.`lifespan_id`
JOIN `reactors` rea ON fact.`reactor_id` = rea.`reactor_id`
JOIN `reactor_types` rty ON rea.`reactor_type_id` = rty.`id`
WHERE sta.`type` = 'Operational' AND ctr.`name` = "People's Republic of China"
AND lif.`operational_from` > '1991-12-15'
ORDER BY lif.`operational_from`;
```

Output:

Nuclear Power Plant	Reactor Type	Latitude	Longitude	Capacity (in MWe)	Started Operating
Qinshan-2-1 (Qinshan 2)	Pressurised Water Reactor	30.440000	120.950000	610	2002-04-15
Lingao-1	Pressurised Water Reactor	22.606000	114.550000	950	2002-05-28
Qinshan-3-1 (Qinshan 6)	Pressurised Heavy Water Reactor	30.440000	120.950000	677	2002-12-31
Lingao-2	Pressurised Water Reactor	22.606000	114.550000	950	2003-01-08
Qinshan-3-2 (Qinshan 7)	Pressurised Heavy Water Reactor	30.440000	120.950000	677	2003-07-24
Qinshan-2-2 (Qinshan 3)	Pressurised Water Reactor	30.440000	120.950000	610	2004-05-03
Tianwan-1	Pressurised Water Reactor	34.690000	119.456000	990	2007-05-17
Tianwan-2	Pressurised Water Reactor	34.690000	119.456000	990	2007-08-16
Lingao-3	Pressurised Water Reactor	22.609000	114.552000	1007	2010-09-15
Qinshan-2-3 (Qinshan 4)	Pressurised Water Reactor	30.440000	120.950000	619	2010-10-05

Figure 4.8: Dicing Query #2 Output

Explanation: 42 row(s) returned. There are a total of 42 power plants built by China. One notable pattern derived from the output is China highly favours pressurised water as the reactor type. All expect two plants use Pressurised Water Reactor, while the other two use Pressurised Heavy Water Reactor.

4.5. Drill-across

There are no drill-across operations that can be performed on the data warehouse. Drill across entails making queries from two or more fact tables in which headers for every query are identical to the chosen attributes. The data warehouse only has one fact table and its data is fairly limited to develop more fact tables. Therefore, drill-across operations cannot be performed.

4.6. Pivoting

4.6.1. Status against Reactor Type

Table 4.9: Pivoting #1 (1/2)

Purpose	Display number of nuclear power plants and total capacity for every status based on reactor type.
Dimensions	Power Plants and Reactors
# of Columns	13
Conditions (If applicable)	-
Justifications	OLAP gives an overview of the all number of nuclear power plants based on their current status for a given reactor type.

SQL Command:

```sql
SELECT
rty.`description` AS 'Type of Reactor',
COUNT(CASE WHEN sta.`id` = 0 THEN 1 ELSE null END) AS 'Unknown',
COUNT(CASE WHEN sta.`id` = 1 THEN 1 ELSE null END) AS 'Planned',
COUNT(CASE WHEN sta.`id` = 2 THEN 1 ELSE null END) AS 'Under Construction',
COUNT(CASE WHEN sta.`id` = 3 THEN 1 ELSE null END) AS 'Operational',
COUNT(CASE WHEN sta.`id` = 4 THEN 1 ELSE null END) AS 'Suspended Operation',
COUNT(CASE WHEN sta.`id` = 5 THEN 1 ELSE null END) AS 'Shutdown',
COUNT(CASE WHEN sta.`id` = 6 THEN 1 ELSE null END) AS 'Unfinished',
COUNT(CASE WHEN sta.`id` = 7 THEN 1 ELSE null END) AS 'Never Built',
COUNT(CASE WHEN sta.`id` = 8 THEN 1 ELSE null END) AS 'Suspended Construction',
COUNT(CASE WHEN sta.`id` = 9 THEN 1 ELSE null END) AS 'Cancelled Construction',
COUNT(fact.`location_id`) AS '# Total',
SUM(fact.`capacity`) AS 'Total Capacity (in MWe)'
FROM `nuclear_power_plant_facts` fact
JOIN `power_plants` pow ON fact.`power_plant_id` = pow.`power_plant_id`
JOIN `status` sta ON pow.`status_id` = sta.`id`
JOIN `reactors` rea ON fact.`reactor_id` = rea.`reactor_id`
JOIN `reactor_types` rty ON rea.`reactor_type_id` = rty.`id`
GROUP BY rty.`id`;
```

Output:

Type of Reactor	Unknown	Planned	Under Construction	Operational	Suspended Operation	Shutdown	Unfinished	Never Built	Suspended Construction	Cancelled Construction	# Total	Total Capacity (in MWe)
Pressurised Heavy Water Reactor	0	0	4	49	0	8	0	0	0	0	61	29217
Pressurised Water Reactor	0	2	43	312	0	54	0	0	2	2	415	371984
Fast Breeder Reactor	0	0	1	3	0	8	0	0	0	0	12	4010
Light Water Graphite Reactor	0	0	0	13	0	11	0	0	0	0	24	17172
Boiling Water Reactor	0	0	4	73	0	42	0	0	0	0	119	92651
Gas-Cooled Reactor	0	0	0	14	0	38	0	0	0	0	52	16861
Heavy Water Gas Cooled Reactor	0	0	0	0	0	4	0	0	0	0	4	286
High-Temperature Gas-cooled Reactor	0	0	1	0	0	4	0	0	0	0	5	879
Heavy Water Light Water Reactor	0	0	0	0	0	2	0	0	0	0	2	398

Figure 4.9: Pivoting #1 Output (1/2)

Explanation: 11 row(s) returned. The output tallies each type of reactor based on status. Then, the total tally with the total capacity is displayed.

4.6.2. Reactor Type against Status

Table 4.10: Pivoting #1 (2/2)

Purpose	Display number of nuclear power plants and total capacity for every reactor type based on status.
Dimensions	Power Plants and Reactors
# of Columns	28
Conditions (If applicable)	-
Justifications	OLAP gives an overview of the all number of nuclear power plants based on their current status for a given reactor type.

SQL Command:

```
SELECT
sta.`type` AS 'Status',
COUNT(CASE WHEN rty.`id` = 1 THEN 1 ELSE null END) AS 'Advanced Boiling Water Reactor',
COUNT(CASE WHEN rty.`id` = 2 THEN 1 ELSE null END) AS 'Advanced Power Reactor',
COUNT(CASE WHEN rty.`id` = 3 THEN 1 ELSE null END) AS 'Advanced Pressurised Water Reactor',
COUNT(CASE WHEN rty.`id` = 4 THEN 1 ELSE null END) AS 'Advanced Gas-cooled Reactor',
COUNT(CASE WHEN rty.`id` = 5 THEN 1 ELSE null END) AS 'Boiling Water Reactor',
COUNT(CASE WHEN rty.`id` = 6 THEN 1 ELSE null END) AS 'Evolutionary Power Reactor',
COUNT(CASE WHEN rty.`id` = 7 THEN 1 ELSE null END) AS 'Fast Breeder Reactor',
COUNT(CASE WHEN rty.`id` = 8 THEN 1 ELSE null END) AS 'Gas-Cooled Reactor',
COUNT(CASE WHEN rty.`id` = 9 THEN 1 ELSE null END) AS 'High-Temperature Gas-cooled Reactor',
COUNT(CASE WHEN rty.`id` = 10 THEN 1 ELSE null END) AS 'High Temperature Reactor - Pebble Module',
COUNT(CASE WHEN rty.`id` = 11 THEN 1 ELSE null END) AS 'Heavy Water Gas Cooled Reactor',
COUNT(CASE WHEN rty.`id` = 12 THEN 1 ELSE null END) AS 'Heavy Water Light Water Reactor',
COUNT(CASE WHEN rty.`id` = 13 THEN 1 ELSE null END) AS 'Heavy Water Organic Cooled Reactor',
COUNT(CASE WHEN rty.`id` = 14 THEN 1 ELSE null END) AS 'Lead-cooled Fast Reactor',
COUNT(CASE WHEN rty.`id` = 15 THEN 1 ELSE null END) AS 'Liquid Metal Fast Breeder Reactor',
COUNT(CASE WHEN rty.`id` = 16 THEN 1 ELSE null END) AS 'Liquid Metal Fast Reactor',
COUNT(CASE WHEN rty.`id` = 17 THEN 1 ELSE null END) AS 'Light Water Graphite Reactor',
COUNT(CASE WHEN rty.`id` = 18 THEN 1 ELSE null END) AS 'Molten Salt Reactor',
COUNT(CASE WHEN rty.`id` = 19 THEN 1 ELSE null END) AS 'Organic Cooled Reactor',
COUNT(CASE WHEN rty.`id` = 20 THEN 1 ELSE null END) AS 'Pressurised Heavy Water Reactor',
COUNT(CASE WHEN rty.`id` = 21 THEN 1 ELSE null END) AS 'Pressurised Water Reactor',
COUNT(CASE WHEN rty.`id` = 22 THEN 1 ELSE null END) AS 'High Power Channel-Type Reactor (Reaktor Bolshoy Moshchnosti Kanalniy)',
COUNT(CASE WHEN rty.`id` = 23 THEN 1 ELSE null END) AS 'Sodium-Graphite Reactor',
COUNT(CASE WHEN rty.`id` = 24 THEN 1 ELSE null END) AS 'Steam Generating Heavy Water Reactor',
COUNT(CASE WHEN rty.`id` = 25 THEN 1 ELSE null END) AS 'Traveling-Wave Reactor',
COUNT(fact.`location_id`) AS '# Total',
SUM(fact.`capacity`) AS 'Total Capacity (in MWe)'
FROM `nuclear_power_plant_facts` fact
JOIN `reactors` rea ON fact.`reactor_id` = rea.`reactor_id`
JOIN `reactor_types` rty ON rea.`reactor_type_id` = rty.`id`
JOIN `power_plants` pow ON fact.`power_plant_id` = pow.`power_plant_id`
JOIN `status` sta ON pow.`status_id` = sta.`id`
GROUP BY sta.`id`;
```

Output:

Status	Advanced Boiling Water Reactor	Advanced Power Reactor	Advanced Pressurised Water Reactor	Advanced Gas-cooled Reactor	Boiling Water Reactor	Evolutionary Power Reactor	Fast Breeder Reactor	Gas-Cooled Reactor	High-Temperature Gas-cooled Reactor	High Temperature Reactor - Pebble Module
Shutdown	0	0	0	0	42	0	8	38	4	0
Under Construction	0	0	0	0	4	0	1	0	1	0
Operational	0	0	0	0	73	0	3	14	0	0
Cancelled Construction	0	0	0	0	0	0	0	0	0	0
Planned	0	0	0	0	0	0	0	0	0	0
Suspended Construction	0	0	0	0	0	0	0	0	0	0

Figure 4.10: Pivoting #1 Output (2/2)

Explanation: 6 row(s) returned. The output tallies each type of status based on reactor type. Then, the total tally with the total capacity is displayed.

4.6.3. Decades against Top Nuclear Nations

Table 4.11: Pivoting #2 (1/2)

Purpose	Display number of nuclear power plants and their total capacity constructed by in top nuclear nations for every decade since 1950.
Dimensions	Locations and Lifespans
# of Columns	11
Conditions (If applicable)	-
Justifications	OLAP can tell how many power plants constructed by the top nuclear nations, as well as the total capacity for every decade.

SQL Command:

```sql
SELECT
ctr.`name` AS 'Country',
COUNT(CASE WHEN lif.`construction_start_at` BETWEEN '1950-01-01' AND '1960-01-01' THEN 1 ELSE null END) AS '1950s',
COUNT(CASE WHEN lif.`construction_start_at` BETWEEN '1960-01-01' AND '1970-01-01' THEN 1 ELSE null END) AS '1960s',
COUNT(CASE WHEN lif.`construction_start_at` BETWEEN '1970-01-01' AND '1980-01-01' THEN 1 ELSE null END) AS '1970s',
COUNT(CASE WHEN lif.`construction_start_at` BETWEEN '1980-01-01' AND '1990-01-01' THEN 1 ELSE null END) AS '1980s',
COUNT(CASE WHEN lif.`construction_start_at` BETWEEN '1990-01-01' AND '2000-01-01' THEN 1 ELSE null END) AS '1990s',
COUNT(CASE WHEN lif.`construction_start_at` BETWEEN '2000-01-01' AND '2010-01-01' THEN 1 ELSE null END) AS '2000s',
COUNT(CASE WHEN lif.`construction_start_at` BETWEEN '2010-01-01' AND '2020-01-01' THEN 1 ELSE null END) AS '2010s',
COUNT(CASE WHEN lif.`construction_start_at` > '2020-01-01' THEN 1 ELSE null END) AS '2020s',
COUNT(fact.`location_id`) AS '# Total',
SUM(fact.`capacity`) AS 'Total Capacity (in MWe)'
FROM `nuclear_power_plant_facts` fact
JOIN `locations` loc ON fact.`location_id` = loc.`location_id`
JOIN `countries` ctr ON loc.`country_code` = ctr.`code`
JOIN `lifespans` lif ON fact.`lifespan_id` = lif.`lifespan_id`
WHERE ctr.`code` IN ('US', 'RU', 'GB', 'FR', 'CN', 'IN', 'PK')
GROUP BY ctr.`code`
ORDER BY ctr.`name`;
```

Output:

Country	1950s	1960s	1970s	1980s	1990s	2000s	2010s	2020s	# Total	Total Capacity (in MWe)
France	4	7	40	18	1	1	0	0	71	68538
India	0	4	4	4	2	9	6	0	29	11304
Pakistan	0	1	0	0	1	1	3	0	6	2353
People's Republic of China	0	0	0	3	7	22	25	0	57	53777
Russian Federation	3	9	19	11	0	7	5	0	52	37078
United Kingdom of Great Britain and Northern Ireland	20	18	2	5	0	0	1	0	46	16855
United States	13	64	60	0	0	0	4	0	137	113350

Figure 4.11: Pivoting #2 Output (1/2)

Explanation: 7 row(s) returned. The output tallies the top nuclear countries against how many each of them builds every decade since 1950. Then, the total number of plants and their total capacity are summed up.

4.6.4. Top Nuclear Nations against Decades

Table 4.12: Pivoting #2 (2/2)

Purpose	Display number of nuclear power plants and their total capacity constructed by in top nuclear nations for every decade since 1950.
Dimensions	Locations and Lifespans
# of Columns	10
Conditions (If applicable)	-
Justifications	OLAP can tell how many power plants constructed by the top nuclear nations, as well as the total capacity for every decade.

SQL Command:

```
SELECT
CONCAT(SUBSTRING(lif.`construction_start_at`, 1, 3), 0, 's') AS decade,
COUNT(CASE WHEN ctr.`code` = 'US' THEN 1 ELSE null END) AS 'United States',
COUNT(CASE WHEN ctr.`code` = 'RU' THEN 1 ELSE null END) AS 'Russia',
COUNT(CASE WHEN ctr.`code` = 'GB' THEN 1 ELSE null END) AS 'United Kingdom',
COUNT(CASE WHEN ctr.`code` = 'FR' THEN 1 ELSE null END) AS 'France',
COUNT(CASE WHEN ctr.`code` = 'CN' THEN 1 ELSE null END) AS 'China',
COUNT(CASE WHEN ctr.`code` = 'IR' THEN 1 ELSE null END) AS 'India',
```

```
COUNT(CASE WHEN ctr.`code` = 'PK' THEN 1 ELSE null END) AS 'Pakistan',
COUNT(fact.`location_id`) AS '# Total',
SUM(fact.`capacity`) AS 'Total Capacity (in MWe)'
FROM `nuclear_power_plant_facts` fact
JOIN `locations` loc ON fact.`location_id` = loc.`location_id`
JOIN `countries` ctr ON loc.`country_code` = ctr.`code`
JOIN `lifespans` lif ON fact.`lifespan_id` = lif.`lifespan_id`
WHERE lif.`construction_start_at` > '1950-01-01'
GROUP BY decade
ORDER BY lif.`construction_start_at`;
```

Output:

decade	United States	Russia	United Kingdom	France	China	India	Pakistan	# Total	Total Capacity (in MWe)
1950s	9	3	20	4	0	0	0	43	4260
1960s	64	5	18	7	0	0	1	140	71598
1970s	60	19	2	40	0	1	0	254	213919
1980s	0	11	5	18	3	0	0	107	99212
1990s	0	0	0	1	7	0	1	36	31836
2000s	0	7	0	1	22	0	1	53	47868
2010s	4	5	1	0	25	0	3	60	60867

Figure 4.12: Pivoting #2 Output (2/2)

Explanation: 7 row(s) returned. The output tallies every decade since 1950 against the top nuclear nations. Then, the total number of plants and their total capacity are summed up.

Chapter 5: Conclusion

In conclusion, the proposed. The constructed data warehouse uses the snowflake schema to retain the normalised form of the lookup tables from the original database in order to prevent data redundancy and save up on storage space and memory. From the constructed data warehouse, hierarchies can be formed in each dimension tables. Using the hierarchies, OLAP queries are performed to extract interesting patterns, display the nuclear plants information for better insights, and find correlations between real-life events and how nuclear power plants are built, retained, or decommissioned over the years.

In my opinion, though the proposed OLAP application does a good job in displaying the nuclear power plants data in a clear tabular format, it can still be further improved by incorporating Spatial OLAP, or SOLAP for short. SOLAP integrates map generalisation to significantly improve the analysis capabilities of spatial multidimensional. Map generalisation produces various depictions of geographic information at various granularities and/or scales. Since it enables displaying geographical phenomena and geographic objects according to several versions, it can aid in the spatial decision-making process. Each of them communicates various information at various levels of abstraction, enabling decision-makers to concentrate exclusively on crucial decision-related factors and assisting them in the creation and validation of hypotheses. To summarise, the proposed OLAP application can be turned into SOLAP in which it displays the nuclear power plant locations using map generalisation, rather than just display them on tables.

References

Chandra, P., & Gupta, M. K. (2017, December 15). Comprehensive survey on data warehousing research. *International Journal of Information Technology, 10*, 217–224. doi:https://doi.org/10.1007/s41870-017-0067-y

Chhabra, M. (2022, February 3). *Understanding EDW (Enterprise Data Warehouse) Simplified 101*. Retrieved from Hevo Data Inc: https://hevodata.com/learn/enterprise-data-warehouse-edw-simplified/

Dancuk, M. (2021, June 10). *Star vs. Snowflake Schema*. Retrieved from phoenixNAP: https://phoenixnap.com/kb/star-vs-snowflake-schema

Data Channel. (2020, October 8). *The Best Applications of Data Warehousing*. Retrieved from Data Channel: https://datachannel.co/blogs/best-applications-of-data-warehousing/

Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques 2nd Edition.* San Francisco: Morgan Kaufmann.

IBM Cloud Education. (2020, June 18). *What is OLAP?* Retrieved from IBM: https://www.ibm.com/cloud/learn/olap#:~:text=OLAP%20(for%20online%20analytical%20processing,other%20unified%2C%20centralized%20data%20store.

Karmani, M. Z., Mustafa, G., Sarwar, N., Wajid, S. H., Nasir, J., & Siddque, S. (2020). A Review of Star Schema and Snowflakes Schema. *Intelligent Technologies and Applications*, 129-140. doi:10.1007/978-981-15-5232-8_12

Kramer, G., & Nisbet, S. (n.d.). *The Practical Union of OLAP Analysis and Geographic Mapping*. Retrieved from Esri: https://www.esri.com/news/arcuser/0206/olap1of2.html

Nanda, A., Gupta, S., & Vijrania, M. (2019). A Comprehensive Survey of OLAP: Recent Trends. *Proceedings of the Third International Conference on Electronics Communication and Aerospace Technology [ICECA 2019]*, (pp. 425-430). doi:10.1109/ICECA.2019.8822203

Oracle. (2022). *What is a Data Mart?* Retrieved from Oracle: https://www.oracle.com/autonomous-database/what-is-data-mart/#:~:text=A%20data%20mart%20is%20a%20simple%20form%20of%20a%20data,fewer%20sources%20than%20data%20warehouses.

Oracle. (n.d.). *MySQL Workbench*. Retrieved from MySQL: https://www.mysql.com/products/workbench/

Panoply. (n.d.). *Data Warehouse Architecture: Traditional vs. Cloud*. Retrieved from Panoply Ltd: https://panoply.io/data-warehouse-guide/data-warehouse-architecture-traditional-vs-cloud/#:~:text=A%20virtual%20data%20warehouse%20is,line%20specific%20reporting%20and%20analysis.

Shanghai Nuclear Office. (n.d.). *China Nuclear Power*. Retrieved from Shanghai Nuclear Office: http://www.smnpo.cn/zghd.htm

Watson, H. J., Goodhue, D. L., & Wixom, B. H. (2002, May). The benefits of data warehousing: why some organizations realise exceptional payoffs. *Information & Management, 39*(6), 491-502. doi:10.1016/S0378-7206(01)00120-3

Wikipedia. (n.d.). *Nuclear power in the Soviet Union*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Nuclear_power_in_the_Soviet_Union

Wikipedia. (n.d.). *Visual Paradigm*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Visual_Paradigm

World Nuclear Association. (2022, August). *Nuclear Power in the USA*. Retrieved from World Nuclear Association: https://world-nuclear.org/information-library/country-profiles/countries-t-z/usa-nuclear-power.aspx

World Nuclear Association. (2022, August). *Nuclear Power Reactors*. Retrieved from World Nuclear Association: https://world-nuclear.org/information-library/nuclear-fuel-cycle/nuclear-power-reactors/nuclear-power-reactors.aspx#:~:text=of%20nuclear%20reactor-,Pressurised%20water%20reactor%20(PWR),more%20employed%20for%20naval%20propulsion.