



KHO DỮ LIỆU & HỆ HỖ TRỢ QUYẾT ĐỊNH

Xây dựng Data Warehouse thời tiết và hạn - mặn các tỉnh Nam Bộ

GVHD: Đỗ Thanh Thái

Nhóm 17: 2210535 - Thái Thành Duy
2112394 - Vũ Ngọc Thuận
2115319 - Lê Hoàng Anh Vũ



Nội dung chính

Tổng quan hệ thống 1

Chi tiết hệ thống 2

Hiện thực kho dữ liệu 3

Kết quả và hình ảnh mô phỏng 4

Tài liệu tham khảo 5

1

Tổng quan hệ thống

**Yêu cầu kho
dữ liệu**

**Vấn đề được
giải quyết**

**Cấu trúc hệ
thống**



- **Tích hợp và lưu trữ dữ liệu:** về thời tiết cùng độ hạn - mặn trên địa bàn các tỉnh Nam Bộ.
- **Cung cấp dữ liệu:** mạch lạc và nhất quán sử dụng được trong báo cáo và phân tích về thời tiết và độ hạn - mặn.

1

Tổng quan hệ thống

Vấn đề được giải quyết

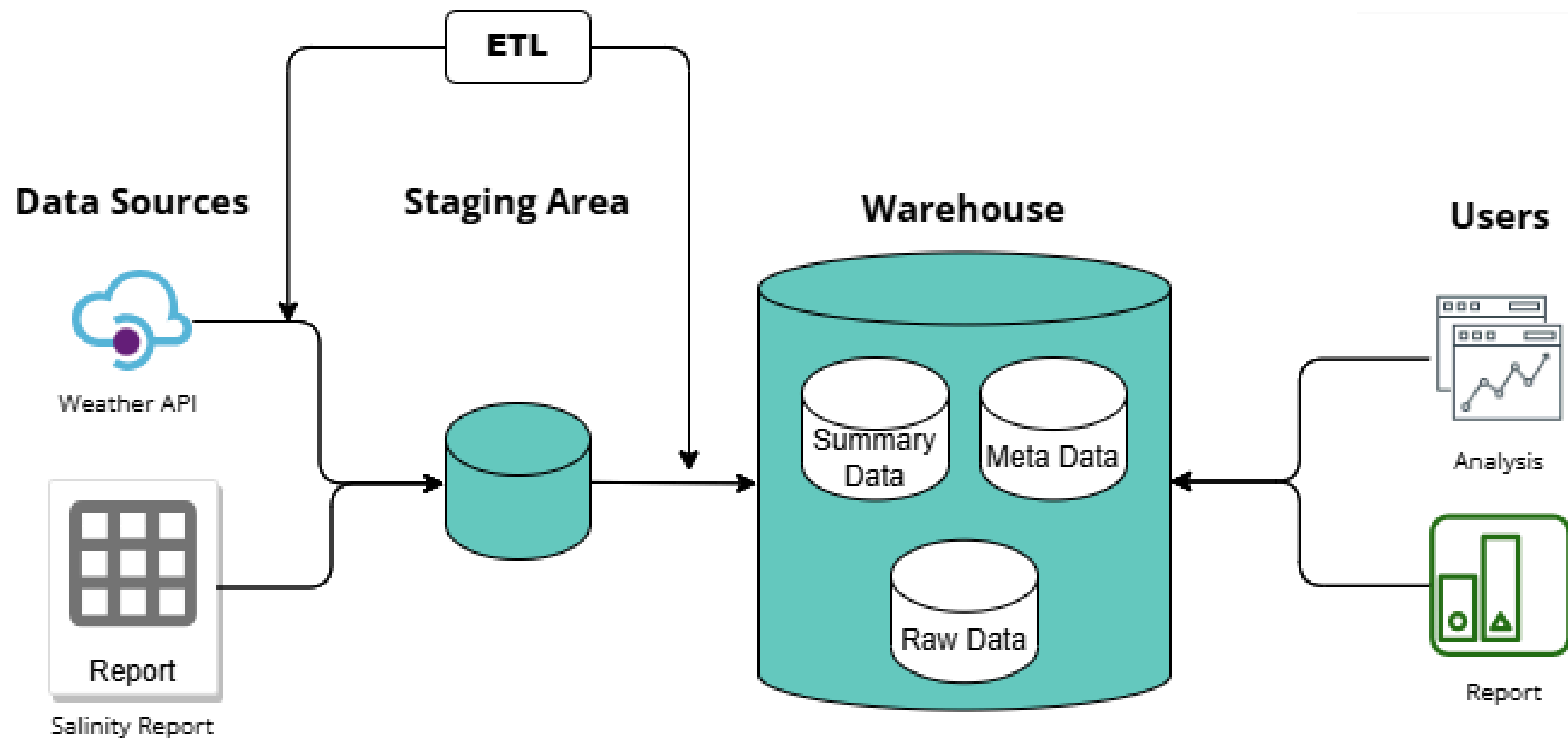
Xây dựng báo cáo về thời tiết và độ hạn - mặn



Phân tích mối quan hệ giữa thời tiết và độ hạn mặn

Phân tích xu hướng thay đổi của thời tiết và độ hạn mặn

Đưa ra kế hoạch để ứng phó với hiện tượng xâm nhập mặn





**Kiến trúc kho
dữ liệu**

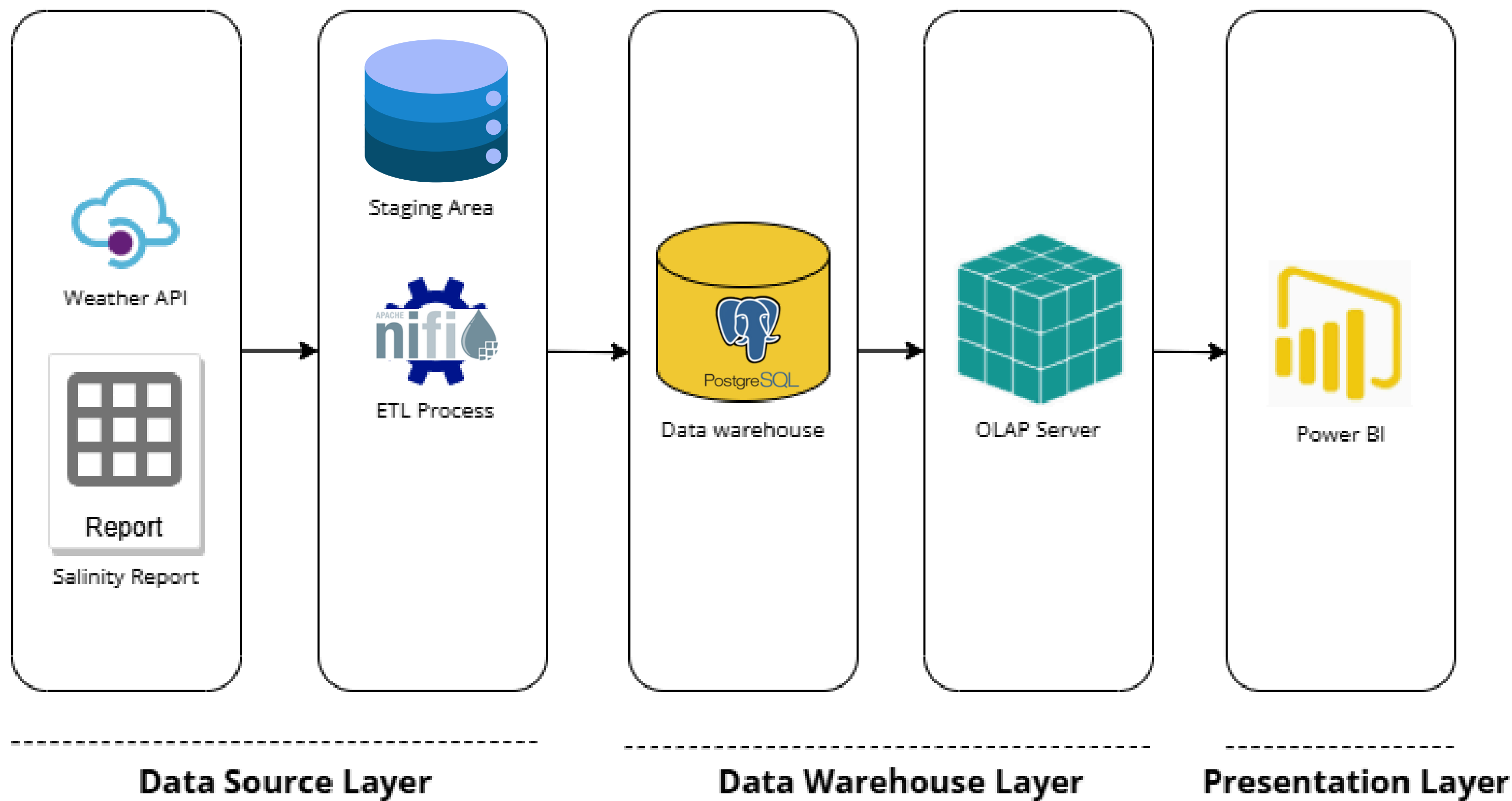


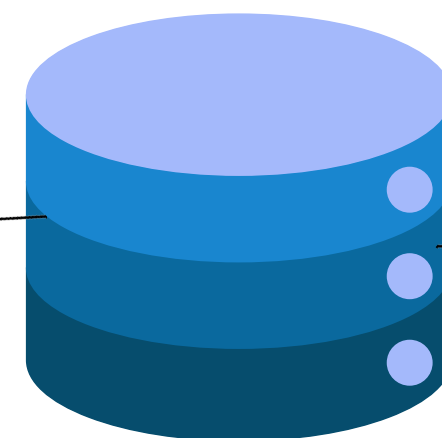
Nguồn dữ liệu

1

Tổng quan hệ thống

Kiến trúc kho dữ liệu





Dữ liệu thời tiết

Open Metro
MeteoSource API



Dữ liệu độ mặn đất

Trung tâm Khí tượng Thủy văn
Khu vực Nam Bộ (SRHMC)



Dữ liệu thời tiết

Open Metro
MeteoSource API



- Lấy hàng ngày từ API
- Dữ liệu về: vị trí địa lý, nhiệt độ, độ ẩm, lượng mưa, độ che phủ mây

```
"daily": {
  "day": "2025-04-06",
  "icon": 2,
  "summary": "Sunny. Temperature 5/15 °C. Wind",
  "all_day": {
    "temperature_min": 4.8,
    "temperature_max": 14.8,
    "wind_speed": 4.1,
    "probability": {
      "precipitation": 0
    }
  },
  "astro": {
    "sun": {
      "rise": "2025-04-06T06:24:24",
      "set": "2025-04-06T19:43:19"
    }
  }
},
```

Dữ liệu độ mặn đất

Trung tâm Khí tượng Thủy văn
Khu vực Nam Bộ (SRHMC)



- Lấy hàng tuần từ dữ liệu của website
- Tổng hợp thành một file báo cáo **.csv**



	A	B	C	D	E
1	province	start_date	end_day	avg_salinity	
2	An Giang	4/7/2025	4/13/2025	0.15	
3	Ba Ria - Vun	4/7/2025	4/13/2025	23.12	
4	Bac Lieu	4/7/2025	4/13/2025	25.73	
5	Ben Tre	4/7/2025	4/13/2025	18.33	
6	Binh Duong	4/7/2025	4/13/2025	2.56	
7	Binh Phuoc	4/7/2025	4/13/2025	0.82	
8	Ca Mau	4/7/2025	4/13/2025	29.87	
9	Can Tho	4/7/2025	4/13/2025	0.32	
10	Dong Nai	4/7/2025	4/13/2025	0.35	
11	Dong Thap	4/7/2025	4/13/2025	0.55	
12	Hau Giang	4/7/2025	4/13/2025	9.76	
13	Ho Chi Minh	4/7/2025	4/13/2025	14.98	
14	Kien Giang	4/7/2025	4/13/2025	8.71	
15	Long An	4/7/2025	4/13/2025	5.83	
16	Soc Trang	4/7/2025	4/13/2025	15.23	
17	Tay Ninh	4/7/2025	4/13/2025	1.2	
18	Tien Giang	4/7/2025	4/13/2025	13.65	
19	Tra Vinh	4/7/2025	4/13/2025	16.39	
20	Vinh Long	4/7/2025	4/13/2025	5.46	
21					

3

Hiện thực kho dữ liệu

**Quy trình ETL
(Extract-
Transform-
Load)**

**Thiết kế
Staging Area**

**Mô hình dữ
liệu của kho
dữ liệu**

Apache Nifi

- Một nền tảng phần mềm mã nguồn mở tự động hóa và quản lý luồng dữ liệu giữa các hệ thống
- Hỗ trợ quy trình ETL với giao diện trực quan
- Hỗ trợ theo dõi và ghi lại từng bước của quá trình ETL



3

Hiện thực kho dữ liệu Staging Area

- **Bảng dữ liệu thô:** chứa dữ liệu thô lấy được từ *Nguồn dữ liệu*.

- **Bảng dữ liệu đã được xử lý:** chứa dữ liệu đã được xử lý và định dạng từ Bảng dữ liệu thô

climeweather
staging_weather_raw
id serial
source character varying(50)
raw_data jsonb
received_at timestamp with out time zone
processed boolean

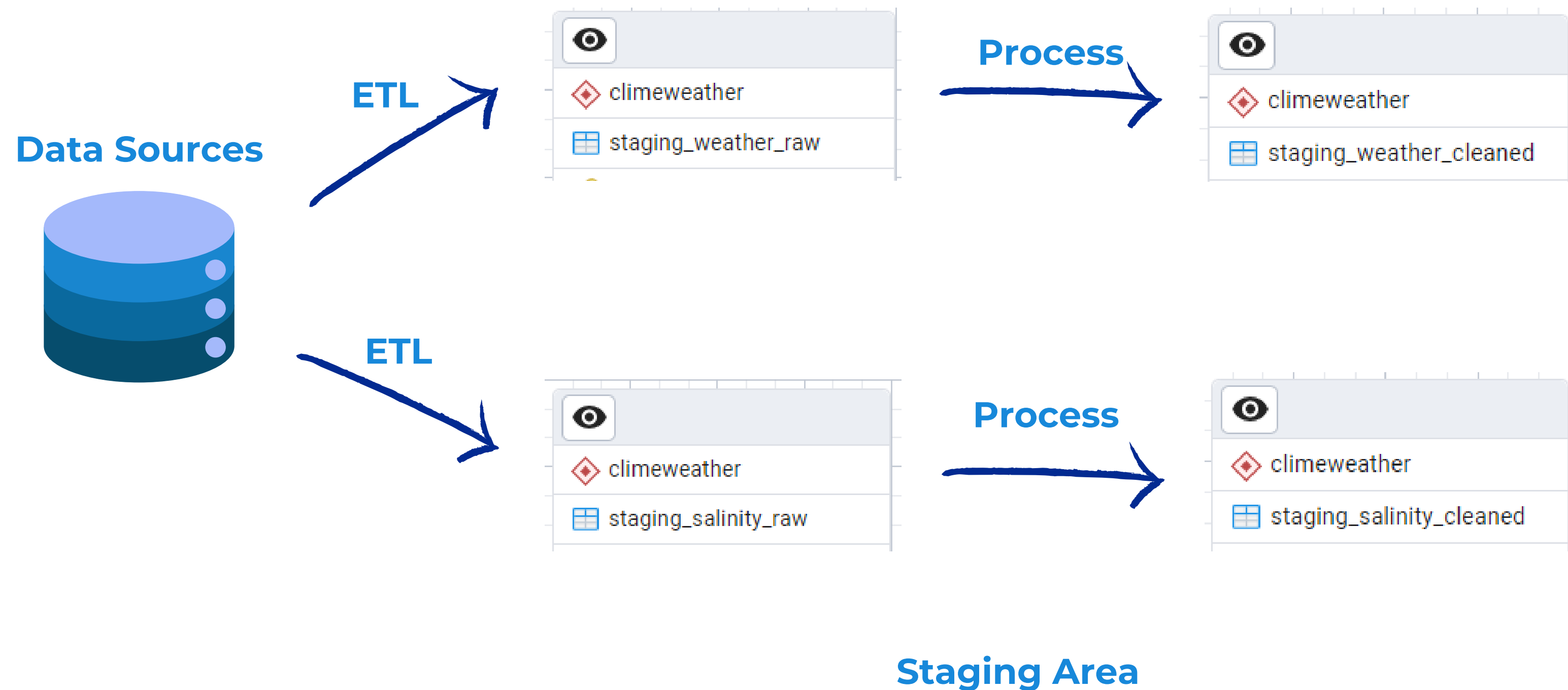
climeweather
staging_salinity_raw
id serial
source character varying(50)
raw_data jsonb
received_at timestamp with out time zone
processed boolean

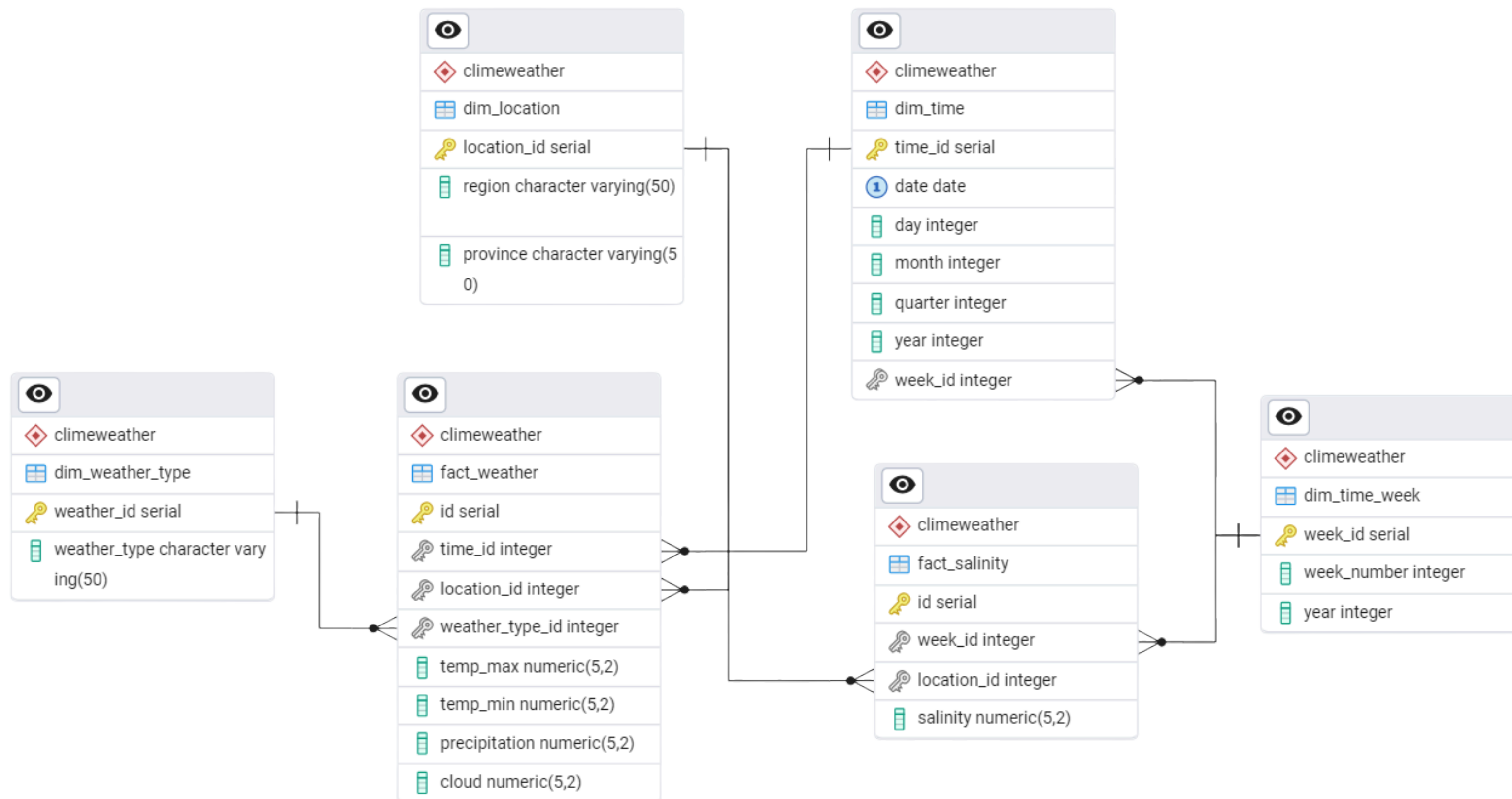
climeweather
staging_weather_cleaned
id serial
time date
location character varying (50)
weather_type character varying(50)
temp_max numeric(5,2)
temp_min numeric(5,2)
precipitation numeric(5,2)
cloud numeric(5,2)
processed_at timestamp without time zone

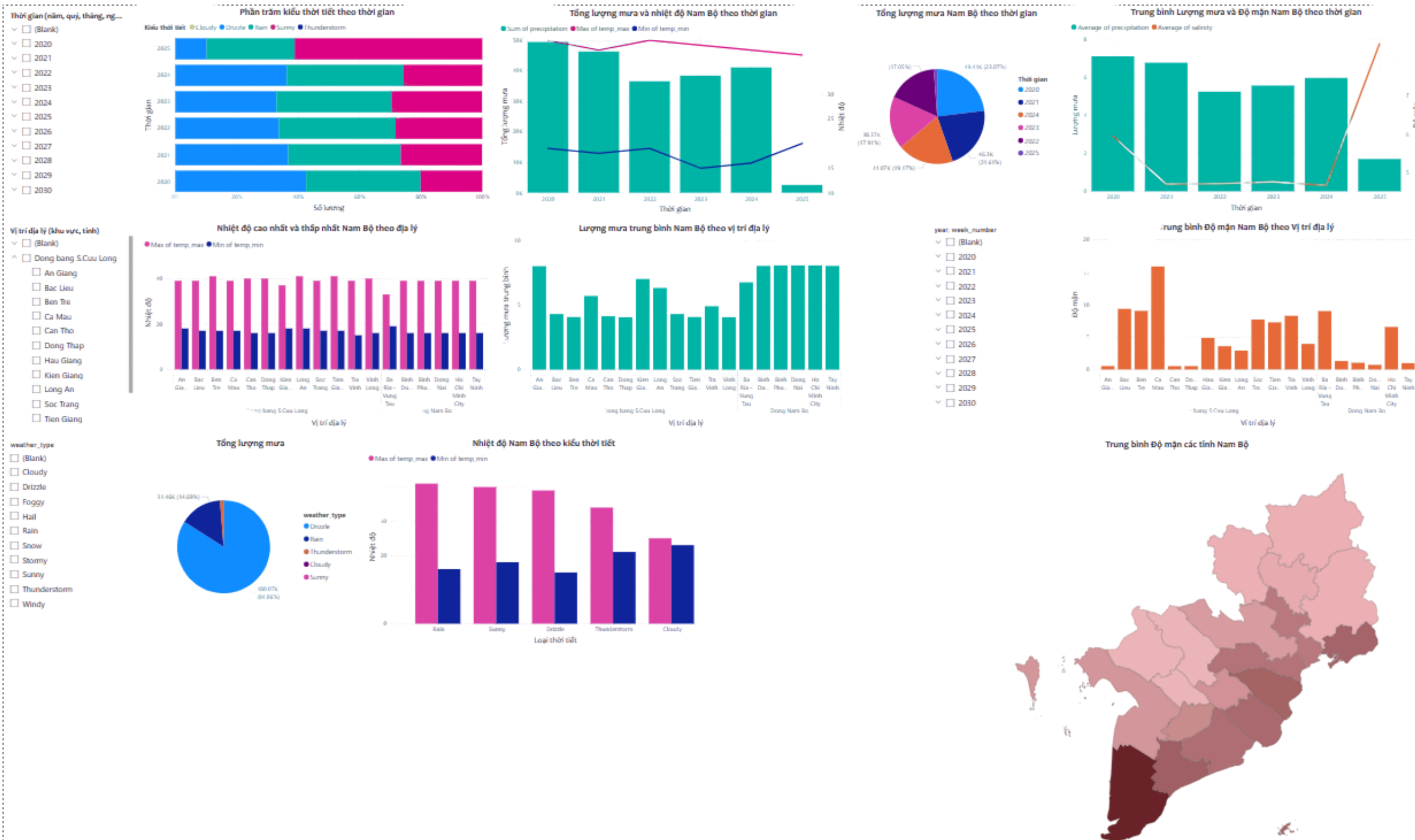
climeweather
staging_salinity_cleaned
id serial
week_number integer
year integer
location character varying (50)
avg_salinity numeric(5,2)
processed_at timestamp without time zone

3

Hiện thực kho dữ liệu Staging Area







5**Tài liệu tham khảo**

- [1] William H. Inmon, “Building the Data Warehouse,” Fourth Edition, Wiley Publishing, Inc., 2005..
- [2] R. Kimball, M. Ross, "The Data Warehouse ToolKit", 3rd Edition, Wiley Publishing, Inc., 2013.
- [3] M. Golfarelli, D. Maio, and S. Rizzi, “Conceptual Design of Data Warehouses from E/R Schemes,” the Proceedings of the Hawaii International Conference On System Sciences, January 6-9, 1998, Kona, Hawaii, pp. 334-343..



ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC BÁCH KHOA

**CẢM ƠN MỌI NGƯỜI
ĐÃ LẮNG NGHE**

Thành phố Hồ Chí Minh, Tháng 4/2025