

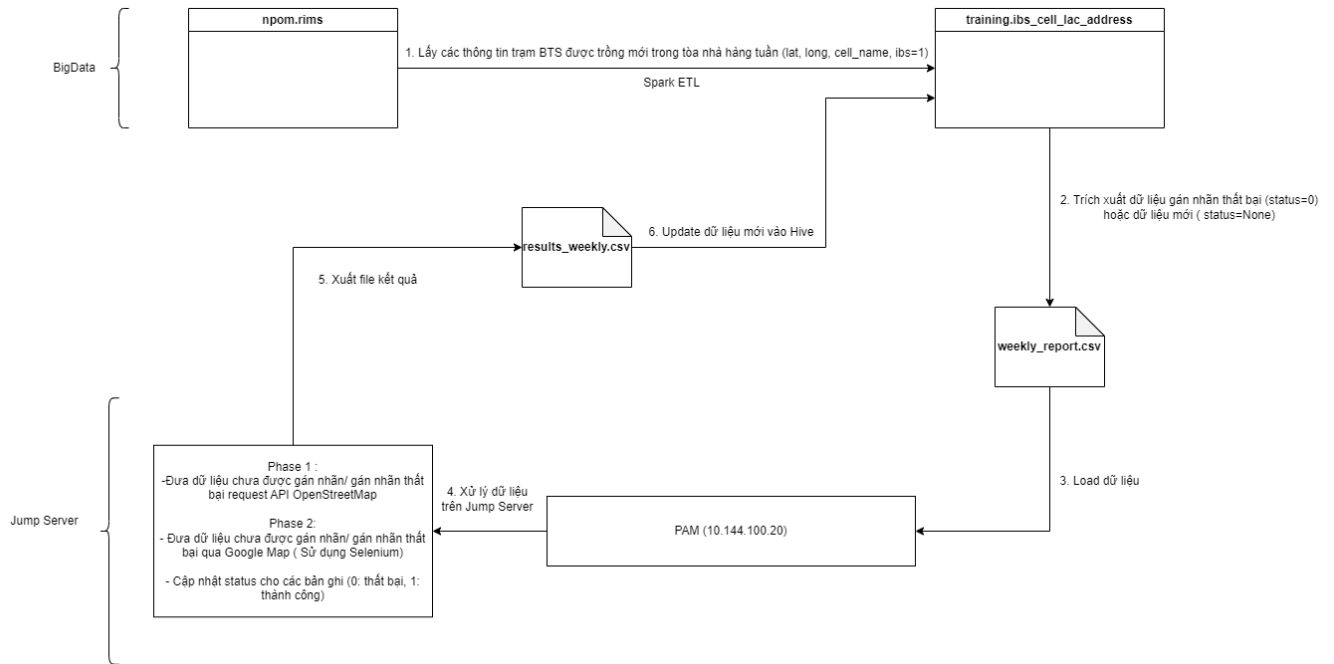
TÀI LIỆU MÔ TẢ TIẾN TRÌNH GÁN NHÃN TRẠM VIỄN THÔNG (BTS)

I. Format bảng dữ liệu training.ibs_cell_lac_address:

Mô tả:

1	cell	INT	CELL ID
2	lac	INT	Mã khu vực
3	latitude	DOUBLE	Vĩ độ đặt trạm BTS
4	longitude	DOUBLE	Kinh độ đặt trạm BTS
5	name	STRING	Tên Trạm BTS
6	address	STRING	Địa chỉ tương ứng với trạm, có 2 trường hợp: <ul style="list-style-type: none">- Tên tòa nhà, bệnh viện, trường học,... (Nếu dữ liệu được gán nhãn thành công)- Địa chỉ chi tiết (Tên đường, phường, quận,... nếu dữ liệu chưa được gán nhãn thành công)
7	status	INT	0: Gán nhãn không thành công 1: Gán nhãn thành công None: Dữ liệu mới được cập nhật

II. Mô tả luồng xử lý dữ liệu



1. Dữ liệu sẽ được trích xuất từ bảng npom.rims hàng tuần để lấy các thông tin về trạm BTS (Base Transceiver Station) được trồng mới trong các tòa nhà (ibs_status =1), những dữ liệu này sẽ được thêm mới vào bảng **training.ibs_cell_lac_address**
2. Từ bảng training.ibs_cell_lac_address, những bản ghi của các trạm chưa được gắn nhãn thành công (status=0) hoặc bản ghi vừa được thêm mới ở bước 1 (status=None) sẽ được trích xuất ra file **weekly_report.csv** có địa chỉ **/home/ducanh/bts_ibs_address_report/weekly_report.csv**
3. Lấy file dữ liệu từ server Big Data đến PAM (10.144.100.20) vào thư mục có địa chỉ **E:\Data\ducanh\test_request\weekly_report.csv**
4. Xử lý dữ liệu hiện tại trên Jump Server theo 2 phase:
Proxy: 10.144.26.20:3129

Phase 1: Cập nhật bảng dữ liệu với dữ liệu đầu vào là latitude, longitude, thông qua API của **OpenStreetMaps** cập nhật cột address với giá trị là tên tòa nhà, trường học, bệnh viện,... giá trị cột status được cập nhật bằng 1 (thành công)

Phase 2: Các giá trị chưa được gắn nhãn sau Phase 1, sẽ được cập nhật qua **Google Maps**, sử dụng Seleniums WebDriver. Nếu kết quả trả về là tên tòa nhà, bệnh viện, trường học,... cột status sẽ được cập nhật bằng 1, trong trường hợp trả về địa chỉ chính xác (tên đường, phường, quận, ...) thì cột status được cập nhật bằng 0.

Quá trình xử lý dữ liệu sẽ được thực hiện bởi file **label_bts_ibs.exe**, kết quả đầu ra là một file csv đã được cập nhật cột 'address' và cột 'status' có địa chỉ

E:\Data\ducanh\test_request\results_weekly.csv, file log quá trình chạy sẽ được lưu vào **bts_ibs_label.log**

5. Đẩy file kết quả lên server Big Data
6. Insert dữ liệu mới nhất từ file **results_weekly.csv** (đã được cập nhật cột ‘address’ và ‘status’) vào Hive

III. Thời gian đẩy dữ liệu (dự kiến)

- Hàng tuần (Thứ 6)
- Thời gian: 12h00