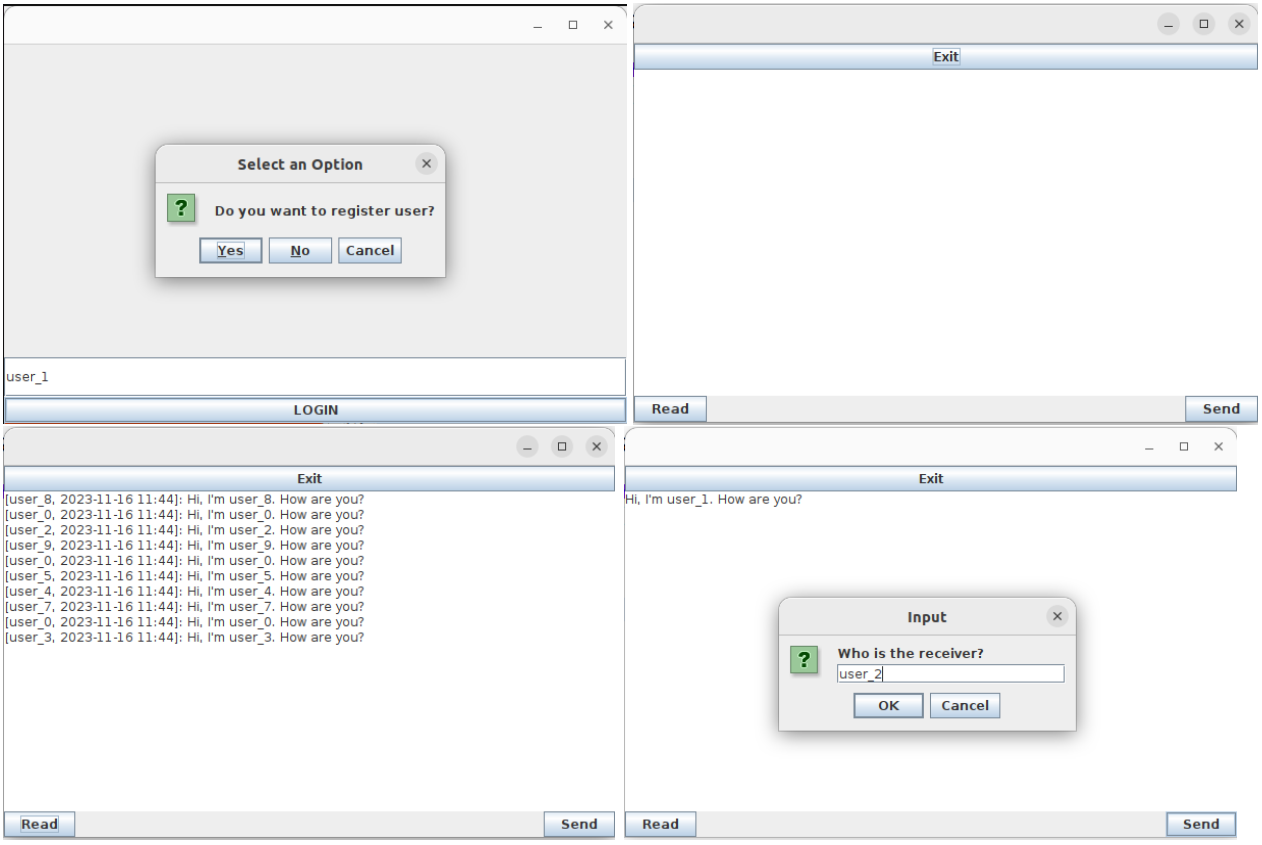
Intermediate Test - Big Data Analytics and Reasoning

Problem 1

We want to implement simple chat that uses hdfs as storage. In particular, we want to have different a folder for each user named with its `username`. Each user folder contains a subfolder `messages` that contains all the messages received by the user.
For this task you are provide with a java swing app described by the following snapshots:



You are asked to implement the following methods:

- LoginPanel::createNewUser(String username); This method creates the user folder with messages subfolder

- LoginPanel::existsUser(String username); This method checks user folder existence
- UserPage::getMessages(); This method reads all messages in the user/messagges folder
- UserPage::send(Message m,String receiver); This method write the current message in the receiver user folder in a file named with sender username

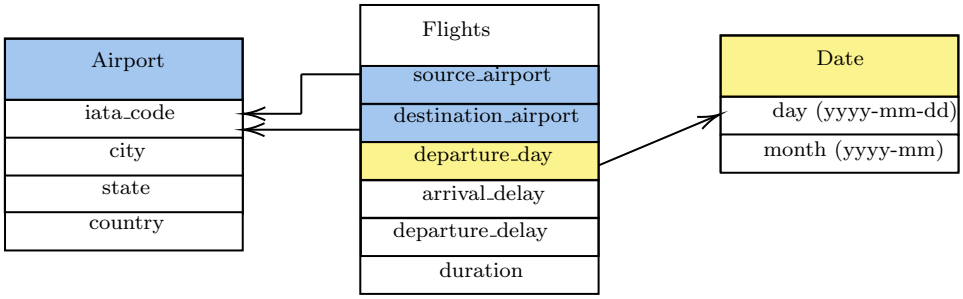
More details are provided in comments for each method

Problem 2

We want to build a data warehouse, for storing data about flights and airports. In particular, we have a data source that is compliant with the following schema:

- `flights(flight_id,year, month, day, day_of_the_week, flight_number, origin_airport, destination_airport, scheduled_departure_time, departure_delay, duration, arrival_delay, cancelled, weather_delay)`
- `airports(iata_code, airport, city, state, country, lat, long)`

We want to build a data warehouse with the following schema:



Build the requested data warehouse and populate it with data stored in the data source. In particular, flights having one of the following field empty has to be filtered out: `year`, `month`, `day`, `origin_airport`, `destination_airport`, and . Note that the company is interested in performing analysis in fixed time zones (*morning 08:00-14:00, evening 15:00-21:00, and night 22:00-07:00*). Thus, provide a way to perform analysis in such time zones. Moreover, you are asked to perform the following query:

- For each country, compute the average duration of morning flights that leave and land in the same country.
- For each state, compute the annual average of the number flights the leave and land from/in a given state.
- Find the average duration for flights landing in an airport with at least 1 landing flight for each month. (Bonus)