Department of Electrical and Computer Engineering
Iowa State University
Spring 2015

Purpose

The goal of this lab is to design a system for helping monitor the location of young children in a shopping center with the help of appropriate sensors. Your task is to design the backend data processing system using Infosphere Streams. Along the way, you will learn the concept of windows.

Submission

Create an archive with the following and hand it in through blackboard.
- The operators that you have used, and the data flow diagram between the operators.
- Commented Code for your program. Include all source files needed for compilation.

Resources

You will need the following before you start the lab.
- IP Address of the Node, your login id and private key file
- Documentation about SPL. It may be especially useful to use the Windowing features of streams. Refer to the Window clause for details

Setting

A shopping center has a new feature to improve the safety of young children within its premises. Each child entering the shopping center is provided with an ID with a chip on it (such as an RFID tag), and each ID has a unique identification number that can be mapped to the identity of the child. The child carries this ID as long as he/she is within the center.

The center has 27 rooms, 0 to 26, (we consider corridors as rooms too) and an additional room number "E" to mark the exit of the child from shopping center. A door connects one room to another. On each door, there is a sensor; if any child with an ID passes through this door, the sensor records the timestamp, the ID of the child passing through, the ID of the sensor itself, the room number from where the child is coming from, and the room number where the child is going to. Each sensor produces a continuous stream of events with all this information, and sends it to a central analyzer.

The data stream provided to you has been made 50x faster so that the duration for which you run the experiment is confined to 15 minutes. So, a minute's data in the stream provided to you is equivalent to data generated over a period of 50 minutes in real time.

Task (80 points)

Write an application that does the following.

1. At the end of every 2 minutes block from the time you start running the program, output the location of each child that has entered the shopping center. The update should be of the form: <timestamp, child ID, current room>.
   If for some reason, the current room of a child cannot be determined, then use "1000" for the current room. Write this output into a file "LocationLog.txt".

2. If a child has not been tracked by any sensor for more than 15 seconds, then generate an alert on a "missing child" stream, with the following format:
   <current time, child ID, time last seen, room last seen, parent phone number>.
   This stream must be written to the output file "MissingChildAlert.txt", in csv format.

3. It is unsafe for kids to spend a long time in room 26. If a kid spends more than 1 minute in room 26, an alert should be generated and a phone call must be made to the parent of the kid. The alert should have the following information <timestamp, child ID, parent phone number>, and should be written to the output file "UnsafeRoomAlert.txt".

4. In addition, output the following statistics the end of every 5 minutes block from the time you start running the program.

   a. For each room, the number of distinct children who visited the room over the last 5 minutes. This should be written to the file "RoomStats.txt", in the format
   <timestamp, room number, number of distinct children>

   Note that if a room already had a child "A" in it at the beginning of the 5 minutes block, "A" counts as a child that visited the room during the 5 minutes block.

   b. For each child, the number of distinct rooms the child visited over the last 5 minutes. This should be written to the file "ChildStats.txt", in the format
   <timestamp, child number, number of distinct rooms>

   Note that if a child was already in room "X" at the beginning of the 5 minutes block, "X" counts as a room that the child visited during this 5 minute.

Input Format

All input data are stored in "/datasets/Lab10" folder. There is a file named "ChildDetails.csv", which has the following information for each child, in csv format. This data is **not a stream** (i.e. in real-life, this data is relatively static), and should be read before processing the actual event stream.
- Child ID
- Parent's name
- Child Name
- Parent's phone number

The streaming data contains the stream of readings from all sensors.

- Timestamp
- Sensor ID
- Child ID
- From room
- To room

P.S. For the final results, run the program for at least 15 minutes, starting from the top of 15 minutes block in any hour. For instance, you may run the program at 12:00am, 12:15am, 12:30am, 12:45am, 1:00am, 1:15am, 1:30am, .., 11:00pm, 11:15pm, 11:30pm, 11:45pm. The clock synchronization has been taken care of.