Assignment 1

IN5040

TA: Espen Volnes
Email: espenvol@ifi.uio.no

Important information

- Deadline: Tuesday 6 October – 11:59
- Deliverable: PDF file containing:

  -> Query and results for questions: 1 -3 - 4 – 6 – 7 – 8

  -> Query, results and a short written answer for question 2

  -> Short written answer for question 5

Overview

- In this assignment we simulate a meteorological institute which receives summarised daily weather recordings from two weather stations

- Period: 2014 – 2018

- Weather stations: JFK airport – New York
  San Francisco Airport

Purpose of the Assignment

- To gain an understanding of some of the basic techniques used for analysing stream data.

- For this task, we will be using Esper, which is a software for Complex event processing (CEP) and Stream processing

Esper

- Language, compiler and runtime for DSM & CEP
- Basic version is a library that can be imported in Java or .NET
- Event Processing Language (EPL) enables rich expressions over events and time
- We will be using version 8.2.0
- http://esper.espertech.com/release-8.2.0/reference-esper
  - Look at page 3 points 2.1 - 2.10 and 2.17 for some of the most
    - important points for the assignment

Documentation: http://esper.espertech.com/release-8.2.0/reference-esper
Try online: https://esper-epl-tryout.appspot.com/epltryout/mainform.html

How to learn syntax and features?

1. Documentation
   - Focus on the basic concepts to begin with
2. Esper notebook
   - https://www.esperonline.net/
3. Esper EPL online
   - https://esper-epl-tryout.appspot.com/epltryout/index.html

http://esper.espertech.com/release-8.2.0/reference-esper/pdf/esper_reference.pdf

Getting Started

- Download assignment1.tar.gz from the IN5040 course page
- Extract:

```
tar –xzvf assignment1.tar.gz
```

Assignment 1 contents

- Assignment1.pdf
- Assignment1.java
- WeatherTuple.java
- Makefile
- jfk.csv
- san_francisco.csv
- query1 – 8.epl
- lib folder

Compilation

- make all:

  Cleans java class files
  Compiles java files

- make runX

  Runs the java files and the queryX, where X is the query number.

Documentation: http://esper.espertech.com/release-8.2.0/reference-esper
Try online: https://esper-epl-tryout.appspot.com/epltryout/mainform.html

Assignment1.java

- You don't have to make changes here, but it's useful to understand what's happening

- What the program does
  - creates runtime system and deploys the selected query
  - Reads from the two data set files
  - Sets up two threads, one for each data set
  - The threads sleep between each processed event
    - You can comment out the sleep invocation in the java program to speed up the execution, but don't do it for the last query (query 8)
    - Reason is that query 8 looks at both data streams

Structure of a WeatherTuple

- timestamp
- weatherStation
- stationName
- averageTemperature
- minimumTemperature
- maximumTemperature
- averageWindSpeed
- precipitation
- weather

Documentation: http://esper.espertech.com/release-8.2.0/reference-esper
Try online: https://esper-epl-tryout.appspot.com/epltryout/mainform.html

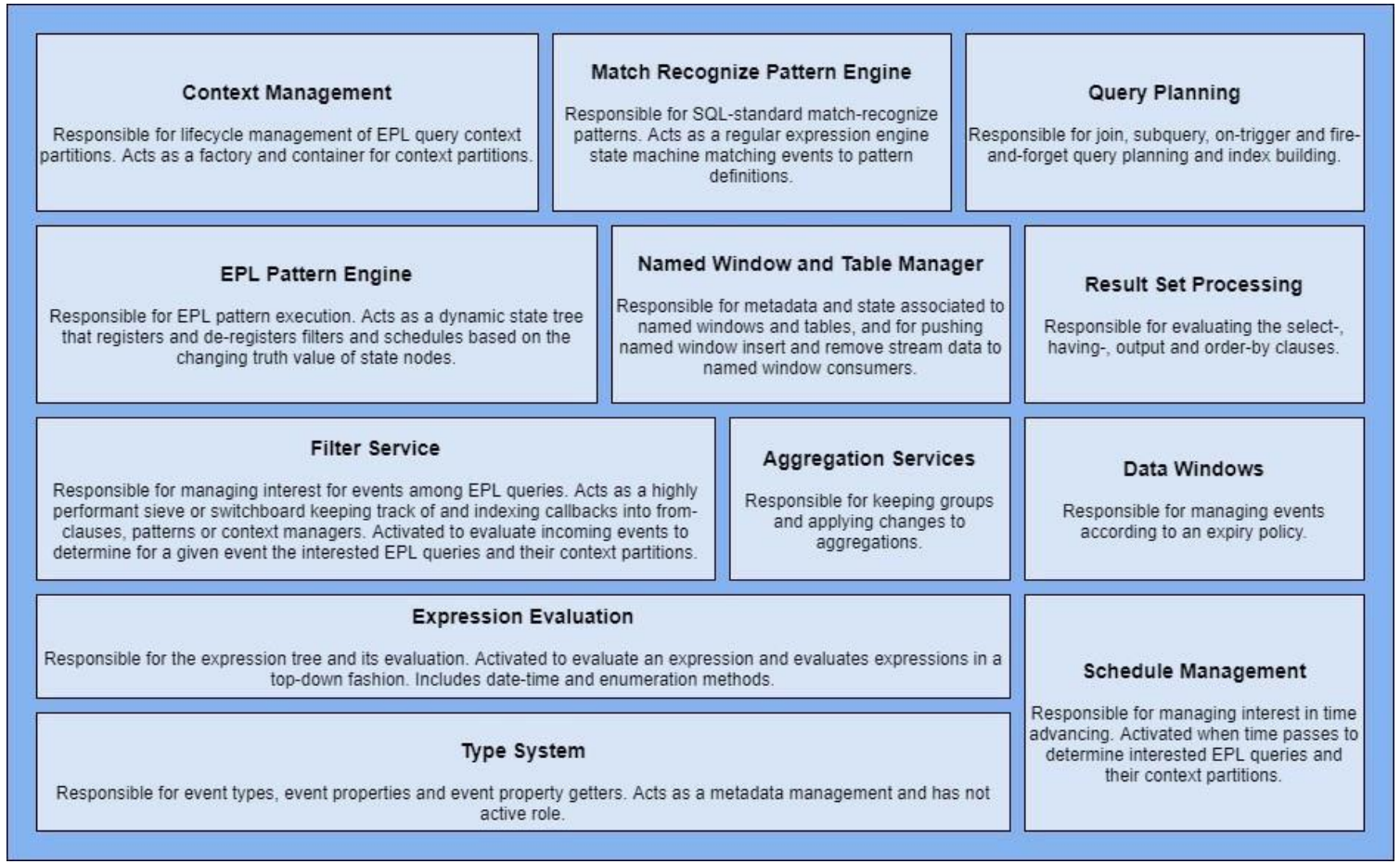
Query Structure

- SELECT *
  FROM jfk

- SELECT *
  FROM san_francisco

Documentation: http://esper.espertech.com/release-8.2.0/reference-esper
Try online: https://esper-epl-tryout.appspot.com/epltryout/mainform.html

Esper architecture



Documentation: http://esper.espertech.com/release-8.2.0/reference-esper
Try online: https://esper-epl-tryout.appspot.com/epltryout/mainform.html

EPL Syntax

- Create schema

```
create schema SchemaName(field_name1 type, field_name2 type)
```

- Data Stream Management:

```
select ID as sensorId, sum(countTags) as numTagsPerSensor
from AutoIdRFIDExample#win:time(60 seconds)
where Observation[0].Command = 'READ_PALLET_TAGS_ONLY'
group by ID
```

- CEP:

```
select a.custId, sum(a.price + b.price)
from pattern [every a=ServiceOrder ->
                    b=ProductOrder(custId = a.custId)
              where timer:within(1 min)]#win:time(2 hour)
where a.name in ('Repair', b.name)
group by a.custId
having sum(a.price + b.price) > 100
```

Documentation: http://esper.espertech.com/release-8.2.0/reference-esper
Try online: https://esper-epl-tryout.appspot.com/epltryout/mainform.html

Window

- Called a view in Esper
- Choose correct window when writing EPL queries!
- Examples of windows/views:

View	Syntax	Description
Time window	win:time(time period)	Sliding time window
Externally-timed window	win:ext_timed(timesta mp expression, time period)	Sliding time window, based on the millisecond time value supplied by an expression
Time batch window	win:time_batch(time period[,optional reference point] [, flow control])	Tumbling window that batches events and releases them every specified time interval, with flow control options
Externally-timed batch window	win:ext_timed_batch(*timestamp_expression*, *time_period* [,*optional_reference_point*])	Tumbling window, based on the millisecond time value supplied by an expression

Documentation: http://esper.espertech.com/release-8.2.0/reference-esper
Try online: https://esper-epl-tryout.appspot.com/epltryout/mainform.html