筑波大学
University of Tsukuba

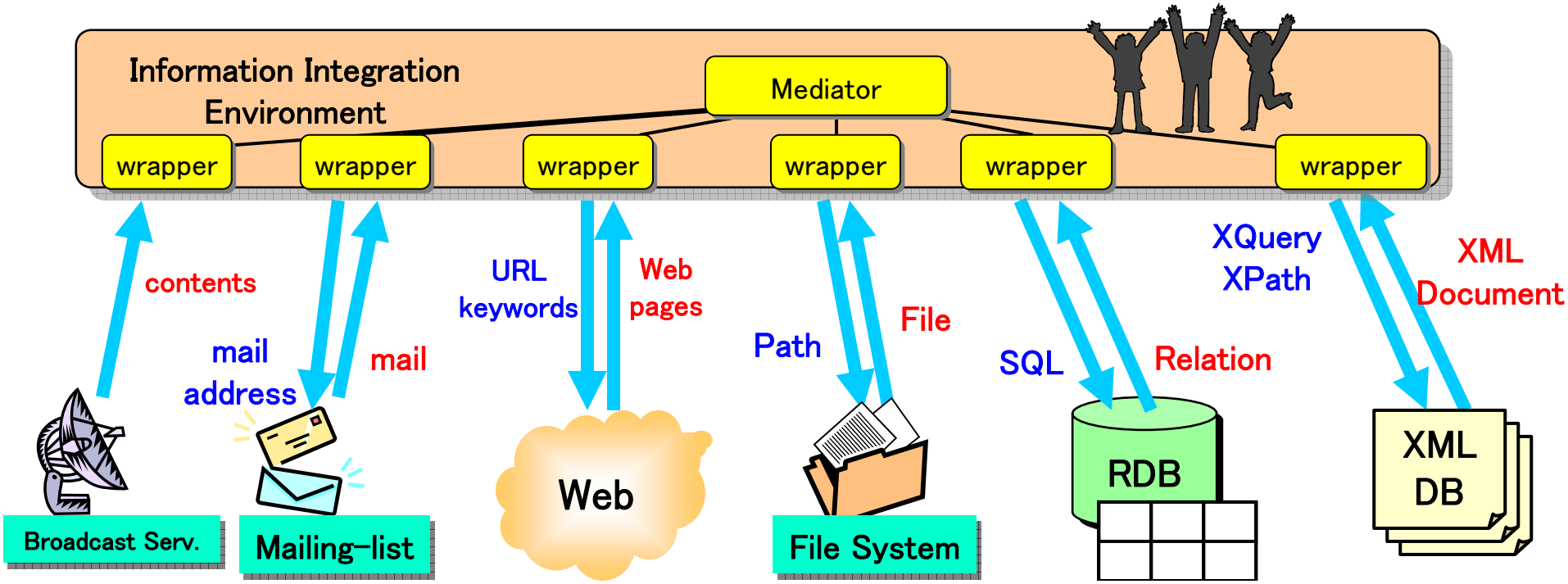
Information Integration
- An Overview of StreamSpinner Project -

Yousuke WATANABE
(JST/CREST)

Kitagawa Data Engineering Laboratory,
Center for Computational Sciences, University of Tsukuba
October 31, 2007

Background

- Various types of information sources
  - Data formats, access methods, and query languages
    - Structured data (RDB), Semi-structured data (XML), Plain text
    - Pull-based system, Push-based system
- Information integration is quite important
  - Provides an uniform access method to users



筑波大学 University of Tsukuba

Sensing Devices in Real World

Sensor-network

Mote (Crossbow)

AirSense (Hitachi)

Location Information

AirLocation (Hitachi)

Ekahau (Ekahau)

GPS

RFID (OMRON)

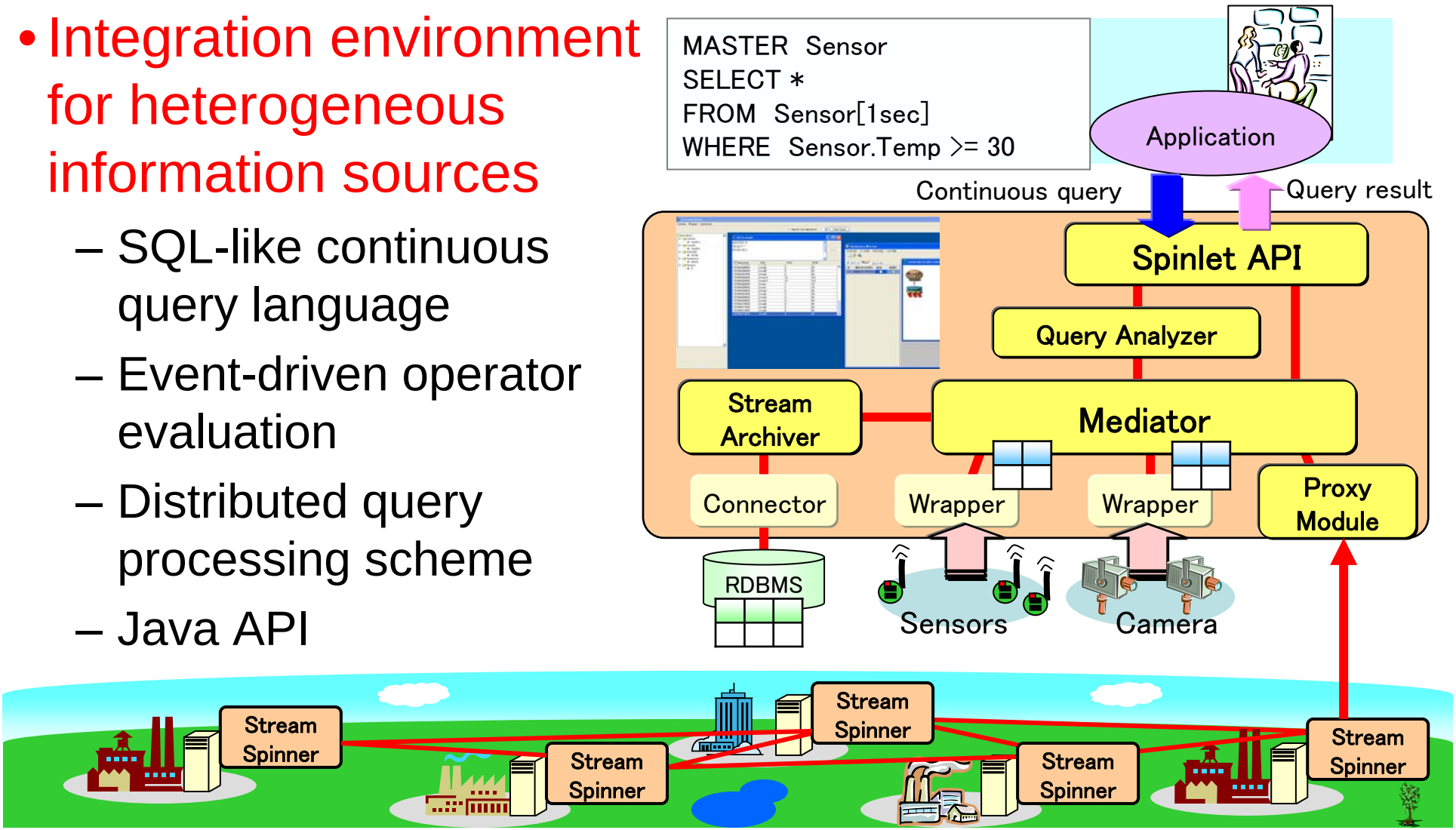
Camera

Micro phone

筑波大学 University of Tsukuba

StreamSpinner

- Integration environment for heterogeneous information sources
  - SQL-like continuous query language
  - Event-driven operator evaluation
  - Distributed query processing scheme
  - Java API



Example 1: simple filtering

Deliver sensor data when the temperature sensor(ttxd11) becomes greater than 25℃

MASTER Turbine

SELECT timestamp, fsr
FROM Turbine[1]
WHERE Turbine.ttxd11 > 25

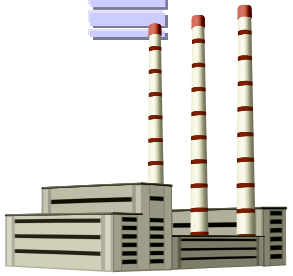
timestamp	ttxd11	ttxd12	tnhrpm	fsr
9867664…	23.3	23.9	1234	87.2
9867664…	23.4	24.1	1234	87.3
9867664…	23.6	24.1	1235	87.5
9867664…	24.1	24.4	1244	87.8

Turbine stream

(ttxd11 expresses a value of the temperature sensor)

Temperature
RPM
…

Sensors monitor behavior of a gas turbine



Plants
(Gas turbines)

Example 2: integration of stream and database

筑波大学 University of Tsukuba

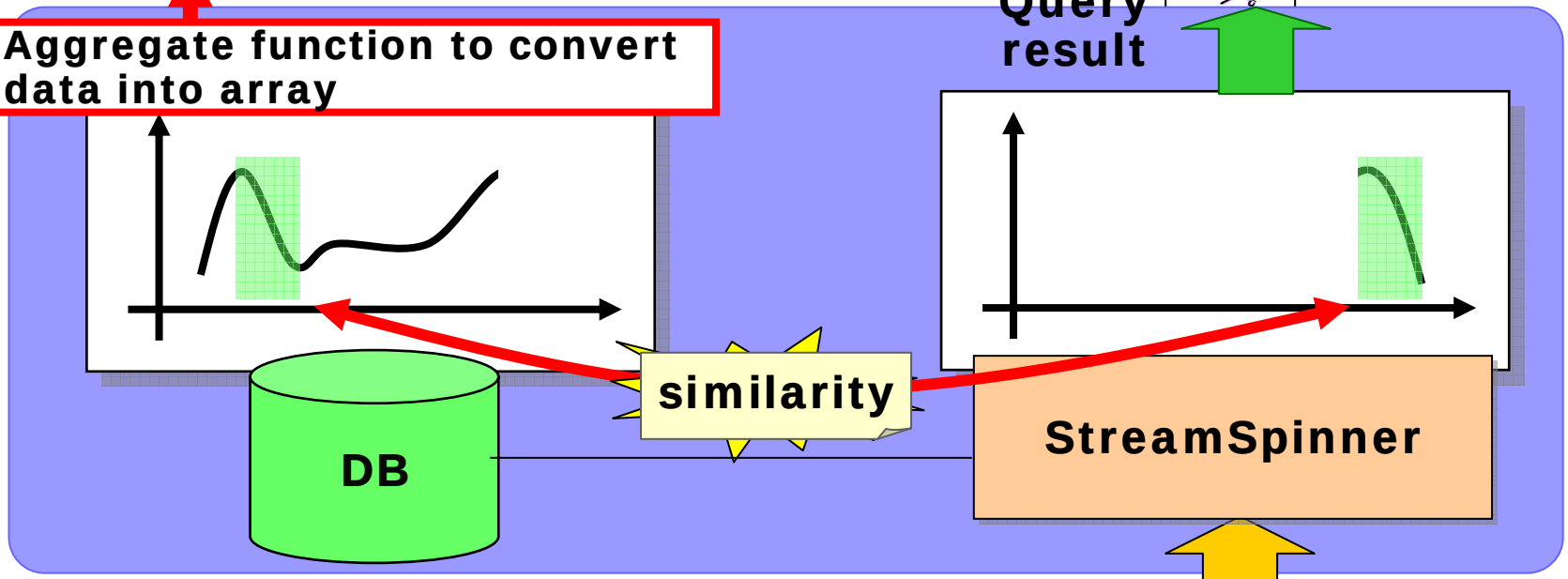
```
MASTER Clock_1minute
SELECT dist(S1.v, S2.v)
FROM
( SELECT array(value) AS v  FROM Turbine[60s]) AS S1,
( SELECT array(value) AS v  FROM TurbineDB) AS S2
```

Function to compute similarity of two arrays

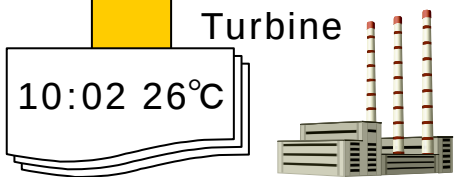
Report similarity between the current pattern and the pattern in the database

Aggregate function to convert data into array

Query result

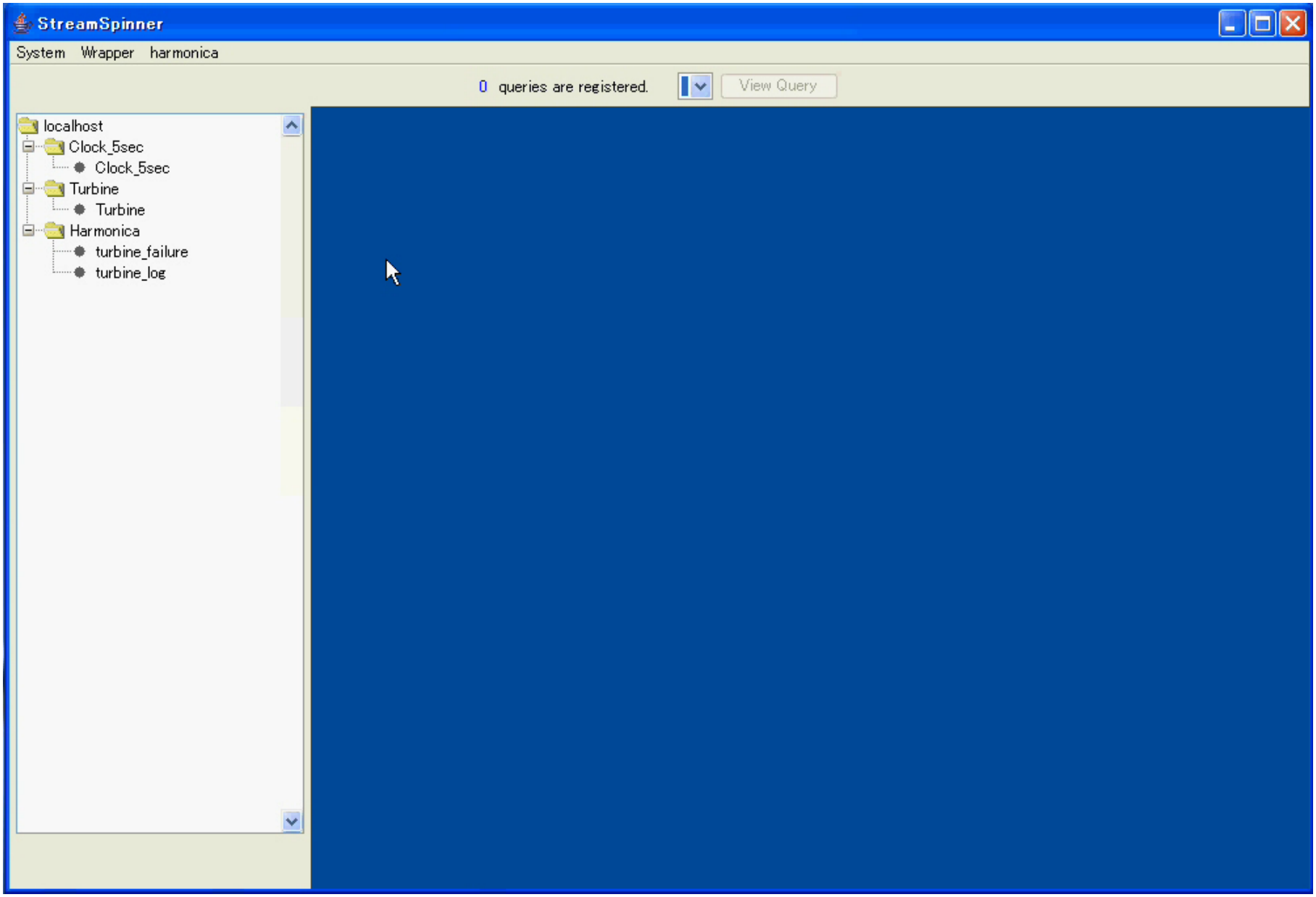


Historical sensor data stored in DB



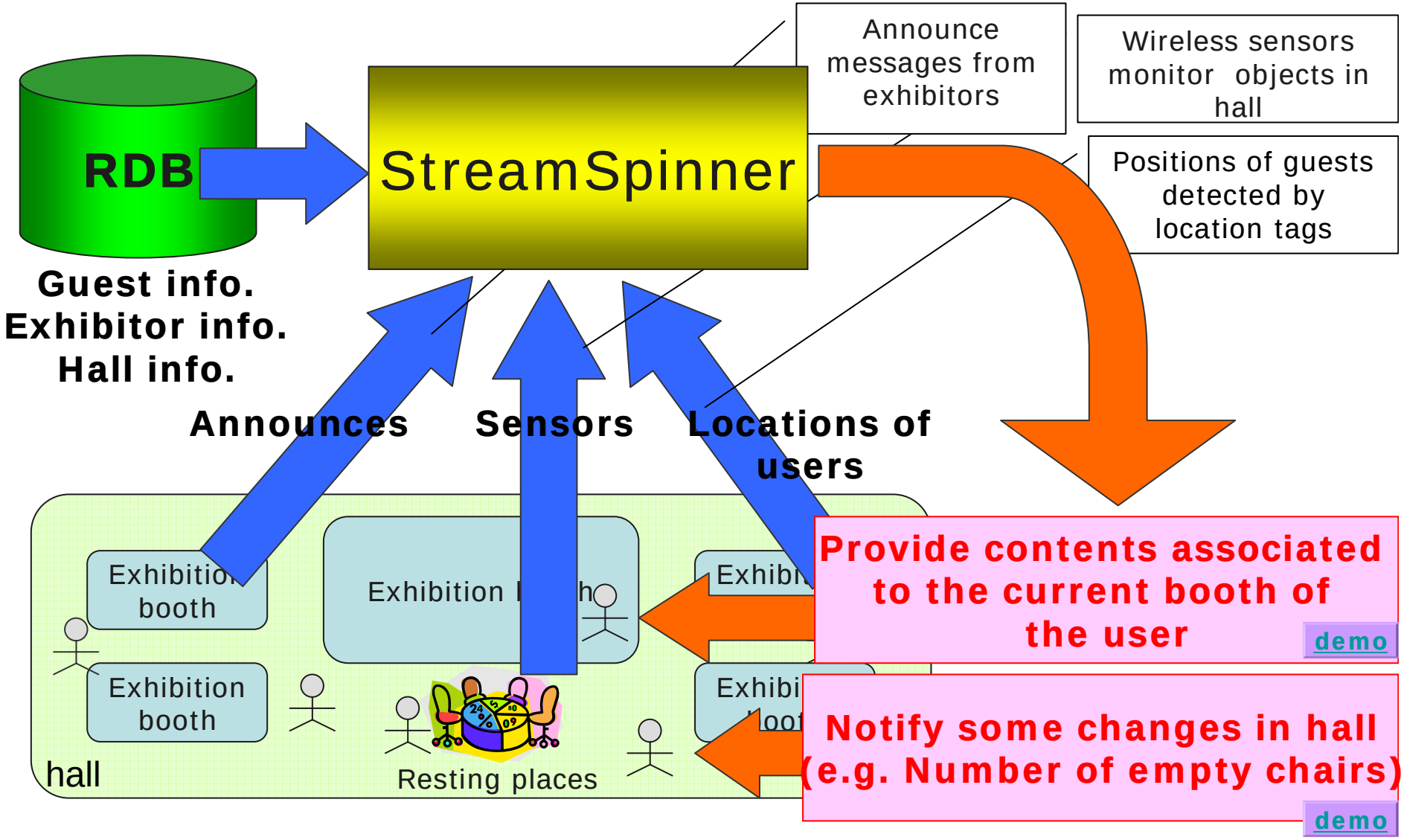
筑波大学
University of Tsukuba

Demonstration



筑波大学 University of Tsukuba

Information delivery in exhibition hall



Distributed Query Processing

Requirement collaborating multiple StreamSpinners

Execute a monitoring program which detects fire, and gather sensor data to the program

Application program

Query

```
MASTER  Sensor1
SELECT  *
   FROM  Sensor1 [1s], Sensor2 [1s], Sensor3[1s]
WHERE  Sensor1.temp > 50  AND Sensor2.temp > 50
          AND Sensor3.temp > 50
```

Definition of output stream

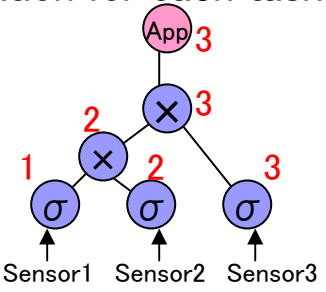
```
CREATE STREAM  Alarm (Bldg_id long, Temp long)
```

Application-specific task (written in java)

Receiving data, detecting fire, sending result

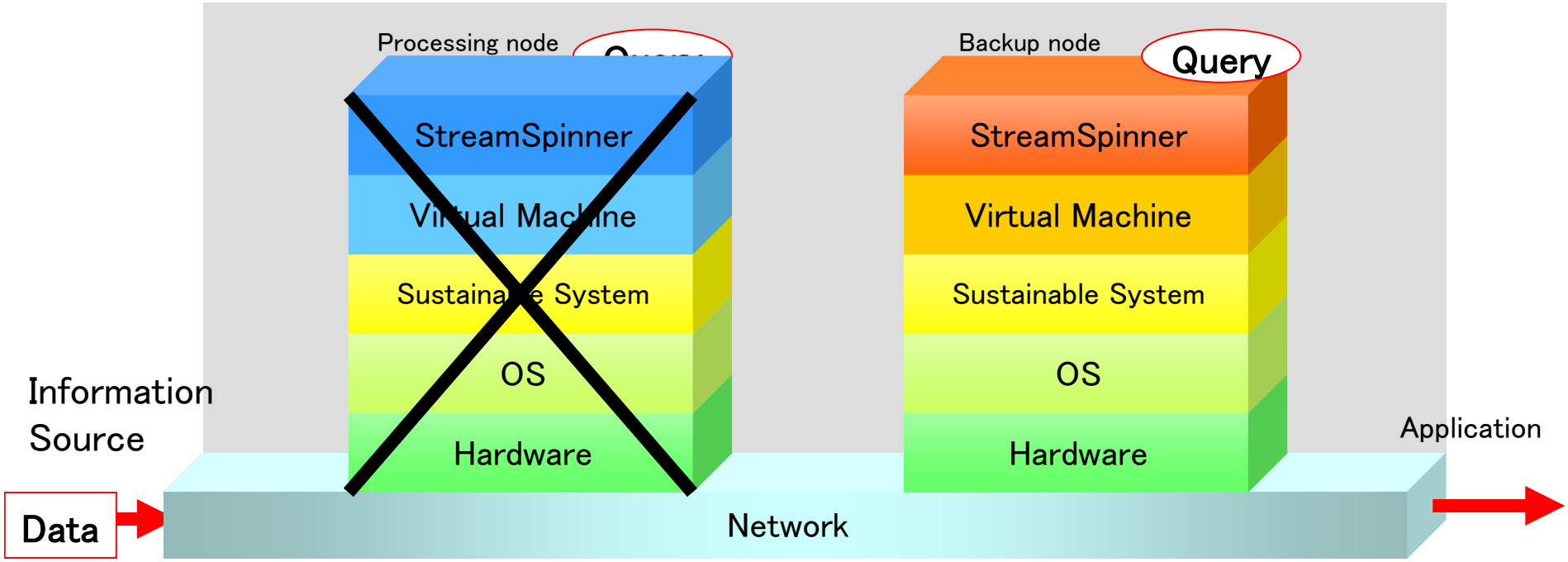
Management Tool for Distributed Environment

Deciding the optimal location for each task

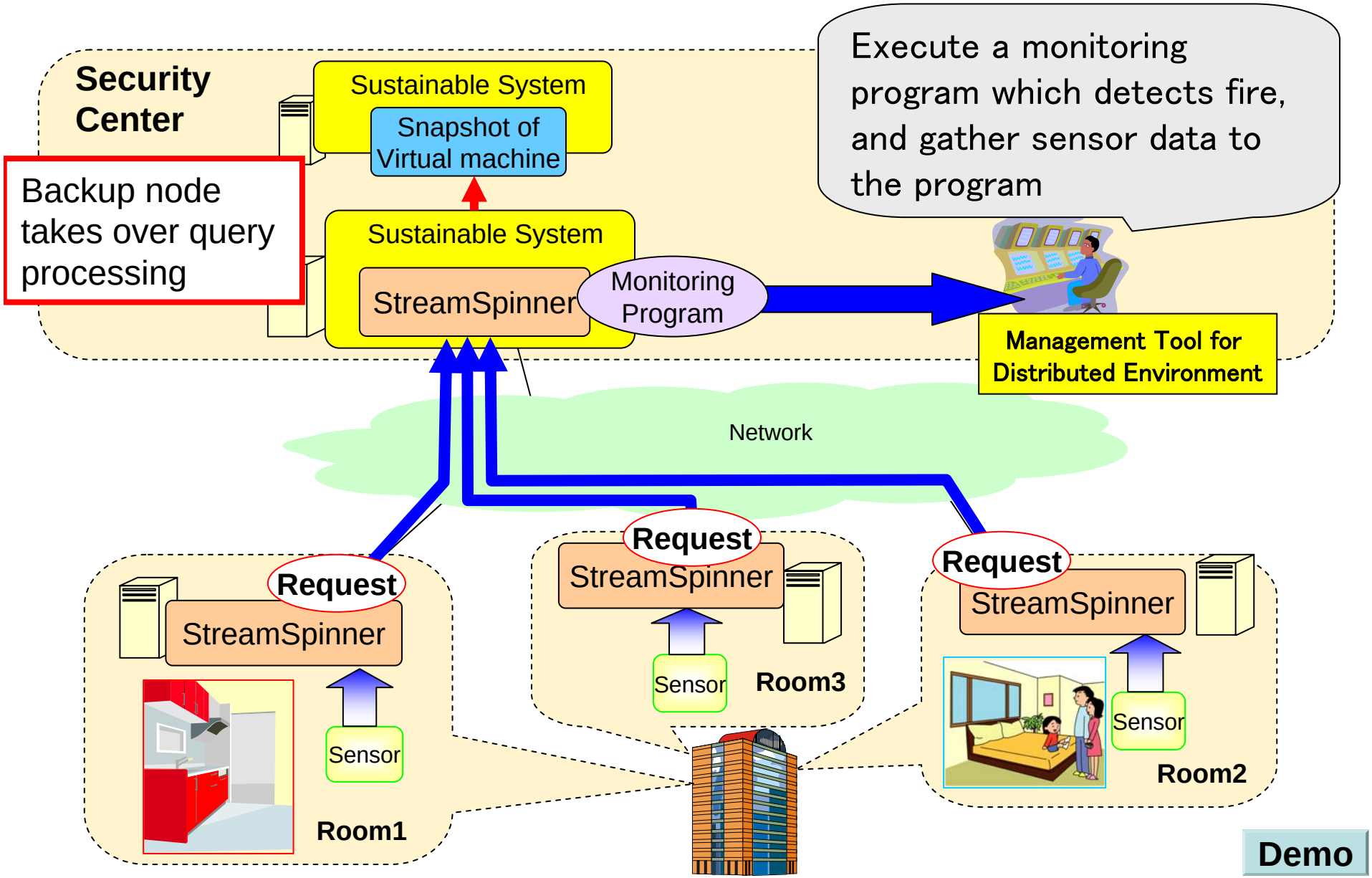


Sustainable Query Processing

- Dependable systems are required to execute queries in long term
- Collaboration with “Sustainable System” (developed by other group)
  - Sustainable system periodically creates a snapshot of running virtual machine, and distributes the snapshot to other nodes
  - When the node fails, backup node recovers the virtual machine from the snapshot



Example: monitoring remote buildings



筑波大学 University of Tsukuba

Conclusion

- StreamSpinner
  - Integration environment for heterogeneous information sources
- Future work
  - Experimental evaluation about distributed query processing
  - Development of real applications
  - Open source