Towards Integration of Big Data Analytics in Internet of Things Mashup Tools

Tanmaya Mahapatra, Ilias Gerostathopoulos & Christian Prehofer

Technische Universität München,
Fakultät für Informatik,
Software- and Systems Engineering Research Group

The Seventh International Workshop on the Web of Things
Contents

1. Internet of Things (IoT) & Application Development
2. Mashup & Mashup Tools
3. Big Data Analytics Tools
4. Why integration is necessary?
5. Challenges
6. Discussion
IoT & Application Landscape

*IoT has been defined as the interconnection of ubiquitous computing devices for the realization of value to end users.*

- The true potentials of IoT are realized only by its application landscape.
- Data collection from devices:
  - For analysis to better understand the environmental context.
  - Task automation for time optimization and enhancing the quality of human life.
Usage of IoT Data

*Insights from data can allow the creation of sophisticated & high-impact applications.* E.g. *Traffic congestion can be avoided by using learned traffic patterns.*

- Mining individual and group preferences
- Mining patterns of end-users (mobility models, …)
- Analyzing the state of engineering structures (structural health monitoring, …)
- Predicting the future state of the physical environment (flood prediction in rivers, …)
Difficulties in IoT Application Development

- Unfortunately, the development of software application for IoT is not simple.

- Write complex boiler plate codes to access devices
- Perform data mediation
- Device identification & Co-ordination
Streamlining Application Development for IoT: Mashups

A mashup is a composite application developed starting from reusable data, application logic, and/or user interfaces typically sourced from the web.
IoT Mashup Tools

- **Mashup Tools** typically offer graphical interfaces for specifying the data flow between **sensors, actuators & services** which **lowers** the **barrier** of creating IoT applications for end-users.
IBM Node-RED

- **Node-RED** is an open-source mashup tool developed by IBM.
- Provides a GUI where users drag-and-drop **blocks** that represent **components** of a larger system which can either be **devices**, **software platforms** or **web services** that are to be connected.
- These blocks are called **nodes**.
- A **node** is a visual representation of a block of **JavaScript code** designed to carry out a specific task.
- Additional **blocks(nodes)** can be placed in between these components to represent software functions that house **business logic/data mediation logic**.
- With Node-RED the **time** and **effort** spent on writing **boilerplate code** is greatly **reduced**, and the developer can focus on the **business logic**.
Example of an IoT Mashup in Node-RED

Mashup Design of a custom Intrusion Detection System from firewall logs
A number of matured tools have emerged to assist in the analysis of huge data sets.

Employ parallelized data analysis and machine learning algorithms to operate on data sets that reside in large clusters of commodity machines in a cost-effective way.

Big Data Analytics:

- manipulating and querying in parallel large amounts of data residing in clusters (batch mode)
- accommodating and analyzing large amounts of incoming data as they come (streaming mode)
Prominent Big Data Analytics Tools

HIVE

USE database;
SELECT from_columns FROM table WHERE conditions;

Pig

A = load 'passwd' using PigStorage(':');
B = foreach A generate $0 as id;
store B into 'id.out';
Integration of Both Worlds: Focal Point

• IoT and Big Data have stood apart from each other.
• IoT is used for collecting data into storage and writing business logic while Big Data for analysis.
• However, in many scenarios, business logic of mashups may need input from Big Data jobs i.e. mashups may require to trigger data analysis. Similarly, after the execution of Big Data jobs there may arise need to perform some additional task i.e. may need to trigger a flow in mashup.

Full fledged integration would enable mashup developers to specify Big Data analytics jobs and consume their results within a single application model.
Mashup involving Big Data Analytics for Traffic Management
Mashup involving Big Data Analytics for Traffic Management
Mashup involving Big Data Analytics for Route Optimization
Challenges

- Blocking execution and synchronous communication in mashups
- Single-threaded mashups
- Visual Programming Limitations
- End-user focus in mashup tools
- Lack of RESTful APIs for Big Data Analytics
Concluding Remarks

**Focus:** Full blown Integration can enhance the development of IoT mashups that continuously harness the value out of sensed data in their operation.

**Problems:** Integration entails enhancement of existing mashup as well as big data tools.

**Current Agenda:** Work on the challenges, starting from lifting the limitations of existing mashup tools that stand in the way of integration with Big Data analytics.

Thanks -- Questions? Ideas?