The NebulaStream Platform: Data and Application Management for the Internet of Things

Steffen Zeuch, Ankit Chaudhary, Bonaventura Del Monte, Haralampos Gavriilidis, Dimitrios Giouroukis, Philipp M. Grulich, Sebastian Breß, Jonas Traub, Volker Markl
Our View the IoT:

The Internet of Things (IoT) presents a novel computing architecture for data management: A geo-distributed, highly dynamic, and heterogeneous environment of massive scale.
NebulaStream

An IoT data management system to process thousands of queries over millions of sensors.
Three questions that everyone should answer before creating a new system

(Is there a need for an IoT system?)
1) What's the **new** thing about IoT applications?

- Continuous Operation under Constant Evolution
- Highly Dynamic Execution Environment
- Heterogeneity, Distribution, and Volume of Data and Compute

Legend:
- Control Flow
- Base Station
- Data Flow
- Processing Node
- Vehicle/Sensor Node
- Disconnected Vehicle
2) Is there even a market for an IoT system?

By 2020, Gartner estimates internet-connected things will outnumber humans 4-to-1, creating new dynamics for marketing, sales and customer service.

Leading the IoT (Gartner 2017)

The Internet of Things offers a potential economic impact of $4 trillion to $11 trillion a year in 2025.

<table>
<thead>
<tr>
<th>Nine settings where value may accrue</th>
<th>Size in 2025, $ trillion*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factories</strong>—eg, operations management, predictive maintenance</td>
<td>Low estimate: 1.2  High estimate: 3.7</td>
</tr>
<tr>
<td><strong>Cities</strong>—eg, public safety and health, traffic control, resource management</td>
<td>Low estimate: 0.9  High estimate: 1.7</td>
</tr>
<tr>
<td><strong>Human</strong>—eg, monitoring and managing illness, improving wellness</td>
<td>Low estimate: 0.2  High estimate: 1.6</td>
</tr>
</tbody>
</table>

Unlocking the potential of the Internet of Things (McKinsey Report 2015)
3) Can’t we just use existing (cloud-)solutions?

There is no general-purpose, end-to-end data management system for the IoT with similar functionality as Flink or Spark.
The answers to the questions are:

Yes IoT is something new

Yes there is a big market/demand

No we cannot just use existing systems
What are the new research challenges of IoT?
Query Submission in the IoT

**C1:** Come up with any plan within a reasonable amount of time

**C2:** Maintain an overview of the topology

**C3:** Find a suitable base for decisions

**C4:** Partial re-optimization and alternative plans

[Diagram showing Global Query Plan, Execution Plan, and IoT Topology]

Do we really have to chase the chickens **centrally**?

**NebulaStream**: A hybrid approach to meet different requirements.

https://www.youtube.com/watch?v=TVF2sQ10vaQ

C: Provide guarantees in this environment (e.g. ACID)
Query execution in a **highly dynamic** environment

Failures are the common case in IoT.

**NebulaStream**: Autonomous node design:

- Local
- Neighborhood
- Global

https://homediy.eu/schuetzen-sie-ihr-huehner-vor-raubtieren/
How can we cope with a highly **heterogeneous** IoT?

**NebulaStream**: Hardware-tailored code generation and a dynamic execution engine.
How to handle “IoT Scale”? 

Key Principles: 
- Reduce complexity 
- Reduce data volumes 
- Use resources efficiently 

NebulaStream: A system for a unified sensor-fog-cloud environment.
Summary

Yes we need a new system

NebulaStream
A general-purpose, end-to-end data management system for the IoT.
http://www.nebula.stream