# Decentralized Query Processing in Data Management Systems for the IoT
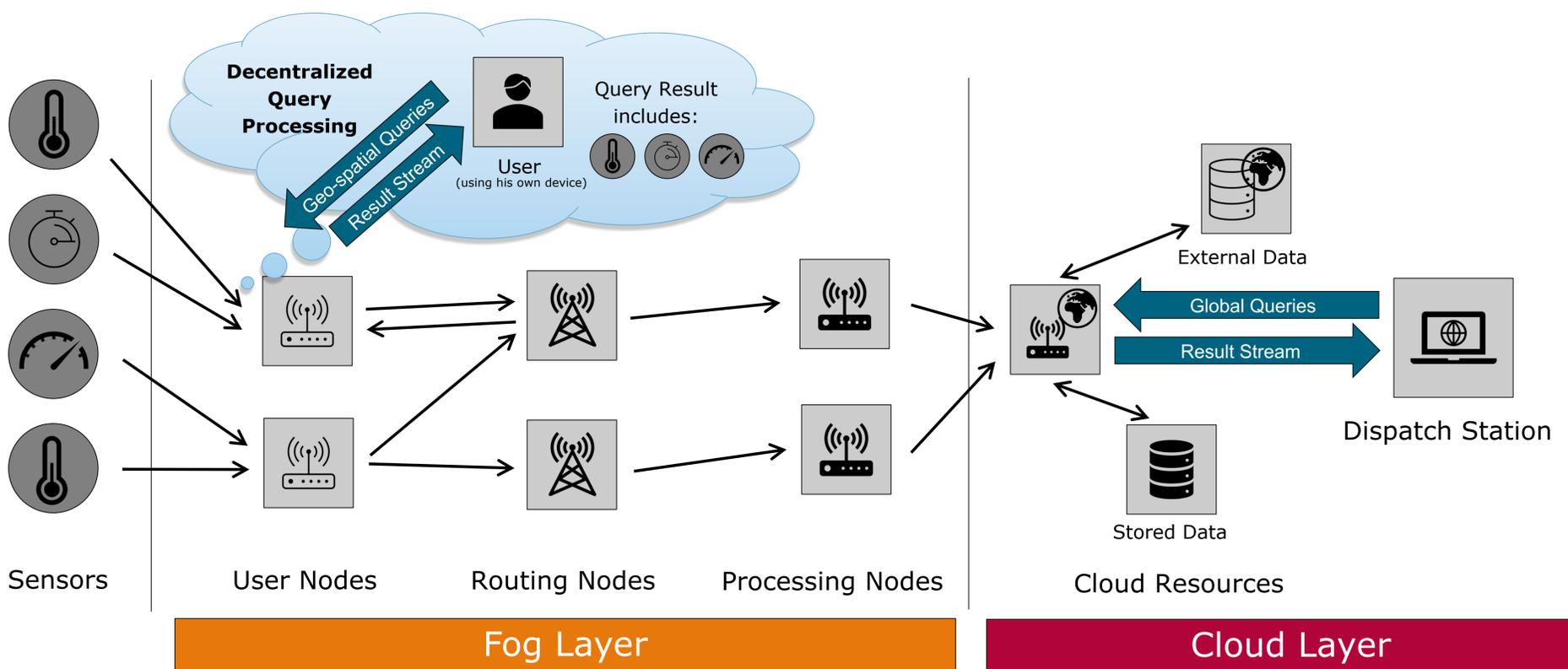
## Challenges in the IoT

The Internet of Things (IoT) combining large data centers with edge devices presents a novel computing architecture for data management: a distributed, highly dynamic and heterogeneous environment of massive scale. Applications for this architecture face the challenges of integrating fog and cloud computing and unifying sensor networks into one environment.

## Connection to Lecture Series on Database Research

The NebularStream Platform (NES) in the lecture "Data Infrastructures" by Volker Markl introduces a general purpose, end-to-end data management system for the IoT. It provides solutions for the challenge of heterogeneous and distributed computation and data and supports diverse data and programming models, deals with potentially unreliable communication and enables constant evolution under continuous operation [1].

## Topology of proposed data management system for the IoT



Decentralized Query Processing · Geo-spatial Queries · Result Stream · User (using his own device) · Query Result includes:

Sensors · User Nodes · Routing Nodes · Processing Nodes · Cloud Resources · External Data · Global Queries · Result Stream · Dispatch Station · Stored Data

**Fog Layer** · **Cloud Layer**

## Problem

Technologies exist addressing the challenges of data management in IoT. While many of them are not fully exploiting the potential of combining fog and cloud computing, the novel NES technology presents a solution that meets this criteria. At the moment, it allows users to insert queries at one centralized dispatch station. This centralized interface leaves unused potential in decentralized query processing.

## Goal

A data management system for IoT should enable users to send queries directly from their own devices. Each device should be able to process queries using available data, which can result in only including nearby data and thereby in geo-spatial query processing. For global queries, the cloud layer must still be available.

## Solution

A data management system for the IoT following the design principles of NES combining fog and cloud computing and allowing decentralized query processing must be researched.

The architecture of this system needs to support query processing on every user node in the fog layer. This implicitly demands the design of a node engine supporting such a behavior and the transfer of data from near user nodes to the node executing a geo-spatial query.

[1] Zeuch, S. et al. (2020). *The NebulaStream Platform: Data and Application Management for the Internet of Things.* In: 10th Biennial Conference on Innovative Data Systems Research (CIDR '20) January 12-15, 2020, Amsterdam, Netherlands.

**Fabian Lange**
Student M. Sc. Data Engineering

Lecture Series on Database Research
Hasso Plattner Institute, Potsdam, Germany

Universität Potsdam

HPI Hasso Plattner Institut