FOG-engine: Towards Big Data Analytics in the Fog

Farhad Mehdipour

Tech Futures Lab & Unitec Institute of Technology
Auckland, New Zealand
farhad@techfutureslab.com

Bahman Javadi

Western Sydney University,
Sydney, Australia

Aniket Mahanti

University of Auckland,
Auckland, New Zealand

The 2nd International Conference on Big Data Intelligence and Computing, Auckland, New Zealand, August 8-10, 2016

Introduction

- **Challenges of the current cloud-based platforms**
  - The cloud physically located in a distant datacenter

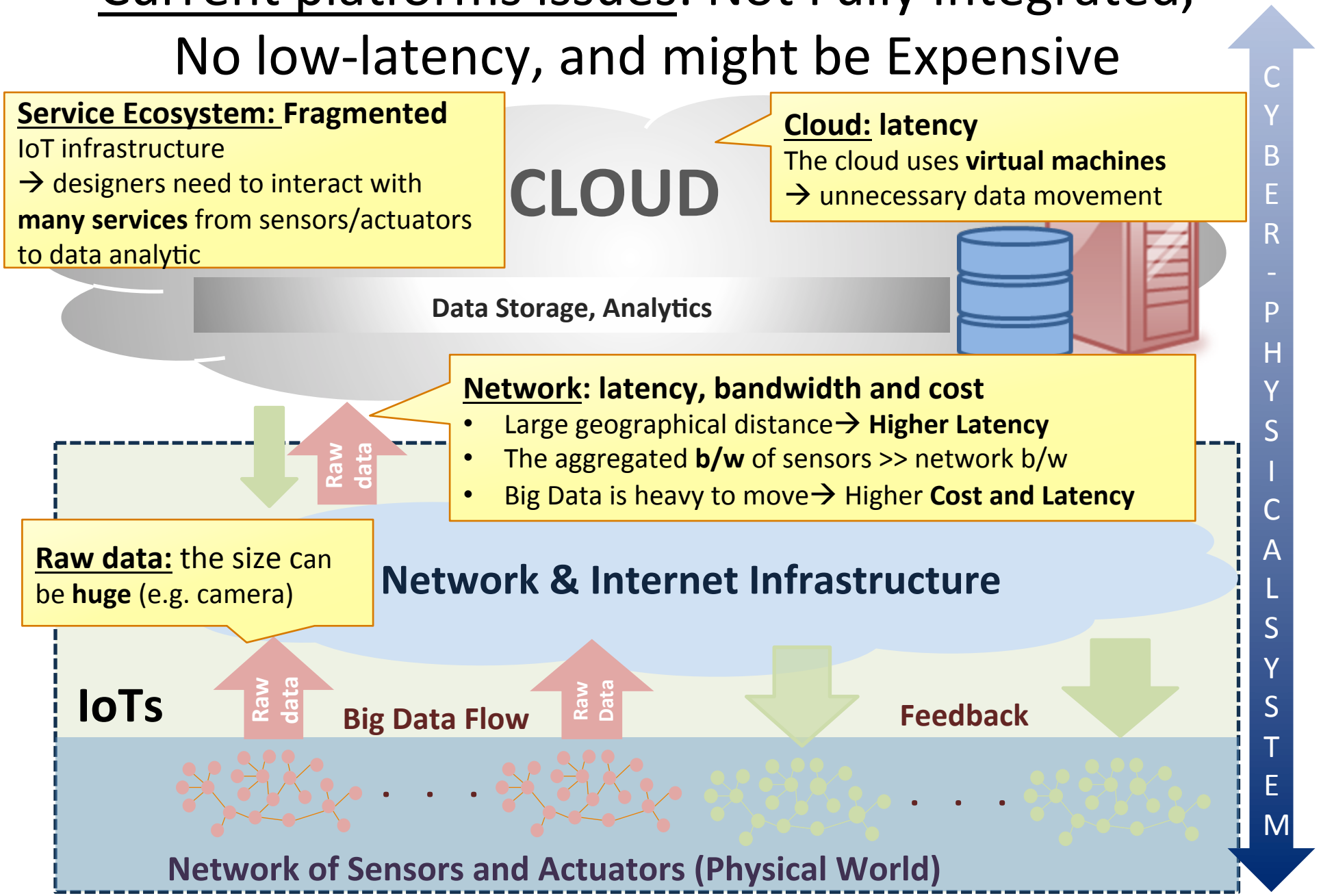
    → Latency
  - Vertically fragmented
  - Real-time processing large quantities of IoT data

    → more security, capacity, and analytics challenges
  - Incapability of current cloud for efficient Big Data Analytic

Our solution

- An on-premise and real-time data analytic engine (FOG-Engine) located near where data is generated
- Collaboration and proximity interaction between IoT devices in a distributed and dynamic manner

Current platforms issues: Not Fully Integrated, No low-latency, and might be Expensive



Fog Computing

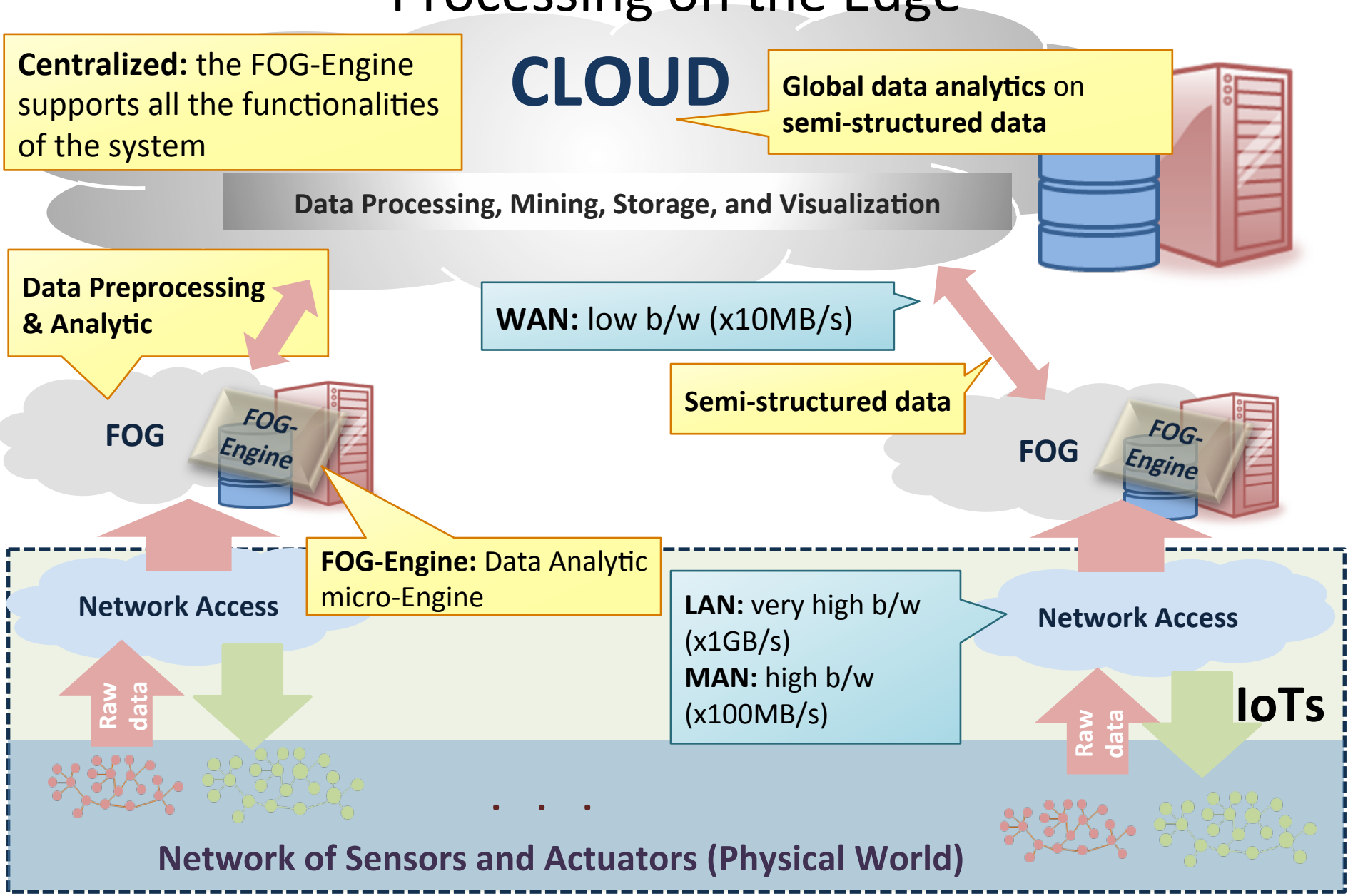
- **The Fog**
  - extends the cloud computing paradigm to the edge of the network,
  - enables a new breed of applications and services
  - an appropriate solution for the applications and services that fold under the umbrella of the IoTs.

- **Benefits**
  - low latency
  - location awareness
  - widespread geographical distribution
  - mobility support
  - the strong presence of streaming and real-time applications
  - heterogeneity

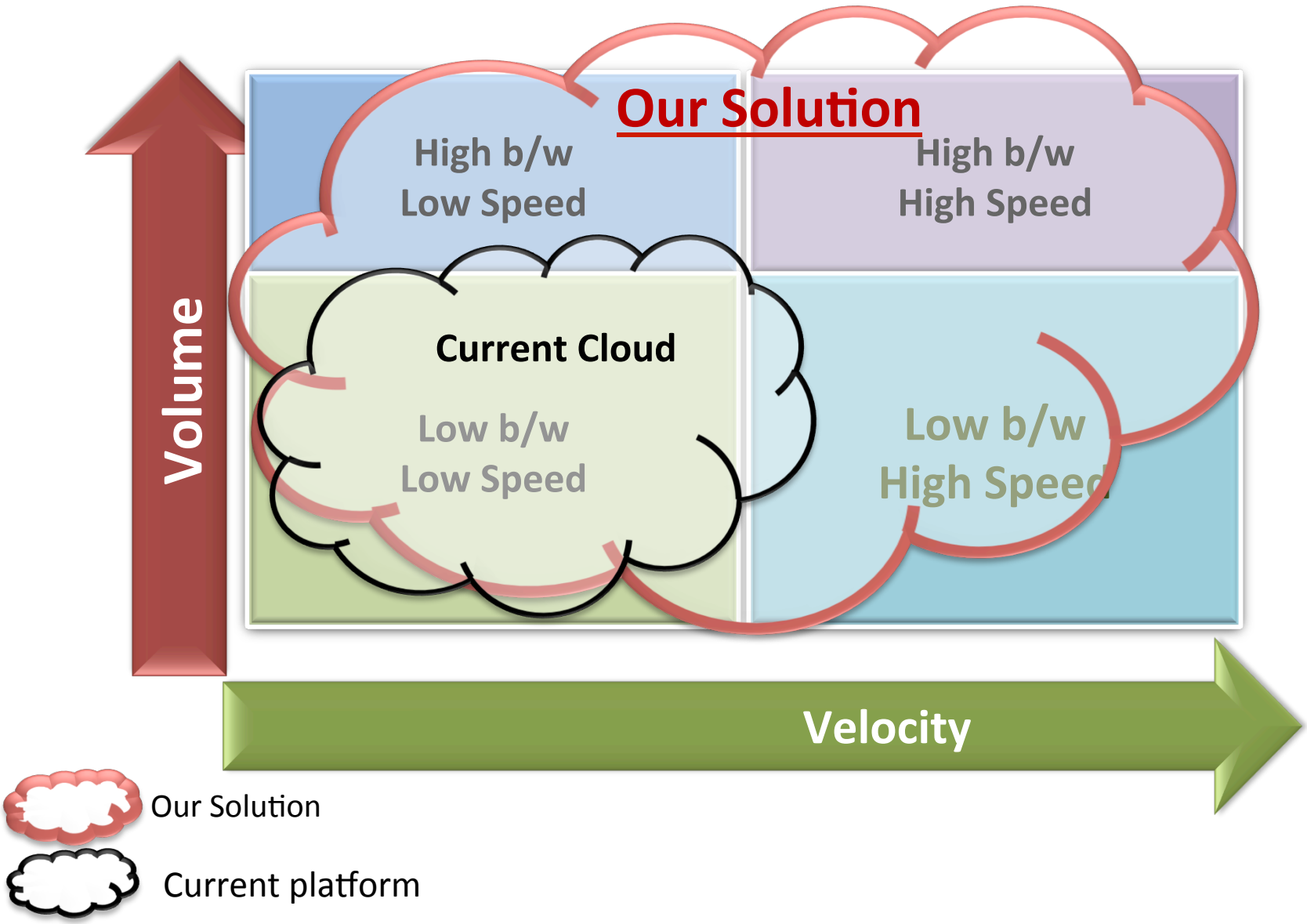
Related Works

	AWS	Microsoft	IBM	Google	Alibaba
Service	AWS IoT	Azure IoT Hub	IBM Watson IoT	Google IoT	AliCloud IoT
Data Collection	HTTP, WebSockets, MQTT	HTTP, AMQP, MQTT and custom protocols (using protocol gateway project)	MQTT, HTTP	HTTP	HTTP
Security	Link Encryption (TLS), Authentication (SigV4, X.509)	Link Encryption (TLS), Authentication (Per-device with SAS token)	Link Encryption (TLS), Authentication (IBM Cloud SSO), Identity management (LDAP)	Link Encryption (TLS)	Link Encryption (TLS)
Integration	REST APIs	REST APIs	REST and Real-time APIs	REST APIs, gRPC	REST APIs
Data Analytics	Amazon Machine Learning model (Amazon QuickSight)	Stream Analytics, Machine Learning	IBM Bluemix Data Analytics	Cloud Dataflow, BigQuery, Datalab, Dataproc	MaxCompute
Gateway Architecture	Device Gateway (in Cloud)	Azure IoT Gateway (on-premises gateway, beta version)	General Gateway	General Gateway (on-premises)	Cloud Gateway (in Cloud)

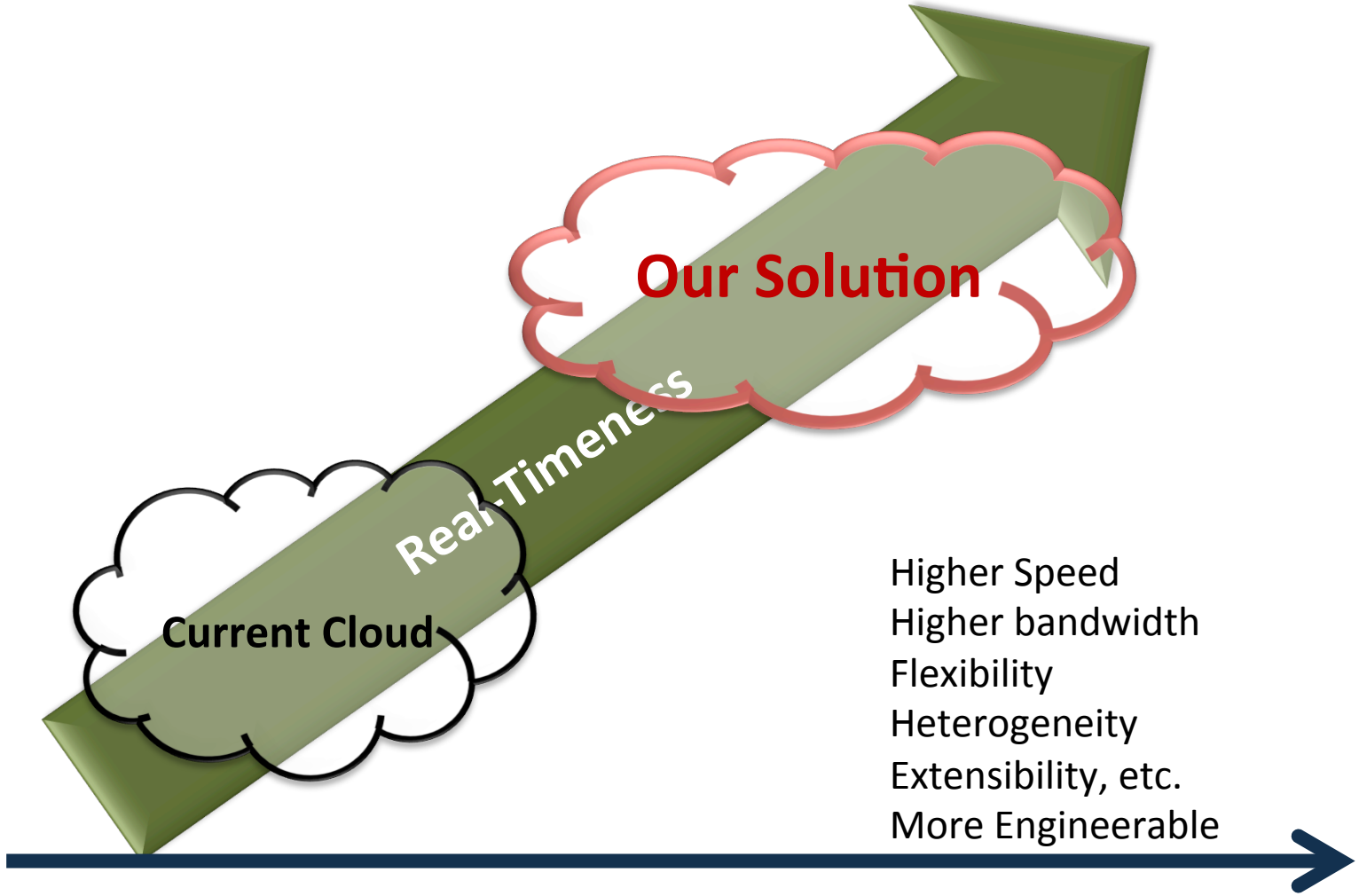
Our View: Decentralized Hierarchical Big Data Processing on the Edge



How to Realize Big Data's Vs: Velocity, Volume, …



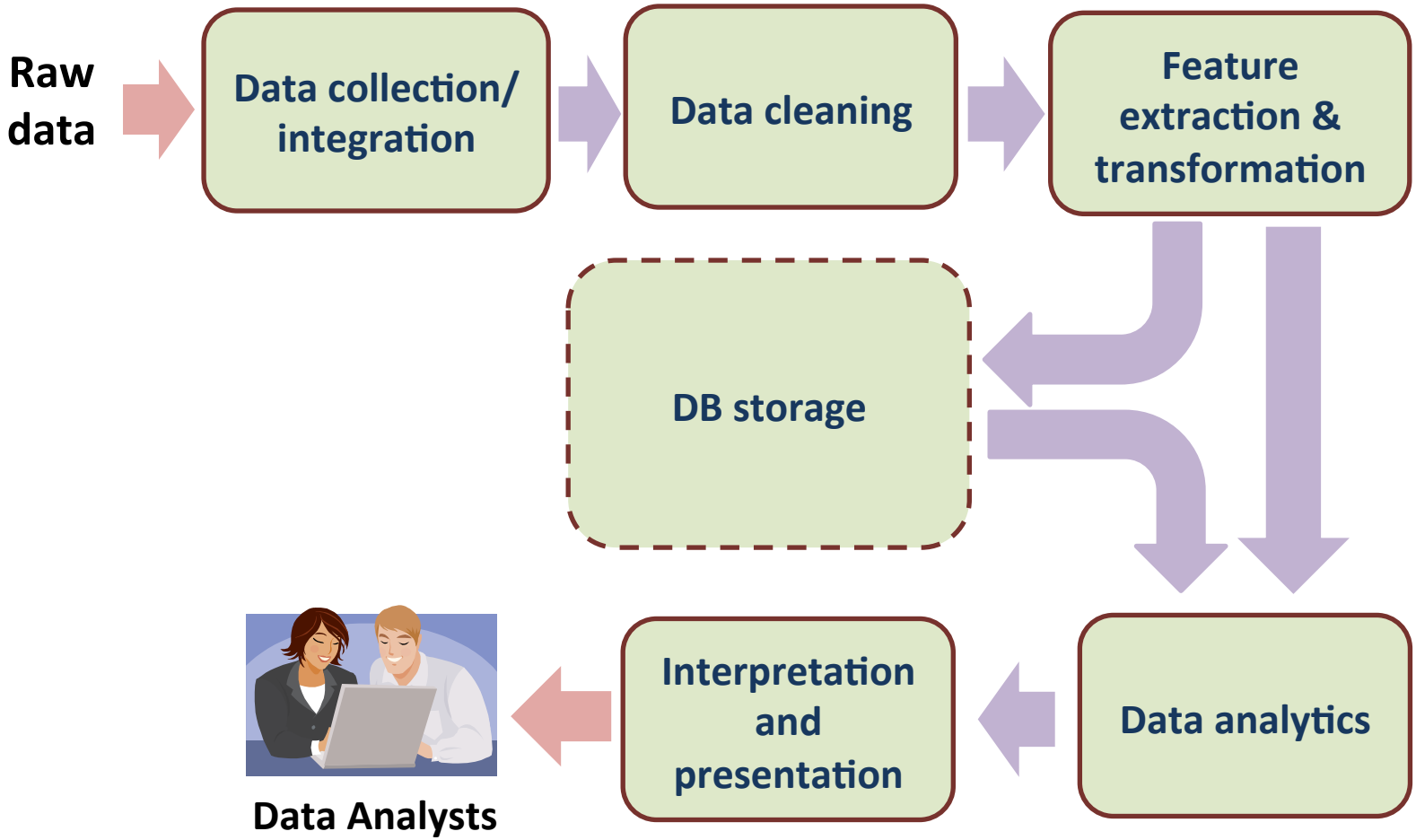
Towards Real-Timeness, with Added Values



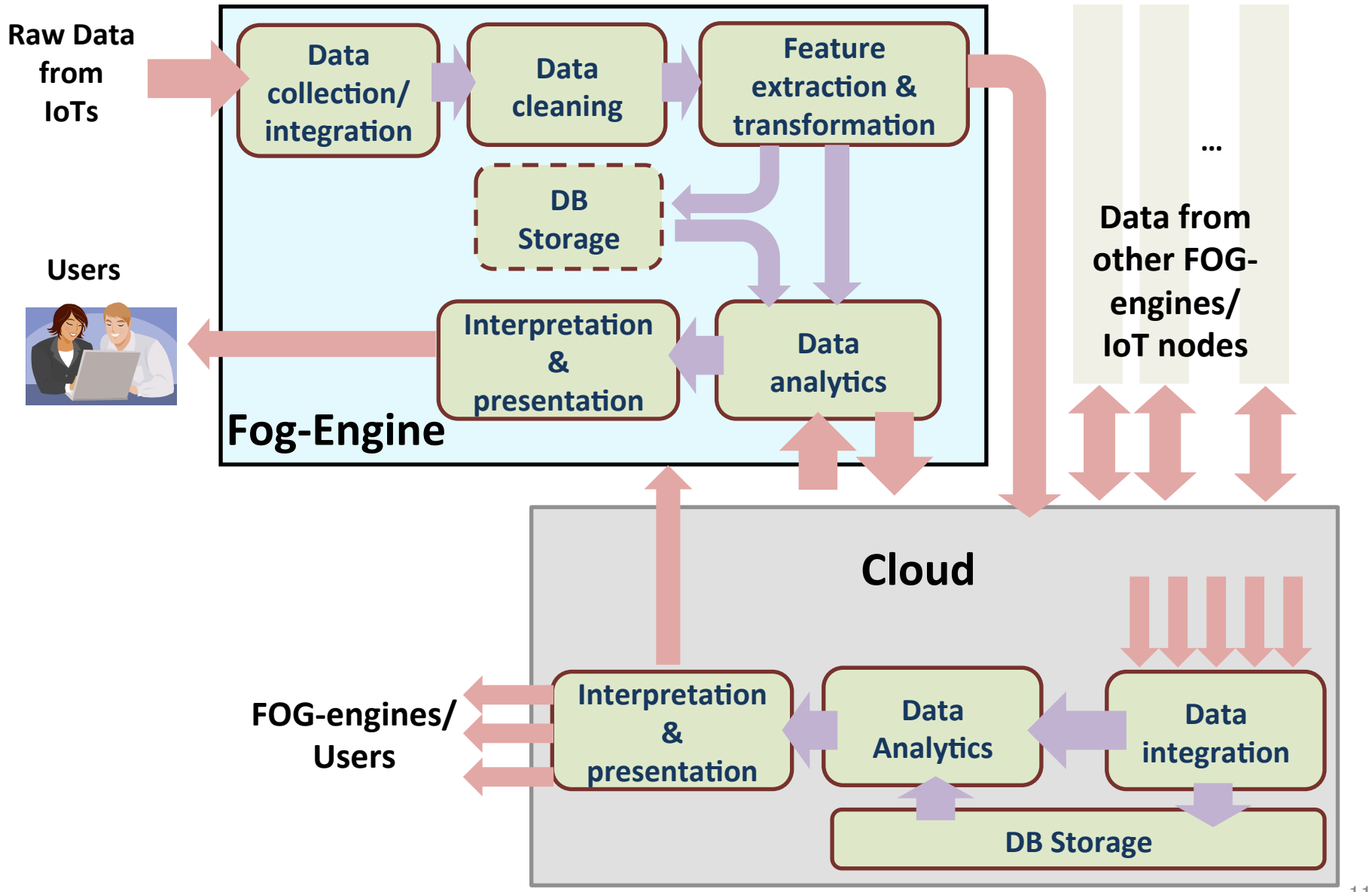
FOG-Engine vs. Cloud

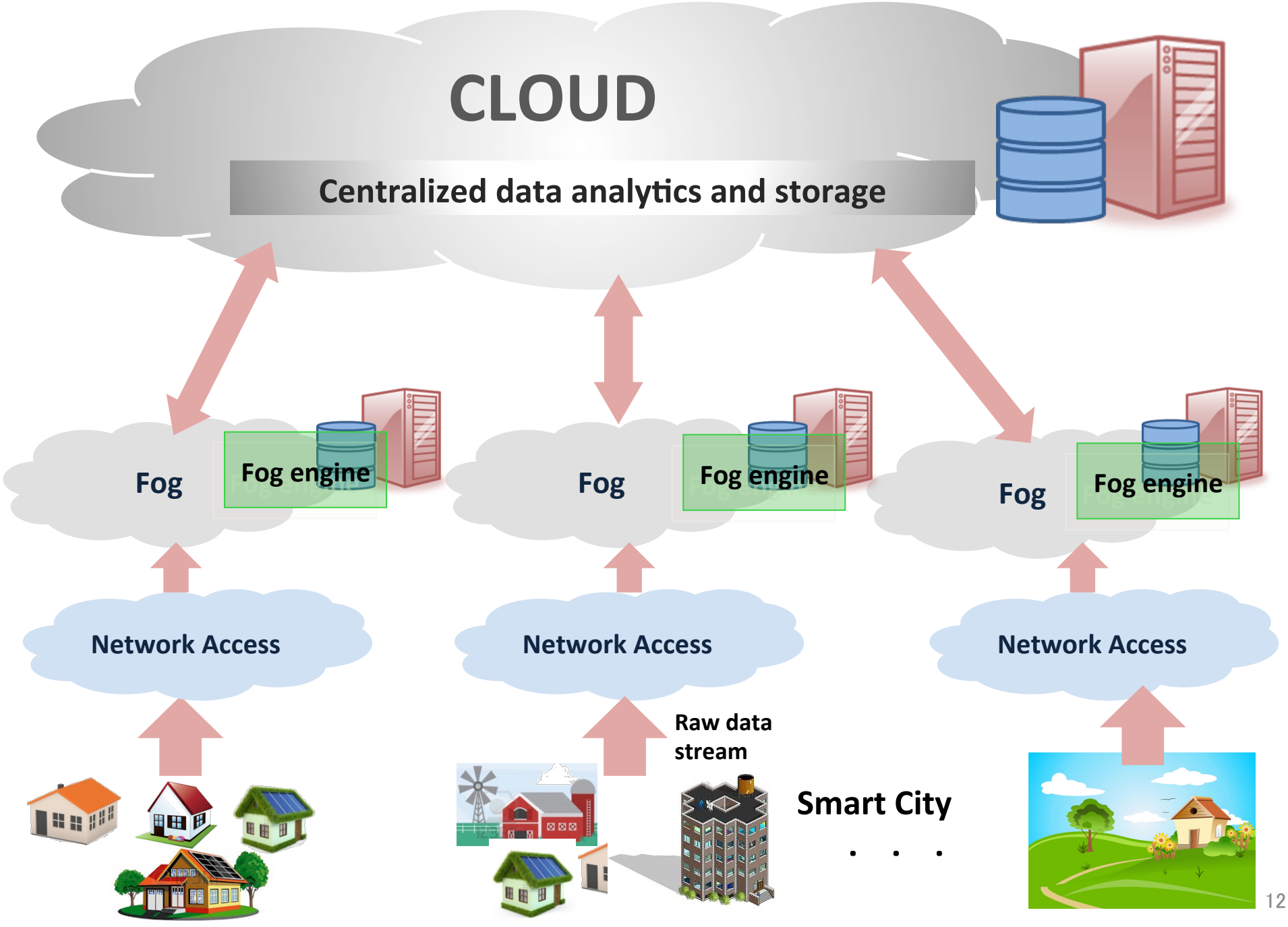
Characteristic	FOG-engine	Cloud platform
Processing hierarchy	Local data analytics	Global data analytics
Processing fashion	In-stream processing	Batch processing
Computing power	GFLOPS	TFLOPS
Network Latency	Miliseconds	Seconds
Data storage	Gigabytes	Infinite
Data lifetime	Hours/Days	Infinite
Fault-tolerance	High	High
Processing resources	Heterogeneous (e.g. CPU, FPGA)	Homogeneous (Data center)
Versatility	Only exists on demand	Intangible servers
Provisioning	Limited by the number of FOG-engines in the vicinity	Infinite, with latency
Mobility of nodes	May be mobile (e.g. in the car)	None

A Typical Data Analytic Flow

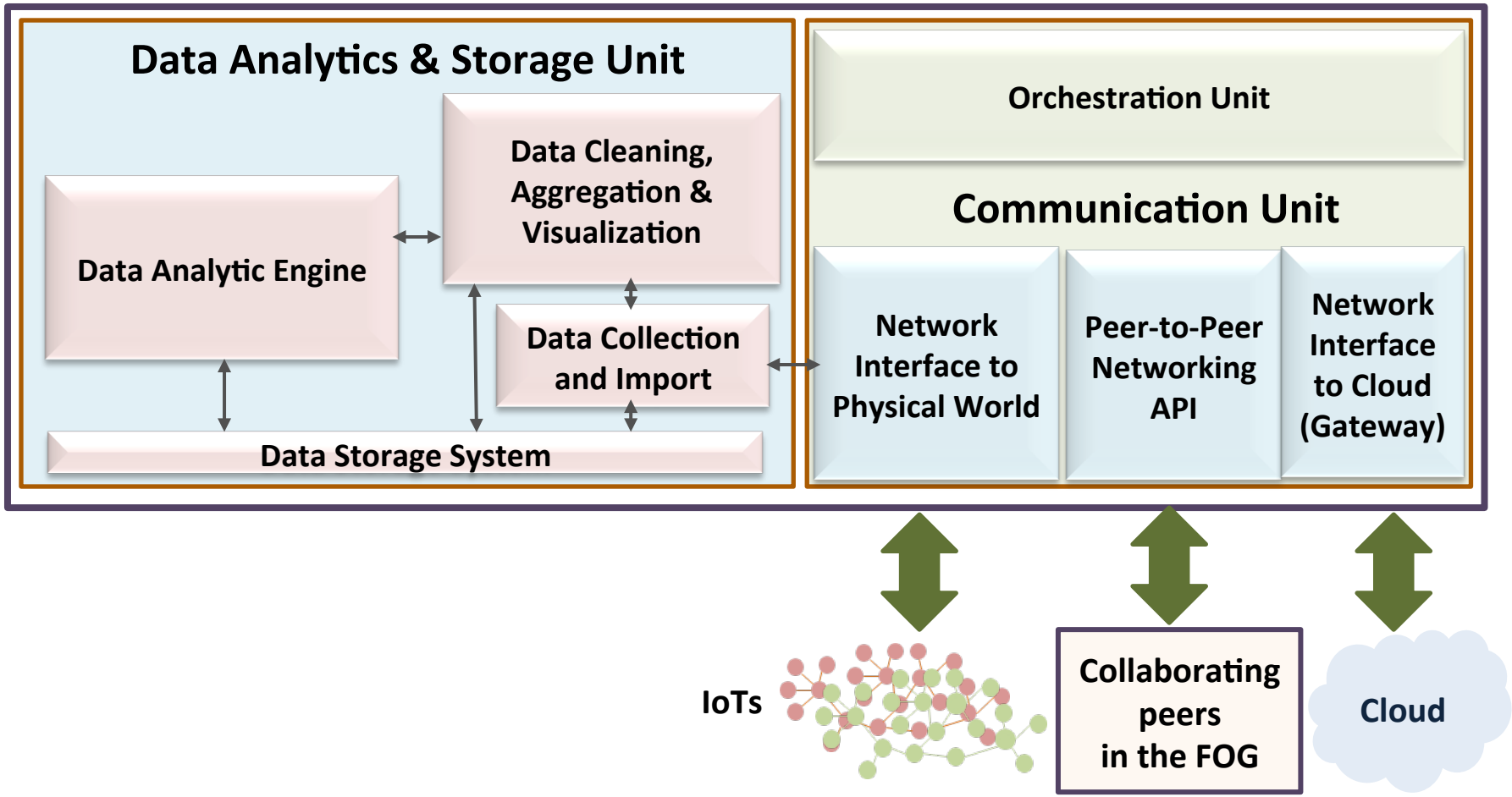


A Modified Data Analytic Flow

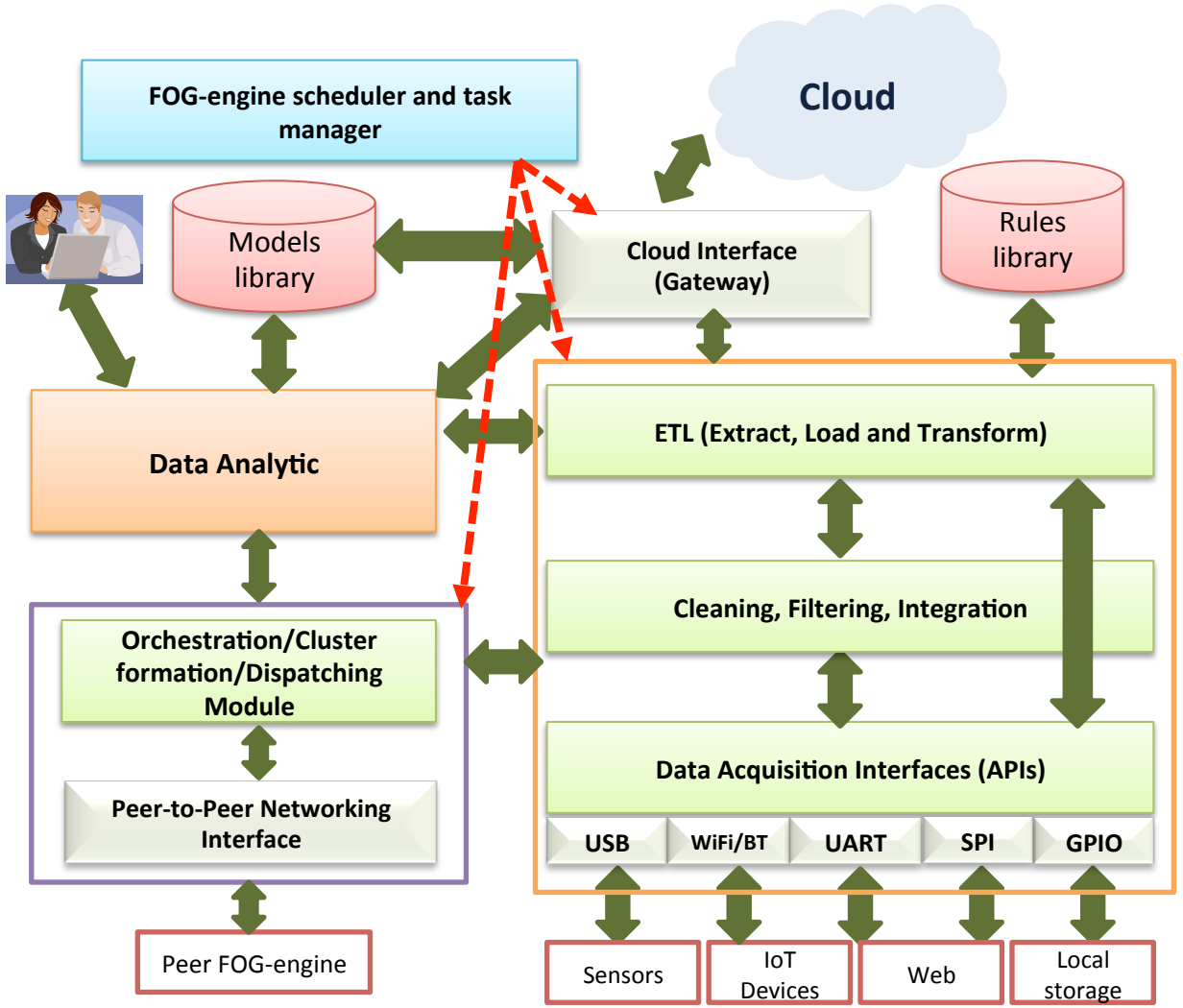



CLOUD
Centralized data analytics and storage
Fog
Fog engine
Fog
Fog engine
Fog
Fog engine
Network Access
Network Access
Network Access
Raw data stream
Smart City

General Architecture of FOG-Engine



Detailed Architecture of FOG-engine



USB: Universal serial bus
BT: Bluetooth
UART: Universal Asynchronous Receiver/Transmitter
SPI: Serial Peripheral Interface Bus
GPIO: General-purpose input/output pins

Preliminary Results

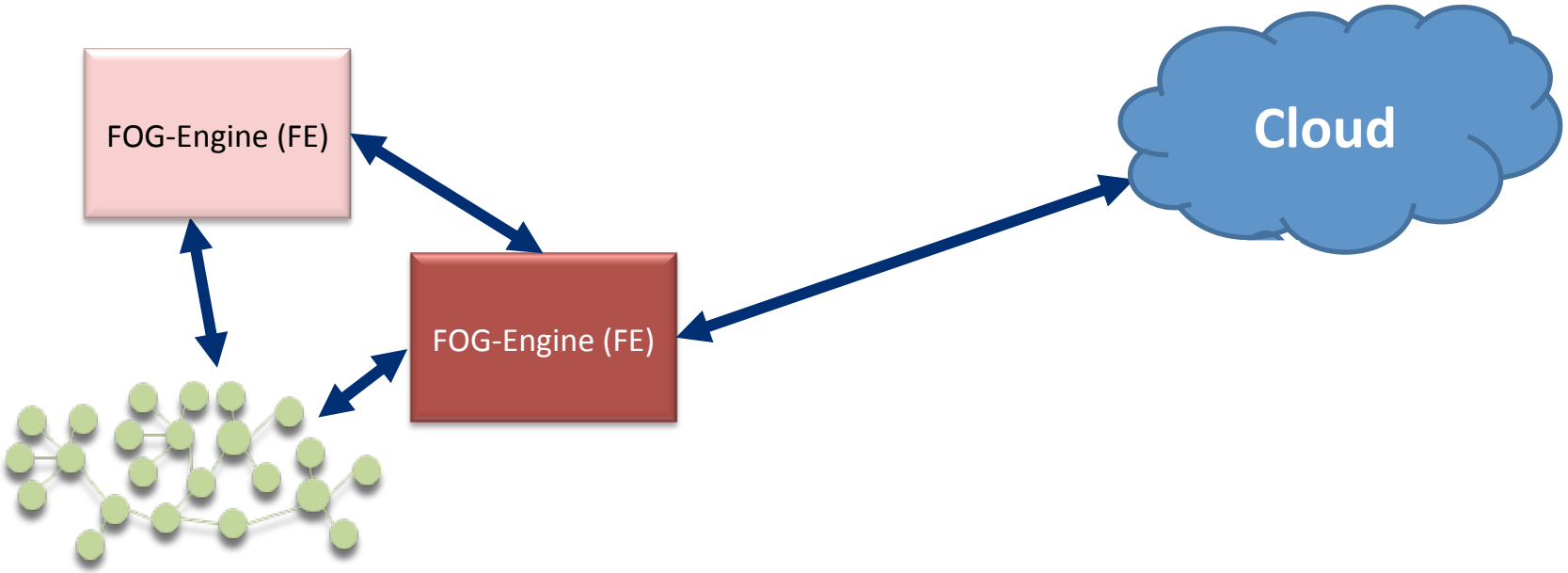
- Implementation platform: Raspberry Pi 2.0 and 3.0

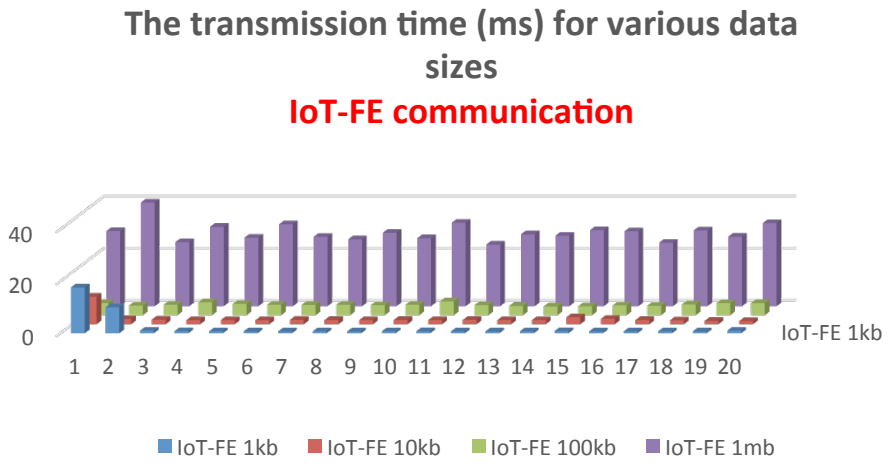
- Scenarios
  1) **Multiple receivers, multiple analysers, and multiple transmitters scenario**
  2) **Multiple receivers, multiple analysers, and single transmitter scenario**
  3) **Multiple receivers, single analyser, and single transmitter scenario**

Scenario II

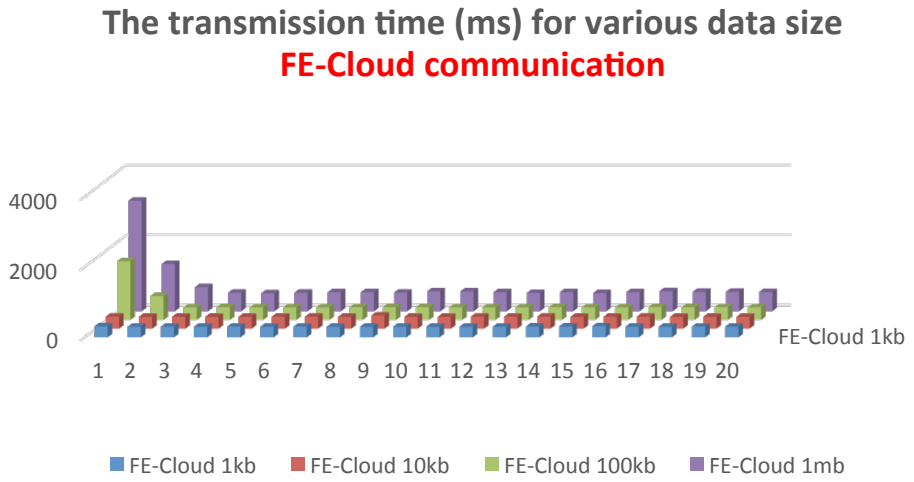
Multiple receivers, multiple analysers, and single transmitter scenario:

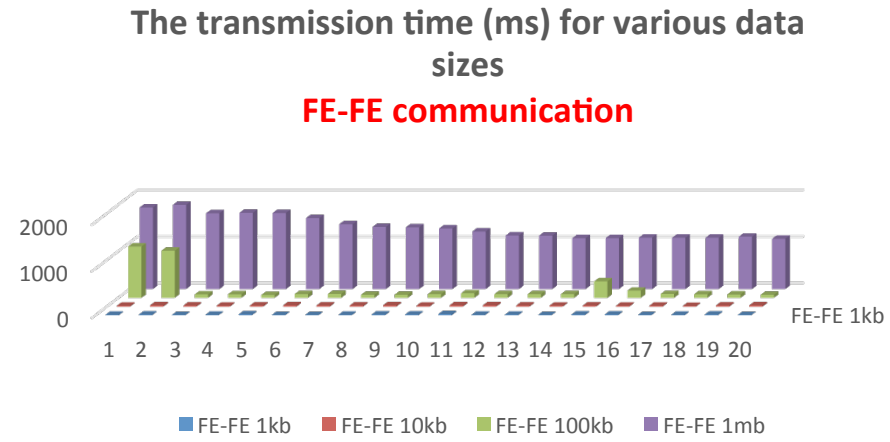
- Multiple FOG-engines receive and analyse data individually,
- FEs data is transmitted to the cloud via one of them which acts as a cluster head

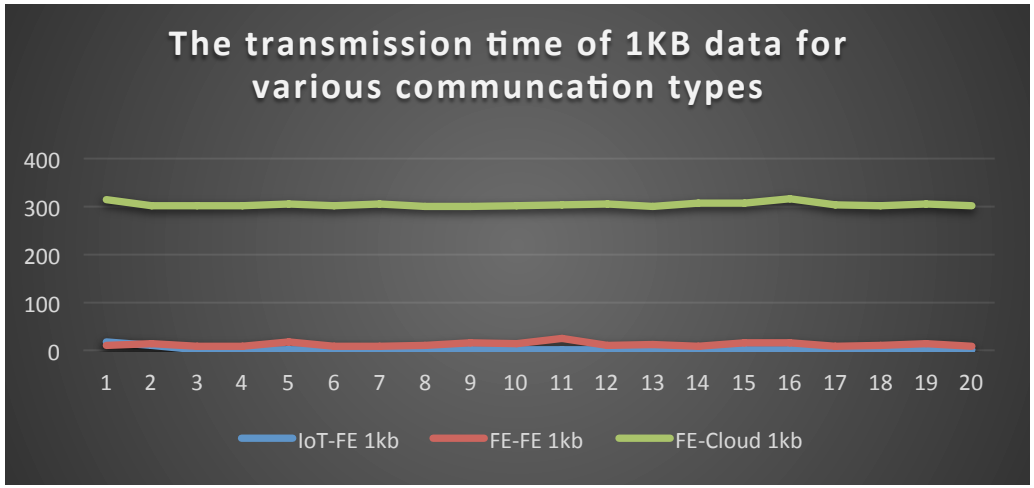


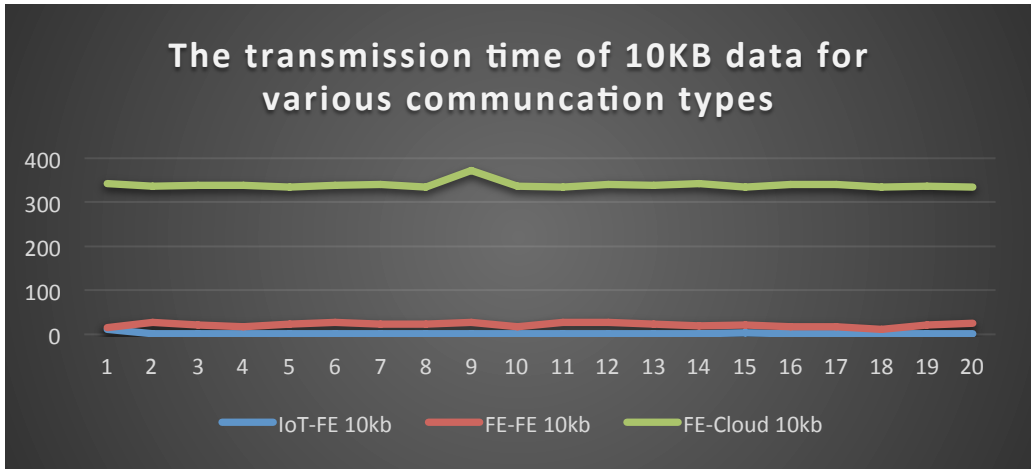


FE to Cloud communication time is significant compared to IoT-FE and FE-FE communication times (as expected).

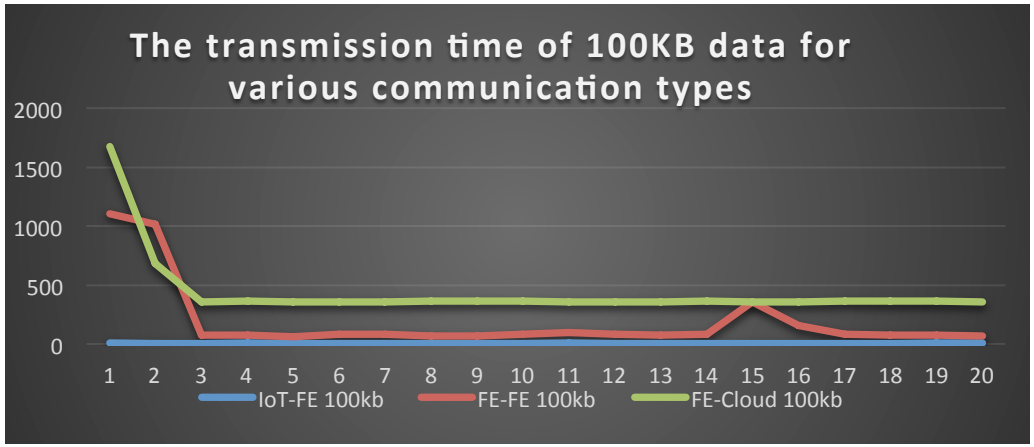








Data transmission time increases by increasing the size of data for all communication types



Thank you

Any Questions/Comments?