Big Data Analytics in Fog Computing

Dr Bahman Javadi
School of Computing, Engineering and Mathematics, Western Sydney University, Australia
Outline

• Introduction
• Integration of IoTs in Cloud Computing
• Big Data Analytics in Fog Computing
• FOG-Engine: Fog Computing Realization
• Case Study: Smart Nutrition Monitoring System using Fog Computing
• Open Issues and Challenges
Computing Paradigm Shift
More Connected Devices on the Planet Today Than People

GROWTH OF THE IoT
THE NUMBER OF CONNECTED DEVICES WILL EXCEED 50 BILLION BY 2020

BILLIONS OF DEVICES

2009
2012
2014
2015
2016
2017
2018
2019
2020

2009 IoT INCEPTION

2009
2012
2014
2015
2016
2017
2018
2019
2020

SOURCE: CISCO

4
Internet of Things Applications
IoTs in Sydney Harbour Bridge

- 1,149m length, 134m height
- 160,000 cars per day
- Moving joints <18cm>
  - 2400 vibration sensors
  - Machine learning for structural monitoring
Global IoT Market

![Graph showing the growth of the global IoT market from 2000 to 2019. The market value increases significantly over the years, reaching 1710 billion in 2019.](image)
Architecture of the Internet of Things

Application Layer
- Merchandise Tracking
- Environment Protection
- Intelligent Search
- Telemedicine
- Intelligent Traffic
- Smart Home

Network Layer
- Mobile Telecom Network
- The Internet
- Information Network

Sensing Layer
- RFID
- Sensor Network
- GPS
- RFID Label
- Sensor Nodes
- Road Mapper

Cloud Computing Platform
## IoT in Public Clouds

<table>
<thead>
<tr>
<th>Service</th>
<th>AWS</th>
<th>Microsoft</th>
<th>IBM</th>
<th>Google</th>
<th>Alibaba</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Collection</strong></td>
<td>HTTP, WebSockets, MQTT</td>
<td>HTTP, AMQP, MQTT and custom protocols (using protocol gateway project)</td>
<td>MQTT, HTTP</td>
<td>HTTP</td>
<td>HTTP</td>
</tr>
<tr>
<td><strong>Security</strong></td>
<td>Link Encryption (TLS), Authentication (SigV4, X.509)</td>
<td>Link Encryption (TLS), Authentication (Per-device with SAS token)</td>
<td>Link Encryption (TLS), Authentication (IBM Cloud SSO), Identity management (LDAP)</td>
<td>Link Encryption (TLS)</td>
<td>Link Encryption (TLS)</td>
</tr>
<tr>
<td><strong>Integration</strong></td>
<td>REST APIs</td>
<td>REST APIs</td>
<td>REST and Real-time APIs</td>
<td>REST APIs, gRPC</td>
<td>REST APIs</td>
</tr>
<tr>
<td><strong>Data Analytics</strong></td>
<td>Amazon Machine Learning model (Amazon QuickSight)</td>
<td>Stream Analytics, Machine Learning</td>
<td>IBM Bluemix Data Analytics</td>
<td>Cloud Dataflow, BigQuery, Datalab, Dataproc</td>
<td>MaxCompute</td>
</tr>
</tbody>
</table>
Current platforms issues: Not Fully Integrated, No low-latency, and might be Expensive

Service Ecosystem: Fragmented
IoT infrastructure
→ designers need to interact with
many services from sensors/actuators
to data analytic

Cloud: latency
The cloud uses virtual machines
→ unnecessary data movement

Network: latency, bandwidth and cost
• Large geographical distance → Higher Latency
• The aggregated b/w of sensors >> network b/w
• Big Data is heavy to move → Higher Cost and Latency

Raw data: the size can be huge (e.g. camera)

Network & Internet Infrastructure

IoTs

Big Data Flow

Feedback

Network of Sensors and Actuators (Physical World)
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
FOG-Engine: Decentralized Hierarchical Big Data Processing on the Edge

CLOUD

Data Processing, Mining, Storage, and Visualization

FOG

Data Preprocessing & Analytic

WAN: low b/w (x10MB/s)

LAN: very high b/w (x1GB/s)
MAN: high b/w (x100MB/s)

Network of Sensors and Actuators (Physical World)

IoTs

Network Access

Raw data

micro-Engine

FOG-Engine: Data Analytic

micro-Engine
# FOG-Engine vs. Cloud

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>FOG-Engine</th>
<th>Cloud platform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing hierarchy</td>
<td>Local data analytics</td>
<td>Global data analytics</td>
</tr>
<tr>
<td>Processing fashion</td>
<td>In-stream processing</td>
<td>Batch processing</td>
</tr>
<tr>
<td>Computing power</td>
<td>GFLOPS</td>
<td>TFLOPS</td>
</tr>
<tr>
<td>Network Latency</td>
<td>Miliseconds</td>
<td>Seconds</td>
</tr>
<tr>
<td>Data storage</td>
<td>Gigabytes</td>
<td>Infinite</td>
</tr>
<tr>
<td>Data lifetime</td>
<td>Hours/Days</td>
<td>Infinite</td>
</tr>
<tr>
<td>Fault-tolerance</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Processing resources</td>
<td>Heterogeneous (e.g. CPU, FPGA)</td>
<td>Homogeneous (Data center)</td>
</tr>
<tr>
<td>Versatility</td>
<td>Only exists on demand</td>
<td>Intangible servers</td>
</tr>
<tr>
<td>Provisioning</td>
<td>Limited by the number of FOG-engines in the vicinity</td>
<td>Infinite, with latency</td>
</tr>
<tr>
<td>Mobility of nodes</td>
<td>May be mobile (e.g. in the car)</td>
<td>None</td>
</tr>
</tbody>
</table>
How to Realize Big Data’s Vs: Velocity, Volume, ...

Proposed Solution

Current Cloud

Low b/w
Low Speed

High b/w
High Speed

Low b/w
Low Speed

High b/w
High Speed

Volume

Velocity
A Typical Data Analytics Flow

Raw data → Data collection/integration → Data cleaning → Feature extraction & transformation → DB storage

Data Analysts → Interpretation and presentation → Data analytics

- Data cleaning
- Feature extraction & transformation
- DB storage
- Data analysts
- Interpretation and presentation
- Data Analysts
- Raw data
- Data collection/integration
A Modified Data Analytics Flow

Raw Data from IoTs → Data collection/integration → Data cleaning → Feature extraction & transformation → DB Storage → Interpretaion & presentation → Data analytics → Cloud → Data from other FOG-engines/IoT nodes.

FOG-engines/Users → Interpretation & presentation → Data Analytics → Data integration → DB Storage.

Users

Fog-Engine
General Architecture of FOG-Engine

Data Analytics & Storage Unit
- Data Analytic Engine
- Data Cleaning, Aggregation & Visualization
- Data Collection and Import
- Data Storage System

Orchestration Unit

Communication Unit
- Network Interface to Physical World
- Peer-to-Peer Networking API
- Network Interface to Cloud (Gateway)

IoTs

Collaborating peers in the FOG

Cloud
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
FOG-Engine Prototyping

• Raspberry Pi 3.0
• Python Library
• Results
  – 40% reduced in data size
  – 60% increase in data efficiency
Smart Nutrition Monitoring System using Fog Computing

Overweight and obesity in Adults
- US: 70%
  - $200B
- Australia: 63%
  - $14B
- India: ~20%
  - 3rd most obese country in the world

Obesity and Overweight Increasing Worldwide

- 37% of the world’s adult population that is overweight or obese
- 0 countries succeeding in decreasing obesity in last 33 years
- 3.4 million deaths caused by overweight and obesity
- 14% of overweight or obese children and adolescents worldwide
- 62% of the world’s obese living in developing countries
- Obesity and overweight increased 27.5% in adults and 47.1% in children since 1980

Middle Eastern countries experiencing some of the largest increases in obesity globally:
- Saudi Arabia, Bahrain, Egypt, Kuwait, and Palestine

Obesity and overweight contribute to:
- Cardiovascular Disease
- Diabetes
- Cancer
- Joint Pain
Smart Nutrition Monitoring System

• Project Scope
  – Take away food (50M meals out each week in Australia)

• Non-invasive
  – Minimizing the amount of direct input and actions from users

• High data accuracy and reliability
  – Heterogeneous IoT sensors

• Scalability
  – Cloud and Fog Computing
Architecture of Smart Nutrition Monitoring System

Cloud Servers

Food Nutrient Database

Smart Nutrition Monitoring Engine

Collections Management

Data Analytics

Visualization

User

Dietitian

IoT sensors

Data collection points (kiosks)
System Prototype

AgiSoft PhotoScan Pro (3D Image)

Cloud servers (private cloud)

Intranet

Internet

FOG-Engine

Mobile App

bluetooth

Kiosk with embedded sensors

Weight

Volume

Structure

Smart Nutrition Monitoring Engine

public cloud
Open Issues and Challenges

- Innovative Machine Learning
- Decentralized Resource Scheduling
- Reliability and Power Efficiency
- Information and Security Management
References


Thank You

Home Page: http://staff.scem.uws.edu.au/~bjavadi/

Email: b.javadi@westernsydney.edu.au