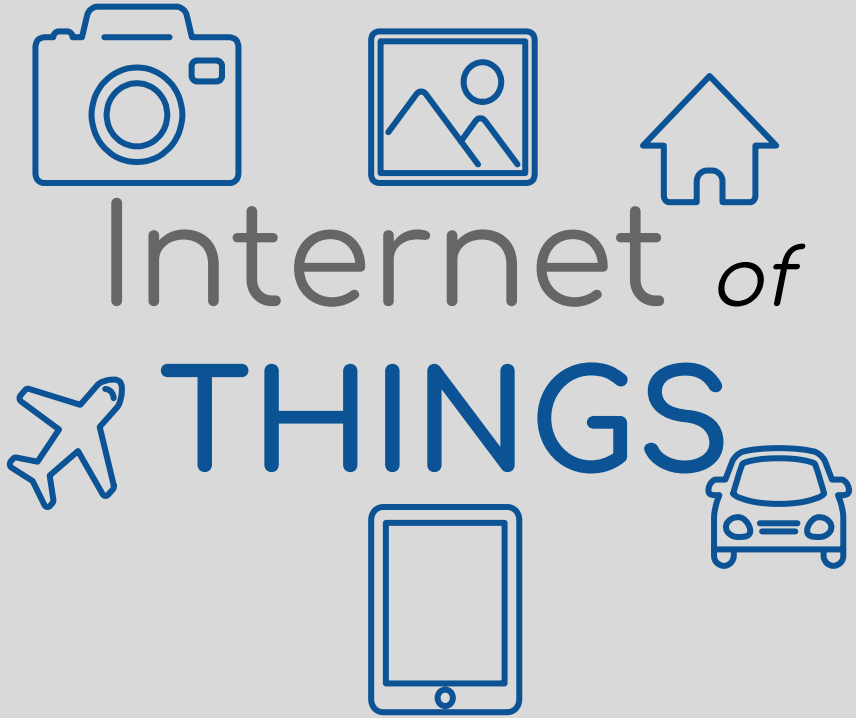


"Breaking down the infrastructure behind IoT ecosystems"

Quang-Duy Tran
Marc Garnica Caparrós

17th of May, 2018
Cloud Computing & Big Data Analytics

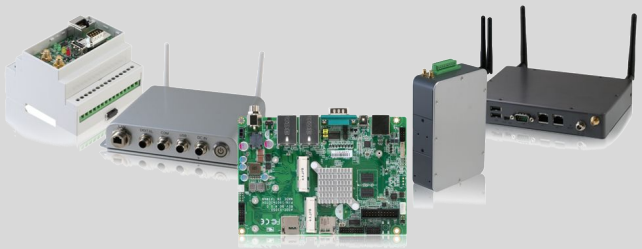
Behind the curtain?

Architecture

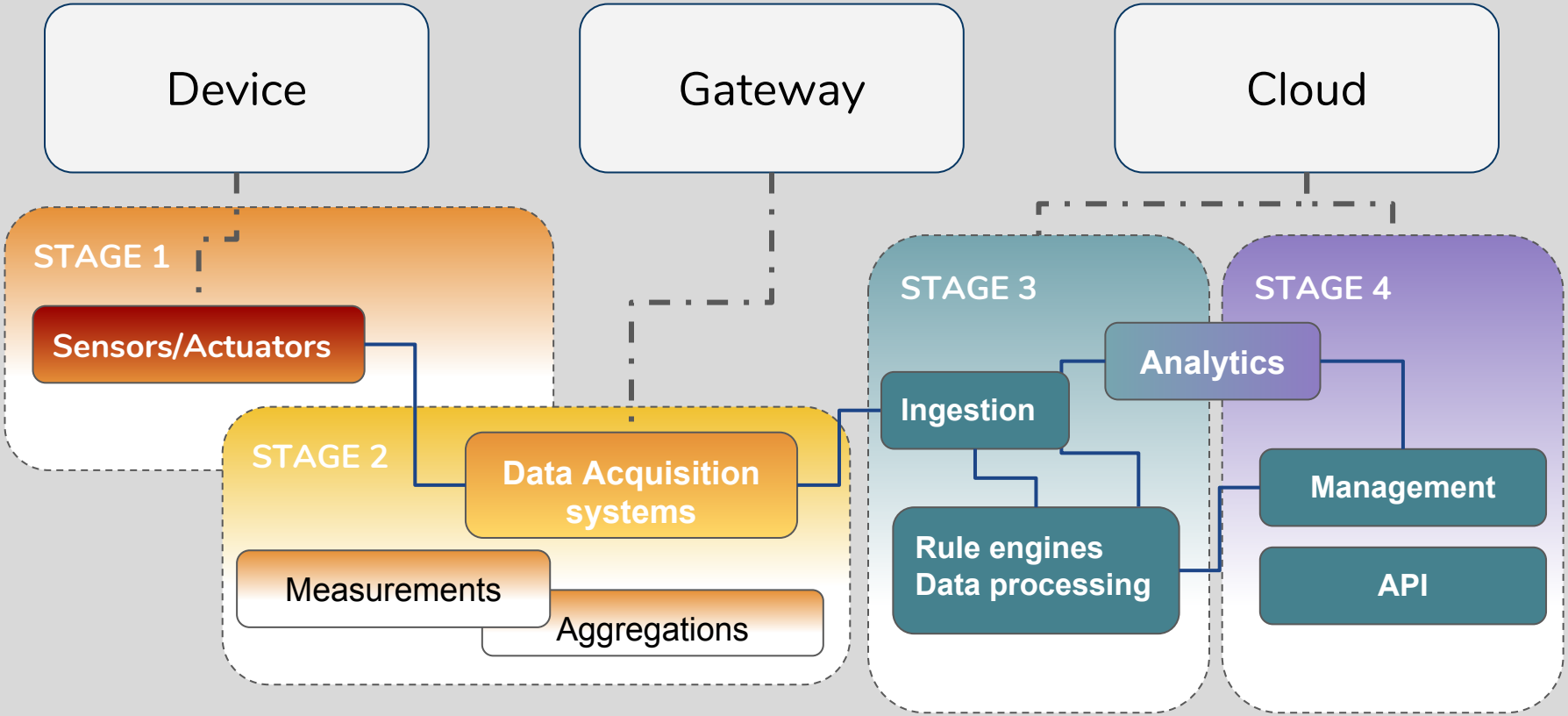
Device

Gateway

Cloud



Architecture - Dataflow



Architecture - Dataflow

Device

Gateway

Cloud

STAGE 1

- First level of interaction
- Type of information
- Metadata and commands
- Device management

Architecture - Dataflow

Device

Gateway

Cloud

STAGE 1

- First level of interaction
- Type of information
- Metadata and commands
- Device management


```
{
  "DeviceProperties": {
    "DeviceID": "deviceid1",
    "HubEnabledState": null,
    "CreatedTime": "2016-04-25T23:54:01.313802Z",
    "DeviceState": "normal",
    "UpdatedTime": null
    },
  "SystemProperties": {
    "ICCID": null
  },
  "Commands": [],
  "CommandHistory": [],
  "IsSimulatedDevice": false,
  "id": "fe81a81c-bcbc-4970-81f4-7f12f2d8bda8"
}
```

Architecture - Dataflow

Device

Gateway

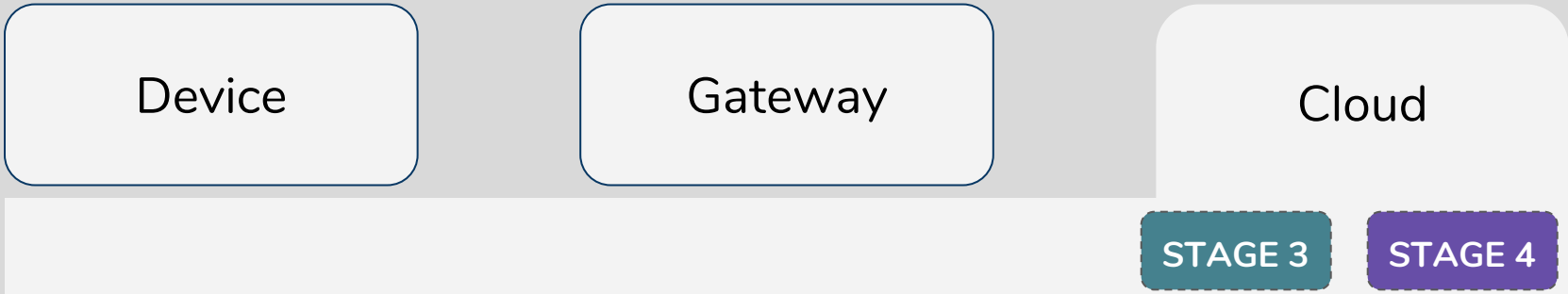
Cloud

STAGE 1

STAGE 2

- Usually underrated
- Simple operations and aggregations. ⚠
- Protocols conversion

Architecture - Dataflow



- Ingestion
- Storage
- Rule processing and stream analytics
- Analytics
- Archival storage

Architecture - Dataflow

Device

Gateway

Cloud

Ingestion:

- Process of importing, transferring, loading and processing the data
- Data consumed in high volume and velocity, from different sources
- Ingest data in real-time or by batches

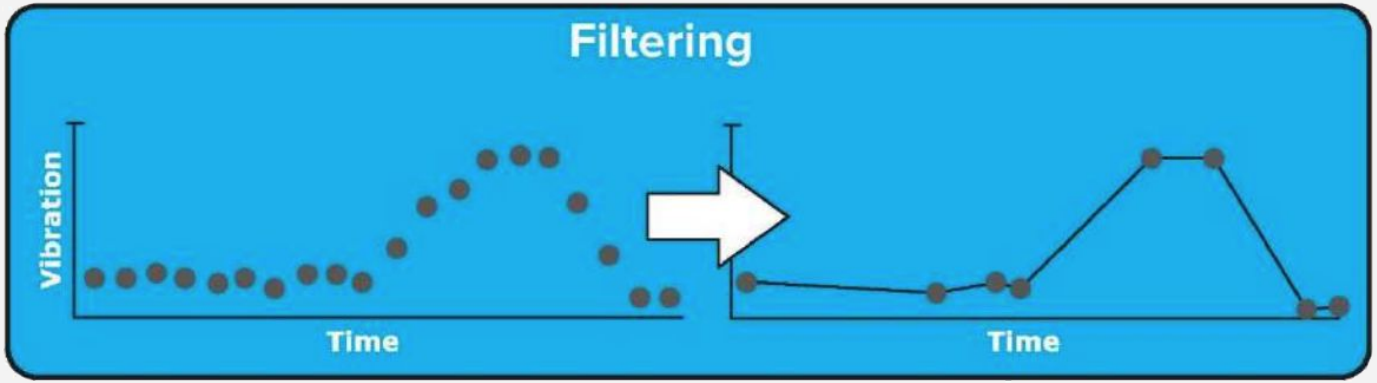
Architecture - Dataflow

Device

Gateway

Cloud

Ingestion: filtering data



Architecture - Dataflow

Device

Gateway

Cloud

Storage:

- Sensor data: small in size, generated frequently, could be millions per second.
  Use NoSQL databases: HBase, Cassandra
- Multimedia data: text, sound, video. Large in size
  Store using BLOB (Binary Large OBject)

Architecture - Dataflow

Device

Gateway

Cloud

Stream analytics:

- (Near) real-time analytics, analyze and make decisions quickly
- Used in cyber security, energy, supply chain, marketing, etc
- Stream operation: aggregation, pattern detection, correlations
- Apache Storm: distributed real-time computing system, can process millions of tuples per second

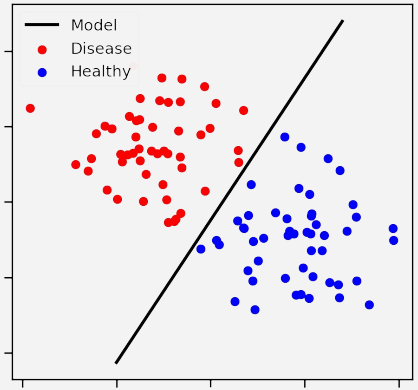
Architecture - Dataflow

Device

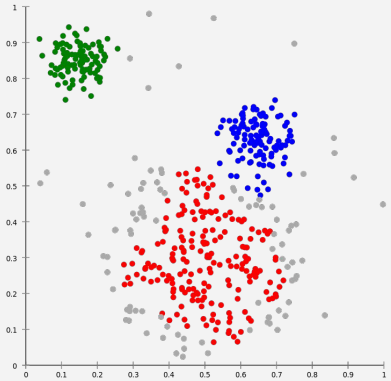
Gateway

Cloud

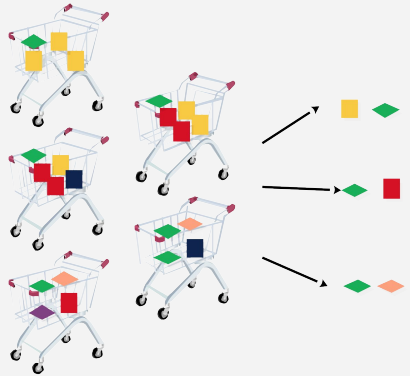
Analytics: data mining techniques



Classification



Clustering



Association rules

Architecture - Dataflow

Device

Gateway

Cloud

Archival storage

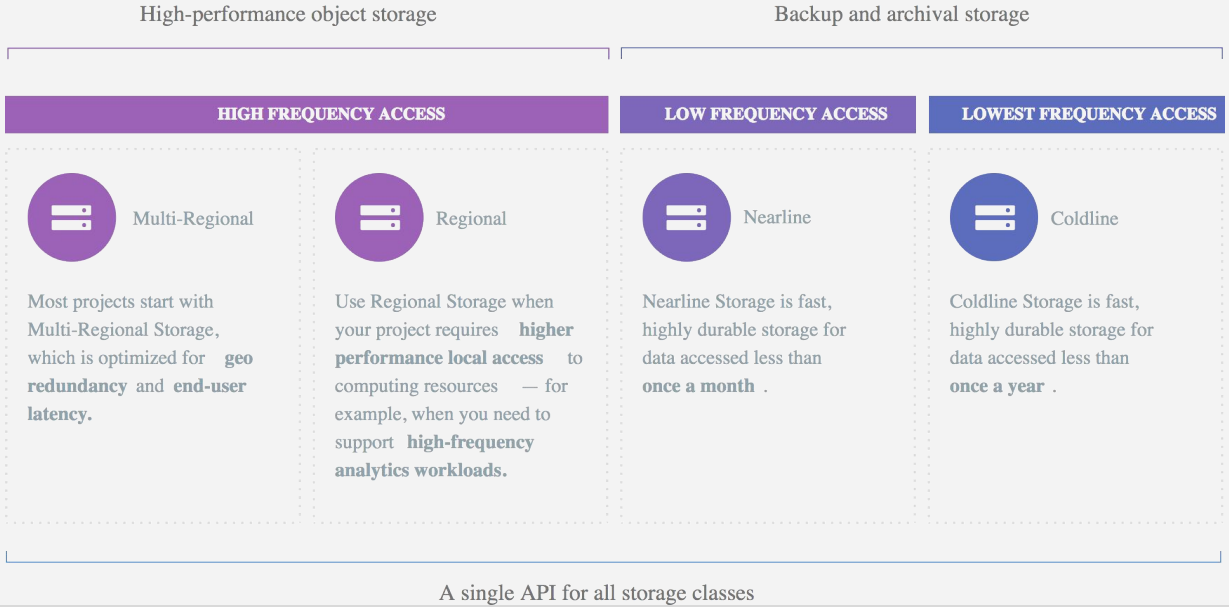
- Infrequently accessed data are stored in another region
- Used for references, keeping older version of changed file

Architecture - Dataflow

Device

Gateway

Cloud



Cloud providers

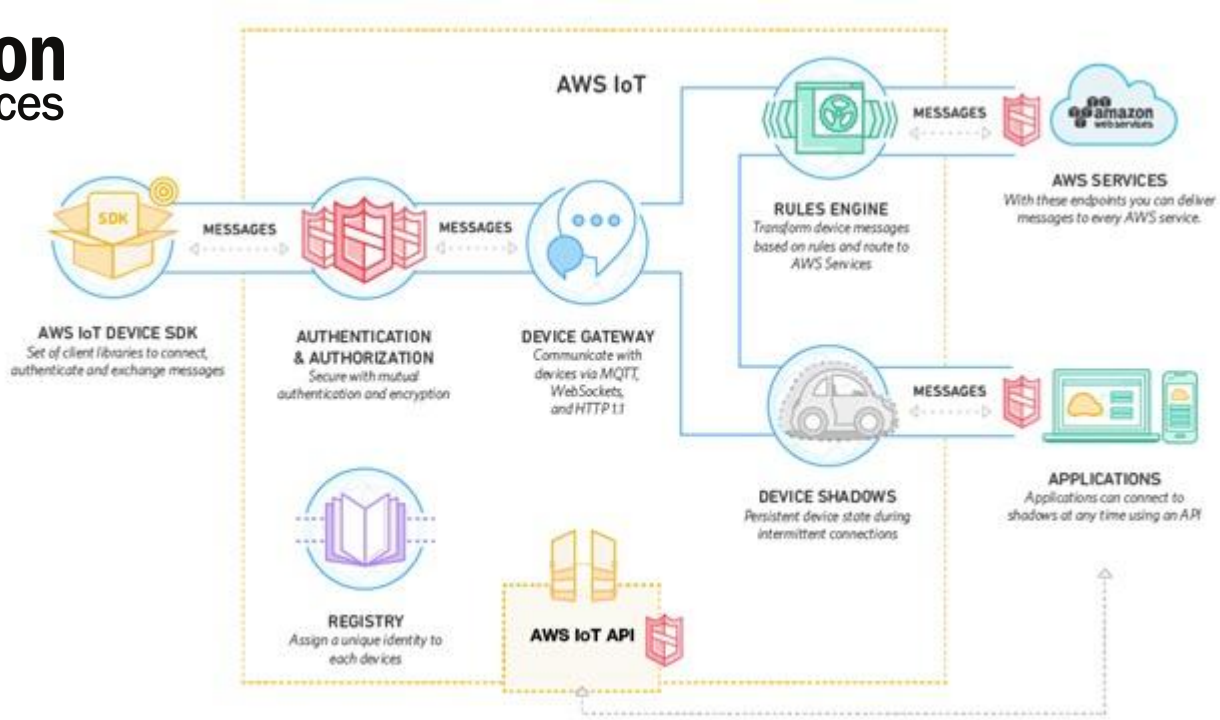
Features

Experience

Integration

Pricing

amazon web services



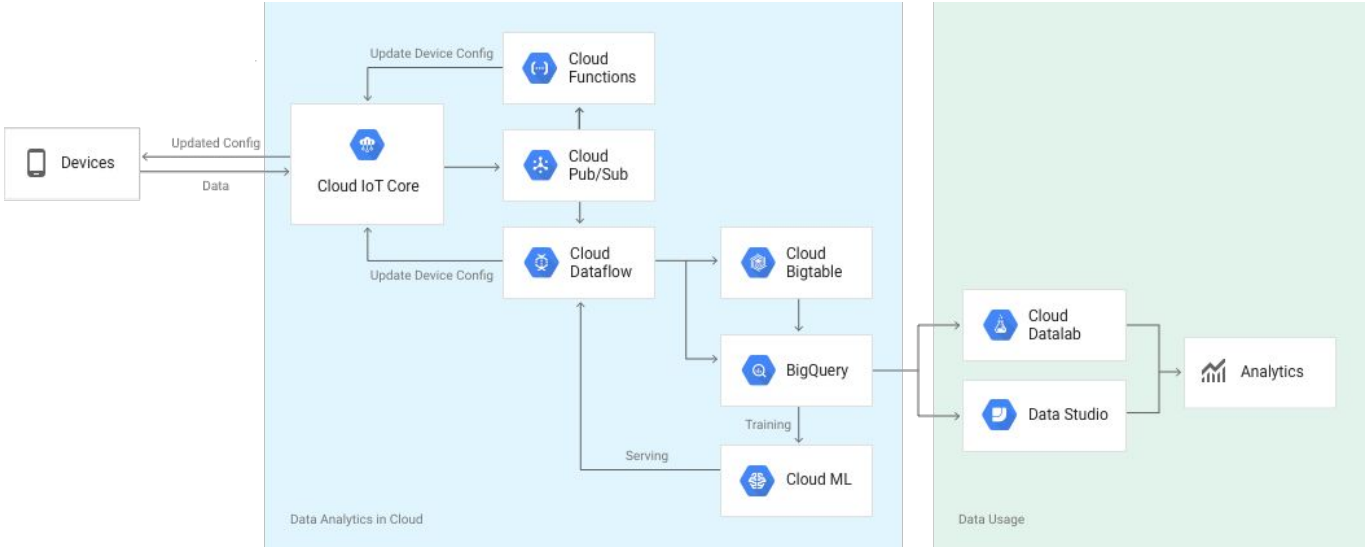
Features

Experience

Integration

Pricing

Google Cloud



Features

Experience

Integration

Pricing

Google Cloud

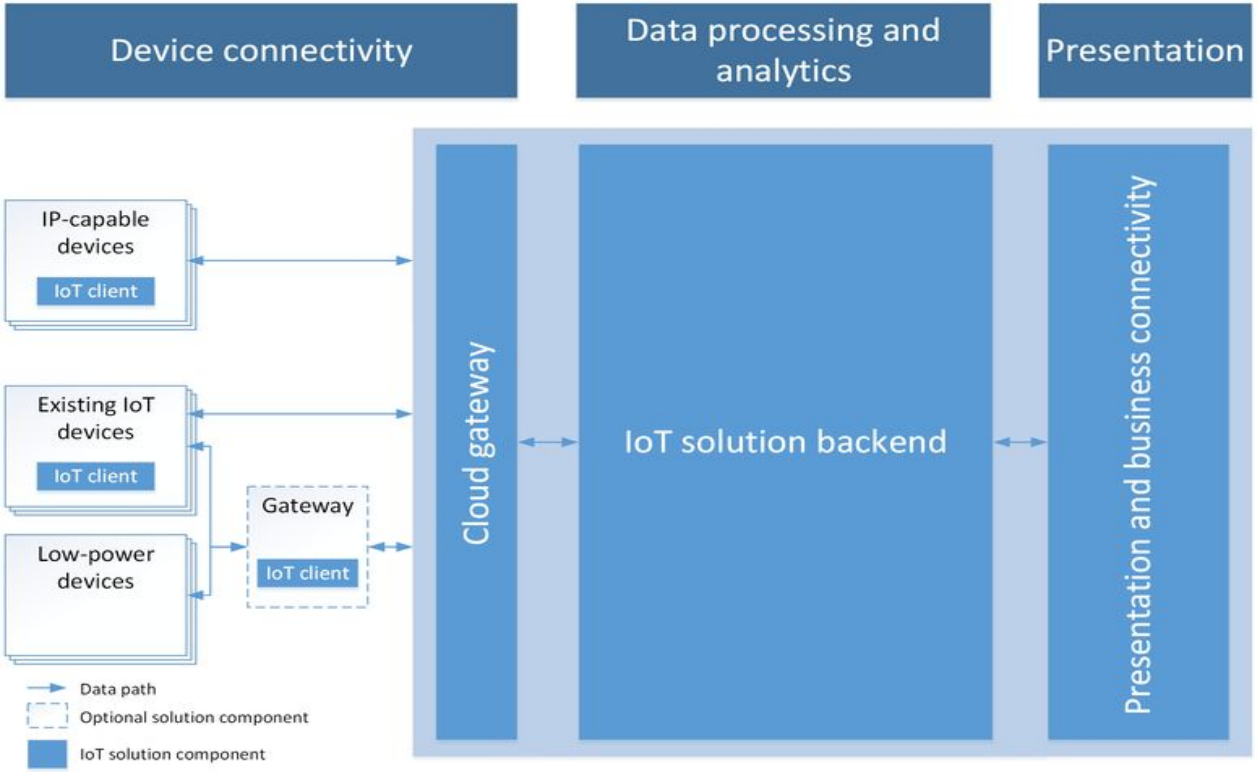
Features

Experience

Integration

Pricing

Microsoft Azure



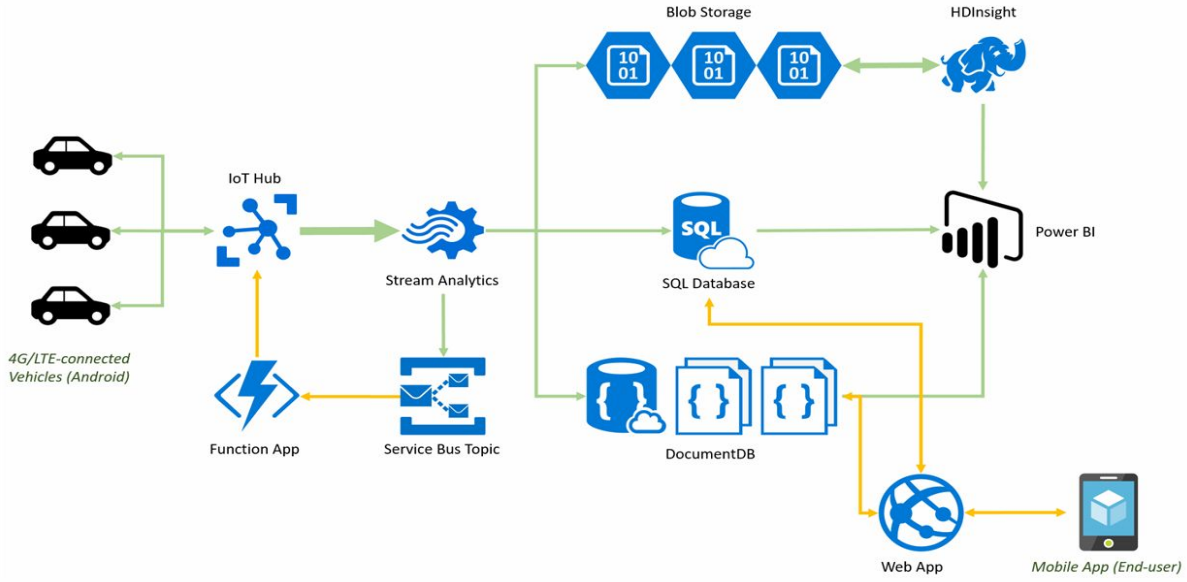
Features

Experience

Integration

Pricing

Microsoft Azure



Features

Experience

Integration

Pricing

Conclusions

Q&A