IoT – Impact On Storage Architecture

SDC India

Girish Kumar B K
NetApp
24th May 2018
IoT - Agenda

1) Introduction
2) Data growth and compute model
3) Industrial needs and IoT architecture
4) Data flow and functional architecture
5) Conclusions
Background: Digital transformation

Harness the power of data for competitive differentiation

Digital transformation is the integration of digital technology into all areas of a business, fundamentally changing how you operate and deliver value to customers. It’s also a cultural change that requires organizations to continually challenge the status quo, experiment, and get comfortable with failure.

- Manufacturing
- Agriculture
- Energy Management
- Environment Monitoring
- Building and Home Automation
- Smart Cities
- Transportation
- HealthCare
Internet of Things (IoT): Enabling communication between devices, people & processes to exchange useful information & knowledge that create value for humans. Term was first proposed by Kevin Ashton in 1999.
IoT – Connected Devices Growth

Grand total and estimated growth

~ 5.5 million new things connected every day in 2017
~ 8.6 million new things connected every day in 2017

Grand Total

Business use case

Source: Gartner 2017
IoT – Data growth samples

Enterprise use cases

- **Twin Engine Jet**
  - 5000 sensors/engine,
  - 844 TB/day (10 GB/second)

- **Wind Farm**
  - 20000 Wind turbines each with 150 sensors
  - 30 TB of data to analyze in real time

- **Autonomous Vehicle**
  - 4 TB/day (50 MB/s)

- **Oil and Gas exploration**
  - 1 TB/day (12.5 MB/s)

Sources: Cisco, Intel, AviationWeek, Big Data in the Oil & Gas Upstream Industry
IoT – Compute Model

IT landscape is evolving from a centralized to a distributed computing model.

Centralized compute model has traditionally supported the Mobile – Cloud architecture (data flows between device to cloud).

As the number of connected devices / sensors and data volume explodes, the landscape is shifting to a distributed compute model (data flows between device to edge gateway to cloud).

IoT – Heat map of industrial sector wise needs
Manufacturing, Smart building, Smart Grid and Consumer

<table>
<thead>
<tr>
<th>Need</th>
<th>Industry sector</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Manufacturing</td>
</tr>
<tr>
<td>Mobility</td>
<td>55</td>
</tr>
<tr>
<td>Ultra Low latency (&lt;10ms)</td>
<td>95</td>
</tr>
<tr>
<td>Autonomy</td>
<td>95</td>
</tr>
<tr>
<td>Security</td>
<td>100</td>
</tr>
<tr>
<td>Local Network Bandwith</td>
<td>100</td>
</tr>
<tr>
<td>WAN Network Bandwidth</td>
<td>35</td>
</tr>
<tr>
<td>Peer-to-Peer Communication</td>
<td>80</td>
</tr>
<tr>
<td>Prioritization</td>
<td>100</td>
</tr>
<tr>
<td>Self-organization, discovery</td>
<td>60</td>
</tr>
<tr>
<td>Artificial Intelligence/</td>
<td>100</td>
</tr>
<tr>
<td>Machine Learning</td>
<td></td>
</tr>
</tbody>
</table>

Source: IIC Journal of Innovation 5th Edition
IoT - Need evolution

Commercial IoT device hierarchy of needs

- Edge intelligence
- Cloud offload
- Rich device functionality
- Security update control, lifecycle management, device management
- Long-life hardware, long-term OS support, security updates
IoT – Why edge?

Factor affecting the IoT architecture

Sources: IIC Journal of Innovation 5th Edition
IoT – Why edge?

Factor affecting the IoT architecture

95% data reduction at edge
IoT – Architecture

- Low latency decision making (<10ms)
- Autonomy (disconnected operations)
- Peer to Peer (edge to edge data sharing)
- Prioritization (Importance of the data)
- AI/ML near real time inferences

- Long term data retention
- Batch processing
- AI/ML model building, refining
- Data governance
- Policy based data lifecycle
IoT – Architecture

Storage requirements
- Incremental capacity scale (1 to 10s of TB)
- High bandwidth with low latency
- Optimized for time-series large blobs
- Storage optimized for containers (CI/CD)

<table>
<thead>
<tr>
<th>Network Type</th>
<th>4G</th>
<th>5G</th>
<th>Fiber N/w</th>
<th>Edge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Available Bandwidth</td>
<td>200 Mbps</td>
<td>1 Gbps</td>
<td>10 Gbps</td>
<td>Infinite</td>
</tr>
<tr>
<td>Time required to transmit 1 TB</td>
<td>~ 15 hrs</td>
<td>~ 3 hrs</td>
<td>~ 20 minutes</td>
<td>Real time</td>
</tr>
</tbody>
</table>
Case study: Active IQ (telemetry data from deployed systems)

The new Active IQ IoT architecture leverages AFF and the Data Fabric.
## Value Proposition

Agility, flexibility, efficiency and lower cost

### Reduced our own cost to run Active IQ environment by $1.25M per year

<table>
<thead>
<tr>
<th></th>
<th>Traditional Hadoop on-premises with DAS</th>
<th>Active IQ platform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to deploy new data pipeline</td>
<td>6 months</td>
<td>1 month</td>
</tr>
<tr>
<td>Time to deploy new Hadoop cluster</td>
<td>6 months</td>
<td>5 minutes</td>
</tr>
<tr>
<td>TCO</td>
<td>High</td>
<td>Moderate</td>
</tr>
<tr>
<td>Infrastructure for AI and machine learning</td>
<td>Expensive</td>
<td>Moderate</td>
</tr>
<tr>
<td>Cloudera Hadoop nodes needed</td>
<td>120</td>
<td>40</td>
</tr>
<tr>
<td>Storage required</td>
<td>12 PB</td>
<td>1.3 PB</td>
</tr>
</tbody>
</table>

Hybrid cloud analytics is up to 20% faster than a pure cloud based solution, while keeping the data secure and accessible from multiple clouds
Conclusions

1) Security, low latency, autonomy, prioritization and edge intelligence (AI/ML) continue to drive the enterprise IoT storage architecture

2) Hybrid cloud analytics is faster than pure cloud based solutions

3) Edges footprint offers opportunity to add generic or specific data management services for different industrial verticals
Thank You
Backup Slides
IoT – Architecture: Data Flow at the Edge
IoT – Architecture: Data flow at the lake

Real-time Data Analysis

Command and Control of IoT Env

Batch-oriented Data Analysis

Refinement of Analytical Models

Real-time Data Analysis

Command and Control of IoT Env

Batch-oriented Data Analysis

Refinement of Analytical Models

Other Data Sources

Application Space

Data Services

Persistence

Persistence

Edge Health Doctor

Policy-driven Caching/ Tiering

Data Anonymization

Data Sovereignty

Format-aware Storage Efficiency

© 2018 NetApp, Inc. All rights reserved.
## IoT – Architecture: Types of Data at the Edge

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Properties</th>
<th>Purpose</th>
<th>Format (typical)</th>
<th>Per unit size</th>
<th>Cardinality</th>
<th>Storage IO Patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device Registry</td>
<td>Devices at the edge that need enumeration</td>
<td></td>
<td>Structured (Relational)</td>
<td>1KB</td>
<td>nDevices</td>
<td>Read-only W - writes: re-provisioning</td>
</tr>
<tr>
<td>Device Configuration</td>
<td>Operational config of each device</td>
<td></td>
<td>Structured (JSON)</td>
<td>&lt;5KB</td>
<td>nDevices</td>
<td>Read-only W - writes: re-config</td>
</tr>
<tr>
<td>Device Identity</td>
<td>Cryptographic data for secure communication with devices</td>
<td></td>
<td>Structured (Key-Value)</td>
<td>1KB</td>
<td>nDevices</td>
<td>Read-only W - writes: permission revocation</td>
</tr>
<tr>
<td>Device State (Operational)</td>
<td>Device/ sensor data sent periodically</td>
<td></td>
<td>Record: {DeviceID,TimeStamp,Version,&lt;Blob&gt;}</td>
<td>&lt;1KB for industrial sensors &lt;3MBps for HD media streaming</td>
<td>nRecords (frequency varies on conditions)</td>
<td>Append-only R - read for Batch transfer</td>
</tr>
<tr>
<td>Edge Configuration</td>
<td>Operational config of IT resources at the edge</td>
<td></td>
<td>Structured (JSON)</td>
<td>&lt;1MB</td>
<td>nModules</td>
<td>Read-only W - writes: re-config</td>
</tr>
<tr>
<td>Digital Exhaust Data</td>
<td>Troubleshooting and Audit</td>
<td></td>
<td>Semi-structured (Logs, Counters)</td>
<td>10MB</td>
<td>nModules (and depends on nRecords)</td>
<td>Log structured R - read for Batch transfer</td>
</tr>
</tbody>
</table>
End of Slides