IoT – Impact On Storage Architecture

SDC India

Girish Kumar B K
NetApp
24th May 2018

NetApp
Data Driven

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

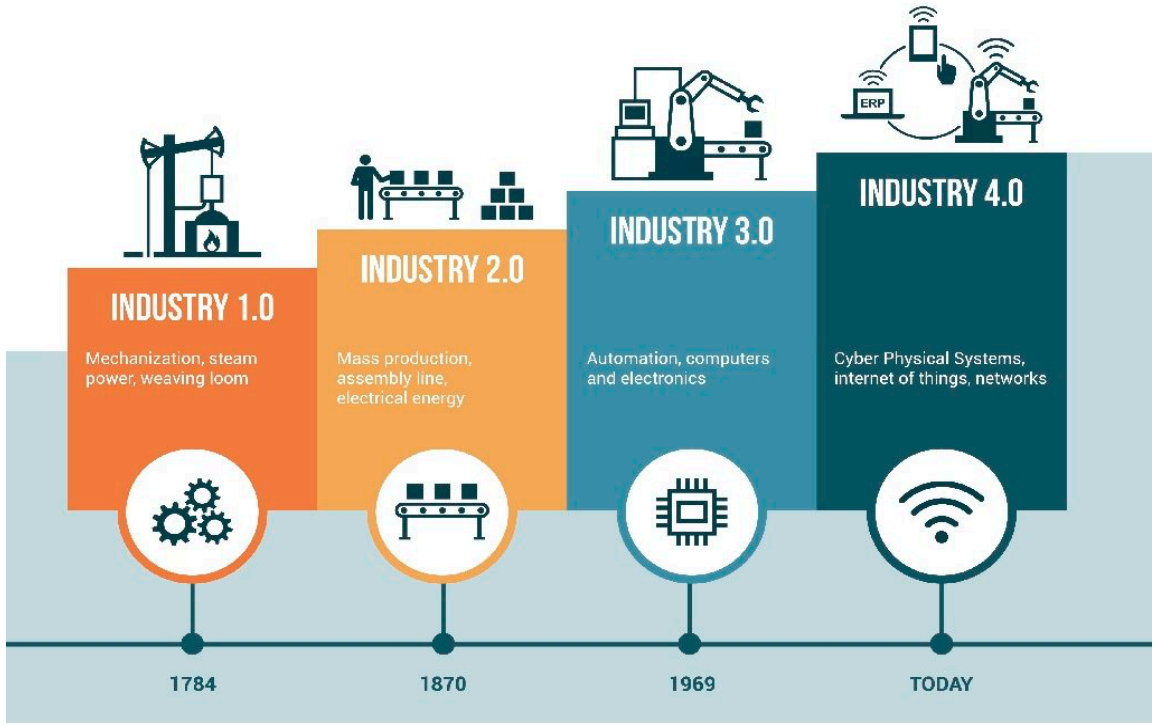
Harness the power of data for competitive differentiation

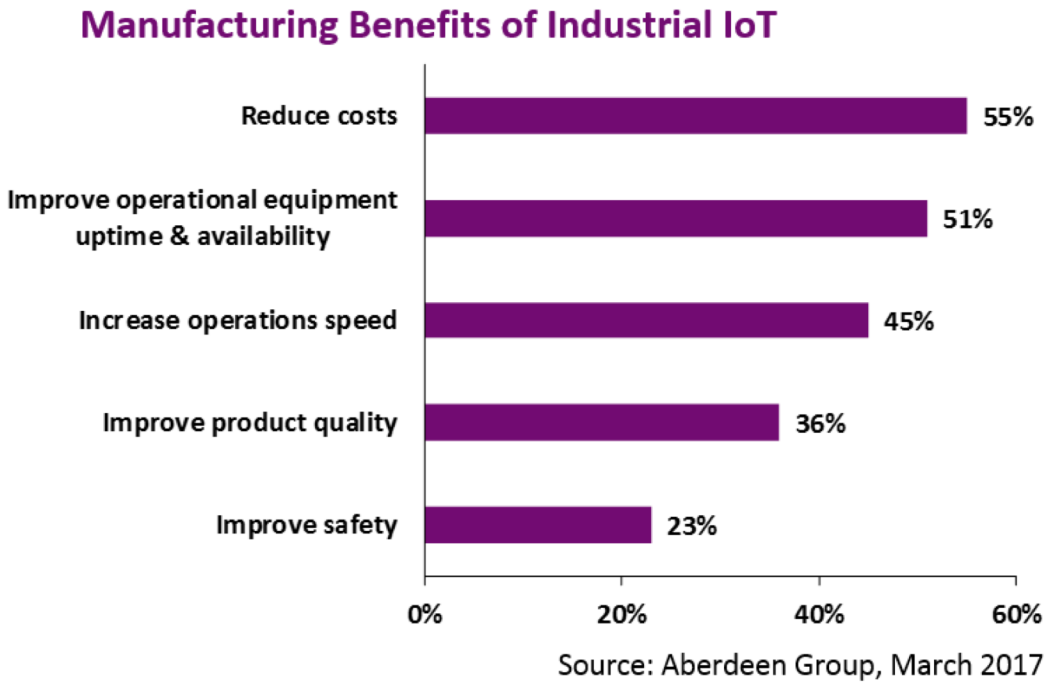
*Digital transformation is the integration of digital technology into all areas of a business, fundamentally **changing how you operate and deliver value to customers**. It's also a cultural change that requires organizations to continually challenge the status quo, experiment, and get comfortable with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

- Manufacturing
- Agriculture
- Energy Management
- Environment Monitoring
- Building and Home Automation
- Smart Cities
- Transportation
- HealthCare

NetApp

IoT : Industry evolution and benefits



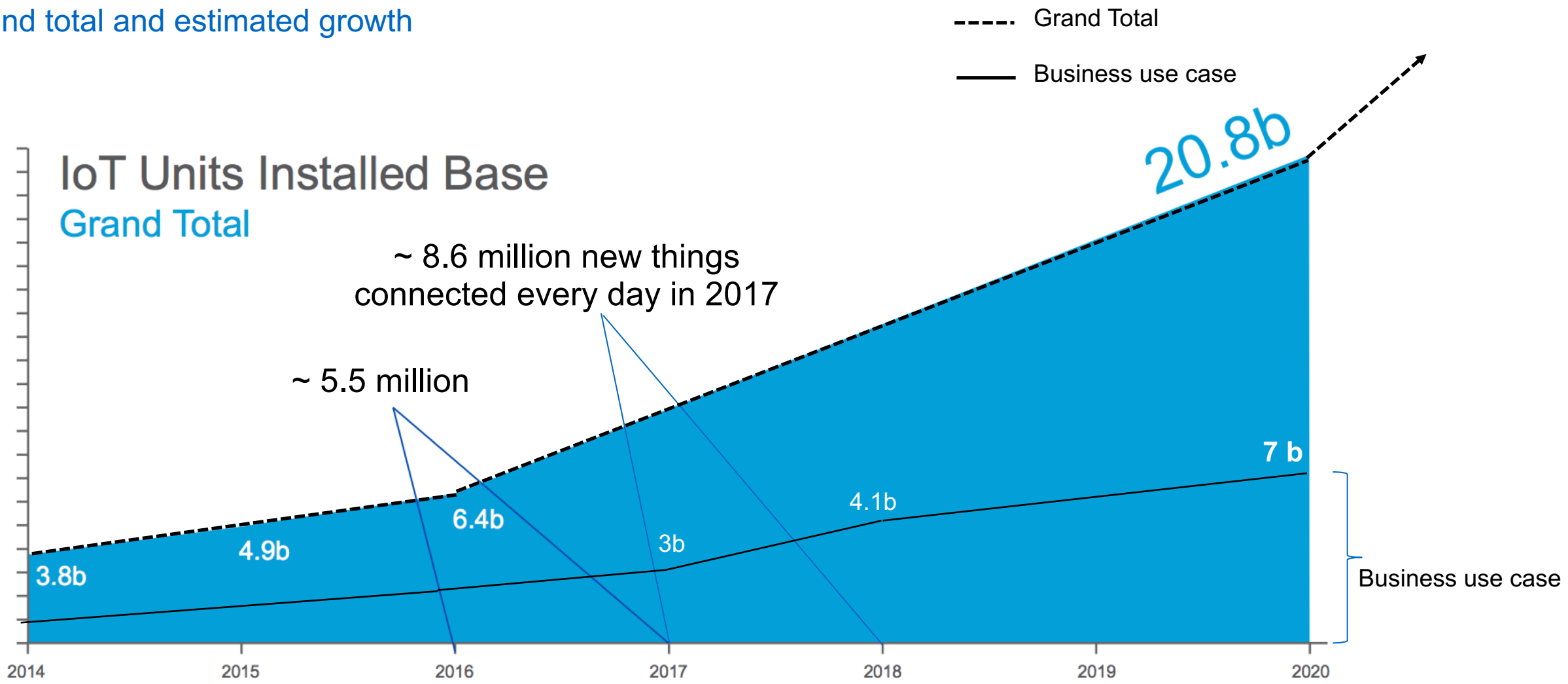


Internet of Things (IoT): Enabling communication between devices, people & processes to *exchange useful information & knowledge that create value for humans*
Term was first proposed by Kevin Ashton in 1999

NetApp

IoT – Connected Devices Growth

Grand total and estimated growth



Source: Gartner 2017

NetApp

IoT – Data growth samples

Enterprise use cases

Twin Engine Jet

5000 sensors/engine,
844 TB/day (10 GB/second)

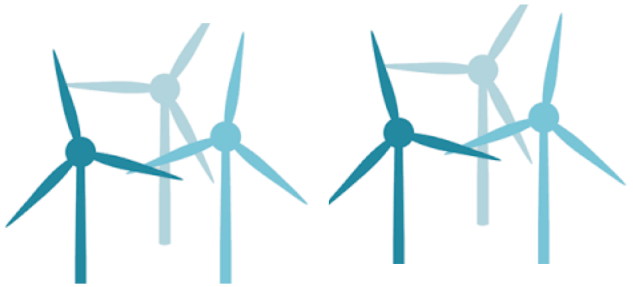
Autonomous Vehicle

4 TB/day (50 MB/s)

Oil and Gas exploration

1 TB/day (12.5 MB/s)

Wind Farm

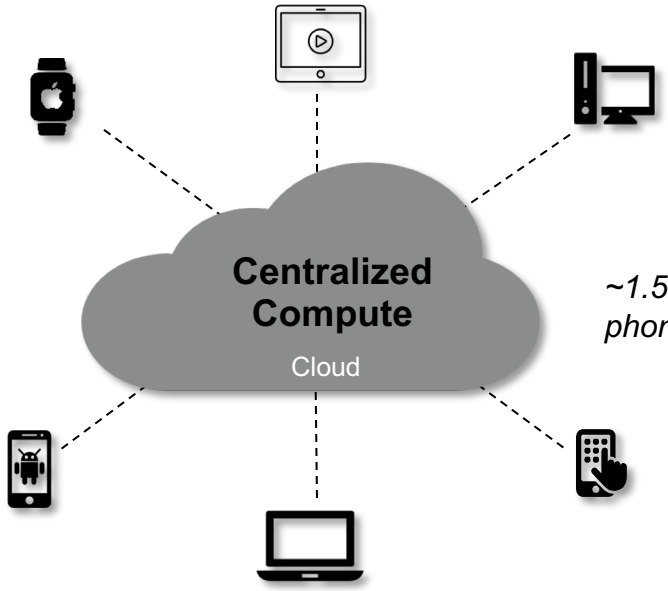


20000 Wind turbines each with 150 sensors
30 TB of data to analyze in real time

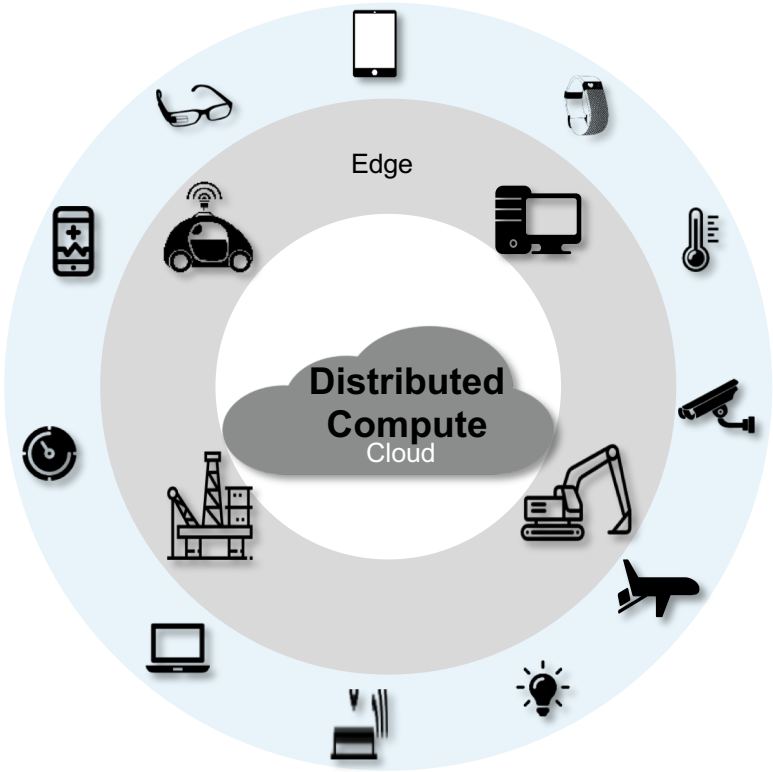
Sources: Cisco, Intel, AviationWeek, Big Data in the Oil & Gas Upstream Industry

IoT – Compute Model

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



~1.5B smart phones in 2017



20B+ connected devices with 1,000B+ sensors by 2020

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**)

Source: PwC research. TechCrunch, Andreseen Horowitz, Gartner, Mary Meeker Tech Trends

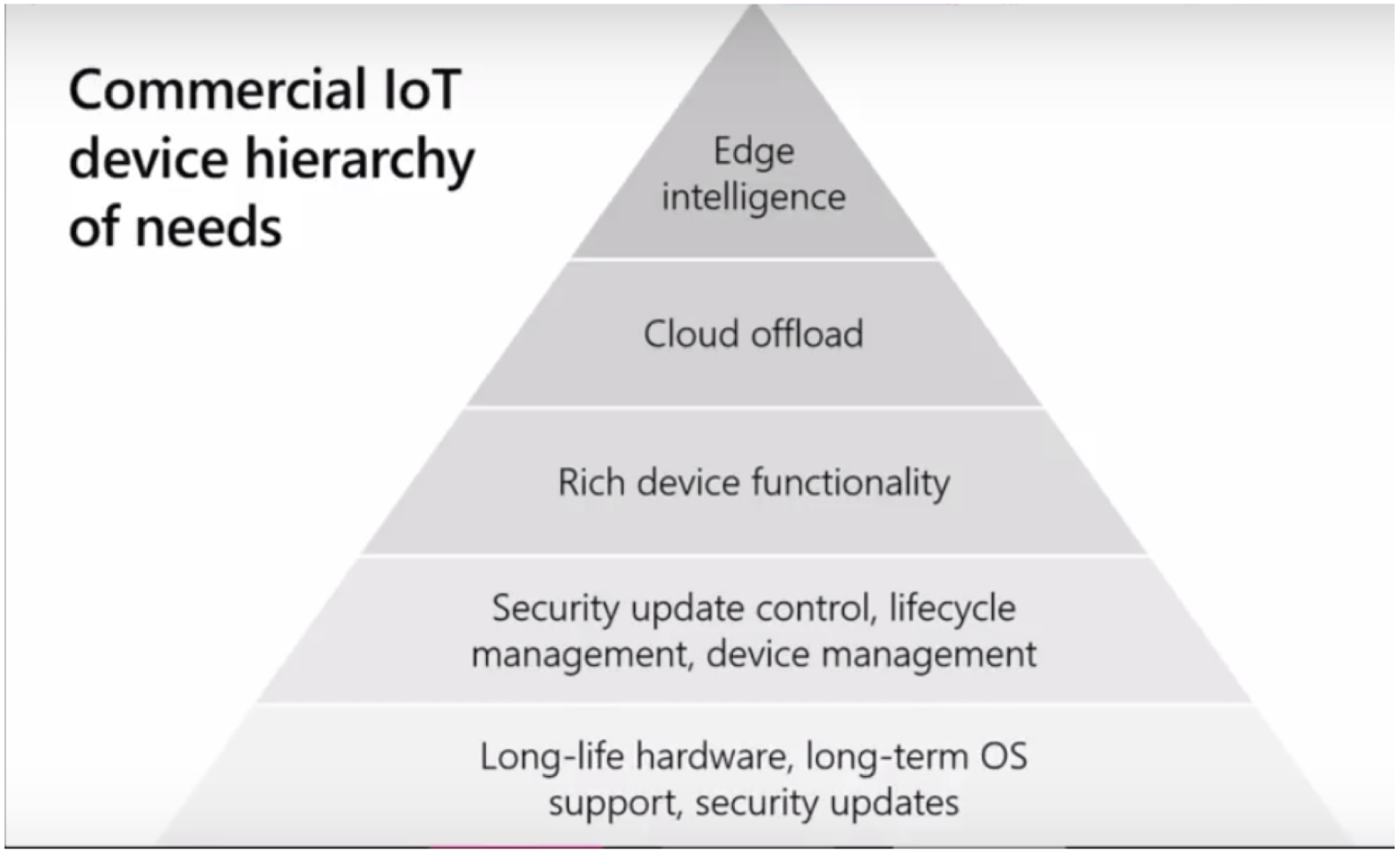
IoT – Heat map of industrial sector wise needs

Manufacturing, Smart building, Smart Grid and Consumer

Need	Industry sector			
	Manufacturing	Buildings & Facilities	Energy & Utilities	Consumer & Home
Mobility	55	10	10	55
Ultra Low latency (<10ms)	95	85	5	15
Autonomy	95	100	100	50
Security	100	100	100	25
Local Network Bandwith	100	90	10	35
WAN Network Bandwith	35	55	10	55
Peer-to-Peer Communication	80	85	50	90
Prioritization	100	15	90	10
Self-organization, discovery	60	20	40	65
Artificial Intelligence/ Machine Learning	100	100	85	45

Source: IIC Journal of Innovation 5th Edition

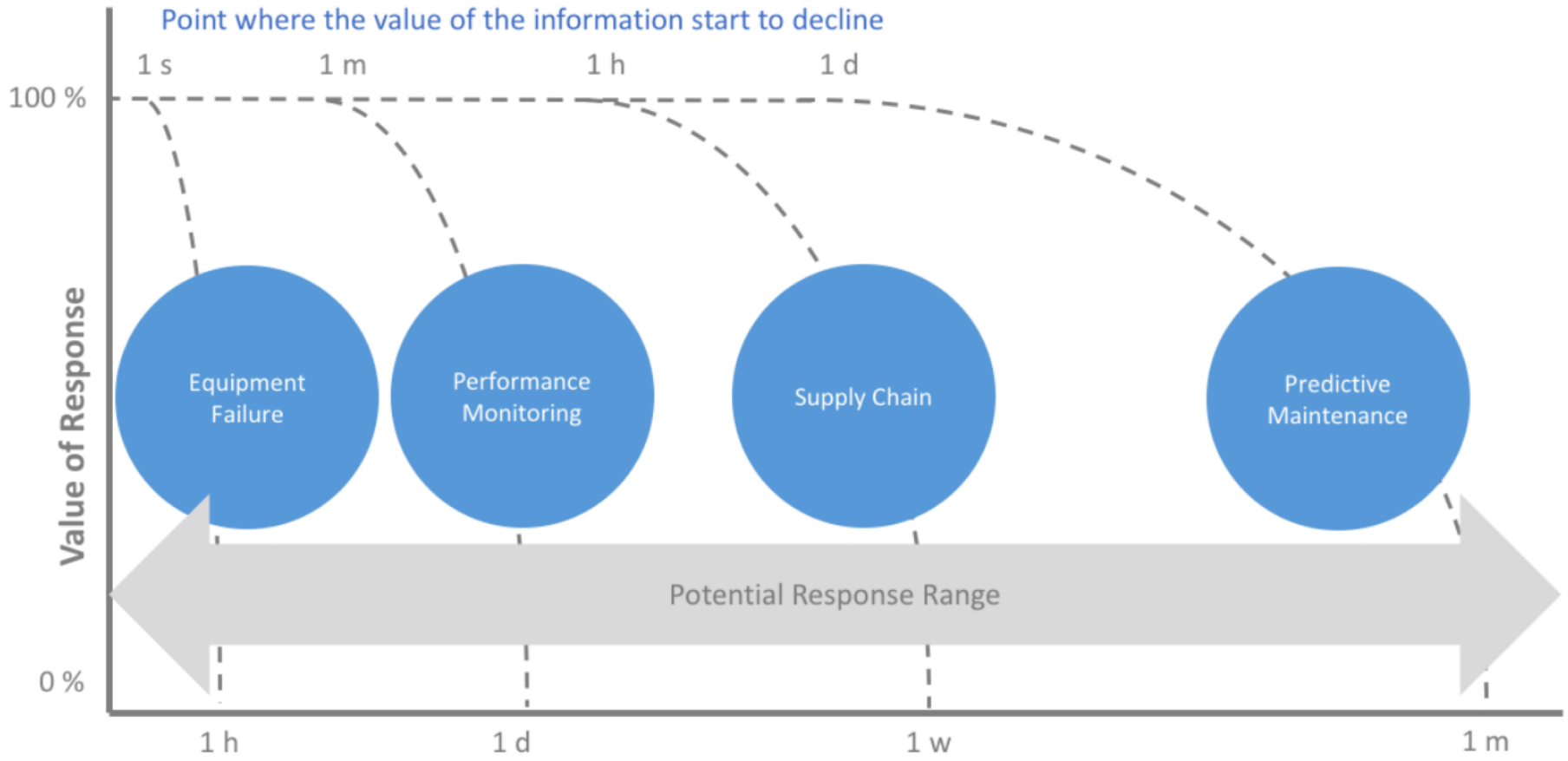
IoT - Need evolution



NetApp

IoT – Why edge ?

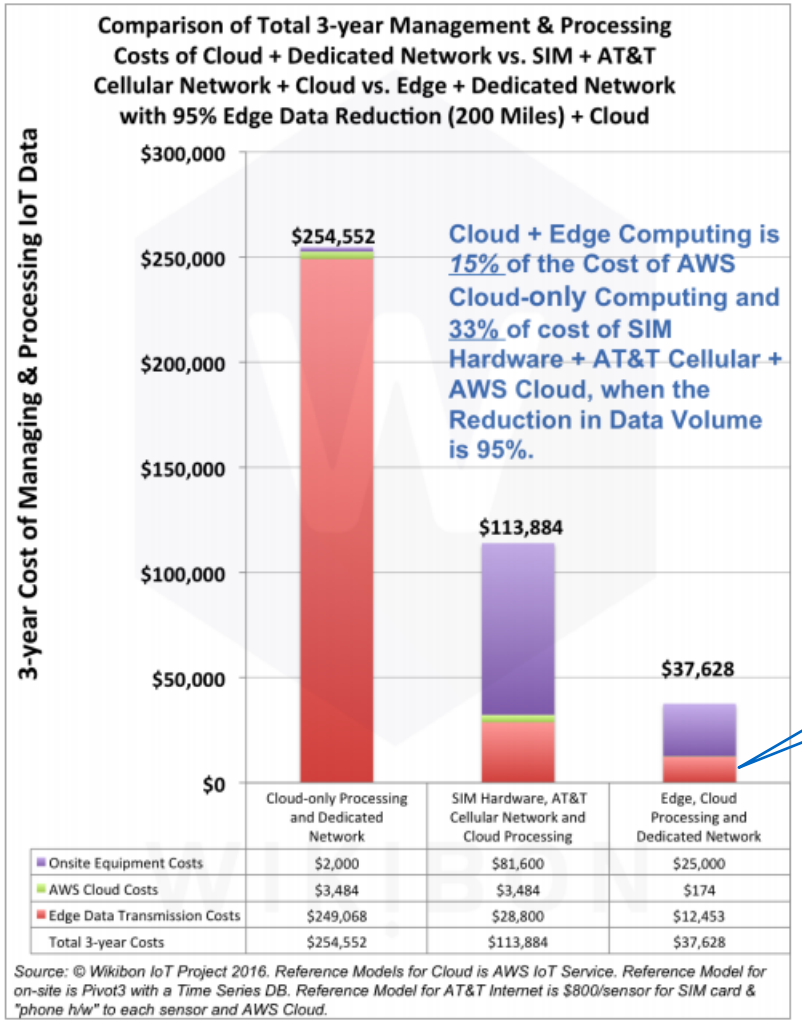
Factor affecting the IoT architecture



Sources: IIC Journal of Innovation 5th Edition

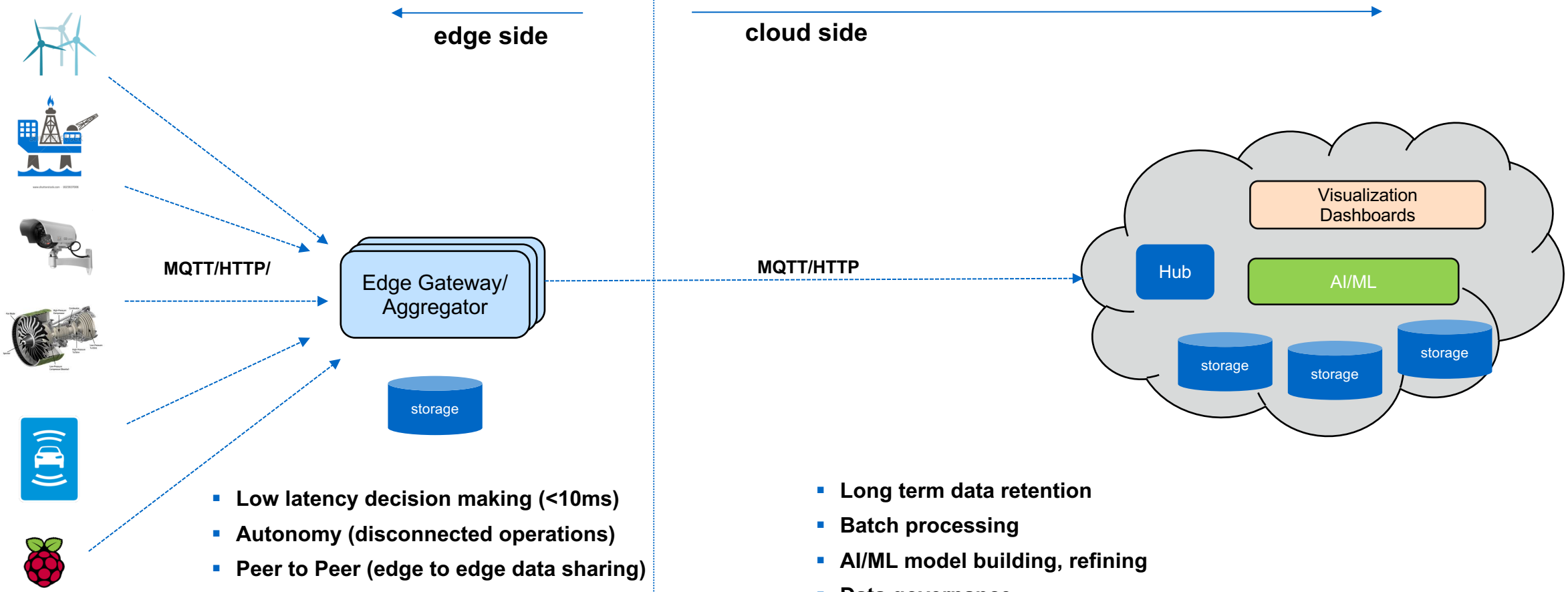
IoT – Why edge ?

Factor affecting the IoT architecture



95% data reduction at edge

IoT – Architecture

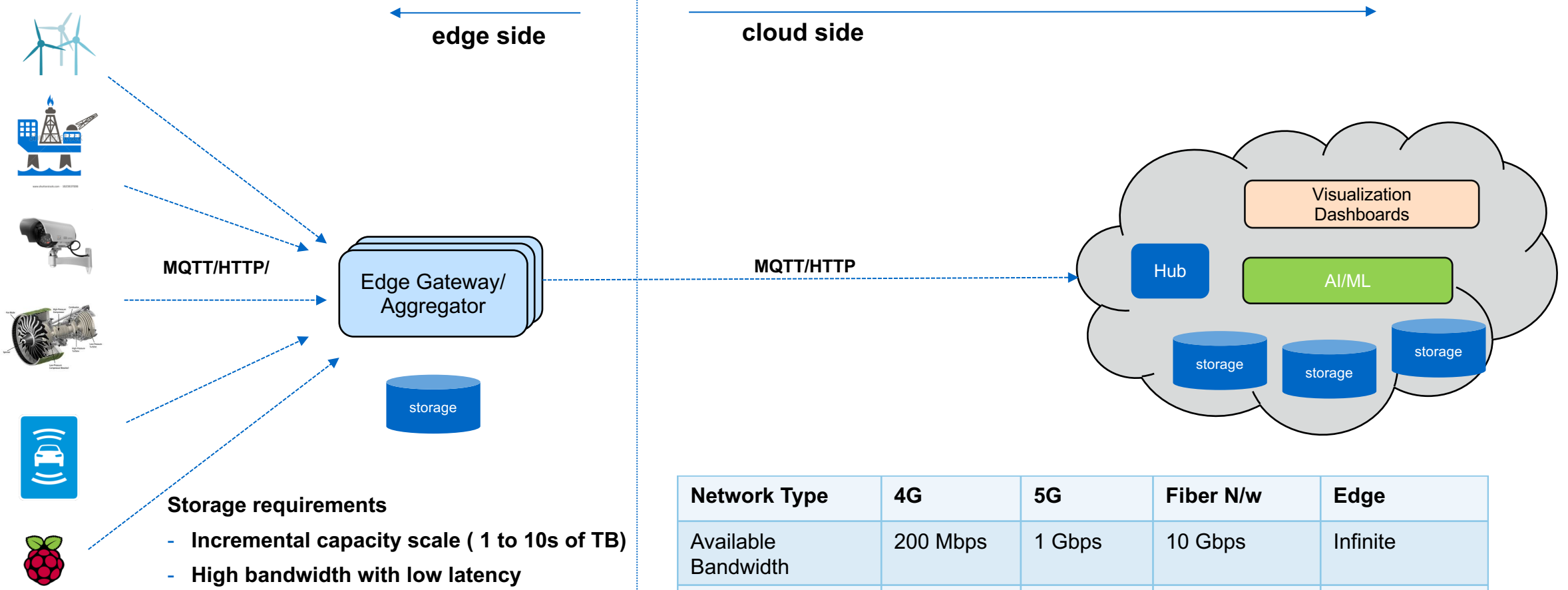


- Low latency decision making (<10ms)
- Autonomy (disconnected operations)
- Peer to Peer (edge to edge data sharing)
- Prioritization (Importance of the data)
- AI/ML near real time inferences

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

NetApp

IoT – Architecture



Storage requirements

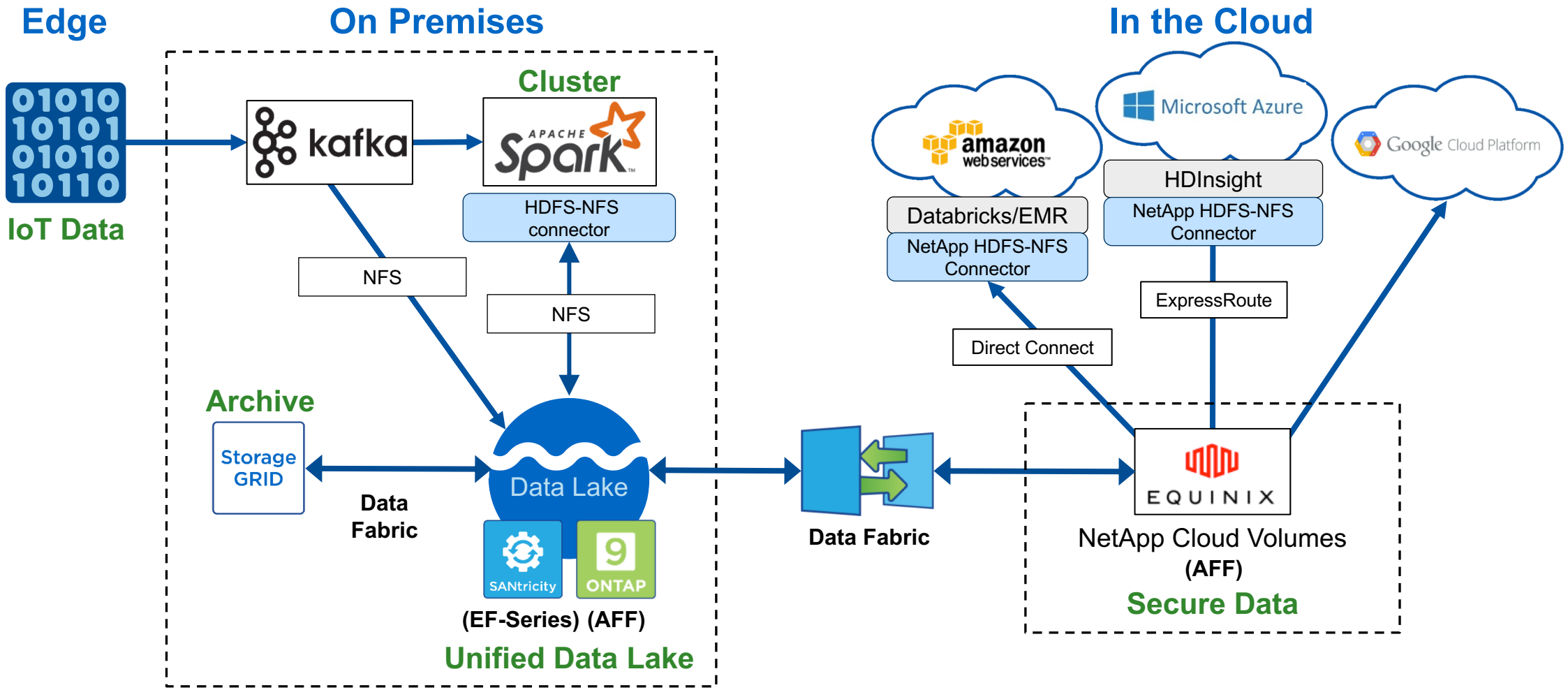
- **Incremental capacity scale (1 to 10s of TB)**
- **High bandwidth with low latency**
- **Optimized for time-series large blobs**
- **Storage optimized for containers (CI/CD)**

Network Type	4G	5G	Fiber N/w	Edge
Available Bandwidth	200 Mbps	1 Gbps	10 Gbps	Infinite
Time required to transmit 1 TB	~ 15 hrs	~ 3 hrs	~ 20 minutes	Real time

NetApp

Case study: Active IQ (telemetry data from deployed systems)

The new Active IQ IoT architecture leverages AFF and the Data Fabric



NetApp

Value Proposition

Agility, flexibility, efficiency and lower cost

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

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

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

NetApp

Conclusions

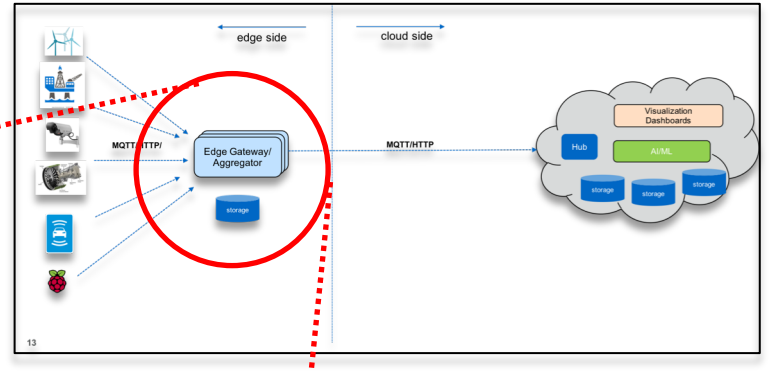
1) Security, low latency, autonomy, prioritization and edge intelligence (AI/ML) continue to drive the enterprise IoT storage architecture
2) Hybrid cloud analytics is faster then pure cloud based solutions
3) Edges footprint offers opportunity to add generic or specific data management services for different industrial verticals

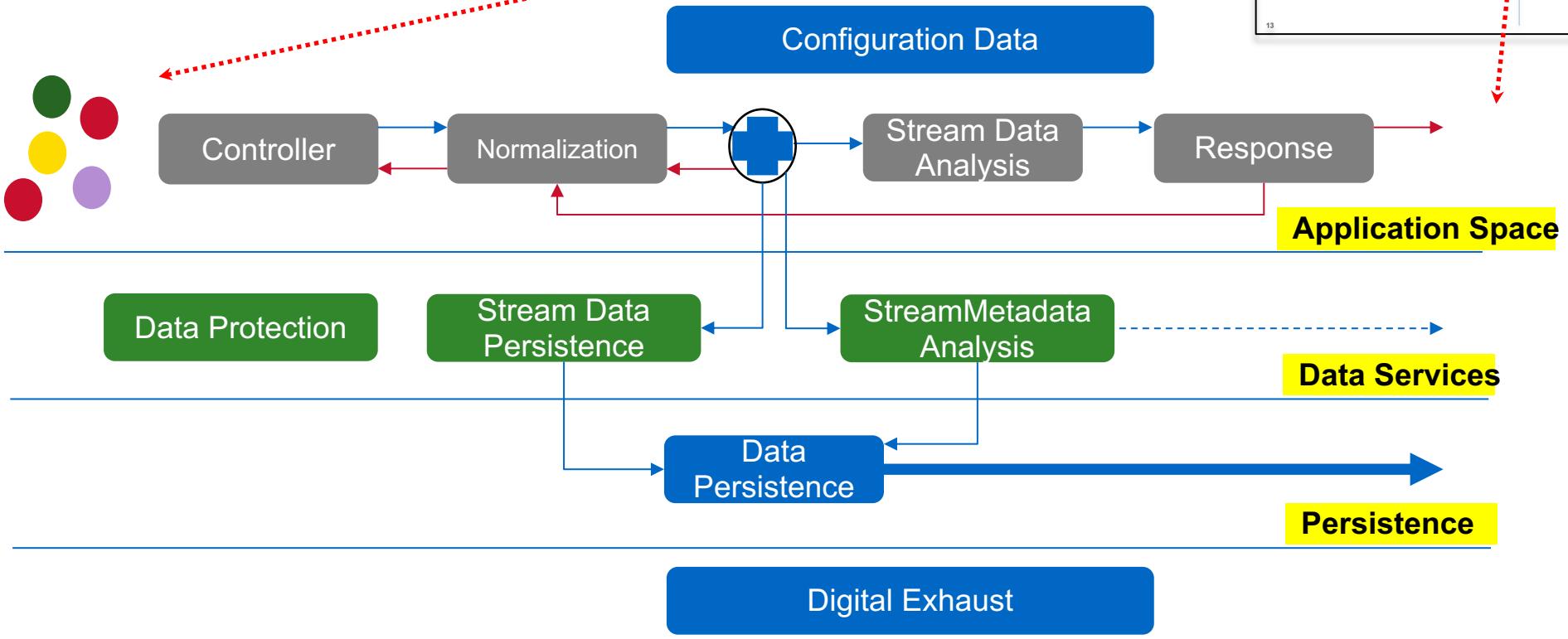
Thank You

NetApp

Backup Slides

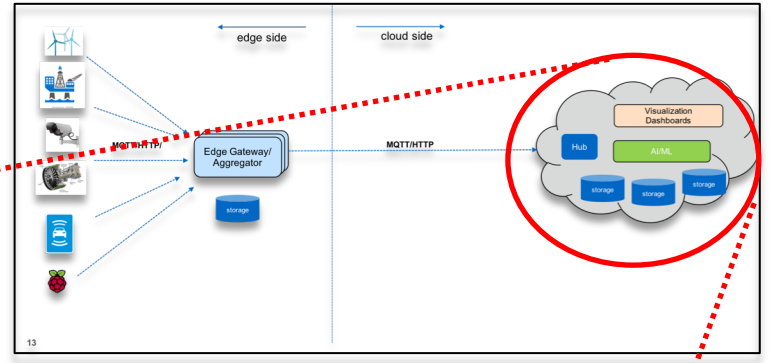
IoT – Architecture : Data Flow at the Edge

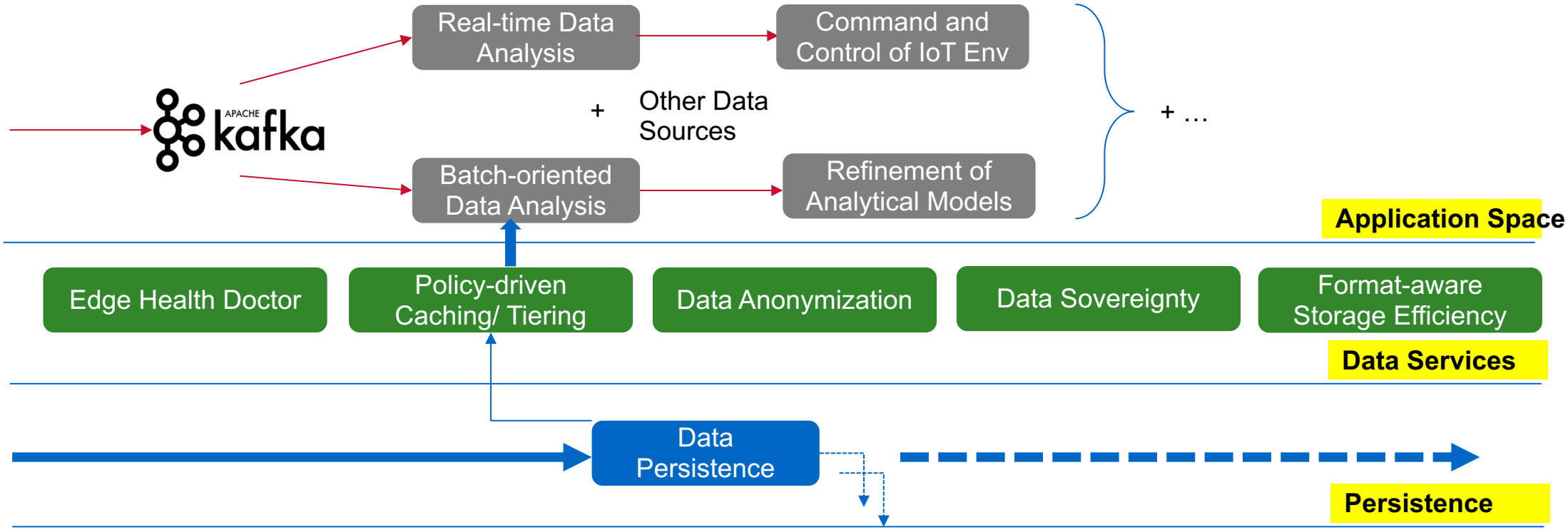




NetApp

IoT – Architecture : Data flow at the lake





NetApp

IoT – Architecture : Types of Data at the Edge

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

End of Slides