SNIA. | CLOUD STORAGE CSTI | TECHNOLOGIES

Storage Implications at the Velocity of 5G Streaming

Live Webcast

October 21, 2020

10:00 am PT

Today's Presenters



Michael Hoard Intel Moderator



Steve Adams Intel Presenter



Chip Maurer Dell EMC Presenter



SNIA Legal Notice

- The material contained in this presentation is copyrighted by the SNIA unless otherwise noted.
- Member companies and individual members may use this material in presentations and literature under the following conditions:
 - Any slide or slides used must be reproduced in their entirety without modification
 - The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA.
- Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- The information presented herein represents the author's personal opinion and current understanding
 of the relevant issues involved. The author, the presenter, and the SNIA do not assume any
 responsibility or liability for damages arising out of any reliance on or use of this information.

NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

SNIA-At-A-Glance









50,000 IT end users & storage pros worldwide





What

We

Educate vendors and users on cloud storage, data services and orchestration



Support & promote

business models and architectures: OpenStack, Software Defined Storage, Kubernetes, Object Storage



Understand Hyperscaler requirements Incorporate them into standards and programs



Collaborate with other industry associations

Agenda

- The nature of disruption & our data problem
- Streaming Data and Real-time Edge AI
- Implications to Edge HW & SW
- Storage impact of 5G use cases
- Kappa and Lambda architectures
- 5G data journey
- Streaming storage
- Analytics options for streaming and static data
- Moving your data from the edge to the core data center



Premise: Major Disruption Occurs When 3 Vectors of Innovation Align to Meet Unmet Needs

Technology Innovation

- AI, 5G, Edge, + distributed storage
- Cloud Native Architectures (e.g. µServices)

Business Model Innovation

- On-demand
- XaaS
- Scales Technology & Business
 - Federated Composable Distributed Systems
 - On-demand Composable Distributed Systems

Next Disruption, will come from a data problem: Today, turning Data into Insights is too slow, costly, and leaky





Problem: The Ratio of "Data-Creation" to "Action" is Way Out of Whack



* Forrester



Goal: Turn "Unused" \rightarrow "Used", Which Requires Mastering <u>Streaming Data</u>



* Forrester



Edge Computing + 5G + Cloud Native = Match Made in Heaven



Common Programming Model







Streaming Data vs. Batch Data

- Use cases for stream processing:
 - Fraud detection
 - Surveillance
 - Social media sentiment analysis
 - Log monitoring
 - Analyzing customer behavior
 - Industrial Computer Visioning



Batch processing	Stream processing
Data is collected over time	Data streams continuously.
Once data is collected, it's sent for processing	Data is processed piece-by-piece.
Batch processing is lengthy and is meant for large quantities of information that aren't time-sensitive.	Stream processing is fast and is meant for information that's needed immediately.

 Batch data is sometimes described as Postmortem data, used to take action well after the event of the data creation.



Common Programming Model: Scripts & Queries





SNIA

CLOUD STORAGE

TECHNOLOGIES

Scripts & Queries Extract Value at Multiple Locations



TECHNOLOGIES

Mixed Streaming + Static Data Processing Defines the Real-time, Distributed Edge (Marketecture view)





Streaming + Static Data Processing Drives Evolution of Server Ingredients: Compute/Move/Store Converge



Implications: AI everywhere

- Servers enable real-time AI
- Storage becomes more tiered
- Network moves based upon

data type (not just NW protocols)



The Next Decade: The Era of Federated On-demand Distributed Autonomous Composable Systems (FODACS)







Storage Implications Created by 5G



Chip Maurer (Dell)



5G Is Driving A New Revolution¹

First Industrial Revolution

- Transition of hand production to machines
- Between 1760 and 1820

Second Industrial Revolution (aka Technological Revolution)

- Railroads, telegraph and electricity
- Modern production line
- Between 1871 and 1914

Third Industrial Revolution (aka Digital Revolution)

- Shift from analog and mechanical technology to digital
- Late 20th century

Forth Industrial Revolution (aka Industry 4.0)

- Interconnection machines, devices, sensors, and people
- Information transparency rapid decisions based on information
- Technical assistance machines helping humans
- Decentralized decisions decisions made at the edge

1 Forbes - Why everyone must get ready for the 4th industrial revolution



Storage Impact of 5G Use Cases

Connected Cars

Some cars could generate between 5TB and 20TB per day¹

Surveillance Video

- Low quality: 300 MB per hour²
- 4K UHD: 7 GB per hour²

Sensor Data from Smart Factories, Smart Farms, Smart Grid, etc

- Low estimate
 - Assume 10,000 sensors, each sending 5 1KB readings per second
 - Data requirements
 - 50 KB per second OR 175 MB per hour
- High estimate
 - Assume 10,000 sensors, each sending 1000 1KB readings per second
 - Data requirements
 - 1 GB per second OR 3.6 TB per hour

1 IOTNow Transport - Data storage is the key to autonomous vehicles future 2 cctvcalculator.net



We Were Already Drowning In Data

- 90% of the data in the world today has been created in the last 2 years¹
- Over 40% growth projected over the next 2 years
 - 175 zettabytes by 2025²
- What good is all that data if you can't process it
 - Wasted \$ on storage costs
 - Lost \$ on unperformed analytics





1 IBM - What is big data? 2 Seagate – Data Age 2025

5G Data Journey From Edge Nodes



Core Data Center(s) Compute + Storage



Kappa and Lambda Architectures

Lambda¹

- Not to be confused with AWS Lambda or Python/Java Lambda functions
- Used to perform both near real-time (streams), and batch (traditional ETL) processing
- Data sources can be streams, objects
 - Typically different code for stream and historical data access
- Typically supports much larger datasets than Kappa architecture





Kappa and Lambda Architectures

Kappa¹

- All data is event stream (append only)
- Used to perform real-time processing
 - Real time by reading stream (e.g. Kafka topic)
 - Historical analytics also possible by reading stored stream data
 - Typically permits same code for stream and historical data access







Edge Node Characteristics

- Use a streaming input for 5G data like video, sensor data, etc
 - Suitable for unbounded data
 - Object data like images can go to stream or directly to storage
 - Low cost storage
 - Likely direct attached, but cloud options available
 - High performance
 - Need to be able to process large quantities of data in short period of time
 - Low latency

⊎\$⊎

- Real time decisions based on 5G data is critical
- Connected nodes
 - Able to collaborate with other Edge devices
- Potentially disposable data
 - Not all 5G data must be kept
- Secure
 - 'Nuff said



Streaming Data Ingestion Options

Stream ingestion (short term storage)

- Data remains for configurable time, or configurable size is reached
- Examples
 - Apache Kafka
 - RabbitMQ
- Stream ingestion (durable storage)
 - Data remains on storage until removed
 - Examples
 - Dell Technologies Pravega
 - Oracle OCI Streaming Service



Processing Data At Edge Nodes



Always working on the most recent data

Low latency stream processing tools

- Apache Flink (true streaming)
- Spark Streaming (micro-batches)
- Knative (Kubernetes)
- AWS Lambda, Azure Functions, Google Cloud Functions (event driven, serverless)

Filter and reduce data

- Keep the important data, discard the rest
- Custom and specific data
 - Each data stream and use case is unique
- AI and ML working on streams of raw 5G data
 - Well suited to process large volumes of data



Public Cloud Edge Solutions

- Public cloud providers have edge solutions
- Bringing cloud data processing and analysis closer to end-points



- AWS
 - AWS IoT *
 - AWS IOT Core, AWS IOT Greengrass, AWS IOT Analytics, etc
 - Lambda@Edge



- Azure
 - Azure Stack Edge
 - Azure Arc



Google Cloud Platform

- Google Cloud Platform
 - Global Mobile Edge Cloud



Core Site 5G Data Storage Destination

- Ingest incoming 5G data from edge nodes using stream, object or file
- Plan storage based on latency requirements
- Plan storage based on retention requirements
- Well suited for batch oriented analytics
- Send larger, less often referenced datasets to cloud based storage
 - Public
 - Private
 - Hybrid
 - Multicloud
- Send smaller, more frequently referenced datasets to direct attached storage
 - HDFS
 - Relational DB
 - No SQL DB



Cloud 5G Data Storage Destination

- Low cost
- No maintenance
- Long retention
 - Corporate and state/federal regulations
- Built-in protection
 - AWS S3
 - Four 9s availability
 - Eleven 9s durability
- Automatic tiering options
- Virtually infinite scalability
- Also well suited for batch oriented analytics
- Private cloud benefits
 - Cost savings at scale
 - Ownership and control
 - Compliance for sensitive data

5G Solution Options Can Be Overwhelming!

Chances are, your storage vendor has 5G edge storage and analytics solutions that meets your needs¹

	B PLUSCOM: Scheme Constraints and Scheme Con	SALES BARTONICO CONTRACTOR CONTRA
	Mills SCIENCI SOTIONOS Mills SCIENC	USAL Matter Distance Matter Distance Matte
Fit. Josts Data SCIENCING Data SCIENCING Data SCIENCING TAMANG/ONLETON Banka Domain Source Operation Source Operation Source Operation State & Source Operation Source Operation Source Operation Source Operation Source Operation State & Source Operation Source Operation Source Operation Source Operation Source Operation State & Source Operation Source Operation Source Operation Source Operation Source Operation State & Source Operation Source Operation Source Operation Source Operation Source Operation State & Source Operation Source Operation Source Operation Source Operation Source Operation State & Source Operation Source Operation Source Operation Source Operation Source Operation State & Source Operation Source Operation Source Operation Source Operation Source Operation	CONFUSE VECON CONFUSE VECON CONFUSE	APPUCATIONS INDUSTRY
ADDATI INCOLOREDAS BUILDING CONTRACTOR DATA BUILDING CONTRACTOR DATA	SHARH SH	HENCHE
Version 1.0 - September 2020 © Matt Turck (@mattturck) & FirstMark (@firstmarkcap) mattturck.com/data2020 FFIRST ACC VIRIUME CARY		

1 mattturck.com - Resiliency and Vibrancy: The 2020 Data & AI Landscape



Additional Resources

- 5G Americas White Paper: 5G at the Edge
- SNIA Educational Library
- SNIAVideo YouTube Channel



- 5G's 3 phases, enhanced mobile broadband, URLLC (ultra reliable low-latency communication), and mMTC (massive machine type communication) will connect billions of new things, which will stream data 24/7 needing to be processed in real-time at the Edge.
- 5G storage solutions will require upfront planning to understand and support the data flow, both now and future
- Data Centers will decentralize their boundaries to incorporate the explosion of edge nodes to support 5G enabled devices
- Reduce data content on edge nodes where possible
- Low latency stream processing is critical on edge nodes for real-time analytics
- Data centers will continue to use a mix of local and cloud storage based on data requirements



After This Webcast

- Please rate this webcast and provide us with feedback
- This webcast and a copy of the slides will be available at the SNIA Educational Library <u>https://www.snia.org/educational-library</u>
- A Q&A from this webcast will be posted to the SNIA Cloud blog: <u>www.sniacloud.com/</u>
- Follow us on Twitter @SNIACloud



Thank you!

