The 5G Revolution

Executive Overview: *The 5G Revolution*

In the coming months, Communication Services Providers (CSPs) across the globe will be busy rolling out the 5G network. 5G is fundamentally different from 4G, LTE, or any other network the telecommunications (telecom) industry has ever seen before. It promises data rates 100x of 4G, network latency of under 1 millisecond, supports 1 million devices/sq. km., and 99.999% availability of the network. 5G will generate data at an unprecedented velocity and immense volume. This "fast data" will fuel a wide range of data-driven services and digital business models.

However, to capitalize on this opportunity, your organization needs to operationalize fast data and make informed business decisions in-event and in real-time.

In this paper, we will examine the 5G revolution from a data perspective - what fast data is, what are the opportunities it creates, and what challenges exist in operationalizing fast data in real-time. It will cover all of the essentials attributes required for operational, transactional & analytical fast data processing - and how VoltDB can help you do just that. It is a challenge that CSPs must meet if they hope to succeed in the race to monetize 5G.

VOLTDB

The Fast Data Imperative

Why is 5G data different?

Fast data refers to the velocity of data so high that legacy IT systems and database technologies can no longer manage and effectively process it. 5G data will stream at significantly higher speeds, volumes, and at lower latency that ever before. Database technologies architected in the 70ies were simply not designed to process the vast deluge of fast data generated by the 5G network core and microservices. While NoSQL were a notch better at scale than legacy databases, they fall short on linear scale at low latency and ACID requirements. The successful monetization of 5G necessitates new approaches and technologies.

This requires a real-time actionable decisioning system to fix this fast data problem. IT will need to move away from traditional store and query models that were designed for big volume to more intelligent stream processing database technologies that can take real-time decisions on incoming event data.

Popular 5G Use Cases

CSPs have already started rolling out 5G across select cities all over the world. By 2020 5G is expected to reach full capacity. And while there are a vast multitude of use cases that 5G will unleash, here are some of the most popular.

Widespread Adoption of IoT

Although IoT and MtoM communication has been around for a while, 5G will enable embedded sensors into every device (Industrial, commercial, and even residential). Use cases such as smart metering, smart cities, vehicle telematics, asset tracking, etc will finally become a reality.

Reliable Low-Latency Communication

This will provide the ability to remotely control a smart-grid, self-driving vehicles, robotics, industrial automation, etc.

Fixed Wireless

Wired residential internet accent will be obsolete. As 5G powered wireless technology will offer significantly greater enhancements, such as: millimeter wave (mmWave) spectrum and beamforming.

Superior Mobile Broadband

In addition to turbo charging traditional mobile broadband computing with ultra high speeds and uninterrupted internet connections for devices, 5G will also drive the adoption of virtual reality (VR) and augmented reality (AR).

Enhanced Event Experience

With 5G, event organizers and marketers can offer unparalleled VR experiences to increase market penetration and revenue.

Once 5G is rolled out, CSPs can cater to hundreds of additional industry vertical specific use cases.

New Business Opportunities via. Network Slicing

5G promises low latency, unprecedented scale, enormous throughput, high availability, security and reliability. Every industry vertical demands services with attributes that are customized to their unique use case.

With 5G, CSPs can create customized network slices that cater to the unique needs of each industry vertical, and focus of all the required attributes for each unique use case. A IoT solutions provider would have their own dedicated network slice that would provide the low latency and reliability required for MtoM communication. While an AR vendor would have a network slice with high bandwidth and low latency that is necessary to transmit large video files.

Network slicing reduces network operational costs and also provides CSPs with new opportunities to monetize customized services for a wide variety of use cases.

Network Functions Virtualization (NFV)

NFV enables network slicing by replacing network functions on appliances such as routers, load balancers, and firewalls with virtualized software instances that run on commodity hardware. With NFV, multiple virtual networks (each supporting different Radio Access Networks) can be created on top of a shared physical infrastructure. Each virtual network dedicated to an unique customized services can be spun up easier and faster than ever before. The control pane and the user pane of network slices is distinct, hence capable of delivering an customized user experience with each individual network slice.

Virtualized Network Functions (VNFs), when coupled with the distributed cloud (which allows for multiple data centers to appear as one unified virtual data center) enables customized deployment of VNFs based on attributes such as: latency, bandwidth, availability, resiliency, etc.

Software Defined Networking (SDN)

SDN is complementary to NFV, it utilizes network management to separate the control plane from the forwarding plane. SDNs enable programmable network controls, while abstracting the underlying infrastructure from the apps and network services. Centralized and controllable, SDNs provide the agility required to adapt to the evolving needs of 5G apps and services.

Fast Data Challenges

Network slicing, NFV, and SDN are new approaches adopted by CSPs to drive operational efficiency, agility, and deliver new profitable business models. However, another key piece of the new business models monetization puzzle is generating value from all of the data that is available.

5G will support higher than ever connection densities: up to 1 million connected devices per .38 square miles, compared to around 2,000 connected devices per .38 square miles with 4G. 5G will also carry a lot more data and transfer it much faster than 4G.

The high device density and much faster network speed will generate massive volumes of data streaming in at blazing fast speeds. 5G apps and services will have to keep-up, by performing transactional processing and analytics on streaming data in real-time.
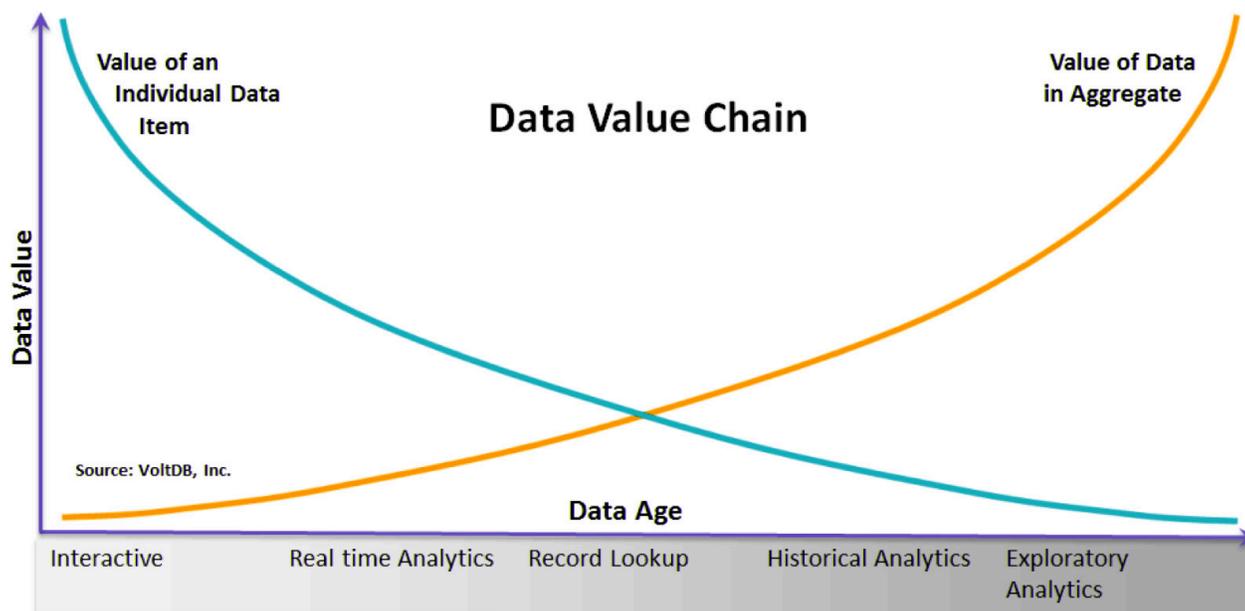

Figure 1: Diminishing Value of Data

The business value of data diminishes as the data gets stale. This phenomenon is extenuated even further for 5G data processing, where a lag of even a few milliseconds could lead to not just loss of revenue (in use cases such as: customer experience management, customer churn, fraud detection and others) but potentially even human life (in IoT use cases such as self driving cars). Real-time actionable analysis is the future in the era of 5G. Action oriented decision making while the data is fresh is the next key differentiator for data driven CSPs. Real-time now means milliseconds, 5G demands moving from post event reconciliation to in-event data processing. If action is not taken immediately, the opportunity to monetize an event is lost.

Unlike 4G, the 5G control pane is data intensive. 5G core functions are split out into granular components, each with specific functions. These become highly scalable and distributed apps that are rolled out in an hybrid cloud model, to be deployed as necessary. Each can be customized to deliver on the latency and throughput profile required for the individual 5G network slice. In this architecture, real-time interaction between the components and with dynamic third party components is critical. 5G microservices need to be spun up at a moments notice to meet demand.

Each microservice performs a dedicated function and is capable of communicating with other services through stateless interfaces. With microservices the complex 5G ecosystem can be broken down into consumable individual components. Each lightweight component can be easily and quickly created, distributed, scaled up or down, and updated without consuming significant development or computing resources.

While 4G was introduced to meet the demand of the growing number of internet connected smart phones. 5G is expected to enable zillions of connected devices (i. e. things) powering IoT. Sensor data from industrial equipment / products, buildings, utility meters, vehicles, residential devices, wearables, etc will be generated at volumes too large to transport and store in a centralized data center. Additionally, a vast number of Industrial IoT use cases, such as predictive maintenance demand processing data in real-time, the risk of a delay can mean equipment downtime and loss of revenue. This demand

for real-time, high volume data processing can only be met by edge computing.

Database Requirements

As discussed in the previous section; 5G demands software capabilities and methodologies that are new to CSPs and telecom software solution providers. In this section we will enumerate the critical database components required from the new 5G database, to be able to implement and deploy the control plane elements.

High throughput & Low Latency

5G ITU IMT-2020 specifications demands speeds up to 20 gig per second and a target latency of 1 ms. This unprecedented combination of high speed and low latency will open the flood gates on data streaming in from fixed wireless, IoT, Video on Demand, Virtual Reality, and other apps. Provisioning and completion decisions will need to scale with the new higher volumes that 5G enables. The new business models driven by 5G: QoS/Provisioning models, and opportunities to deploy new cloud microservices closer to the edge will drive new requirements on the existing infrastructure. Legacy databases & disparate NoSQL data stores are just not capable of performing complex transactions and analysis on this high throughput data (millions of tps) at low latency (in ms).

Linear Scale

As discussed earlier, 5G networks will generate vast multitudes of streaming data from not only 5G enabled apps, but also from the network, subscribers, enterprise users, network operators and call processing. It would be essential for the 5G database to scale linearly at a moments notice while maintaining the high performance and low latency requirements all along.

HA / DR / XDCR

The 5G network promises a network availability of 99.999%. If the network goes down even for a second the microservices / apps, along with the consumer

subscribers / enterprise users that rely on them incur significant revenue loss, and in some IoT and healthcare use cases even potential human life could be lost. The database powering the apps and microservices is expected to have High Availability (HA) especially across multiple geographies (XDCR), and Disaster Recovery (DR) built-in. Additionally, 5G's latency requirements are causing CSPs to move to active-active XDCR, as the speed of light makes a single large data center too slow for end users.

Virtualization / Cloud / Dockerization

The scale, elasticity, agility, responsiveness and rich software functionality required for 5G microservices can only be achieved in the cloud. While lower CAPEX & OPEX is a additional benefit of being in the cloud. Today NFV is done on VMs, and VMs will be continue to be utilized in an 5G environment, containers will shoulder the bulk of the load in building and deploying 5G microservices. Containers offer the agility to spin up or spin down microservices and enable a DevOps culture that is necessary in the 5G era. Containers can also be easily orchestrated using open-source tools such as Google's Kubernetes. The modern 5G database need to seamlessly work in the cloud, and also be orchestrated in a containerized environment.

The real challenge for CSPs and software solution providers is to fulfill all of 5G's fast data SLAs with one underlying database.

VoltDB — The Real-Time Database Architecture for 5G

VoltDB is an open source SQL-based relational database that was designed to power apps which require real-time intelligent actionable decisions on streaming data, without compromising on ACID requirements. Here are the features that enable VoltDB to execute fast transactional processing:

- **In-memory Architecture** — Traditional disk based databases are extremely slow for 5G microservices and apps. Reading and writing data from/on disk is a laborious process. Memory however; is much faster than disk/flash drives. With historically low memory prices, the time is ripe for in-memory databases.

- **Automatic Partitioning** — Sharding (partitioning) data and queuing commands to the relevant partition, distributes the work and enables scale across multiple CPUs. Limitless scale can be obtained by adding multiple partitions to a server and then multiple servers to a cluster allowing the system to consume many CPU cores, each running per-partition commands.

- **Full ACID Transactions** — Partitions run independently, while providing ACID semantics for the commands processed from its associated command queue.

  - *Atomicity* — While executing the commands, the program maintains in-memory undo logs so that aborted commands can roll back, ensuring atomicity.

  - *Consistency* — Constraints and data types are enforced during the execution of each command guaranteeing consistency.

  - *Isolation* — Commands run one at a time without overlap, providing for strict serializable isolation.

  - *Durability* — If all of the commands have deterministic side-effects (running the same queue of commands in the same order against the same starting dataset is promised to produce the same ending dataset), then writing (and fsync-ing) the command queue to disk before executing the commands makes all data changes durable.

VoltDB's stored state data and strict serialization can help take some of the complexity out of SDN by maintaining state and ensuring 100% ACID on the network operations data. This enables the app/microservices developer to not worry about programming for ACID, saving valuable time and thus bringing the microservices to the market faster.

- **Replication** — Partitions are replicated to provide for fault tolerance. If a server crashes and its partitions are terminated, the cluster can continue to operate using the other copies available on the surviving servers.

- **Stored Procedures, User Defined Functions, and Materialized Views**:

  - *Stored procedures* — Pre-compiled server-side stored procedures make the in-memory database execution times even faster; by running complex transactions at scale and with ease.

  - *User defined functions (UDF)* — UDFs enable the user to run custom logic inside a SQL statement, simplifying apps and allowing developers to create new apps faster.

  - *Materialized views* — Users can declaratively pre-aggregate query results on a mutating dataset with materialized views, resulting in answering complex aggregation queries in under a millisecond.

- **Embedded Machine Learning** — The 5G database must support operationalization of ML models. In an production environment, the model needs to continually ingest, train on historical data and operationalize in real-time at very low latency. The operational database is able to convert a ML model into an executable process as a user-defined function (UDF) that is then embedded into the database to take real-time actions on high velocity streaming data, and deliver the desired business outcomes.

- **Kubernetes Orchestration** — The 5G operational database needs to run in an distributed container-ized environment. Up until now, orchestrating SQL databases in Kubernetes had been very challenging as operational database systems have state and can't just be spun up or spun down on a moment's notice. Recent developments ensure container orchestration with Kubernetes, turning many tedious and complex tasks into something as simple as a declarative config file, and allowing for continuous and frequent deployment.

- **Cloud Ready** — 5G powered microservices and apps will be cloud native. The operational database needs to provide horizontal scale by clustering on commodity hardware, with hardware and networking fault-tolerance built-in.

Key Takeaways

With the emergence of 5G the need for real-time actionable decisoning is greater than ever before. The perfect solution for a 5G powered world; is a fast and scalable in-memory operational database that can ingest millions of transactions per second, either in the cloud or on-premises. With a real-time database powering next-generation apps & microservices; developers can monetize their fast data to create a competitive advantage. Transforming their infrastructure from post-event to in-event and actionable.

Test drive VoltDB at: https://www.voltdb.com/try-voltdb/download-enterprise/

18 December 2018

VOLTDB