

Building a Network Data Fabric at Hyperscale

VelociData delivers a cohesive data solution to a broadband network provider by making use of Intel® technologies.

What is a data fabric?

A data fabric can first be thought of as coherent many-to-many interconnectivity among multiple data sources and multiple data consumers. In this case, the data sources are the millions of devices in the network generating telemetry data, and the data consumers are the people and systems that seek to extract different kinds of business value from that data.

But a data fabric represents more than just a vast number of data connections. By analogy, if device data is the raw material (wool) that is transformed into usable information streams (yarn), then the data fabric emerges when all these streams are woven together to form a tapestry—a complete, time-coherent picture of diverse network data as a whole. This big-picture data fabric is what enables operations, engineering, marketing, and business personnel to gain full network visibility and insights that provide business value.

A major broadband network operator with millions of customer endpoints faced challenges in trying to extract value from the telemetry data that its network equipment generated. VelociData helped the operator overcome those challenges using a data-fabric solution built on a wide range of Intel technologies. Now the network operator can harness vast amounts of streaming network data in real time to fight fraud, reduce capital expenditures (CapEx), recover revenue, and improve quality of service (QoS).

The promise and challenge of harnessing network data

Broadband providers are in the business of delivering fast, reliable Internet services to millions of customers. Accomplishing this requires millions of network devices, from customer endpoints to cable modem termination systems (CMTSS) to gateways to the giant routers that form the backbone of the network. These devices not only steer data traffic to and from customers, but they also generate massive streams of real-time data about device usage and state. Harnessing all that network data can provide extremely valuable insights to help guide company operations, from capital-investment decisions to fraud detection and customer support.

One such network operator pursued a vision that the network itself could provide significant insight into what was going on within the network: how busy it was, where it was busy, how it was being used, what kind of traffic was being transported, and where traffic volume was likely to grow. In order to get real value from this data, the company realized it needed an architected data fabric designed to make the entire network data stream usable for analytical modeling, artificial intelligence (AI), and decision science.

Working with real-time streaming data at this scale—not millions but tens of billions of rows of data every day—would be difficult under the best of circumstances. The challenge is not just to collect the data but also to unpack it, transform it, and analyze it in real-time at Internet scale, without losing any valuable information, even if there are slowdowns or failures in the components or interconnections.

This challenge is made more difficult by a heterogeneous network of diverse hardware, software, and data protocols that were not designed to work together, and that don't always follow established standards.

Building a solution with VelociData

The network operator turned to VelociData for a comprehensive approach to acquiring, curating, delivering, and analyzing the nonstop firehose of data from network elements, which represents a large and strategically important piece of its overall data fabric.

Case Study | Building a Network Data Fabric at Hyperscale

VelociData set out to enable timely, actionable insights from the never-ending stream of complex data in many different formats and structures from devices across the network. The solution needed to accomplish the following objectives in real time:

- Acquire data at fast speed and high scale
- Parse the data's diverse and complex protocols
- Maintain full-time coherence across multiple streams
- Format the data for usability in analytical processing
- Deliver the data to human and digital consumers
- Analyze and present the data in the most actionable way

All of this must be done with data in constant motion, and the solution must retain data integrity and consistency no matter what happens within the system. Slowdowns, failure conditions, and data gaps must all be dealt with in real time, or the data's informational value will be lost.

VelociData delivered the solution in the form of appliances built on standard Dell Technologies servers, such as the Dell EMC PowerEdge R740 server. The servers are equipped with advanced Intel technologies, which VelociData utilizes fully to provide an end-to-end solution with performance and resiliency.

Key Intel technologies used in the solution include:

- [Intel® Arria® 10](#) field-programmable gate arrays (FPGAs) for data parsing and indexing at 10 gigabits per second (Gb/s) rates, providing usable data formats such as comma-separated values (CSV), XML, and JavaScript Object Notation (JSON). Intel-based FPGA SmartNICs are on the VelociData roadmap.
- Intel® Xeon® Scalable processors (initially [Intel Xeon Gold 6150](#) processors) with specialized Streaming SIMD Extensions 4.2 (SSE4.2) vector instructions, especially for data parsing (tokenizing the input stream).

- NUMA-optimized streaming algorithms in memory to fully utilize available bandwidth by putting data close to the processor that needs to use it.
- Intel Ethernet Converged Network Adapter XL710 network-card hardware for packet capture up to 40 Gb, and Data Plane Development Kit (DPDK) for packet capture and inspection—this replaces an interrupt-driven driver with a well-implemented poll-mode driver that enables higher throughput while helping to lower CPU overhead.

VelociData spent significant effort optimizing its applications to make use of the advantages of the hardware platform. Developers ensured tight alignment between the compute and storage elements, made careful use of the vector instruction set, and optimized parallel processing for low latencies. VelociData also built its own FPGA Direct Memory Access (DMA) engine, driver, and libraries that resulted in extremely low-latency communications between the FPGA and the CPU memory.

The appliances run two different VelociData software packages known as Raptor and Vortex, as illustrated in Figure 1. Raptor understands the languages of the network devices, and Vortex speaks the languages of the data-analytics and AI tools.

Raptor is the data-collection part of the solution. It connects to all the different elements in the network (such as routers and CMTSs) to extract data using protocols that include Simple Mail Transfer Protocol (SMTP), IP Detail Record (IPDR), NetFlow, streaming telemetry, and others. Raptor understands all the different protocol formats, data fields, and nuances of various brands of equipment. It parses all the fields of data from all the network elements and transforms the data into VelociData's proprietary common data format for use in network monitoring, reporting, and automation applications.

Vortex integrates data from the Dell Technologies servers running Raptor with data from other systems and application sources to create and deliver a comprehensive

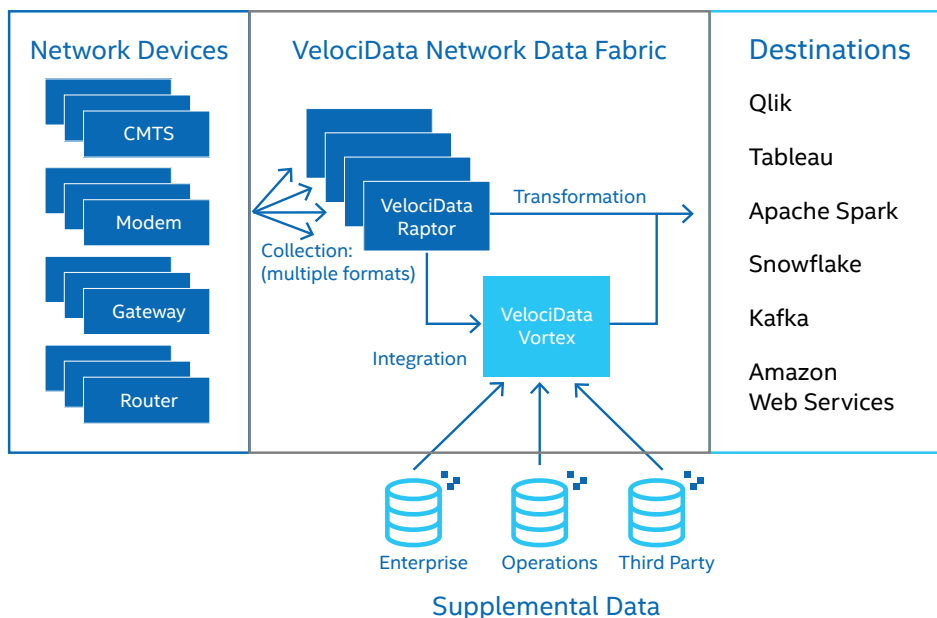


Figure 1. VelociData network data-fabric architecture

network-business view to multiple destinations including Amazon Simple Storage Service (Amazon S3) and Kafka in a standardized format for consumption by machines and humans alike.

About VelociData

VelociData is a leading innovator in real-time streaming data collection, processing, and delivery, and it is working closely with Intel to help customers get the most out of their systems. Offering software that is easy to use, exceptionally reliable, extremely cost-effective, and backed by a deep portfolio of fundamental streaming data patents, VelociData provides the world's largest, fastest enterprises with powerful real-time data solutions for financial services, healthcare, network management, and more.

The benefits of a well-architected data fabric

The VelociData solution for this network operator has been running reliably for several years now. The operator has benefited from a broad set of services and insights resulting from the availability of this data fabric at scale.

Broadband is a capital-intensive business, and the data fabric has given this network operator an improved ability to predict future utilization and make capital-investment decisions accordingly. For example, the data fabric can help in decision making about how the company should deploy more fiber—in what geographies, for which users, at what cost, and with what expected return on the investment (ROI). The company has developed advanced models to help answer these kinds of questions, based in part on the data provided by the VelociData solution.

Another kind of modeling has helped with fraud detection. Hacked cable modems, or “clones,” that spoof the network to look legitimate are a significant problem for broadband providers.

Analysis using machine learning (ML) has enabled this operator to identify and disable clones, and in some cases even increase revenue by selling legitimate modems and contracts to the people who were using the clones.

A number of the operator's data applications revolve around diagnostics. Customer support and its infrastructure is a huge expense for any broadband provider, so applications that can pre-emptively identify and diagnose network problems represent a major opportunity for cost reduction. Suppose, for example, a streaming service is pixelating for multiple customers. Data from network devices might identify the problem before it's even reported by a customer. And based on where the problem is occurring on the network, automated diagnostics might be able to determine if the source of the problem is more likely in an apartment building or at a node connecting a larger number of customers. This real-time visibility across the entire network allows the operator to conduct a more granular and capable set of diagnostics that can reduce costs. It is costly, for example, to send a crew to troubleshoot at an apartment building when the problem is located elsewhere.

The network operator has discovered numerous other uses for its data fabric. The operator has built customer personas based on the usage data, for example, and used those personas to improve the process of new product development. The operator has also been able to model the impact on QoS that is likely to result from rolling out a new product or service—which is far more efficient than the old empirical method of rolling the product or service out and then watching to see what happens.

Network data stream harnessed

VelociData has taken advantage of a broad array of technologies and systems from Intel to create a nonstop data system that provides insights into, and greater optimization for, business decisions and operations for a large broadband network provider. VelociData delivered millions of dollars of increased value realized by the network operator from its real-time data by giving the company the ability to optimize capital planning, detect fraudulent behavior, improve QoS, and contain operational expenses.

Learn more

Read how VelociData reduces data friction: <https://velocidata.com/>

For more information about Intel programmable logic: <https://blogs.intel.com/psg/>

Visit the Intel Acceleration Hub: intel.com/content/www/us/en/programmable/solutions/acceleration-hub/solutions.html



Intel technologies may require enabled hardware, software or service activation.

No product or component can be absolutely secure.

Your costs and results may vary.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.

Printed in USA

0121/TM/PRW/PDF

Please Recycle 345596-001US