White Paper

Network Security

Privacy and Data Confidence in Edge Clouds

The growth in AI and edge networks is increasing demand for measuring trust in real time. Dell Technologies and Intel are jointly exploring Data Confidence Fabrics (DCFs) to enable trust to be implemented and measured in edge ecosystems.

Authors

Steve Todd
Dell Technologies

Paul O'Neill
Intel

Ben McCahill
Intel

Sanjay Bakshi
Intel

Table of Contents

Executive Summary

From grammar corrections to image recognition, from traffic routing to driver assistance, AI inferences and judgments are everywhere in our lives. These pervasive tools are now being used as aids in critical judgments. AI acts on very large data sets, some of which are beyond the abilities of humans to review. The impact of AI on decisions made by business, government, and devices—without the firewall of a human review of these massive data sets—makes it very important that the data and the AI algorithm are trustworthy.

Data privacy is a requirement for good AI and a critical trust issue. A recent Omdia[1] survey of nearly 300 enterprises across all geographies reported that 60% of end-user respondents say lack of data privacy is slowing or significantly slowing their AI initiatives. AI technologies deliver considerable benefit, but trust in that technology must be earned by AI providers in order for customers to fully "lean in" to the technology and reap the benefits.

Until now, encryption has been the most-used privacy tool for data in transit across public networks or the internet to the data center. But with the advance of AI and other technologies, it is becoming increasingly important to develop additional data privacy and confidence tools that provide an independent audit and measurement capability for the data being used, the underlying heuristics, as well as for any inferences generated. This capability must be designed to preserve privacy and confidentiality with transparency regarding actions of involved parties.

Another factor that weighs into the need for more trust in networks is the trend toward locating data processing at the network edge in order to handle increasing data volumes and the impact of emerging data economies as well as to comply with data locality regulations. These edge cloud servers are located in points of presence, at base stations, and at other edge locations where the processing can take place closer to the data source. Edge data processing offers the promise of reduced data movement and reduced transportation latency, while also addressing data privacy and confidentiality concerns and supporting real-time insights.

For example, one of the leading requirements for edge networking is privacy-preserving predictive analytics. This requires very small decision latencies that are different to data processed in the cloud: this processing typically occurs outside the boundaries of data center firewalls and beyond the reach of dedicated data center security teams, potentially creating new data trust issues.

Processing trustworthy data means creating an agreed-upon source of truth in a complex operating environment. This depends upon data providers and data consumers agreeing on the provenance of the information and policies governing the information. This requires a digital supply chain audit capability where participants understand and agree on what happened, when, and why—without breaking compliance and privacy rules.

Measuring trust at the network edge where typical usage involves multiple parties with and without pre-existing business relationships is something that has never been done before. Dell Technologies and Intel are tackling this challenge by jointly creating the industry's first Data Confidence Fabric and offering communications service providers (CoSP) the ability to trial this technology in their own network.

Extending Data Center Trust to the Network Edge

The very large data sets and use of AI make data trust measurement important to the success of AI and services that are based on its decisions. There are a wide range of factors that are driving the need to quantify trust, including:

- **Temporal:** Complex environments with multiple parties often require very low decision latency. How do you expose the "working data" of a partnership without exposing trade secrets, valuable data sources, or IP? How do you audit the results? How can parties with no prior interaction transact in real time at scale?
- **Societal:** An accelerating erosion of trust as centralized platforms mishandle and monetize data in inappropriate ways. These concerns have caused some in the industry to pull back on the use of facial recognition,[2] and have hampered access to health data that could help with diagnoses and cures. The latest example of this is contact tracing applications designed to help slow the spread of contagious disease.
- **Technological:** Detecting or preventing the use of fakes and the ability to weaponize misinformation. Language models are ushering in the age of false writing. Similarly, "deep fake" videos are so good they are resulting in the spread of misinformation. Countering the increasing use of training in data silos that can confer uneven competitive advantage and where training is not subject to open scrutiny for resulting bias.
- **Financial:** "Trust friction" refers to silos of information owned by different parties in a transaction that require constantly repeated checks and a complicated value chain. Existing systems contain manual respondents that slow things down and add little value. A single shared version of "what is true," along with real-time settlement of this capability, allows for monetization of data along with insights in real time once the data is validated and tokenized.
- **Autonomous:** In the machine-to-machine economy that is driven by exchanging IoT data, machines will autonomously take actions without human intervention.

The number of connected IoT devices continues to grow, generating more data from trusted and untrusted sources that applications need to ingest and process. This can result in unforeseen costs—both in time and the costs of managing the data transfers. In some situations, a data source's reputation will be used as a factor in deciding how much trust we can place on that data. Given the multitude of sources, there is a need for a "just in time" data trust mechanism that can vet data sources in real time.

This has crystallised the view that the fundamental challenges to increasing trust in data are 1) a hesitancy to push business logic and IP outside the firewall into edge servers, 2) that insights are risk-filled without measurable trust in data, and 3) the data and the code must be trustworthy—making the insights more valuable and monetizable.

Dell Technologies and Intel have been here before—providing trusted data delivery and compute within the walls of enterprise data centers via workloads running on Intel®-based Dell servers, stored as implicitly trusted data in Dell storage systems.

But now the two companies are taking this experience and working together to deliver the same trusted environment in edge computing by studying the trust problems inherent in edge-based ecosystems.

Why Distributed Trust Is Such a Challenge

Why is trustworthiness at the edge such a challenge? In a data center, it is very simple to draw a straight line between enterprise storage system data and applications consuming that data. This is shown on the left side of Figure 1, which highlights the traditional strength of Intel and Dell Technologies in the data center. From the perspective of the data owner, this framework builds trust based on 1) a well-defined and managed perimeter around data processing, 2) data is encrypted at the source before being shipped, 3) the data owner and consumer have a pre-existing direct business relationship, and 4) managed perimeter security to address compliance concerns.

Also, in data center environments, the delivery of trusted data is more securely managed and tuned by a dedicated IT team. At the network edge, however, bringing trusted data and applications together is beyond the scope and capabilities of a data center security team. Also, for many emerging edge application scenarios 1) the data owner and consumer may not have a direct pre-existing business relationship, and 2) processing occurs on servers with weak or unknown perimeter security, resulting in a potential compliance exposure.

The right-hand side of Figure 1 illustrates the complex mesh of east-west permutations as decentralized data and applications are brought together. The intermediate nodes may represent physical nodes or virtual compute instances, each under the control of different parties. Because these nodes are operating on the data (i.e., they are not serving as simple data routers), there must be an assurance that the data is trustworthy and is not misused in any way.

A related problem involves the distribution of these applications onto edge processing platforms. The applications must be delivered securely without revealing critical IP. And the trustworthiness of the application itself must be measured. The potential combinations of decentralized data sources found on the edge is staggering. Any application that wishes to aggregate and analyze data from these sources faces a challenge determining where the data originated, who owns it, where it has been, and what policies are associated with it. The data may have been vetted a *priori,* but this vetting may have been lost.

Data insight consumers must be able to trust that their application is being executed per their expectations because they are no longer in control of their deployment environment. This lack of control is further compounded by application logic being distributed/replicated or distributed/partitioned on edge nodes that are under the control of different entities.

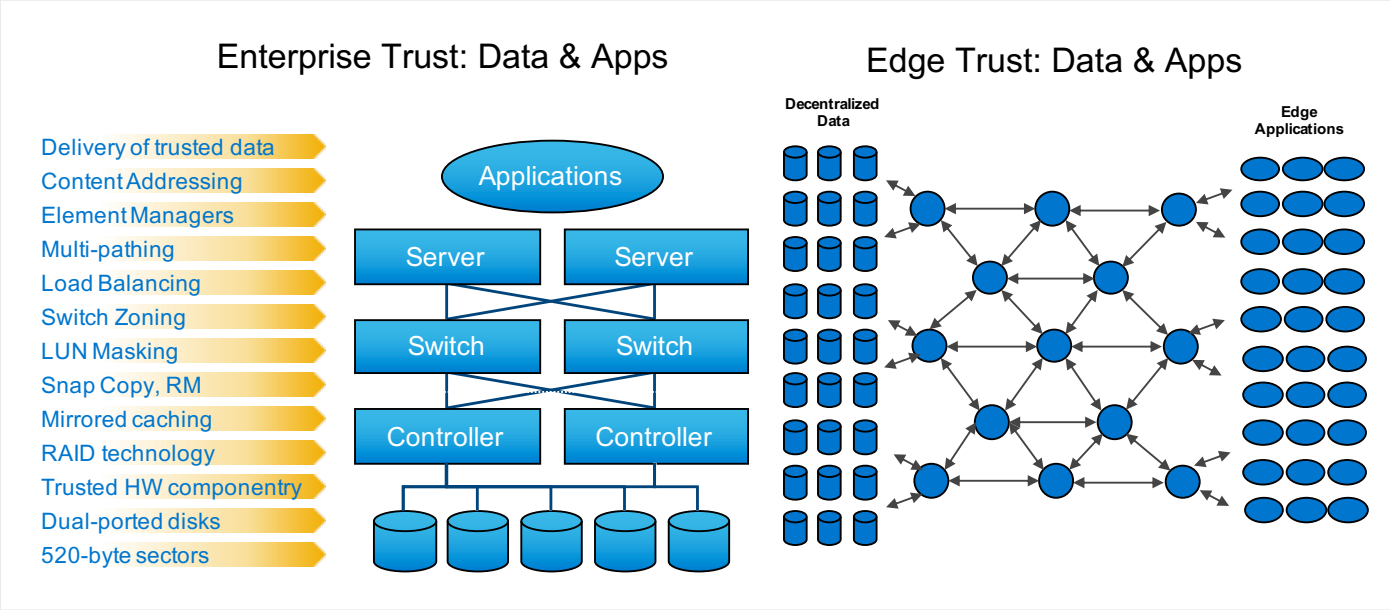

Figure 1. Enterprise data center trust vs. edge networks trust.

Finally, data owners must trust that the application is doing what it is supposed to do. Data can be shared among numerous applications and it is not scalable or acceptable to expect data owners to formally vet the sources of all the applications along the processing path.

Consider a simplified cross-section of the edge environment depicted in Figure 2. The diagram shows data flowing from a sensor, to a gateway, to an edge server, before arriving at a cloud application. During this flow, each node performs some sort of processing operation before the data arrives at the application.

Note that it is not feasible to ship all data to the cloud; the data volume makes it uneconomical to transport all of that data. The application needing the data may live at any point along the data's path.

No matter where the application lives, problems can occur as the gateway, edge server, and cloud perform pipeline processing operations on the data. The list below highlights the different problems that can arise:

- What is the data's provenance (originating device)?
- Who is the owner of the data?
- Who is the authority attesting to the data's provenance and ownership?
- What was the security profile at the point of ingest (e.g., on the gateway)?
- Did the ingesting device previously go through a secure onboarding process?
- Have any of the operating environments along the path been tampered with?
- Is the lineage of the data (the path travelled) available for inspection throughout its journey?
- Who is attempting to transform or look at the data along its journey? Are they authorized?
- Who is attempting to store the data, for how long, and why?
- How are applications securely installed along the path?
- How do these applications execute in a "roomed" (e.g., isolated) environment?
- How are analytic results "tokenized" into a monetizable result?
- How can payment for analytic services along this path also be monetized?
- How can the data journey and analytic processing be audited and governed?
- How can frictionless edge configuration processes enable all of the above?
- How does data privacy, competitive exposure, and regulatory/geo restrictions impact sharing data?

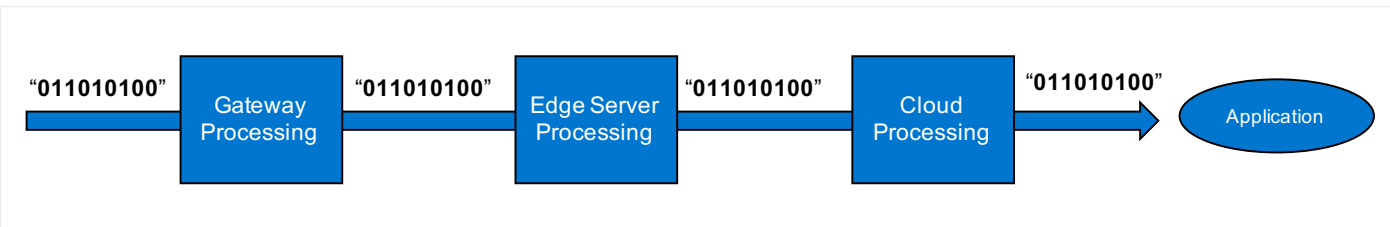

Figure 2. Raw data arrives with unknown history and processing transparency

What is needed is a mechanism that can establish trust and confidence in real time, programmatically, meaning that data, and the resulting insights, can be monetized in a fluid manner.

Given this objective, the fundamental assumptions needed for a solution are 1) the hesitancy to push business logic outside firewall, 2) that insights are problematic—risk-filled and without measurable trust in the raw data, 3) that data and code cannot be explicitly assured as trustworthy, which makes the insights more valuable and monetizable, 4) that new liability exposures and opportunities are constantly coming into being which allow for data misuse.

Four Principles of Establishing Data Trust

Given this computing environment and the assumptions needed for a solution, there are four basic principles for establishing and measuring data trust—at scale and in real time—at the edge:

Attest: Establish the provenance of data from a given network, storage, compute, or data element using publicly auditable and interoperable protocols.

Transform: Execute an operation while ensuring that the only operations allowed on a given data or inferencing element comply with all locally applicable security and compliance statutes, using publicly auditable and interoperable protocols. This shall include full separation of data from the algorithms and underlying compute infrastructure that is being used to process that data, in a way that ensures all participating parties have full assurance that their IP, ownership, and privacy rights are strictly enforced.

Annotate: Create a permissioned log chain consisting of data plus metadata and references to a permissioned list of the operations allowed for, and performed on, that data. Examples of allowed operations may include model training, analytics, anonymizing, encrypting, compressing, or other operations that operate upon the data.

Audit: Implement a standardized, open, permissioned, and distributed measurement fabric to ensure compliance with the above principles.

By focusing on these principles and applying them to edge ecosystems, it is possible to create a solution that will deliver on the following minimum requirements:

- Analytic processing requires auditable trust in data and code provenance
- Establishing trust requires both data and code
- Trust establishment needs local, customizable governance to setup and operate new entrants, execution modes, roadmap, and revenue distributions
- Trust establishment must work when parties have no prior relationship
- Trust establishment needs to be automated and fleeting
- Trust establishment needs to deliver an auditable supply chain of inferencing and provenance
- Trust establishment does not need a single fabric; however, solving the problem once in an open forum and scaling is far more secure, efficient, and valuable

The effort is worth it: measuring trust makes the insights more valuable and monetizable—the value of a "single source of truth" coming from attested platform/code integrity. It is possible to deliver on the above principles and requirements using a combination of hardware and software library elements. Specifically, we propose that a privacy-preserving edge compute framework combined with a Data Confidence Fabric (DCF) makes it possible to address these requirements. A DCF solution delivers open and auditable metrics to generate and measure trust, producing monetizable insights at scale, in real-time, in a way that unlocks data potential and delivers high quality, auditable decision making.

Implementing Distributed Trust

Figure 3 highlights the architecture of a DCF solution section and outlines the key architectural components to meet the attest, transform, annotate, and audit principals for a data stream:

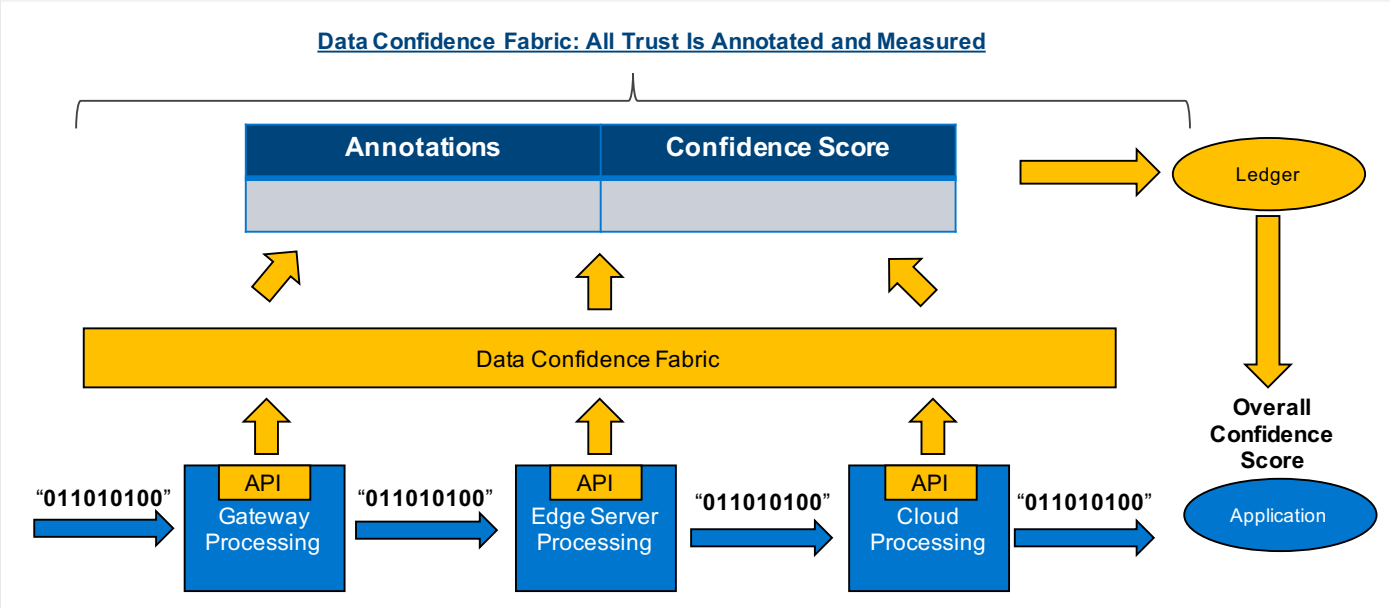

Figure 3. Overview of Data Confidence Fabric.

Any edge node that participates in a DCF calls an industry-standard API that supports annotation (what type of trust operation is being performed during transit) and scoring (how much confidence is being added). As each trust operation is annotated and scored, the DCF collects this additional metadata that travels along with the packet through the data path similar to a "motorcycle sidecar." Ultimately, all annotations and scores are stored in a distributed ledger, where they are designed to be permanently, securely, and immutably associated with the corresponding data stream.

Applications that have edge-based access to the distributed ledger can now analyze the trustworthiness of the data stream by inspecting the annotations and the scores. Consider a real-world example in which Dell and Intel hardware and software assets are deployed onto the edge and configured to function as part of a DCF.

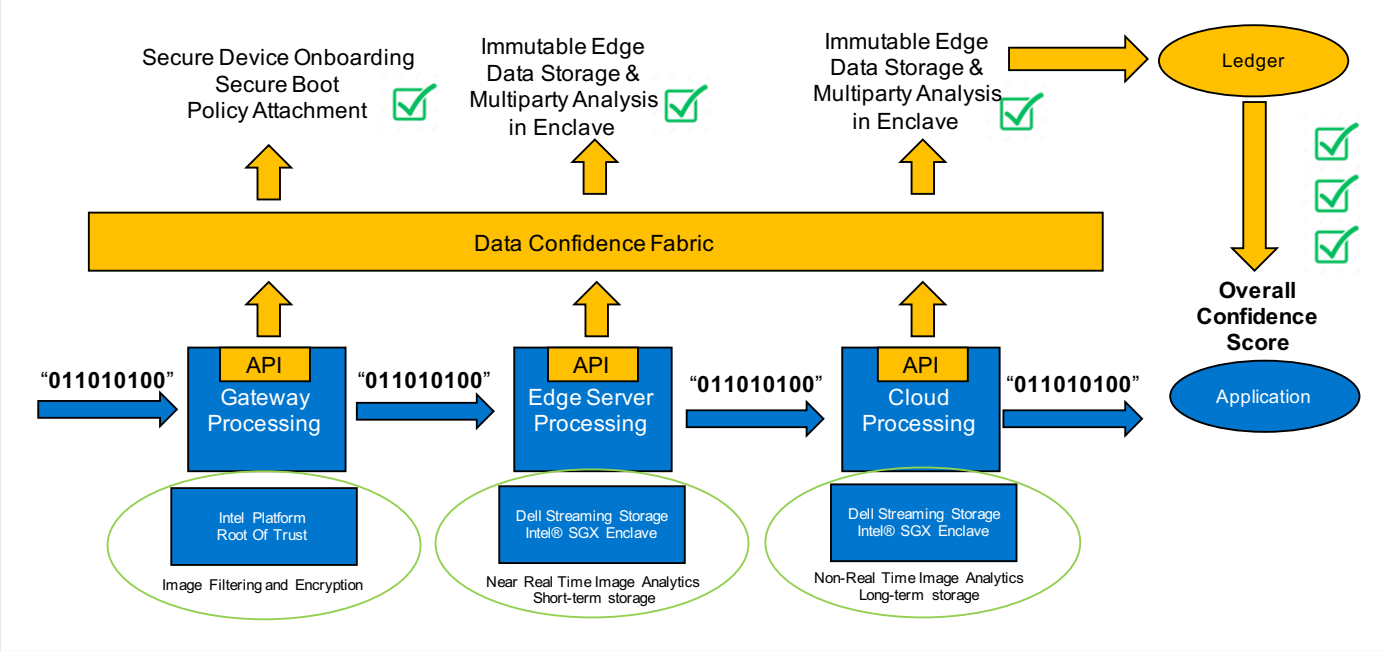

Figure 4. Cross section of DCF implementation enabling for privacy-preserving computer vision for smart cities

Privacy-Preserving Computer Vision for Smart Cities

The example in Figure 4 shows how data can be better protected as it is transformed or modified or used as it moves from the gateway to the cloud via an edge server. As a use case, consider how a local government agency might use images and video captured by cameras deployed in a city for public safety, traffic management, and other uses.

Depending upon the usage scenario, the security camera system does object recognition for vehicle detection or license plate detection, or facial recognition of drivers, passengers, or pedestrians. In an example deployment, cameras are placed on traffic light poles at various intersections. These cameras are connected to gateways located in roadside units. The gateways are connected to a service provider's edge servers hosted by an edge cloud provider. Finally, the edge servers are connected to cloud servers hosted by public cloud provider.

The cameras capture the images and transfer them to the gateways. Depending upon the usage scenario, artifacts in the image (e.g., faces or license plates) that should not be disclosed are blurred by the gateways. The gateways then encrypt the updated images along with location of the camera and date and time of capture and transmit to the edge servers.

The edge server decrypts the received images, and depending upon the usage scenario, does vehicle detection or license plate detection or facial recognition of the driver, passenger, or pedestrian of interest. Finally, depending upon the usage scenario, the resulting images of interest and associated details may be re-encrypted and sent to cloud servers for long-term storage. Such a service can be used during accident claim settlement or a stolen vehicle search while sharing and storing only the needed details with authorized servers.

To deploy this service, a hardware security module (HSM) is deployed into the cameras while gateways and edge servers utilize Intel® Software Guard Extensions (Intel® SGX). Trusted workers are launched in gateways and edge servers as containers enhanced with Intel SGX to help protect the un-encrypted data processing and the logic that processes that un-encrypted data.

The trusted worker in cameras uses the HSM to encrypt the camera capture. The gateway along with location, date, and time of capture writes a hash of captured image and relevant root of trust (RoT) platform data to the DCF.

Depending upon the usage, rules can be written to have the gateway automatically blur artifacts that are not needed for the usage scenario and should not be disclosed. The trusted worker on the edge server uses an Intel SGX enclave to decrypt the data received from a gateway for inferencing as per the usage scenario. Intel SGX is used to help protect the unencrypted data and AI model IP. Upon successful detection, it also writes a data hash to log its operation on the DCF.

Confidential compute, for example, Intel SGX and distributed ledger-based DCFs together deliver trust at the network edge and a fluid data economy that lowers risks from data theft, data misuse, and applications IP theft.

Technology Used in This Case Study

The Dell and Intel technologies used to create the DCF in this example include the following:

PowerEdge R640 Rack Server: Dell's flagship server offers scalable computing and storage in a 1U, 2-socket platform with an ideal mix of performance, cost, and density for most data centers. Available in single processor or dual-processor configurations.

PowerEdge XE2420 Edge: A Dell server designed for low-latency high performance edge in flexible configurations. It is designed for demanding retail and analytics applications at the edge.

Dell EMC ECS EX 300 Object Storage: The ECS can grow from 60 GB to exabyte-scale. The ECS is the leading object-storage platform from Dell EMC and has been engineered to support both traditional and next-generation workloads. Deployable in a software-defined model or as a turnkey appliance, ECS boasts scalability, manageability, resilience, and economics.

Dell EMC Streaming Data Platform: This data center-grade software platform empowers organizations to harness their real-time and historical data in a single, auto-scaling infrastructure and programming model. Using the Streaming Data Platform, organizations can achieve innovation throughout their entire ecosystem through use of their unstructured data.

2nd generation Intel® Xeon® Scalable processors: Both Dell servers are equipped with these CPUs that provide the foundation for a powerful data center and network edge platforms delivering both agility and scalability. This innovative processor platform converges capabilities across compute, storage, memory, network, and security. The Intel Xeon Scalable platform is designed for data center modernization to drive operational efficiencies that lead to improved total cost of ownership (TCO) and higher productivity for users. For DCF, these servers offer Intel® Key Protection Technology (Intel® KPT) to help secure keys in hardware and enable a high throughput hardware security module (HSM) on the server.

Root of Trust: Cryptographic keys are more protected by hardware with an ID that is inseparable from the hardware.

Secure Boot: Designed to ensure that only authorized firmware images are installed and run on a hardware platform to mitigate the risk of attacks targeting low-level, highly privileged platform components.

Intel® Software Guard Extensions (Intel® SGX): Hardware based security that includes new instructions to increase the security of application code and data from disclosure or modification. Developers can place sensitive applications and data into enclaves, which are areas of protected memory, providing increased security protection while data is being executed on.

Summary

In this paper we have endeavoured to show that while AI insight tools are critical to the future of society and a prudent use of scarce resources, such tools require independent, permissioned audit and oversight in order to allow industries to safely expand their adoption of these emerging technologies.

Intel and Dell Technologies have created the DCF framework that fulfills the four principles of data trust—attest, transform, annotate, and audit. Through this framework, CoSPs will be able to more securely expand their AI-based services or even create new services such as a distributed marketplace, where all parties to a transaction are better protected and real-time value discovery becomes a reality.

Next Steps

Intel and Dell Technologies invite CoSPs, CSPs, ISVs, and other ecosystem partners to work with us on the next phase of this emerging framework—especially those who are eager to move quickly in the edge services space and wish to do so with the benefits of a measured trust framework. To participate contact Paul O'Neill at paul.oneill@intel.com.

About the Authors

Steve Todd
Steve Todd is the Vice President of Data Innovation and Strategy in the Dell Technologies Office of the CTO. He is a long-time inventor in high tech, having filed over 300 patent applications with the USPTO, and his innovations and inventions have generated tens of billions of dollars in global customer purchases. Steve actively researches the value of data and is currently the Head of the Data Office focused on extracting new forms of value from internal data assets. Most recently, Steve co-founded and launched Project Alvarium, an open-source research platform for valuing trustworthy data in a Data Confidence Fabric (DCF). Steve is a Dell Technologies' Fellow with bachelor's and master's degrees in Computer Science from the University of New Hampshire.

Paul O'Neill

Paul is Director of Strategic Business Development in Intel's Confidential Computing Group. He is responsible for eco-system development and coordination of Confidential Computing solutions like Intel SGX and strategies across market segments from Enterprise, Government, Cloud, Healthcare and IoT with strategic partners. Paul has been working at Intel since 2014 and has 20+ years' experience in technology companies from start-ups to large enterprises.

Ben McCahill

Ben works in the strategy and planning teams for network edge platforms at Intel. His areas of expertise are in fixed and mobile technologies, internet ecosystems and cloud solutions. Ben has over 25 years industry experience within multiple startups, system vendors, and telco operations. Ben has held roles in engineering, sales, and senior management. He has a Bachelor's in English (with Honors) from University College in Dublin and a Master's in Science from Heriot Watt University in Edinburgh.

Sanjay Bakshi

Sanjay Bakshi is an Intel Principal Engineer and a multi-discipline technologist. He has 15+ years' experience leading teams across multiple geographies through ideation, prototyping, and development of technologies that have been shipped broadly by Intel. He has delivered several successful projects showcasing Intel silicon in areas of distributed confidential computing, blockchain, secure identity and biometrics, USB power management, WIMAX security, IP MPLS, and IP DiffServ QoS. Most recently, Sanjay is leading the Intel initiative round Privacy Preserving Edge Cloud and serves on the board of Enterprise Ethereum Alliance. He has over 50 issued patents and a Bachelor's degree in Computer Science from NIIT India.

More Information

Power Edge XE2420:
https://www.dell.com/en-us/work/shop/povw/poweredge-xe2420/techspecs

Power Edge R640: https://www.dell.com/en-us/work/shop/povw/poweredge-r640

Dell EMC ECS EX300: https://www.delltechnologies.com/en-us/storage/ecs/index.htm#accordion0

Dell EMC Streaming Data Platform: https://www.delltechnologies.com/en-us/storage/streaming-data-platform.htm

Intel® Software Guard Extensions: https://software.intel.com/sgx

Intel® Key Protection Technology: https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/key-protection-technology-paper.pdf

Intel® Xeon® Processors: https://www.intel.com/xeon