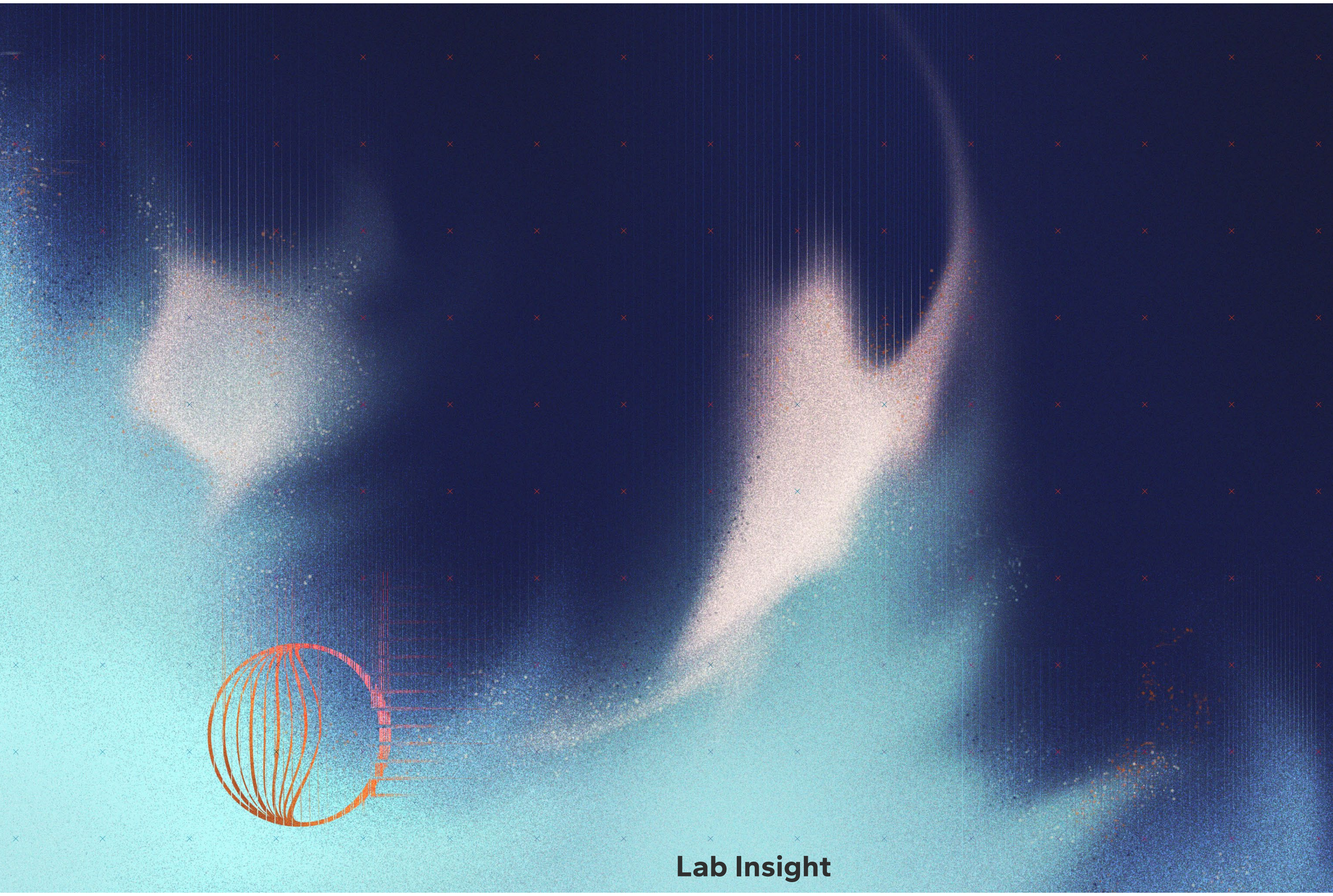
signal65



Lab Insight

Dell Enables an Industrial Digital Twin Proof of Concept with Artificial Intelligence Technology

AUTHOR

Mitch Lewis
Research Analyst | The Futurum Group

MARCH 2024

Introduction

As part of Dell's ongoing efforts to help make industry-leading AI workflows available to its clients, this paper outlines a sample AI solution for manufacturing environments. The proof-of-concept solution presented in this paper demonstrates an AI powered application for detecting hazardous events and equipment failures in a factory setting. The reference solution demonstrates how such a solution can be brought to life by using a digital twin Metaverse, autonomous mobile robots (AMRs), and readily available hardware from Dell™ and Broadcom™.

Recent developments in AI technology have brought significant interest to the potential of AI powered solutions across a broad range of industries. Much of the recent focus on AI is related to generative AI applications and large language models, however, the field of AI encompasses many different techniques and capabilities that organizations can leverage to build industry specific solutions. While new AI solutions can offer significant value, in many cases organizations are unfamiliar with the requirements of deployment. Manufacturing environments, in particular, present a unique challenge for developing and applying AI applications and technologies, presenting an opportunity to leverage additional emerging technologies.

The reference solution outlined in this paper leverages Dell PowerEdge servers, 3D graphics technologies, AMR robots, and AI to enable the detection of chemical spills in the real world by training on synthetic data in the Metaverse. It additionally uses factory sensor data and machine learning techniques to predict equipment failures. The proof-of-concept demonstrates the ability to quickly train and deploy a manufacturing focused AI solution by using readily available hardware from Dell with NVIDIA™ L40S GPUs and Broadcom Ethernet adapters. The solution can be further expanded upon or customized to meet the individual requirements of manufacturing organizations.

Additional information including configuration details, performance test results, and source code have been made available in a GitHub repository by Dell.

Importance for the Manufacturing Market

AI technology holds significant potential to impact the manufacturing industry, and the use of AI in manufacturing is growing, with an estimated growth of 25% in 2024 according to Futurum Intelligence[1]. Globally, there are millions of factories running 24/7, often with highly complex processes. By leveraging AI technology, organizations can further optimize and monitor these processes, resulting in safer factory environments, less downtime, and greater quality control. Possible solutions include AI-powered quality control and defect detection, predictive maintenance for machinery, or intelligent production demand insights.

These AI driven solutions require input data from a variety of sources, such as video streams or machine sensor data. Due to the complex and dynamic layout of many factory environments, video streams provided from stationary cameras may be obstructed from certain views and may not be capable of capturing the full environment. In such cases, other technologies such as AMRs may be useful in dynamically scanning the factory environment. To leverage machine sensor data, AI solutions must integrate with standard machine communication protocols used in manufacturing environments.

For certain AI models which may seek to capture anomalous conditions in manufacturing environments, creating adequate training data may be challenging. In situations such as detecting fires, broken machinery, or unsafe conditions, it is impractical to create these conditions on a live factory floor. Meanwhile, the desired solution needs to be trained and tested on the specific environment that it will be deployed in to ensure accuracy. This presents an ideal use case for Metaverse technology.

These challenges, along with general unfamiliarity with AI technology, may lead to hesitancy in deploying AI solutions amongst manufacturing organizations, and an overall loss in the value they provide. Despite challenging conditions, practical AI solutions for manufacturing environments are achievable when leveraging the correct technologies. The following PoC solution demonstrates one such strategy for a readily deployable AI-powered manufacturing solution.

[1] https://futurumgroup.com/intelligence/artificial-intelligence-data-analytics-iq-dashboard/

Solution Overview

To demonstrate an AI-powered application focused on manufacturing, Scalers AI™, in partnership with Dell, Broadcom, and Signal65 implemented a proof-of-concept for detecting hazards in a factory setting as well as providing predictive maintenance for factory equipment.

The hazard detection capabilities achieved in this proof-of-concept solution demonstrate significant potential value for manufacturing organizations. Factories may be subject to an array of hazardous events that pose significant risk to both factory operations and worker safety. By utilizing AMRs and AI image detection, hazards such as fires, broken pipes, gas leaks, and chemical spills can be efficiently detected without exposing workers to such hazardous conditions. This PoC solution specifically targets detection of chemical spills, however, utilizing the same fundamental approach, the solution could be expanded to detect additional hazardous events.

Machine learning powered equipment fault detection capabilities are also demonstrated in the PoC, showcasing the potential for AI to detect machinery issues, enable predictive maintenance, and prevent costly downtime. The PoC specifically focused on detecting bearing faults in industrial compressors.

Solution Highlights

- Chemical factory digital twin leveraged to train and test an AI powered solution
- Streaming video from AMR robots and AI image detection leveraged to detect hazardous chemical spills
- Factory equipment sensor data utilized alongside machine learning to achieve predictive maintenance capabilities
- Scalable solution with separate inference and visualization services.

An overview of the PoC solution is as follows:

- A factory floor digital twin is created with NVIDIA Isaac Sim, part of NVIDIA Omniverse. This PoC simulated a chemical factory used in the oil and gas industry. Chemical spills are placed throughout the simulated factory floor.
- Digital AMR robots are deployed within the factory floor Metaverse. The AMRs are assigned to designated zones within the factory, with waypoints set to ensure total coverage of the environment.

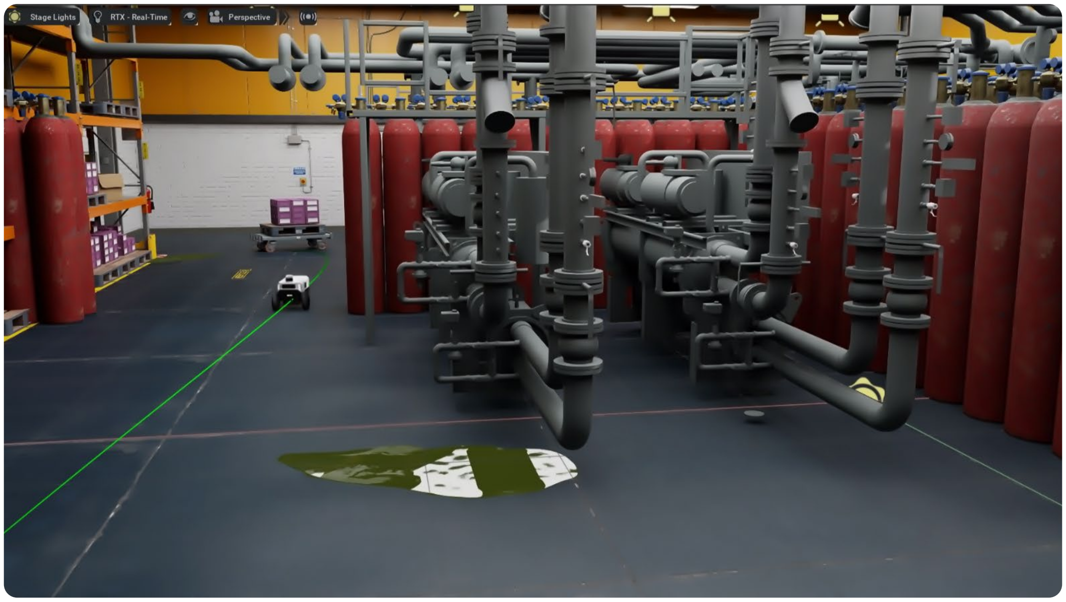


Figure 1: *View of Metaverse (Source: Scalers AI)*

- As AMRs navigate the factory floor, video is streamed to an AI pipeline deployed on a Dell PowerEdge server. Video streams are decoded and inferenced with an AI image detection model trained to detect chemical spills. An image of an AMR approaching a chemical spill within the metaverse can be seen in Figure 1.
- Sensor data from simulated compressors deployed in the Metaverse is collected at regular intervals and published over OPC Unified Architecture (OPC UA). A machine learning powered analytics module analyzes the data to detect bearing faults.
- A visualization dashboard provides monitoring and alerting of both chemical spill detection and compressor failure. The dashboard includes direct views from AMR cameras, a map of the factory floor with live AMR locations, chemical spill incident logs, and a time series graph of compressor failures. Alerts of hazards are displayed on the dashboard and can additionally be sent as mobile notifications. The visualization dashboard can be seen in Figure 2.

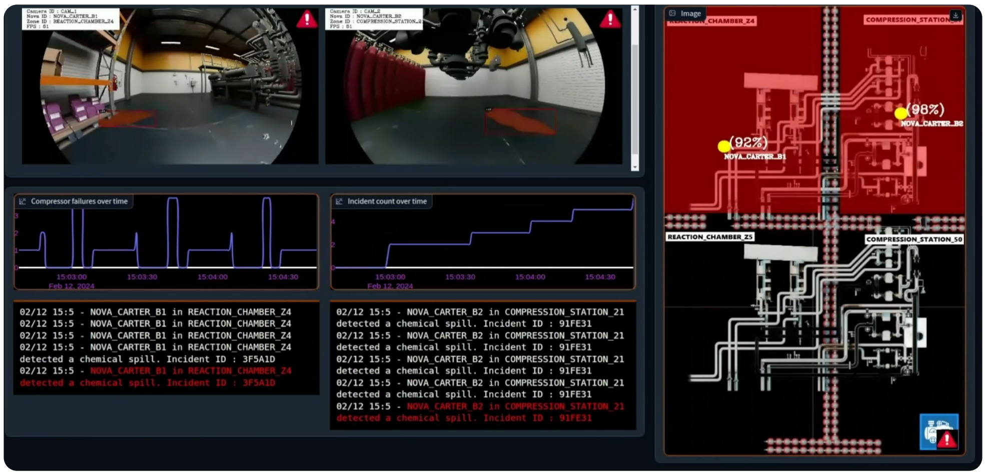


Figure 2: *Visualization Dashboard (Source: Scalers AI)*

A high-level overview of the PoC solution can be found in Figure 3.

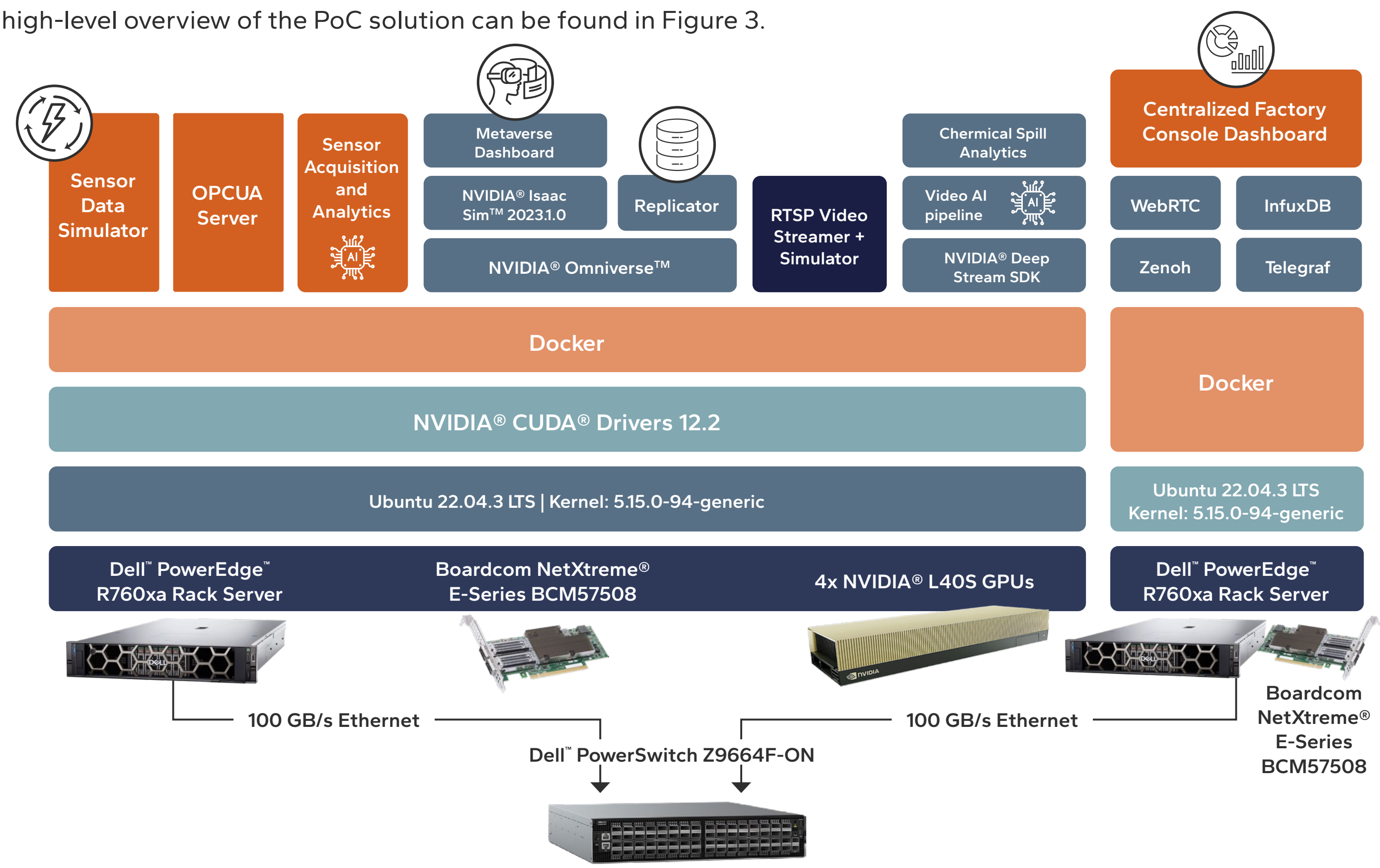
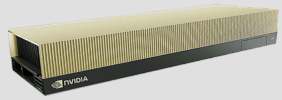


Figure 3: *Solution Overview (Source: Scalers AI)*

The key hardware components used in the solution included the following:

Dell PowerEdge XR7620 Server



NVIDIA® L40S GPUs

Broadcom NetXtreme E-Series BCM57508 Ethernet Controllers

Dell PowerSwitch Z9664F-ON

Highlights for AI Practitioners

This PoC demonstrates a manufacturing focused AI application that can serve as a reference solution for AI practitioners in the industry. In particular, the use of Metaverse technology to enhance AI workflows should be noted. In addition, the solution demonstrates tools and protocols that can be leveraged to build AI solutions with data sources common to manufacturing environments, such as video streams and machine sensors.

The use of Metaverse digital twins enables development of AI solutions that can be translated to specific, real-world environments, such as factory floors, without the cost or complexity of utilizing a real environment or a physical replica. It should be noted that the digital twin environment is used both to test the solution, as well as to create training data. By capturing images of chemical spills placed throughout the Metaverse factory, a training dataset can be created quickly and safely while remaining accurate to the specific factory environment. This training dataset is then used to fine tune the pre-trained YOLOV8s image detection model, quickly achieving an AI model capable of detecting chemical spills. The Metaverse environment is then used to test the solution, which can be further deployed to the digital twin's physical counterpart for a real-world solution.

AI practitioners should also note the PoC's utilization of NVIDIA DeepStream to accelerate development of a video-based AI pipeline. NIVIDIA DeepStream is a software development kit targeted at computer vision and streaming video analytics applications. DeepStream is made available for free by NVIDIA, and is a powerful tool for developing a wide range of image, video, or sensor data driven AI applications on NVIDIA hardware. This PoC provides an example of DeepStream's capabilities, which can be further utilized by AI practitioners in expanding the PoC or developing similar AI applications that leverage streaming video data.

Key Highlights for AI Practitioners

- AI implementation enabled by leveraging NVIDIA Omniverse for both generating a training dataset and testing the solution. A final solution is transferable to a physical factory environment.
- NVIDIA DeepStream used to accelerate development of video driven AI application.
- Machine learning leveraged sensor data published over common OPC UA protocol.

Along with utilizing video data to detect hazardous chemical spills, the PoC additionally enables AI powered predictive maintenance by detecting bearing faults in industrial compressors. To enable this, the PoC captures sensor data from simulated compressors and analyzes it with a machine learning module. Key to this process is publishing the sensor data over OPC UA, which is an open-source, standard method for exchanging machine-to-machine data, commonly utilized in manufacturing environments. Incorporating standard machine-to-machine protocols such as OPC UA is crucial for AI developers in manufacturing environments where AI solutions are commonly driven by machine generated data.

Considerations for IT Operations

A key challenge for IT operations in supporting new AI applications is an understanding of hardware requirements, as AI encompasses a broad range of solutions with varying levels of computational performance and data requirements. This PoC demonstrates a particularly demanding AI application that must support a combination of video processing, AI inferencing, and Metaverse simulation.

The factory hazard detection PoC was deployed using Dell PowerEdge R760xa servers with NVIDIA L40S GPUs. The L40S is an ideal choice for this deployment as it is a powerful, general-purpose GPU that is well fit for both the video and simulation components of the solution, as well as the AI training and inferencing. The L40S is a modified version of NVIDIA's L40 GPU which is designed for data center graphics and virtualization. The L40S shares the same Ada Lovelace architecture as the L40, but improves upon it with additional optimizations for AI. This provides a GPU that is well suited for solutions requiring AI computations and Metaverse graphics.

Dell PowerEdge R760xa servers are additionally well suited for this deployment due to their high compute density and flexible GPU support. The PoC uses two distinct PowerEdge R760xa server configurations, one with four GPUs to handle the intensive metaverse and AI workloads, and a second without GPUs to support the visualization dashboard. The R760xa servers deployed in this PoC demonstrate the flexibility to support both services, including the high GPU capacity required for compute intensive graphics and AI workloads. Specifications of the Dell PowerEdge R760xa servers deployed in this PoC can be found in Figure 4. Full configuration details are available in the appendix.

Service	Server	CPU	RAM	Disk	GPUs
Metaverse and AI Pipeline	Dell PowerEdge R760xa	Intel® Xeon® Plavtinum 8592+	v500 GB	500 GB	4xNVIDIA® L40S GPUs
Visualization Dashboard	Dell PowerEdge R760xa	Intel® Xeon® Platinum 8470Q	500 GB	1 TB	

Figure 4: *Dell PowerEdge R760xa Configuration Overview*

Key Highlights for IT Operations

- Dell PowerEdge R760xa servers with NVIDIA L40s GPUs provide flexibility for video processing, Metaverse simulation, and AI inferencing
- Modular architecture and high bandwidth Ethernet provide scalability and ease integration with common network infrastructure of existing production environments

The separation of distinct services should also be noted by IT operations for the scalability it provides. By separating the Metaverse and AI pipeline service from the visualization dashboard, either component can be scaled individually as needed. This architecture provides flexibility for IT operations; however, it requires a high bandwidth connection to ensure performance. To maintain performance, the two services were connected utilizing scalable, high bandwidth Ethernet from Broadcom.

Solution Performance Observations

This PoC involves several computationally demanding processes combined into a single solution, including deploying a Metaverse, decoding video streams, inferencing, and publishing video streams to a visualization dashboard. To test the overall performance and scalability of the solution, the PoC deployment was tested with an increasing number of video streams.

Each test was conducted for 5 minutes, with the number of concurrent video streams increasing from 1 to 24. Figures 5 and 6 display the throughput and bandwidth results of the performance testing. Full results are available in the appendix.

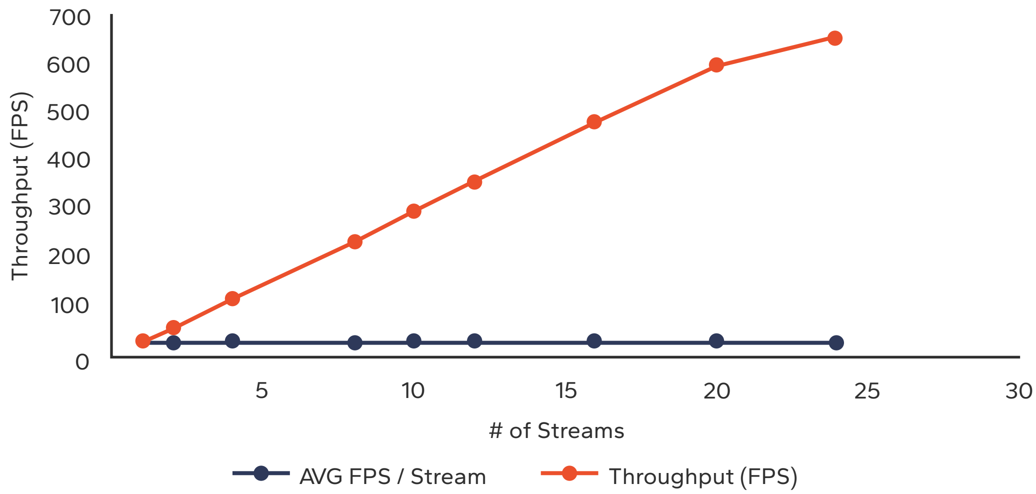


Figure 5: *Throughput Measurements*

The performance testing demonstrated that the solution is capable of scaling throughput as the number of video streams increase. The measurement of FPS per stream in this scenario includes video decoding, inferencing, post processing and publishing of uncompressed streams over a network. The system achieved a total throughput of 29.98 FPS with 1 stream, up to 657.8 FPS with 24 concurrent streams. The average FPS per stream remained steady around 30 FPS, which is considered an industry standard for video workloads. This average was maintained by scaling the number of GPUs used for inferencing as needed. One of the four NVIDIA L40S GPUs was utilized by the Isaac Sim Metaverse, while the remaining three were available for the inferencing pipeline. Initial tests utilized only one of the available NVIDIA L40S GPUs for inferencing, with a second one used beginning at 8 concurrent streams, and the third utilized beginning at 18 concurrent streams. This demonstrates both the power of an individual NVIDIA L40S GPU to support smaller deployments, as well as how the solution can be scaled with additional GPUs.

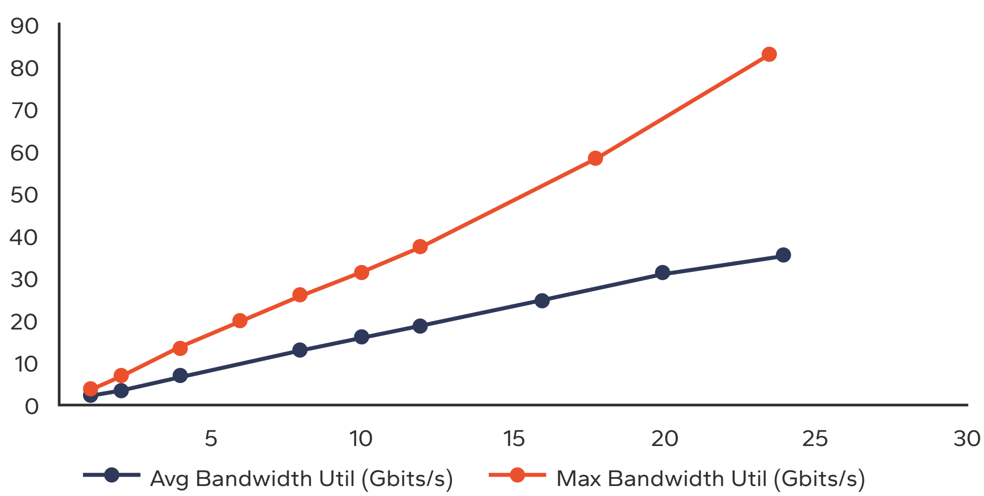


Figure 6: *Bandwidth Utilization*

Testing the solution with an increasing number of video streams additionally shows the solution's ability to scale the connection between the AI pipeline and the visualization dashboard. As the number of concurrent video streams increased, so did both the average bandwidth utilization and the maximum bandwidth utilization. The average bandwidth utilization scaled from 1.56 Gbits/s with one stream, to 34.3 Gbits/s with 24 streams. Meanwhile, the maximum bandwidth utilization scaled from 3.14 Gbits/s to 83.9 Gbits/s. The increase in bandwidth utilization highlights the need to utilize scalable, high bandwidth Ethernet in this PoC to maintain connection between the AI pipeline and the visualization dashboard.

The performance of this solution is crucial, as rapid detection and alerting of hazards within a factory setting can mitigate damage and minimize factory downtime. To achieve this, the solution needs to be powerful enough to inference multiple incoming video streams and efficiently publish data to the visualization service for monitoring and alerting. Support for concurrent video streams is vital as it signifies the solution's ability to support video input from multiple AMR robots. The number of AMRs required will vary between organizations depending on the size of the factory being monitored and the urgency of detecting hazards.

Training and testing the solution involves additional performance requirements to support the Metaverse simulation. The performance results validate the design of the PoC solution and demonstrate its ability to be further scaled to unique manufacturing demands. By utilizing Dell PowerEdge servers and NVIDIA GPUs the PoC is capable of providing the performance necessary to maintain industry standard throughput for a complex AI workload, while simultaneously supporting the Isaac Sim Metaverse application. Additionally, by utilizing Broadcom NetXtreme E-Series BCM57508 Ethernet adapters, the PoC solution is able to handle the increased bandwidth demands of additional concurrent streams, maintaining connection to the visualization service for monitoring and alerting of detected hazards.

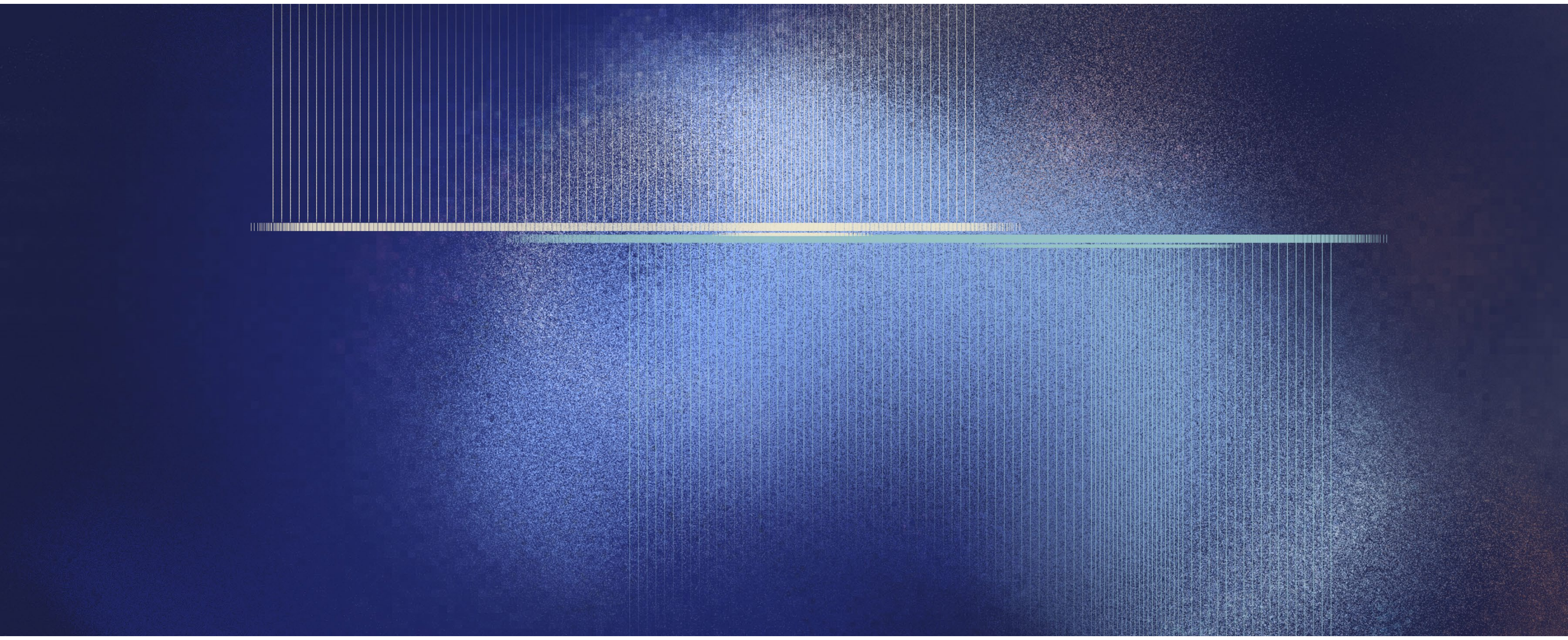
Final Thoughts

AI technology has the potential to provide innovative new solutions to enhance and optimize the manufacturing industry. Manufacturing processes are complex and detailed monitoring of the equipment and environments involved is highly valuable to ensure normal operations and maintain workplace safety. In many cases, this detailed level of monitoring may not be possible for humans to achieve manually. By leveraging AI technology and computer vision, manufacturing organizations can gain new insights, optimize processes, and quickly detect potential issues.

This PoC showcases one such AI powered solution for manufacturing, built using readily available hardware from Dell and Broadcom. The solution addresses a major challenge in building applications for manufacturing environments by utilizing an NVIDIA Omniverse digital twin to replicate a physical environment. The solution can be trained and tested in the Metaverse environment and deployed into a real factory setting to detect hazardous events, minimize factory downtime and damage as well as improve worker safety and compliance. The approach provides a path for organizations to quickly begin leveraging AI in their factory environments where it may otherwise be too costly, timely, or complex to build such an application.

The PoC additionally showcases how machine learning can be utilized alongside factory equipment sensor data to detect failures and enable predictive maintenance. This functionality leverages standard protocols commonly found in manufacturing environments and demonstrates how AI can be utilized to avoid downtime and maintain factory operations.

The solution presented in this paper was designed as a reference for manufacturing organizations to leverage in building their own applications. The PoC solution can be scaled to meet various needs as well as altered to detect different hazardous conditions and machine failures beyond the chemical spill and bearing fault use cases. The strategies and infrastructure utilized in this PoC, including the use of digital twins and AMR robots, serve as a strong foundation for building AI applications for manufacturing. Additionally, the PoC provides insights into how Dell PowerEdge hardware with NVIDIA GPUs and Broadcom Ethernet can be used to create a powerful and scalable AI solution. The findings from this PoC can be further leveraged by organizations to build new innovative solutions to further advance AI adoption in the manufacturing industry.



Appendix

The chart in Figure 7 displays full performance testing results.

GPU SKUs	Number of Streams	AVG FPS / Stream	Throughput (FPS)	Avg Bandwidth Util (Gbits/s)	Max Bandwidth Util (Gbits/s)	Avg System GPU Memory Util (%)	Avg System GPU Util (%)
1x NVIDIA® L40S GPU	1	29.98	29.98	1.56	3.14	4.9	7.67
1x NVIDIA® L40S GPU	2	29.65	59.3	3.08	6.16	5.34	7.53
1x NVIDIA® L40S GPU	4	29.78	119.12	6.17	12.6	6.21	11.6
1x NVIDIA® L40S GPU	6	29.65	177.9	9.25	19.5	7.04	16.6
2x NVIDIA® L40S GPU	8	29	232	12.4	25.5	7.95	16.7
2x NVIDIA® L40S GPU	10	29.7	297	15.5	32.1	8.8	20.2
2x NVIDIA® L40S GPU	12	29.8	357.6	18.5	37.8	9.62	24.8
3x NVIDIA® L40S GPU	18	29.5	531	27.7	58	12.1	33.4
3x NVIDIA® L40S GPU	24	27.4	657.6	34.3	83.9	14.69	41.8

Figure 7: *Performance results*

Figures 8 and 9 provide additional details of the Dell PowerEdge servers used in the PoC deployment. The server in Figure 8 includes NVIDIA GPUs and was utilized for the Metaverse and inferencing services. The server detailed in figure 9 was utilized for the visualization service.

Dell™ PowerEdge™ R760xa		
Device Name		Dell™ PowerEdge™ R760xa
Motherboard		Dell Inc 0CRRF2 X30
CPU	Model Name	Intel® Xeon® PLATINUM 8592+
	Number Of Cores per Socket	64
	Number Of Sockets	2
	Thread(s) per core	1
	Stepping	2
	LLC Cache	L1d: 6 MiB (128 instances) L1i: 4 MiB (128 instances) L2: 256 MiB (128 instances) L3: 640 MiB (2 instances)
Memory	Slot	System Board or Motherboard
	Speed	5600 MT/s
	Size	500 GB
BIOS	Vendor	Dell Inc
	Version	2.1.0
	Release	11/15/2023
	Size	32 MB
Storage	Size	500 GB
OS	Name	Ubuntu 22.04.3 LTS
	Kernel	5.15.0-94-generic
GPU	Product	4 x NVIDIA® L40S
	Vendor	NVIDIA® Corporation
	Memory	48 GB
	Width	64 bits
	Clock	33MHz
	Product	BCM57508 NetXtreme-E 10Gb/25Gb/40Gb/50Gb/100Gb/200Gb Ethernet
	Vendor	Broadcom Inc. and subsidiaries
	Version	11
	Capacity	25Gbit/s
	Width	64 bits
	Clock	33 MHz
	Speed	100 Gbps

Figure 8: *Dell PowerEdge R760xa for Visualization Service*

Dell™ PowerEdge™ R760xa		
Device Name		Dell™ PowerEdge™ R760xa
Motherboard		Dell Inc 0CRRF2 X30
CPU	Model Name	Intel® Xeon® Platinum 8470Q
	Number Of Cores per Socket	52
	Number Of Sockets	2
	Thread(s) per core	2
	Stepping	6
	LLC Cache	L1d: 4.9 MiB (104 instances) L1i: 3.3 MiB (104 instances) L2: 208 MiB (104 instances) L3: 210 MiB (2 instances)
Memory	Slot	System Board or Motherboard
	Speed	4800 MT/s
	Size	500 GB
BIOS	Vendor	Dell Inc
	Version	2.1.0
	Release	11/15/2023
	Size	32 MB
Storage	Size	1 TB
OS	Name	Ubuntu 22.04.3 LTS
	Kernel	5.15.0-94-generic
	Product	BCM57508 NetXtreme-E 10Gb/25Gb/40Gb/50Gb/100Gb/200Gb Ethernet
	Vendor	NVIDIA® Corporation
	Memory	48 GB
	Width	64 bits
	Clock	33MHz
	Product	BCM57508 NetXtreme-E 10Gb/25Gb/40Gb/50Gb/100Gb/200Gb Ethernet
	Vendor	Broadcom Inc. and subsidiaries
	Version	11
	Capacity	25Gbit/s
	Width	64 bits
	Clock	33 MHz
	Speed	100 Gbps

Figure 9: *Dell PowerEdge R760xa for Visualization Service*

Figure 10 details the key software components and versions used in the PoC.

Software	Version
Docker	24.0.7
NVIDIA CUDA	V12.2
NVIDIA® Isaac Sim	2023.1.0
DeepStream	6.4
Zenoh	0.10.1rc0
OPC UA	1.0.6
InfluxDB	1.7.10
Telegraf	1.22.3

Figure 10: *Software Components*

Important Information About this Report

CONTRIBUTORS

Mitch Lewis
Research Analyst | The Futurum Group

PUBLISHER

Daniel Newman
CEO | The Futurum Group

INQUIRIES

Contact us if you would like to discuss this report and Signal65 will respond promptly.

CITATIONS

This paper can be cited by accredited press and analysts, but must be cited in-context, displaying author's name, author's title, and "Signal65." Non-press and non-analysts must receive prior written permission by Signal65 for any citations.

LICENSING

DISCLOSURES

Signal65 provides research, analysis, advising, and consulting to many high-tech companies, including those mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.

signal65

ABOUT SIGNAL65

Signal65 is an independent research, analysis, and advisory firm, focused on digital innovation and market-disrupting technologies and trends. Every day our analysts, researchers, and advisors help business leaders from around the world anticipate tectonic shifts in their industries and leverage disruptive innovation to either gain or maintain a competitive advantage in their markets.

signal65

CONTACT INFORMATION

Signal65 I signal65.com