Building the Intelligent Warehouse with Digital Twins and Agentic AI

White Paper

Contents

Executive summary

Warehouses sit at the center of modern supply chains. Yet many facilities still operate as fixed cost centers that react to problems after they occur. Planners work in spreadsheets and offline tools while real operations change rapidly.

Digital twins and agentic AI change this dynamic. Digital twins create a virtual replica of the warehouse that stays in sync with the physical floor. Agentic AI systems perceive what is happening, reason about options, and trigger actions that align with business goals. Together, they allow the warehouse to function as a live, learning system rather than static infrastructure.

This white paper describes a practical way to build that system. It introduces a six-stage operational flywheel: sense, simulate, reason, optimize, train and deploy, and measure. Data flows into the twin, AI systems test options, optimization engines compute plans and orchestration tools apply them in production. Each loop compares results to business goals and feeds new insight back into the next cycle.

Organizations that adopt this architecture use digital twins to de-risk layout and automation decisions, increase throughput without expanding footprint, and reduce operational firefighting.

Finally, the white paper outlines a phased roadmap that starts with a focused pilot and scales to multi-site deployment.

Industry challenge: Static tools in a dynamic environment

Warehouses operate in an environment that shifts constantly. Order profiles change with promotions, channels, and seasons. Inventory positions move across the network. Labor supply varies by day and shift. In addition, automation introduces new dependencies between robots, conveyors, shuttle systems, and control software.

Despite this volatility, many facilities still run on:

- Static slotting rules that rarely reflect current demand
- Fixed routing logic for workers and robots
- Planning models that update on weekly or monthly cycles

As a result, there is a gap between the plan and the reality on the floor. Leaders and supervisors often detect issues after they've occurred, and the response depends on local experience rather than a system-level view. This reactive pattern has several consequences:

- Longer queues and order cycle times because action starts late
- Idle or overstressed assets because static rules do not balance flows
- Higher labor costs because recovery requires manual interventions
- High risk in layout and automation decisions because teams test only a narrow set of scenarios before committing capital

In recent years, enterprises have begun using NVIDIA Omniverse and OpenUSD to build digital twins of their factories and warehouses. These twins create realistic virtual environments where teams can design and test changes before they touch the physical floor.

This is why enterprises need a clear operating model and shared architecture that links warehouse design, day-to-day decisions, and incident response into a single continuous loop.

Solution overview: The intelligent warehouse flywheel

An intelligent warehouse treats operations as a system that senses, thinks, and acts in alignment with business goals. First, data from warehouse systems, automation, sensors, and cameras flow into a digital twin. Next, agentic AI interprets objectives, examines options, and uses optimization engines to compute concrete plans. Finally, orchestration tools apply those plans to robots, workflows, and systems.

The core behavior follows a six-stage flywheel:

1. **Sense**: The warehouse ingests live data from the warehouse management system, control systems, robots, IoT sensors, and cameras. It builds a current picture of orders, inventory locations, asset status, and queues.
2. **Simulate**: The data drives a digital twin of the facility. The twin captures layout, equipment, control logic, and flows. It can replay historical days and run “what if” scenarios without disturbing live operations. NVIDIA Omniverse and OpenUSD provide the foundation for these industrial digital twins.
3. **Reason**: Agentic AI converts business goals into concrete constraints and targets. Language and vision models understand objectives such as service level commitments, labor limits, and safety rules. They translate these into inputs that optimization engines can use.
4. **Optimize**: NVIDIA cuOpt solves routing, batching, and task assignment problems under those constraints. It focuses on mixed integer programming, linear programming, and vehicle routing problems. It also scales to large and complex warehouses.
5. **Train and deploy**: Teams validate strategies in simulation using NVIDIA Isaac Sim. They test robot behavior, new layouts, and control policies in a virtual copy of the warehouse. Once validated, orchestration services (Isaac Mission Control) apply the changes on the physical floor.
6. **Measure**: The warehouse measures the impact of each loop against operational and financial KPIs. It feeds those measurements back into the “Sense” stage. Models, rules, and playbooks update based on observed results.

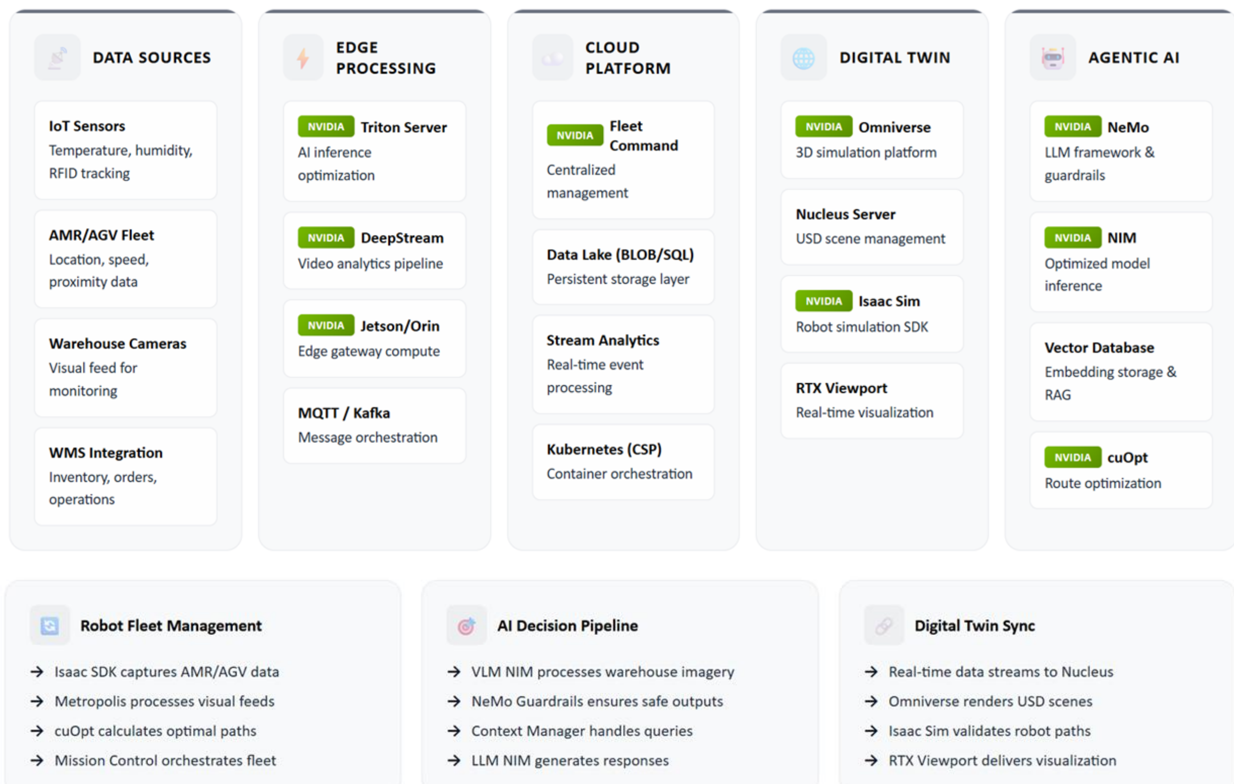
Over time, this flywheel turns design and operations into a continuous cycle. The same system that supports daily decisions also tests medium term changes and informs long-term investments.

Solution deep dive: Technology architecture

An intelligent warehouse needs more than isolated tools. It relies on a layered architecture that connects data, simulation, AI, optimization, robotics, and vision in a single flow, so each layer builds the previous one instead of running a separate project. This structure makes it easier to scale pilots, enforce governance, and reuse capabilities across sites and use cases.

The architecture below describes an end-to-end intelligent warehouse system that integrates real-time edge data from sensors and robots into a cloud-based digital twin powered by NVIDIA Omniverse. It uses an advanced "Agentic AI workflow" featuring vision and large language models (LLMs) to process operational data. This facilitates automated decision-making and gives users actionable insights through a dashboard.

Exhibit: Logical architecture diagram illustrating an intelligent warehouse ecosystem



Architecture layers

Five main layers support the flywheel:

1. **Data sources and integration:** This layer captures operational truth from the warehouse. It ingests signals from IoT sensors, AMR and AGV fleets, warehouse cameras, and WMS integrations. The output is raw sensor data and operational events that reflect orders, inventory state, asset status, and queue conditions.
2. **Edge processing:** This layer performs low-latency ingestion and preprocessing close to the floor. It normalizes telemetry, filters noise, and runs real-time video and sensor pipelines where needed. It packages data for secure, reliable transport to the cloud, using lightweight messaging and streaming patterns.
3. **Cloud platform:** This layer provides centralized storage, streaming, and orchestration. It synchronizes incoming telemetry into a governed data foundation, including a data lake and stream analytics. It also hosts the runtime platform for services and workloads, typically on Kubernetes, to ensure scale, reliability, and lifecycle management.
4. **Digital twin platform:** This layer converts synchronized operational data into a live, shared 3D representation of the facility. It maintains a continuously updated USD scene of the warehouse layout, assets, and flows. It supports physics-based simulation and replay, so teams can test scenarios and evaluate changes without disrupting production operations.
5. **Agentic AI reasoning and optimization:** This layer turns business intent into executable decisions. Agentic AI interprets goals, constraints, and policies, then combines them with current operating conditions. Optimization engines generate concrete plans such as routing, batching, and task assignments. Guardrails apply policy and safety checks before recommendations are released to operations.

The value comes from how these layers interact. The digital twin supports layout design, daily optimization, and incident response. On the other hand, the AI and optimization components support both planning cycles and real time adjustments.

Envisioning Digital Twins with NVIDIA Omniverse and OpenUSD

The digital twin sits at the center of this architecture. NVIDIA Omniverse provides a collaboration and simulation platform, while OpenUSD acts as the scene description standard that represents complex industrial environments.

For a warehouse, the twin can:

- Import CAD layouts, rack designs, and conveyor models into a shared 3D scene
- Represent robots, dock doors, staging areas, and stations as objects with behaviors
- Connect live data streams from warehouse systems and automation to those objects

The twin starts as a design model and evolves into a "digital shadow" as it starts to mirror real operations. Operations teams can view current state, stress test scenarios, and review proposed changes in a common visual environment.

AI reasoning with NVIDIA NIM and guardrails

Agentic AI uses models that understand goals, constraints, and context. NVIDIA NIM provides containerized inference microservices that host these models across cloud, data center, and edge environments.

In warehouse scenarios, NIM can support:

- Language models that interpret business objectives and encode them as optimization targets
- Vision models that classify and contextualize events detected by cameras
- Specialized models that propose routing or batching strategies for specific flows

Guardrail components enforce safety, policy, and compliance rules around AI outputs. Plans that conflict with defined limits, such as safety zones or labor regulations, do not pass through to execution.

Optimization with NVIDIA cuOpt

Optimization converts goals and constraints into executable plans. NVIDIA cuOpt is a GPU accelerated decision optimization engine that addresses mixed integer programming, linear programming, and vehicle routing problems at scale.

Within the intelligent warehouse, cuOpt can:

- Compute pick paths and task sequences for workers and robots
- Reroute fleets in response to congestion, incidents or changing priorities
- Rebatch orders to reduce travel distance while protecting cut off times

These capabilities allow the warehouse to respond to changing conditions without manual recalculation of routes and assignments.

Robotics simulation and orchestration with NVIDIA Isaac

NVIDIA Isaac Sim runs on Omniverse and uses OpenUSD scenes to provide realistic simulation for robots.

Warehouse teams can use Isaac Sim to:

- Test new robot behaviors in accurate copies of their facilities
- Validate path planning and collision avoidance strategies
- Generate synthetic data to train perception models

Once behaviors and workflows prove safe and effective in simulation, orchestration tools such as Isaac based fleet managers coordinate robots in production. The same optimization outputs that cuOpt generates feed directly into these controllers.

Vision AI with NVIDIA Metropolis

NVIDIA Metropolis provides a platform and ecosystem for visual AI agents that operate from edge devices to cloud infrastructure.

In a warehouse, Metropolis can:

- Monitor queue lengths and detect unusual build ups
- Identify conveyor stoppages and product spills
- Track compliance with safety zones and restricted areas

These visual signals become events that feed the flywheel. When the system detects a growing queue or a stopped lane, it triggers analysis in the twin and optimization engines rather than waiting for manual observation.

Integrated stack for continuous improvement

When these layers come together, the warehouse gains a shared digital environment and a consistent way to move from data to action. Data integration keeps the twin current. AI and optimization generate options. Simulation and orchestration validate and apply those options. Vision AI keeps the system aware of events that the data layer alone might miss.

This integrated stack supports both strategic questions, such as how to design a new layout, and operational questions, such as how to respond to a sudden spike in express orders.

Business impact

An intelligent warehouse architecture influences three main areas:

1. Operating expense and productivity
2. Capital efficiency and design risk
3. Resilience and safety

Operating expense and productivity

It focuses on how architecture improves daily throughput, labor use, and the effectiveness of automation. Better visibility and optimization reduce waste in daily operations.

Enterprises can:

- Cut unproductive travel by planning routes and tasks around current congestion and order mix.
- Use robots and automated systems where they provide clear benefit rather than applying static rules across all flows.
- Lower the amount of manual firefighting by detecting risks earlier and applying tested responses.

These changes translate into more orders processed per labor hour, fewer unplanned overtime spikes, and more predictable use of automation.

Capital efficiency and design risk

It focuses on how well the warehouse uses space and equipment and how confidently it makes large design and investment decisions. Digital twins and simulation change how teams make design and investment decisions.

Teams can:

- Compare multiple layouts and automation options using realistic demand and operating scenarios.
- Size fleets of AMRs, AGVs and other assets based on data from simulations rather than simple rules of thumb.
- Validate changes to control logic and operating policies before deployment on the physical floor.

This approach reduces the likelihood of late design changes, physical prototypes, and on floor experiments that disrupt operations and increase risk.

Resilience, safety and collaboration

This covers the ability to absorb shocks, protect people, and enable better decisions across operations, engineering, IT, and finance. The architecture also supports more resilient and safer operations.

Organizations gain:

- A shared visual environment for operations, engineering, IT and finance to review current performance and potential changes.
- Vision based monitoring and simulation validated rules help reduce unsafe patterns and near misses.
- Clear links between daily operational decisions, medium term process changes and long-term investments in network design and automation.

These gains do not come from a single feature. They're due to the compounding effect of many small decisions made with better information and tested options.

Key use cases

The intelligent warehouse architecture applies to several high value operational challenges. The following four use cases illustrate how it works in practice:

1. Warehouse layout and automation planning
2. Autonomous fleet management and training
3. Real-time operations and incident response
4. Continuous policy and process improvement

Warehouse layout and automation planning

Design teams often rely on CAD layouts, spreadsheets, and vendor models. They evaluate only a handful of options and rely heavily on experience. After construction or installation, major changes become difficult and expensive.

With a digital twin on Omniverse and OpenUSD, teams can import CAD layouts and historical warehouse management data into a single scene. They build several candidate layouts and automation configurations, then run “day in the life” simulations that reflect realistic peaks and product mixes.

This process highlights bottlenecks, supports tradeoff discussions, and builds confidence in capital decisions. This enables stakeholders to evaluate the impact of choices rather than debate them in abstract terms.

Autonomous fleet management and training

Many AMR (Autonomous Mobile Robot) and AGV (Automated Guided Vehicle) deployments start with static routing and basic traffic management. Robots cluster in popular aisles, and human workers adjust around them. Operations teams often test new behaviors on the live floor.

When the fleet connects to the intelligent warehouse stack, the digital twin represents the warehouse and robot fleet, and cuOpt computes dynamic routes based on current tasks and congestion. Isaac Sim hosts a virtual copy of the site where teams test new strategies and safety scenarios.

As a result, the fleet uses capacity more evenly, robots and humans interact more predictably, and new behaviors move into production only after passing simulation tests.

Real-time operations and incident response

Supervisors typically notice issues when queues already stretch into aisles or when downstream teams raise alarms. Responses often require urgent labor shifts or expedited shipping.

With Metropolis, cameras monitor queues, workstations, and critical conveyors. When the system detects a growing queue or a stalled lane, it generates an event. Agentic AI uses the twin to estimate the impact on service levels, tests routing and staffing options, and uses cuOpt to compute updated assignments.

Leaders receive early, structured options instead of late, ad-hoc choices. They can reduce the number and severity of incidents that lead to missed service levels and costly recovery actions.

Continuous policy and process improvement

Many warehouses only change their wave plans, picking rules and batching policies during scheduled review meetings. In those meetings, teams often rely on basic reports and informal feedback from supervisors instead of detailed data. After they release a new rule, they usually leave it in place until the next formal review.

In the intelligent warehouse, the flywheel uses operational data to detect patterns that suggest policy changes. It tests new rules in simulation and controlled experiments and measures impact on throughput, cost, and service. The system keeps a record of which policies perform better under which conditions.

This approach turns process changes into a continuous, evidence-based activity rather than a periodic project.

Implementation roadmap

Moving to an intelligent warehouse is a journey, not a single project. Most enterprises already have running facilities, existing systems, and live service commitments. A phased roadmap helps teams make progress without disrupting operations.

Foundation and targeted pilot

Integration and closed loop control

Agentic AI and multi-site scaling

Each phase focuses on a specific goal, produces a concrete outcome, and sets up the next phase.

Phase 1: Foundation and targeted pilot

Objective: Establish data foundations and prove value in one focused area.

Key steps:

- Select a specific area or flow with clear pain and measurable outcomes, such as outbound staging or a high-volume pick zone.
- Integrate data from the warehouse management system, key automation systems, and relevant sensors.
- Build the initial digital twin for the selected area in Omniverse and validate it against historical performance.
- Run offline simulations to explore improvements in routing, layout, or policies.

Outcome: The organization ends this phase with a working twin, basic data pipelines, and an initial set of quantified improvements that demonstrate the value of the approach.

Phase 2: Integration and closed loop control

Objective: Keep the twin in sync with live operations and use it to drive real decisions in a controlled way.

Key steps:

- Connect streaming data feeds so the twin reflects near real-time conditions.
- Deploy cuOpt for a subset of routes or tasks and wire optimization outputs into existing systems.
- Introduce Isaac Sim for robotics simulation where automated fleets operate and link it with the twin.
- Deploy Metropolis based vision AI in areas with camera coverage to detect queues and incidents.

Outcome: At least one area of the warehouse now operates in a closed loop. The system senses conditions, tests options, optimizes plans, applies them and measures results.

Phase 3: Agentic AI and multi-site scaling

Objective: Introduce agentic AI for higher level decision support and roll out the architecture across more areas and sites.

Key steps:

- Deploy NVIDIA NIM microservices to host language and vision models tuned to warehouse scenarios.
- Define and implement guardrails for safety, compliance, and business policies.
- Document patterns, templates, and playbooks that describe how to apply the stack in new facilities.
- Align governance, change management, and training so operations, IT, and engineering teams can use the new capabilities effectively.

Outcome: The intelligent warehouse stack becomes a standard part of the enterprise toolkit. New projects use established patterns, and continuous improvement becomes part of regular operations rather than a one-off initiative.

Conclusion

Warehouse logistics faces sustained pressure from demand volatility, labor constraints, and rising service expectations. Static planning tools and isolated automation can ease some pain, but they force leaders to react to issues only after those problems appear on the floor.

Digital twins and agentic AI enable operations teams to run warehouses as intelligent systems. By combining a high-fidelity twin, AI reasoning, optimization, robotics simulation, and vision AI in a closed loop, operations teams can make better decisions, respond faster to events, and design more effective facilities.

Progress starts with a focused pilot, not a sweeping transformation. Teams can target a single area with clear pain points, build a basic twin, and deploy a simple optimization loop to demonstrate value and build trust. From there, enterprises can expand the approach to more flows, more automation, and more sites.

An intelligent warehouse does not depend on a single breakthrough. Instead, teams drive improvement through repeated cycles of sensing, simulating, deciding, and acting. This white paper describes an architecture that gives enterprises a practical way to run those cycles with greater clarity, safety, and business alignment.

fractal.ai

info@fractal.ai @fractalai Linkedin.com/company/fractal-analytics