Connecting AI to Decisions

/01 DATA
/02 LOGIC
/03 ACTION
↘
DECISIONS

/02
/01
/03
/01
/02
/03

→ PALANTIR.COM

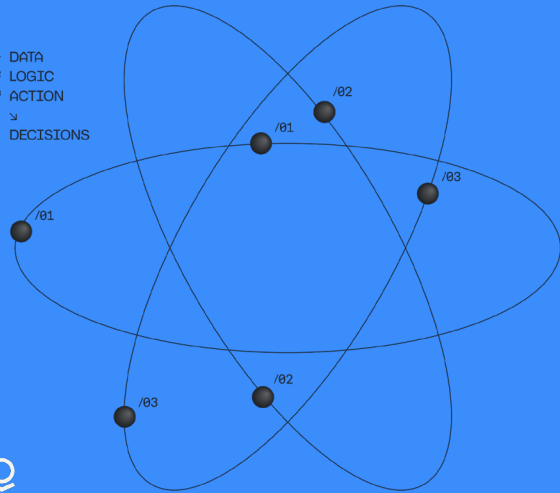## WITH THE PALANTIR ONTOLOGY

by Akshay Krishnaswamy, Chief Architect, Palantir

The Palantir Ontology

Palantir's software powers real-time, AI-driven decision-making in many of the most critical commercial and government contexts around the world. From public health to battery production, our customers depend on Palantir AIP to safely, securely, and effectively leverage AI in their enterprises—and drive operational results.

While many factors contribute to achieving and scaling operational impact—including our AIP Bootcamps, where customers are hands-on-keyboards and achieving outcomes with AI in a matter of hours—the key differentiator is a software architecture which revolves around the Palantir Ontology.

The Ontology is designed to represent the *decisions* in an enterprise, not simply the data. The prime directive of every organization in the world is to execute the best possible decisions, often in real-time, while contending with internal and external conditions that are constantly in flux. Traditional data architectures do not capture the reasoning that goes into decision-making or the action that results, and therefore limit learning and the incorporation of AI. Conventional analytics architectures do not contextualize computation within lived reality, and therefore remain disconnected from operations. To navigate and win in today's world, the modern enterprise needs a decision-centric software architecture.

To understand the value of the Ontology, let's start by considering the three elements of any decision:

- **Data,** or the information used to make a decision
- **Logic**, or the process of evaluating a decision
- **Action**, or the execution of the decision

The Ontology integrates these three constituent elements of decision-making into a scalable, dynamic, collaborative foundation which reflects the ever-changing conditions and ambitions of the organization as they evolve in real time.
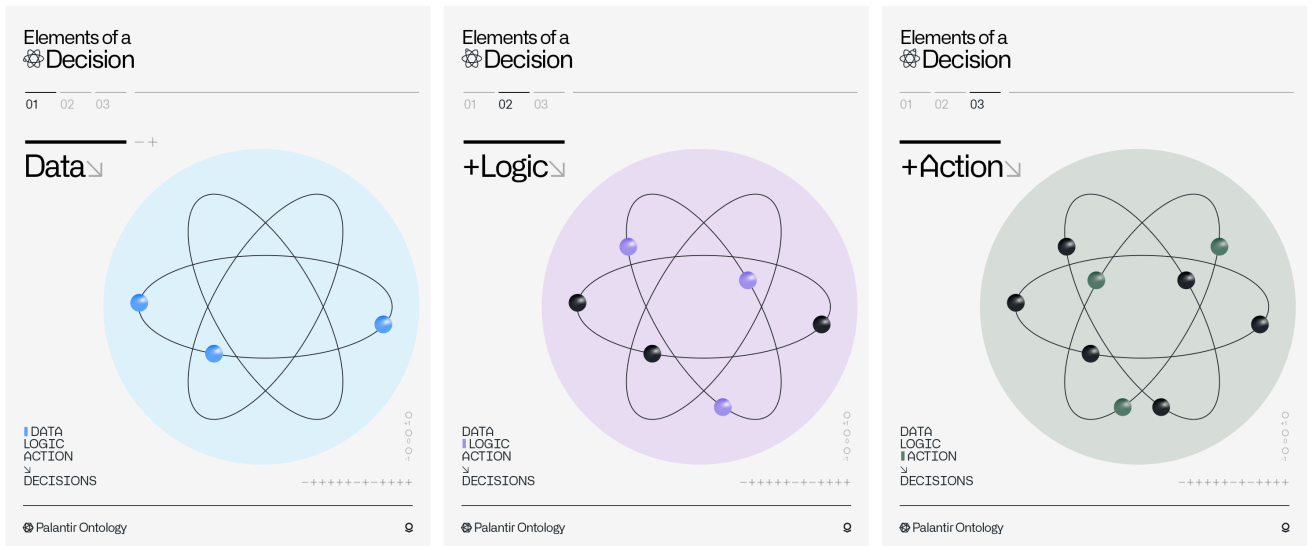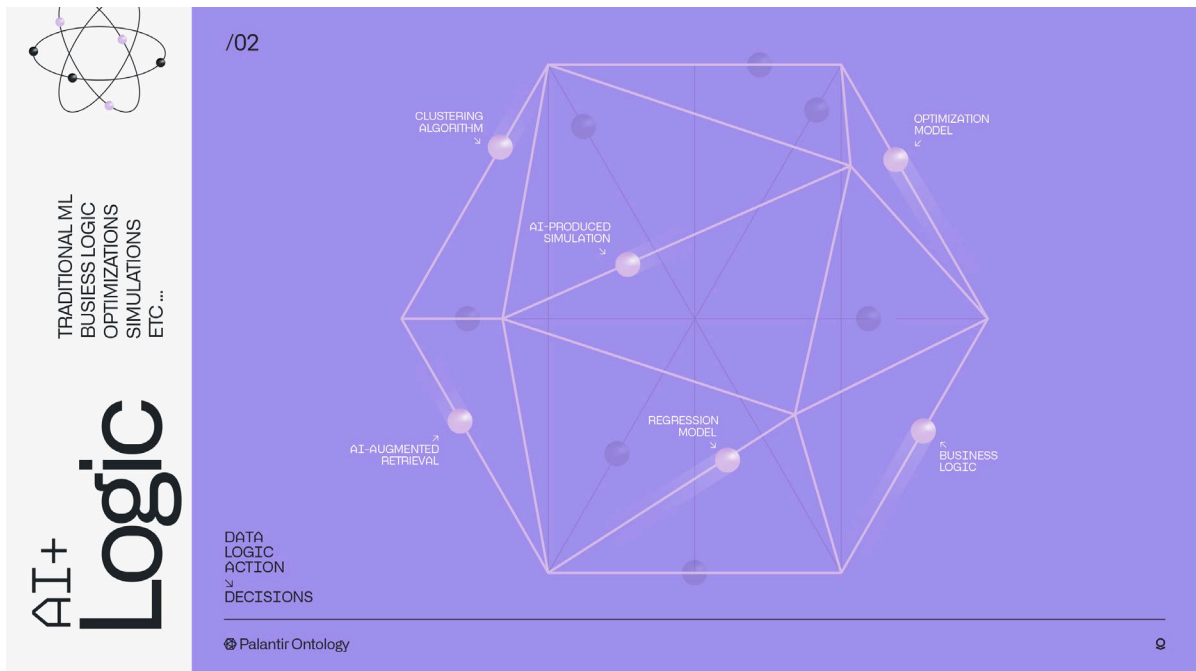

At a fundamental level, every decision is comprised of data (the information used to make a decision), logic (the process of evaluating a decision), and action (the execution of the decision).
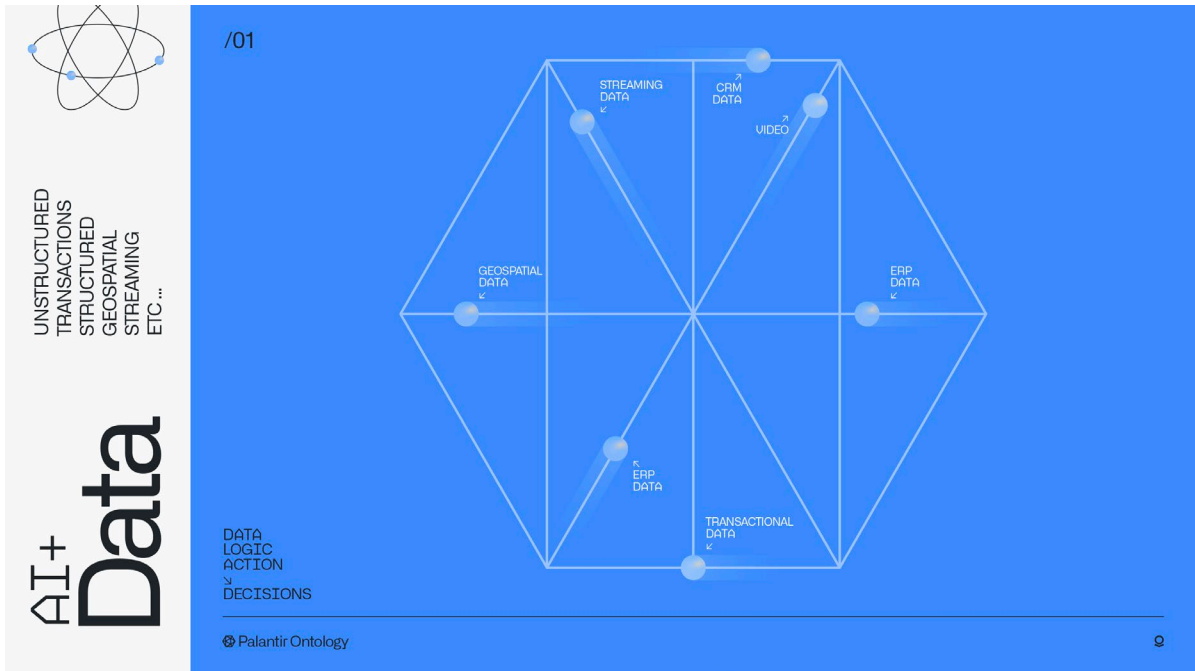
The Ontology integrates all modalities of data into a full-scale, full-fidelity semantic representation that captures the constantly evolving reality of the enterprise and serves as the foundation for powerful AI-driven workflows.

Today's organizations are inundated with unprecedented amounts of data. The volume, variety, and velocity of data sources is not only increasing, but *accelerating* over time. While plenty of ink has been spilled on the virtues of cleaning and unifying data, in the age of AI the principal problem is *relevance*. Relevant data of course includes the full range of enterprise data sources — structured data, streaming and edge sources, unstructured repositories, imagery data, and more — but it also includes the data that is generated by end users as decisions are being made. This "decision data" contains the context surrounding a given decision, the different options evaluated, and the downstream implications of the committed choice. Generative AI provides a breakthrough ability to synthesize learnings from the full scale of decision data, and continuously enrich both human- and AI-driven workflows. Naturally, integrating the full range of enterprise data with the fluid landscape of decision data requires a very different architecture than a classical database management solution that is optimized for reporting and analytics.
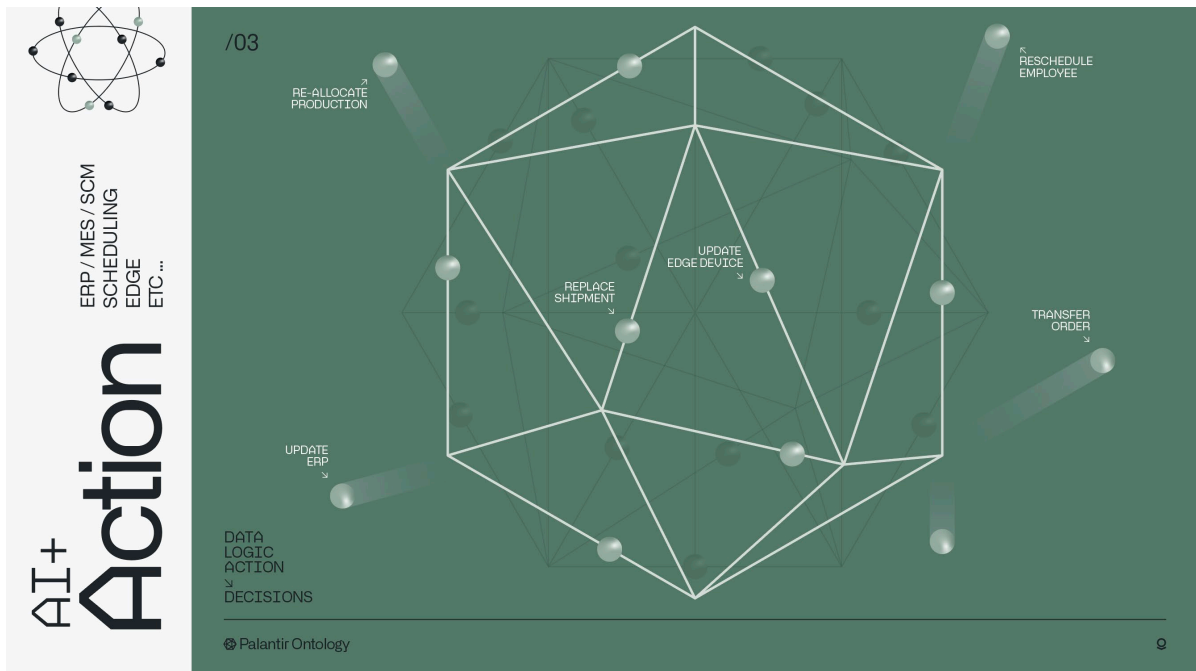
The Ontology integrates all modalities of data into a full-scale, full-fidelity semantic representation of the enterprise. The wide range of operational data sources (ERPs, MES, WMS, et al.) can be synchronized and contextualized alongside data streams from IoT and edge systems, the relevant sections of unstructured data repositories, geospatial data stores, and more. The Ontology unites and activates these fragmented pools of data, and surfaces them in the language of the enterprise. Instead of dealing with golden tables that flatten the richness of operations into narrow schemas, the full expanse of the enterprise comes to life in the form of objects, properties, and links which evolve in real-time, and are designed to be embedded directly into decision-making workflows. Critically, the Ontology is designed to safely capture the decision data that is produced by operational users as they carry out daily work (e.g., within supply chains, hospital systems, customer service centers). The end-to-end "decision lineage" of when a given decision was made, atop which version of enterprise data, and through which application, is automatically captured and securely accessible to both human developers and generative AI. This provides the comprehensive foundation that is required to power AI-driven learning at scale.

AI+ Logic

/02

TRADITIONAL ML
BUSIESS LOGIC
OPTIMIZATIONS
SIMULATIONS
ETC…

CLUSTERING ALGORITHM

OPTIMIZATION MODEL

AI-PRODUCED SIMULATION

AI-AUGMENTED RETRIEVAL

REGRESSION MODEL

BUSINESS LOGIC

DATA
LOGIC
ACTION
↓
DECISIONS

✪ Palantir Ontology

The Ontology enables users to construct workflows that incorporate and combine heterogeneous logic assets. Ultimately, this means that AI-driven reasoning can be securely introduced into increasingly complex decision-making contexts.

While data is foundational, it is only one dimension of the decision-making process; it must be complemented by the reasoning, or logic, that determines when and how to make a given decision. The logic that underpins a decision can be a simple piece of business logic within a core business system, a forecast model that is maintained using a cloud data science workbench, an optimization model that uses several data sources to produce an operational plan — among myriad possibilities. In real-world contexts, human reasoning is often what orchestrates which logical assets are utilized at different points in a given workflow, and how they are potentially chained together in more complex processes. With the advent of generative AI, it is now critical that AI-driven reasoning can leverage all of these logical assets in the same way that humans have historically. Deterministic functions, algorithms, and conventional statistical processes must be surfaced as "tools" which complement the non-deterministic reasoning of large language models (LLMs) and multi-modal models.

The Ontology enables the full set of logic assets — the calculations and processes that dictate *how* decisions are made — to be connected and contextualized for both human and AI users. This includes business logic pertaining to customer interactions often found in CRMs and ERPs; the modeling logic that drives conventional machine learning, which is spread across data science environments; and the planning, optimization, and simulation algorithms that are typically intertwined with domain-specific tools. The Ontology's flexible "logic binding" paradigm provides a consistent interface for constructing workflows that seamlessly incorporate and combine heterogeneous logic assets — which may all live in very different environments (e.g., on-premises data centers, enterprise cloud environments, SaaS environments, the Palantir platform). Ultimately, this means that AI-driven reasoning can be smoothly introduced into decision-making contexts which leverage diverse sets of logic, and which have been traditionally steered exclusively by human users.

◉ Palantir

/03

RE-ALLOCATE
PRODUCTION

RESCHEDULE
EMPLOYEE

UPDATE
EDGE DEVICE

REPLACE
SHIPMENT

TRANSFER
ORDER

UPDATE
ERP

ERP / MES / SCM
SCHEDULING
EDGE
ETC...

AI+
Action

DATA
LOGIC
ACTION

DECISIONS

✸ Palantir Ontology

The Ontology natively models actions within a cohesive, decision-centric model of the enterprise, enabling human and AI-driven actions to be safely staged as scenarios, governed with the same access controls as data and logic primitives, and securely written back to every enterprise substrate.

With both information (the data) and reasoning (the logic) incorporated into a shared representation, the remaining piece to model is the execution and orchestration of the decision itself (the action). Closing the action loop as decisions are made in real-time is what distinguishes an operational system from an analytical system. Since Palantir's inception, the execution of decisions has been as critical a consideration as the synthesis of data, or the incorporation of analytics. This has required the design and implementation of a broad set of functionality which includes how to safely capture decisions which might be happening simultaneously and are potentially in conflict; a collaborative model that segments those who can explore possible decisions, those who can stage decisions for review, and those who can commit those decisions; and an extensive framework for synchronizing decisions to existing databases, edge platforms, and rugged assets.

The Ontology natively models actions within a cohesive, decision-centric model of the enterprise. If the data elements in the Ontology are "the nouns" of the enterprise (the semantic, real-world objects and links), then the actions can be considered "the verbs" (the kinetic, real-world execution). With every Ontology-driven workflow, the nouns and the verbs are brought together into complete sentences through human- and/or AI-driven reasoning, which incorporates various pieces of logic. While uniting data within a semantic model is itself valuable, and while it is imperative to stitch together the logic required to holistically evaluate possible decisions — it is all ultimately of limited value *unless* the executed decisions are synchronized with operational systems. The Ontology enables human and AI-driven actions to be safely staged as scenarios, governed with the same granular access controls as data and logic primitives, and securely written back to every enterprise substrate — transactional systems, edge devices, custom applications, et al.

In short, the Ontology brings together data, logic, and action into a decision-centric model of the enterprise, which can be jointly leveraged by both humans and AI. Everything from data integration, to application building, to end user workflows is driven through a battle-tested, modular architecture — enabling human users and AI-driven copilots and automations to query, reason, and act across a shared operational foundation.

Operational Example

Let's step through a notional example to unpack how the Ontology is enabling organizations across 50+ sectors to activate AI-driven workflows in days.

Titan Industries, a fictional manufacturer of medical equipment, produces a range of finished goods, from syringes to surgical masks, each of which requires moving a precise set of materials through an associated manufacturing process. A diverse set of teams is managing everything from supplier relations, to warehouse operations, to production of the finished goods, to distribution to end customers; decisions are interdependent, and constantly adapting to changing circumstances. In short, every day brings unique challenges when operating the business.

In this example, Titan Industries is faced with an unexpected disruption with one of their major suppliers, who provides the key raw materials needed to produce surgical masks. Given the tight production schedules across Titan's manufacturing plants and the escalating demand from customers for surgical masks, this disruption is poised to create serious issues with fulfilling outstanding customer orders. Fortunately, Titan's operational leadership has connected the wide array of data sources, logic assets, and systems of action into their enterprise ontology — and has the ability to swiftly respond.
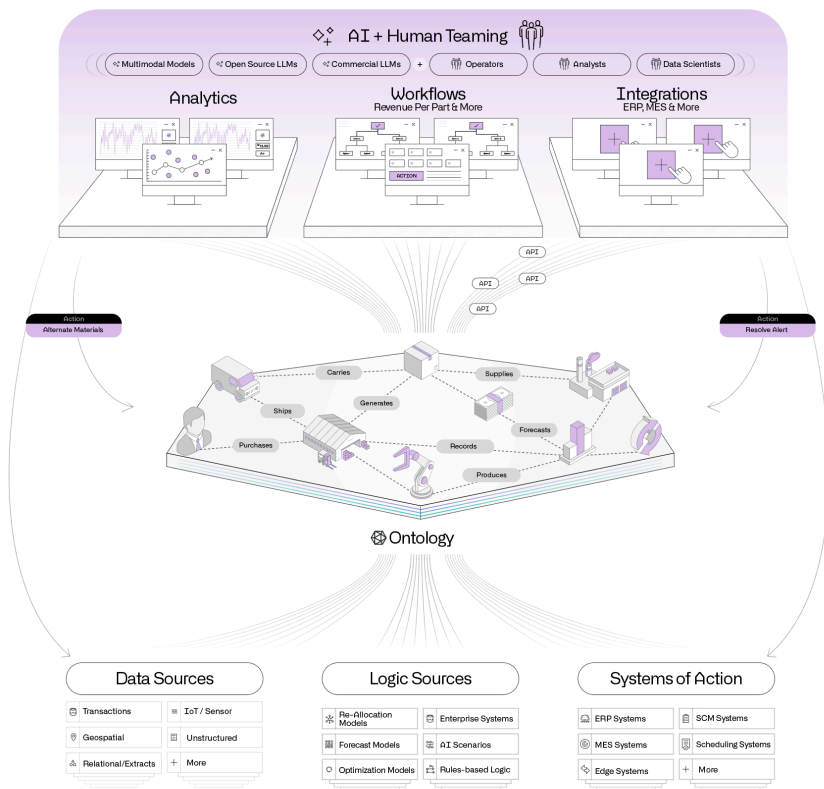
Example →
Titan Industries

Elements of a Raw Materials ⊛ Decision

Full visibility into revenue at stake for each shortage

Prioritizing shortages based on revenue at stake

AIP-driven intelligence and resolutions

Continuous learning informs future decisions

⊛ Palantir Ontology

Titan's ontology brings together all decision-making elements necessary to navigate this raw materials disruption: it provides full visibility into revenue at stake for each shortage to inform prioritization, allows for AI-driven recommendations and ultimate resolutions which account for the enterprise's operational reality, and drives writeback and continuous learning to not only keep systems current, but also optimize future decisions.

Titan will start by assessing the immediate impact of the supplier shortage, and will then employ AI to assess possible reallocation strategies across production lines, before finally translating their decisions into a set of connected actions that will simultaneously update warehouse processes, production schedules, and fulfillment routes.

Titan's ontology provides real-time, end-to-end visibility into the operations happening across each interdependent part of the business — enabling both leadership and on-the-ground teams to quickly understand the supplier disruption. The vital data systems pertaining to supplier management, warehouse operations, production activity within plants, distribution center processing, and customer fulfillment are all synthesized into semantic objects and links, which reflect the language of the business. In a few clicks, an operations leader is able to pinpoint the surgical mask production that is at risk due to the raw material shortage, and through the connections in their ontology, navigate to every outstanding customer order that is now also at risk. The Ontology's granular security model ensures that more sensitive data elements (e.g., financial metrics) are automatically hidden by default, as the response widens to include more teams across the enterprise.
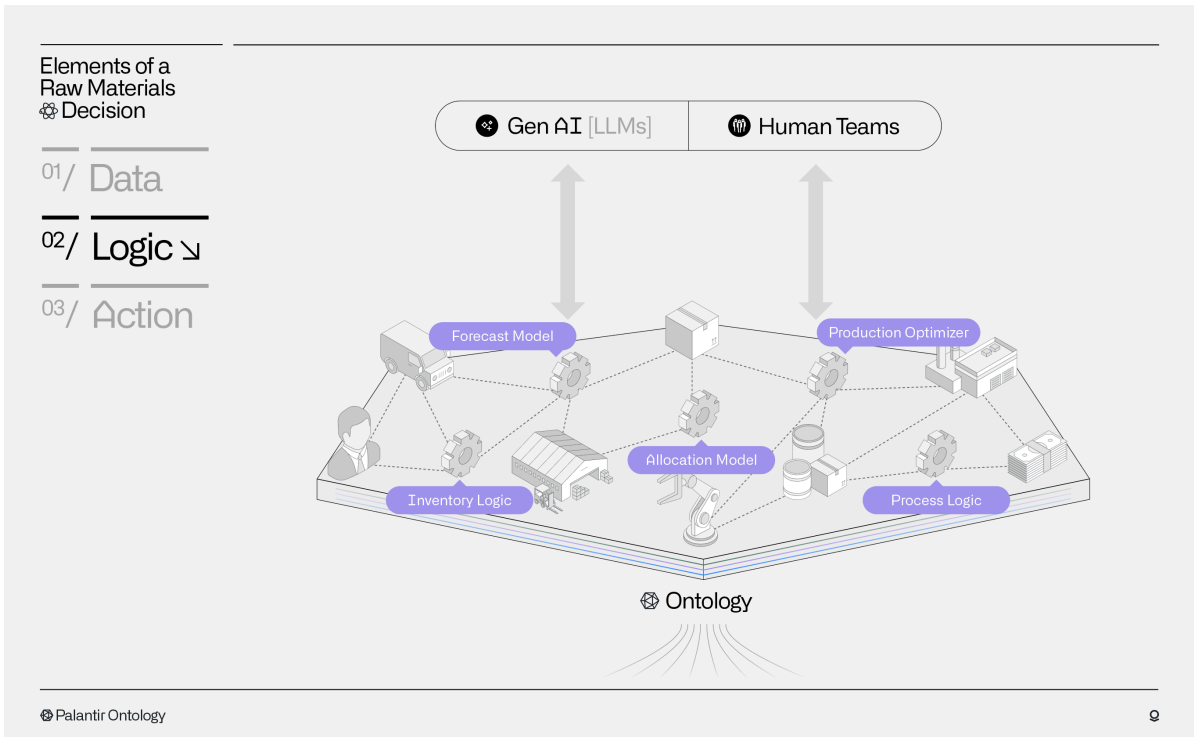
While it is seamless for operational users to navigate the Ontology through intuitive [Workshop](#)- and [SDK](#)-driven applications, the inclusion of large language models (LLMs) is a force multiplier for Titan Industries. AI-powered copilots, which leverage both open-source and proprietary LLMs, are able to fluidly navigate across supplier information, stock levels, real-time production metrics, shipping manifests, and customer feedback all contained within the organization's ontology. Critically, all AI activity is controlled with the same security policies that govern human usage — ensuring that Titan engineers always have precise control over what the LLMs can query, recommend, and act upon. Each constructed and deployed AI copilot can be considered a new team member, who is gradually granted a wider purview as Titan team members gain confidence in its performance.
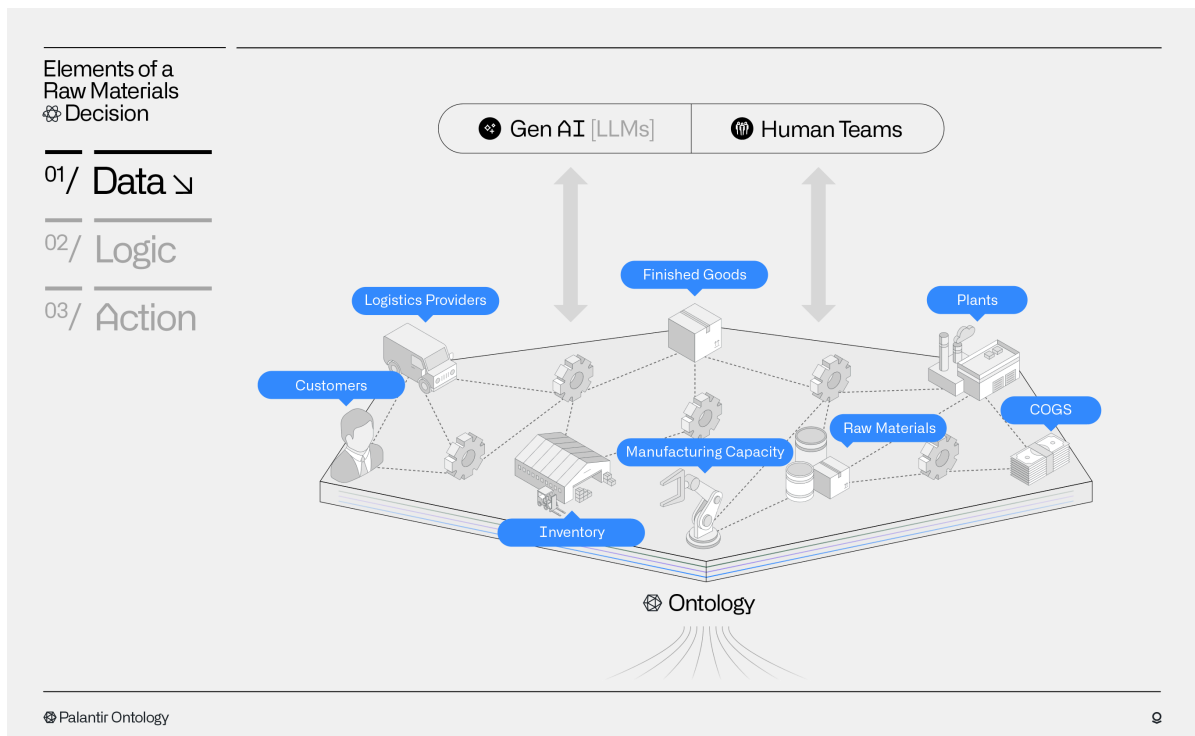
Titan's ontology integrates data from the organization's vital systems, synthesizing it into semantic objects and links which provide real-time, end-to-end visibility into operations and allow both leadership and on-the-ground users to rapidly assess the full impact of the disruption.

Situational awareness is only the tip of the ontological iceberg; Titan Industries needs to rapidly identify solutions to deal with the supplier disruption, and explore the tradeoffs inherent with each possible decision. Fortunately, the diverse set of forecast models, allocation models, production optimizers, and other logic assets have been connected into Titan's ontology, alongside the aforementioned data sources. This enables supply chain analysts to quickly run a battery of simulations that detail the consequences of the different possible material substitutions. The connected, real-time nature of the Ontology is key at this stage, since substituting raw materials will potentially have downstream implications for the other products (e.g., syringes, gloves) being produced from the same materials. As the simulations are run, the simulated outputs are staged as ontology scenarios, which safely package the proposed changes into a sandboxed subset of the Ontology — enabling teams to safely explore and analyze the implications of the decision before committing to it.

The true game-changer for the Titan team is that AI-driven copilots and automations can securely leverage the full range of logic assets, and the same scenarios framework. The Ontology enables LLMs to go beyond the data-centric limitations of retrieval-augmented generation, and instead interface with the interconnected data, logic, and action primitives in the Ontology through an extensible tools paradigm. This means that as Titan's analytics and data science teams are creating new machine learning models in their cloud workbenches, tuning optimization algorithms within enterprise systems, and fine-tuning LLMs using Palantir's open model building framework, the Ontology securely surfaces all of these logic assets as AI-ready tools. In this case, Titan has created a tuned AI copilot, "Disruption Bot," that is able to use a set of Ontology-driven tools to scan across the full range of enterprise data sources, the after-action reports on prior courses of action taken in similar situations, and the potentially applicable material reallocation models. Because of the rich, dense context provided through the Ontology, Disruption Bot is able to surface a novel reallocation plan, which uses a newer model that the supply chain analysts had not yet considered. With the consequences of the plan safely staged in a scenario, the AI's proposed decision is handed off to a human analyst for final review.

Elements of a Raw Materials ⚙ Decision

01/ Data
02/ Logic ↘
03/ Action

⊗ Gen AI [LLMs]   🖐 Human Teams

Forecast Model
Production Optimizer
Inventory Logic
Allocation Model
Process Logic

⚙ Ontology

⚙ Palantir Ontology

The Ontology securely surfaces Titan's logic assets — from machine learning to optimization models — as AI-ready tools, providing rich, dense context for both human- and AI-driven workflows.

 Palantir

With a viable plan to address the material shortage identified, Titan Industries needs to rapidly and safely push the decision to the operational systems that run the constituent processes. Given that the enterprise has grown through acquisition, and contains a diverse and delicate mix of critical operational systems, the Titan IT team is vigilant about which processes can write back to these systems, and under which conditions. Fortunately, the Ontology applies the same rigorous control and validation to actions as it does to data and logic; enabling fine-grain control over who can invoke a given action, test-driven frameworks for publishing changes, the ability to stage and review changes in batch, and detailed logging for every event. In this case, the execution of the material reallocation plan automatically orchestrates a set of writeback routines, each tuned for the receiving system: the warehouse management system receives an API-driven update; the three ERP systems each receive updates via native Ontology-driven connectors, which abide by the safeguards in each system; and the production planning system receives a consolidated flat file, which it ingests asynchronously. As actions are executed, the Titan IT team can monitor system responses, and always has the power to audit past activity.

The Ontology provides the guardrails needed for AI to safely take action within permitted boundaries. Alongside data and logic, actions can be automatically surfaced as tools for AI-driven copilots and automations. The scope of an action can be limited to simply reflecting a given change (e.g., an edit to an object, or the creation of a new object) in the Ontology itself; or can write back to single, or multiple systems. In Titan's context, they have granted Disruption Bot and the handful of other production AI copilots access to a handful of actions. In the default case, these actions (e.g., changing the status of a work order, or pushing a reallocation plan) can only be staged by the AI, and are then handed off to a human for final review. However, with the granular logging and operational instrumentation provided by the Ontology (and the wider Palantir platform), Titan is able to surgically choose which trusted, well-worn AI processes can automatically close the action loop without human review. As conditions evolve, the latitude given to AI can be expanded or contracted — and instantly reflected across all Ontology-driven workflows.
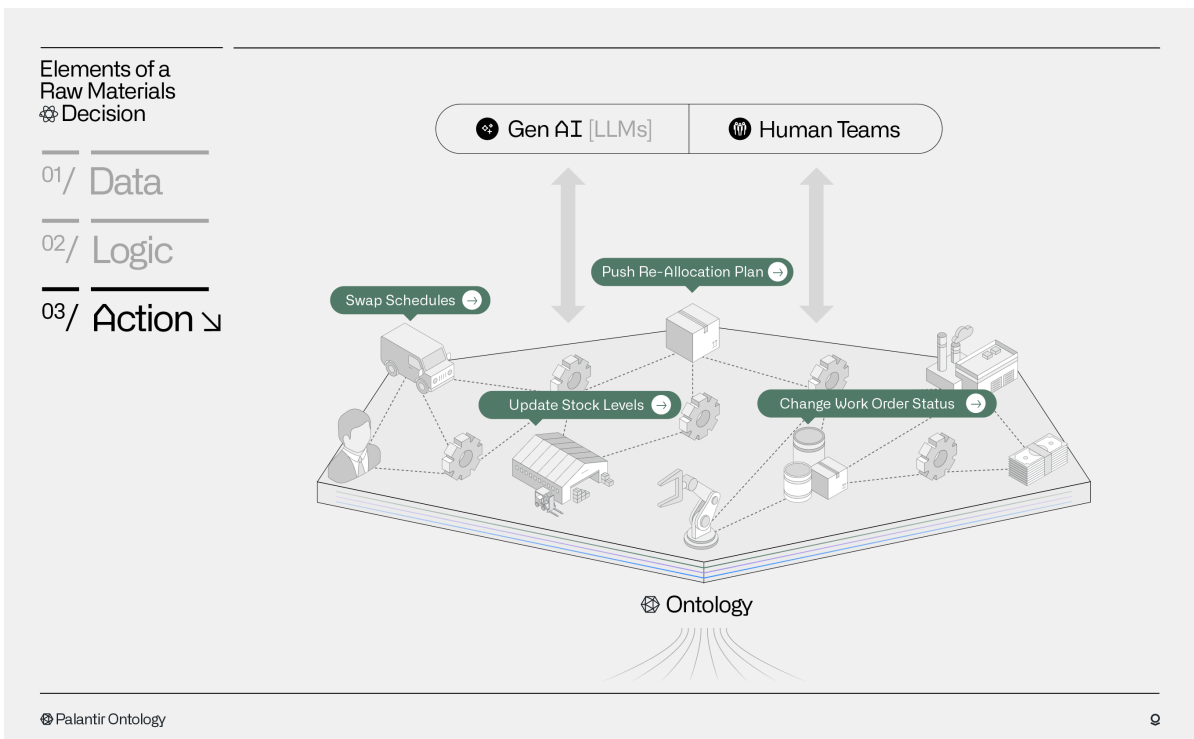
The Ontology allows Titan to automatically surface actions as tools for AI-driven copilots and automations while providing the necessary guardrails for AI to safely take action within predetermined boundaries.

What comes after the crisis? With data, logic, and action all connected into Titan's ontology, the organization has the ability to conduct powerful decision-centric learning. The human-AI teaming that produced a specific solution to the material shortage also revealed generalizable workflows, which the organization will want to memorialize and surface in the future. Every data element, logic asset, and action assessed is captured in end-to-end decision lineage — which serves as rich, contextual fuel for optimizing the performance of AI. The aggregate decisions made by thousands of users throughout Ontology can be securely leveraged as training data when fine-tuning models, and can be distilled into targeted principles that are called upon during LLM prompting. The tribal knowledge that has been traditionally trapped in the seams of workflows can be illuminated by AI, in order to improve the application of AI.
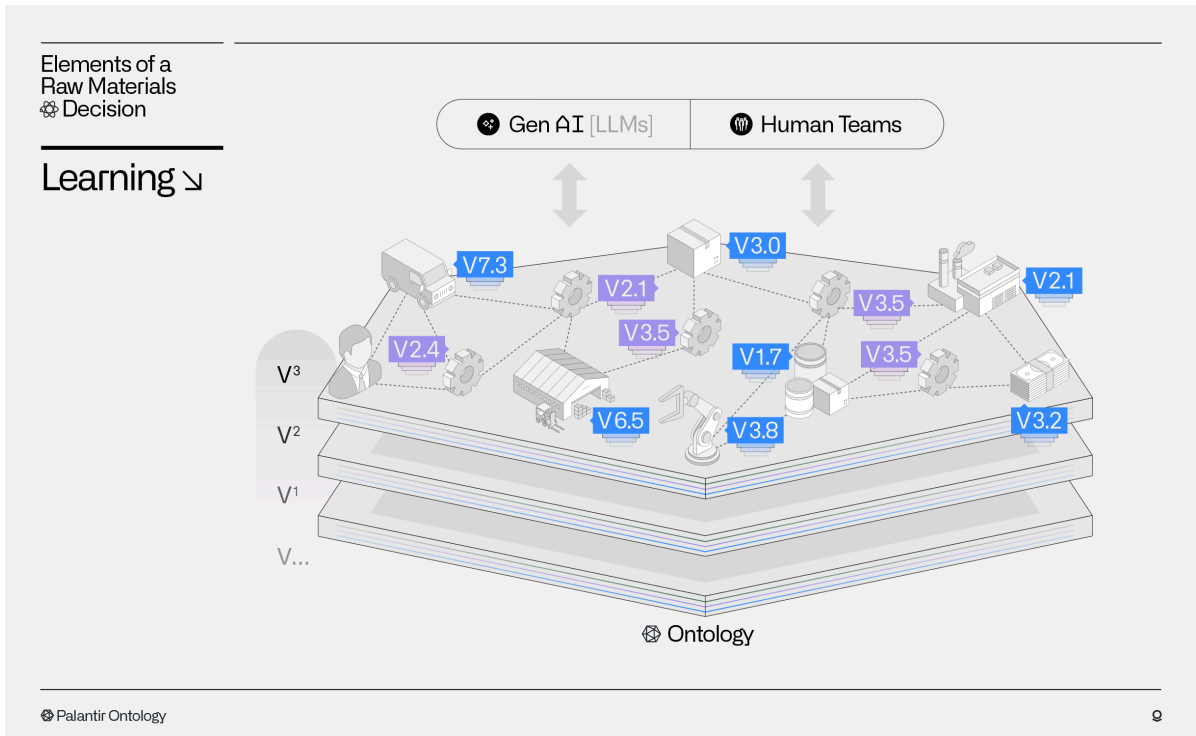
The Ontology captures updates to every data element, logic asset, and action as decisions are made — which serves as rich, contextual fuel for optimizing the performance of AI over time.

Onward with the Ontology

Ultimately, the Ontology allows each organization to connect AI directly into their core operations, and precisely control how and when AI-driven recommendations, augmentations, and automations can be utilized in frontline contexts. This is uniquely possible because the Ontology is decision-centric, not simply data-centric; it brings together the constituent elements of decision-making — data, logic, and action — within a single software system. New data can be rapidly integrated into a full-fidelity semantic representation; new algorithms and business logic can be seamlessly surfaced for both human and AI users; and robust action integration is achieved through real-time connections with the full range of operational systems. Each organization's ontology is a real-time pulse on the changing conditions, ambitions, and decisions being made across teams — ensuring that AI is always anchored in the reality of the enterprise.

Thanks for reading through this introduction to the Ontology. This has only scratched the surface on the Ontology's underlying decision-centric architecture; the system's native simulation and scenario-building capabilities; the extensibility provided through the Ontology SDK; the various ways of connecting commercial and open-source generative AI models to secure subsets of data, logic, and action; and the methods of scaling human-AI teaming across the entire enterprise.

Stay tuned, and reach out to us if you'd like to learn more.

Real-World Examples

- See how J.D. Power uses their ontology to bring AI to the auto industry
- See how HCA Healthcare uses their ontology to help power AI-driven hospital operations
- See how Jacobs uses their ontology to build AI copilots for their engineers