The Next Data Stack

A working thesis on real-time capture, programmatic surfacing, and predictive action at the edge

By Dylan Ferrara · May 2026

1 · CAPTURE	2 · QUERY	3 · ACT
Real-time, sensor-fed	**Programmatic surfacing**	**Predictive at the edge**
IoT sensors: $13B in 2023, $106B by 2030	NLP analytics: 27.5% CAGR through 2030	30B+ connected endpoints by 2030

Premise. The next decade of data businesses will run on three principles working in concert. Continuous capture from real-world sensor networks. Programmatic structuring and surfacing of that data with optional human oversight. Predictive models that act on it at the device level rather than after a round-trip to the cloud. Each is replacing a layer of the legacy systems-of-record stack, where data was sampled, queried by hand, and acted on by people. The consolidation has already started. What follows is a view of where it goes.

Real-time, fed from sensor networks

As machines proliferate throughout industries and the economy, so too will the frontier for physical datasets directly tied to manufacturing processes, transportation patterns, and personal consumption. These data troves have always existed, but in a dimension that we have been unable to capture. That is until now, thanks to an abundance of edge devices encircled by performant sensors that capture, monitor, and predict actions, all in real time.

Autonomous vehicles are probably the best example of this today. On average they carry over 30 sensors, generate between 1 and 19 terabytes of data per hour, and operate at end-to-end response latencies of 50 to 100 milliseconds. By comparison, a human driver takes roughly 1.6 seconds to brake for an unexpected road hazard.[1] Mobileye sits at the center of this build-out, with its ADAS perception stack deployed across most major automakers. And the build-out is not confined to easy environments. Purpose-built, durable sensor stacks are now being broadly deployed from underwater acoustic navigation (Advanced Navigation) to real-time pathogen and contamination detection in food and hospital settings (HyperSpectral), instrumenting territory that was previously dark. The market is following: the IoT sensors market alone is on track from ~$13B in 2023 to $106B by 2030.[2]

The most telling signal is who is paying to own the capture layer. Salesforce buying Informatica. IBM buying Confluent. Publicis buying LiveRamp. And most recently, Travis Kalanick's Atoms acquiring Anthony Levandowski's Pronto to anchor its mining division.[3] Incumbents are bidding for the streams of structured, proprietary, real-world data those companies sit on, because that is increasingly where the durable asset lives.

Programmatic, structured and surfaced for the moment of need

It is obvious that more data will be collected in the years to come. Perhaps an equally important phenomenon is the ease of how that vital information will get called, structured, and presented on the

[1] Tuxera, Premio, and Label Your Data on AV sensor counts and data generation rates; DriveBlocks on end-to-end response latency targets; MIT News (2019) on human hazard-response times.

[2] Grand View Research, IoT Sensors Market Report 2030 (36.8% CAGR).

[3] Atoms / Pronto.ai press materials (April 2026); International Mining and InvestMETS coverage of the Atoms Mining division formation.

fly. Downstream this sits in two camps: formats intuitive to humans, and clean machine-interpretable context. Both have value and different use cases. In this piece we focus on the human component.

Put simply, having a pulse on what is going on, on how things are performing, will get a whole lot easier. Data troves will no longer require massive lead times for data science teams to package and map datasets to analytical software platforms. Instead, key insights will be surfaced automatically, highly relevant to business processes, problem sets, and user activities. Should end users want other information or comparisons, those data points can be crunched, structured, and presented natively in whatever surface they are engaging with at that point in time. Displays will be intuitive, commands will be in natural language, and outputs will be generative and highly precise.

This is more than a natural-language wrapper over a SQL engine. It is an intelligent layer that decides what matters now and brings it forward without being asked. As a market, NLP-driven analytics is forecast to grow from $36B in 2024 to $156B by 2030, with adopters reporting analysts redirect nearly half their time away from query construction toward actual analysis.[4]

Predictive, with inference at the edge

As critical data gets collected and reasoned over, this will inevitably evolve into continuous, business-specific prediction models. These systems will utilize datastreams as they come in, run logic on top of them, reference historical patterns, and actively adjust to optimize outputs in the most efficient and cost-effective manner known. Think self-optimized assembly lines for industrial equipment, or intelligent statewide drone-delivery networks running on a perpetual loop to drive results and favorable customer outcomes.

As alluded to earlier, these devices and networks will overwhelmingly run on the edge. They will independently process and make sense of inputs as they come in, and use those critical nuggets of knowledge to guide future performance. The implication is that we still have a long way to go in discovering more efficient compute architectures, ushering in new memory paradigms and device form factors, and developing scalable context-window compression and management techniques alongside the hardware harnesses that house them.

Inference now accounts for roughly 99.8% of all edge AI compute, and the device base it runs on is on track to exceed 30 billion connected endpoints by 2030. Every one of those endpoints generates a stream of structured, time-stamped, geospatial telemetry that did not exist a decade ago.[5]

Where this leads

There is tremendous opportunity in all of these areas to innovate. As these capabilities develop, we will start to see novel wearable devices doing things that are tough to imagine today. They will solve complex problems and surface actionable, geo-located preference data before you even think to ask for it. A simple analogy is how your Nest thermostat already knows the temperature you like at a given point in the day. We will see a step-function shift here, and this is just the consumer side. Industrial-scale automation is a whole different beast, factoring real-time production and supply-chain signals into wholesale material negotiation in ways that fundamentally compress the procurement cycle.

Sources: Grand View Research, MarketsAndMarkets, IoT Analytics; Tuxera, Premio, Label Your Data, DriveBlocks, MIT News; Atoms / Pronto.ai press materials; Promethium and Querio adoption surveys. Sector positioning informed by the FirstMark MAD Landscape (2026).

---

[4] Aggregated NLP analytics forecasts (BlazeSQL, Querio, industry trackers); Promethium and Querio enterprise adoption surveys, 2024–2026.

[5] IoT Analytics, State of IoT 2024–2025 (global connected IoT device base forecast at ~27–30 billion by 2030); Grand View Research, Edge AI Chips Market (inference ~99.8% of edge AI compute volume).