

WINTERCORP REPORT

BY RICHARD WINTER

W I N T E R C O R P O R A T I O N

PROJECTING ANALYTIC DATA WORKLOADS FOR AGENTIC AI

.....

*Why Enterprise Scale Agentic
Workloads Demand a Different Kind
of Data Platform*

.....

SPONSORED
RESEARCH
PROGRAM

EXPERTS IN ANALYTIC DATA MANAGEMENT AND AI AT SCALE



PROJECTING ANALYTIC DATA WORKLOADS FOR AGENTIC AI

.....

*Why Enterprise Scale Agentic
Workloads Demand a Different Kind
of Data Platform*

RICHARD WINTER



WinterCorp

www.wintercorp.com

42 DERBY LANE, TYNGSBORO, MA
617-695-1800

S Summary

ENTERPRISES ARE ON THE THRESHOLD of a new era in which intelligent agents, powered by generative AI and operating with significant autonomy, will be doing much of the routine work. In a report released in April 2026, IDC projected that agentic AI would deliver 22.5 trillion dollars in business value by 2031 via productivity gains, new revenue models and business transformation.¹

While ROI expectations vary widely with the use case, 62% of organizations surveyed in one recently published report were expecting an average ROI of 171% on their agentic AI investments.² Many expect that the number of intelligent agents will far exceed the number of employees in most companies, within a few years.

Widespread deployment of autonomous agents can produce enormous benefits, but enabling changes are required. Much of what these agents will do involves the use of enterprise data managed by analytic data platforms. Agents will be analyzing that data to solve business problems and answer business questions.

COLLECTIVELY, the agents will impose a significant new workload on the analytic or operational data platform, typically much larger — and very different — than what we see today. As agents are deployed into production use, that workload can grow and change quite rapidly. In addition, intelligent agents will need business context in order to use data effectively.

Therefore, deployment of agents at scale imposes three critical requirements:

1. The data platform must become a **knowledge** platform, capable of storing, curating and retrieving on demand the business context of the enterprise data;
2. The data platform must be **autonomous**, in order to manage the dynamic and largely unpredictable demands generated by the intelligent agents; this means it must self-optimize and self-manage in response to intensive, rapidly changing workloads; and,
3. The resulting knowledge platform must be **highly scalable and efficient** for processing **enormous volumes of short queries**, which were not present in such volumes prior to the large-scale use of agents.

In short, the enterprise of the agentic era will need an autonomous knowledge platform to manage its data. •

WinterCorp — an independent expert in data, analytics and AI at large scale — has been focused on the demanding requirements of data-intensive enterprises for more than thirty years. As the author of this report, we are sounding an alarm regarding these issues. From what we have seen, many organizations are planning large-scale deployments of agentic AI without asking whether their existing data platform is able to handle the workload.

In this report, we describe how agents interact with the data; we provide realistic examples of agents, workloads and requirements; we recommend strategies for projecting the likely architectural requirements of, and workloads resulting from, implementing agentic AI; and, we recommend steps to take, so that agentic AI strategies will be successful.

Particularly when agentic AI will be implemented at large scale, we recommend a multi-step process of progressively more accurate projections of the data platform workload. The simplest first step is to use a workload projection tool to estimate how the workload may change and grow.³ Initial estimates available via such tools are typically based on broad industry experience and will not take into account the specific

¹ “IDC Highlights New Research Directions for 2026,” Morningstar, April 2026 [[IDC Highlights New AI Research at Directions 2026 on Economic Impact, Agentic Buyers and the Rise of AI Agents](#) | Morningstar]

² “Autonomous AI Agents: Adoption, Stats, ROI Insights,” Arcade, 2026 [[Autonomous AI Agents: Adoption Stats, ROI Insights](#)]

³ Such a tool may be available from your vendor; contact WinterCorp for up to date information on such tools.

Methodology

• • • • •

PURPOSE AND METHODOLOGY FOR THIS REPORT

This *WinterCorp Research*

Report is based on independent work by WinterCorp and its affiliates. In connection with this work, WinterCorp was briefed by Teradata, in its capacity as a sponsor, concerning its ongoing efforts to project agentic AI workloads and requirements. Teradata reviewed this report with respect to facts. However, the opinions and conclusions stated here are exclusively those of WinterCorp.

• • • • •

strategies and requirements of your organizations agentic AI program. They are therefore rough estimates, to be relied on only until it is practical to use the more accurate methods we discuss in *Section 4*.

Once you have an initial estimate, consider what it is telling you about the risks you face and the steps that may be necessary for success. If the risks look sizable — as we believe they often will be — consider then refining the estimate via more rigorous analysis and testing at the earliest possible time. In *Section 5* of this report, we describe the WinterCorp approach to such analysis and testing. Regardless, make sure that you have a platform that can handle the extraordinary new workload that agents will impose on your data layer — and that you have it in place before agentic AI solutions are in production at scale.

KEY RECOMMENDATION: On the basis of its independent research, WinterCorp recommends that enterprises implementing or considering agentic AI take all necessary steps to ensure that they have in place an **autonomous knowledge platform** able to: (a) Curate, manage and retrieve on demand the business context required for agents to exploit enterprise data; (b) Handle the large, demanding, unprecedented query workloads expected from deployment of intelligent agents at scale; and, (c) Manage its own operations with a high degree of autonomy.

Note that this recommendation calls for a specific level and type of scalability often not present in popular analytic data platforms — the ability to handle concurrently and efficiently very large numbers of short queries — while also efficiently processing other heavy workloads involving intensive computation, high volume update and complex queries and analytics. •

1 Agentic AI and the Data (Autonomous Knowledge) Platform

Intelligent agents powered by generative AI are expected to perform tasks autonomously — tasks that previously required skilled people. These agents will plan and reason toward an objective. Just as a self-driving car can navigate a street network, traveling around obstacles that were not present when the trip began, so intelligent agents are expected to work out their path to a goal without having been explicitly programmed for each intermediate decision. As agents repeat tasks, they may improve processes, increase efficiency, reduce time to value, or advance other predefined objectives.

1.1 THE BRINK OF THE AGENTIC ERA

Agentic AI promises remarkable new benefits to the organization. Many workers will be able to turn routine tasks over to such agents and thereby spend more of their time and energy on higher value tasks. A large return on the investment required is anticipated, as indicated in the IDC report previously cited.

Beyond that, agentic AI has the potential to dramatically transform business processes and change the dynamics of entire industries. Just as robots are transforming manufacturing and drones are revolutionizing agriculture, autonomous agents will bring large, structural changes to many organizations: in financial services, in healthcare and in most other sectors of the economy.

1.2 PERSONAL AI VS ORGANIZATIONAL AI

While most companies are using or experimenting with AI, few have realized the returns they expect from their investment.

Personal AI. Why? Because most projects have focused on individuals employing AI in a localized work process. When a software engineer uses an AI to generate code for a business application, many elements of the overall work process are unchanged. Human engineers must still specify the solution; evaluate the quality of the generated code; revise; test; integrate; then manage the deployment into production; etc. Success in such efforts may increase productivity for a particular set of employees, but it does not typically result in large-scale returns or transformational changes in companies or industries.

Organizational AI. The larger returns will come from organizational AI: major changes to the way larger business processes are structured or organized, empowered by the use of enterprise data.

An example is congestive heart failure (CHF) monitoring. A typical large health system manages 20,000 or more outpatients with congestive heart failure.

Prior to the use of AI, nurses checked on a subset of patients with the highest risk factors daily and responded when other patients were brought to their attention.

Now, an AI agent **continuously** monitors all 20,000 patients for:

- Lab results
- Data from remote monitoring devices
- Medication compliance
- Notes/findings from outpatient visits to doctors or other hospital departments
- Recent admissions and emergency visits.

When the AI detects a potential deterioration, it acts autonomously. The AI notices such conditions as:

- Patient weight up to 5 pounds in 4 days
- Rising blood pressure
- Medical professional noting a heartbeat irregularity in the patient record
- Missed diuretic refill
- Indigestion or arm pain — symptoms that might not otherwise be cause for concern but that could indicate a developing cardiac event
- Increased shortness of breath reported through a patient portal.

The AI then:

- Identifies the patient as potentially elevated risk.
- Reviews the patient's history.

- Generates a clinical risk assessment.
- Creates a summary explaining why intervention may be required.
- If indicated, schedules an urgent telehealth appointment in an available slot.
- Drafts outreach messages to the patient.
- Prepares recommendations for the care team.

The AI has done most of the work; the human supervises. The nurse receives:

“High-risk alert generated. Telehealth visit scheduled tomorrow at 9:00 AM. Recommended review of fluid-retention status and medication adherence.”

The nurse can approve, modify, or cancel those actions.

With the AI agent, the level of care is significantly elevated because agents poll the entire set of patients continuously for alerts; they monitor for a much broader array of indicators; and react much more rapidly than nurses otherwise could, given their typical workload. Now, nurses can focus their efforts on those patients referred to them by the AI because they warrant special attention. Expert nursing effort is thus focused more effectively; all patients get more attention when it would make a difference; outcomes are better; and there is a prospect that lives will be saved as a result.

1.3 AGENTS AND THE AUTONOMOUS KNOWLEDGE PLATFORM

While the promise of agentic AI is indeed extraordinary, there will be challenges. Some of those challenges are widely discussed elsewhere — governance, ethics, organizational change management, data quality and how to actually realize the target business outcomes. These issues are indeed critical.

But there is another challenge that has received less attention: the remarkable new and different demands on the platform that stores and manages data for analytics and AI. The platform itself must now evolve into something new: an **autonomous knowledge platform**. As shown in *Table 1*, this new type of platform must take on several new capabilities that have not been expected of the data platforms in use over the last decade.

HOW A KNOWLEDGE PLATFORM IS DIFFERENT		
USAGE MODE	BUSINESS INTELLIGENCE	AGENTIC ARTIFICIAL INTELLIGENCE (AI)
CAPABILITY REQUIRED	ANALYTIC DATA PLATFORM (Pre-2026)	AUTONOMOUS KNOWLEDGE PLATFORM (2026+)
BI Query & Reporting	YES	YES
Classic ML	YES	YES
Business Context	NO	YES
Agentic Workloads at Scale	NO	YES*
Vector Datatype & Indexing	NO	YES
Semantic Search	NO	YES
Natural Language Query at Scale	NO	YES
Autonomous Operation	NO	YES

*Requires rapid response to very large numbers of concurrent short queries.

Table 1: The Analytic Data Platform Must Meet Demanding New Requirements to Enable Agentic AI

So, the autonomous knowledge platform must store and manage the business context that makes it possible for an AI to link a business question to the data that can answer it. The autonomous knowledge platform must handle agentic AI workloads at large scale. It must support vector embedding, vector indexing and semantic (vector) search. And it must handle natural language query workloads — typically feeding in through agents — at scale as well.

AGENTIC AI WORKLOADS AT LARGE SCALE are an urgent issue for many organizations — because they are potentially massive and they are squarely on the critical path to success for most agentic programs: **The impact of agentic AI on the data platform is a change that is unprecedented in the history of digital data and analytics.** •

When agentic AI is fully implemented in larger and more data-intensive enterprises, the resulting workloads will be much greater — and much different — than anything that has ever been run on these analytic data platforms before. In August of 2025 I published a research note warning about this issue, with a title based on a line from the 1975 movie, *Jaws*. The title of that note is “You’re Gonna Need a Bigger Boat.”



Figure 1: You’re Gonna Need a Bigger Boat

IN THE MOVIE, a shark has been attacking and killing people in the waters around a small town on the New England seacoast. Though the shark becomes a **large problem**, the people involved believe they are dealing with the type of shark they are familiar with, perhaps a 12-foot hammerhead shark. Such an animal can pose a lethal threat to people swimming nearby and one would have good reason to fear it.

But, when the shark finally reveals itself to the town sheriff and the movie audience, we are all **shocked by its enormous size and ferocity**. It is a 35-foot great white shark — far larger and more ferocious than anyone in the movie imagined — and previously unknown in the waters of New England. The town sheriff, backing away from the sight in shock, turns to the captain and says, “You’re gonna need a bigger boat.” In other words, **“We weren’t thinking big enough.”** •

That's my metaphor for AI and the data platform: **No matter what platform you are using now for data and analytics, it is up for a big challenge as the use of agentic AI scales up in your company. And these are not the workloads called "heavy" in the past. Like the shark in *Jaws*, they are *much larger, different, and more challenging to manage than what has been seen in the past.***

In the new workloads, queries are much more frequent and most are very short. Many of these workloads will have an impact on the data platform that is well outside of previous experience. Most organizations are not set up for these workloads because they are so much larger and different from what they have faced in the past.

Most organizations are "gonna need a bigger boat" — and many will need to introduce a different platform capable of processing large numbers of short queries efficiently — at the same time that more complex queries and analytics are processed on the same copy of the data, which may be changing rapidly as a result of time sensitive business-critical updates.

This report is about the impact of agentic AI on the data platform and how to prepare for it with realistic measures that actually work.

1.4 THE CHALLENGE OF PREPARING FOR AGENTIC AI

So, you or someone reporting to you is responsible for your organization's data platform.

Chances are, the top executives in your company are anxious to see agentic AI implemented and delivering value for the business. If the company doesn't move fast enough with AI, competition will leave it behind. In most companies, large-scale AI strategies are in the works already.

From this report, you know that agentic AI has the potential to create new, unprecedented workloads for the data platform. You might be thinking: "Well, we'll cross that bridge when we come to it." Maybe there will be no problem at your company. Best not to go looking for trouble.

Unfortunately, "We'll cross that bridge when we come to it" is a recipe for deep trouble. That approach works when what you need is a moderate adjustment to configuration or budget. When the potential scale of the change is much larger — when the nature of the workload is dramatically different — it becomes impractical to react effectively at the moment the problem arises. You must prepare well in advance of the crisis.

By analogy, when you get an alert that a great white shark has been spotted nearby, you don't keep swimming until you can see the shark: you get out of the water right away.

But here's the problem with agentic AI: What, exactly, can you do? There is little production experience with large scale implementation. Your data platform faces different, new and larger workloads.... How can you prepare when you don't know how large they are going to be?

The next few sections answer that question. We'll go through some concrete examples of agent implementation; we'll cover how agents progress through levels of maturity and why that matters; we'll describe how agents look to the data platform; how the workloads are different and how they grow; we'll discuss what you need in a data platform; how to estimate the size and nature of the workload; and, how to approach the decisions to be made.

2 Agentic Maturity

The benefits of AI are not going to arrive all at once. Most organizations will progress through four levels of maturity on their way to realizing a robust version of agentic intelligence. At each level, more is expected from the agents — there are correspondingly larger changes to human behavior and business processes — and there are likewise larger benefits delivered to the organization.

2.1 LEVELS OF MATURITY

A **maturity model**, detailed by Wakefield Research⁴, characterizes how customer deployment of agentic AI is likely to mature over time. The model is shown in *Figure 2*:

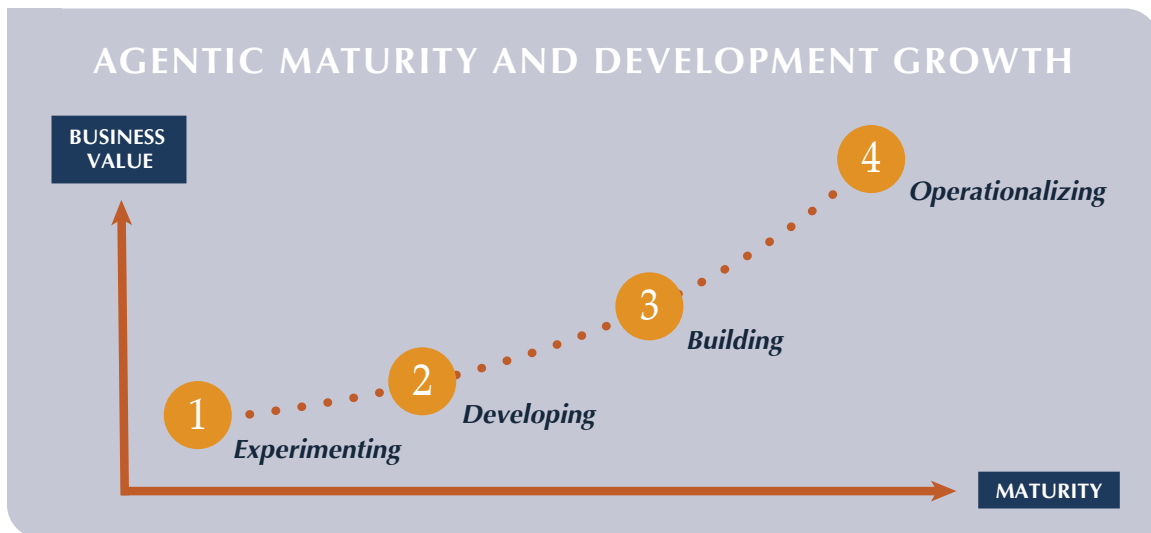


Figure 2: The Four Levels of Agentic Maturity

Customers will pass through four levels of maturity as they deploy agents in progressively more advanced and comprehensive ways. The business value they provide grows exponentially as they mature: slowly at first, but with increasing speed as they progress.

For the sake of guiding customers who do not yet have specific deployment timetables, each level of maturity is assumed to occupy a calendar year (customers can adjust this interval once they have specific plans in place). The higher levels deliver the largest benefits but require the largest changes in organization, business process, culture, governance and human behavior. To control downside risk, most enterprises will “walk before they run” as they work their way to full scale use of autonomous agents...meaning, they will start with *Level 1*.

Each level of maturity involves the use of agents on a larger scale and in a more autonomous manner. At each level, agents do more work; they provide more benefit; and they interact more with enterprise data on the data platform. The four levels are:

1. **Experimenting:** Organizations here are exploring localized pilot projects. They are learning how AI helps individuals work better and are beginning to map out broader data strategies;
2. **Developing:** These companies have successful models but no shared mechanisms to assemble the right context for the right agent on behalf of the right business or work process. Each team has its own definitions, its own data, its own rules. Navigating this personal-to-organizational transition is the defining challenge at this level;

⁴ “Arrested Automation: Why Agentic AI Stalls at the Enterprise Level,” Wakefield Research [www.wakefieldresearch.com]

3. **Building:** These enterprises have structured local governance and automated basic workflows, but the data foundation underneath hasn't been built for organizational scale. The barrier here isn't frameworks – it's whether the data layer can support agents operating across functions, not just within them;
4. **Operationalizing:** These organizations have made deliberate architectural investments in harmonizing significant portions of their data, established dynamic safety rules, and are starting to see AI confidently execute multi-step workflows. They are actively developing *Autonomous Knowledge*: data with enough context, lineage and governance for agents to act on it reliably. Early results show measurable impact but extending that foundation across the full enterprise is ongoing work.⁵

In its recent survey of 1,000 enterprise executives, Wakefield found that 68% of the market is presently in Level 1 or 2. Only 7% of organizations surveyed are in Level 4.

An example of an agent operating at Level 4 is the management of equipment and maintenance in a semiconductor manufacturing plant. Every day there are:

- Equipment failures;
- Yield fluctuations;
- Rush orders;
- Material shortages;
- Energy management constraints; and
- Maintenance windows.

Historically, these decisions required teams of planners and production controllers. A *Level 4* multi-agent AI system instead:

- Monitors equipment telemetry;
- Tracks production queues;
- Predicts bottlenecks;
- Predicts probable equipment failures;
- Recalculates optimal production schedules;
- Reassigns lots to alternate equipment;
- Reschedules maintenance; and
- Updates ERP, MES, and workforce systems automatically.

Humans are notified but do not approve routine decisions.

Eventually, we expect to see agents not only operating a business process autonomously, but continuously discovering new opportunities and modifying the process to better achieve business objectives.

For example, a network of agents is implemented to manage a supply chain for a manufacturer. The agentic network might discover:

- A growing demand trend in Southeast Asia for *Product Family A*;
- Increasing shipping delays through a major port used by *Supplier X*; and/or
- Improving economics of an alternate *Supplier Y*.

The system then develops and evaluates responses. First, the system defines an opportunity:

“Demand for Product Family A is likely to grow 23% over the next 18 months while Supplier X faces increasing disruption risk.”

⁵ Definitions are from Wakefield Research.

Next, the system identifies options:

- Increasing inventory;
- Add a second supplier;
- Expand factory capacity in Mexico; and/or
- Shift production to a different region.

Then, the system simulates outcomes. Agent teams perform:

- Cost modeling;
- Revenue modeling;
- Risk analysis;
- Sustainability analysis; and
- Workforce impact analysis.

The agents negotiate among themselves to find the optimal balance. Then, they execute a plan. Within predefined governance limits, the agents could:

- Initiate supplier qualification;
- Negotiate contracts through procurement agents; and/or
- Realign logistics networks

without waiting for executives to identify the problem or approve the solution.

Clearly, agents would need to be very thoroughly tested before they would be authorized to operate like this. Presumably, the organization would test this level of autonomy first with lower impact processes and gradually progress to higher impact processes.

Nonetheless, it is likely that we will see — in five years or less — autonomous AI agents realigning supply chains to avoid disruptions and/or anticipate changes in demand.

3 How Agentic AI Looks to the Data Platform

As the customer organization moves to higher levels of agentic maturity, more work is shifted from humans to agents. Much of this work involves the use of data and analytics, and therefore this shift affects the workload of the analytic data platform.

Understanding this change requires examining how agents use enterprise data. They are quite different from humans.

3.1 ILLUSTRATION: ONE BUSINESS QUESTION

When a human has a business question, and when no AI agent is involved, the human submits a database query. For example, a store manager may ask, “How did sales in our store yesterday compare to sales in the same day last week?” Since this question comes up regularly in the manager’s daily work, they probably have arranged for a skilled business analyst to create and store a report or dashboard that answers this question. The manager clicks an icon on their phone or desktop; an SQL query goes to the database; and the answer comes back.

However, when an AI agent asks the same question, the process is very different. An autonomous agent typically generates more interactions with enterprise data than a traditional query application or a simple chatbot, because it often operates in an iterative loop rather than a single request/response cycle. Thus, research on autonomous-agent architectures shows that agents typically operate through iterative reasoning and tool-use cycles involving multiple retrieval and validation steps rather than a single request-response interaction. As agents become more autonomous and perform more complex planning, execution, validation and coordination tasks, the number of tool invocations and data accesses per business question increases substantially.⁶

For example, when our store manager asks, “How did yesterday’s sales compare to sales in the same day last week?” an autonomous agent may:

1. Retrieve business context and metadata;
2. Search for relevant documents and prior analyses;
3. Query analytical data stores;
4. Retrieve historical memory or prior agent results;
5. Evaluate whether the results are sufficient;
6. Issue follow-up queries to fill gaps;
7. Perform calculations or analytics;
8. Validate findings; and/or
9. Generate a final conclusion.

This “reason $\Rightarrow$ act $\Rightarrow$ observe” cycle is a common agent pattern. One often sees agents alternating between reasoning and executing tool calls, where each tool call may access a database or other enterprise data source. In addition, agents can reformulate searches, execute additional queries, and self-correct based on intermediate results.

The key point is that, for one business question:

- **Traditional BI query:** often one SQL query, one result.
- **Autonomous agent:** potentially many retrievals, queries, tool invocations, validations, and retries before producing an answer.

⁶ [ReAct Agents: Reasoning, Acting, and Tool Calls | NVIDIA DLI](#)

3.2 HOW QUERY COUNT GROWS

As an organization progresses to higher levels of agentic maturity, agent behavior becomes more complex. A Level 1 agent may generate tens of data platform interactions for one business question. Additionally, the number of data-platform interactions associated with a single business question can increase dramatically. Advanced agents often employ iterative planning, tool-use, reflection, self-correction, verification, and multi-agent collaboration. Consequently, a task that might require only a few database accesses in a traditional application can generate dozens to hundreds of tool and data interactions when executed by a sophisticated autonomous agent.^{7,8,9,10}

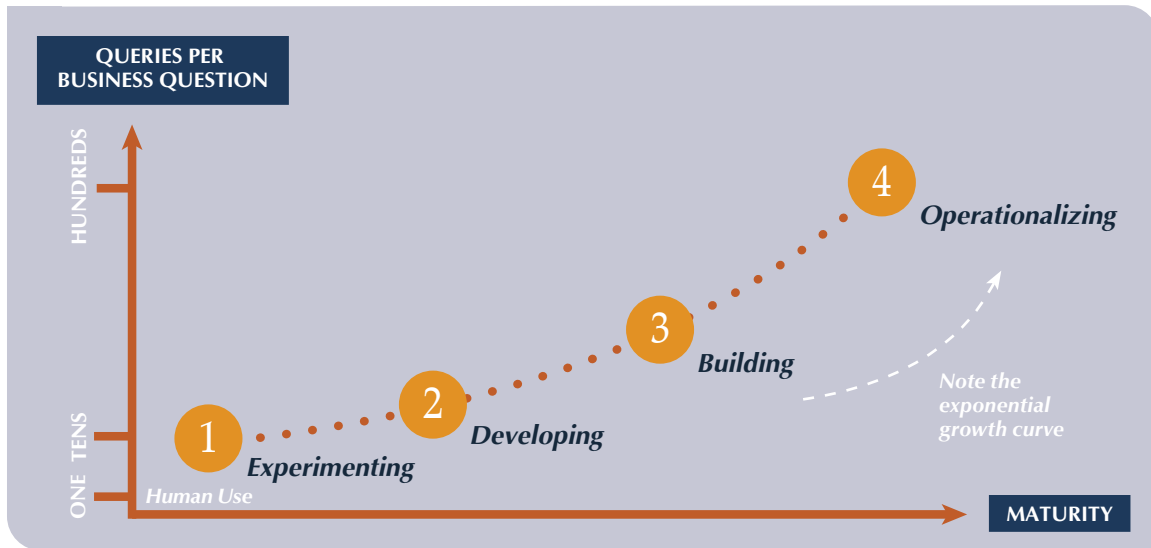


Figure 3: Database Query Count Grows Exponentially as Agents Mature

3.3 THE WORKLOAD SHIFT

There are three major implications from this pattern:

1. A large increase in the number of database queries;
2. Most of those additional queries are short probes retrieving a single piece of business context; a single piece of metadata; or a small number of data values from a single column or data object; and
3. The number of queries grows rapidly as the customer organization progresses through the maturity levels.

Therefore, as work is shifted from humans to AI agents we have two large impacts on the workload of the analytic data platform as the number of agents increases and the level of maturity of agents grows over time:

- A large increase in the number of database queries, estimated to start at 20x and grow rapidly; and
- A large shift, so that a large majority of all queries are short queries, retrieving a single value or small number of values.

The resulting workload is much larger and substantially different from the workloads those data platforms have handled previously.

⁷ Yao, S. et al. (2023). *ReAct: Synergizing Reasoning and Acting in Language Models*. International Conference on Learning Representations (ICLR).

⁸ Xu, H. et al. (2026). *The Evolution of Tool Use in LLM Agents: From Single-Tool Call to Multi-Tool Orchestration*. arXiv:2603.22862

⁹ Kalappurackal Mansoor, I., Phadke, A., & Rana, P. (2026). *Verified Tool Calls Improve LLM Agent Reliability Under Non-Atomic Failures*. arXiv.

¹⁰ Garcia, J. & Malarme, P. (2024). *Baseline Agentic AI Systems Architecture*. Microsoft Azure AI Foundry Blog

3.4 MATURITY MODEL AND WORKLOAD GROWTH

As indicated earlier in this Section, as the agentic AI program matures, progressing through *Levels 1 through 4*, the data platform workload continues to grow, and the query mix tilts ever further in the direction of short queries. Even if the number of business questions were to remain constant, the number of database queries would increase at each level of maturity. But it is likely that both will grow: the volume of business activity handled by agents will increase at each level of maturity; in addition, the number of database queries per business question will also increase.

The net effect is rapid growth in the query count and a continued rapid shift in the workload toward an ever-larger percentage of short queries. It will still be critically important for the platform to handle long queries efficiently. However, the count of short queries will grow much faster than the count of long queries.

4 The WinterCorp Perspective

WinterCorp brings a particular perspective to the subject of AI and the data platform. This results from our thirty years of vendor-independent work focused on analytic data management at large scale. We specialize in helping customers who face challenging requirements with respect to scale: we have particularly focused on quantitative analysis of architectures; testing and measurement; cost modeling and cost management; and related matters. This has resulted in extensive experience projecting, quantifying and analyzing analytic data workloads, including the workloads that analytic data platforms must handle under the most demanding circumstances.

We view database scale challenges from a multi-dimensional perspective. That is, we consider database size, query/update volume, database structure, query/update complexity, query concurrency, data latency, compute intensity, service level requirements and other factors. Fundamentally, any engineering factor that challenges the limits of data platform performance, management or operation can play a role in the analysis and testing we do.

In this section we address the workload challenges of agentic AI from that perspective.

4.1 THE BEST FIRST STEP: USE A WORKLOAD PROJECTION CALCULATOR

At WinterCorp, we have a rigorous process for quantifying business and architectural requirements once the AI strategy is well-defined. Applied with skill, our approach produces consistently excellent results for customers. We describe that approach briefly in the following sections.

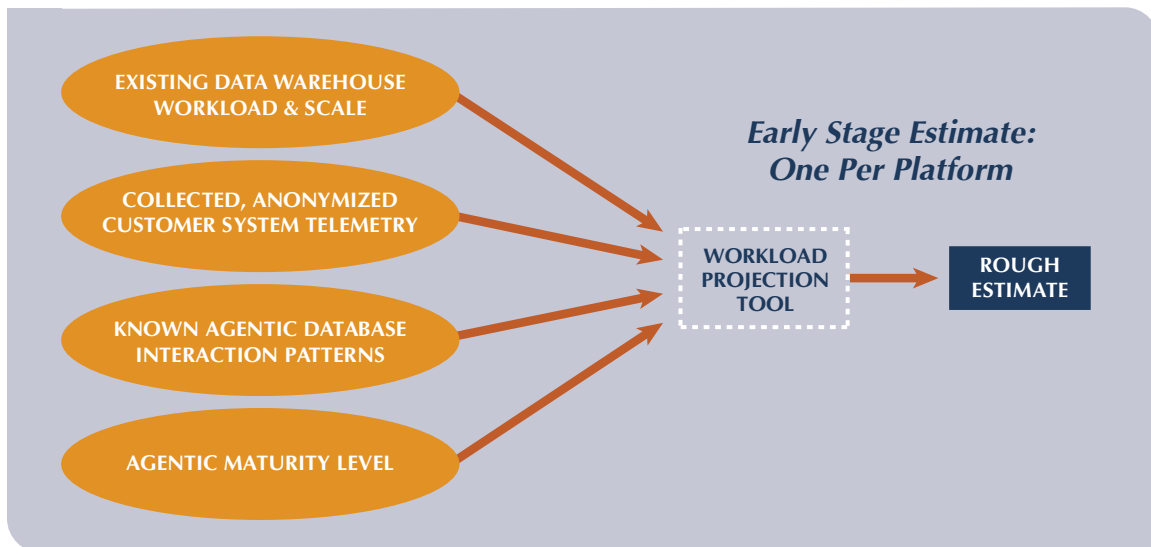


Figure 4: Early-Stage Workload Estimator: Projection Tool

Often one must begin preparing for larger workloads, when the organization's specific AI strategies are not yet clearly defined. We see situations where large investments are being made in AI, but what is going to be implemented — and at what scale — is not yet known.

In that situation, it helps to use a workload projection calculator that is based on general industry experience. Such a calculator can take into account such factors as:

- The scale of your current data warehouse/lakehouse workload;
- The maturity level of your agentic AI program;
- The structure of AI agents at each level of maturity and the database query patterns that result; and
- Other top-level factors you may factor in based on your knowledge of the situation.

Such a calculator provides a starting point for planning and budgeting, to be confirmed as you progress, either with testing, pilot programs or small-scale production. You might check with your vendor whether such an estimate can be provided, bearing in mind that it is only a rough estimate.

4.2 WORK FROM THE TARGET BUSINESS OUTCOME

So, a workload projection tool has provided you with an initial rough estimate. That is helpful on an interim basis, but it does not take into account the specific strategies that will be pursued in your organization.

As soon as your organization's AI strategies are reasonably clear, you want to work toward estimates that are actually grounded in their specific requirements.

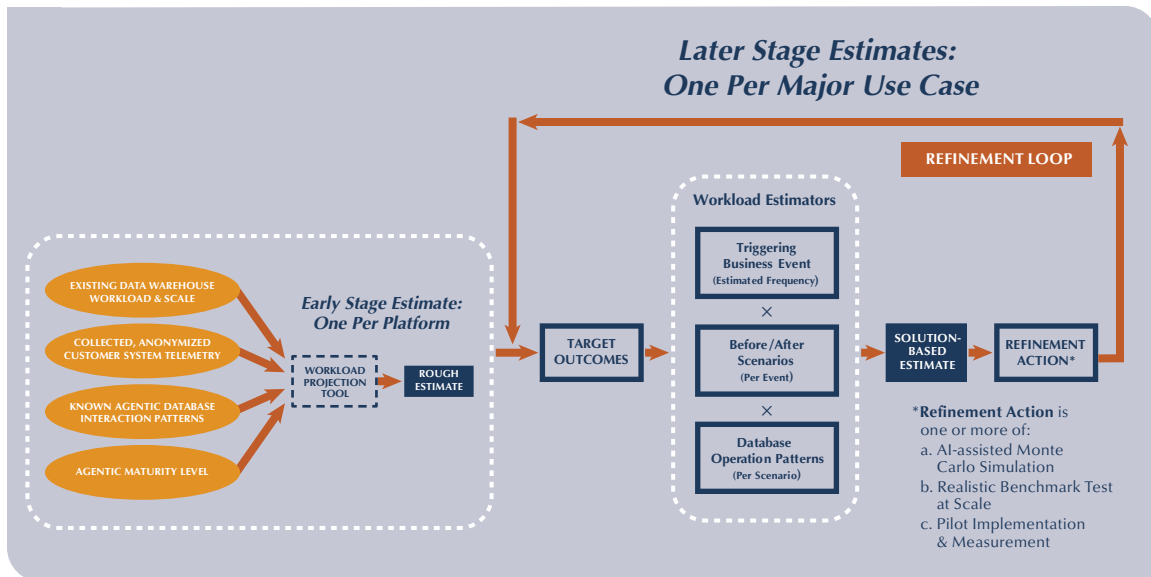


Figure 5: Later Stage Estimate for Each Major Use Case: WinterCorp Method

The best approach is:

1. Define the target business outcomes
2. Identify the key stakeholder(s) for each target outcome
3. With the involvement of stakeholders, define before/after scenarios
 - » How the business process works today (Before)
 - » How the business process will work with AI (After)
4. Determine the expected frequency of each major scenario

What's an example of this? Let's use the example of agentic AI in *Section 1.1*. The healthcare organization has 20,000 patients with congestive heart failure. The AI agent will be responsible for monitoring them for each of several types of events. When an event fitting the profile occurs, the agent is responsible for taking certain actions.

By working with the nurses who perform this monitoring function now — and by analyzing the data currently available on the frequency of various events in the patient population, you can estimate: (a) the frequency with which the agents will run their monitoring queries in the new continuous-monitoring regime; (b) the frequency with which they will need to take action; and (c) the nature of the questions to be answered in each instance. For this case, you have then defined the metrics of the business requirements. For example, such an analysis might show that: (a) the status of each of the 20,000 patients is checked against available data sources 10 times a day; (b) risk assessments are expected to identify 50 patients a day who should be seen by a cardiac nurse within a week; and (c) risk assessments are expected to identify

five patients a day who should be seen within 48 hours by a cardiologist. Given these inputs from the medical staff, and given a concept for how the agents will perform their assessments, it will be possible to estimate the query pattern of the agent(s) and thereby generate an estimator of the workload that can be fed into a model or benchmark.

We typically repeat this process for each use case that we believe could have major impact on the architectural requirements. The result of this step is a set of major use cases; one or more business scenarios for each use case; and a set of estimated frequencies for each scenario.

4.3 DERIVE QUANTIFIED ARCHITECTURAL REQUIREMENTS

A recurring error, even in major enterprises, is defining architectural requirements without quantifying them. It is quite common to define the desired data architecture in strictly logical or qualitative terms.

Because the typical WinterCorp engagement entails an engineering challenge, we always seek to quantify — at least roughly — our architectural requirements. This clarifies matters both for the client and for any vendors making proposals.

We think of quantified architectural requirements as typically including:

- Metrics of database scale;
- Macro structure of the database (size & row count of the most important tables; major data relationships among them). Usually, we find there are between five and 25 tables that are most important to the architecture, operation and management of the database; among these are usually the largest, most intensively used and/or fastest changing tables.
- Macro structure of the key workloads; develop these from the query patterns of the most important business scenarios; and
- Required service levels.

Early in the process it can be impossible to know such values precisely. Our approach is to estimate them as is practical; and to be realistic about the uncertainty.

So, if the daily volume of query type *X* may vary by a factor of three, depending on inputs whose value won't be known for another six months, then use that information and make allowances when appropriate for both highest likely and lowest likely values.

4.4 PROJECT WORKLOADS, THEN MODEL AND TEST TO VALIDATE

With:

1. a set of key business scenarios and metrics; and
2. quantified architectural requirements,

you have the inputs to a model and the ability to construct a benchmark test.

You can then:

1. Derive the query pattern for an occurrence of each scenario;
2. Use items 1 and 2 to generate database workload projections;
3. Create realistic simulations of the database and the workload that will exist at full scale production;
4. Run simulations or realistic benchmarks to confirm that your data platform can handle the workload and generate estimates of the configuration needed.
5. If your current data platform/architecture cannot process the workload with satisfactory performance and cost, evaluate options for changing the data architecture or modifying the target business outcomes.

At WinterCorp, we employ this approach and/or modifications of it, when customers retain us to advise them on major new strategic initiatives. We find that the keys to success are accurately understanding the

business drivers (Section 5.2); correctly generating the workload projections (Steps 1-2); making the benchmark fully realistic; and then thoughtfully analyzing what the benchmark tells you about platform and architecture.

4.5 NATURAL LANGUAGE QUERY (NLQ)

We know from the factors discussed in Section 4 that there will be a large impact on data platform workloads from the implementation of agentic AI. The effects we have discussed so far are a direct result of the structure of AI agents: how they are designed and how they go about the business of interacting with enterprise data for analytic purposes.

Agentic AI and NLQ. But there is an aspect of agentic AI that will affect the analytic data platform in yet another way: *natural language query*. The natural language queries will typically flow through AI agents, thus becoming part of the agentic query workload. As companies make natural language query available to large user populations, we expect the flood of business questions that result will often create a different set of unprecedented demands on the data platform (that is, different from the short queries generated by the agents). We believe that the availability of NLQ will stimulate many more business questions than have actually been submitted to the data platform in the past.

In many organizations, we have business strategists planning large AI programs without taking into account how they may affect the data architecture. Data architects who could address the problem are often excluded from planning. This may be in part because those driving the program are pre-occupied with upstream problems such as token economics, governance or organizational change management.

Why do we expect NLQ to have an additional large impact? Because we believe there is often a large pent-up demand for answers to business questions. In many organizations today, end users rely primarily on predefined queries, reports and dashboards that have been created for them. Typically, these answer only the most predictable questions. While in theory many end users can go to a designated business analyst or IT specialist for ad hoc queries and analytics, in practice many do so only occasionally. Therefore, a large latent demand for business data and analytics will be unleashed in many organizations when natural language query is made available to large end user populations.

We expect that the volume and variety of end user queries will far exceed anything most data platforms have received before. We expect the number of business questions asked every day, as a legitimate aspect of performing daily business functions, will be much larger than current query volumes. In general, we apply the same methods to the projection of workloads from NLQ as we have described in the preceding sections, with this difference: the key issue is estimating the frequency of business questions likely to be generated by a given population of end users. This is usually subject to considerable uncertainty. We recommend simulating and/or testing for a range of frequencies (lowest likely, highest likely and most likely) and keeping architectural options open until some experience is gained.

We have conducted private simulations for a number of businesses in an effort to anticipate and plan for programs to equip a large population of users with NLQ. The resulting query volumes are at times startling — and much larger than previously expected. We recommend planning with care; rolling out such programs incrementally, with careful management; and recognizing that some populations are likely to have a large pent-up demand for better, more timely answers to practical business questions.

5 What Your Data Layer Technology Needs to Deliver

So, most organizations are planning on widespread use of agentic AI to increase productivity, create new opportunities and/or transform the business. In some cases, this will result in dramatic structural changes that will greatly change economics or business outcomes for an entire industry. As our discussion demonstrates, this is likely to substantially increase the data platform workload and stress it in ways it has not been stressed before. This report has discussed the nature of these coming changes; techniques for estimating their impact; and other aspects of preparation.

Why are these changes required? Because agentic AI really is changing the way organizations operate. For many organizations, strategies and architectures that were fine for the last five years are not going to work well — or are not going to deliver reasonable economics — once the organizations are making significant use of agentic AI. This section of our report wraps up with a summary of the critical platform requirements for the coming era of agentic AI.

5.1 SCALABILITY

Database scalability is no longer a feature. With agentic AI, a specific type of scalability is a requirement.

Once substantial numbers of agents are active in your company, using enterprise data, your data platform workload will be drastically different. Query concurrency will be much higher. Short queries — sometimes called ‘tactical’ and ‘sub-tactical’ queries — will dominate. Recalling our analysis in *Section 4: at Level 1*, an intelligent agent generates tens of short queries for each business question; at *Level 4*, the typical agent generates hundreds of such short queries. And these require fast response. With hundreds or thousands of agents active in the system at any one moment, the deluge of short queries is remarkably unlike the workloads that most analytic data platforms have handled before.

And it matters which platform you select. In private benchmarks conducted on behalf of enterprise clients, WinterCorp has seen and produced independent benchmark results demonstrating that some widely used data platforms are hundreds of times more efficient than others when handling this class of queries. That efficiency will no longer be optional. Similarly, some platforms are far more cost effective than others when handling large numbers of concurrent queries. That ability to manage high concurrency will no longer be optional.

Finally, there is the “bigger boat” issue: What will happen when your company empowers a large user population (employees, channel partners, suppliers, customers) with natural language query (NLQ)? It is likely that your data platform will start seeing a far greater query volume and variety...large numbers of queries of relatively high complexity and variety. Some widely used data platforms are not designed for this.

BOTTOM LINE: Many organizations will need to reassess and consider bringing in a platform with more scalability, able to handle the heavy, unprecedented workloads of agentic AI. Those that do not will face the likelihood that their large investment in agentic AI will fail. •

Soon, we won’t be calling them data platforms, because we will all need an autonomous knowledge platform. In this report, we have used the term “data platform” at times because readers are familiar with it. But, with agentic AI, what organizations really need is an *autonomous knowledge platform*. **An autonomous knowledge platform contains business context in addition to the data and the schema.** The schema is a technical description of the data: it gives each data object a technical name; a data type; and certain other information. But it typically does not tell you which combination of tables and columns answers a given business question or corresponds to a given business concept.

So, when the head of sales asks for a projection of commission costs for the coming quarter, how does an AI know which sales projection to look at? How does it know which commission plan to apply? How does it know how many new salespeople will start earning commission that quarter? When it comes to answering real business questions in real organizations, there are many subtleties to the data and many

complications in how the data relates to the business processes. Skilled people work out these questions and come up with good answers. For AI to do that, it needs access to organized and curated business knowledge. Autonomous knowledge platforms will manage that knowledge; facilitate the processes of retention and curation; and maintain the relationships between the business and the data. They will do all that in addition to meeting the demanding new performance and scale requirements of agentic AI.

Finally, we assert in this report that the knowledge platform must be largely or entirely autonomous in its operation. By this, we mean that it is adaptable, scalable and self-optimizing. It must not only optimize individual queries but also workloads: sets of database operations that must be performed at once or rapid succession. With large numbers of autonomous agents — each acting on its own tasks and goals — workloads will grow, shrink and shift rapidly — and there will be a large number of them. Only a system which can handle such demands with little or no human intervention can possibly keep up with such demands. And, only a system with efficient ways of handling diverse workloads at once will be successful.

6 Summary of Platform Architectural Requirements

In summary, an autonomous knowledge platform to manage the context and data for agentic AI must meet the following requirements:

- Store, curate and efficiently access business context that links business questions to relevant data of all forms
- Store business context in at least two forms as native data types:
 - » Knowledge graph(s) precisely defining the relationships between words, concepts and data objects;
 - » Vectors expressing less precisely the similarities between concepts (e.g., dogs and cats are both animals, but cats are more similar to bobcats than dogs are);
- Process enormous volumes of short queries with rapid response time and at near-zero cost;
- Process large numbers of queries **concurrently** (e.g., hundreds to thousands of short, concurrent queries at once);
- Autonomously and efficiently manage dynamic, mixed workloads in which both short queries and long queries occur commonly at the same time; such a system must scale up and down as needed; self-optimize; and self-manage;
- If NLQ is offered to large populations of users (e.g., thousands or more), then the platform must also be able to concurrently process queries of considerable complexity, as real business questions often join data from multiple sources in complex ways; and
- Handle large volumes of online updates concurrently with large query workloads.

These requirements are inherent in the agentic AI strategies being pursued in most medium-to-large data-intensive organizations. Virtually all AI strategists in such organizations picture hundreds to thousands, if not more, of continuously active agents at work throughout the workday if not around the clock. Further, many AI strategists assume that the data available to AI agents is business real time data — updated in the last minute, if not the last second. Why? Because the agents are often making decisions with operational impact. Such agents — at least as currently understood with today's technology — would generate very large numbers of database interactions on a frequent basis.

Customers anticipating such strategies thus must assess their data architecture for its ability to process the expected workloads with reasonable response time and at reasonable cost.

KEY RECOMMENDATION: So, yes, it is going to be essential to have an **efficient, scalable platform** able to handle a much larger volume of queries — with a **different query mix** in which many of them are short — and it is going to be essential that it be an **autonomous knowledge platform**. The key requirements for that platform are, in addition to handling today's analytic data workloads effectively:

- Perform very short queries with rapid response at low cost;
- Scale easily to handle such queries in extremely high volumes, while still processing complex queries and analytics at the same time on the same data — data which may be changing rapidly due to time sensitive, business critical updates;
- Handle very large numbers of such queries concurrently;
- Often, also handle many natural language queries from large, active populations of end users at the same time;
- Efficiently and autonomously manage many concurrent, rapidly changing, workloads; self-optimizing as needed;
- Manage, curate, and retrieve business context on demand, meeting similar requirements for concurrency, query volume and efficiency.

That is the highly scalable autonomous knowledge platform you'll need for the coming era of agentic AI. •

WinterCorp is an independent consulting firm expert in the strategy, architecture and scalability of the modern analytic knowledge ecosystem, focusing on enterprise AI at scale.

Since our founding in 1992, we have architected and engineered solutions to some of the toughest and most demanding analytic and AI challenges worldwide.

We help customers define their AI and knowledge-related business interests and vision; quantify their AI-related workloads and platform requirements; develop their knowledge strategies and architectures; select their knowledge platforms; and engineer their solutions to optimize business value and obtain otherwise infeasible business outcomes. We often help customers with the special architectural and engineering challenges that are involved in mission-critical systems, including those with the most extreme availability requirements.

Our customers get business results with analytics and AI, in which their return is often ten or more times their investment.

When needed, we create and conduct simulations, benchmarks, proofs of concept, pilot programs and system engineering studies that help our clients manage profound technical risks, control costs and reach business goals.

With our in-depth knowledge and experience, we deliver unmatched insight into the issues that impede scalability and into the technologies and practices that enable business success.



WinterCorp

www.wintercorp.com

42 DERBY LANE, TYNGSBORO, MA

617-695-1800