Prophecy

How AI agents are redefining data preparation and analysis

AI agents will collaborate with data teams to accelerate and automate tasks, bringing a step change in productivity

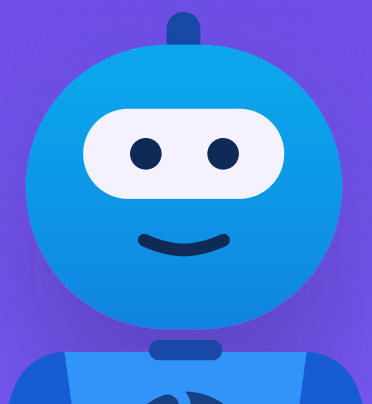
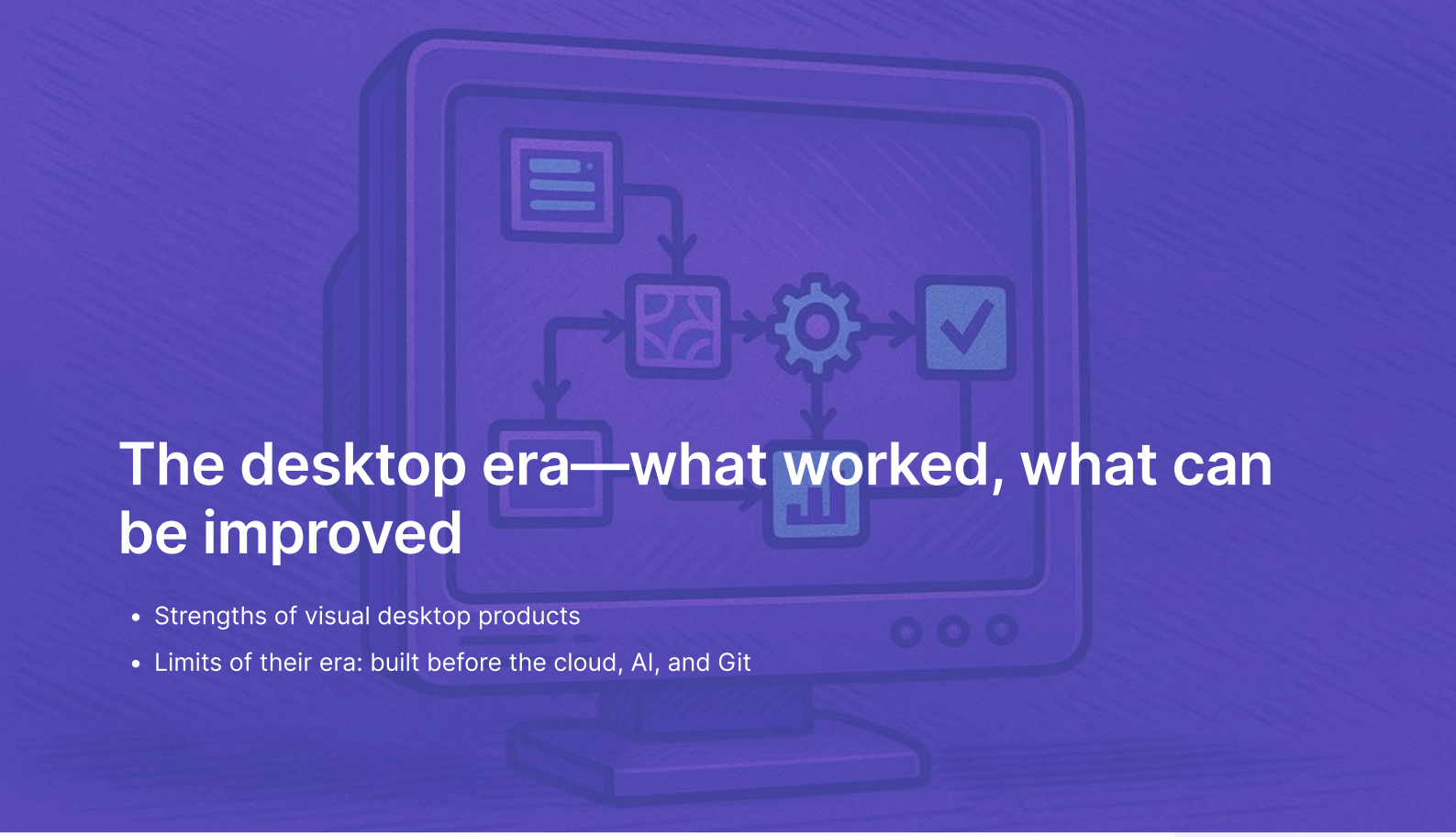


Table of Contents



The desktop era—what worked, what can be improved

- Strengths of visual desktop products
- Limits of their era: built before the cloud, AI, and Git

Strengths of visual desktop products

Desktop data preparation products (such as Alteryx) solved an important problem for business data teams: enabling them to prepare data for analysis without relying on others. Users consistently emphasize two qualities—ease and speed. As these products have became ubiquitous, they defined the table-stakes features for end-to-end data prep that users now expect.

Data input and output

Read from common sources such as files (Excel, CSVs), databases, SharePoint, and application APIs

Write to common targets such as Excel, Tableau / Power BI, databases, reports, and documents

Manipulate data

Prepare: filter, clean, sort, de-duplicate, and handle complex data types and operations

Blend and join: join, union, lookup

Transform: pivot, group, aggregate, and compute more complex derived fields

Analytics and reporting

Advanced and spatial analytics

Report-generation tools for PDFs, charts, and tables

Egress to files and tables with append and merge semantics

Automation and scheduling

Automate workflows through scheduling for recurring runs

Parameterize apps to create dynamic, reusable workflows

Limits of their era: built before Git, the cloud, and AI

Desktop data preparation predates cloud computing and Git-based version control for collaboration. Cloud data platforms—Databricks, Snowflake, and BigQuery—now provide processing at scale, and most enterprises use one or more of them.

This shift creates new opportunities for users in this era.

Data scale

Handling large datasets shouldn't mean hours-long or days-long desktop runs; pipelines can execute natively on cloud data platforms when scale is needed.

Path to production

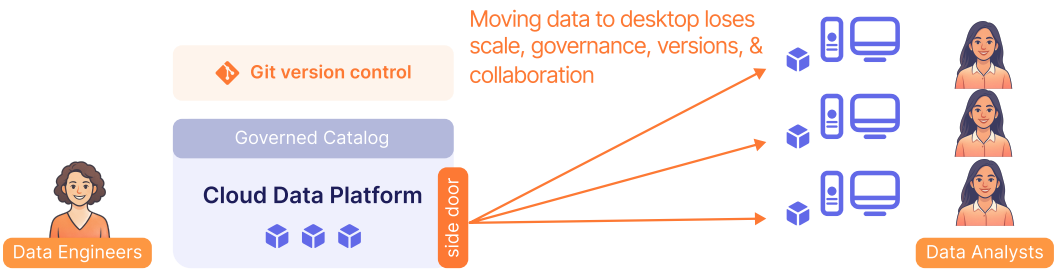
Data teams shouldn't have to rewrite data pipelines built by business teams to run on the cloud data platform; the same logic should promote to production.

Git

Data pipelines should take advantage of Git for version control and collaboration, using the robust development practices teams rely on.

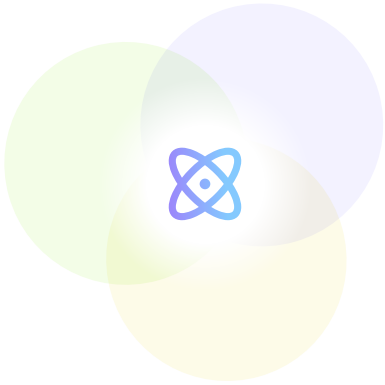
Governance

Copies of data on desktops create security risks. Access controls defined on the cloud data platform should be honored; you shouldn't have to recreate or bypass them elsewhere.



AI limitations for desktop data preparation

Desktop data-preparation tools are not AI-native and their fragmented architecture make it difficult to apply AI. Users of these tools will miss this step-change in productivity, as AI becomes a must-have.



The AI sweet spot

As a specialized form of code, data pipelines sit squarely in the sweet spot for LLM-powered agents. Agentic data prep needs:

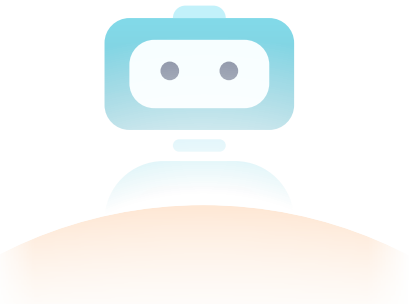
1. The understanding of data;
2. For every specialized task, a specialized agent that supplies the right context and knows how to iteratively get good results, and;
3. The delivery of results in visual interfaces, that enable business and data experts to refine and iterate, taking 80% results to 100%.

Adaptation or replacement

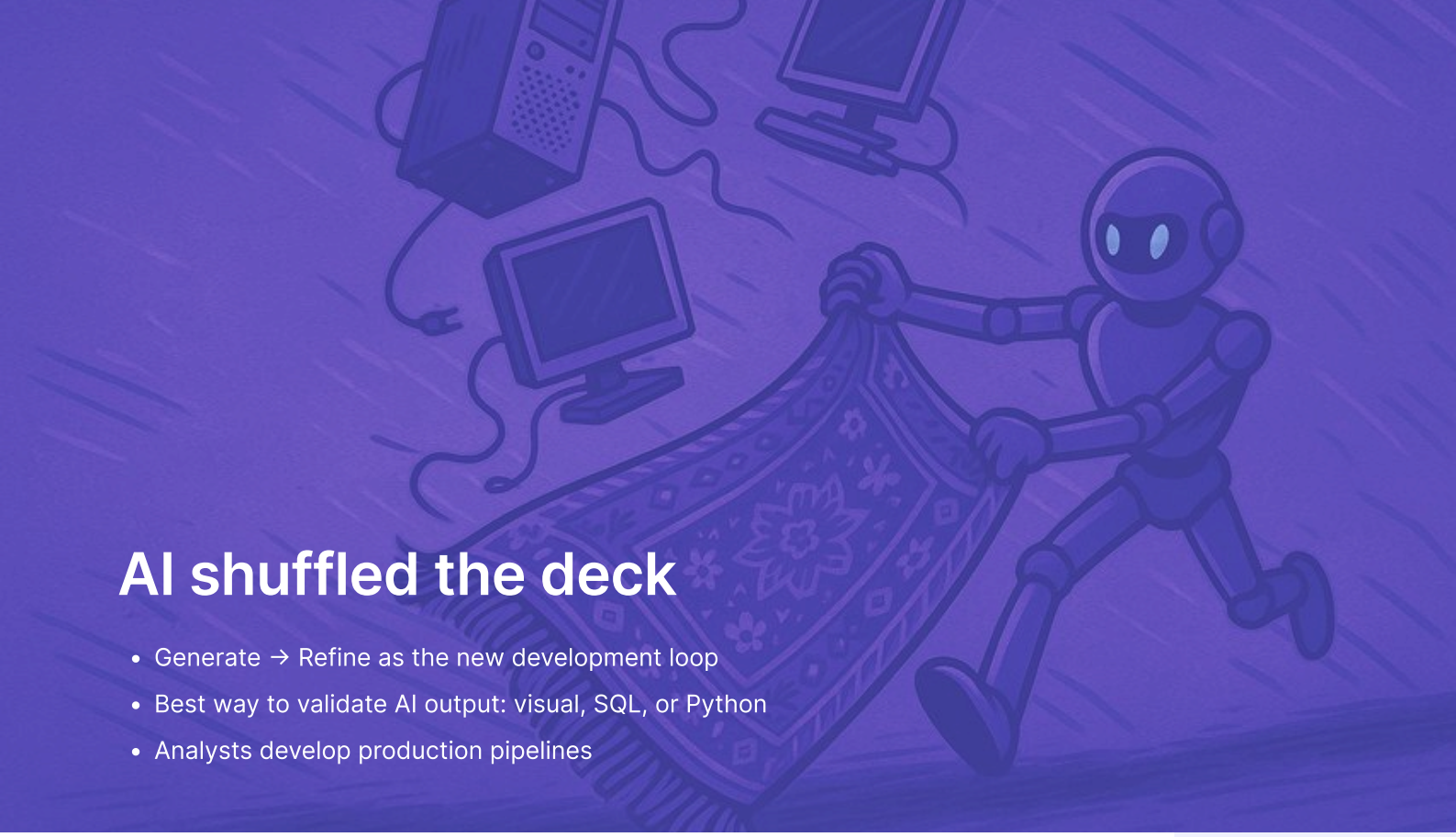
When a shift is incremental, replacement rarely wins out. Desktop data prep has weathered many technical waves by adjusting and adapting.

AI is different: it isn't incremental. Existing products will be replaced since their architecture prevents the full adoption of AI and their operations do not leverage cloud scale.

However, this does not mean overnight replacement; AI-native products will coexist with the desktop products for the transition.



Pace of change—Given the velocity of AI, these changes will come quickly. It took 6–12 months for software development to move to AI-first default for many engineers. Data preparation is following a similar timeline. Adopt AI-native products side-by-side, so you're ready to make change at the right pace.



AI shuffled the deck

- Generate → Refine as the new development loop
- Best way to validate AI output: visual, SQL, or Python
- Analysts develop production pipelines

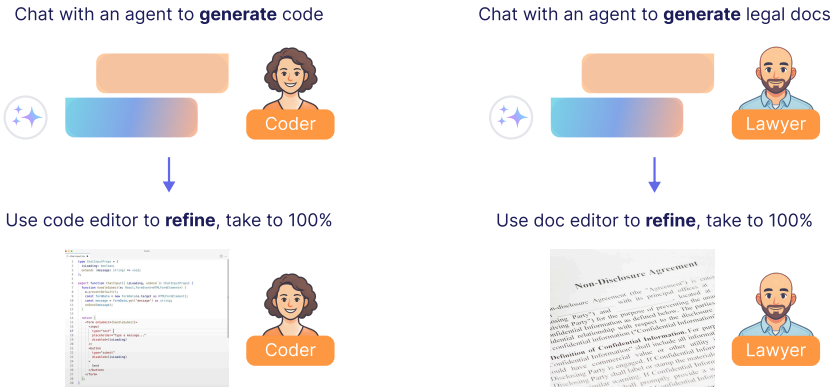
Development mechanism of the future

The emergence of AI assisted work in every field is converging on a new pattern, **Generate → Refine**, where:

1. Conversations with AI agents generate the first draft
2. The draft is produced in a format that the users understand
3. Users can refine (edit) the draft to reach the desired final result

Examples where this excels include text—work is produced as marketing or legal copy that users can edit or code—work is produced as code that developers can edit.

However, there are numerous domains such as images in ChatGPT where a prompt may generate something close to what the user wants, but there is no mechanism to edit the image for the final result.



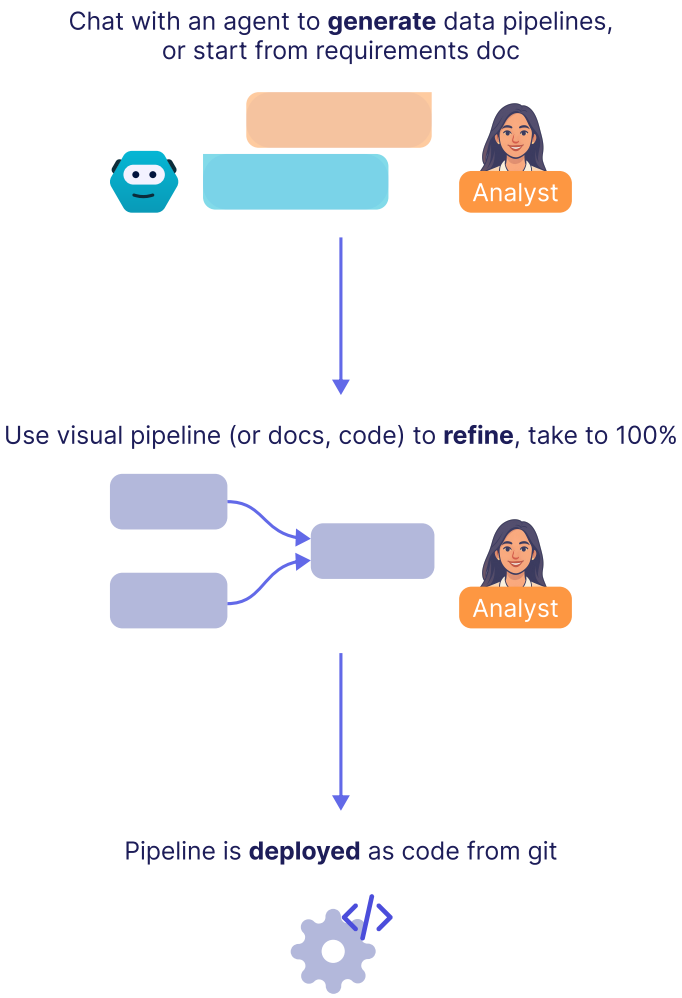
Mechanism applied to data preparation

Business data teams span marketing, finance, accounting, and more, and bring unique needs. Data users (analysts) are comfortable with visual pipelines and spreadsheets, while business domain experts often prefer documents.

The new G**enerate → Refine → Deploy** mechanism for business data work will bring incredible changes:

- AI generates results as visual pipelines that are easy to inspect, edit, and refine
- Documents and document authors are first-class citizens in the data process
- Generated pipelines are backed by production-ready code for deployment

This brings significant productivity gains by uniting all users around one shared, reviewable artifact.



Best way to validate AI output—visual, SQL, or Python

Data users often favor a particular way of developing data pipelines—visual pipelines (low-code/no-code), SQL, or Python.

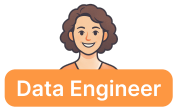
Does this need a rethink?

As AI generates more of the pipeline, the optimal format will be the one that is best suited to validate AI responses, not the one that's easiest for authoring.

Our team has tried different formats and find that visual pipelines are the best suited to validating AI output.

Analysts can develop production pipelines

The dynamic in many enterprises today is that:

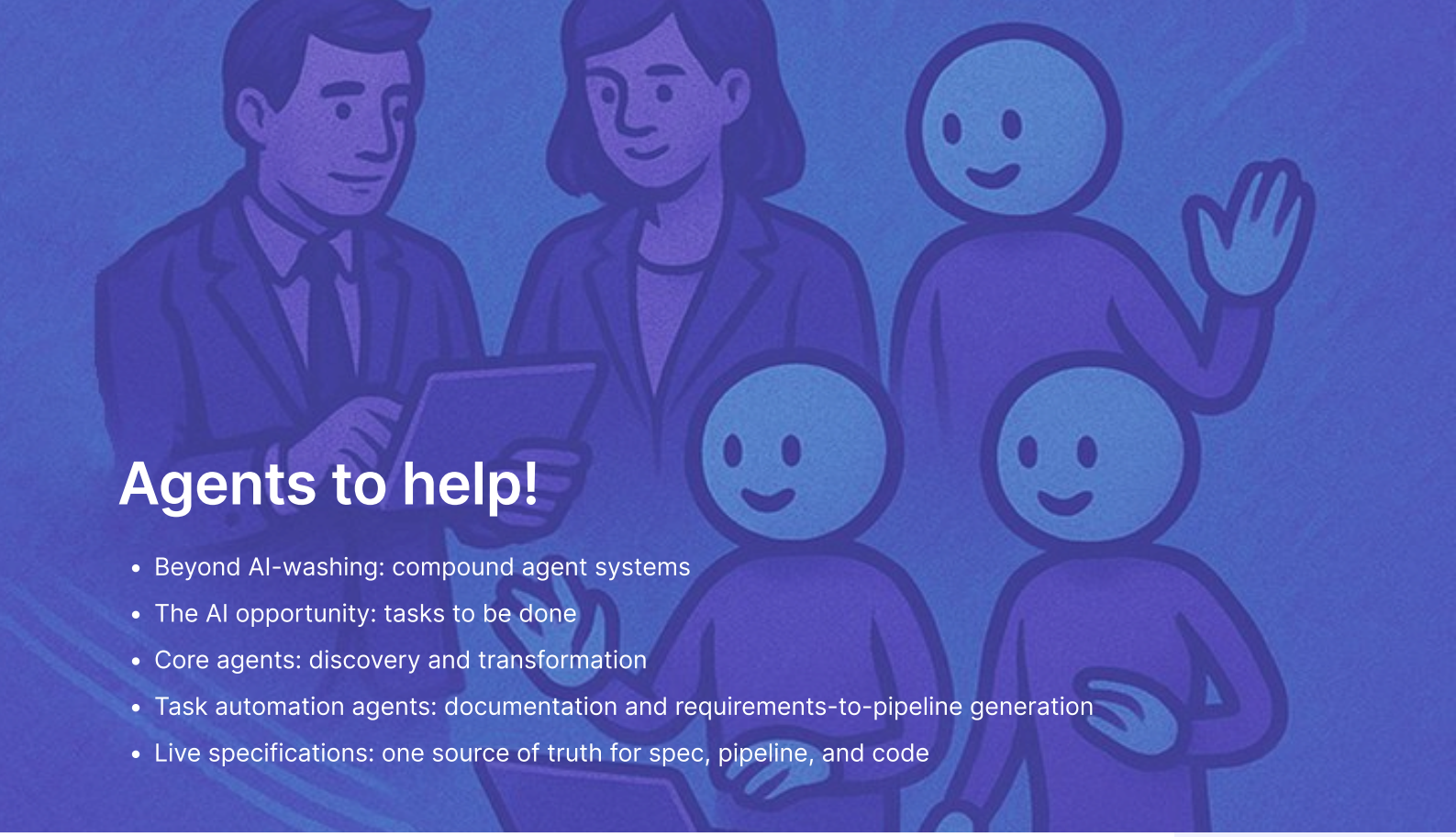


Data engineers build pipelines on platforms such as Apache Spark, typically in Python, and tune them for performance and reliability.

Business data users understand the business needs but aren't deep in the platform stack, so they rely on those engineering teams.

With AI, business data users will be able to build production-grade pipelines on the same platforms, greatly accelerating time to market for business-critical data.

Visual pipelines—as the best format for validating AI output, and the same as high performance code—should be the primary format for business data teams.



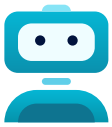
Agents to help!

- Beyond AI-washing: compound agent systems
- The AI opportunity: tasks to be done
- Core agents: discovery and transformation
- Task automation agents: documentation and requirements-to-pipeline generation
- Live specifications: one source of truth for spec, pipeline, and code

Agents will quickly emerge as the primary workers in data preparation and analysis—acting on behalf of users and collaborating with them. Let’s go a level deeper, to look beyond AI-washing, and delve into the capabilities to expect in the next few months from AI-native products.

Beyond AI-washing: compound agent systems

A typical business analysis workflow—preparing data and generating dashboards or reports—includes a series of narrow, specialized tasks: dataset discovery, transformation, error correction, documentation, and summarization. These tasks are well-suited to AI, but high-quality results require more than LLMs.



Manager agent (orchestrator/controller)

Interprets user intent, routes work to the right specialists, coordinates their outputs, and assembles the final result. Sometimes this agent simply guides the user; other times it decomposes complex requests and manages execution end to end.

Worker agents (specialists)

Domain-aware agents are built on AI models, and understand your data context, specialize in generating best results for a narrow task such as generating documentation for a data pipeline, and return results in the right format.

These are just some examples of the integrated approach. Sticking with legacy vendors that resort to 'AI-washing' with a layer of AI that isn't integrated into core workflows and a roadmap for upcoming magic will not give you productivity. Instead it will delay your productivity by 1-2 years, putting you behind your peers.

The AI opportunity: tasks to be done

Business analyses acquire multiple datasets, then clean, join, and aggregate them to produce data products such as reports and dashboards. These processes are deeply embedded in organizations, but many steps reflect historical technical limits that AI can change.

The analysis process

Requirements from the business user specify the goal.

Visual pipelines implement these requirements.

Analysts add and clarify business logic on top of the initial requirements.

Execution at production scale often requires a rewrite as code.

Cloud data platforms (Databricks, Snowflake, BigQuery) expect code.

Data engineers duplicate analyst work in a machine-oriented format (e.g., Python/SQL).

Regulatory documentation must describe the actual execution.

Regulators care about what was implemented; because requirements drift, the document is often written from scratch.

AI opportunity

AI can assist in two ways.

Productivity

Make every step more productive by maximizing automation from AI. This can be adopted immediately and delivers a significant boost for data users.

Unification

Unify formats into a single, collaborative specification (a “live spec”), removing hand-offs and rewrites. This is a ground-up rethink using advanced AI and yields further gains through fewer steps, shared context, and reduced iteration.

Let’s walk through these.

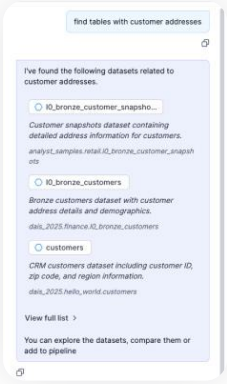
Productivity: agents for data discovery & transformation

AVAILABLE IN PROPHECY

Data preparation or analysis has two core tasks—finding the right datasets, and to prepare, combine, and transform them into the right data products—tables, reports of dashboards. There are tremendous productivity gains to be had with AI here. Data preparation is a special case of programming and as we have seen, AI is very successful and getting better each day—but this must be adapted to data.

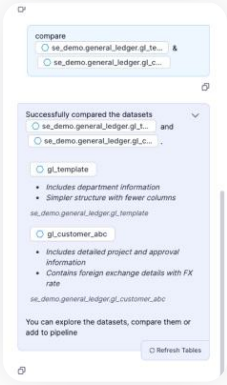
AI Data Discovery

Sometimes you know the data you need to work with intimately, and sometimes you must find it. Here is what finding and discovering data with AI will look like:



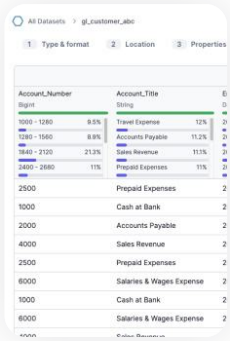
Find datasets

You can find data across all data platforms by attribute, not just table and column names.



Compare datasets

There are often multiple copies of a dataset, and the user must compare them to understand which one is right.



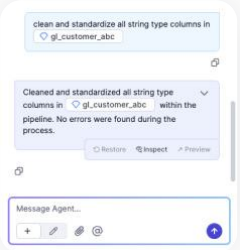
Explore datasets

When the summary information is not enough, the user can inspect the dataset as a table with summary statistics and slice & dice data with filters or plot charts, and ask questions using a chat with an agent.

AI for data transformation: create visual pipelines

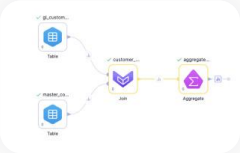
AVAILABLE IN PROPHECY

AI makes steps of transformation much simpler, where many times you do not need to specify the exact changes, but instead what you want to do with data, and the steps are encoded in the pipeline for you to inspect and validate.



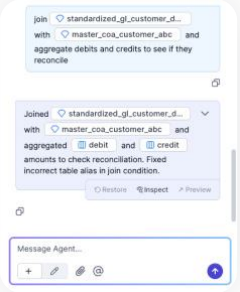
AI makes cleaning data simple

Use chat to transform raw input into well-formatted, ready-to-use data. Execute complex tasks like un-bundling in overloaded fields, normalizing dates and times, and creating geo points to for spatial analysis.



Use AI to build pipelines

AI agents can analyze selected data sets and execute multiple steps to get that data to the desired outcome. Using natural language, all data users can generate full pipelines, but as the pipelines get long, the accuracy reduces. A very reliable mechanism is to build the pipeline step-by-step, where one as a chat with an agent to describe the next logical step, and the agent generates one or a few visual components to achieve it.



Inspect results and refine

To get accurate results, users must understand each step of the agent's output. With Inspect, users can see every step of the pipeline and test outputs. They can then refine the process, resolve errors, and further iterate with the agent, refining to deliver the exact data that they need.

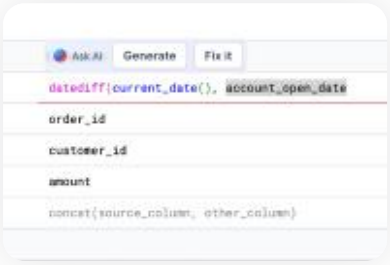
AI for data transformation: refine visual pipelines

AVAILABLE IN PROPHECY

Inspect and Refine

At every stage of the process, Prophecy enables users to inspect the business logic along with the data before and after each step.

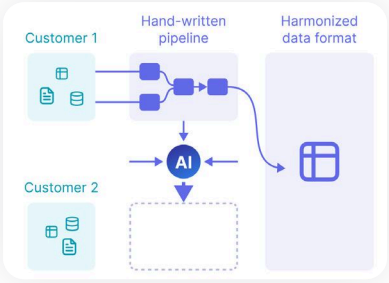
Users navigate steps with a visual representation of the pipeline, making it easy to understand the process and the incremental results along the way.



There are helper agents to generate tests and fix errors

As users develop pipelines, errors and warnings are automatically generated. Users can navigate quickly to the gem or expression to correct the error.

Users can also leverage AI to propose corrections across whole pipelines. This is particularly powerful for production systems where AI agents produce fixes for support team without the immediate input and expertise of the original team.



COMING SOON IN PROPHECY

Automatic harmonization of datasets

A very common pattern is to onboard data from a variety of sources across your customers and convert it into a single harmonized format to process it in a unified way. New datasets are similar, but not identical, forcing the need to create slightly different pipelines through very repetitive work.

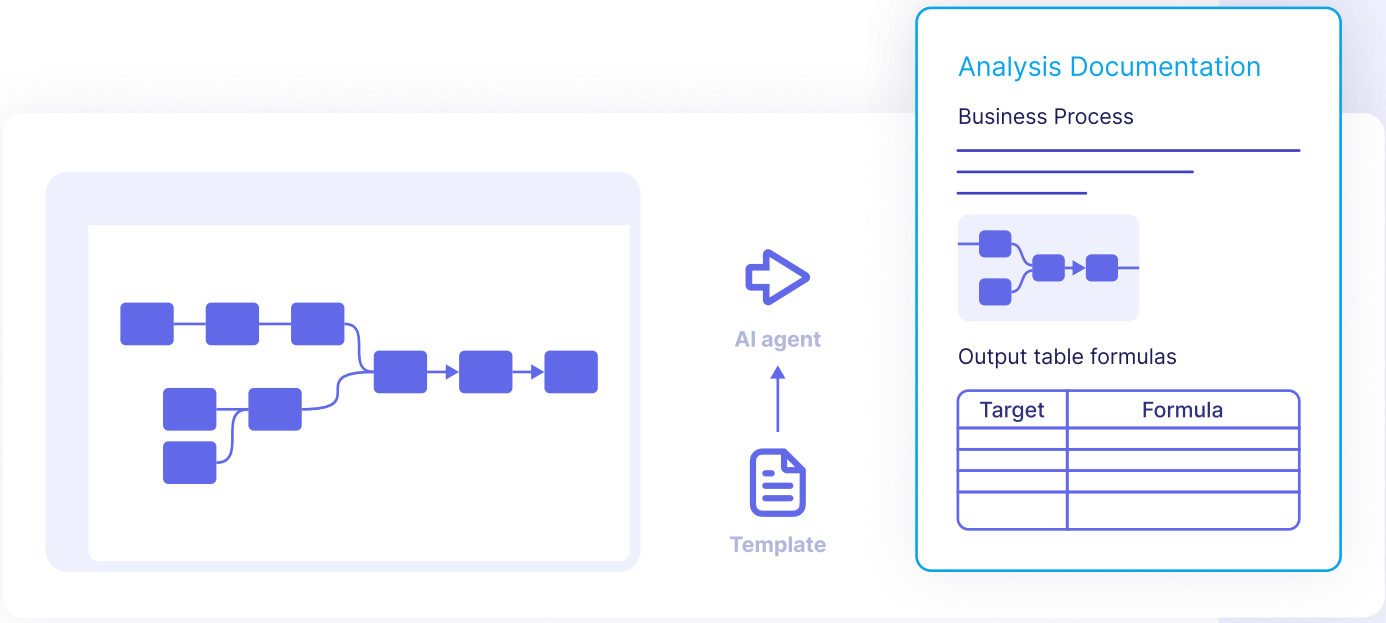
AI knows the harmonized format, and has a specification, or example pipelines that describe example transforms. Starting from these examples, a pipeline specific to the new source data can be generated.

Productivity: agents for documentation

COMING SOON IN PROPHECY

Once an analysis has been developed, many businesses need to formally document the implementation, often for regulatory purposes. Examples of this include accounting, finance, or pharmaceutical companies with audit requirements. The document often includes:

- Summary of the process (including screenshots of pipelines)
- Summary of business logic for target data

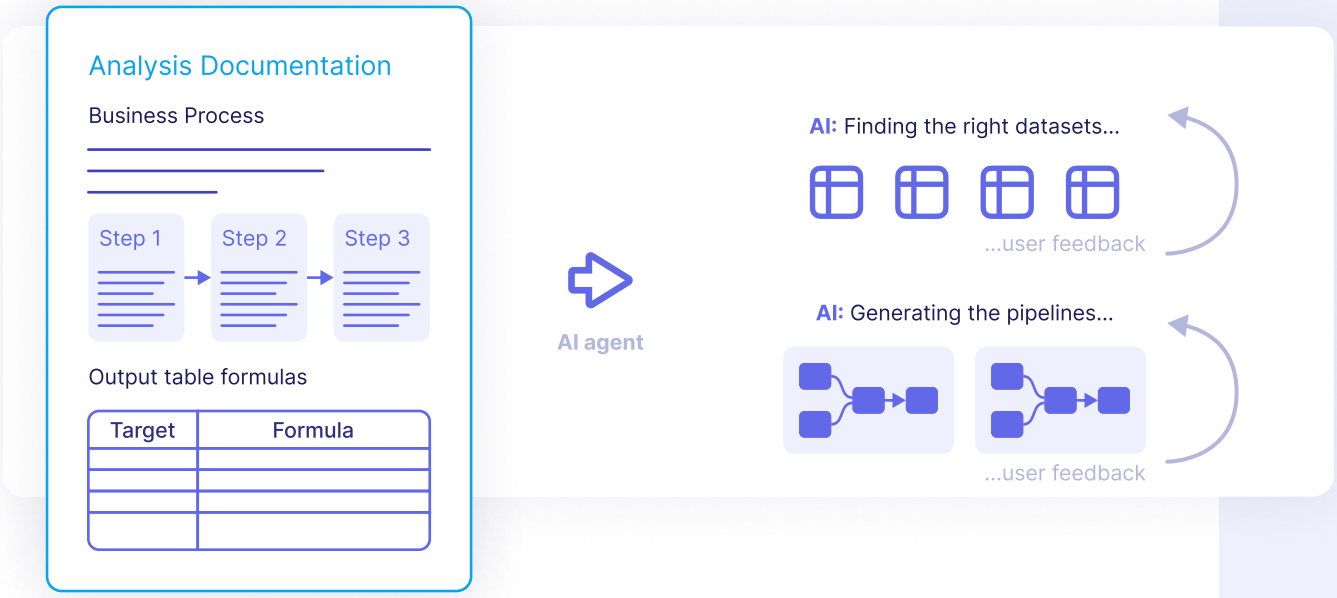


The AI agent can take the data pipeline and the desired template as inputs, and produce the document automatically. This can be reviewed by an analyst. Sometimes the output formats require answering questions and this process can be AI guided with suggested answers.

Agents for requirements-to-pipeline generation

COMING SOON IN PROPHECY

Many data analysts start with a business user, an internal or external customer, providing a requirements document that describes the business logic of one or more analyses. Data pipelines can be generated from the specification. As changes are made to the data pipelines, the requirements document is correspondingly updated.

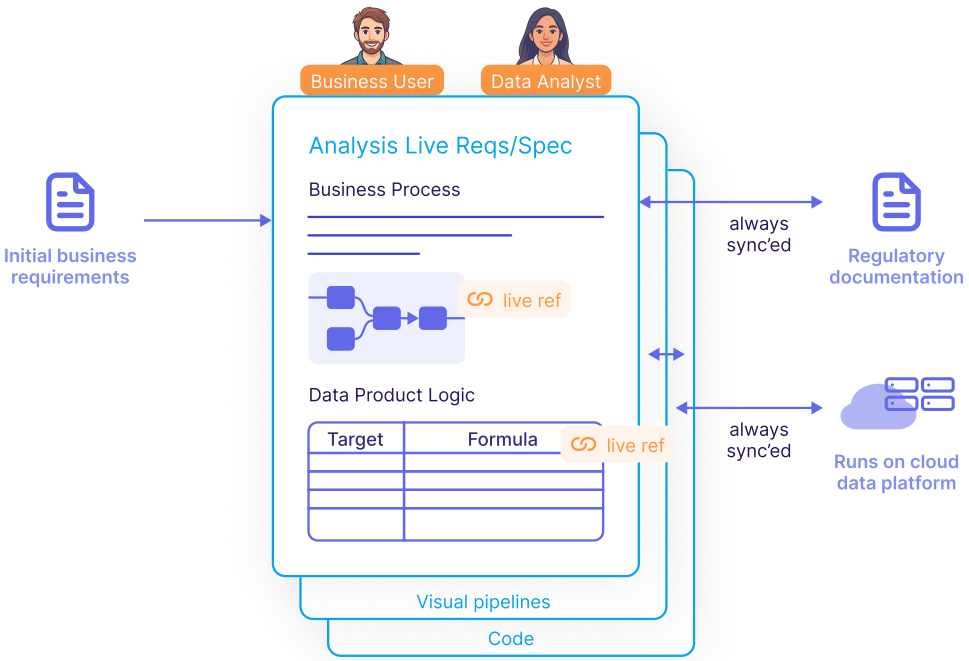


Live specifications: one source-of-truth for spec, pipeline, and code

If we rethink from scratch, with advanced AI along side us, would we build the analysis business processes the same way, or would we change the process? There are two main inefficiencies that we reimagine.

There are many formats and translations: docs for business users, visual pipelines for implementation by data analysts, optional code for running on cloud data platforms, and implementation documentation.

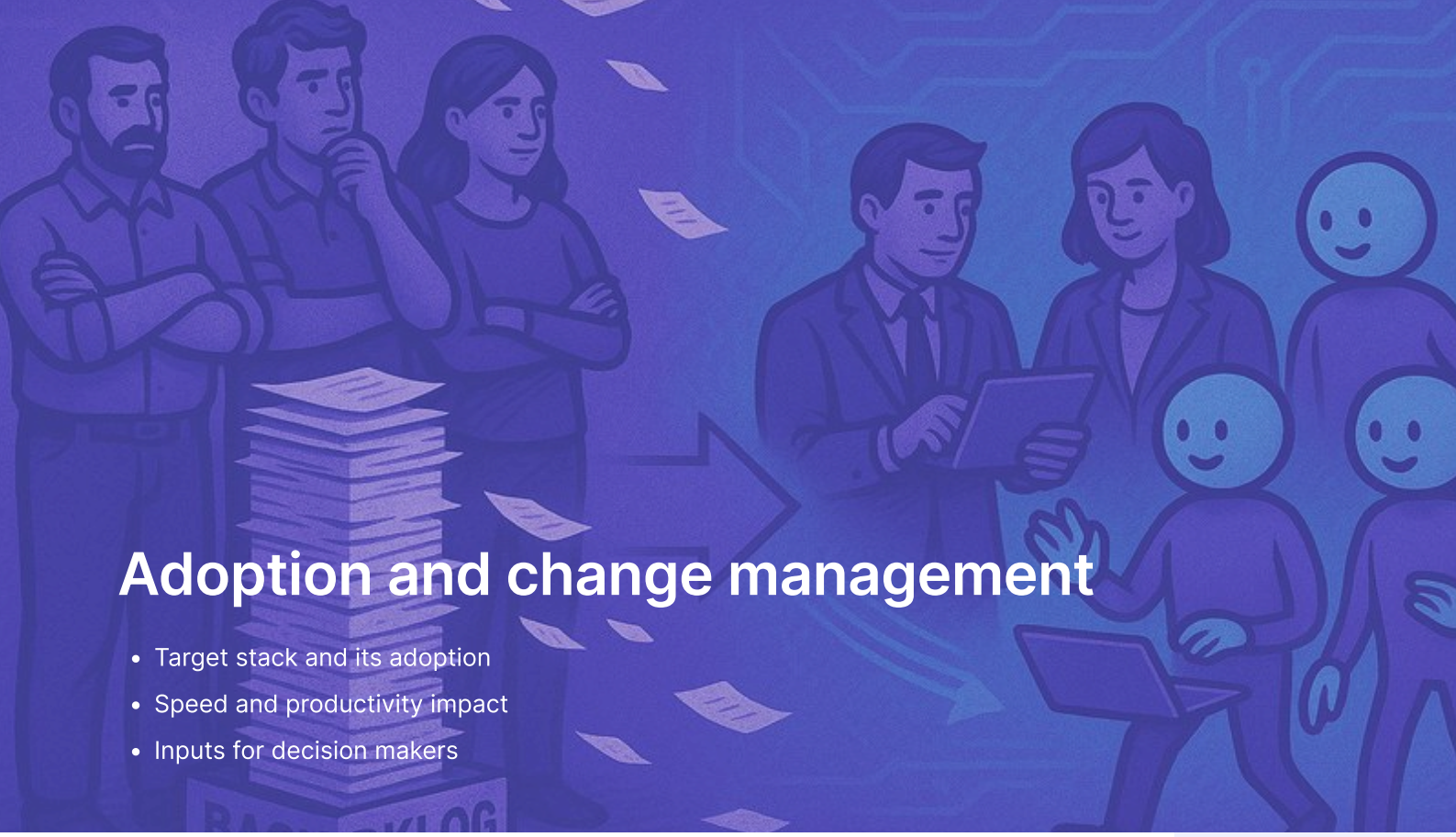
There is iteration: requirements initially are incomplete or inexact, requiring users across skillsets to go back and forth.



With AI, we can have a single specification that includes text from the business users, with embedded business processes and data logic tables that are parts of the implementation. In this single format both text and implementation can be edited and the other representations immediately updated.

With this approach, the documentation and code are just all formats of the live specifications, and are always 100% in sync with one another, with no effort from the user.

To generate complex business-impactful data, a typical pipelinecreation processes takes many back-and-forth steps just to translate between the desired outcome and the technical language needed to deliver the data. Radically reducing that friction is just one way AI will reinvent data pipeline creation.



Adoption and change management

- Target stack and its adoption
- Speed and productivity impact
- Inputs for decision makers

Adoption and change management has been a central topic in Prophecy conversations with current and prospective customers. Below is a summary of the issues and the solutions we're proposing.

Main questions to consider

? What does the target **product stack** look like?

? How do we adopt the new stack in a **phased way**?

? How does it change our **speed and productivity**?

? Will business data users do more, and **rely less** on central teams, reducing their backlogs?

? How much **repetitive work** can be automated, freeing analysts for higher-value work?

Let's look at each of these questions—and how Prophecy addresses them—so the answers are grounded in reality.

What does the target stack look like?

The transition from desktop to cloud consists of three layers.

Source stack (today)

Data analysts use a desktop data preparation product (such as Alteryx)

alteryx

Cloud platform

Databricks, Snowflake, or BigQuery where data engineers write code, establish central governance, and maintain common datasets.

databricks

snowflake

Google BigQuery

Cloud data preparation and analysis platform

AI driven development

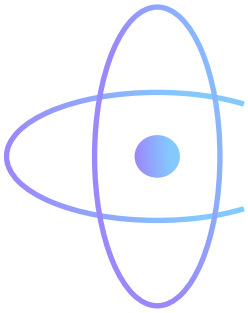
Resulting visual pipelines are also represented as document and code.

In-memory processing

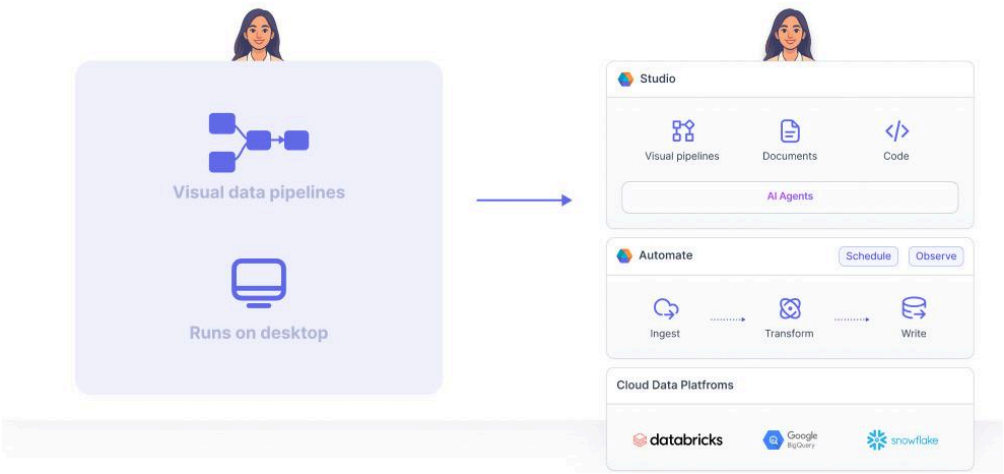
Ad hoc pipelines run in memory to avoid unnecessary cloud spend as business users shift to the platform.

Native execution with governance

For large pipelines, we connect to the cloud data platform, respect central governance, and run at scale in production without rewrites.



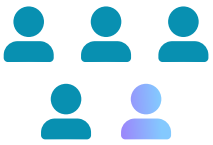
Prophecy



How do we move to the new stack?

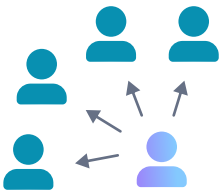
Moving to the new stack all at once creates a high risk and lengthy project. A better plan is to plan for a transition with coexistence and migration.

1 Gain experience on the new stack (3-6 months)



Designate first teams and workloads to be the early adopters of the new stack

Choose a greenfield project when possible. If a greenfield project is not available, choose a medium-sized existing project. Prophecy can import existing Alteryx pipelines and convert them to Prophecy pipelines, though this will require some manual effort for last mile changes. This gets you setup for new development.

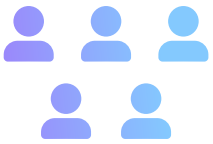


The first team to adopt Prophecy will develop best practices for your organization

They will gain experience in developing AI first pipelines for maximum productivity, and you'll setup integrations with central data stack.

You might work with Prophecy if you have specialized tasks like generating reports where the AI needs to do a better job on your specific workload.

2 Plan phased modernization



Once there is experience and confidence in the new product, various teams can plan and scope out their own modernization journeys.

How will it change speed and productivity?

AI agents improve time-to-market and productive capacity by cutting time to start projects, reducing hand-offs, speeding iteration, and eliminating rewrites to reach production.

Speed (time-to-market) and productivity

Faster starts

A goal statement produces a first draft (pipeline + tests + docs) in minutes.

Fewer hand-offs

Everyone edits a single live spec (visual, SQL, or Python view), so changes don't bounce between teams.

No rewrite for scale

Pipelines run natively on the cloud data platform; promotion is a Git review, not a re-implementation.

Built-in quality

Agents generate tests and docs with each change, reducing late rework.

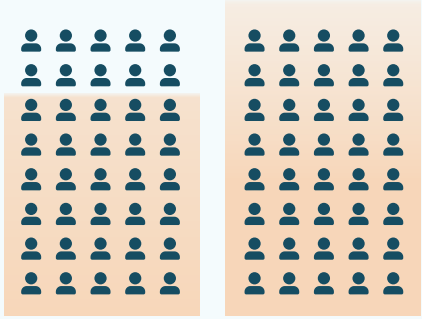
What to expect: cycle time drops as the initial pipeline arrives sooner, alignment happens in one place, and promotion is a review —not a rewrite. The result? A dramatic improvement in productivity.

Illustrative Math

The combination of these factors can have a dramatic impact on costs and productivity.

For example, a firm with 1,000 analysts at $100K average compensation (~$100M total) would see:

- A **25%** productivity lift ≈ **$25M** in annual capacity redirected to higher-value work.
- A **50%** lift ≈ **$50M**.



Inputs to decision making forleaders

AI is getting adopted very quickly and successful data leaders will be those that stay ahead of the curve. Here are some factors toconsider.

AI data prep will come fast and be excellent

Data analysis and preparation is adjacent to programming, just a few quarters behind.

Programming is a board-level topic with CEOs from technology companies such as Microsoft and Google to finance companies such as JPMorganChase—each touting 30% gains that they're already getting across tens of thousands of developers.

AI is a competitive must-have, don't get left behind

The consensus view is not that AI agents will replace coders or data analysts, but that coders with AI will replace those without.

Similarly, very soon your peer in another company will publish a case study detailing how they're getting the same work done, faster, and at half the cost—saving tens of millions to the organization. You don't want to be the manager who did not adopt AI quickly enough.

Do’s and don’ts of AI data prep

✓ Start soon, be a leader in delivering AI in delivering value with AI	✗ Move slow and have your peers publish millions in productivity gains before you
✓ Start with coexistence, bring new AI product to one use case, gain experience	✗ Don’t try to run a year long migration project into the new product
✓ Judge success by the ability of a new team to succeed in delivering their project	✗ Don’t have a technical team do feature-by-feature comparison, the legacy product will always have more features and will delay your modernization
✓ AI will work well on basics, but you might have high-value use cases such as your data harmonization, or documenting data pipelines that can be automated. Become a design partner, so that AI in the product works on your high value use case.	✗ Don’t expect AI to work out of the box for complex and specific use cases. Don’t miss the bandwidth for design partnership that tool vendors have today, before they hit scale.

How do I get started?

Prophecy provides Enterprise Express, designed to get you started fast and get your team to value within 3 months.

It includes:

- The ability to buy from marketplaces
- Success framework: identify team, project, and success criteria
- Onboarding support: the Prophecy team will work hand-in-hand with yours to help them deliver

Prophecy