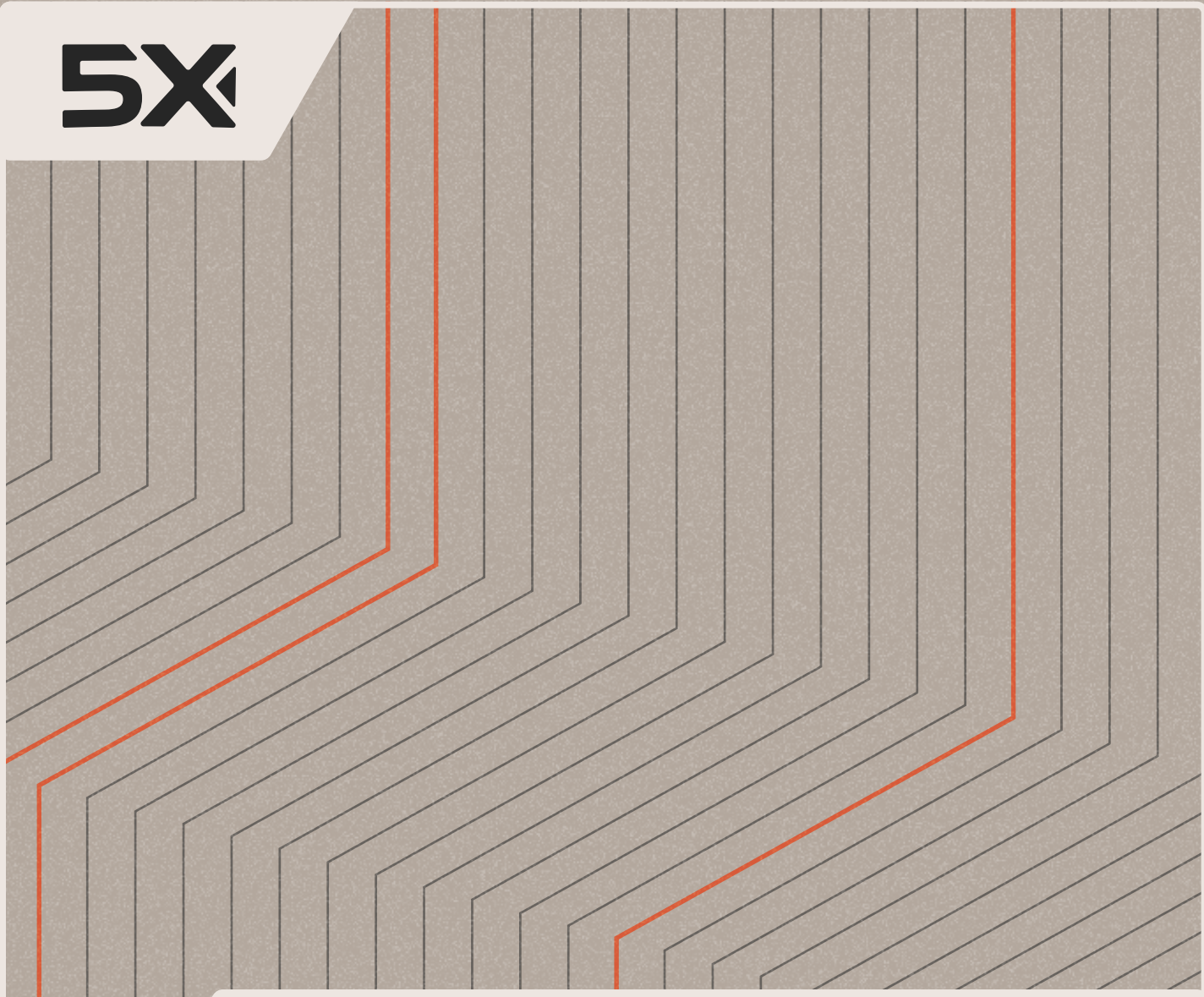
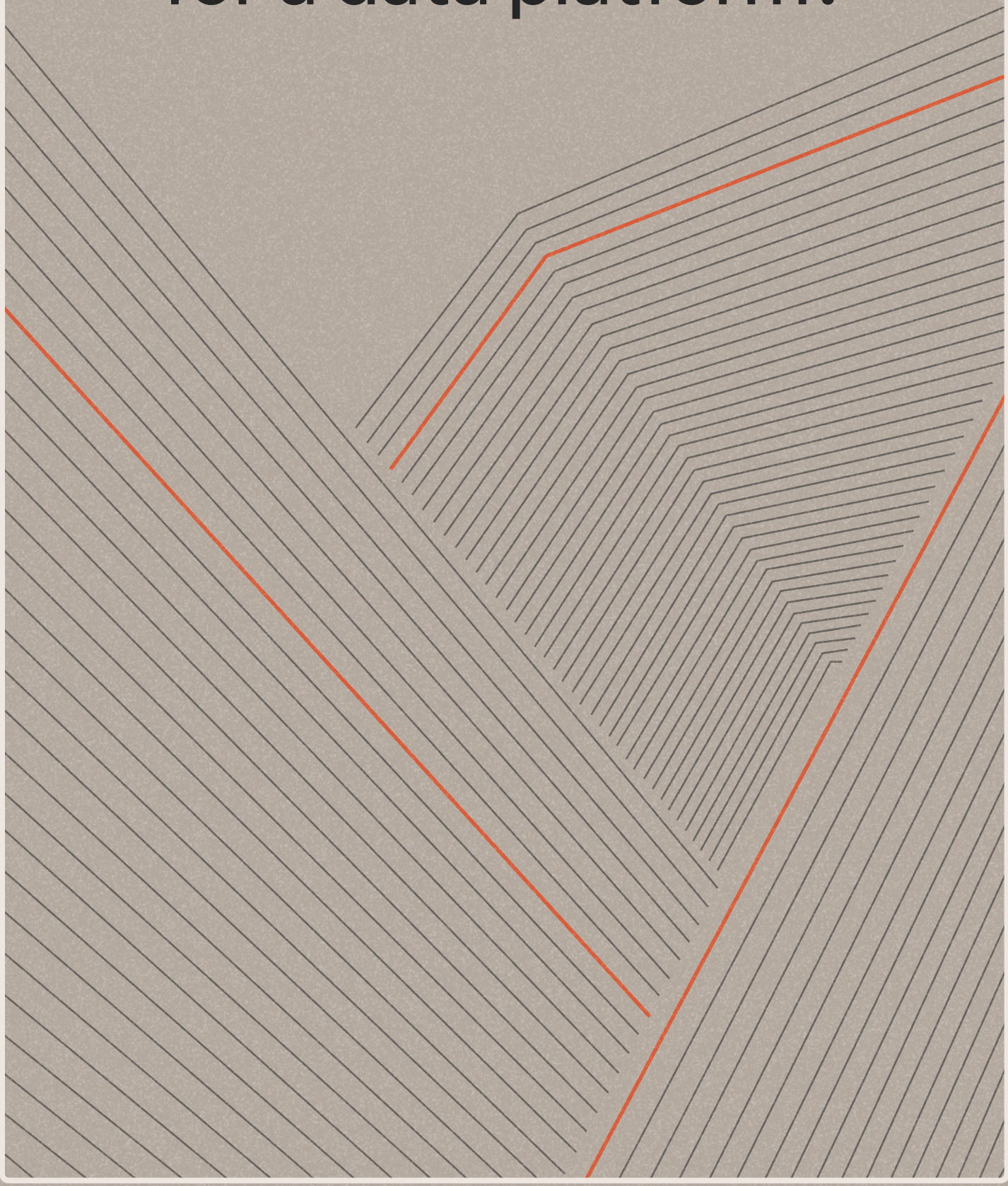


2025 Data Platform Buyer's Guide

The guide every vendor hopes you don't read.

Chapter 1

Are you even ready for a data platform?



Most companies shop for data platforms the same way they buy enterprise software; with optimism, good intentions, and a sprinkle of vibes. Unfortunately, that's the enterprise equivalent of impulse-buying an espresso machine you'll never learn to use.

It's not a tooling problem. It's a timing problem.

Before you invest in a modern data stack, you need to know whether your organization is actually ready for one. That means understanding what kind of data problems you have, how mature your infrastructure is, and where the biggest bottlenecks live—your tools, your team, or both.

That's not something you figure out in a brainstorming session. It's something you measure.

Start with a Data & AI Maturity Assessment

This is what we recommend to every company, whether it's just starting its data journey or halfway through re-platforming for the third time.

This Data and AI Maturity Assessment is a fast, objective way to benchmark where you stand today. Built on proven frameworks from Gartner and McKinsey, it evaluates your org across six core dimensions:

- Data infrastructure
- Data modeling & quality
- BI & dashboards
- Governance
- Predictive & prescriptive analytics
- AI applications

In less than 10 minutes, you'll get placed into one of five maturity tiers—**Nascent, Developing, Operational, Advanced,** or **Leading**—and receive a custom scorecard with tailored recommendations for what to fix, what to improve, and what to ignore (for now).

Why does this matter?

Because readiness isn't subjective. It's architectural. Strategic. Budgetary. Organizational. And until you understand where you are, every platform will look the same, and none of them will deliver what you expect.

The goal here isn't to rank you. It's to help you:

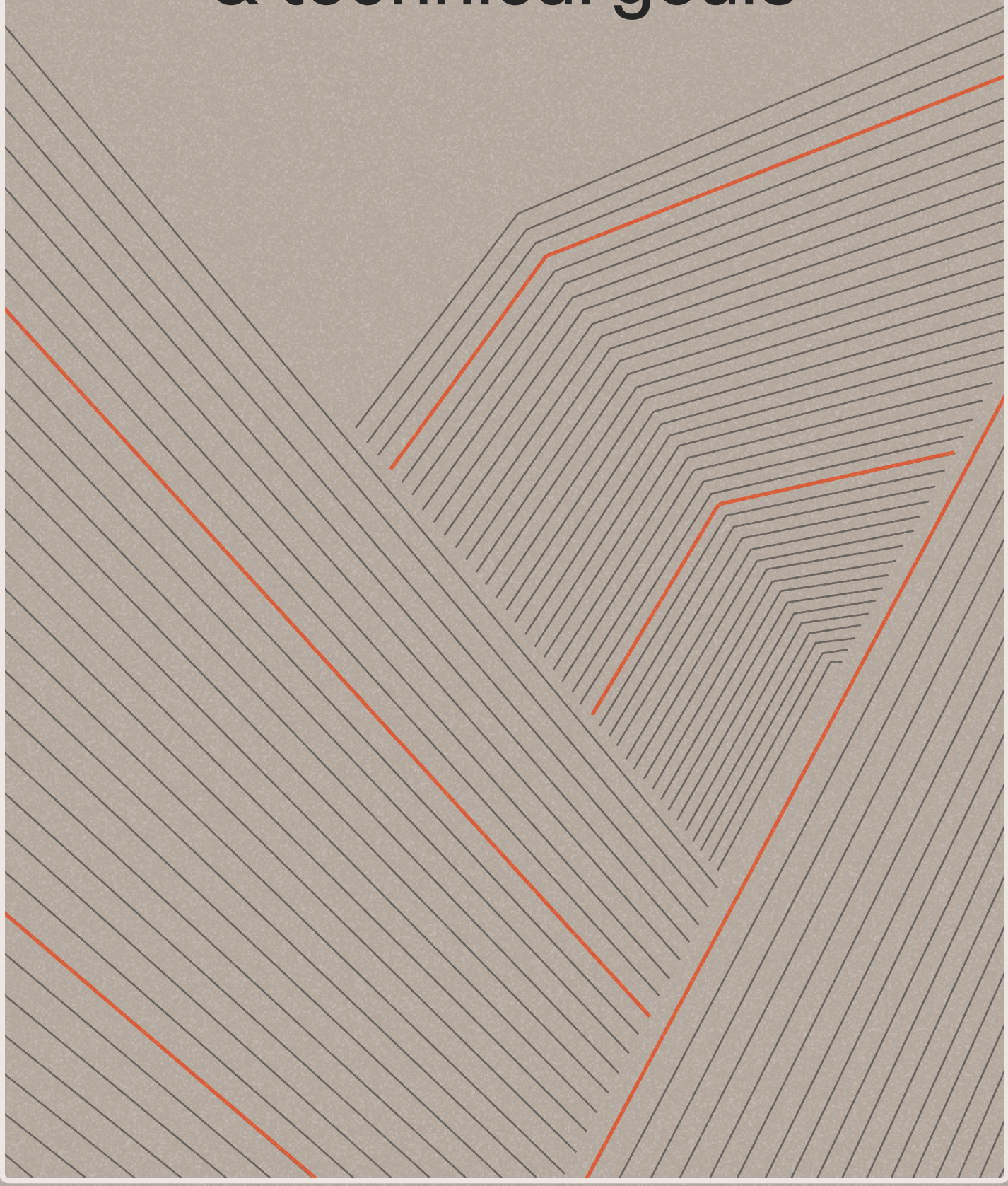
- Identify your real bottlenecks
- Align stakeholders around realistic priorities
- Choose a platform that fits where you are—not just where you want to be

Take the Data and AI Maturity Assessment

Complete it in under 10 minutes and get your personalized action plan. Because the only thing worse than picking the wrong platform is picking one before you're ready.

Chapter 2

Define your business & technical goals



Now that you know your data maturity level, it's time to zoom in. The assessment told you where you are. This next step is about where you want to go.

Before you start comparing platforms, you need to be crystal clear on **why you're even in the market.**

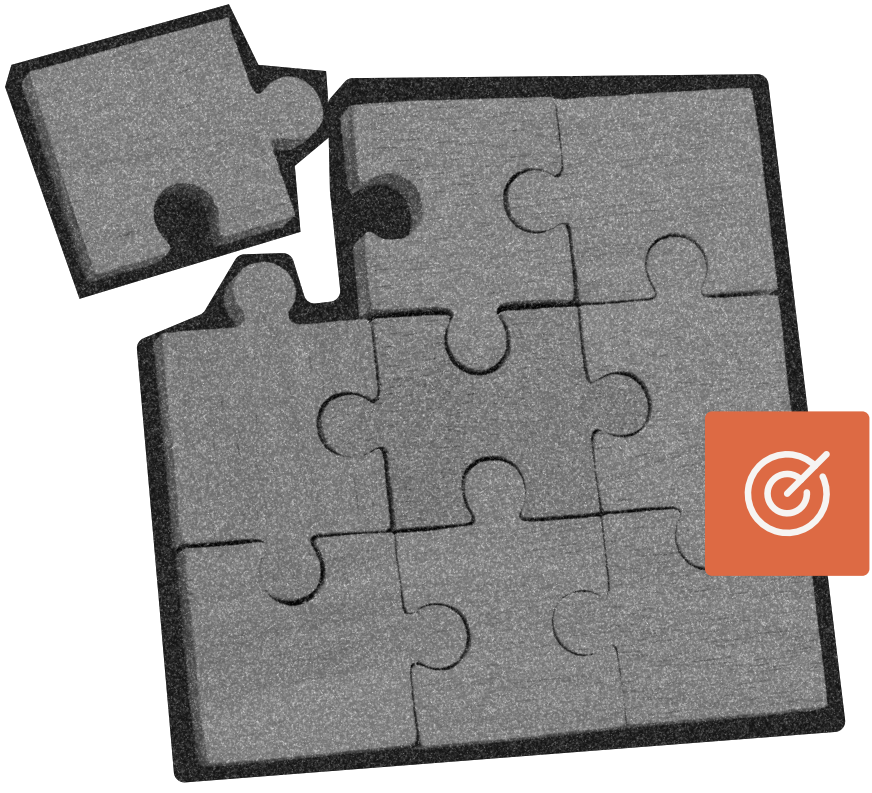
Not high-level answers like "to become data-driven." That means nothing until it means something. We're talking real, measurable outcomes across your business.

Every organization has its own mix of pressures. Your CEO wants better forecasting. Your product team wants real-time metrics. Your finance team is begging for cost visibility. And your data team? They're just trying to keep the damn pipelines from breaking.

So take a step back.

This Chapter is about setting goals that align across the business. Goals that translate into platform requirements.

Because if you can't connect the two, you'll end up with a tool that looks great in a demo and dies in production.



Break your goals into two parts:

1. What do you need the platform to do?

Think capabilities, actual use cases, and outcomes.

- Unify fragmented data sources?
- Power real-time dashboards for ops?
- Run AI/ML workflows on fresh data?
- Enable reverse ETL to push metrics into Salesforce?
- Reduce infra spend?
- Improve data governance and lineage?

List the things that will move the needle.

Then, rank them by priority.

2. What business outcomes are you solving for?

Now anchor each goal to a tangible business objective:

- Increase conversion rate by improving funnel visibility
- Reduce CAC by enabling better audience segmentation
- Improve renewal forecasting for CS teams
- Build embedded analytics to drive revenue from your product
- Speed up monthly closes by automating finance reporting

Tie everything to dollars, efficiency, or risk.

Many 5X customers have used this step to prioritize low-lift, high-impact use cases (like reverse ETL or cost optimization) before scaling into AI or ML territory.

Use the table below...

To map your top business goals against the data capabilities required to support them. This gives you a practical way to identify what your platform must deliver—and which features are non-negotiable based on your priorities. It also helps you quickly see where multiple goals share the same technical needs, so you can consolidate requirements instead of overengineering.

Business Goal	Unify data sources	Real-Time dashboards	AI/ML Workflows	Reverse ETL	Infra cost reduction	Governance & Lineage
Improve marketing ROI	✓	✓		✓		✓
Increase ops efficiency	✓	✓		✓		
Product personalization		✓	✓	✓		✓
Launch AI features			✓			
Lower data warehouse cost	✓	✓			✓	
Enable self-serve analytics	✓	✓	✓			✓

A note of caution

If your goals sound like this:

- "Centralize our data"
- "Make reporting easier"
- "Get off our old stack"

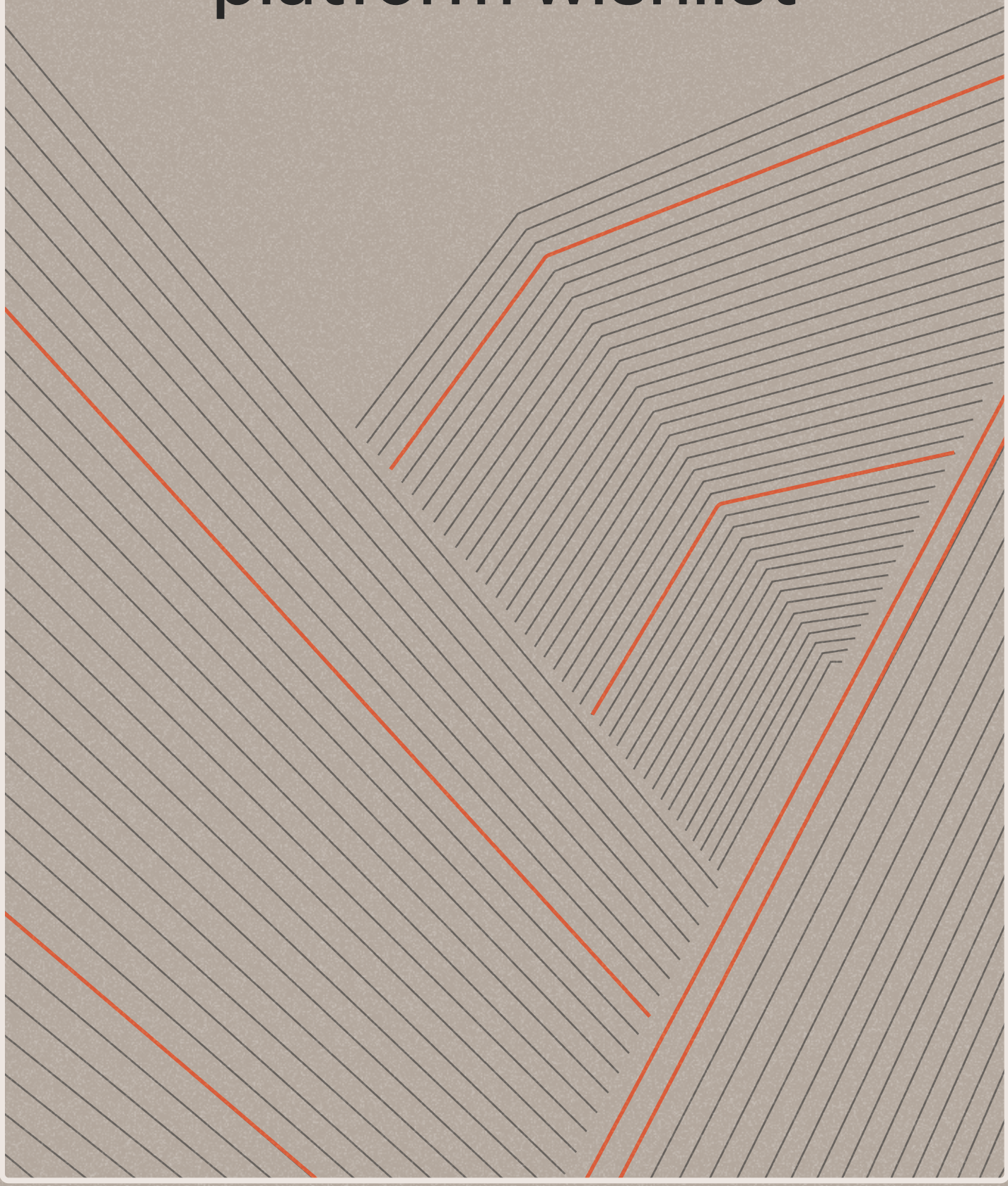
That's a red flag. These aren't goals. They're symptoms.

Push for clarity. Real platforms solve real problems. If you don't know what you're solving, you'll buy the wrong solution.

Use this chapter to get everyone aligned before you move to feature checklists or vendor pitches.

Chapter 3

Build your ideal data platform wishlist



Once you know where you are—and what you're solving for—the next step is to define what the right platform looks like.

Not what your peers are buying. Not what the market is hyping. What your team actually needs.

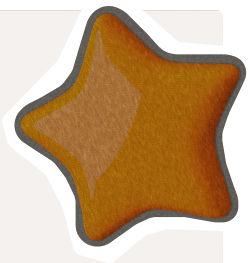
Every platform promises a full-stack dream. But when you start peeling back layers, most fall short in one of two places: either they're stitched together with duct tape, or they force you into proprietary black boxes you can't escape later.

This chapter is about narrowing your focus to the features that actually matter, so you don't end up with shelfware and regrets.

We've broken down the modern data platform into six functional layers. Each one includes critical questions to help you figure out what's essential, what's flexible, and what your team will realistically use. For each layer, we've also highlighted who benefits most—so you can identify the right stakeholders and how this platform will unlock value for them.

So before we walk through every layer of a modern data stack, here's a shortlist of priorities that consistently separate successful implementations from shelfware.

What to prioritize



Across dozens of implementations, the 5X team has helped customers stay focused on the fundamentals that actually move the needle.

If we had to pick the non-negotiables? It'd be these:

- **Cloud-agnostic, secure deployment:** Run securely in your VPC, on-prem, or hybrid
- **Open-source core:** Open formats and open-source components ensure zero vendor lock-in and full data sovereignty
- **Enterprise-grade support:** Including built-in governance and compliance

This wishlist framework is vendor-neutral—but we've designed 5X to tick these boxes by default.

Explore 5X platform

1. Data ingestion & integration

- Do you need native connectors, or will your team build/maintain custom ones?
- Does the platform support both real-time streaming and batch processing?
- Is Change Data Capture (CDC) support built-in or layered on?
- How much prebuilt integration exists with your sources (databases, APIs, SaaS tools)?

Who benefits most:

- **Data engineers:** Faster onboarding, fewer custom builds
- **Analytics engineers:** Clean pipelines = reliable insights
- **DevOps:** Less maintenance overhead

2. Orchestration & pipelines

- Is the orchestration engine declarative (like dbt/Prefect) or code-first (like Airflow)?
- Do workflows support CLI, UI, and SDK triggers?
- Can you manage retries, alerts, SLAs, and dependencies cleanly?
- How does it integrate with your existing CI/CD tooling?

Who benefits most:

- **Data platform teams:** Simplified job monitoring and scaling
- **Data engineers:** Fewer manual reruns, better automation

3. Storage & warehousing

- What engines does the platform support—Snowflake, BigQuery, Redshift, Lakehouse, etc.?
- Does it optimize for scale, latency, cost—or all three?
- Does it support open formats (Parquet, Iceberg, Delta)?
- How portable is your data if you switch vendors later?

Who benefits most:

- **Infra/Ops teams:** Manage cost and performance predictably
- **Finance:** Forecast TCO and reduce surprise overages
- **CTOs:** Strategic flexibility without lock-in

4. Transformation & modeling

- Does the platform integrate natively with dbt?
- Can you modularize and reuse logic across models?
- How does it handle data quality checks and tests?
- Is lineage tracking automatic?

Who benefits most:

- **Analytics engineers:** Faster model iteration, more reuse
- **Data quality teams:** Built-in validation
- **Finance stakeholders:** Confidence in reporting integrity

5. Activation & reverse ETL

- Does it offer out-of-the-box syncs to tools like Salesforce, HubSpot, Marketo?
- How often can data be refreshed in destination systems?
- Is data pushed, pulled, or event-based?

Who benefits most:

- **Marketing:** Real-time personalization and campaign syncs
- **RevOps:** Lead scoring, pipeline updates
- **Customer success & product:** More responsive playbooks

6. AI readiness

- Can it handle both structured and unstructured data?
- Does it support embedding stores, vector databases, or hybrid search?
- Is there tooling for RAG, fine-tuning, and LLMOps?
- Can you easily feed model outputs back into the platform?

Who benefits most:

- **Data scientists:** Faster prototyping, integrated deployment
- **Product teams:** Build LLM-powered experiences
- **AI/ML teams:** Native support for experimentation and scaling

Your ideal platform wishlist

Use this worksheet to turn abstract goals into concrete platform requirements. For each capability, align internally on:

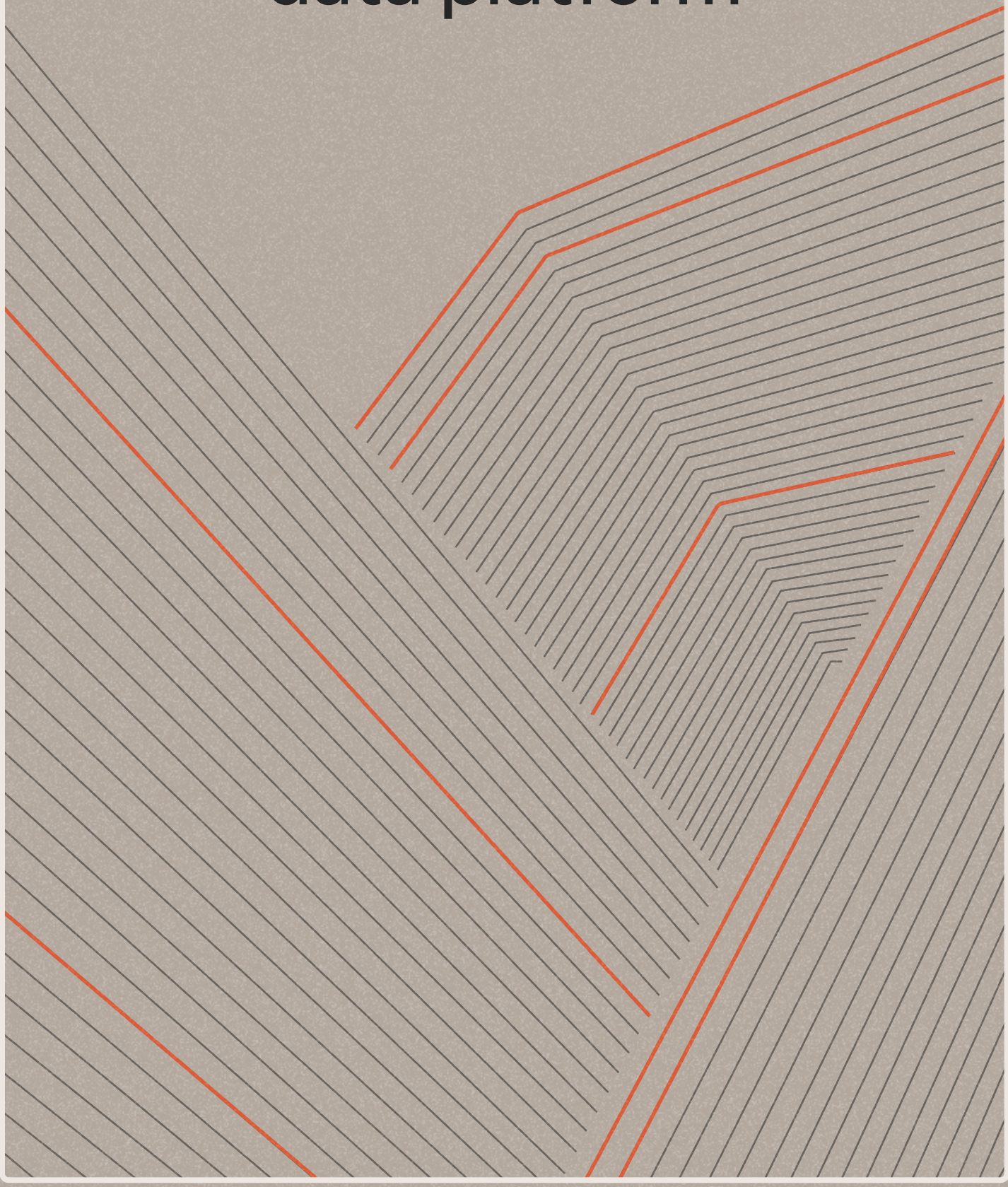
- Whether it's a must-have or nice-to-have
- Who's responsible for driving the decision
- How critical it is to the business
- Any open questions to resolve

Capability	Must have	Nice to have	Owned by	Team priority	Questions to clarify
Cloud-agnostic deployment	☑		Infra / CTO	High	Can we deploy in our cloud / on-prem stack?
Modular, open-source core	☑		Data Platform Lead	High	Can we swap out orchestration, storage, etc.?
Full-stack coverage	☑		Data Eng / Product	High	Can we avoid stitching 5 tools together?
Enterprise-grade support	☑		Engineering	High	Do we have a named success contact?
Predictable pricing	☑		Finance / Procurement	High	Are we avoiding compound vendor markups?
AI-ready architecture	☑		Data Science / AI	Medium	Can we build GenAI use cases natively?
Native data connectors	☐	☐	Data Engineering	High	Do we need ELT or real-time sync?
Real-time ingestion support	☐	☐	Infra / Data Eng	Medium	Are streaming sources on our roadmap?
Change Data Capture (CDC)	☐	☐	Data Engineering	Medium	Do we need CDC for transactional sources?
Declarative orchestration	☐	☐	Platform Lead	High	Are we standardizing on dbt?
Retry logic & SLAs	☐	☐	Infra	High	Are there hard SLAs for refreshes?
Open format compatibility	☐	☐	CTO / Data Architect	Medium	Do we plan to use Parquet / Iceberg?
Reverse ETL syncs	☐	☐	RevOps / Marketing	High	What destinations are essential?
Business user self-service	☐	☐	Product / Ops	Medium	Do business teams own sync config?
Embedding store integration	☐	☐	AI / ML team	Medium	Are we building GenAI features this year?

Download the checklist

Chapter 4

Choose the right data platform



Choosing a platform is as much a business decision as it is a technical one. But most evaluation cycles lose steam fast—bogged down by feature checklists, siloed input, and slide decks that all sound the same.

This chapter gives you a structured, step-by-step approach to run a high-confidence evaluation. You'll get clear guidance on how to scope your needs, align stakeholders, and run a transparent selection process that works across business, data, and tech teams.

1. Scope what you're solving for

Before you issue a single calendar invite or vendor form, get clarity on what problem you're solving.

Ask:

- What are your top 2–3 business goals?
- Where is your current data setup failing you?
- What's the maturity level of your team? (Use the Data & AI Maturity Assessment here)
- Are you replacing something or starting fresh?
- What does success look like in 6–12 months?

Tip: Not every company needs a full-stack platform. Your scope should fit your maturity and growth plan and not what's trending on LinkedIn.

2. Identify and align stakeholders

Include voices across the stack. Who's using the platform? Who's buying it? Who's responsible for outcomes?

Create a working group of:

Stakeholder	Role	What they care about
CTO / CDO	Budget holder, strategy	Scalability, extensibility, ROI
Head of Data	Day-to-day owner	Flexibility, integration, governance
Engineering	Infra/operator	Deployability, support, performance
Marketing / RevOps	End users	Reverse ETL, campaign data
Finance	Buyer partner	TCO, licensing, forecastability

3. Evaluate vendors against real needs

A common trap in vendor evaluations? Letting the vendor lead.

They show up with a polished deck, a pre-scripted demo, and a promise to "circle back" on your real questions. Meanwhile, your team nods along, unsure whether what you're seeing is product, prototype, or PowerPoint.

That's why you can't afford to evaluate reactively. This is your stack, your data, your business. You drive.

Here's how.

→ Show, don't tell: demo the way you work

Most vendors will walk you through a happy path that shows everything working perfectly. That's not useful.

Instead, take control of the session. Ask them to walk through your actual use case—live.

Examples:

- "Show me how we'd ingest data from [your source] into your platform."
- "Can you build a simple pipeline to clean and model this?"
- "Let's sync that into Salesforce. How many steps? How much code?"
- "If a pipeline fails at 2am, how are we notified? Can we retry automatically?"

Pro tip: Send them the scenario ahead of time and ask for a tailored walkthrough. The good ones will say yes. The great ones will be excited. The rest? That tells you something too.

→ Ask hard, non-generic questions

You're not here to be sold. You're here to uncover fit.

Here's a list of pointed, practical questions that map back to the key evaluation themes from Chapter 3:

On platform capability

- Does your platform support both batch and real-time ingestion?
- What kind of transformations are supported—SQL, Python, dbt-native?
- Do you support Change Data Capture (CDC) out of the box?

On AI & ML readiness

- Can you support structured + unstructured data pipelines?
- How do you manage embedding stores, vector search, RAG workflows?
- Are LLM outputs integrated natively into your stack or stitched in?

On cloud & deployment

- Can I deploy securely in my cloud (BYOC)? On-prem? Hybrid?
- What are your security compliance certifications (HIPAA, GDPR, etc.)?

On lock-in

- How securely can I export pipelines, metadata, and models?
- Are you using open-source standards or building proprietary black boxes?

On support & services

- Who will handle onboarding? Is it a partner, or your team?
- Do we get Slack-based support with access to engineers?
- Can we schedule architecture reviews as our needs evolve?

On pricing & TCO

- How is pricing structured—by usage, users, or modules?
- Is pricing predictable at scale, or does it spike?
- Are there "gotcha" fees for storage, compute, or egress?

On cloud & deployment

- Can I deploy securely in my cloud (BYOC)? On-prem? Hybrid?
- What are your security compliance certifications (HIPAA, GDPR, etc.)?

On lock-in

- How securely can I export pipelines, metadata, and models?
- Are you using open-source standards or building proprietary black boxes?

→ Keep your own scorecard

Demos move fast. Promises blur. Notes get lost in inboxes.

That's why you need a **shared scorecard.** It lets you:

- Track each vendor's answers and gaps
- Record what was actually shown live
- Note red flags, surprises, or areas that need deeper follow-up
- Compare side-by-side, asynchronously, with your working group

Use a simple template in Notion or Google Sheets. For each capability, assign:

✓ Fully supported

! Partial / workaround

✗ Not supported

And assign internal owners to follow up on each open item.

Scorecard template

Evaluation category	Evaluation question	Vendor response	Support level ✓ ! ✗	Internal notes	Follow up needed? (Y/N)
Platform capability	Does the platform support both batch and real-time ingestion?				
Platform capability	Do you support Change Data Capture (CDC) out of the box?				
AI & ML readiness	Can you support structured + unstructured data pipelines?				
AI & ML readiness	How do you manage embedding stores, vector search, RAG workflows?				
Cloud & deployment	Can I deploy in my cloud (BYOC)? On-prem? Hybrid?				
Cloud & deployment	What's the typical setup time in a VPC environment?				
Lock-in	What happens if we decide to migrate off your platform?				
Lock-in	Can we export all pipelines, models, and metadata in open formats?				
Support & services	Do we get Slack-based support with access to engineers?				
Support & services	Can we schedule architecture reviews as our needs evolve?				
Pricing & TCO	Is pricing predictable at scale, or does it spike?				
Pricing & TCO	Are there "gotcha" fees for storage, compute, or egress?				

Download the scorecard template

→ After 3–4 demos, patterns emerge

You'll start to notice:

- Which vendors show up prepared vs. scrambling
- Who actually solves your use case vs. tap dances around it
- What's native vs. bolted on vs. vaporware
- Where complexity hides (and who will pay for it later)

Don't rely on a "gut feeling." Document everything. Review in working sessions. Use this as the basis for your RFP scoring.

4. Final decision & rollout planning

You've scoped your goals. Built your wishlist. Led the demos. Compared vendors.

Now comes the part where most teams stall.

Too many data platform evaluations drag into endless cycles of "one more meeting," "one more stakeholder," or "one more reference call." Meanwhile, pipelines remain a mess, dashboards stay broken, and the team loses momentum.

This is where you draw a line.

You're not just buying software. You're committing to a long-term partner that will power your infrastructure, your insights, your AI. So how do you make the final call?

→ Lock in on alignment, not just features

By now, most vendors probably looked impressive on paper. Maybe even during demos. But your goal isn't to pick the most feature-rich option. Your job is to choose the one that aligns best to your priorities.

Here's a final checklist to stress-test your short list:

1. Alignment to business priorities

- Which vendor maps cleanly to the "must-have" capabilities your team defined?
- Who solves for your top use cases out of the box—without weeks of engineering lift?
- Are they addressing your current and next 12-month roadmap?

2. Deployment & architecture fit

- Can you run this in your cloud (BYOC)?
- Is there on-prem or hybrid flexibility if needed down the line?
- What tooling do they require you to adopt—and what can you swap out?

3. Lock-in and exit plan

- Can you export pipelines, metadata, models easily?
- Are they using open-source standards or building black boxes?

4. Support & onboarding

- Is there a named team for you post-sale?
- Do they offer hands-on architecture guidance, regular check-ins, Slack support?
- Is it service, or is it just software?

5. Total cost of ownership

- What's included in the contract vs. what's going to be an upcharge later?
- How does cost scale with volume, users, and complexity?

6. AI & future-readiness

- Is AI baked in or bolted on?
- Do they support structured + unstructured data, vector stores, RAG pipelines?
- Can their platform handle both dashboards and LLM use cases?

→ Don't buy blind—test with real data

Before signing anything, ask for a **trial environment.**

Your team should be able to:

- Connect a real data source
- Run a basic pipeline
- Model and visualize one use case
- Sync to a real destination
- Validate performance, UX, and operational complexity

At 5X, we offer a **fully deployable trial environment.**

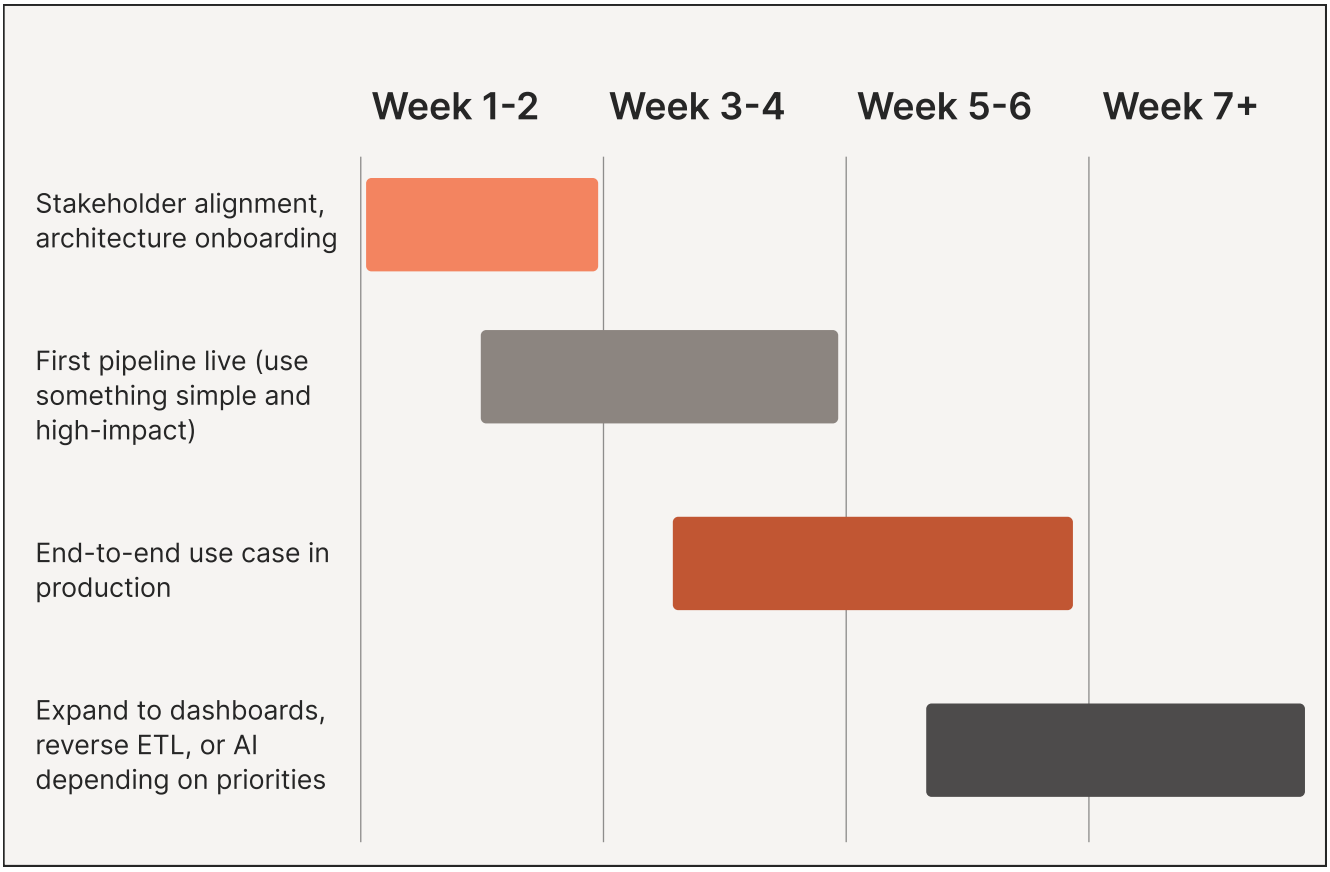
No mock data. No locked features. Just a working stack, built on your real workflows, so you can evaluate based on truth.

Start a free trial

Rollout planning: keep momentum, avoid chaos

Once the decision is made, don't wait 6 weeks for a kickoff.

Build a basic implementation plan:



Set clear roles—who owns what, who drives velocity, who handles training. Give teams small wins fast. Create champions early.

Chapter 5

Most platforms show you features. 5X shows you results

If you've made it this far, you've done the hard part. You've mapped your maturity. Prioritized your needs. Scoped internal alignment. Built a rigorous RFP. Sat through more demos than you care to admit.

Now what?

It's time to make a decision that won't just check today's boxes, but scale with tomorrow's bets.

But most platforms are built to showcase features. Not to solve for speed, flexibility, security, or stakeholder alignment. That's why so many teams end up stuck with bloated costs, vendor lock-in, and dashboards nobody uses.

5X exists to change that.

We didn't build another "data tool." We built an **outcome infrastructure platform**—one that lets you start anywhere (a dashboard, an AI app, a reverse ETL sync) and scale as you go.

The gaps traditional platforms leave behind

By now, you've probably seen it all:

- ✕ Demos full of AI promises—until it's time to deploy your actual use case
- ✕ Composable stacks with 5 tools, 3 contracts, and no real support
- ✕ Platforms that need 90 days just to spin up a working pipeline
- ✕ "Open" architectures that still trap your data in their UIs

That's why 5X isn't just a collection of features. It's a rethink of what modern teams need to succeed.

What makes 5X different (and why it matters)

The 5X principle	What it means	Why it matters
Outcome-first	Infrastructure aligned to business use cases.	Launch securely from any starting point—dashboard, pipeline, or AI app.
Speed-to-value	Live in days, not quarters.	Prove value fast, even if you're early in your data maturity.
Open & modular	Built on open-source standards: dbt, Superset, etc.	No lock-in, no black-box components
Zero lock-in	BYOC or on-prem; full exportability.	You're never trapped; leave without losing your pipelines or data models.
Managed + supported	One contract. One support team. One outcome.	No ticket queues or boilerplate responses.

So who is 5X for?

You don't need a 200-person data team. You need a clear path to business outcomes and a partner that meets you where you are. Here's who thrives with 5X:

CIOs / Heads of Data
Looking to reduce vendor sprawl, unify tooling, and stay compliant without slowing down.

VPs of Analytics / Data Platform Leads
Want dbt-native pipelines, plug-and-play orchestration, and no infra maintenance.

Go-to-Market & RevOps
Care about data as outcomes: dashboards, syncs, apps, not warehouse architecture.

AI/ML Leaders
Want rapid experimentation and deployment with RAG/LLM support built-in.

Real teams, real outcomes

You don’t need to build the perfect stack from Day 1. But you do need a secure foundation that won’t punish you for evolving.

With 5X, you can start lean, scale when you're ready, and still sleep at night knowing you’ll never have to re-platform in a year.

See what your stack would look like on 5X.

Book a demo