

SURVIVAL

OF THE **DATA-FITTEST.**

A guide for developing data strategy **with muscle.**

Get data-fit. **Pentaho+**

Take a breath



If you find yourself trying to move increasingly *massive* amounts of data around your business in increasingly *minuscule* amounts of time, this guide is your sign to take a breath.

If you're finding quick-fix point solutions aren't delivering data-fitness – and are, in fact, slowing you down – you're in the right place.

We're Pentaho. People who have been around long enough to recognize real data-fitness: hard-won, hard-earned and rock solid. Long enough to have honed our technologies and strengthened our solutions to handle the challenges ahead.

**How you perform tomorrow,
is how you prepare today.
Read on to see how we can help.**

**Get data-fit.
Get Pentaho+.**

044 16.00 > W0116 > D1 > 000000016XX 105 11
8304-649305045 =
12305 0L 16 77 99%
F135506558195045: FGFSD > 16 77 90%
URLWEB DOCUMENT:
LEFT_COUNT_SIGN=0

<DIV CLASS=LEFT_SETTINGS onclick=“MENUSETTINGS(0)”>
<DIV CLASS=LEFT_SETTINGS onclick=“MENUSETTINGS(0)”>
000000000000000002541
7535-3325-2305-1105-4051
GLK-OFF //

CLASS= LEFT_COUNT_WRAP_FL_R
LEFT_VOID><SPAN
CLASS=INL_BL
LEFT_COUNT_SIGN=0

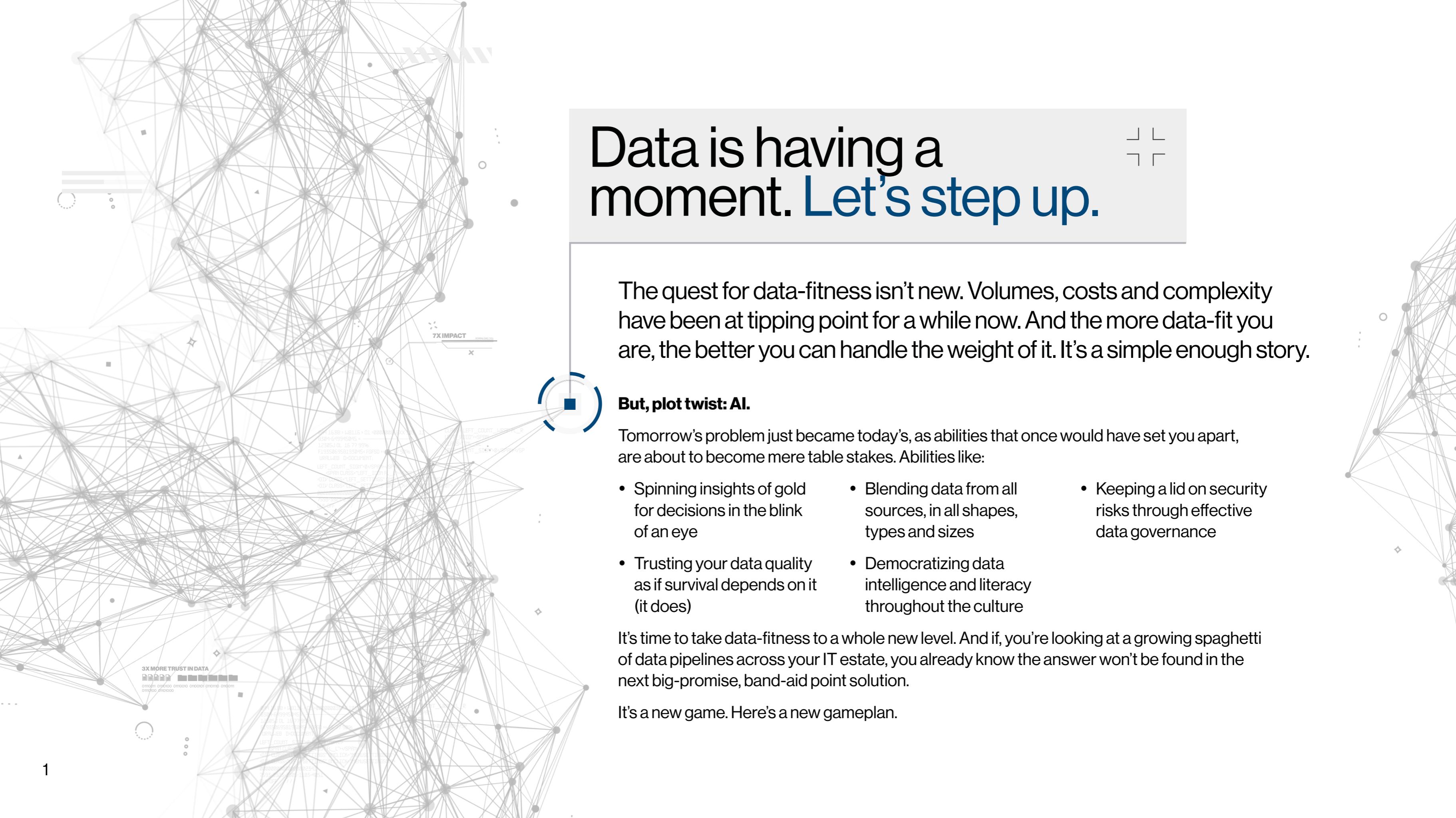
<DIV CLASS=LEFT_SETTINGS onclick=“MENUSETTINGS(0)”>
000000000000000002541
7535-3325-2305-1105-4051
GLK-OFF //

Contents



- Data is having a moment. Let's step up. **1**
A new game needs a new gameplan. **2**
Strength: lift the weight of data. **3**
Precision: sweat the small stuff. **6**
Agility: spot the gap, seize the moment. **9**
Efficiency: more race, less drag. **12**
Speed: faster out of the blocks. **15**
How data-fit are you? **17**





Data is having a moment. Let's step up.

The quest for data-fitness isn't new. Volumes, costs and complexity have been at tipping point for a while now. And the more data-fit you are, the better you can handle the weight of it. It's a simple enough story.

But, plot twist: AI.

Tomorrow's problem just became today's, as abilities that once would have set you apart, are about to become mere table stakes. Abilities like:

- Spinning insights of gold for decisions in the blink of an eye
- Blending data from all sources, in all shapes, types and sizes
- Keeping a lid on security risks through effective data governance
- Trusting your data quality as if survival depends on it (it does)
- Democratizing data intelligence and literacy throughout the culture

It's time to take data-fitness to a whole new level. And if, you're looking at a growing spaghetti of data pipelines across your IT estate, you already know the answer won't be found in the next big-promise, band-aid point solution.

It's a new game. Here's a new gameplan.

A NEW GAME NEEDS A NEW GAMEPLAN

One glance at the data landscape today and you'll see plenty of quick fixes within data management and data ops, each offering self-contained, short-lived boosts of the skin-deep kind. But real fitness isn't *how do I look today*, it's *how do I feel today*.

That's why at Pentaho we focus on foundational data-fitness – the kind that runs wide and deep through your people, processes and technology – the kind that gets you firing on all cylinders and performing at the top of your game.

We measure foundational data-fitness against five pillars, both capabilities and outcomes:

Strength

Your ability to make light work of heavy data volumes and varieties from all sources without overwhelming the system.

Precision

Your ability to observe and trust the quality, completeness, and accuracy of your data; access to breadth, depth and granularity.

Agility

Your ability to adapt in response to change, threat or opportunity, quickly and simply, free from high overhead or vendor lock-in.

Efficiency

Your ability to manage and scale your data estate with economy and cost-certainty, getting maximum return on investment.

Speed

Your ability to be quicker out of the blocks across all key performance indicators; time to value, innovation, and growth.



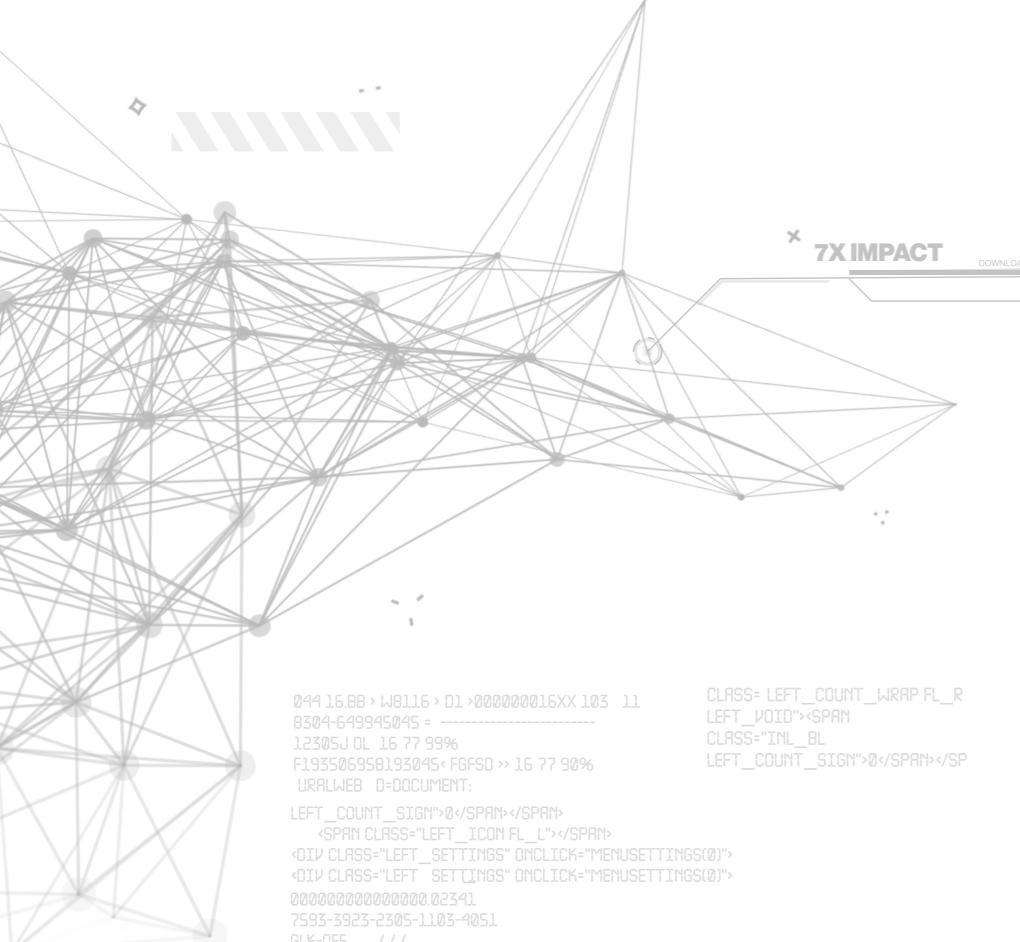
STRENGTH

Lift the weight of data

The great data irony.

The very thing that should make businesses light on their feet is what's weighing them down most of all. Data-fit strength allows you to shoulder the burden of high-volume, high-complexity data and move faster – not despite it, but because of it.

Strength is the foundation of data-fitness – and by strength, we don't mean the quick-win, flex-in-front-of-the-mirror, look-good-in-a-t-shirt kind. We mean foundational strength.



STRENGTH



Let's compare the capabilities of those enterprises that have it (Contenders) with those that don't (Pretenders).

Contenders

- ✓ Automatically ingest and curate any source of data into their data estate.
- ✓ Handle new data or formats when they need it with little or no coding.
- ✓ Handle massive data volume and variety without relying on expensive specialist solutions.

Pretenders

- ✗ Labor under a heavy manual workload to handle different data formats.
- ✗ Spend a lot of time stitching together pipelines to accommodate a series of point solutions.
- ✗ Bring in costly bits of kit each time they want to scale up capacity or handle new data types.

STRENGTH

How to get data-fit strength: 3 practices

Strength is in consistency

No matter how good they are, point solutions can only ever offer strength in one area. And with or without costly integration or coding, you'll end up with a series of weak spots or gaps where disparate solutions cannot communicate with each other. Choose modular data platforms where you can add in new functionality as you need it.

Don't be someone else's beta tester

A lot of data intelligence tech is still in development when you buy it with irregular patches to fix what they forgot. You won't get sure footing when engineers are playing catch up with the tools you gave them. Choose tried-and-tested, battle-hardened, enterprise-grade solutions, for foundations you can rely on from day one.

Go with the word on the street

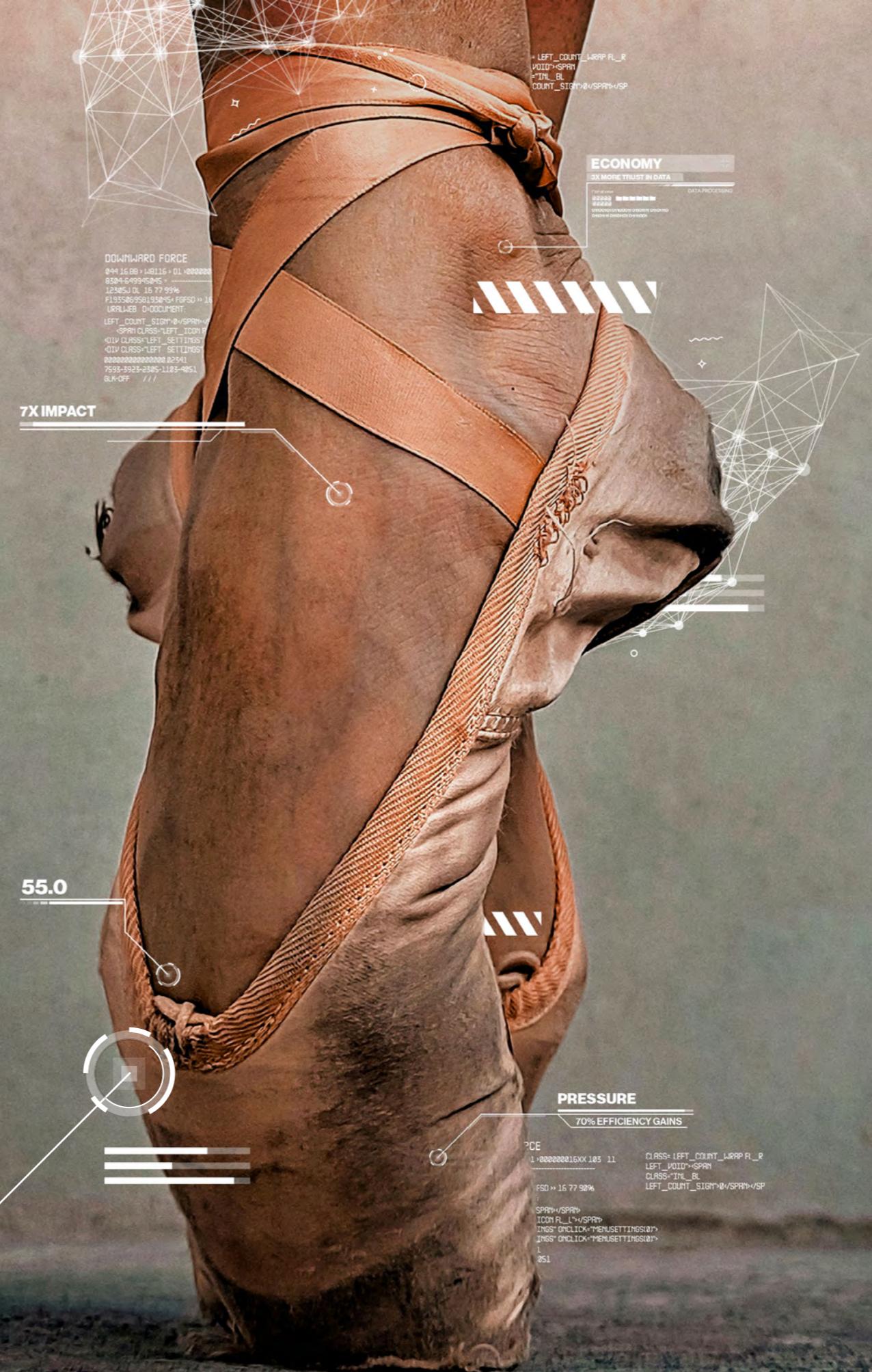
If it's hard to choose between different vendors, look for those with credentials in sectors like financial services and government. You can be sure that enterprises in these verticals have gone above and beyond with due diligence on the stability, reliability and compliance-compatibility of any data platforms they're investing in.

PRECISION

Sweat the small stuff

The ability to make split-second decisions under pressure doesn't make you data-fit. The ability to do so *with confidence* makes you data-fit. And that comes down to trust in data quality. There's only one thing more dangerous to a business than *no* data: *unreliable* data.

When people are moving at speed they don't have time to double-check every report, every analysis, every number. But when your people doubt your data, they slow down. And everything slows down with them, including the pace of decisions, transformation, growth, and innovation.



PRECISION

Here's the difference between data-fit Contenders and data-fit Pretenders:

Contenders

- Organize data with a trusted AI/ML-enabled catalog for automated quality monitoring.
- Have full observability of their data and pinpoint problems with rapid root cause analysis.
- Know the upstream and downstream lineage of their data with policy-driven controls.

Pretenders

- Struggle to back up business decisions with data, even with manual checks.
- Have their engineers manually checking for defects, bottlenecks and problems.
- Aren't sure which users/AI have touched which PII, exposing the business to compliance risk.

There's only one thing more dangerous to a business than *no data*: *unreliable data*.

PRECISION

How to get data-fit precision: 3 practices

Quality starts at source

Most data coming into a business is ‘dark’: dirty, non-standardized and lacking context. The upfront effort to make it usable swallows a lot of resources. Look for tools with AI-driven discovery and classification of structured, semi- and unstructured data for quicker, more accurate onboarding and a foundation for quality.

Trusting the data means understanding the data

Labeling is an under-estimated mission-critical data management capability. Look for tools that can be set up to label data accurately and automatically from any source, using not just a generic business glossary, but your business glossary, built from terms your execs and LoB managers are already using.

Let the tech take care of the detail

Integrated platforms, or modular product suites, will always offer a tighter net for catching data quality defects. For example: your catalog flags issues with a particular data source and automatically kick-starts a cleansing job. The more manual this process, the more room for human error.

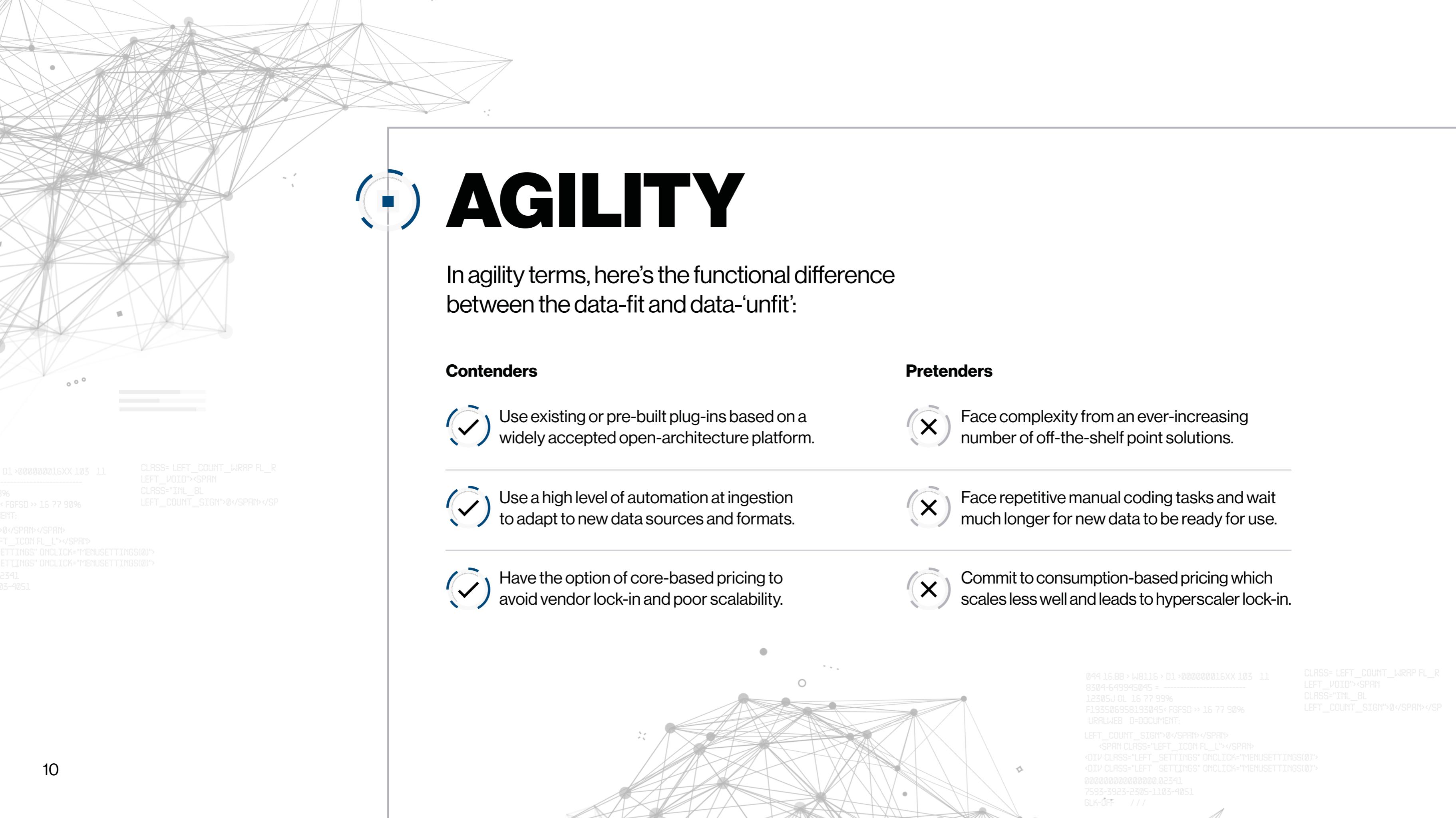


AGILITY

Spot the gap,
seize the moment

You don't know how data-fit you are until you step up to compete. When you do, agility is your ability to respond in the moment to unforeseen challenges or opportunities such as new business requirements, competitive pressure, regulatory change, or emerging data sources. This strategic flexibility is a key measure of data-fitness.

Data management sits at the heart of agility, notably how well you are able to take on fresh data sources and enable quick access and rapid analysis for informed and confident decision-making.



AGILITY

In agility terms, here's the functional difference between the data-fit and data-'unfit':

Contenders

-  Use existing or pre-built plug-ins based on a widely accepted open-architecture platform.
-  Use a high level of automation at ingestion to adapt to new data sources and formats.
-  Have the option of core-based pricing to avoid vendor lock-in and poor scalability.

Pretenders

-  Face complexity from an ever-increasing number of off-the-shelf point solutions.
-  Face repetitive manual coding tasks and wait much longer for new data to be ready for use.
-  Commit to consumption-based pricing which scales less well and leads to hyperscaler lock-in.

AGILITY

How to get data-fit agility: 3 practices

The more embeddable, the more value

Choose API-friendly platforms that allow you to easily create, customize and deploy analytics in a variety of environments in your own enterprise (executive reporting dashboards) and embed them externally (in customer-facing websites or applications).

Serve up insights sooner

You've onboarded a new set of data sources and you want to be able to start squeezing value from them. Look for an ingestion process that quickly reads metadata at scale across similar-format sources. Your data will be business-ready much sooner, and it'll leave you with a simpler set of pipelines.

Look for no-code/ low-code tagging

If you've got a complex data source coming in with thousands of fields, it takes a lot of dev time to go through and label them all. Choose data catalog tools that use AI and ML to automate labeling and integrate your data much faster. But don't trade agility for quality: be sure to ask vendors for their accuracy stats on this.

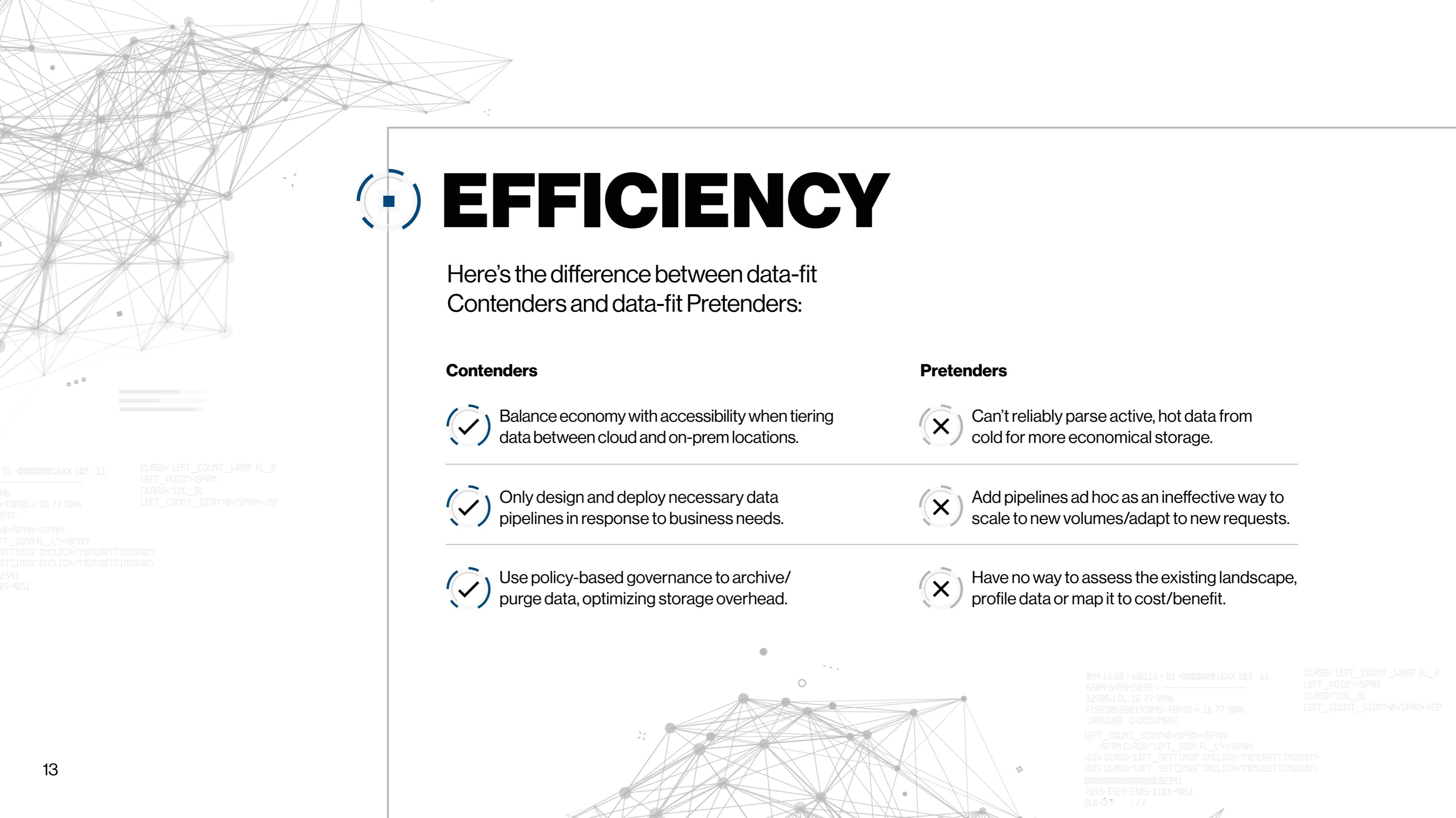
A dynamic photograph of a swimmer performing the butterfly stroke. The swimmer's body is partially submerged in clear blue water, creating a spray of droplets. Overlaid on the image are several futuristic data visualization elements. In the upper left, a circular performance metric shows "5.45". To the right, a network graph with nodes and connections is labeled "7X IMPACT". In the lower left, another circular metric shows "70% EFFICIENCY GAINS". On the right side, a rectangular box is labeled "RESILIENCE" with the subtext "3X MORE TRUST IN DATA" and "DATA PROCESSING", featuring binary code at the bottom. The background is a dark, moody blue.

EFFICIENCY

More race,
less drag

Peak performance economy comes when an athlete expends the minimum level of energy to produce the maximum level of performance. Efficiency in data-fitness is the same: optimizing your data investment to produce the best business return.

Clogged with point solutions and spiraling storage and compute costs – efficiency is a priority for most businesses, as are tools that give them back control over the cost of storing and moving their data – managing high volume and high throughput *without* suffering latency issues.



EFFICIENCY

Here's the difference between data-fit Contenders and data-fit Pretenders:

Contenders

- ✓ Balance economy with accessibility when tiering data between cloud and on-prem locations.
- ✓ Only design and deploy necessary data pipelines in response to business needs.
- ✓ Use policy-based governance to archive/purge data, optimizing storage overhead.

Pretenders

- ✗ Can't reliably parse active, hot data from cold for more economical storage.
- ✗ Add pipelines ad hoc as an ineffective way to scale to new volumes/adapt to new requests.
- ✗ Have no way to assess the existing landscape, profile data or map it to cost/benefit.

EFFICIENCY

How to get data-fit efficiency: 3 practices

Automate data lifecycle management	Look for scalable licensing/core-based pricing	Use a data-tiering storage strategy
<p>There's an ongoing conflict of interest in data optimization: business users decide on data value, but IT manages the storage and foots the bill. Resolve it – and reduce costs – with an automated policy-driven solution that takes both aspects into account as well as data risk, sensitivity and regulatory requirements.</p>	<p>Consumption/seat-based pricing models are the norm for SaaS but when it comes to foundational data intelligence tools, you don't want them limited to a few data engineers. You want them out there in as many business hands as possible if you're going to create a self-service data democracy – and keep a lid on costs.</p>	<p>Choose tools that use business rules and policies to automatically identify, tag and move data to the correct location according to cost, risk and sustainability criteria. Efficiency lies in your ability to quickly offload, archive, and move data to wherever it needs to be available for apps or analysis.</p>



SPEED

Faster out of
the blocks.
Performance
unlocked.



SPEED

The ultimate outcome for data-fitness. Businesses that develop their data-fitness by building strength, precision, agility and efficiency find across-the-board acceleration towards strategic performance indicators including:

Expedited time-to-insight

High-quality, well-prepared data, timely analysis and actionable intelligence enable stakeholders to make nimbler (and better) decisions, reducing decision-making cycles and enabling faster responses to opportunities and threats.

Expedited time-to-innovation

By fostering a culture of data-driven experimentation, data leaders can access relevant and trusted data sets efficiently, test hypotheses, and iterate rapidly to develop new products, services, and business models.

Expedited time-to-value

Using automation to streamline data processes, ingest, organize and make data readily accessible to users, businesses can develop a collaborative, data-literate data democracy, unlocking performance across the workforce.

How data-fit are you?

How many of these data-fitness capabilities do you currently have? Are you able to:

Automatically ingest and curate any source of data into your data estate?

Layer on new functionality when you need it with little or no hand-coding?

Strengthen data literacy with business glossaries and easy-to-use, self-service?

House data in a trusted AI/ML-enabled catalog automated for quality monitoring?

Have full observability of your data and pinpoint problems with rapid root-cause analysis?

Know the lineage of your data with policy-driven controls?

Use custom-built plug-ins based on a widely accepted open architecture platform?

Incorporate a high level of automation in your data ingestion and labeling process?

Connect to any storage environments for streamlined data flows across hybrid landscapes?

Balance economy with accessibility when tiering data between cloud and on-prem locations?

Rapidly deploy data pipelines to effectively scale up to business requirements?

Use policy-based governance to archive and purge data, optimizing infrastructure expenses?

Your score

Give yourself one point for every item you checked:

1–4: Indicates poor data-fit. Let's prioritize data quality checks, cleansing at ingestion, along with basic governance practices. We can begin by documenting your current data processes and workflows to lay the groundwork for improvement.

5–8: Suggests some data-fit. Next steps would be to implement automated data quality, metadata management and lineage tracking. Let's also look at enhancing your governance with clear roles, responsibilities, and processes.

9–12: Good levels of data-fit. You can benefit from advanced analytics for insights and comprehensive data governance. We can help you foster a culture of continuous improvement in data management.

› **How did you score? We'd love to help you reach your data-fit goals. Visit the website today.**

pentaho.com



Hitachi® and Hitachi Vantara® are registered trademarks of Hitachi, Ltd. in the U.S. and other countries. Pentaho® is a registered trademark of Hitachi Vantara LLC in the U.S. and other countries.

Pentaho+ is a data intelligence platform consisting of four tried-and-tested products: *Pentaho Data Catalog*, *Pentaho Data Optimizer*, *Pentaho Data Integration* & *Pentaho Business Analytics*. Together these solutions form a reliable, modular, and scalable platform, *Pentaho+*. It's for businesses who want to develop data-fitness muscle and gain the performance benefits that come with it.

Unlike point solutions that solve only part of the problem, Pentaho+ connects right into your existing environment and allows you to decide where to start without costly integration or coding. Get in touch today to unlock the power of your people – data practitioners and business leaders alike – by giving them simple tools and direct access to mission-critical insights, driven by 100% trusted data.

www.pentaho.com

© 2024

Get data-fit. **Pentaho+**