# Shortlisting a Data & ML Vendor

A Procurement Brief

Datahike · October 2026 · contact@datahikeanalytics.com

*Public briefing for buyer and supplier-management teams. It is not a product catalog, not a security questionnaire, and not legal advice. Use it to force comparable evidence before anyone signs a statement of work.*

## 1 Why vendor shortlists fail

Feature lists look objective and are not. Two vendors can both claim "enterprise MLOps," "real-time pipelines," and "generative AI" while delivering incompatible things: a notebook that never leaves a laptop, a platform that needs a year of integration, or a model that cannot be explained to an auditor.

Shortlists fail for four recurring reasons.

1. **Outcomes are missing.** The RFP asks *what the stack can do*, not *which decision will change* if the work succeeds.
2. **PoCs are not comparable.** Each bidder chooses a dataset, a metric, and a clock. Procurement is left with three demos and one story each.
3. **Delivery is treated as a later problem.** Integration, ownership, and what happens when the champion leaves sit in an appendix — or never.
4. **Risk language is decorative.** "LGPD-ready" and "secure by design" appear without a claim you could test: where data lives, who can see it, how a model is retired, how a failure would be described to the board.

A usable shortlist is a table of *evidence*, not a table of *logos*. The rest of this brief is a way to build that table in a few pages.

## 2 Must / Should evidence

Score every bidder on the same three columns. Use Must for deal-breakers; Should for differentiators you will actually pay for.

| Lens | Must (examples) | Should (examples) |
|---|---|---|
| Capability | Names the business question; shows a baseline vs proposed method; states the metric and the failure cases that matter to operators. | Relevant prior pattern (anonymized); ability to say “this will not work” with a reason. |
| Delivery | Named owner after week four; handoff note a maintenance team can read; time box and frozen data extract. | How the next source system would be added without a rewrite. |
| Trust & posture | Written statement of where data is processed; who can access it; how personal data is minimized (LGPD/GDPR *as a claim you can ask about*, not a badge). | Independent references who will talk about a missed deadline, not only a win. |

*Do not* treat certifications, slide counts, or model-name lists as Must items unless your legal team has already made them contractual.

## 3 Questions to put in an RFI

Ask the same questions of every vendor. Freeze the answers in writing.

**Question (capability)**

- Which decision of ours changes if your work succeeds? Answer in one sentence.
- What baseline will you beat, and with which metric? Who owns the metric definition?
- Show the last time you were wrong in production. What did operators see?

**Question (delivery)**

- Who is on the account in weeks 1, 4, and 12? What happens if that person leaves?
- What do you need from us that is not in the extract (access, people, exceptions)?
- What will we still be able to run if we stop paying you in month six?

**Question (trust)**

- Where is our data stored and processed? Which subprocessors?
- How is personal data minimized, retained, and deleted on request?
- Who can prompt or query a model that has seen our data?

Refuse oral-only answers. A vendor who cannot write this down will not survive your second audit.

## 4 How to read a PoC proposal

A proposal that can be compared has six parts. If any is missing, send it back.

1. **Question** copied from your brief, not rewritten into a product pitch.
2. **Data** — the extract they will use, including known defects.
3. **Method** — simple enough that a skeptical engineer can sketch it.
4. **Metric and baseline** — including how ties and missing labels are handled.
5. **Clock** — calendar dates, not “sprints.”
6. **Exit** — what you keep (code, docs, trained artifacts, nothing).

Red flags: a demo environment you cannot inspect; a metric invented after seeing the test set; a requirement that you buy the platform before the pilot starts; unbounded “discovery” hours; humor at the expense of the operators who will live with the result.

Green flags: they ask to shrink the question; they price a stop/go after two weeks; they document assumptions in a table you can print.

## 5 One-page scorecard

Copy this table into a spreadsheet. One row per vendor. Score 0–2 (missing / partial / evidenced). A Must of 0 is a no.

| Item | A | B | C |
|---|---|---|---|
| Business question restated without product language | | | |
| Baseline named and beatable | | | |
| Metric owned by a named role on our side | | | |
| Frozen extract accepted (no vendor-curated sample) | | | |
| Time box $\leq$ 8 weeks with a mid-point stop/go | | | |
| Handoff artifact listed in the exit clause | | | |
| Data location and access written, not implied | | | |
| Personal-data minimization described as a procedure | | | |
| Named delivery owner through week 12 | | | |
| Reference who will discuss a miss, not only a win | | | |
| **Sum (Must items first)** | | | |

Total the Must rows first. Do not average them with Should rows. If two vendors tie on Must, use Should and price. If none pass Must, buy nothing.

## 6 A worked shortlist (anonymous)

Three bidders, one question: *Can we forecast weekly demand well enough to cut emergency purchases by 15% in two distribution centers within one quarter?* Frozen extract: two years of orders, stock-outs flagged, no customer PII. Clock: six weeks, stop/go at week three.

| | **Vendor A** | **Vendor B** | **Vendor C** |
|---|---|---|---|
| Question | Restated cleanly. | Rewrote into a platform pitch. | Restated, then added a second question you did not ask. |
| Baseline | Last year's buyer rule. | "Industry benchmark" with no source. | Seasonal naive forecast, documented. |
| Extract | Accepted defects. | Asked for a "gold" sample. | Accepted; listed three quality risks. |
| Must sum | Pass. | Fail (baseline + extract). | Pass. |

Result: B is out on Must. A and C both pass. C is more expensive and more honest about data quality. Procurement can now argue price against evidence instead of against slide count. That is the whole point of this brief.

## 7 Anti-patterns in data and ML RFPs

- **The kitchen-sink annex.** Fifty "nice to have" integrations, none tied to the decision. Vendors pad; you cannot score.
- **The live bake-off on production.** You leak operations and still cannot compare methods because the clock and the data moved.
- **The unpaid marathon.** Six-week "free" PoCs select for firms that can afford theater, not for firms that can hand off.
- **The metric after the demo.** Accuracy quoted on a set the vendor chose after seeing results. Demand the split in the RFI.
- **The license before the lesson.** Platform SKUs as a prerequisite for a four-week look. That is a sale, not a pilot.
- **The missing owner.** "A blended team" with no named person after week two. Delivery risk is then yours, silently.

Write the inverse of each anti-pattern into the RFI as a constraint: one question, frozen extract, paid time box, metric declared up front, no license gate, named owner.

## 8 What “good enough” looks like at week three

A stop/go review should be boring. You want a document, not a keynote.

1. The question is still the question (or a documented shrink).
2. Baseline vs candidate on the agreed metric, with a plot or table of errors on the operator-relevant slice — not a global average that hides the expensive misses.
3. A list of assumptions that broke, in plain language.
4. A recommendation: continue, stop, or re-scope — with a cost to finish.

If week three is only architecture diagrams, you are not in a pilot. You are in a pre-sales cycle. Stop or rewrite the brief.

## 9 Security and personal-data posture — as claims

You are not running a full assurance audit in a six-week look. You *are* allowed to demand claims you can later test.

Ask, in writing:

- Where is data processed (region, subprocessors)?
- Is personal data required for the question? If yes, what is the minimum set, retention, and deletion path?
- Who can query a model or notebook that has seen the extract?
- How is the extract destroyed at the end of the pilot?

Treat “we are LGPD-compliant” as incomplete until those four answers exist. If the vendor will only discuss certificates, involve your privacy counsel before you share data — not after a demo.

## 10 Pricing the comparison without buying a platform

Pay for the clock, not for hope.

- A fixed fee for the time box, payable on delivery of the week-three packet even if you stop.
- No consumption surprises (GPU, tokens, connectors) outside a cap written in the purchase order.
- An optional paid next step with its own question — not an automatic renewal of the same theater.

Vendors who cannot quote that structure are telling you they sell licenses, not answers. That can be fine for a later phase. It is the wrong instrument for a first shortlist.

## 11 Sample cover note for the RFI

*Subject: Frozen question and extract — data/ML shortlist (time-boxed)*

Please answer the attached RFI against the question in section 1. Do not substitute a different question. Use only the extract described in section 2. Proposals that require a platform purchase before week three, or that replace the extract with a vendor sample, will not be scored. Delivery of a week-three packet is the acceptance criterion for this fee. Email written answers to the procurement contact on the cover sheet.

You may copy that paragraph. Length is a feature.

## 12 Glossary for mixed buyer–engineering rooms

**Baseline**
The thing you already do (or a simple model) that a vendor must beat. If there is no baseline, you are not measuring.

**Extract**
A frozen copy of data for the pilot. Changing it mid-flight resets the comparison.

**Grain**
The row meaning (store-week, invoice-line, sensor-hour). Mixing grains is how metrics become fiction.

**Handoff**
Artifacts a team that was not in the pilot can run or retire.

**Must / Should**
Deal-breaker vs differentiator. Never average them.

**Stop/go**
A scheduled decision to continue, stop, or shrink — with evidence, not enthusiasm.

## 13 Ninety days after you pick a vendor

The shortlist is not the project. If you award a next step, freeze a second brief on day one of that step:

1. Days 1–15: confirm extract, access, named owner, week-three packet date.
2. Days 16–45: run the method against the frozen question; no new scope without a written shrink or expand.
3. Days 46–60: handoff draft, operator walkthrough, retire/keep decision.

4. Days 61–90: either a paid production path with its own contract, or a clean stop with artifacts listed in the exit clause.

If day 90 is “we are still aligning on the use case,” the first brief failed. Restart from the question — do not buy more hours of alignment.

## 14 Extended RFI bank (optional pages)

Use only what you will score. Extra questions that nobody grades become noise.

### Capability

- What would make you decline this work? Write two sentences.
- Which operator role must sit with you weekly, and for how many hours?
- If labels are delayed or wrong, how does the metric change?

### Delivery

- Show a one-page plan for adding a second source system after the pilot.
- What do we own if three of your people change in a month?
- How do you version the notebook or pipeline we will inherit?

### Commercial

- Fixed fee for the time box, including the week-three packet if we stop.
- Cap on usage costs, in the purchase order, not in a slide.
- Price of a named next question (not a blanket retainer).

## 15 What to send legal before any extract leaves the building

A one-page memo is enough:

- The business question and why personal data is or is not required.
- The extract description (tables, grain, date range, known defects).
- Vendor answers on location, access, subprocessors, and destruction.
- The stop/go date and the exit clause (what is deleted vs returned).

If legal cannot review that memo in the same week you send the RFI, you are not ready to share data. Delay the extract. Do not delay the written questions — vendors can answer capability and delivery without your files.

## 16 Facilitation notes for the scoring meeting

Room: procurement, one operator, one engineer, optional privacy. Ninety minutes. No vendor in the room.

1. Ten minutes: read Must scores aloud. Any 0 on Must is a no; do not debate taste.
2. Twenty minutes: remaining Must ties — evidence only, not brand.
3. Twenty minutes: Should differentiators you will actually fund.
4. Twenty minutes: price vs remaining two (or one).
5. Twenty minutes: write the award or the no, including the next question if you proceed.

Capture the no in the file. Future RFPs get faster when "we declined because Must item 4 failed" is written down.

## 17 One-page close-out checklist

Print this with the scorecard.

- Question frozen in one sentence.
- Extract frozen; defects listed; personal data justified or excluded.
- Same RFI sent to every bidder; oral answers rejected.
- Must scores complete; any 0 documented as a no.
- Week-three packet date in the purchase order.
- Exit clause: destroy, return, or keep — named artifacts.
- Award or no written, with the next question if you proceed.
- Email on file: `contact@datahikeanalytics.com` if you want a second reader on the brief (not a demo).

## 18 How to reach Datahike

Datahike is a consultancy in data science, data engineering, data consultancy, applied AI engineering, and delivery leadership. We will not pretend a pilot is a platform sale.

If you have a question a supplier should answer — not a wish list — write to contact@datahikeanalytics.com. Send the decision you need to make, the time box you can defend, and whether you already have a frozen extract. We will tell you if we are the wrong firm.

Companion note (public Insights): *Scoping an ML pilot so procurement can compare vendors*, at `https://datahikeanalytics.com/blog/scoping-an-ml-pilot/`.

Datahike · datahikeanalytics.com · Brief v1 · October 2026. No internal service names, hosts, or staff tools appear in this document because they do not belong on a public procurement surface.