

Databricks in *Plain English*

A short guide for executives — no technical background needed



What is it?

Databricks is **one secure place where a company gathers all of its information and turns it into answers** — accurate reports, forecasts, and smart tools that spot patterns for us. Think of it as a central library with a skilled research team inside it, instead of papers scattered across many filing cabinets.

The problem it solves

- **Information is scattered** across finance, sales, operations and customer systems, and across spreadsheets.
- **Numbers don't always agree.** Two teams ask the same question and get two different answers.
- **Reports are slow and manual.** People spend days copying and cleaning data before anyone can use it.
- **Looking ahead is hard.** Most tools explain what already happened; few help predict what comes next.

One home for all our data: the “data lake”

Today our information lives in many places. Databricks connects to all of them and brings a copy into **one central store, called a data lake**. Picture a real lake: many streams flow in, and everything ends up in one place where it can be used. Nothing has to be forced into a rigid shape first.



What is “unstructured” data? There are two kinds of information in every business:

Structured — neat rows and columns	Unstructured — everything that does not fit in rows and columns
Like a spreadsheet: invoices, sales figures, policy lists, payments, stock levels.	Emails, contracts, PDFs, scanned forms, photos, call recordings and chat messages.

Most of a company's information is unstructured, yet ordinary reporting tools can only use the first kind. Databricks keeps **both kinds together**, so a customer's record can be read alongside their emails, documents and call notes — and AI can read those documents for us.

Databricks as a “data kitchen”: preparing the data

Raw data is like raw ingredients: it must be washed, sorted and prepared before it is useful. Databricks does this preparation (called **data transformation**) **automatically, on a schedule**, instead of people doing it by hand in spreadsheets.

1. Raw	2. Cleaned	3. Ready to use
Data arrives exactly as received, untouched — a safe original copy.	Duplicates removed, dates and names made consistent, the same customer matched across systems, errors flagged.	Business rules applied once (for example, how “revenue” is calculated) so every report uses the same figures.

Technical teams call these stages Bronze, Silver and Gold. The benefit is one reliable preparation line instead of several separate tools and manual steps.

Data governance simply means clear rules about who may see which data, and who is accountable for it. Think of an office building with key cards: staff can enter only the floors they need, and a visitor log records every entry. Databricks applies the same idea to data, in one place, for people and for AI alike.

For people

- **Access by role.** Finance sees finance figures; salaries or personal details are limited to those who need them.
- **Sensitive details hidden.** For example, an ID number can show as ***** to most users.
- **Everything is recorded.** We can see who looked at what, and when.
- **A “family tree” for every number.** We can trace where a figure came from and how it was calculated.
- **A searchable catalogue.** Each dataset has an owner and a plain definition, so people trust what they find.

For AI tools

- **Same rules as the person asking.** An AI assistant can only use data that the person asking is allowed to see.
- **Approved data only.** AI works from our prepared, labelled, quality-checked data, so answers are more reliable.
- **Stays under our control.** Our data is processed inside our own secure environment, not scattered across outside AI websites.
- **Every question and answer is logged,** so AI use can be reviewed and audited.
- **One set of rules.** We do not need separate security for people and for AI.

What it does for the business

- **One trusted version of the truth** — everyone works from the same numbers.
- **Faster answers, better decisions** — reports refresh automatically, with forecasts and early warnings (such as which customers may leave).
- **Ready for AI, and grows with us** — governed data lets AI be used safely, and we pay only for what we use.

“Why not just use a reporting tool like Power BI or Tableau?”

A reporting tool is the **dashboard in the car**: it shows speed and fuel level. Databricks is the **engine and fuel line**: it collects, cleans and protects the information so the dashboard has something reliable to show. They do different jobs, and work well together.

Question	Reporting tool alone	Databricks (with a reporting tool on top)
Where does the data come from?	Needs data already gathered and tidied by someone else.	Gathers and cleans data from all our systems automatically.
Can it handle emails, PDFs and recordings?	Not really; it expects rows and columns.	Yes — stores and reads both kinds of data.
Do all teams get the same numbers?	Often not; each report may be built its own way.	Yes — one shared, controlled source.
Who may see what, and what about AI?	Basic access settings inside each report.	One set of rules for all data, people and AI.
Can it predict the future?	Mostly shows the past.	Built for forecasting and AI.

Fair note: if a company only needs a few simple reports from one or two systems, a reporting tool alone is enough. Databricks earns its place when data comes from many sources, volumes are growing, and leaders want trusted forecasts and AI.

The bottom line: Databricks does not replace our reporting tools — it makes them trustworthy and lets us go further. It turns data from a pile of paperwork into a dependable, well-protected asset that supports faster, better-informed decisions.