Engineer

Data analytics

processing data systematically to find patterns and trends

What it is

Turning recorded facts into answers someone can act on is the job of data analytics, often just called analytics. In the big data framework published by NIST, the US standards institute, analytics means working through data methodically, reshaping it where needed, to bring out patterns, links between data points and trends over time, and to try to predict what happens next. NIST places that step inside a wider *data analytics life cycle* that turns raw data into knowledge an organisation can use: collecting, preparing, analysing, visualising and giving access.

It is a kind of data processing, set apart from the transactional kind by what it reads: an application writes one order at a time, while an analytical query reads millions of rows to return a handful. It also differs from data engineering, which builds the pipelines and storage that analytics reads from.

The word covers quite different work:

- **Reporting**: the same figures delivered on a schedule, as a PDF or an email.
- **Dashboards**: those figures kept live in a business intelligence tool such as Power BI.
- **Ad-hoc analysis**: a SQL query written for a question nobody asked before, often against a warehouse such as BigQuery.
- **Geospatial analysis**: the same work over map data in a GIS, asking where as well as how much.
- **Analytical applications**: a model or a statistical method behind a form, for people who write no code.
- **Web and product analytics**: counting the visitors to a website or the users of a product, and what they do there.

The last kind carries a privacy cost, and it is the one meant when this site's privacy, cookies, credits and security pages say "no analytics": no analytics tag, no tracking pixel, no cookie and no account with an analytics company. The site still measures itself. Its web server keeps an access log, and a program turns it into daily counts and a monthly ranking of pages read, referring sites and crawlers, kept as numbers and never joined into any one visitor's path.

Use the concept in a decision

Put "What is data analytics?" to work in a real decision. We can help relate the concept to the system you are building or buying. Dashboards and reports showing the handful of figures a decision actually turns on, refreshed without being asked.

Explore our services Get in touch →

Test the assumptions behind the term

Use the explanation and linked work to frame questions for your project. We can examine the assumptions with you before a tool is chosen. Brand communication informed by audience response — brand communication informed by audience response

Brand communication informed by audience response Get in touch →

Why it matters

Two choices settle most of what analytics will cost: where the queries run, and how data about people is collected.

Analysis on the transactional database is the cheapest start: the data is current and nothing is copied, but every report competes with the application for memory and disk. A read replica moves that load off the primary for the price of a second server. A warehouse such as BigQuery keeps storage apart from computing, and on its on-demand pricing bills each query by the bytes it reads, which rewards narrow queries, and it needs a pipeline to copy the data in.

Lock-in hides in the definitions. A BI tool keeps its measures in its own formula language, and a year later the meaning of an active customer lives in one file only that tool opens. Definitions kept as SQL in the repository survive a change of tool.

Web and product analytics adds a legal question. The usual hosted tool works through a tag: a script in every page that sets a cookie or a similar identifier in the browser and reports each page view to the vendor. Article 5(3) of the EU's ePrivacy Directive makes consent the condition for putting information on a visitor's device or reading it back, with two exceptions: storage or access whose sole purpose is carrying a communication over a network, and storage or access strictly necessary to deliver a service the visitor explicitly requested. Once events

can be tied to a person, even through a cookie identifier or an IP address, the GDPR applies too: a lawful basis, a stated purpose, a retention period. And the vendor is another company, now holding a record of what your visitors did.

So the real choice is among a hosted tool that is quick to add and comes with dashboards, a first-party collection you build and run yourself, and counts from logs your server already writes, which answer fewer questions and carry far less.

What it looks like in practice

For a professional networking platform, the company built product analytics inside PostgreSQL in place of a third-party tag: a schema of 7 tables and 47 functions, event tables partitioned with 22 partition definitions that prune themselves on a schedule, and visitor identity pseudonymised behind a rotating salt. The client sends nothing until the visitor has consented, and funnels are computed in the database next to the data, which NIST calls data locality. The page names the price: the rotating salt gives up long-horizon cohort analysis, and no free dashboards come with it.

On the reporting side, a Python reporting system pulled the data, ran the analysis and delivered the same PDF report on a predictable schedule to the people making commercial decisions. The page credits that reporting with raising quarterly software revenue by 400%.

This site is the third example. Its privacy page sets out the measurement: an access log that zeroes the final part of each IP address before writing a line, the referring site kept as a name only, and lines that live about thirty days. A program reads the log into a daily tally and a monthly ranking, both kept indefinitely as counts with nothing that could join them back into one visit. With no tag, no cookie and no third party, the page shows no consent banner and gives the reason as “nothing to consent to”. It is open about the limit too: while a line exists, a rare browser on a small network narrows a visitor down further than the site would like. The trade is fewer answers, and no funnels, for nothing stored on the reader’s device and nothing sent to anyone else.

Where it goes wrong

Two tools report two visitor counts

The hosted tool shows one number for last month and the server log twice as many, each right about something different. A tag counts only browsers that ran its script and, behind a banner, only visitors who agreed; a log counts every request, crawlers included, until someone filters them out. Write each metric's definition down once, in SQL, with what it includes and excludes.

Reports slow the application down

Every Monday the application slows while the weekly report scans whole tables on the transactional database. Move the report to a replica or a warehouse, or roll raw rows into a small summary table each night and run window functions over that:

```
SELECT day,
       signups,
       sum(signups) OVER (ORDER BY day
                          ROWS BETWEEN 6 PRECEDING AND CURRENT ROW)
         AS last_7_days
FROM daily_counts
ORDER BY day;
```

The banner goes and the tracking stays

A team swaps cookies for a "cookieless" product, takes the consent banner down and goes on following visitors. Article 5(3) concerns access to information on the device, whatever the mechanism, and the EDPB's guidelines on its technical scope say that fingerprinting from HTTP headers, or tracking navigation by IP address alone, can fall under it too, while recalling that the rule applying does not by itself mean consent is needed. Judge a tool by whether it links one request to the next, not by whether it sets a cookie.

Consent that regulators would reject

A banner blocks the page until the visitor accepts, treats scrolling as agreement, or makes withdrawing harder than agreeing was. The EDPB's consent guidelines say a cookie wall leaves consent not freely given and that scrolling is never a clear affirmative act, and Article 7(3) of the GDPR makes withdrawing as easy as giving. Give refusing the same weight as accepting, and send nothing before the answer.

Pseudonymous events kept forever

The event table grows by millions of rows a month, and the hashed user IDs in it are treated as anonymous. Recital 26 of the GDPR treats pseudonymised data that extra information could attribute to someone as personal data, and recital 30 names IP addresses and cookie identifiers among the online identifiers whose traces, combined with other information, may be used to profile and identify someone. Set the retention period before the first event arrives, partition by time so an expired month is dropped whole, and keep long-term figures as aggregates without identifiers.

Further reading

- NIST SP 1500-1r2, Big Data Interoperability Framework: Definitions — the definitions of analytics, the data analytics life cycle and data locality that this article starts from.
- Directive 2002/58/EC, consolidated — Article 5(3), the consent rule for anything stored on or read from a visitor's device, with its two exceptions.
- Regulation (EU) 2016/679, the GDPR — the definition of consent, Article 7 on withdrawing it, and recitals 26 and 30 on pseudonymised data and online identifiers.
- EDPB Guidelines 05/2020 on consent — cookie walls, scrolling and withdrawal, each with worked examples.
- EDPB Guidelines 2/2023 on Article 5(3) — why IP-based tracking, fingerprinting and tracking pixels can fall under the same rule as cookies.
- PostgreSQL tutorial: Window Functions — rankings and running totals over related rows, computed inside the database.

Glossary

Choose for your actual constraints

Bring a representative dataset, the questions it must answer and the current bottleneck. We can agree how to measure quality, throughput and recovery.

See what it costs Get in touch →

Make the terminology practical

Ask how “What is data analytics?” applies to your organisation. A concrete example from your current system is enough to start the discussion.

Explore our services

welcome@engineer.company +45 50 10 61 84 Copenhagen, Denmark

Take this page with you

Contact card

Carefully Crafted With Love by D.

CVR 45969703 · Tagensvej 105C, 3. th, 2200 København N

Copyright © 2025 – present · Engineer ApS · All rights reserved.

Printed from https://engineer.company/glossary/analytics/