

Ngobrol Karir Data & AI

Peta Dunia Data

Peran, alat, istilah, dan jalur belajar di bidang data dan AI

Versi 3 Oktober 2026



Penyusun

Cara membaca handout ini



Untuk siapa

Pemula di bidang data, termasuk yang pindah karier dari luar IT.



Isinya

Bagian 1–4: alur data, peran, konsep. Bagian 5–10: alat, belajar, pasar kerja, kamus.



Tanggal data

Angka lowongan per 3 Oktober 2026, bisa berubah.

Label ada di baris oranye di atas judul, sumber rinci di notes.

Tiga label sumber

Data

Hitungan lowongan 3 Okt 2026, atau sumber resmi (O*NET, dokumentasi vendor).

Praktisi

Pengalaman narasumber atau host sesi DA dan DE, belum statistik.

Penyusun

Penilaian penyusun, belum diuji data.

Penyusun

Peta isi handout



1. Perjalanan data

Dari kasir sampai keputusan



2. Peta peran

Siapa memegang lapisan mana



3. Kartu peran

Kerja, alat, dan jalur tiap peran



4. Konsep kunci

Sebelas konsep dalam gambar



5. Landscape tools

Alat per lapisan



6. Roadmap belajar

Tangga belajar per jalur



7. Pasar kerja

Lowongan kerja di Indonesia



8. Era AI

Yang dibantu AI, yang tetap manusia



9. Memilih jalur

Pohon keputusan dan langkah minggu ini



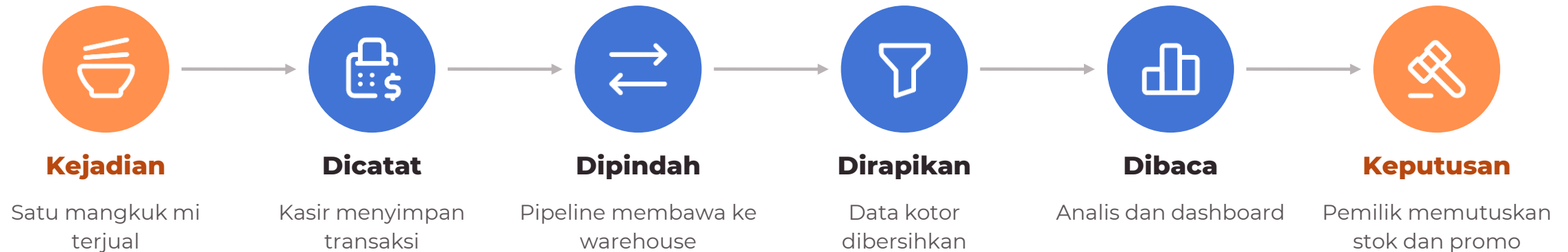
10. Kamus

Istilah, arti, dan contohnya

Bagian 11: sumber dan batasan data.

Penyusun

Dunia data itu perjalanan data, dari **kejadian** sampai **keputusan**.



Setiap peran di dunia data memegang satu potong perjalanan ini. Di sini terlihat potongan mana dipegang siapa, alat apa yang dipakai, dan cara mulai belajar.



Bagian 1

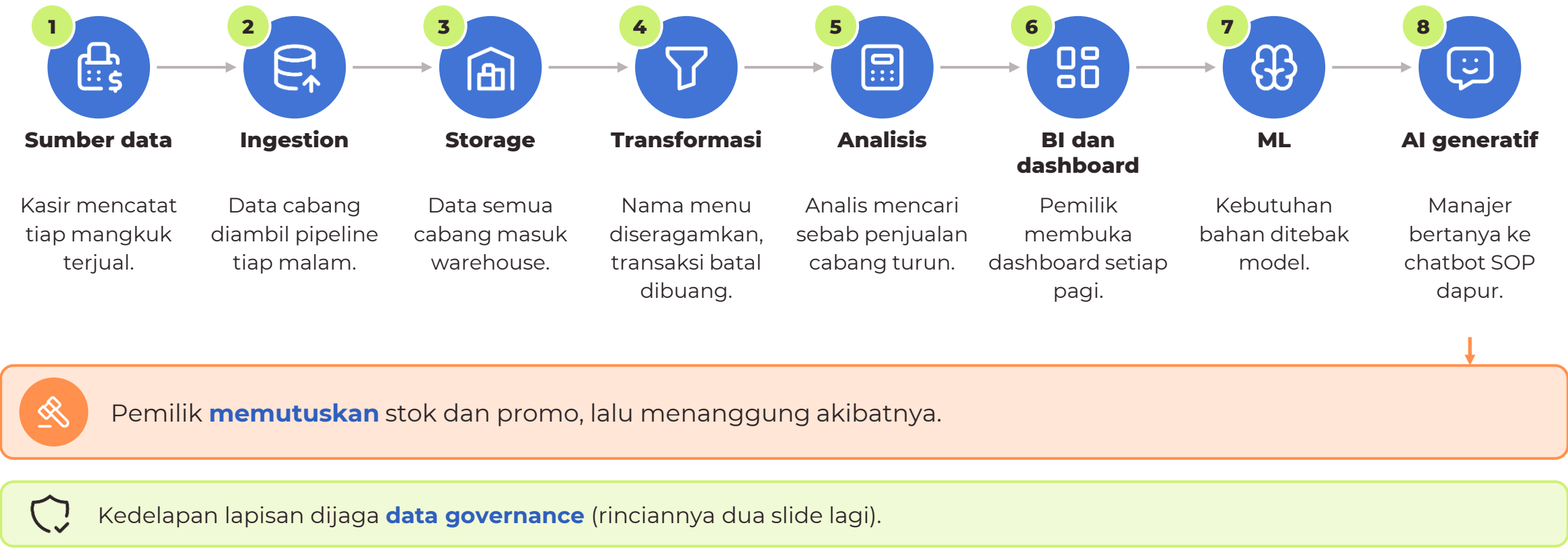
Perjalanan data

Satu cerita dari kasir restoran mi sampai keputusan pemiliknya, dalam delapan lapisan.



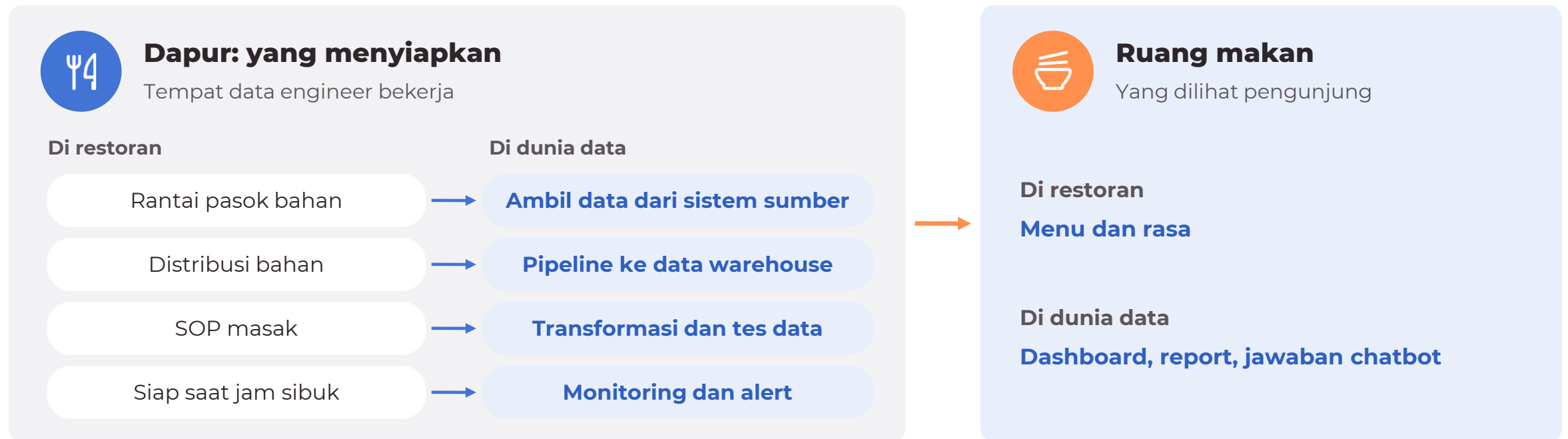
Penyusun • contoh fiktif

Delapan lapisan dari kasir sampai keputusan



Praktisi · sesi Data Engineer

Kerja data yang tidak terlihat pengunjung



Kata narasumber DE Pengunjung hanya melihat menu dan rasa, jadi kalau dapurnya beres, orang tidak sadar pentingnya dapur. Kerja data engineer juga begitu.

Penyusun

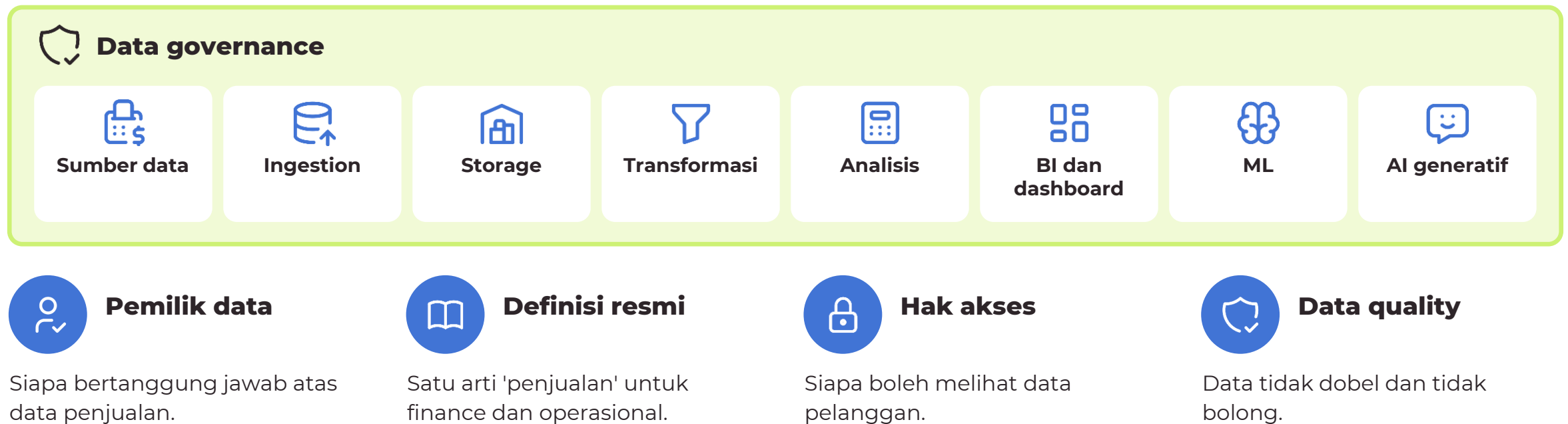
Satu pertanyaan untuk setiap lapisan



Alat di tiap lapisan bisa berganti dari tahun ke tahun; petanya ada di bagian 5.

Data • AWS • Praktisi • sesi DE

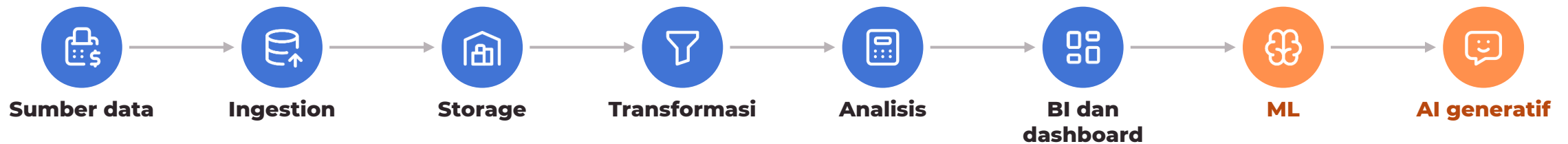
Semua lapisan berada di bawah data governance



Menurut AWS, data governance berisi proses dan kebijakan supaya data tetap layak dipakai bisnis. Narasumber DE: angka finance dan operasional bisa beda karena sudut pandangnya beda.

Praktisi · sesi DA dan DE

Di mana AI masuk ke perjalanan data



AI sebagai asisten kerja, membantu orang di lapisan itu

Menyelesaikan
error sintaks

Dokumentasi
struktur tabel

Query standar

Draft query, cek
logika rumus

AI sebagai produk, lapisan yang dibangun dengan AI

Model prediksi

Chatbot
pembaca
dokumen

Narasumber DE: data bersih dari lapisan 1–4 itulah **bahan bakar** AI.

Host: AI tidak bisa bertanggung jawab, jadi **keputusan** tetap di tangan manusia.

Penyusun · ringkasan bagian 1

Satu transaksi dipakai banyak peran





Bagian 2

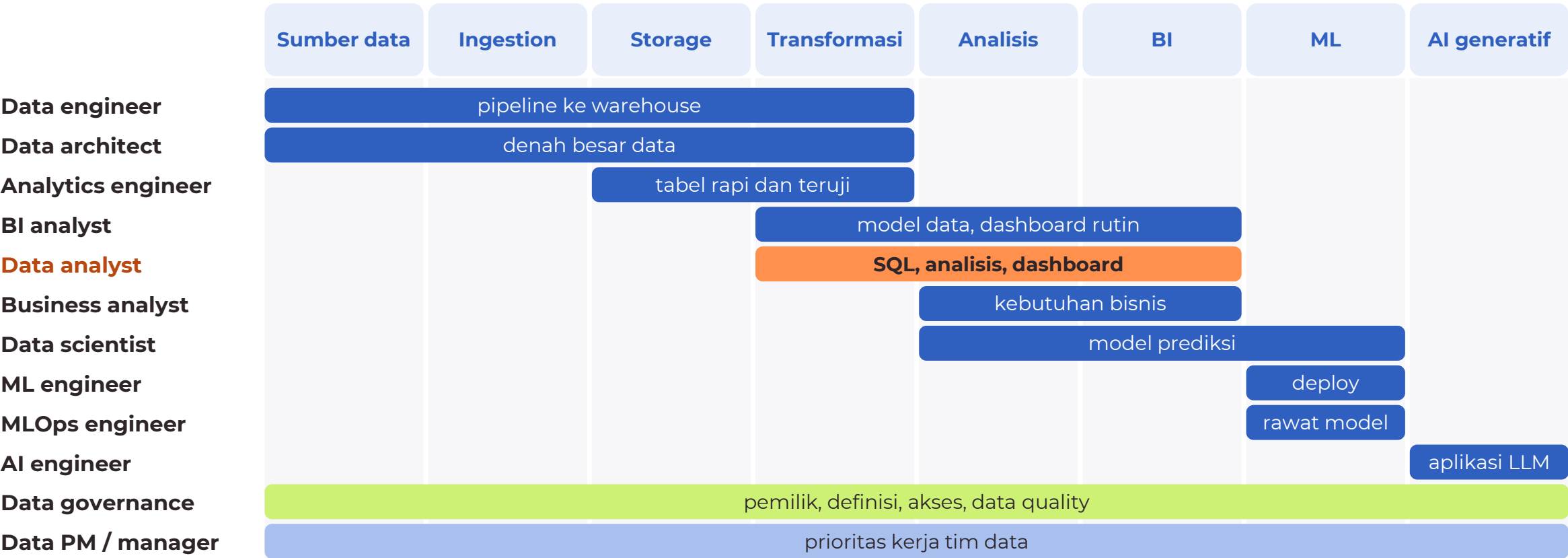
Peta peran

Dua belas peran di atas delapan lapisan: siapa memegang apa, siapa bekerja dengan siapa, dan mana yang ramah pemula.



Penyusun

Siapa memegang lapisan mana



Batang = lapisan utama tiap peran.

Penyusun

Dua belas peran dalam satu halaman



Data analyst

Mengolah data untuk keputusan bisnis.



BI analyst / developer

Merawat dashboard rutin.



Business analyst

Kebutuhan bisnis jadi sistem.



Analytics engineer

Merapikan dan menguji tabel.



Data engineer

Pipeline ke data warehouse.



Data architect

Denah besar data perusahaan.



Data scientist

Model untuk memprediksi.



ML engineer

Memasang model ke produk.



MLOps engineer

Otomasi retraining dan monitoring.



AI engineer

Aplikasi di atas LLM.



Data governance

Pemilik, definisi, akses, data quality.

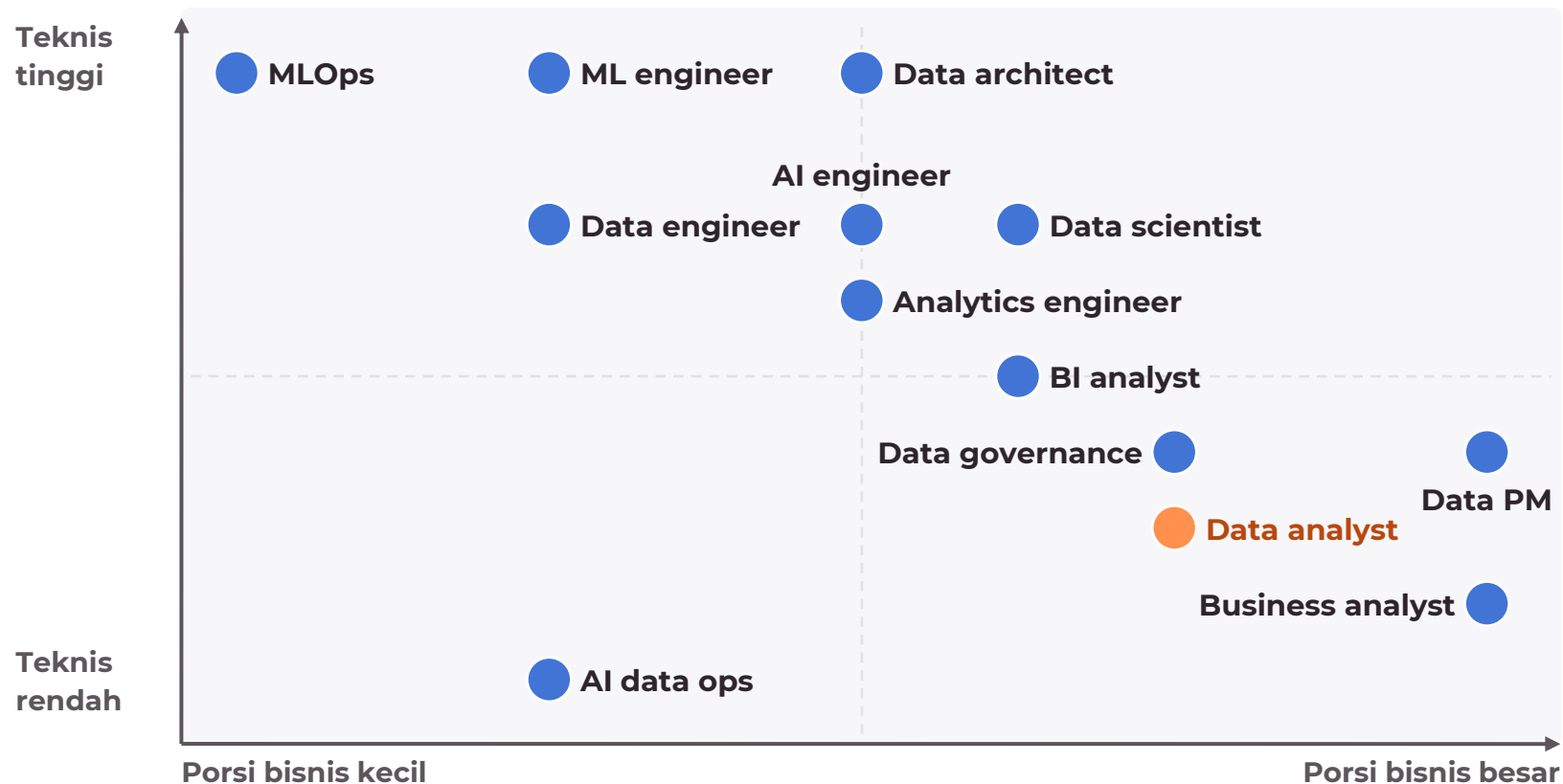


Data PM / analytics manager

Memilih yang dikerjakan dulu.

Penyusun · penilaian

Seberapa teknis, seberapa dekat ke bisnis?



Cara membaca

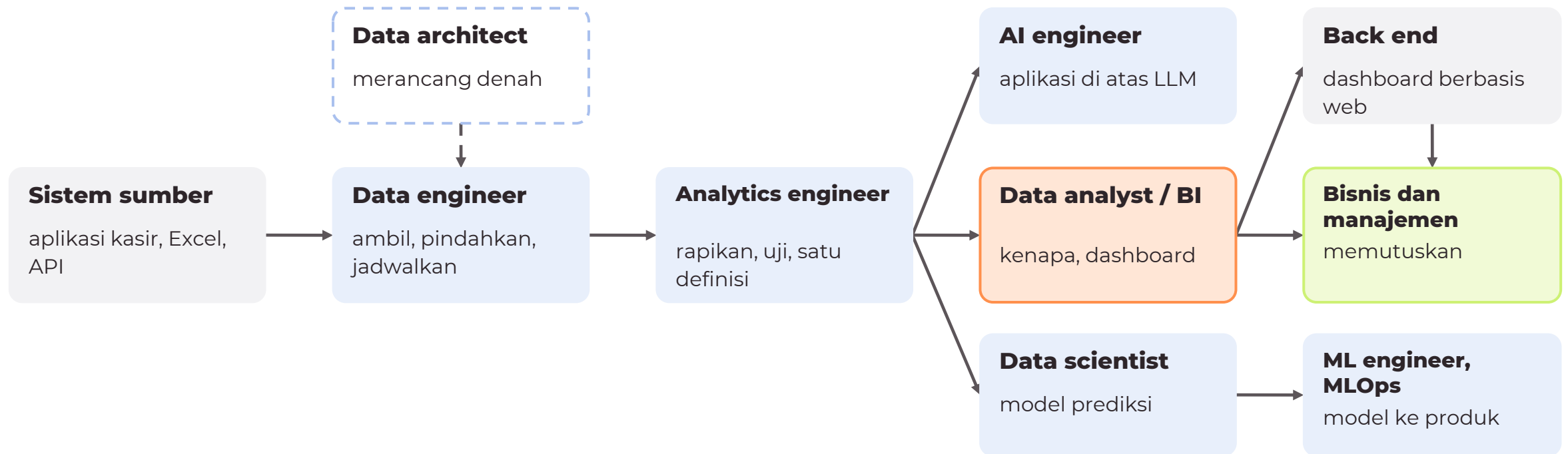
Kiri atas: membangun sistem, jauh dari meja bisnis. Kanan bawah: dekat dengan meja bisnis, alatnya lebih ringan.

Pindah karier dari non-IT? Lihat kanan bawah, yang paling dekat dengan pengalaman kerja sebelumnya.

Posisi titik kira-kira, karena di kantor nyata peran bisa bergeser.

Praktisi · sesi DA dan DE

Siapa bekerja dengan siapa, dari hulu ke hilir



Di sepanjang alur: data governance menjaga aturan data · business analyst menerjemahkan kebutuhan · data PM memilih prioritas

Praktisi • sesi Data Engineer

Empat rasa data engineer



Infrastructure / platform

Mengurus server dan cluster tempat data berjalan.

Dekat dengan
IT infrastructure, DevOps



Pipeline

Menjaga aliran data tetap jalan dan memasang alert kalau error.

Dekat dengan
inti pekerjaan DE



Analytics engineer

Sisi bisnis, sering berdiskusi dengan orang bisnis.

Dekat dengan
DA, BI



ML-focused

Memasang ML dan AI di pipeline, lalu mengurus modelnya.

Dekat dengan
DS, ML engineer

Semua rasa mengurus pipeline; ada pekerjaan yang memborong keempatnya.

Praktisi · sesi DA dan DE

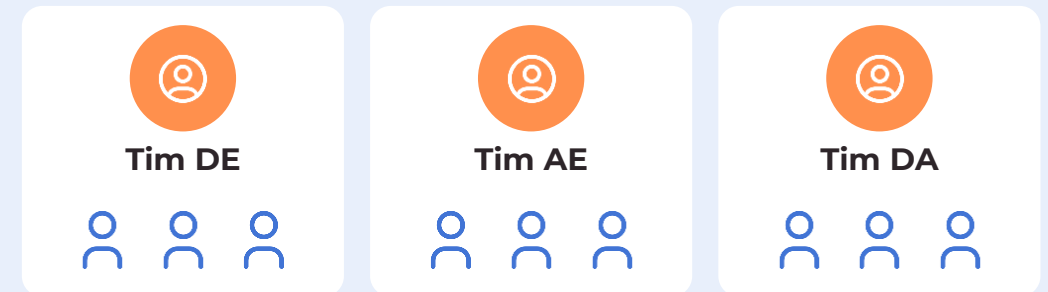
Tim kecil dan tim matang membagi kerja berbeda

Tim kecil



Satu orang memegang beberapa lapisan sekaligus. Host: ekspektasi sekarang "lebih ke fullstack data".

Tim matang



DE, AE, dan DA bisa punya tim dan manajer sendiri. Di perusahaan teknologi besar, tim data bisa puluhan sampai ratusan orang.

Penyusun: kalau melamar ke kantor kecil, siapkan diri memegang lebih dari satu lapisan.

Penyusun • Data • lowongan 3 Okt 2026

Judul role yang muncul di lowongan kerja

Data analyst

- Analisis Data
- Market Analyst

BI analyst

- BI Developer
- BI Specialist

Business analyst

- System Analyst
- Process Analyst

Analytics eng.

- Data Modeler
- BI Engineer

Data engineer

- ETL Developer
- Data Platform Engineer

Data scientist

- Applied Scientist
- Statistician

ML engineer

- Applied ML Engineer
- AI Engineer

AI engineer

- LLM Engineer
- GenAI Engineer

Data governance

- Data Steward
- Data Quality Analyst

Data PM

- Analytics Manager
- Head of Data

9%

lowongan berjudul Data Analyst isinya tidak berupa analisis data (14/151).

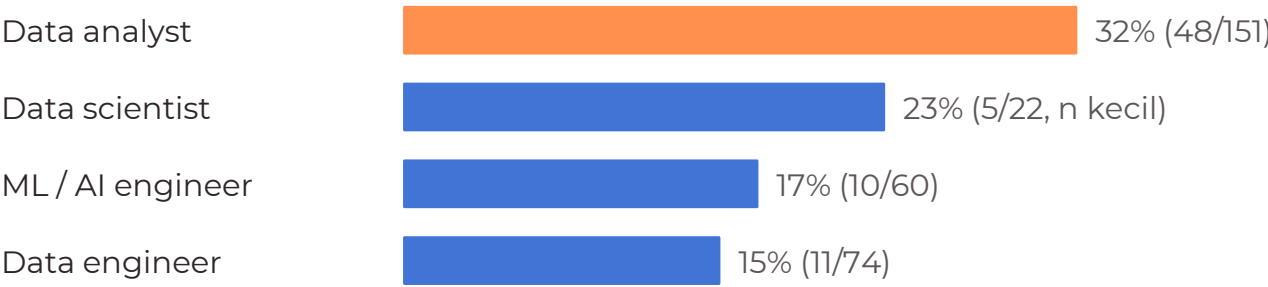
Baca **isi lowongan**, jangan hanya judulnya.

Penyusun • Data • lowongan 3 Okt 2026

Ramah pemula atau tidak

Terbuka	Sedang	Rendah-sedang	Rendah	Bukan pemula
Data analyst	BI analyst	Data engineer	ML engineer	Data architect
AI data ops	Business analyst	Data scientist	Analytics engineer	MLOps engineer
Admin data	Data governance	AI engineer		Data PM

Terbuka untuk fresh graduate (deskripsi lengkap)



Banyak lowongan DA tetap minta pengalaman **1-3 tahun**.

AI data ops: 16/31 terbuka untuk fresh grad, sering kontrak vendor.



Bagian 3

Kartu peran

Dua kartu per peran, tiga untuk DA dan DE. **Kartu A:** kerjanya apa. **Kartu B:** alat, skill awal, dan jalur masuk.



Praktisi • sesi Data Analyst

Data Analyst membantu tim bisnis memutuskan lewat data



Kerjanya apa

Mengolah data untuk tim bisnis dan menjelaskannya ke yang memutuskan, sebagai penghubung bisnis dan teknis.

Analogi Dokter umum: mendengar keluhan, memeriksa, lalu memberi diagnosis dan saran.

Pertanyaan khasnya

- Kenapa penjualan bulan ini turun?
- Cabang mana di bawah target, dan dibanding apa?
- Promo kemarin berhasil atau tidak?

Keseharian

- Menggali kebutuhan user
- Eksplorasi, validasi, cleansing
- Query, measure, visualisasi
- Presentasi dan rekomendasi

Output

- Dashboard
- Report rutin
- Memo rekomendasi

Hulu Data engineer, pemilik sistem sumber



DA



Hilir Manajemen, marketing, sales, back end

Data · lowongan 3 Okt 2026 · Praktisi

Alat, skill awal, dan jalur masuk Data Analyst

Alat umum

Excel / Sheets

SQL

Power BI atau Tableau

Python (nilai tambah)

Judul role lain

Analisis Data

Reporting Analyst

Market Analyst

MIS Officer



Skill minimum junior

Excel lancar, SQL sampai CTE, satu BI tool, statistik deskriptif.



Ide portfolio

Satu minggu satu portfolio: dashboard plus memo rekomendasi dari data asli.



Dibantu AI

Draft query, cek logika rumus, dokumentasi.



Kesalahan umum

Cockpit syndrome, dashboard tanpa pembandingan, confirmation bias (kartu 3).



Jalur masuk

Lowongan terbanyak: 221 dari 538. Terbuka untuk fresh grad 32% (48 dari 151).



Tetap di manusia

Menggali kebutuhan, memvalidasi query multi-tabel, tanggung jawab.

Praktisi • sesi Data Analyst

Siklus kerja DA, dan tiga kesalahan yang sering muncul



70% = pengalaman narasumber, belum hasil studi.

Tiga kesalahan dari sesi DA

- 1 Cockpit syndrome**
Semua grafik masuk satu dashboard, tanpa hierarki.
- 2 Dashboard tanpa cerita**
Beri pembandingan: target, bulan lalu, tahun lalu.
- 3 Confirmation bias**
Mencari data pendukung asumsi saja, lupa pola musiman.

Data · O*NET · Penyusun

BI Analyst / Developer merawat dashboard rutin



Kerjanya apa

Membuat dan merawat dashboard dan report rutin, supaya orang bisa melihat angka sendiri tanpa bertanya.

Analogi Pemasang panel dashboard mobil: pengemudi cukup melirik, tidak perlu membuka mesin.

Pertanyaan khasnya

- Bagaimana KPI hari ini dibanding target?
- Report apa yang dibutuhkan tiap divisi?
- Kenapa angka di dua dashboard berbeda?

Keseharian

- Membuat report untuk stakeholder
- Merawat dashboard dan database
- Membantu user memakai report
- Menulis spesifikasi dashboard

Output

Dashboard

Report rutin

Semantic model

Hulu Data engineer, analytics engineer



BI



Hilir Manajemen, divisi pengguna dashboard

Penyusun • Data • lowongan dan O*NET

Alat, skill awal, dan jalur masuk BI Analyst

Alat umum

Power BI (DAX)

Tableau

Looker

Metabase

SQL

Star schema

Judul role lain

BI Developer

BI Specialist

Reporting Analyst

Dashboard Developer



Skill minimum junior

SQL kuat, satu BI tool sampai measure dan relasi, star schema.



Ide portfolio

Dashboard dengan pembandingan, plus daftar measure dan definisinya.



Dibantu AI

Membuat halaman report, narasi otomatis, dan draft DAX.



Kesalahan umum

Cockpit syndrome. Satu metrik dengan definisi beda di dua dashboard.



Jalur masuk

Lowongan BI hanya 17 (n kecil), jadi umumnya mulai sebagai DA.



Tetap di manusia

Memilih metrik, merapikan model data, memeriksa narasi.

Data • IIBA • Penyusun

Business Analyst mengubah kebutuhan bisnis jadi sistem



Kerjanya apa

Menerjemahkan kebutuhan bisnis menjadi perubahan proses atau sistem yang bisa dibangun dan diuji.

Analogi Arsitek renovasi rumah: mendengar penghuni, lalu menulis gambar kerja untuk tukang.

Pertanyaan khasnya

- Masalah bisnisnya sebenarnya apa?
- Proses sekarang macet di mana?
- Sistem baru harus bisa apa, dan bagaimana mengujinya?

Keseharian

- Wawancara dan workshop user
- Memetakan proses
- Menulis requirement dan user story
- Mendampingi UAT

Output

Dokumen requirement

Peta proses

User story

SOP

Hulu User bisnis, manajemen

BA

Hilir Developer, QA, data engineer, data analyst

Penyusun • Data • lowongan 3 Okt 2026

Alat, skill awal, dan jalur masuk Business Analyst

Alat umum

Spreadsheet

Diagram proses (BPMN)

Manajemen tugas

SQL dasar

BI tool

Judul role lain

System Analyst

Business System Analyst

Process Analyst

Requirements Engineer



Skill minimum junior

Wawancara kebutuhan, menulis jelas, membaca proses, Excel, dan SQL dasar.



Ide portfolio

Studi kasus layanan publik: peta proses, masalah terukur, dan requirement.



Dibantu AI

Notulen, draft requirement, ringkasan dokumen panjang.



Kesalahan umum

Menyalin permintaan user tanpa menggali masalahnya. Requirement yang tidak bisa diuji.



Jalur masuk

31 lowongan (peran terkait). Cocok untuk yang pindah dari operasional.



Tetap di manusia

Menggali masalah tersembunyi, menengahi antardivisi.

Praktisi • sesi DE • Penyusun

Analytics Engineer menyeragamkan angka semua tim



Kerjanya apa

Merapikan dan menguji tabel di data warehouse, supaya semua analis dan dashboard memakai angka yang sama.

Analogi Juru resep di dapur pusat: menulis resep baku supaya rasa di semua cabang sama.

Pertanyaan khasnya

- ❓ Apa definisi resmi 'pelanggan aktif'?
- ❓ Kenapa finance dan operasional beda angka?
- ❓ Tabel mana yang aman dipakai dashboard?

Keseharian

- Menulis model transformasi SQL
- Menulis tes data quality
- Dokumentasi tabel dan metrik
- Berdiskusi dengan orang bisnis

Output

Tabel gold

Definisi metrik

Tes data

Lineage

Hulu Data engineer



AE



Hilir Data analyst, BI, data scientist

Penyusun • Praktisi • Data

Alat, skill awal, dan jalur masuk Analytics Engineer

Alat umum

dbt

SQL

BigQuery / Snowflake

Git

Airflow

Judul role lain

Data Modeler

BI Engineer

Data Engineer (Analytics)



Skill minimum junior

SQL lanjut, data modeling, Git dasar, dan menulis tes data.



Ide portfolio

Proyek dbt: bronze sampai gold, lima tes, satu dashboard.



Dibantu AI

Menulis SQL model dan tes standar.



Kesalahan umum

Menghafal fitur dbt tanpa paham kenapa model dipecah bertahap.



Jalur masuk

Judul langka: 2 dari 538 lowongan. Lebih realistis sebagai skill DA atau DE.



Tetap di manusia

Menyepakati definisi metrik dengan bisnis, yang butuh negosiasi.

Praktisi · sesi Data Engineer

Data Engineer membangun pipeline ke data warehouse



Kerjanya apa

Memindahkan data dari sistem sumber ke data warehouse lalu merapikannya, terjadwal dan otomatis, siap dipakai DA.

Analogi Dapur restoran mi: bahan dan SOP masak siap terus, termasuk saat jam sibuk.

Pertanyaan khasnya

- ❓ Data ada di mana saja, dan bagaimana menyatukannya?
- ❓ Bagaimana data kemarin sudah ada di dashboard pagi ini?
- ❓ Kalau pipeline rusak, siapa yang tahu?

Empat rasa

- Infrastructure / platform
- Pipeline
- Analytics engineer
- ML-focused

Output

Pipeline terjadwal

Tabel di warehouse

Alert

Hulu Pemilik aplikasi sumber, tim IT



DE



Hilir DA, AE, DS, tim AI

Praktisi • Data • lowongan 3 Okt 2026

Alat, skill awal, dan jalur masuk Data Engineer

Alat umum

Python

SQL

CLI Linux

Git

Docker

Airflow

dbt

BigQuery

Judul role lain

ETL Developer

Data Administrator

Data Platform Engineer

Big Data Engineer



Skill minimum junior

SQL, Python dasar, CLI, Git, Docker. "Tidak perlu jago, yang penting tahu dulu."



Ide portfolio

Pipeline end-to-end lokal di GitHub, sampai gold. Ceritakan kesulitannya.



Dibantu AI

Query standar, dokumentasi, error sintaks.



Kesalahan umum

Tool-oriented, lupa komunikasi, biaya cloud bengkak, overkill.



Jalur masuk

120 lowongan. Terbuka untuk fresh grad 15% (11 dari 74). Pipeline disebut di 92% (68 dari 74).



Tetap di manusia

Konteks bisnis, infra, memilih alat.

Praktisi · sesi Data Engineer

Keseharian DE dan ukuran suksesnya

Yang dikerjakan setiap hari



Ada juga

Development

Testing

Deployment

Paling banyak

Monitoring

Debugging

Memasang alert

Fixing

Fixing biasanya perlu saat ada data unik yang merusak pipeline.
Narasumber: komunikasi dengan user sering diabaikan orang IT, padahal penting.

Ukuran sukses menurut narasumber

-  Pipeline stabil
-  Data dipercaya: tidak ada lagi debat angka antartim
-  Stakeholder bisa mencari jawabannya sendiri
-  Biaya cloud terkendali
-  Rekan senang bekerja sama: responsif, tidak menyembunyikan masalah

Data · O*NET · Penyusun

Data Architect merancang denah besar data perusahaan



Kerjanya apa

Merancang di mana data disimpan, bentuknya, jalurnya, dan aturannya, sebelum sistem dibangun.

Analogi Arsitek gedung: tidak memasang pipa sendiri, tapi menentukan letak pipa dan kamar.

Pertanyaan khasnya

- Warehouse, lakehouse, atau cukup satu database?
- Skema apa yang tetap cepat saat data sepuluh kali lipat?
- Standar apa yang dipakai semua tim?

Keseharian

- Mendokumentasikan arsitektur
- Berkolaborasi dengan stakeholder
- Merancang untuk skala dan keamanan
- Menetapkan standar

Output

Dokumen arsitektur

Data model

Standar

Keputusan teknologi

Hulu Manajemen, stakeholder sistem



Architect



Hilir DE, database administrator, AE, tim keamanan

Penyusun • Data • O*NET dan BLS

Alat, skill awal, dan jalur masuk Data Architect

Alat umum

- Oracle
- SQL Server
- PostgreSQL
- MongoDB
- Snowflake
- AWS / Azure

Judul role lain

- Database Architect
- Enterprise Architect
- Information Architect



Skill minimum junior

Bukan peran junior: prasyaratnya beberapa tahun sebagai DE atau database administrator.



Ide portfolio

Dokumen desain kasus fiktif: diagram, data model, alasan memilih alat.



Dibantu AI

Draft dokumen, perbandingan opsi.



Kesalahan umum

Overkill: merancang arsitektur perusahaan raksasa untuk organisasi kecil.



Jalur masuk

Bukan untuk fresh graduate. BLS: database administrator berpengalaman bisa naik jadi architect.



Tetap di manusia

Trade-off biaya dan risiko, dan tanggung jawab kalau desainnya gagal.

Data · BLS dan O*NET · Penyusun

Data Scientist memprediksi dari pola data



Kerjanya apa

Membuat model dari data untuk memprediksi atau mengelompokkan, misalnya permintaan bulan depan.

Analogi Prakirawan cuaca: menghitung peluang dari pola masa lalu, lengkap dengan ketidakpastiannya.

Pertanyaan khasnya

- ❓ Pelanggan mana yang kemungkinan berhenti?
- ❓ Berapa stok yang perlu disiapkan bulan depan?
- ❓ Perubahan ini berpengaruh, atau kebetulan?

Keseharian

- Membersihkan data mentah
- Membangun dan menguji model
- Membandingkan model dengan metrik
- Menyajikan hasil ke stakeholder

Output

Model prediksi

Laporan evaluasi

Analisis eksperimen

Hulu DE, AE, pemilik masalah bisnis

DS

Hilir ML engineer, manajemen, product

Data · lowongan dan Google · Penyusun

Alat, skill awal, dan jalur masuk Data Scientist

Alat umum	Python	R	SQL	pandas	scikit-learn	Jupyter	Git
Judul role lain	Associate Data Scientist		Applied Scientist		Decision Scientist		Statistician

**Skill minimum junior**

Python, SQL, statistik inferensial, dasar ML: regresi, klasifikasi, evaluasi.

**Ide portfolio**

Forecasting data publik dengan baseline sederhana, plus di mana model salah.

**Dibantu AI**

Menulis kode analisis dan model, menjelaskan library.

**Kesalahan umum**

Model rumit tanpa baseline. Data masa depan bocor ke data training.

**Jalur masuk**

47 lowongan. Python diminta di 20 dari 22 lowongan dengan deskripsi lengkap. Fresh grad 5 dari 22 (n kecil).

**Tetap di manusia**

Memilih masalah, mengecek asumsi, menanggung keputusan.

Data · Google Cloud · Penyusun

ML Engineer membawa model dari notebook ke produk



Kerjanya apa

Memasang model ke produk supaya berjalan otomatis, cepat, dan bisa dipakai banyak orang.

Analogi Insinyur pabrik: mengubah resep juara dapur uji jadi lini produksi yang rasanya sama.

Pertanyaan khasnya

- Bagaimana model melayani banyak permintaan sekaligus?
- Bagaimana tahu performa model menurun?
- Kapan model perlu retraining?

Keseharian

- Mengubah notebook DS jadi kode production
- Membangun pipeline ML
- Deploy model sebagai API
- Monitoring model di production

Output

Model di production

API model

Pipeline training

Hulu Data scientist, data engineer



MLE



Hilir Tim aplikasi, product

Data · lowongan dan Google · Penyusun

Alat, skill awal, dan jalur masuk ML Engineer

Alat umum	Python	scikit-learn	TensorFlow / PyTorch	Docker	Kubernetes	Cloud
Judul role lain	Machine Learning Engineer	Applied ML Engineer	AI Engineer (sering tertukar)			

**Skill minimum junior**

Software engineering (Git, testing), Python kuat, dasar ML, Docker, satu cloud.

**Ide portfolio**

Model sederhana dibungkus API di Docker, dengan tes, log, dan skrip retraining.

**Dibantu AI**

Menulis kode dan membantu debugging.

**Kesalahan umum**

Fokus akurasi, lupa infrastruktur.

**Jalur masuk**

Jarang untuk fresh grad. Lowongan ML/AI 117, terbuka untuk fresh grad 17% (10 dari 60).

**Tetap di manusia**

Memutuskan model aman dirilis, responsible AI.

Data · Google Cloud · Penyusun

MLOps Engineer menjaga model tetap sehat



Kerjanya apa

Membuat pipeline otomatis untuk retraining, pengujian, rilis, dan monitoring model tanpa kerja manual.

Analogi Tim perawatan pesawat: jadwal cek, daftar Periksa, dan alarm supaya aman terbang tiap hari.

Pertanyaan khasnya

- Model mana yang berjalan, dan apa data training-nya?
- Performa turun karena data berubah?
- Bagaimana merilis model baru dengan aman?

Keseharian

- Pipeline training otomatis
- CI/CD untuk kode dan model
- Monitoring dan alert
- Mencatat versi data dan model

Output

Pipeline ML

Monitoring model

Model registry

Hulu DS, ML engineer, DE



MLOps



Hilir Tim platform dan aplikasi

Penyusun • Data • Google Cloud

Alat, skill awal, dan jalur masuk MLOps Engineer

Alat umum

Git

CI/CD

Docker

Kubernetes

Airflow

Model registry

Cloud

Judul role lain

ML Platform Engineer

ML Infrastructure Engineer

DevOps Engineer



Skill minimum junior

Bukan peran junior; prasyaratnya software engineering, DevOps, dan dasar ML.



Ide portfolio

Pipeline lokal yang me-retrain model kecil dan mengirim alert.



Dibantu AI

Menulis konfigurasi dan skrip CI/CD, membaca log.



Kesalahan umum

Overkill: platform lengkap untuk satu model yang jarang berubah.



Jalur masuk

Sangat jarang untuk fresh grad; biasanya dari DevOps atau software engineering.



Tetap di manusia

Kapan model dianggap rusak, kapan rollback.

Data • Microsoft dan Anthropic • Penyusun

AI Engineer membangun aplikasi di atas LLM



Kerjanya apa

Membangun aplikasi di atas LLM yang sudah ada, misalnya chatbot yang menjawab dari dokumen perusahaan.

Analogi Kontraktor dapur: tidak membuat kompor, tapi menyusun sambungan dan pengamannya.

Pertanyaan khasnya

- Bagaimana AI menjawab dari dokumen kita, tidak mengarang?
- Seberapa sering jawabannya benar?
- Apa yang tidak boleh dilakukan AI ini?

Keseharian

- Merancang prompt, retrieval, dan tool
- Menyusun set evaluasi
- Mengukur dan memperbaiki berulang
- Memasang guardrail

Output

Aplikasi AI

Set evaluasi

Dokumentasi guardrail

Hulu DE, data scientist, pemilik dokumen



AI eng.



Hilir Developer aplikasi, pemilik proses bisnis

Data · lowongan dan Anthropic · Penyusun

Alat, skill awal, dan jalur masuk AI Engineer

Alat umum

Python

API LLM

Vector database

SDK vendor / LangChain

Docker

Judul role lain

LLM Engineer

Generative AI Engineer

AI Developer

AI Automation Engineer



Skill minimum junior

Python, memanggil API, dasar RAG, menyusun evaluasi.



Ide portfolio

Asisten tanya-jawab dokumen publik, 30 pertanyaan uji, catatan kasus gagal.



Dibantu AI

Menulis kode dan mengusulkan kasus uji.



Kesalahan umum

Agent rumit padahal satu panggilan model cukup, dan tanpa evaluasi.



Jalur masuk

94 lowongan berjudul AI. LLM atau GenAI diminta di 27 dari 44 lowongan dengan deskripsi lengkap.



Tetap di manusia

Kriteria sukses, menilai jawaban yang "hampir benar".

Data · AWS · Penyusun

Data Governance memastikan data punya pemilik dan aturan



Kerjanya apa

Memastikan data punya pemilik, definisi resmi, dan aturan akses, lalu memantau data quality-nya.

Analogi Pustakawan: buku punya nomor katalog, ada aturan pinjam, dan buku rusak diperbaiki.

Pertanyaan khasnya



Angka mana yang resmi?



Siapa pemilik data ini, dan siapa boleh melihatnya?



Data pribadi diperlakukan sesuai aturan?

Keseharian

- Merawat data dictionary dan katalog
- Menulis aturan data quality
- Mengelola permintaan akses
- Menengahi definisi antardivisi

Output

Data dictionary

Kebijakan akses

Report data quality

Hulu Data owner, pemilik sistem



DG



Hilir Semua pemakai data, tim kepatuhan

Data · lowongan dan UU PDP · Penyusun

Alat, skill awal, dan jalur masuk Data Governance

Alat umum

Excel

SQL

DataHub / Purview

Tes data dbt

DAMA-DMBOK

Judul role lain

Data Steward

Data Quality Analyst

Master Data Analyst

Data Protection Officer



Skill minimum junior

Teliti, Excel kuat, SQL dasar, paham dasar UU PDP.



Ide portfolio

Data dictionary, klasifikasi kolom, dan hasil cek data quality dari dataset publik.



Dibantu AI

Draft data dictionary, deteksi anomali.



Kesalahan umum

Kebijakan tebal yang tidak dipakai.
Governance yang menghalangi akses.



Jalur masuk

Lowongan sedikit: 14 (n kecil). Cocok untuk latar hukum, audit, administrasi.



Tetap di manusia

Menetapkan definisi resmi, memutuskan akses, dan menanggung kepatuhan.

Penyusun • seluruh kartu

Data PM / Analytics Manager memilih yang dikerjakan dulu



Kerjanya apa

Memilih pekerjaan data mana yang dikerjakan dulu, untuk siapa, dan memastikan hasilnya dipakai.

Analogi Kepala dapur: menentukan menu, membagi tugas, dan mencicipi sebelum piring keluar.

Pertanyaan khasnya

- Masalah bisnis mana yang paling layak dikerjakan dengan data?
- Dashboard ini dipakai atau tidak?
- Siapa mengerjakan apa?

Keseharian

- Menyusun prioritas dan roadmap
- Review hasil tim
- Rapat dengan eksekutif
- Rekrutmen dan budget

Output

Roadmap data

Standar kerja tim

Keputusan prioritas

Hulu Eksekutif, divisi bisnis



Data PM



Hilir Semua peran data di tim

Penyusun • Praktisi • host sesi DA

Alat, skill, dan jalur Data PM / Analytics Manager

Alat umum

Spreadsheet

Manajemen proyek

BI tool untuk membaca hasil

Judul role lain

Analytics Manager

Head of Data

Data Product Owner

BI Manager

Data Lead



Skill minimum junior

Paham semua tahap alur data, komunikasi dengan eksekutif, dan menyusun prioritas.



Ide portfolio

Belum relevan untuk pemula. Kumpulkan cerita proyek yang hasilnya dipakai.



Dibantu AI

Ringkasan dan draft dokumen.



Kesalahan umum

Lupa komunikasi: pekerjaan bagus yang tidak dipahami pemakainya tidak akan dipakai.



Jalur masuk

Bukan untuk fresh graduate. Host: dari DA, BI manager, sampai Data & AI Lead.



Tetap di manusia

Prioritas, tanggung jawab ke manajemen, dan membina tim.

Data · lowongan 3 Okt 2026 · Penyusun

AI Data Ops dan Admin Data, pintu masuk dekat data



Kerjanya apa

AI trainer dan annotator menilai atau memperbaiki data dan jawaban AI. Admin data menginput dan merapikan data.

Analogi Korektor naskah: memeriksa tulisan sebelum buku dicetak.

- ### Pertanyaan khasnya
- Jawaban AI ini benar menurut rubrik?
 - Data ini lengkap dan konsisten?

- ### Keseharian
- Menilai jawaban dengan rubrik
 - Memberi label data
 - Input dan cek data

- ### Output
- Label data
 - Penilaian jawaban AI
 - Data rapi



Data · lowongan 3 Okt 2026 · Penyusun

Alat, skill, dan catatan AI Data Ops dan Admin Data

Alat umum

Platform anotasi

Spreadsheet

LLM

Judul role lain

AI Trainer

Annotator

Data Entry

Admin Data



Skill

Ketelitian, bahasa, rubrik penilaian, kadang domain seperti hukum atau medis.



Lowongan di snapshot

AI data ops: 31 lowongan, 16 terbuka untuk fresh grad. Admin data dan data entry: 32 lowongan.



Dibantu AI

Bagian input manual paling mudah diotomasi.



Catatan jujur

Sering remote atau kontrak lewat vendor. Stabilitas dan jenjang kariernya belum jelas.



Cara memakainya

Jadikan pintu dan tempat mengumpulkan pengalaman, sambil membangun portfolio DA.



Tetap di manusia

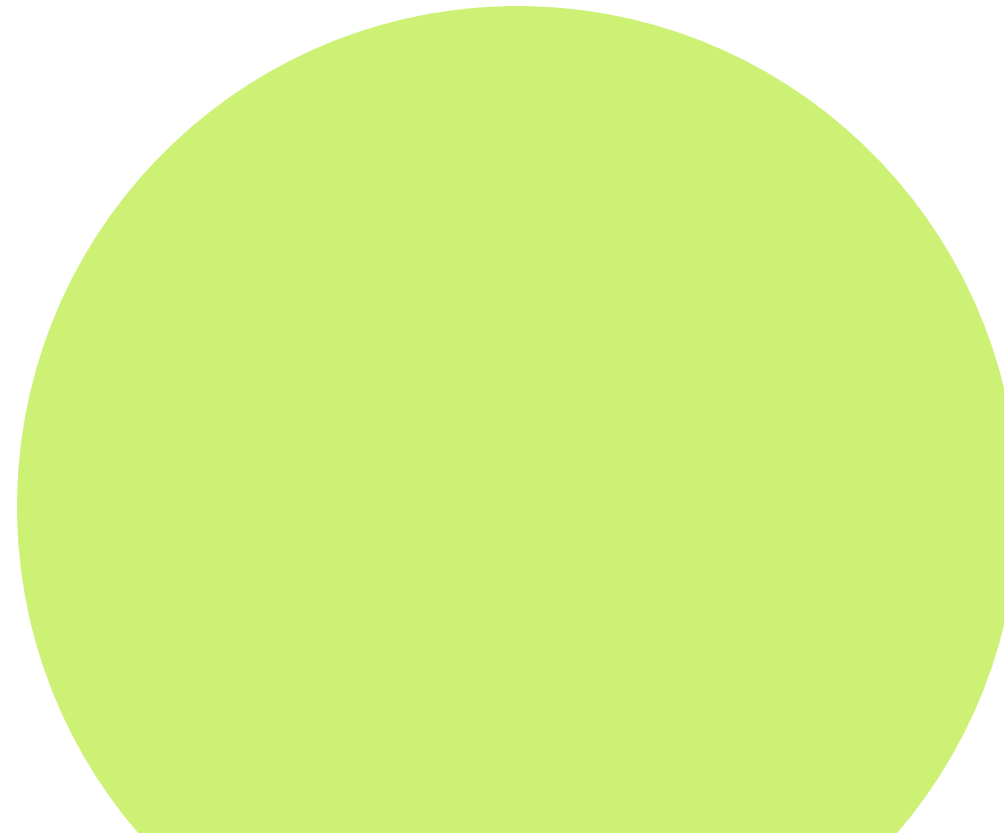
Menilai output AI, karena memang itu inti pekerjaannya.



Bagian 4

Konsep kunci

Sebelas konsep yang sering muncul di sesi dan di lowongan kerja, masing-masing dalam satu gambar.



Praktisi • sesi Data Engineer

OLTP mencatat transaksi, OLAP merekapnya



OLTP: mencatat

Database aplikasi kasir

waktu	cabang	menu	porsi	kasir
—	—	—	—	—
—	—	—	—	—
—	—	—	—	—
—	—	—	—	—
—	—	—	—	—

Satu transaksi = satu baris baru

Cepat menulis satu baris per transaksi, tapi lambat untuk merekap data besar. Contoh: PostgreSQL, MySQL.



OLAP: merekap

Database khusus analytics

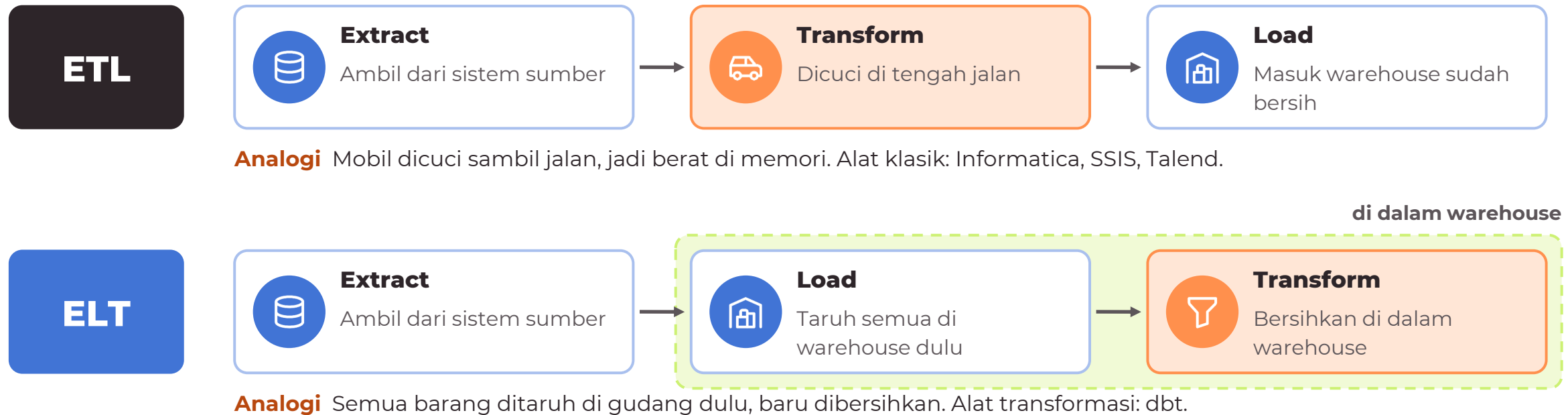
waktu	cabang	menu	porsi	kasir
—	—	—	—	—
—	—	—	—	—
—	—	—	—	—
—	—	—	—	—
—	—	—	—	—

Hanya membaca kolom yang dibutuhkan

Total porsi per cabang sebulan cukup dihitung dari dua kolom, cocok untuk database columnar. Contoh: ClickHouse, BigQuery.

Praktisi · sesi Data Engineer

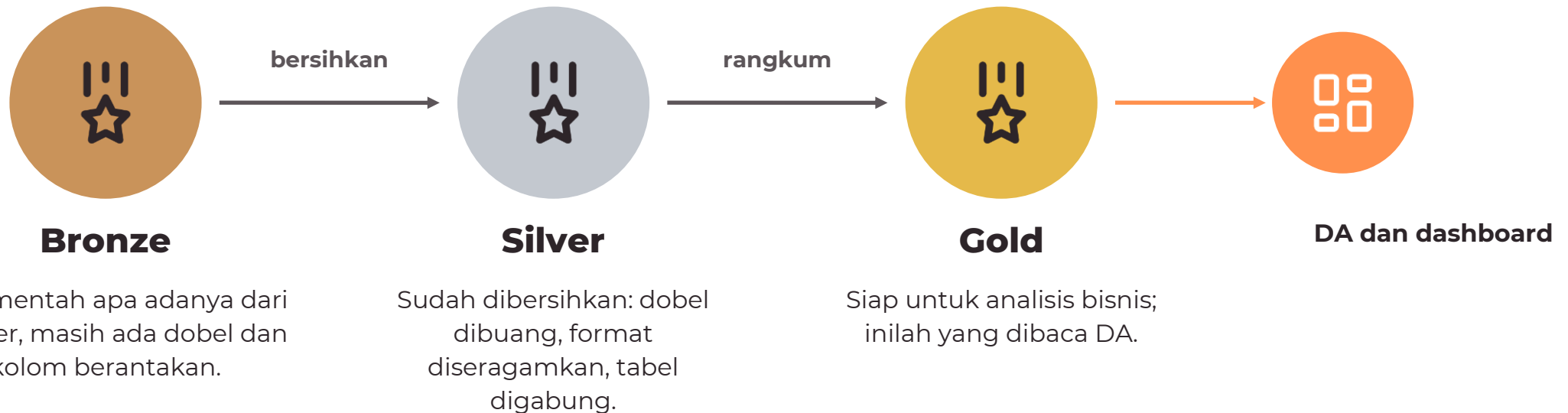
ETL atau ELT, kapan data dibersihkan?



Narasumber DE: dulu ETL populer, kini ELT lebih dipilih karena lebih murah dan rapi. Pengalaman praktisi, belum diukur.

Praktisi · sesi DE · Penyusun

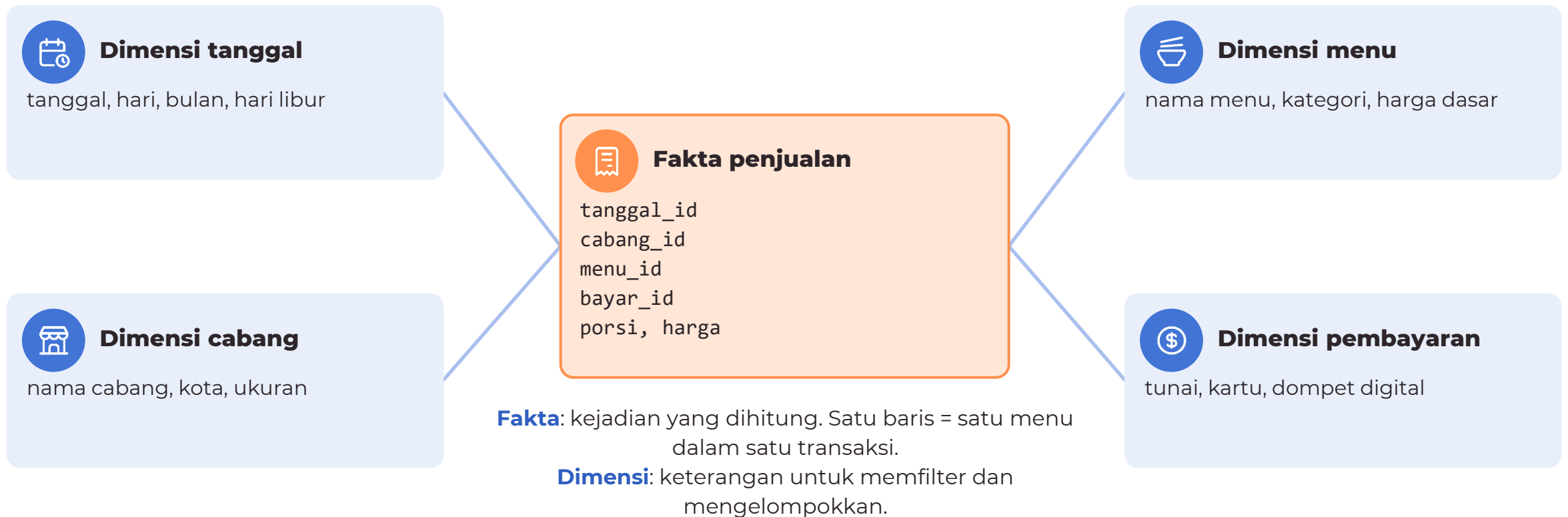
Medallion architecture, pembersihan data bertahap



Contoh restoran mi Bronze: semua transaksi kasir apa adanya. Silver: transaksi batal dibuang, nama menu diseragamkan. Gold: penjualan harian per cabang, siap untuk dashboard pemilik.

Penyusun · contoh fiktif

Star schema, satu tabel fakta dikelilingi dimensi



Penyusun

Warehouse, data lake, dan lakehouse



Data warehouse

Menyimpan

Tabel rapi yang siap dianalisis.

Bayangkan

Gudang dengan rak berlabel.

Cocok untuk

Dashboard dan query SQL rutin.

Contoh

BigQuery, Snowflake, Redshift, ClickHouse



Data lake

Menyimpan

File mentah: tabel, log, gambar.

Bayangkan

Danau: masuk dulu, ditata nanti.

Cocok untuk

Data besar dengan biaya murah.

Contoh

Object storage seperti S3



Lakehouse

Menyimpan

File mentah yang bisa dibaca SQL.

Bayangkan

Danau yang punya rak berlabel.

Cocok untuk

Analisis dan ML dari satu tempat.

Contoh

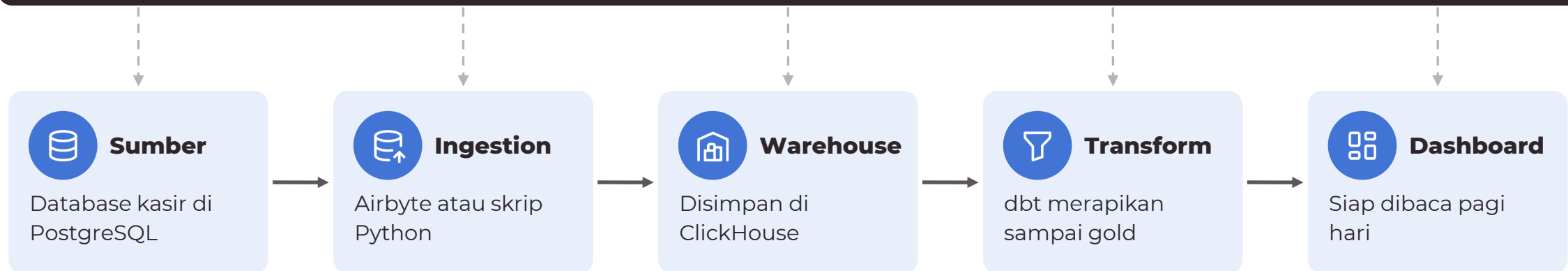
File di object storage yang disajikan sebagai tabel SQL

Praktisi • sesi Data Engineer

Dengan orkestrasi, pipeline jalan sendiri



Langkah-langkah ini dijalankan **orkestrator** (Airflow, Prefect) berurutan **sesuai jadwal**, misalnya setiap pagi.



Kalau satu langkah gagal, langkah berikutnya ditahan dan **alert** dikirim ke data engineer.

Host: data masih kecil? Alat ingestion ringan seperti Airbyte bisa cukup tanpa orkestrasi besar. Jangan **overkill**.

Penyusun

Batch atau streaming?



Batch

Data dikumpulkan, lalu dipindah sekaligus pada jadwal, misalnya tiap tengah malam.

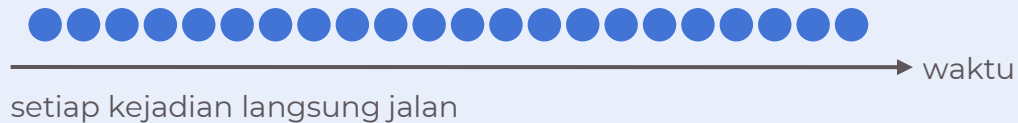


Contoh Laporan penjualan harian restoran. Alat: Airflow dan SQL.



Streaming

Data dipindah dan diolah terus-menerus, begitu kejadiannya terjadi.



Contoh Posisi kurir di aplikasi ojek. Alat: Kafka, Spark.

Narasumber DE menaruh big data dan streaming di level medium-lanjut, jadi mulailah dari batch.

Praktisi · sesi DA · Penyusun

SQL: bahasa untuk bertanya ke database

```
SELECT  cabang,  
        SUM(porsi) AS total_porsi  
FROM    penjualan  
WHERE   tanggal >= '2026-09-01'  
GROUP BY cabang  
ORDER BY total_porsi DESC;
```

SQL

SELECT: kolom cabang dan jumlah porsi.

FROM: tabel penjualan.

WHERE: transaksi sejak 1 September.

GROUP BY: satu baris per cabang.

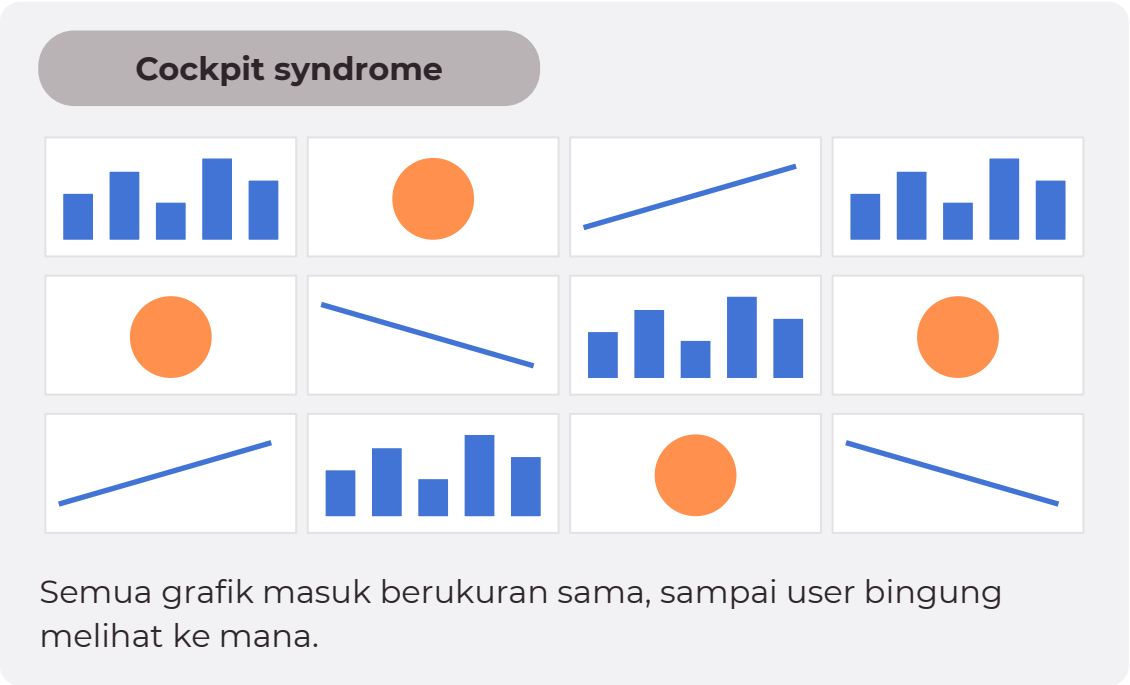
ORDER BY: porsi terbanyak di atas.

Hasil: daftar cabang, yang paling laris di atas.

Host: SQL **bahasa utama orang data**. Narasumber DA: pelajari sampai CTE agar bisa mengecek query buatan AI.

Praktisi • sesi Data Analyst

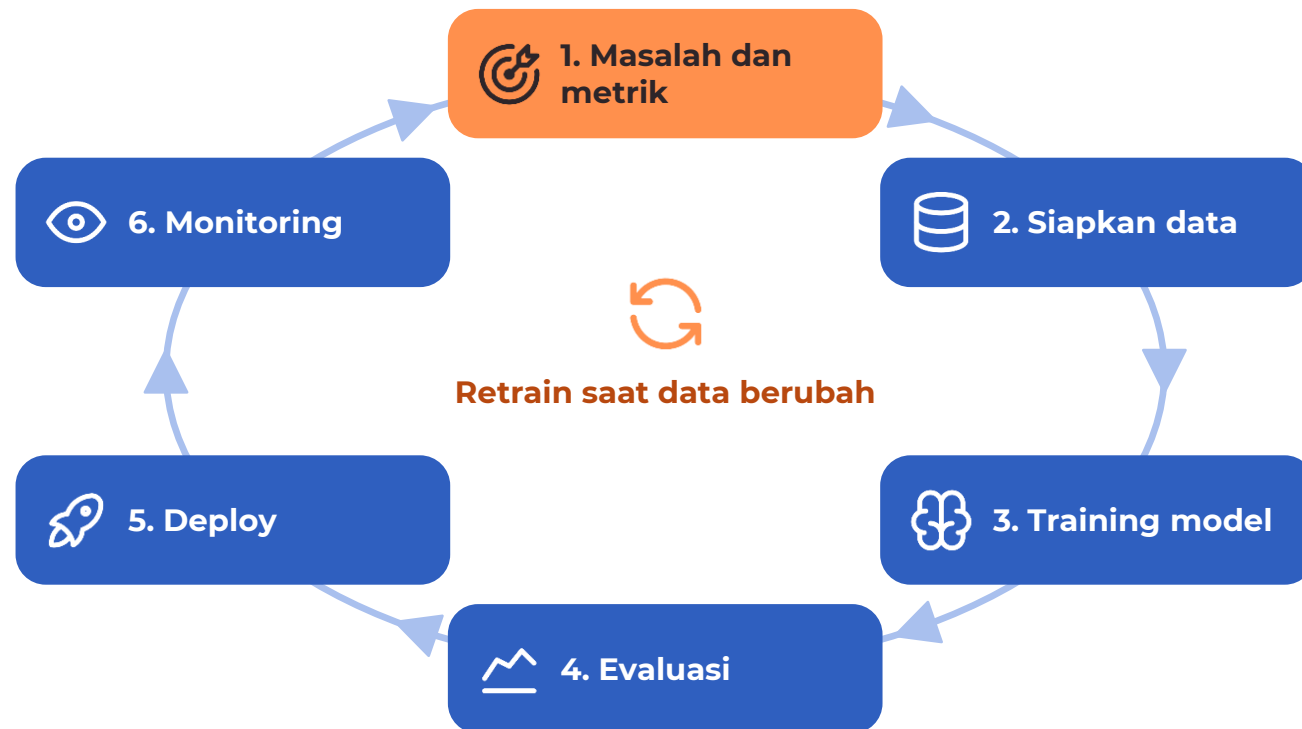
Dashboard cockpit atau dashboard bercerita



Tambahkan konteks: pola musiman, event belanja, biaya iklan. Lihat polanya dari helicopter view sebelum menyimpulkan.

Data · Google · Penyusun

Siklus model machine learning



Mulai dari metrik. Google Rules of ML: rancang metriknya dulu, dan model pertama sebaiknya sederhana.

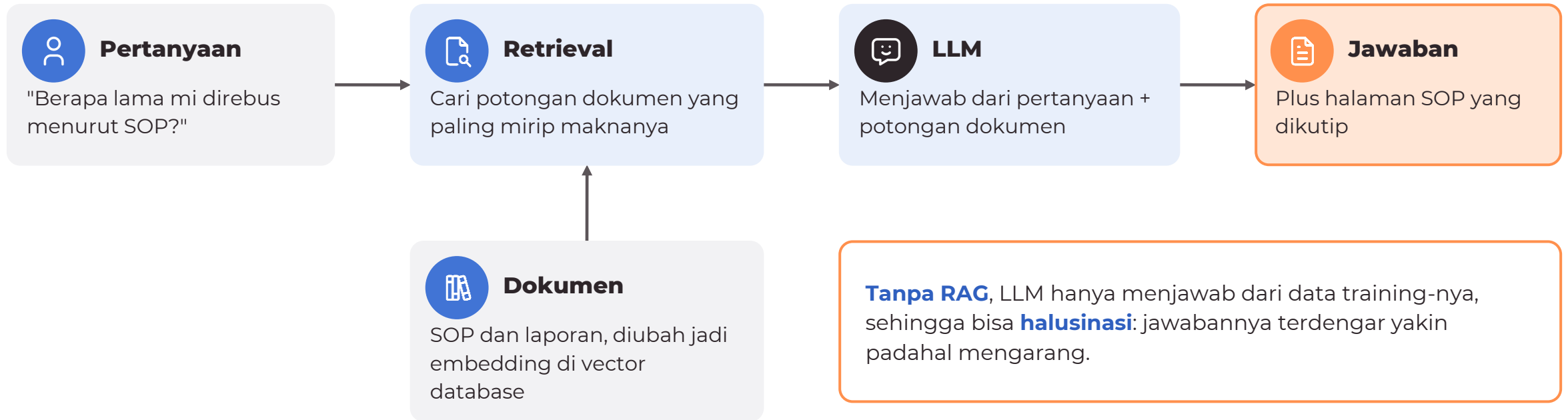
Data scientist biasanya memegang langkah 1–4.

ML engineer memegang deploy. **MLOps** mengotomasi monitoring dan retraining.

Menurut Google, performa model bisa menurun karena profil data terus berubah, sehingga model perlu monitoring dan retraining.

Penyusun • Data • Anthropic

LLM dan RAG dalam satu gambar



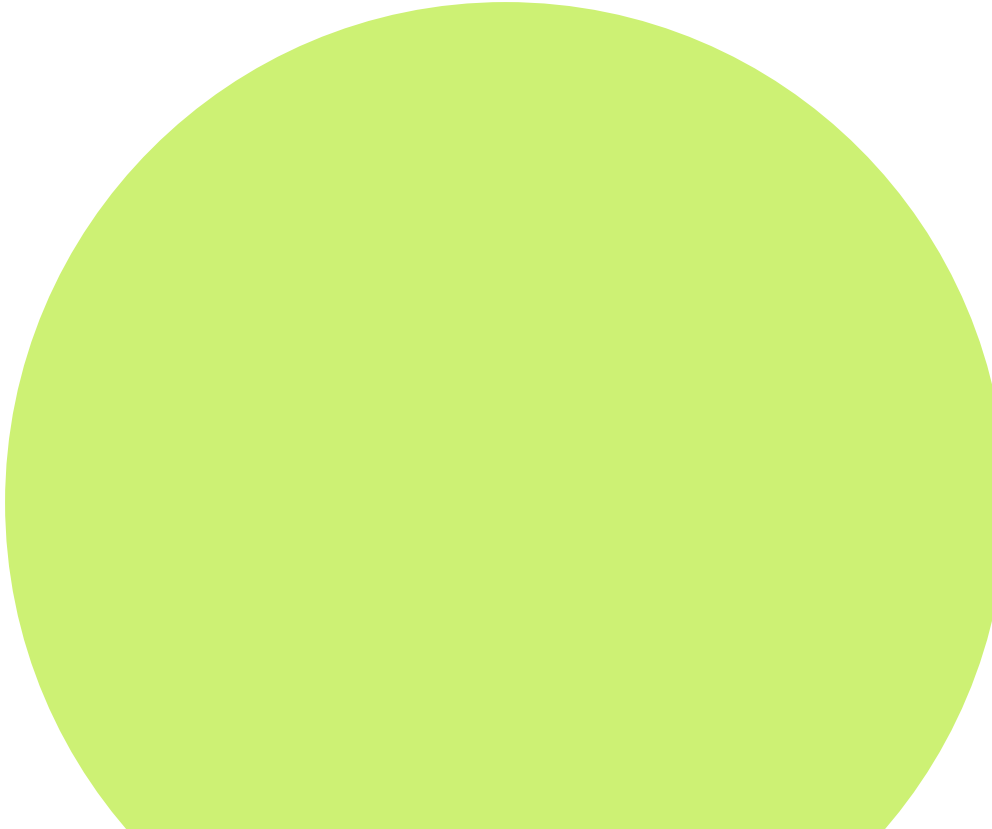
Evaluasi: siapkan daftar pertanyaan dengan jawaban yang benar, lalu ukur berulang. **Guardrail**: tetapkan apa yang tidak boleh dijawab.



Bagian 5

Alat di setiap lapisan

Tiap lapisan punya **satu pertanyaan**; konsepnya bisa dipelajari lewat satu alat gratis.



Penyusun · dicek 3 Okt 2026

13 lapisan alat dalam satu halaman

1. Sumber data	2. Ingestion	3. Orchestration	4. Storage	5. Transformasi	6. Big data, streaming	7. Analisis
SQL Server	SSIS	Airflow	BigQuery	SQL	Spark / PySpark	Python + pandas
PostgreSQL	Pentaho	n8n	Redshift	pandas	Hadoop / Hive	Excel
MySQL	Talend	Dagster	Snowflake	dbt Core	Kafka	R / RStudio
Oracle Database	Script Python	Prefect	ClickHouse	Dataform	Flink	Sheets
8. BI, visualisasi	9. Data science, ML	10. MLOps, deployment	11. AI generatif	12. Data governance	13. Kerja sama	
Power BI Desktop	TensorFlow / Keras	Kubernetes	Claude	Great Expectations	Google Cloud	
Tableau Public	PyTorch	FastAPI	OpenAI API	DataHub	Git	
Data Studio	scikit-learn	Kubeflow	LangChain	Collibra / Atlan	AWS	
Metabase	MLflow	Vertex AI	Gemini API	dbt tests	Azure	

Penyusun · orange = ramah pemula

Data lahir, dipindah, lalu dijadwalkan



1. Sumber data

Data lahir di mana?

Analogi: Mesin kasir: tiap pembelian jadi satu baris.

PostgreSQL

MySQL

SQL Server

Oracle Database

MongoDB

Excel / Sheets

API

Mulai dari: SQL di DuckDB, lalu PostgreSQL di laptop.



2. Ingestion

Bagaimana data berpindah?

Analogi: Kurir yang menjemput bahan dari banyak supplier.

Script Python

Airbyte

Fivetran

SSIS

Informatica

Talend

Pentaho

Mulai dari: Script Python di Colab, dari API ke database.



3. Orchestration

Kapan tiap langkah jalan?

Analogi: Manajer dapur yang memegang jadwal masak.

cron

Airflow

Prefect

Dagster

Cloud Composer

n8n

Mulai dari: Prefect dulu, Airflow sesudah paham scheduling.

64

Penyusun · oranye = ramah pemula

Data disimpan dan diolah jadi tabel rapi



4. Storage

Data dikumpulkan di mana?

Analogi: Gudang pusat: rak tertata atau gudang curah.

BigQuery

Snowflake

Redshift

ClickHouse

Databricks

DuckDB

S3

MinIO

Mulai dari: DuckDB, lalu BigQuery sandbox tanpa kartu kredit.



5. Transformasi

Bagaimana data mentah jadi rapi?

Analogi: Dapur persiapan: cuci, potong, siap masak.

SQL

dbt Core

Dataform

pandas

Power Query

Mulai dari: SQL, lalu dbt Core dengan DuckDB.



6. Big data dan streaming

Data raksasa atau real-time?

Analogi: Ban berjalan: pesanan diproses begitu masuk.

Spark / PySpark

Hadoop / Hive

Kafka

Flink

Pub/Sub

Debezium

Mulai dari: Belakangan, lewat PySpark di Colab.

65

Penyusun · oranye = ramah pemula

Data dipakai untuk analisis dan dashboard



7. Analisis

Apa yang terjadi, dan kenapa?

Analogi: Koki kepala membaca catatan penjualan.

Excel	Sheets
SQL di DuckDB	Python + pandas
R / RStudio	Jupyter / Google Colab
SAS / SPSS / Stata	

Mulai dari: Sheets, lalu SQL di DuckDB, lalu Python.



8. BI dan visualisasi

Bagaimana bisnis membaca hasilnya?

Analogi: Papan skor di dinding restoran.

Power BI Desktop	Tableau Public
Data Studio	Looker (enterprise)
Metabase	Superset
Qlik Sense	Grafana

Mulai dari: Power BI Desktop atau Tableau Public.

Penyusun • oranye = ramah pemula

Data dipakai untuk model dan AI



9. Data science dan ML

Bisakah data lama menebak esok?

Analogi: Juru masak senior menebak porsi besar.

scikit-learn	TensorFlow / Keras
PyTorch	XGBoost / LightGBM
Hugging Face	Kaggle
MLflow	Weights & Biases

Mulai dari: scikit-learn di Colab dengan data Kaggle.



10. MLOps dan deployment

Bagaimana model dipakai orang lain?

Analogi: Cabang baru dengan SOP yang sama.

FastAPI	Streamlit
Hugging Face Spaces	Kubernetes
Kubeflow	SageMaker
Vertex AI	

Mulai dari: Streamlit di Community Cloud.



11. AI generatif

AI menjawab dari dokumen kita?

Analogi: Asisten yang membuka buku SOP dulu (RAG).

Gemini API	OpenAI API
Claude	Ollama
LangChain	LlamaIndex
Chroma / Qdrant	Pinecone

Mulai dari: Ollama atau Gemini API, lalu RAG kecil.

Penyusun • oranye = ramah pemula

Penjaga data dan alat lintas lapisan



12. Data governance dan data quality

Datanya benar? Siapa boleh lihat?

Analogi: Inspektur kebersihan dapur.

dbt tests	Great Expectations
DataHub	OpenMetadata
Collibra / Atlan	Purview

Mulai dari: Cek dobel dan kosong dengan SQL, lalu tes dbt.



13. Kerja sama dan fondasi

Alat yang dipakai di semua lapisan?

Analogi: Git: buku resep bersama. Docker: kotak bekal.

Git	GitHub
Docker Desktop	Linux CLI
Google Cloud	AWS
Azure	

Mulai dari: Git dan GitHub sejak proyek pertama.

Penyusun • semua bisa dimulai gratis

Starter kit gratis per jalur

Kebutuhan	Data analyst	Data engineer	Data scientist / ML	AI engineer
Data latihan	Kaggle*, data pemerintah	API publik, CSV	Kaggle*	PDF publik
Database	DuckDB	PostgreSQL atau Neon Free	DuckDB	PostgreSQL
Alat utama	Sheets, Python di Colab	Python, Prefect, dbt Core	pandas, scikit-learn, MLflow	Ollama atau Gemini API*, Chroma
Tampilan hasil	Power BI atau Tableau Public	Diagram pipeline, Data Studio	Streamlit	Streamlit, Hugging Face Spaces
Kerja sama	GitHub	Git, GitHub, Docker	GitHub	Git, GitHub, Docker

*Syarat belum dicek, cek sendiri. Power BI Desktop hanya untuk Windows. Karya di Tableau Public terbuka untuk umum: jangan upload data pribadi.

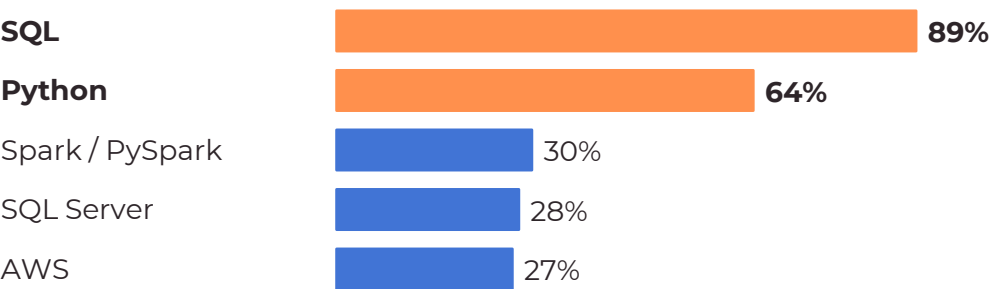
Data • lowongan 3 Okt 2026

Lima alat yang paling sering diminta lowongan, per peran

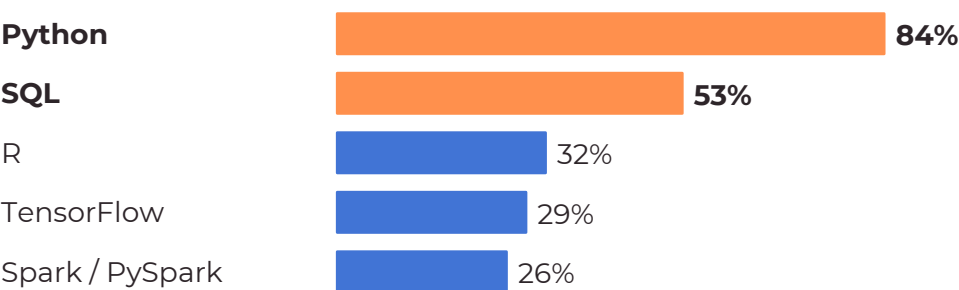
Data analyst n = 151 lowongan



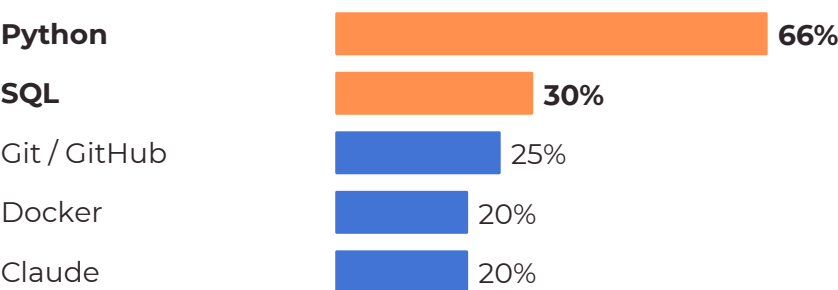
Data engineer n = 74 lowongan



Data scientist / ML n = 38 lowongan, n kecil



AI engineer n = 44 lowongan, n kecil



Oranye: SQL dan Python masuk lima besar di keempat peran.

Praktisi · sesi DA dan DE

Konsep lebih awet daripada alat



Narasumber DA

“Yang terpenting dari data analyst adalah cara berpikir kritisnya.”



Narasumber DE

“Kuasai sistemnya, bukan tools-nya. Airflow terlalu berat? Pakai Prefect. Yang memilih tetap kita.”



Host sesi DE

“Belajar satu tools membuat pindah ke tools lain lebih cepat, karena fondasinya mirip.”

Buktinya: alat memang berganti

Talend Open Studio kini dialihkan ke versi cloud berbayar.

Looker Studio kembali bernama **Data Studio** sejak April 2026.

Repositori **MinIO** kini bertuliskan: tidak lagi dirawat.

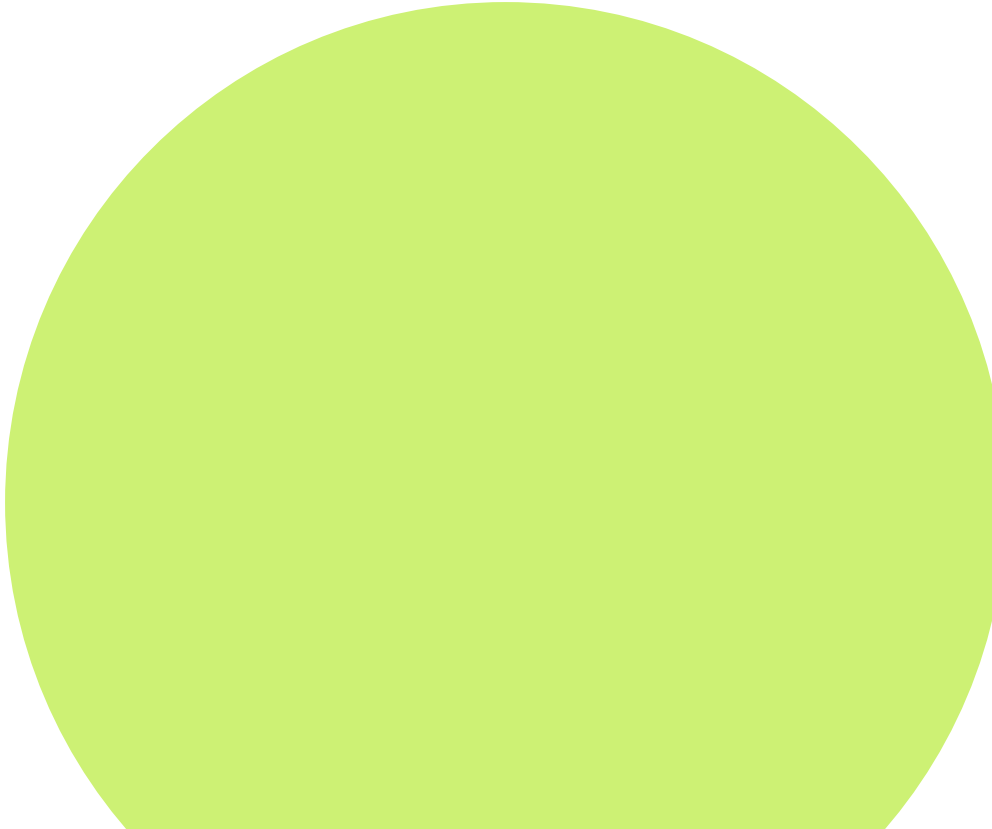
Data · halaman resmi, dibuka 3 Okt 2026



Bagian 6

Roadmap belajar

Fondasi yang sama untuk semua peran, lalu **tiga anak tangga** per jalur: fondasi, inti, lanjut.



Praktisi · sesi DA dan DE

Lima fondasi yang dipakai semua peran



Berpikir kritis

Ubah "tolong tarik data" jadi pertanyaan keputusan. Pakai 5W1H.



SQL

"Bahasa utama orang data". Mulai dari SELECT sampai CTE.



Spreadsheet

Banyak kantor masih hanya memakai Excel. Pivot, lookup, grafik.



Statistik dasar

Rata-rata, median, outlier, korelasi bukan sebab.



Komunikasi

Jelaskan satu temuan dalam tiga kalimat.

Data: SQL diminta di 60% lowongan DA, 89% DE, 82% DS. Snapshot 3 Okt 2026.

Praktisi · sesi Data Analyst

Roadmap data analyst

1 Fondasi

5W1H, spreadsheet, statistik deskriptif.

Cukup kalau: Bisa membuat pivot dan tahu kapan rata-rata menyedatkan.

 AI membuat soal latihan.

2 Inti

SQL sampai CTE, data modeling, satu BI tool.

Cukup kalau: Dashboard dari dua tabel atau lebih, dengan pembandingan.

 AI membuat draft query, kita menguji.

3 Lanjut

Python, dasar ETL, A/B test, satu domain bisnis.

Cukup kalau: Report rutin jalan otomatis.

 Kode pandas dari AI kita uji.

Praktisi · sesi Data Engineer

Roadmap data engineer

1 Fondasi

CLI Linux, Git, Python, Docker, SQL. "Yang penting tahu dulu."

Cukup kalau: Database jalan di Docker, satu file masuk tabel.

 AI menjelaskan error.

2 Inti

Pipeline sederhana: API ke database, lalu query. ETL vs ELT.

Cukup kalau: Satu pipeline end-to-end terjadwal di laptop.

 AI membantu query standar.

3 Lanjut

Cloud warehouse, object storage, streaming.

Cukup kalau: Memilih alat sesuai skala, tanpa overkill.

 Kita yang memilih alat.

Penyusun · dari Google ML Crash Course

Roadmap data scientist dan ML engineer

1 Fondasi

Python, SQL, statistik, aljabar dasar.

Cukup kalau: Membersihkan dan meringkas data di pandas.

 AI menjelaskan konsep.

2 Inti

Regresi, klasifikasi, evaluasi, overfitting.
Mulai dari **baseline**.

Cukup kalau: Membandingkan model dengan baseline.

 Kode dari AI tetap diuji.

3 Lanjut

Neural network, embedding. ML engineer: API, Docker, monitoring.

Cukup kalau: Model dipakai orang lain, lengkap dengan metrik.

 AI membantu debugging.

Penyusun · dokumentasi resmi vendor

Roadmap AI engineer

</> 1 Fondasi

Python, API, JSON, Git, dasar LLM.

Cukup kalau: Memanggil model lewat API.

 AI menulis kode dasar.

✦ 2 Inti

RAG sederhana, set evaluasi, guardrail.

Cukup kalau: Satu aplikasi dengan angka evaluasi.

 AI mengusulkan kasus uji.

3 Lanjut

Workflow vs agent, tool use, biaya, prompt injection.

Cukup kalau: Bisa menjelaskan pilihan workflow atau agent.

 AI membantu riset dokumentasi.

Praktisi • sesi DA dan DE

SQL sampai mana yang cukup?



Data analyst, minimal: sampai CTE, supaya bisa membaca query buatan AI (narasumber DA)

Host sesi DA: intermediate lebih baik

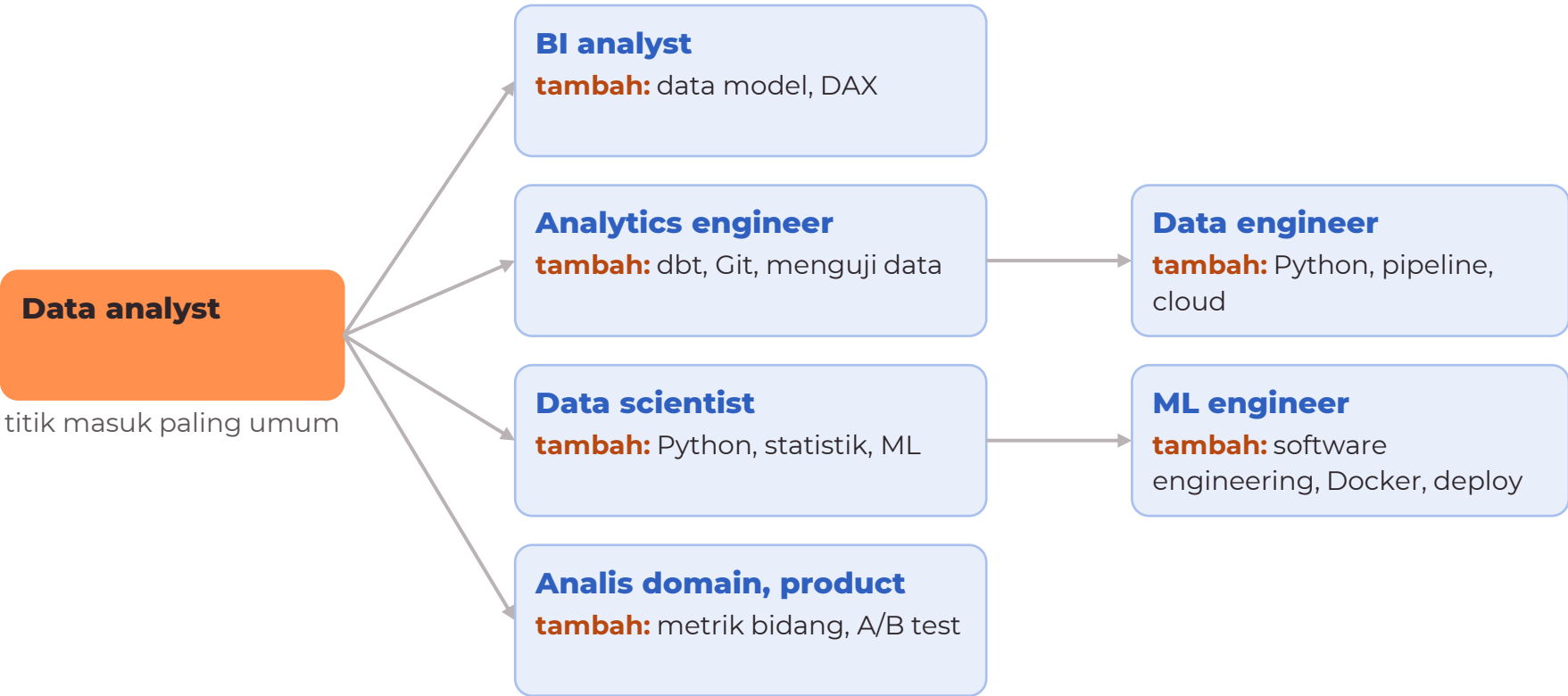
Data engineer: ditambah INSERT, UPDATE, CREATE TABLE (narasumber DE)

Analytics engineer: sampai window function (penyusun)

Data • lowongan 3 Okt 2026. SQL diminta di **60%** lowongan DA (n = 151), **89%** lowongan DE (n = 74), dan **82%** lowongan DS (n = 22).

Penyusun · belum ada data perpindahan

Dari DA bisa pindah ke mana saja?



Data. SQL juga diminta di 89% lowongan DE dan 82% lowongan DS, tapi bukti perpindahannya belum ada.

Praktisi. Narasumber DA masuk lewat market analyst, narasumber DE lewat ETL developer.

Praktisi · sesi Data Analyst

Kapan AI membantu belajar, kapan merugikan



AI sebagai lawan

Seluruh data dilempar ke AI, lalu hasilnya ditelan mentah-mentah.

Data perusahaan ter-upload ke chatbot, jadi rawan bocor.

Kita berhenti berpikir kritis, padahal AI bisa halusinasi.



AI sebagai kawan

AI membuat draft query, lalu kita baca, jalankan, uji.

Kita berdiskusi soal logika query dan formula metrik.

AI mempercepat riset dan dokumentasi teknis.

Narasumber DE: pakai AI sebagai **akselerator**, lalu verifikasi sendiri, karena penentu kebenarannya tetap kita.

Penyusun · dari saran praktisi

30 hari pertama menuju satu portfolio



Minggu 1

Pilih jalur

Jawab pohon di bagian 9, pasang satu alat, kumpulkan 10 lowongan.



Minggu 2

Fondasi

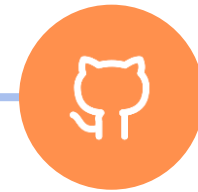
Belajar SQL dasar di DuckDB, atau Python dasar di Colab, sesuai jalur.



Minggu 3

Proyek kecil

Ambil dataset publik dari bidang yang kita kenal, lalu jawab satu pertanyaan bisnis.



Minggu 4

Terbitkan

Dashboard, notebook, atau pipeline di GitHub plus tiga kalimat temuan; bagikan di LinkedIn.

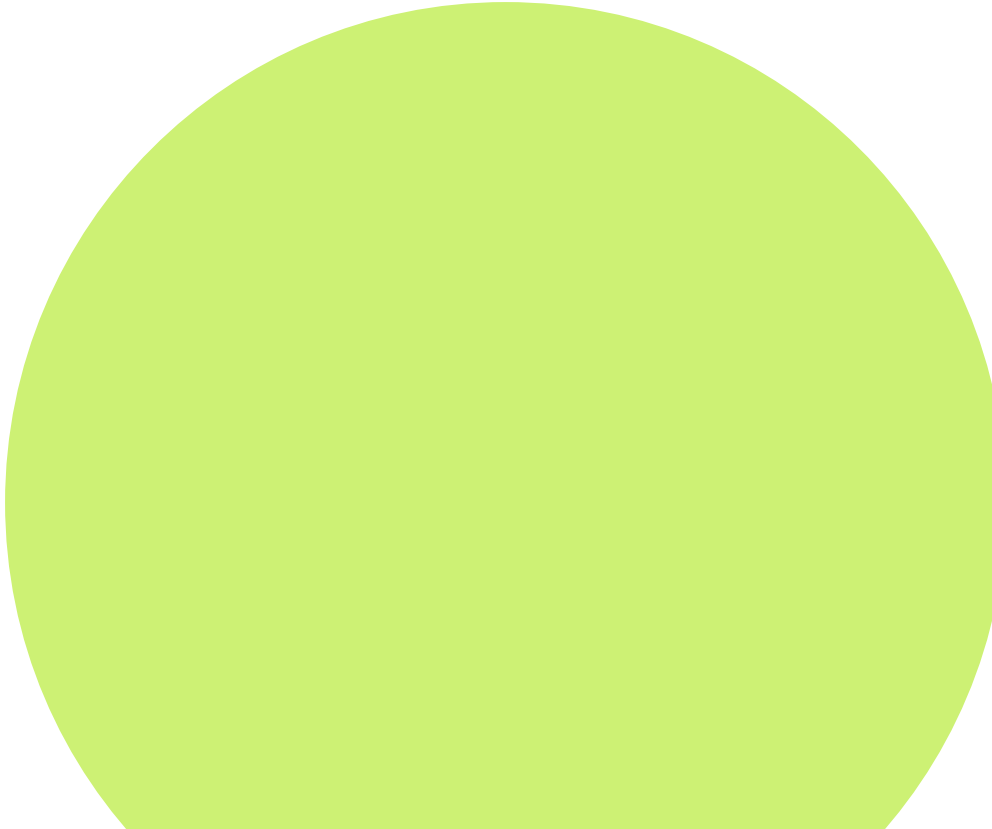
Ulangi minggu 3–4 dengan dataset lain. Tantangan host sesi DA: **satu portfolio per minggu**.



Bagian 7

Pasar kerja Indonesia

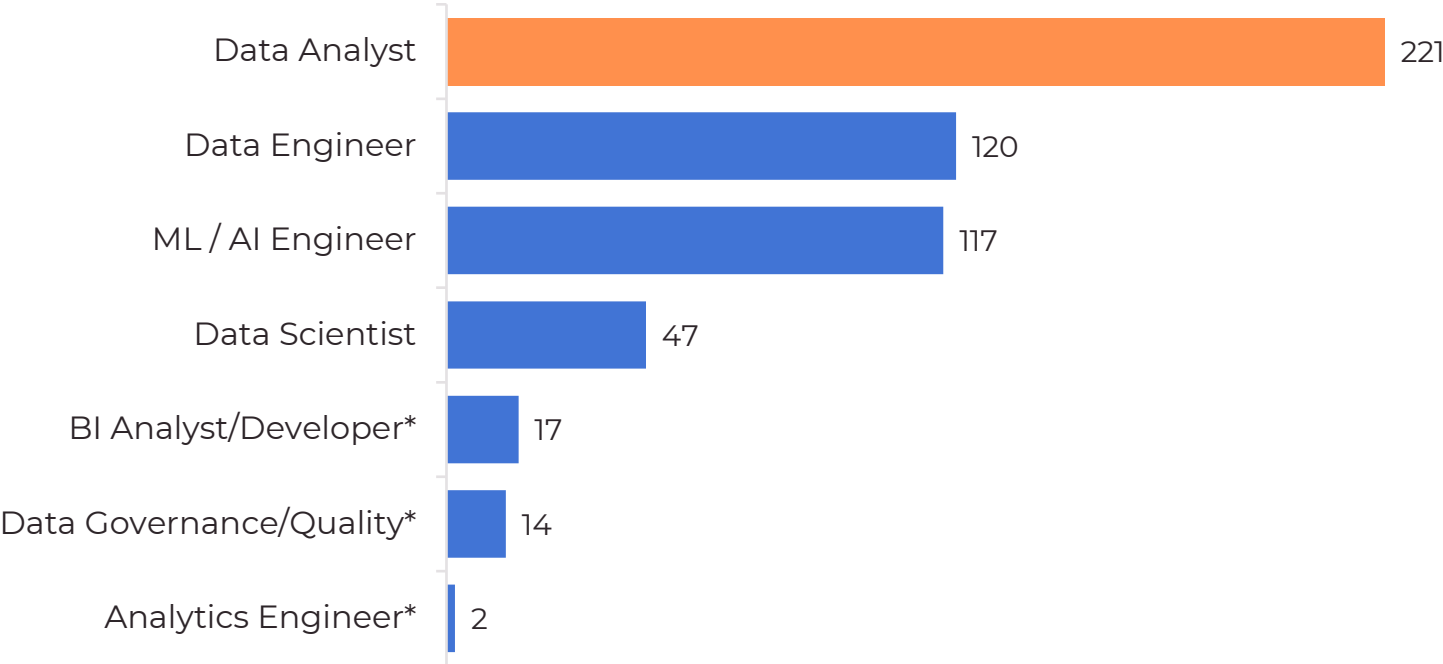
Potret **538 lowongan** peran data pada 3 Oktober 2026,
diambil dalam satu hari tanpa LinkedIn.



Data • lowongan 3 Okt 2026

221 dari 538 lowongan peran data untuk data analyst

Lowongan per peran, 9 situs lowongan publik (n = 538). *n kecil: petunjuk arah saja



Satu hari saja. Angka ini bukan total lowongan di Indonesia.

ML/AI engineer 117. Sebanyak 94 judulnya hanya menyebut AI, tanpa machine learning.

LinkedIn tidak ikut. Potret ini kemungkinan lebih ramah pemula daripada pasar sebenarnya.

Di luar hitungan ini ada **335 lowongan peran dekat data**, misalnya admin data, business analyst, dan AI data ops.

Data • lowongan 3 Okt 2026

Banyak lowongan tetap minta pengalaman 1–3 tahun

Porsi lowongan dengan syarat pengalaman paling lama 1 tahun

Data analyst



Data engineer



ML/AI engineer



Data scientist



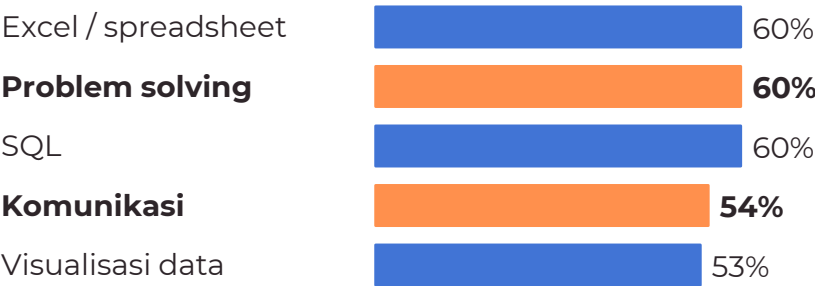
***Kemungkinan terlalu tinggi.** Glints menulis "1–3 tahun", tapi tercatat 1.

Lowongan DA yang menulis syaratnya sendiri: **median 2 tahun** (n = 48). Portfolio bisa menggantikan pengalaman.

Data • lowongan 3 Okt 2026

Skill yang paling sering diminta, per peran

Data analyst n = 151 lowongan



ML/AI engineer n = 60 lowongan



Oranye: skill **non-teknis**.

Data engineer n = 74 lowongan



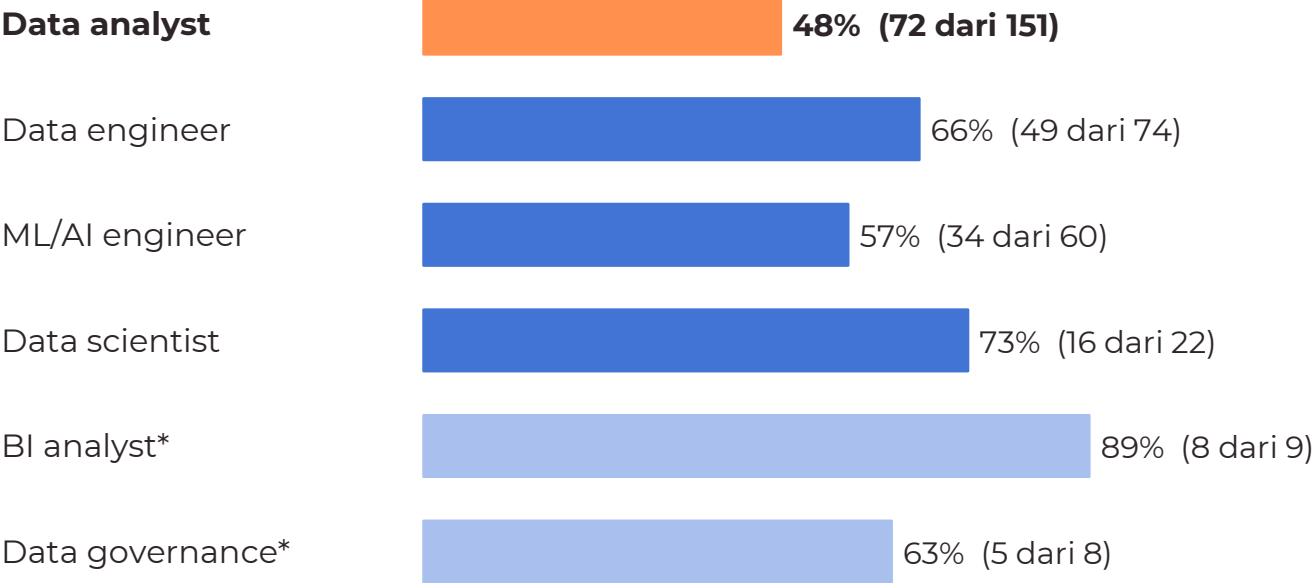
Data scientist n = 22 lowongan, n kecil



Data • lowongan 3 Okt 2026

Hampir separuh lowongan DA ditulis dalam bahasa Inggris

Porsi lowongan berbahasa Inggris



*n kecil: petunjuk arah saja

Ini bahasa **penulisan** lowongan; yang **mensyaratkan** bahasa Inggris di lowongan DA hanya **9%** (14 dari 151).

Penyusun: biasanya membaca lowongan dan menulis CV dalam bahasa Inggris.

Data • lowongan 3 Okt 2026

Seberapa terbuka pintu masuk data analyst?



Syaratnya

Dua pertiga lowongan DA **tidak** terbuka untuk fresh graduate.

15 dari 151 menerima fresh graduate **dengan** magang atau portfolio kuat.

9% lowongan berjudul DA bukan analisis data. Baca isinya.



Yang mendukung

Lowongan terbanyak: 221 dari 538.

Syarat alat paling ringan: Python di 30% lowongan DA, 91% di DS.

Paling terbuka untuk fresh graduate: 32%. DE 15%, DS 23%, ML/AI 17%.

Syarat alatnya ringan, tapi hanya **sepertiga** lowongan DA terbuka untuk fresh graduate.

Penyusun · cara snapshot, versi manual

Cek lowongan sendiri, sebulan sekali



Cari di dua atau tiga situs

LinkedIn, Glints, Jobstreet. Cari juga "intern" dan "magang".



Baca isi lowongan, jangan hanya judulnya

Pekerjaan DA bisa berjudul MIS atau reporting analyst.



Catat 10 lowongan di spreadsheet

Judul, skill, alat, pengalaman, terbuka fresh graduate.



Hitung yang paling sering muncul

Skill yang berulang jadi daftar belajar bulan ini.



Ulangi bulan depan

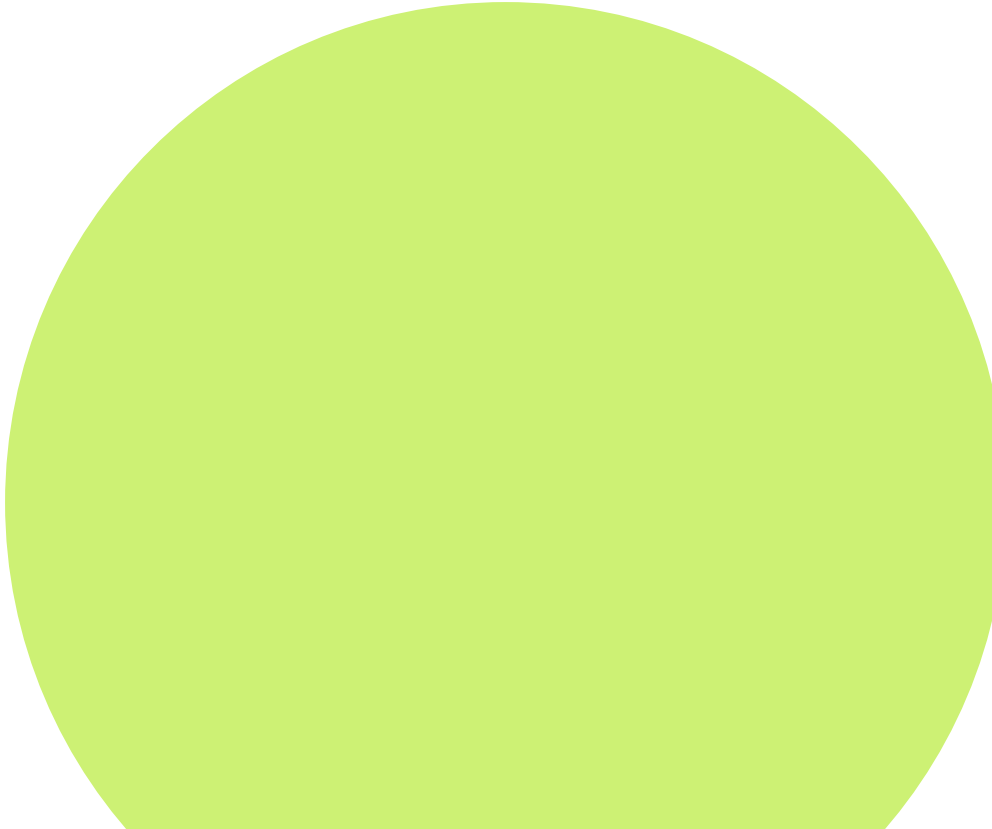
Handout ini hanya potret satu hari.



Bagian 8

Peran manusia di era AI

AI mempercepat pekerjaan, tapi **keputusan dan tanggung jawab** tetap milik kita.



Praktisi dan penyusun • lihat kolom Label

Yang dibantu AI, dan yang tetap di manusia

Peran	Dibantu AI	Tetap di manusia	Label
Data analyst	Draft query, riset, dokumentasi	Menggali kebutuhan, validasi query, keputusan	Praktisi
Data engineer	Query standar, dokumentasi, error sintaks	Konteks bisnis, infrastruktur, memilih alat	Praktisi
BI analyst	Report dan DAX (fitur berbayar)	Satu definisi metrik	Penyusun
Data scientist	Kode analisis dan model	Memilih masalah, mengecek asumsi	Penyusun
ML engineer	Kode dan tes	Keputusan rilis, cek kode "hampir benar"	Penyusun
AI engineer	Kode, usulan kasus uji	Kriteria sukses, batas perilaku sistem	Penyusun
Data governance	Draft data dictionary	Definisi resmi, hak akses	Penyusun

Praktisi · sesi DA dan DE

Teknologi mengambil tugas, belum tentu perannya



ATM dan teller bank

Saat ATM muncul, teller diduga hilang, padahal menurut host masih ada.

Pesan: tugas diambil, peran tetap.



Pilot dan autopilot

Kokpit makin otomatis, pilot tetap ada untuk memutuskan.

Pesan: pegang keputusan.



Kalkulator dan akuntan

Hitungan manual diambil alih, akuntan tetap ada.

Pesan: kuasai pemahamannya.

Analogi dari pengamatan, belum didukung data.

Praktisi · sesi DA dan DE

AI menulis draft, kita yang mengecek dan memutuskan



AI membuat draft

Query, rumus, kode, ringkasan.



Baca dulu

Paham setiap baris? Kalau belum, jangan dipakai dulu.



Uji

Jalankan, cocokkan dengan angka yang kita tahu, hitung dua cara.



Putuskan

Keputusan dan tanggung jawab ke manajemen tetap milik kita.

Host sesi DA: untuk query multi-tabel AI **tidak konsisten**, karena prompt yang sama bisa menghasilkan query berbeda. Menjawab "kata AI begitu" ke manajemen itu blunder.

Praktisi • sesi DA; UU PDP

Aturan aman memakai AI di kantor



Jangan upload data perusahaan atau data pribadi ke chatbot publik

Data rawan bocor. Data pribadi dilindungi UU 27/2022 (PDP).



Anggap jawaban AI sebagai draft

AI bisa halusinasi.



Uji sebelum dipakai

Jalankan, lalu cocokkan dengan angka yang kita tahu.



Simpan versi yang sudah dicek

Prompt sama, hasil bisa berbeda.



Tanggung jawab tetap milik kita

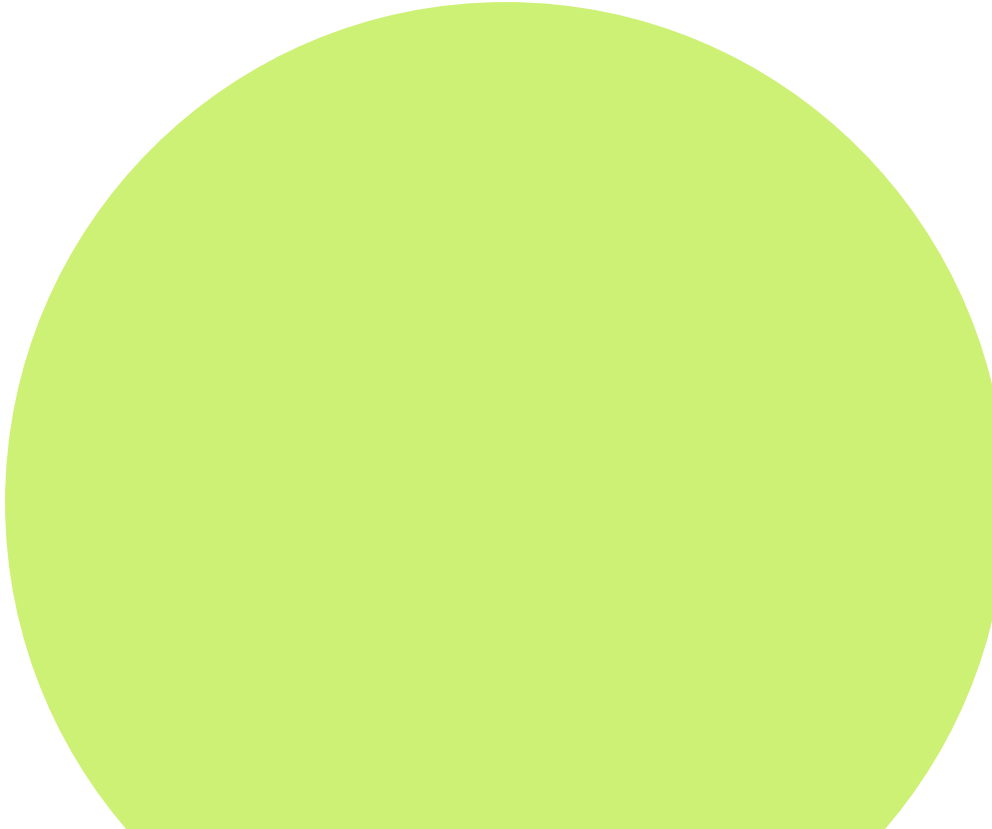
Di depan manajemen, kita yang ditanya.



Bagian 9

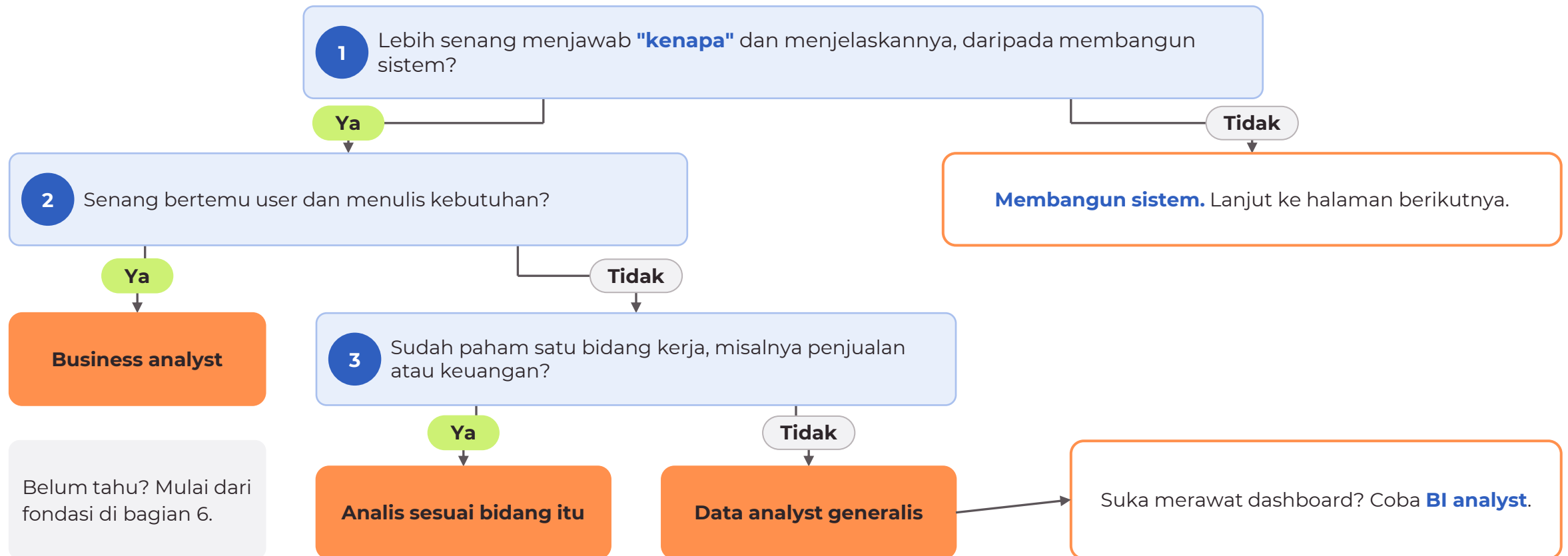
Memilih jalur dan langkah berikut

Jawab beberapa pertanyaan ya atau tidak. Pilih **satu** jalur untuk mulai minggu ini. Pilihan boleh berubah.



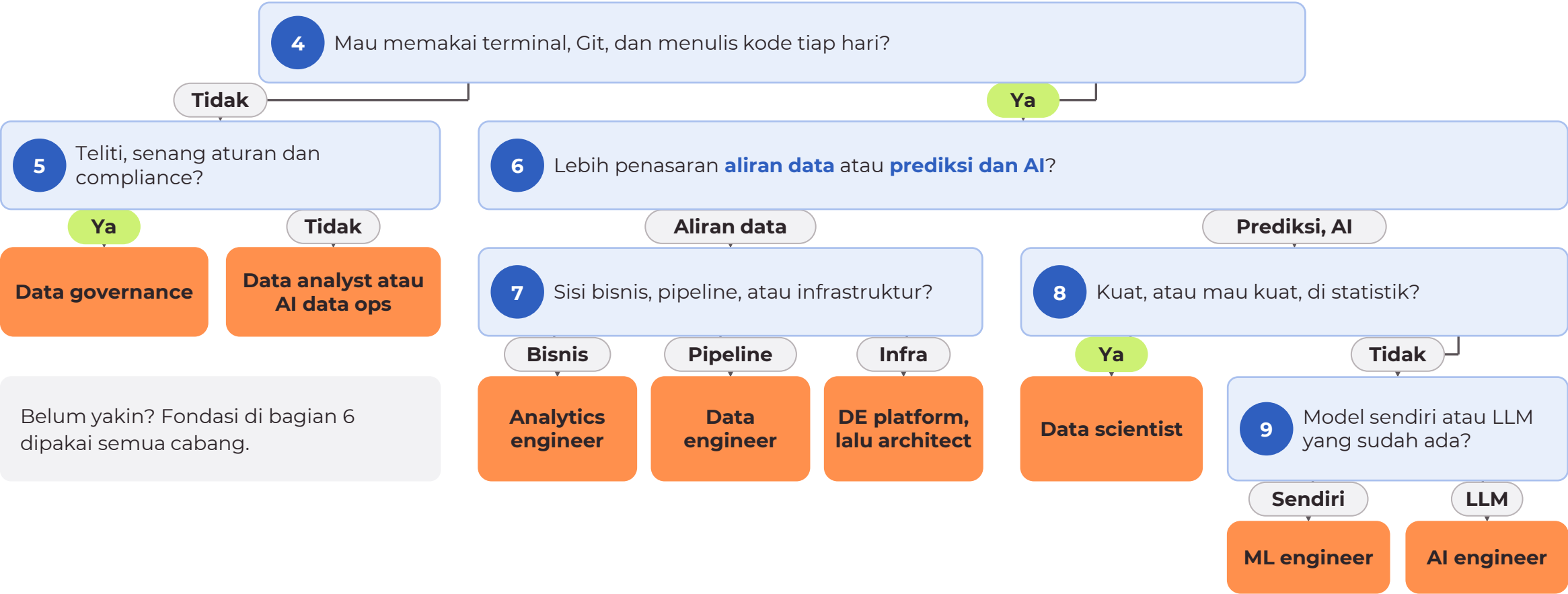
Penyusun • bukan tes psikometri

Cocok di jalur mana? Mulai dari pertanyaan pertama



Penyusun • bukan tes psikometri

Membangun sistem untuk data, model, atau AI?



Penyusun • DA dan DE dari praktisi

Satu ide portfolio per peran



Data analyst

Dashboard dengan pembandingan, plus memo satu halaman.



BI analyst

Dashboard berhierarki, plus daftar measure.



Data engineer

Pipeline end-to-end terjadwal di GitHub.



Analytics engineer

Proyek dbt bronze-silver-gold dengan tes data.



Data scientist

Prediksi dari data publik vs baseline.



ML engineer

Model kecil sebagai API di Docker.



AI engineer

Tanya-jawab dokumen, dengan 30 pertanyaan uji.



Data governance

Data dictionary dan aturan data quality.

Oranye: saran praktisi. Ceritakan prosesnya, termasuk kesulitannya.

Praktisi • sesi DA dan DE

Kesalahan umum lintas peran

Kesalahan	Paling sering di	Isinya
Cockpit syndrome	DA, BI	Semua grafik di satu dashboard, tanpa hierarki
Dashboard tanpa pembanding	DA, BI	Hanya total, tanpa target
Confirmation bias	DA, DS	Hanya mencari data pendukung dugaan
Tool-oriented	Semua	Mengoleksi nama alat, lupa fondasi
Lupa komunikasi	Peran teknis	Hasil yang tidak dipahami tidak dipakai
Biaya cloud bengkak	DE, ML, AI	Server kebesaran, query berat
Overkill	DE, MLOps, AI	Alat besar untuk masalah kecil
Menelan output AI mentah	Semua	Tanpa cek, tanpa tanggung jawab

Praktisi · sesi DA dan DE

Supaya dilirik recruiter



LinkedIn

Rajin bagikan portfolio. Host mendapat pekerjaan berikutnya lewat LinkedIn karena rutin berbagi portfolio.



Magang

Pekerjaan pertama host berawal dari magang kampus; lowongan magang DA masih ada.



Komunitas data

Narasumber DA menyarankan gabung komunitas: dapat insight dan info lowongan.



Cerita portfolio

Interviewer melihat portfolio sekaligus cara kita menceritakannya, termasuk untuk fresh graduate.

Ini pengalaman pribadi praktisi. Pintu masuk tiap orang berbeda.

Penyusun · tiap butir kurang dari sejam

Checklist minggu ini

- ☐ **1** Jawab pohon, pilih **satu** jalur untuk 30 hari.
- ☐ **2** Pasang satu alat gratis, jalankan satu contoh.
- ☐ **3** Kumpulkan 10 lowongan, tandai skill yang berulang.
- ☐ **4** Pilih satu dataset publik dari bidang sendiri.
- ☐ **5** Rapikan LinkedIn, buat GitHub, gabung komunitas data.



Minggu depan

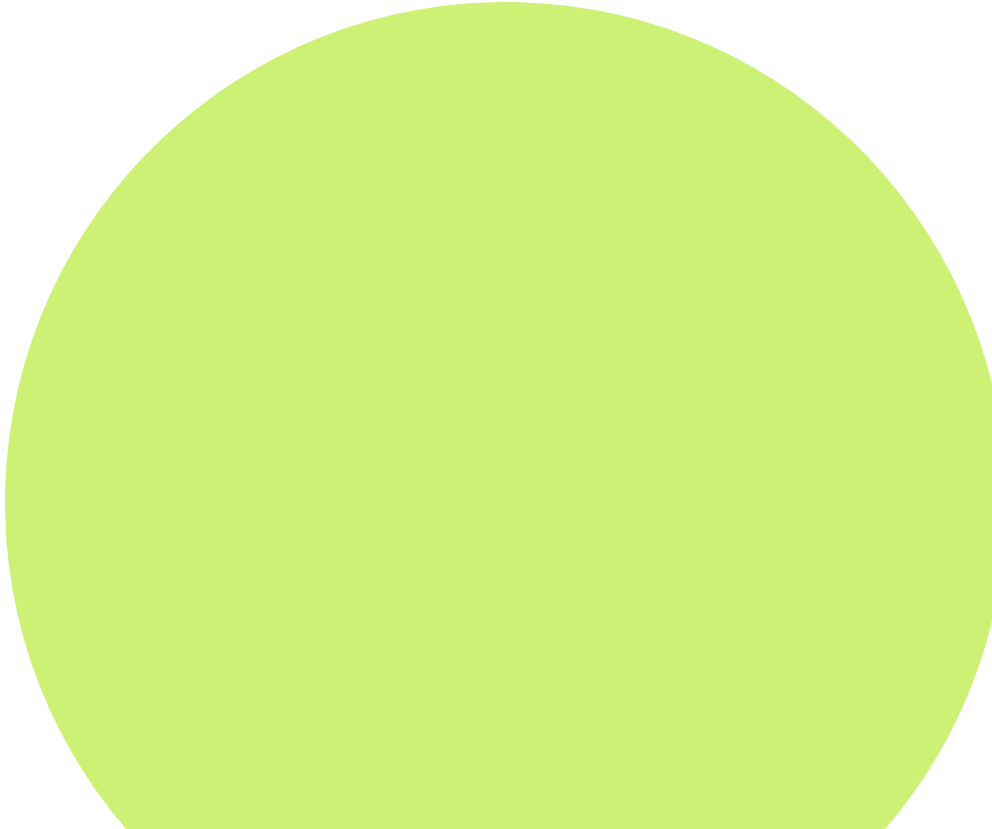
Lanjutkan rencana **30 hari** di bagian 6. Buka **kamus** di bagian 10 saat ada istilah asing.



Bagian 10

Kamus

258 istilah dalam 10 kelompok, dengan arti, contoh, dan peran yang memakainya. Buka saat ada kata yang terasa asing.



Kamus · 258 istilah

Sepuluh kelompok istilah



1

Peran dan dunia kerja data

28 istilah · hal. 103–106



2

Sumber data dan storage

28 istilah · hal. 107–110



3

Pipeline dan infrastruktur

28 istilah · hal. 111–114



4

Data cleaning, SQL, data modeling

33 istilah · hal. 115–119



5

Analisis dan critical thinking

34 istilah · hal. 120–124



6

Visualisasi dan dashboard

20 istilah · hal. 125–127



7

Data science dan machine learning

19 istilah · hal. 128–130



8

AI generatif dan LLM

20 istilah · hal. 131–133



9

Data governance, privasi, MLOps

18 istilah · hal. 134–136



10

Karier dan melamar

30 istilah · hal. 137–141

1 • Peran dan dunia kerja data (1/4)

Istilah	Artinya	Contoh	Peran
Data	Catatan tentang kejadian, orang, atau benda yang bisa dihitung, dibandingkan, dan dibaca ulang.	struk belanja, daftar absensi, catatan harga beras di pasar	Semua
Dataset	Kumpulan data tentang satu hal yang disimpan bersama, biasanya dalam satu atau beberapa tabel.	data harga pangan dari pasar-pasar di Indonesia sejak 2007 (data WFP)	Semua
Insight	Temuan yang menjelaskan kenapa sesuatu terjadi, mengubah pemahaman, dan menunjuk ke hal yang bisa diubah.	'penjualan turun' itu observasi; 'turun karena dua toko terbesar kehabisan stok' itu insight	DA, BI, BA, DS
Stakeholder	Orang yang berkepentingan dengan hasil analisis: yang meminta, yang memutuskan, atau yang terkena dampaknya.	manajer gudang yang meminta laporan stok, dan tim pembelian yang memakainya	Semua
User	Orang yang memakai hasil kerja tim data, misalnya membaca dashboard atau memakai angka di laporan.	manajer toko yang membuka dashboard penjualan setiap pagi	Semua
Konteks bisnis	Paham apa yang dikejar bisnis, kenapa satu angka penting, dan keputusan apa yang menunggu hasil analisis.	tahu bahwa penjualan naik saat promo besar itu wajar, belum tentu karena kampanye baru	Semua
Domain knowledge	Pengetahuan tentang bidang asal data, misalnya kesehatan, perbankan, atau pertanian, yang jadi nilai tambah saat pindah ke data.	mantan perawat lebih cepat paham data rumah sakit	Semua

1 • Peran dan dunia kerja data (2/4)

Istilah	Artinya	Contoh	Peran
Data workflow	Urutan kerja data dari sumber sampai jadi keputusan: diambil, dirapikan, dianalisis, lalu disajikan.	data kasir masuk data warehouse, dirapikan, dianalisis, lalu jadi laporan bulanan	Semua
Upstream dan downstream	Di upstream (hulu), data diambil dan disiapkan; di downstream (hilir), data dipakai orang.	DE bekerja di upstream menyiapkan data; tim aplikasi di downstream menampilkannya ke user	Semua
Backend dan UI	Bagian aplikasi di belakang layar yang mengolah dan menyimpan data disebut backend; tampilan yang dilihat user disebut UI.	DA menulis query metrik, tim backend memasangnya ke dashboard di halaman web	DA, DE
Data analyst (DA)	Orang yang mencari jawaban dari data untuk tim bisnis, lalu menjelaskannya ke yang memutuskan.	mencari cabang yang penjualannya turun, lalu menjelaskan sebabnya ke manajer area	DA
Business analyst (BA)	Orang yang menerjemahkan kebutuhan bisnis jadi kebutuhan proses atau sistem, dan menjembatani tim bisnis dengan tim teknis.	menuliskan kebutuhan laporan tim keuangan supaya tim IT bisa membangunnya	BA
BI analyst	Orang yang membuat dan merawat dashboard serta report rutin, supaya orang bisa melihat angka sendiri tanpa bertanya.	merawat dashboard target vs realisasi penjualan yang dibuka setiap Senin	BI
Product analyst	Analisis yang mengukur cara orang memakai aplikasi dan fitur mana yang berhasil.	membandingkan jumlah user yang selesai mendaftar sebelum dan sesudah formulir diubah	DA

1 • Peran dan dunia kerja data (3/4)

Istilah	Artinya	Contoh	Peran
Market analyst	Analisis yang membaca pergerakan harga, permintaan, dan pasokan pasar, lalu menjelaskan penyebabnya.	menjelaskan harga komoditas melonjak karena stok menipis dan pengiriman tertahan	DA, BA
Financial analyst	Analisis yang membaca laporan keuangan, anggaran, dan risiko untuk membantu keputusan soal uang.	membandingkan biaya operasional tiap cabang dengan anggarannya setiap bulan	DA, BA
Data engineer (DE)	Orang yang membangun dan merawat pipeline dari sumber ke data warehouse, supaya data sampai tepat waktu, lengkap, dan rapi.	memastikan transaksi kemarin sudah masuk sebelum analisis mulai kerja pagi	DE
ETL developer	Sebutan lama untuk DE yang fokus membuat proses ambil, rapikan, dan simpan data; isi kerjanya mirip DE.	membuat proses malam hari yang menyalin data penjualan ke data warehouse	DE
Data infrastructure	Server, storage, dan jaringan tempat data disimpan dan diolah; DE yang fokus di sini menjaga mesinnya tetap hidup.	menambah kapasitas server data warehouse saat volume transaksi melonjak	DE
DevOps	Cara kerja dan peran yang mengurus server, pemasangan aplikasi, dan otomatisasinya, supaya sistem jalan stabil.	menyiapkan server baru dan memastikan aplikasi menyala lagi sendiri kalau mati	DE, MLE
Analytics engineer (AE)	Orang yang merapikan dan menguji tabel di data warehouse, supaya semua analisis memakai definisi angka yang sama.	satu tabel resmi 'pelanggan aktif' yang dipakai semua dashboard, lengkap dengan tes otomatis	AE

1 • Peran dan dunia kerja data (4/4)

Istilah	Artinya	Contoh	Peran
Data scientist (DS)	Orang yang membuat model untuk memprediksi atau mengelompokkan, biasanya dengan Python dan machine learning.	memperkirakan pelanggan mana yang akan berhenti berlangganan bulan depan	DS
ML engineer (MLE)	Orang yang memasang model machine learning ke aplikasi atau pipeline, supaya model jalan otomatis dan terawat.	fitur rekomendasi produk di aplikasi belanja yang di-update setiap hari	MLE
AI engineer (AIE)	Orang yang membangun aplikasi berbasis AI generatif, misalnya chatbot yang menjawab dari dokumen perusahaan.	asisten tanya-jawab untuk karyawan yang membaca buku panduan perusahaan	AIE
Fullstack data	Sebutan untuk orang data yang memegang beberapa peran sekaligus, dari menyiapkan data sampai menganalisis dan menyajikannya.	satu orang mengambil data, merapikannya, lalu membuat dashboard sendiri	Semua
Data practitioner	Orang data yang paham konteks bisnis dan cara kerja sistem secara utuh, tidak hanya mahir satu alat.	tahu kapan cukup memakai alat ringan dan kapan butuh alat yang lebih besar	Semua
Tool-agnostic	Tidak terikat pada satu alat: konsepnya dikuasai, jadi mudah pindah alat sesuai kondisi perusahaan.	paham konsep data warehouse, jadi bisa bekerja di BigQuery maupun Redshift	Semua
Data maturity	Seberapa matang perusahaan mengelola data: dari data tercecer di banyak file sampai punya tim dan sistem yang rapi.	kantor yang analisisnya hanya memakai Excel, dibanding kantor yang punya tim DE sendiri	Semua

2 • Sumber data dan storage (1/4)

Istilah	Artinya	Contoh	Peran
Excel dan Google Sheets	Aplikasi spreadsheet untuk mengetik data, menghitung dengan rumus, dan meringkas tabel; banyak kantor masih memakainya sebagai alat utama.	rekap penjualan bulanan dengan SUMIF dan pivot table	DA, BI, BA
CSV	File tabel berupa teks biasa: satu baris satu catatan, dan nilai antarkolom dipisah koma.	file harga.csv bisa dibuka di Google Sheets, Excel, atau Notepad	Semua
Data source	Tempat asal data sebelum dipindahkan atau diolah: aplikasi, database, file, atau layanan lain.	aplikasi kasir, aplikasi keuangan, dan file Excel kiriman mitra	DE, DA
Data silo	Data yang tercecer di banyak tempat tanpa standar, sampai pemiliknya sendiri lupa menyimpannya di mana.	penjualan di aplikasi kasir, stok di Excel gudang, data promo di email	DE, DG
Open data	Data yang diterbitkan untuk umum dan boleh dipakai ulang, biasanya dengan syarat sumbernya disebut.	data harga pangan WFP berlisensi terbuka, boleh dipakai di portfolio asal sumbernya ditulis	Semua
Kaggle	Situs berisi banyak dataset gratis dan kompetisi data, sering dipakai untuk latihan dan bahan portfolio.	mengambil dataset penjualan ritel dari Kaggle untuk latihan pipeline	DA, DE, DS
Data dummy	Data karangan, bukan dari kejadian nyata. Cocok untuk mencoba alat, tapi kurang meyakinkan untuk portfolio.	angka penjualan karangan untuk mencoba membuat dashboard pertama	Semua

2 • Sumber data dan storage (2/4)

Istilah	Artinya	Contoh	Peran
API	Pintu resmi aplikasi untuk bertukar data dengan program lain secara otomatis.	program mengambil data kurs harian dari API setiap pagi	DE, AIE, DS
Database	Tempat menyimpan banyak tabel yang saling terkait, supaya bisa dicari dan digabung dengan cepat.	database toko online berisi tabel pelanggan, pesanan, dan produk	Semua
PostgreSQL, MySQL, SQL Server	Database yang banyak dipakai aplikasi untuk mencatat transaksi, sehingga sering jadi sumber data bagi analis dan DE.	data pesanan aplikasi toko online tersimpan di PostgreSQL	DE, DA, AE
Server	Komputer yang menyala terus untuk menyimpan data atau menjalankan aplikasi, dan diakses dari komputer lain.	database transaksi toko berjalan di server kantor pusat	DE
OLTP	Online transaction processing: database untuk mencatat transaksi satu per satu dengan cepat. Kuat mencatat, tapi lambat untuk rekap besar.	setiap pembayaran di kasir restoran menambah satu baris di database aplikasi kasir	DE, DA
OLAP	Online analytical processing: database yang dirancang untuk meringkas banyak data sekaligus, misalnya total per kota per bulan.	menghitung total transaksi semua cabang di satu kota sepanjang September	DE, DA, BI
Columnar	Cara database menyimpan data kolom demi kolom. Rekap jadi cepat karena database hanya membaca kolom yang dibutuhkan.	menjumlah kolom nilai transaksi tanpa membaca kolom nama dan alamat pelanggan	DE

2 • Sumber data dan storage (3/4)

Istilah	Artinya	Contoh	Peran
Data warehouse	Database besar khusus analisis, tempat data dari banyak sistem dikumpulkan dan dirapikan.	data kasir, aplikasi, dan stok gudang disatukan supaya bisa digabung dalam satu query	DE, AE, DA, BI
Lakehouse	Gabungan data lake (tempat menampung data mentah apa adanya) dan data warehouse: data mentah murah disimpan, tetap bisa diquery seperti tabel.	file mentah di object storage disajikan sebagai tabel yang bisa diquery dengan SQL	DE
Cloud	Server dan storage milik penyedia seperti Amazon (AWS), Google, atau Microsoft, yang disewa lewat internet.	BigQuery jalan di Google Cloud, jadi kantor tidak perlu membeli server sendiri	Semua
On-premise	Server dan software dipasang di gedung milik perusahaan sendiri, kebalikan dari cloud.	bank menyimpan database nasabah di ruang server kantornya sendiri	DE, DG
Cloud warehouse	Data warehouse sewaan dari penyedia cloud: tanpa membeli server, bayar sesuai pemakaian.	BigQuery, Snowflake, dan Redshift	DE, AE, DA
BigQuery	Cloud warehouse milik Google. Data diquery dengan SQL, dan biayanya mengikuti banyaknya data yang dibaca.	analisis menulis SQL di BigQuery untuk merekap penjualan semua cabang	DE, AE, DA, BI
Redshift	Cloud warehouse milik Amazon (AWS).	perusahaan yang servernya sudah di AWS sering memilih Redshift	DE, AE

2 • Sumber data dan storage (4/4)

Istilah	Artinya	Contoh	Peran
Snowflake	Cloud warehouse yang bisa berjalan di beberapa penyedia cloud. Biaya storage dan compute-nya dihitung terpisah.	tim data menyalakan compute Snowflake hanya saat report dibuat	DE, AE
ClickHouse	Database OLAP open source yang cepat untuk rekap data besar; bisa dipasang sendiri di laptop atau server.	memasang ClickHouse di laptop dengan Docker untuk latihan pipeline	DE
Azure	Layanan cloud milik Microsoft, dengan pilihan database, storage, dan AI sendiri.	kantor yang sudah memakai produk Microsoft menaruh data warehouse-nya di Azure	DE, MLE, AIE
Object storage (S3)	Storage cloud untuk file dalam jumlah sangat besar dan murah. Nama S3, layanan milik Amazon, sering dipakai untuk jenis ini.	file CSV transaksi harian ditaruh di S3 sebelum dimuat ke warehouse	DE
Hadoop	Teknologi lama untuk menyimpan dan mengolah data sangat besar di cluster, yaitu sekelompok server yang bekerja seperti satu mesin.	riwayat transaksi bertahun-tahun disimpan di cluster Hadoop milik bank	DE
Big data	Data yang terlalu besar, terlalu cepat datang, atau terlalu beragam untuk diolah satu komputer biasa.	catatan klik semua user aplikasi belanja setiap detik	DE, DS
Open source	Software yang kodenya terbuka dan boleh dipakai gratis; banyak alat data populer termasuk jenis ini.	PostgreSQL, ClickHouse, Airflow, dan dbt Core bisa dipasang gratis di laptop	Semua

3 • Pipeline dan infrastruktur (1/4)

Istilah	Artinya	Contoh	Peran
Pipeline	Rangkaian langkah otomatis yang memindahkan dan mengolah data dari sumbernya ke tempat tujuan, sesuai jadwal.	setiap dini hari, transaksi kemarin disalin dari aplikasi kasir ke data warehouse	DE, AE, MLE
Ingestion	Tahap memasukkan data dari sumber ke storage untuk analisis, entah sekali, terjadwal, atau terus-menerus.	menarik data pesanan dari PostgreSQL ke ClickHouse setiap pagi	DE
ETL	Extract, transform, load: data diambil, dirapikan di tengah jalan oleh mesin khusus, baru disimpan di tujuan.	format tanggal diseragamkan oleh mesin ETL sebelum data masuk warehouse	DE
ELT	Extract, load, transform: data mentah disimpan dulu di warehouse, baru dirapikan di sana. Banyak tim memilihnya karena lebih murah dan rapi.	semua tabel mentah masuk BigQuery dulu, lalu dirapikan dengan SQL di dbt	DE, AE
On the fly	Diolah sambil jalan, sebelum data sampai di tujuan. Pada ETL klasik, transformasi terjadi di memori mesin sehingga boros.	data dibersihkan di memori mesin ETL saat dipindahkan, belum disimpan di mana pun	DE
Tool ETL berbasis GUI	Software ETL yang dipakai dengan klik dan seret kotak di layar, biasanya butuh memori besar.	Informatica, SSIS, Talend, Pentaho	DE
Scheduling	Mengatur proses supaya jalan otomatis pada waktu tertentu, tanpa dijalankan manual.	pipeline penjualan dijadwalkan jalan setiap dini hari	DE, AE, MLE

3 • Pipeline dan infrastruktur (2/4)

Istilah	Artinya	Contoh	Peran
Batch	Data dipindahkan per kelompok pada waktu tertentu, misalnya sekali sehari.	transaksi sepanjang hari dipindahkan sekaligus setiap tengah malam	DE
Streaming	Data dipindahkan dan diolah terus-menerus begitu kejadiannya terjadi, hampir tanpa jeda.	posisi kurir di aplikasi ojek terus bergerak di peta	DE, MLE
Orchestration	Mengatur banyak langkah pipeline supaya jalan berurutan, sesuai jadwal, dan memberi tahu kalau ada yang gagal.	ambil data, rapikan, refresh dashboard; kalau satu langkah gagal, sisanya ditahan	DE, AE, MLE
Airflow	Tool orchestration open source yang sering disebut di lowongan DE; langkah pipeline-nya ditulis dengan Python.	Airflow menjalankan pipeline penjualan setiap pagi; versi siap pakainya ada di Google Cloud Composer	DE
Prefect	Tool orchestration berbasis Python yang lebih ringan dari Airflow, cocok untuk tim kecil.	tim kecil memakai Prefect karena Airflow terasa terlalu berat	DE
Airbyte	Ingestion tool open source: alat khusus untuk menyalin data dari sumber ke warehouse tanpa banyak menulis kode.	menyalin tabel dari PostgreSQL ke data warehouse tanpa skrip Python	DE
Big data processing	Mengolah data yang terlalu besar untuk satu komputer dengan membagi pekerjaannya ke banyak mesin.	merangkum catatan klik aplikasi bertahun-tahun di banyak server sekaligus	DE, DS

3 • Pipeline dan infrastruktur (3/4)

Istilah	Artinya	Contoh	Peran
Spark dan PySpark	Spark membagi pengolahan data besar ke banyak komputer. Kalau dipakai lewat Python, namanya PySpark.	menulis groupby di PySpark untuk data yang tidak muat di laptop	DE, DS, MLE
CLI dan Linux	CLI: mengoperasikan komputer dengan mengetik perintah teks. Linux: sistem operasi gratis yang banyak dipakai di server.	perintah ls, cd, dan cat untuk menjelajah folder di server	DE, MLE, AIE
Git	Alat pencatat riwayat perubahan kode, supaya versi lama bisa dilihat dan kerja tim tidak saling menimpa.	setiap perubahan query dicatat dengan pesan singkat dan bisa dikembalikan kalau salah	DE, AE, DS, MLE
Repo	Singkatan repository: satu folder proyek yang dicatat Git, lengkap dengan riwayat perubahannya.	satu repo berisi kode pipeline, model dbt, dan README	Semua
Docker	Alat untuk membungkus aplikasi beserta semua kebutuhannya, supaya jalan sama persis di komputer mana pun.	menyalakan ClickHouse dan Airflow di laptop tanpa instalasi rumit	DE, MLE, AIE
Development, testing, deployment	Tiga tahap membangun sistem: menulis kode, mengujinya, lalu memasangnya supaya dipakai sungguhan.	pipeline baru ditulis, diuji dengan data kemarin, lalu di-deploy ke server production	DE, MLE, AIE
Logging	Mencatat apa yang terjadi saat program berjalan, supaya saat error ketahuan langkah mana yang bermasalah.	log mencatat jam mulai, jumlah baris yang dipindah, dan pesan error	DE, MLE, AIE

3 • Pipeline dan infrastruktur (4/4)

Istilah	Artinya	Contoh	Peran
Monitoring	Memantau sistem terus-menerus: apakah pipeline jalan, tepat waktu, dan datanya lengkap.	memeriksa setiap pagi apakah semua tabel sudah terisi data kemarin	DE, MLE, AIE
Alert	Pesan otomatis yang dikirim saat ada yang salah, supaya tim segera tahu.	pesan masuk ke grup chat tim saat pipeline penjualan gagal jalan	DE, MLE
Bug	Kesalahan di kode yang membuat program berjalan tidak sesuai harapan.	pipeline gagal karena ada nama produk berisi tanda kutip yang tak terduga	DE, DA, MLE
Debugging	Mencari dan memperbaiki penyebab error atau bug.	membaca log untuk menemukan baris data mana yang membuat pipeline berhenti	DE, DA, MLE
Pipeline breaking	Pipeline berhenti atau rusak di tengah jalan, sering karena ada data yang bentuknya tak terduga.	kolom tanggal tiba-tiba berisi teks, pipeline berhenti, dashboard tidak terisi	DE
Biaya cloud	Tagihan pemakaian layanan cloud, yang bisa membengkak karena query boros atau server kebesaran.	query yang membaca seluruh tabel setiap jam membuat tagihan BigQuery naik	DE, AE
Overkill	Memakai solusi yang terlalu besar untuk masalah yang kecil, sehingga mahal dan rumit tanpa perlu.	memasang Airflow dan cluster Spark untuk mengolah satu file Excel kecil	Semua

4 • Data cleaning, SQL, data modeling (1/5)

Istilah	Artinya	Contoh	Peran
Data cleaning (cleansing)	Membereskan data sebelum dianalisis: tipe data, nilai kosong, duplikat, penulisan beda-beda, dengan setiap langkahnya dicatat.	membuang baris duplikat, lalu mencatat berapa baris yang hilang dan kenapa	DA, DE, AE, DS
Data kotor	Data yang berisi kesalahan: dobel, kosong, salah ketik, atau formatnya beda-beda. Kalau dipakai mentah, hasil analisis ikut salah.	nama kota ditulis 'Jkt', 'Jakarta', dan 'DKI Jakarta' di kolom yang sama	Semua
Data validation	Mengecek apakah data yang masuk sudah sesuai sumbernya dan masuk akal, sebelum dipakai.	mencocokkan total penjualan di warehouse dengan total di aplikasi kasir	DA, DE, AE
Data investigation	Menelusuri data mana yang relevan dengan masalah dan bagaimana kondisinya, sebelum mulai mengolah.	mencari tabel yang mencatat retur dan sejak kapan kolomnya terisi	DA
Transformasi data	Mengubah bentuk data supaya siap dipakai: membersihkan, menggabung, menghitung kolom baru, atau meringkas.	menghitung omzet dari harga kali jumlah, lalu meringkasnya per bulan	DA, DE, AE
Tipe data	Jenis isi kolom: angka, teks, tanggal, atau benar/salah. Kalau tipenya salah, hitungan ikut salah atau gagal.	gaji yang ditulis 'Rp5 juta' terbaca sebagai teks, jadi tidak bisa dijumlahkan	DA, DE, AE
Missing value	Nilai yang kosong atau tidak tercatat. Bisa tampil sebagai sel kosong, NA, tanda tanya, atau nol yang menyesatkan.	243.007 baris data Online Retail II tidak punya Customer ID	DA, DE, DS

4 • Data cleaning, SQL, data modeling (2/5)

Istilah	Artinya	Contoh	Peran
Duplikat	Catatan yang muncul lebih dari sekali. Kalau tidak dibuang, jumlah dan total ikut membengkak.	transaksi yang sama tercatat di dua file, sehingga totalnya terhitung ganda	DA, DE, AE
Standardisasi	Menyeragamkan cara penulisan nilai yang sebenarnya sama, supaya terhitung sebagai satu kelompok.	'Data Analyst', 'Analisis Data', dan 'DA' dijadikan satu judul baku	DA, DE
Pivot table	Ringkasan banyak baris jadi sedikit baris, dikelompokkan menurut kolom tertentu.	440 pelanggan grosir dirangkum jadi dua baris: 298 Horeca dan 142 Retail (data grosir Portugal)	DA, BI, BA
Data dictionary	Daftar penjelasan setiap kolom: namanya, artinya, satuannya, dan nilai yang mungkin.	Channel: 1 = Horeca (hotel, restoran, kafe), 2 = Retail	DA, DE, AE, DG
SQL	Bahasa untuk meminta dan mengolah data di database, dipakai bersama DA, DE, AE, dan DS.	SELECT tanggal, total FROM penjualan	Semua
Query	Perintah tertulis untuk meminta data tertentu dari database.	'tampilkan harga beras di Jawa Barat sepanjang tahun lalu' ditulis dalam SQL	DA, DE, AE, BI, DS
SELECT dan WHERE	SELECT menyebut kolom yang ingin ditampilkan. WHERE menyaring baris yang memenuhi syarat.	SELECT tanggal, harga FROM harga WHERE komoditas = 'Beras'	DA, DE, AE, BI

4 • Data cleaning, SQL, data modeling (3/5)

Istilah	Artinya	Contoh	Peran
GROUP BY	Mengelompokkan baris yang nilainya sama, lalu menghitung satu ringkasan per kelompok.	GROUP BY provinsi menghasilkan satu baris per provinsi	DA, DE, AE, BI
Agregasi	Meringkas banyak nilai jadi satu angka: jumlah, rata-rata, hitungan, minimum, maksimum.	SUM, AVG, COUNT, MIN, MAX	DA, DE, AE, BI
DISTINCT	Perintah SQL untuk menampilkan nilai yang berbeda saja, tanpa pengulangan.	SELECT DISTINCT kota FROM pelanggan menampilkan daftar kota tanpa dobel	DA, DE
CASE WHEN	Aturan 'kalau ini, maka itu' di dalam query, untuk membuat kolom baru dari kondisi.	CASE WHEN total >= 1000000 THEN 'besar' ELSE 'kecil' END	DA, DE, AE
JOIN	Menggabungkan dua tabel lewat kolom kunci yang sama di keduanya.	tabel pesanan digabung ke tabel produk lewat kode produk supaya nama produknya muncul	DA, DE, AE, BI
Subquery	Query di dalam tanda kurung yang hasilnya dipakai oleh query di luarnya.	mencari pasar yang rata-rata harganya di atas rata-rata semua pasar	DA, DE, AE
CTE	Common table expression: potongan query yang diberi nama dengan WITH, lalu dipakai seperti tabel di langkah berikutnya.	WITH bersih AS (...) SELECT ... FROM bersih	DA, DE, AE

4 • Data cleaning, SQL, data modeling (4/5)

Istilah	Artinya	Contoh	Peran
INSERT, UPDATE, CREATE TABLE	Perintah SQL untuk membuat tabel baru, menambah baris, dan mengubah isi baris yang sudah ada.	CREATE TABLE ringkasan_harian, lalu INSERT hasil rekap setiap pagi	DE, AE
Primary key	Kolom yang nilainya unik untuk setiap baris, jadi bisa dipakai menunjuk satu baris dengan pasti.	NIK di data kependudukan, kode produk di tabel produk	DA, DE, AE
Grain	Apa yang diwakili satu baris tabel; harus jelas sebelum menjumlah atau menggabung.	satu baris = satu harga satu komoditas di satu pasar pada satu bulan	DA, DE, AE, BI
Fan-out	Baris berlipat ganda sesudah JOIN, karena satu baris punya lebih dari satu pasangan. Totalnya ikut membengkak.	total penjualan mendadak naik setelah digabung ke tabel promo yang punya banyak baris per produk	DA, DE, AE
Relasi antartabel	Hubungan dua tabel lewat kolom kunci. Misalnya, satu pelanggan bisa punya banyak pesanan.	tabel pelanggan terhubung ke tabel pesanan lewat ID pelanggan	DA, DE, AE, BI
Skema database	Rancangan isi database: tabel apa saja, kolomnya apa, dan bagaimana tabel saling terhubung.	gambar kotak tabel pelanggan, pesanan, dan produk dengan garis penghubung	DE, DA, AE

4 • Data cleaning, SQL, data modeling (5/5)

Istilah	Artinya	Contoh	Peran
Normalisasi	Memecah satu tabel besar jadi beberapa tabel kecil, supaya keterangan yang sama tidak ditulis berulang dan tetap konsisten.	alamat toko disimpan sekali di tabel toko, tidak diulang di setiap baris transaksi	DE, DA, AE
Data modeling	Merancang bentuk tabel dan hubungannya supaya data mudah dianalisis dan angkanya konsisten.	memutuskan tabel penjualan dipisah dari tabel produk dan tabel toko	DE, AE, BI, DA
Fact table dan dimension table	Fact table berisi kejadian yang diukur; dimension table berisi keterangan untuk mengelompokkannya.	fakta: setiap transaksi dan nilainya; dimensi: produk, toko, tanggal	DE, AE, BI, DA
Star schema	Susunan tabel berbentuk bintang: satu fact table di tengah, dikelilingi dimension table.	fakta penjualan di tengah, terhubung ke dimensi produk, toko, pelanggan, dan tanggal	DE, AE, BI
Medallion architecture	Cara menyusun data di warehouse dalam tiga lapis. Bronze: data mentah. Silver: sudah dibersihkan. Gold: siap untuk dashboard.	bronze transaksi mentah, silver transaksi tanpa duplikat, gold penjualan per toko per hari	DE, AE, DA
dbt	Alat untuk menulis transformasi data dengan SQL di dalam warehouse, lengkap dengan dokumentasi dan tes otomatis.	query pembersihan disimpan sebagai model dbt dan dites setiap pagi	AE, DE

5 • Analisis dan critical thinking (1/5)

Istilah	Artinya	Contoh	Peran
Requirement gathering	Menggali kebutuhan user sebelum mengolah data: masalahnya apa, keputusan apa yang dibantu, metrik apa yang perlu.	berdiskusi dengan manajer pemasaran tentang angka apa yang dia perlukan dan untuk apa	DA, BA, BI, DE
Framing	Merumuskan masalah sebelum menghitung: keputusan apa yang dibantu, siapa yang memutuskan, apa yang perlu dijawab.	'bikin dashboard penjualan' diubah jadi 'toko mana yang perlu tambahan stok sebelum Lebaran?'	DA, BA, DS
Problem statement	Kalimat singkat tentang apa yang salah, seberapa besar, dan kapan harus beres.	response time keluhan memburuk kuartal ini dan harus pulih sebelum akhir tahun	DA, BA
Critical thinking	Kebiasaan mempertanyakan data dan kesimpulan: dari mana asalnya, masuk akal tidak, adakah penjelasan lain.	bertanya apakah penjualan turun karena promo berhenti atau karena memang bulan sepi	Semua
5W1H	Enam pertanyaan dasar untuk menggali masalah: what, who, when, where, why, how.	apa yang turun, di cabang mana, sejak kapan, siapa yang terdampak, kenapa, bagaimana	DA, BA
Root cause analysis	Menelusuri akar penyebab masalah, tidak berhenti di gejala yang terlihat.	harga naik, ternyata stok menipis, ternyata pengiriman tertahan di jalur laut	DA, BA, DE
Connecting the dots	Menghubungkan beberapa informasi yang terpisah sampai terlihat satu penjelasan utuh.	menyambungkan penjualan yang turun dengan biaya iklan yang dipangkas di bulan yang sama	DA, BA

5 • Analisis dan critical thinking (2/5)

Istilah	Artinya	Contoh	Peran
Hipotesis	Dugaan sementara yang diuji dengan data, dan hasilnya bisa terbukti atau gugur.	dugaan: penjualan turun karena iklan dipotong; lalu dicek di data pemasaran	DA, DS
Asumsi	Hal yang dianggap benar tanpa dicek dulu, karena itu perlu ditulis supaya bisa diuji.	menganggap semua cabang buka setiap hari, padahal ada yang tutup hari Minggu	DA, DS
Confirmation bias	Kecenderungan hanya mencari data yang mendukung dugaan sendiri dan mengabaikan data yang membantah.	yakin penjualan turun karena iklan, lalu hanya membuka data pemasaran	Semua
EDA	Exploratory data analysis: menjelajah data sebelum mulai menganalisis, misalnya cek jumlah baris, nilai kosong, dan sebaran.	cek jumlah baris, nilai kosong per kolom, sebaran harga, dan satu grafik tren	DA, DS
Metrik	Ukuran yang disepakati bersama, lengkap dengan definisi cara menghitungnya.	jumlah pesanan per hari, persentase pelanggan yang kembali dalam sebulan	DA, BI, AE, DS
KPI	Key performance indicator: metrik pilihan untuk menilai apakah target tercapai.	persentase pesanan yang terkirim tepat waktu	DA, BI, BA
Target	Angka yang ingin dicapai dalam periode tertentu, sekaligus pembanding untuk menilai angka aktual.	realisasi penjualan bulan ini dibandingkan target bulanannya	DA, BI

5 • Analisis dan critical thinking (3/5)

Istilah	Artinya	Contoh	Peran
Year on year (YoY)	Membandingkan angka satu periode dengan periode yang sama tahun lalu, supaya pengaruh musim tidak menipu.	penjualan Ramadan tahun ini dibandingkan Ramadan tahun lalu	DA, BI
Tren	Arah gerak angka dari waktu ke waktu: naik, turun, atau datar.	harga cabai naik pelan selama beberapa bulan berturut-turut	DA, BI, DS
Seasonality	Naik-turun yang berulang di waktu yang sama setiap periode, misalnya tiap akhir pekan atau tiap Lebaran.	penjualan melonjak setiap promo tanggal kembar di marketplace	DA, DS
Helicopter view	Melihat masalah dari atas secara menyeluruh sebelum masuk ke detail, supaya pola besarnya terlihat.	melihat tren penjualan setahun penuh sebelum membedah satu bulan yang turun	Semua
Distribusi	Gambaran sebaran nilai: di mana kebanyakan berkumpul, seberapa lebar, dan adakah nilai yang jauh menyendiri.	kebanyakan harga beras berkumpul di tengah, beberapa pasar jauh lebih mahal	DA, DS
Rata-rata	Jumlah semua nilai dibagi banyaknya nilai, sehingga mudah terdongkrak nilai ekstrem.	gaji direktur menaikkan rata-rata gaji satu divisi	DA, DS
Median	Nilai tengah setelah semua nilai diurutkan, jadi tidak ikut terdongkrak nilai ekstrem.	kemiskinan kab/kota 2019: rata-rata 11,97%, median 9,7% (INDO-DAPOER)	DA, DS

5 • Analisis dan critical thinking (4/5)

Istilah	Artinya	Contoh	Peran
Outlier	Nilai yang jauh berbeda dari yang lain, entah salah catat atau kejadian nyata.	melek huruf 113,9% di INDO-DAPOER: persen tak mungkin di atas 100, jadi cek dulu	DA, DS
Korelasi	Dua hal cenderung naik-turun bersama, tapi belum ketahuan mana yang menyebabkan.	penjualan payung dan jumlah karyawan terlambat sama-sama naik di musim hujan	DA, DS
Kausalitas	Hubungan sebab-akibat: yang satu benar-benar menyebabkan yang lain.	obat menurunkan demam	DA, DS
Confounder	Faktor ketiga yang tersembunyi dan menggerakkan dua hal sekaligus.	es krim dan kasus tenggelam sama-sama naik di musim panas	DA, DS
Sampel	Sebagian yang diamati untuk mewakili keseluruhan. Makin kecil sampelnya, makin mudah terlihat ekstrem.	harga beras medium Bengkulu Juni 2026 hanya dari 2 pasar (data WFP)	DA, DS
A/B test	Dua versi dibandingkan dengan membagi pengunjung secara acak, lalu dilihat versi mana yang hasilnya lebih baik.	setengah pengunjung melihat judul A, setengah lagi judul B	DA, DS
Funnel	Urutan tahap yang dilalui pelanggan, dari banyak orang di awal sampai sedikit yang selesai.	kunjungan, lihat produk, masuk keranjang, bayar	DA, BI

5 • Analisis dan critical thinking (5/5)

Istilah	Artinya	Contoh	Peran
Retention	Berapa banyak pelanggan yang masih kembali atau masih aktif sesudah waktu tertentu.	pelanggan yang belanja lagi sebulan setelah belanja pertamanya	DA, DS
Churn	Pelanggan yang berhenti, kebalikan dari retention.	pelanggan internet rumah yang tidak memperpanjang langganan bulan ini	DA, DS
Goodhart's law	Begitu satu angka dijadikan target, orang mengejar angkanya sampai angka itu tak lagi mengukur hal yang dimaksud.	tim CS dinilai dari tiket yang ditutup, lalu tiket ditutup walau masalahnya belum selesai	DA, BI, DG
Python	Bahasa pemrograman yang banyak dipakai untuk mengolah data, menulis skrip pipeline, dan membangun model.	beberapa baris Python menghitung ulang total yang dibuat di spreadsheet	DA, DE, DS, MLE, AIE
Notebook (Jupyter, Colab)	Dokumen kerja berisi kode, hasil, dan catatan berurutan, jadi orang lain bisa membaca dan mengulang langkahnya.	notebook EDA berisi sel kode, grafik, dan catatan pengecekan	DA, DS, MLE
Cross-check	Mencocokkan angka atau informasi dengan sumber atau cara hitung lain sebelum dipercaya.	mencocokkan angka dari AI dengan hitungan manual di spreadsheet	Semua

6 • Visualisasi dan dashboard (1/3)

Istilah	Artinya	Contoh	Peran
Visualisasi data	Menyajikan angka dalam bentuk grafik supaya pola dan perbandingan cepat terlihat.	grafik garis harga beras per bulan, grafik batang penjualan per cabang	DA, BI
Chart	Grafik batang, garis, pai, sebar, dan sejenisnya; jenisnya dipilih sesuai pesan yang ingin disampaikan.	grafik garis untuk tren waktu, grafik batang untuk membandingkan cabang	DA, BI
Dashboard	Satu halaman berisi beberapa grafik dan angka kunci yang di-update rutin, untuk memantau atau membantu satu keputusan.	dashboard harian stok gudang yang dibuka manajer gudang tiap pagi	DA, BI
BI tool	Software untuk membuat dashboard dan report dari data: cukup tarik kolom ke kanvas, grafiknya jadi.	Power BI, Tableau, Looker Studio, Metabase	DA, BI
Power BI	BI tool buatan Microsoft untuk dashboard dan report, yang sering diminta di lowongan DA dan BI.	dashboard penjualan bulanan dengan filter cabang dibuat di Power BI Desktop	DA, BI
Tableau	BI tool untuk membuat dashboard interaktif. Versi gratisnya, Tableau Public, menyimpan karya secara terbuka.	dashboard portfolio dari data publik dibagikan lewat link Tableau Public	DA, BI
Looker Studio	BI tool gratis dari Google, kini bernama Data Studio, yang mudah tersambung ke Google Sheets dan BigQuery. Versi berbayarnya Looker.	dashboard kampanye iklan yang datanya diambil langsung dari Google Sheets	DA, BI

6 • Visualisasi dan dashboard (2/3)

Istilah	Artinya	Contoh	Peran
Filter dan slicer	Kontrol di dashboard untuk mempersempit data yang ditampilkan. Di Power BI, filter berbentuk kotak pilihan disebut slicer.	pilihan tahun dan provinsi di bagian atas dashboard	DA, BI
Measure	Rumus ringkasan yang kita tulis sendiri di Power BI, misalnya total atau rata-rata, supaya cara meringkasnya terkendali.	measure Total Penjualan = SUM(penjualan[nilai])	DA, BI
DAX	Bahasa rumus di Power BI untuk menulis measure dan kolom hitungan.	measure penjualan tahun lalu ditulis dengan fungsi DAX SAMEPERIODLASTYEAR	DA, BI
Cockpit syndrome	Dashboard dijejali semua grafik dan angka tanpa urutan penting, seperti kokpit pesawat, sampai user bingung harus melihat ke mana.	satu halaman penuh grafik berukuran sama tanpa judul yang menjelaskan apa pun	DA, BI
Hierarki visual	Mengatur ukuran, posisi, dan warna supaya mata pembaca menangkap yang paling penting lebih dulu.	angka penjualan vs target dibuat besar di kiri atas, detail per produk di bawah	DA, BI
Chart junk	Hiasan di grafik yang tidak menambah informasi dan malah mengganggu pembacaan.	efek 3D, bayangan, dan gambar latar di grafik batang	DA, BI
Pembanding	Angka acuan supaya satu angka punya arti: target, periode lalu, atau rata-rata kelompok.	total sales saja belum bercerita; total sales vs target dan vs tahun lalu baru bercerita	DA, BI

6 • Visualisasi dan dashboard (3/3)

Istilah	Artinya	Contoh	Peran
Data storytelling	Menyusun temuan data jadi cerita yang runtut: apa yang terjadi, kenapa, dan apa yang perlu dilakukan.	dashboard dibuka dengan penjualan turun, lalu penyebabnya, lalu usulan tindakan	DA, BI, DS
Big idea	Pesan utama dalam satu kalimat: sudut pandang kita dan apa yang dipertaruhkan.	'Kita sebaiknya mengunci harga cabai dengan supplier, karena ...'	DA, BA, DS
Action title	Judul slide atau grafik yang langsung menyatakan kesimpulan.	'Titik panas 2015 memuncak di September: 64.080 titik' alih-alih 'Titik panas per bulan'	DA, BI, BA
Rekomendasi	Usulan tindakan yang lahir dari temuan data, lengkap dengan alasan dan risikonya.	tambah stok dua toko terlaris sebelum Lebaran karena tahun lalu keduanya kehabisan	DA, BA, DS
Report dan reporting	Laporan berisi angka, grafik, dan penjelasan yang dikirim rutin atau sekali jalan ke manajemen.	report bulanan pergerakan harga komoditas untuk tim pembelian	DA, BI
Dokumentasi	Catatan tertulis tentang cara kerja query, pipeline, atau analisis, supaya orang lain bisa memahami dan meneruskannya.	README yang menjelaskan isi tabel, sumber data, dan cara menjalankan pipeline	Semua

7 • Data science dan machine learning (1/3)

Istilah	Artinya	Contoh	Peran
Data science	Bidang yang memakai statistik, pemrograman, dan pengetahuan bisnis untuk menemukan pola dan membuat prediksi dari data.	memperkirakan permintaan produk bulan depan dari riwayat penjualan	DS
Machine learning	Cara membuat komputer menemukan pola dari data contoh, lalu memakai pola itu untuk menebak data baru.	belajar dari data pinjaman lama untuk menebak apakah peminjam baru akan lancar membayar	DS, MLE
Model	Rumus atau program hasil belajar dari data contoh, yang dipakai untuk menebak atau mengelompokkan data baru.	model yang menebak apakah satu transaksi kartu kredit mencurigakan	DS, MLE
Training	Proses model belajar pola dari data contoh.	model di-training dengan data pinjaman lama yang sudah diketahui lancar atau macet	DS, MLE
Fitur (feature)	Kolom keterangan yang dipakai model untuk menebak; pilihan fiturnya sangat menentukan hasil.	umur, lama berlangganan, dan jumlah keluhan untuk menebak pelanggan yang akan berhenti	DS, MLE
Label	Jawaban yang ingin ditebak model, dan sudah diketahui di data training.	kolom 'berhenti berlangganan: ya atau tidak' di data pelanggan lama	DS, MLE
Supervised learning	Model belajar dari contoh yang sudah ada jawabannya, lalu menebak jawaban untuk data baru.	belajar dari email yang sudah ditandai spam atau bukan	DS, MLE

7 • Data science dan machine learning (2/3)

Istilah	Artinya	Contoh	Peran
Unsupervised learning	Model mencari pola atau kelompok sendiri dari data, tanpa jawaban yang disiapkan.	mengelompokkan pelanggan berdasarkan kebiasaan belanja tanpa label apa pun	DS
Klasifikasi	Menebak kategori: ya atau tidak, atau salah satu dari beberapa kelas.	menebak apakah peminjam baru akan lancar atau macet membayar	DS
Regresi	Menebak angka, misalnya harga, jumlah, atau durasi.	memperkirakan harga rumah dari luas, lokasi, dan jumlah kamar	DS
Clustering	Mengelompokkan data yang mirip tanpa label, lalu memberi nama tiap kelompoknya.	segmen pelanggan: pemburu promo, pelanggan setia, pembeli musiman	DS, DA
Forecasting	Memperkirakan angka di masa depan dari pola data masa lalu.	perkiraan permintaan stok per gudang untuk bulan depan	DS, DA
Train-test split	Memisahkan data jadi dua: satu untuk training model, satu untuk mengujinya pada data yang belum pernah dilihat.	data sampai tahun lalu untuk training, data tahun ini untuk testing model	DS, MLE

7 • Data science dan machine learning (3/3)

Istilah	Artinya	Contoh	Peran
Overfitting	Model terlalu hafal data training sampai gagal saat bertemu data baru.	nilainya sempurna di data training, tapi tebakannya meleset di data bulan berikutnya	DS, MLE
Akurasi, precision, recall	Ukuran kualitas tebakan model. Akurasi: berapa yang benar. Precision: tebakan positif yang tepat. Recall: kasus positif yang tertangkap.	model penipuan yang tampak akurat tapi melewatkan banyak transaksi penipuan	DS, MLE
scikit-learn	Library Python yang populer untuk machine learning klasik: klasifikasi, regresi, dan clustering.	training model prediksi churn dalam beberapa baris kode	DS
Deep learning	Jenis machine learning dengan jaringan berlapis-lapis, kuat untuk gambar, suara, dan teks. Alatnya antara lain TensorFlow dan PyTorch.	model yang mengenali jenis penyakit daun dari foto	DS, MLE, AIE
Deployment model	Deploy model supaya bisa dipakai aplikasi sungguhan, tidak lagi hanya jalan di notebook.	model rekomendasi dipanggil aplikasi setiap kali user membuka halaman produk	MLE
Fraud detection	Model yang menandai transaksi atau perilaku mencurigakan supaya dicek lebih lanjut.	transaksi besar dari kota yang tak biasa langsung ditahan untuk diperiksa	DS, MLE

8 • AI generatif dan LLM (1/3)

Istilah	Artinya	Contoh	Peran
Artificial intelligence (AI)	Program komputer yang mengerjakan tugas yang biasanya butuh kecerdasan manusia, seperti membaca, menebak, atau menjawab.	pengenal wajah di ponsel, chatbot customer service	Semua
Generative AI (GenAI)	AI yang membuat teks, gambar, atau kode baru dari perintah kita, berbekal pola dari teks yang sangat banyak.	ChatGPT, Gemini, Claude	Semua
LLM	Large language model: model AI yang belajar dari teks sangat banyak, sehingga bisa menulis, merangkum, dan menjawab.	model di balik ChatGPT, Gemini, dan Claude	AIE, DS, MLE
Chatbot	Program yang diajak bercakap lewat teks, kini banyak yang memakai LLM.	chatbot customer service yang menjawab status pesanan	AIE
Prompt	Perintah atau pertanyaan yang ditulis untuk AI. Makin jelas konteks dan batasannya, makin berguna jawabannya.	'Buat query total penjualan per cabang bulan lalu dari tabel penjualan'	Semua
Prompt engineering	Merancang prompt dengan sengaja: memberi peran, konteks, contoh, dan format jawaban yang diinginkan.	menyertakan nama tabel dan kolom saat meminta AI membuat draft query	AIE, DA, DS
Context window	Informasi yang dibaca AI dalam satu percakapan disebut context; batas panjangnya, termasuk prompt dan jawaban, disebut context window.	menempelkan struktur tabel dan definisi metrik sebelum bertanya	Semua

8 • AI generatif dan LLM (2/3)

Istilah	Artinya	Contoh	Peran
Token	Potongan kecil teks yang dibaca dan ditulis LLM. Biaya dan batas panjang AI dihitung dalam token.	kalimat panjang dipecah jadi banyak token sebelum diproses model	AIE
Halusinasi	Jawaban AI yang terdengar meyakinkan tapi salah atau mengarang, termasuk angka dan sumber yang tidak ada.	AI menyebut fungsi SQL yang tidak ada di database kita	Semua
Non-deterministik	Sifat AI yang bisa memberi jawaban berbeda untuk pertanyaan yang sama. Hasilnya perlu dicek setiap kali.	meminta query yang sama berulang kali, hasilnya berbeda-beda dan ada yang keliru	Semua
AI sebagai akselerator	Memakai AI untuk mempercepat pekerjaan yang sudah kita pahami, lalu memeriksa hasilnya sendiri.	AI membuat draft query dan dokumentasi; kita mengecek logika dan angkanya	Semua
Human in the loop	Manusia tetap ikut memeriksa dan memutuskan dalam proses yang dibantu AI; manusia pula yang bertanggung jawab.	analisis membaca dan menguji query buatan AI sebelum hasilnya dilaporkan	Semua
Text-to-SQL	AI yang mengubah pertanyaan dalam bahasa sehari-hari jadi query SQL.	'berapa penjualan bulan lalu per cabang?' diubah AI jadi query	DA, AIE
RAG	Retrieval-augmented generation: AI mencari dulu potongan dokumen yang relevan, lalu menjawab berdasarkan potongan itu.	chatbot HR menjawab soal cuti dengan mengutip buku peraturan perusahaan	AIE

8 • AI generatif dan LLM (3/3)

Istilah	Artinya	Contoh	Peran
Embedding	Teks yang diubah jadi deretan angka, sehingga komputer bisa mengukur kemiripan makna dua teks.	'gaji' dan 'upah' punya embedding yang berdekatan	AIE, DS
Vector database	Database untuk menyimpan embedding dan mencari teks yang maknanya paling mirip dengan pertanyaan.	menyimpan potongan dokumen kebijakan untuk dicari oleh chatbot RAG	AIE
Agent	Program AI yang bisa merencanakan langkah dan memakai alat sendiri, misalnya menjalankan query atau membuka file.	agent yang menulis query, menjalankannya, lalu merangkum hasilnya	AIE
Fine-tuning	Training ulang model yang sudah jadi dengan contoh khusus, supaya gaya atau keahliannya sesuai kebutuhan.	training model dengan contoh jawaban customer service perusahaan sendiri	AIE, MLE
Evaluasi LLM	Menguji jawaban AI secara terukur dengan sekumpulan pertanyaan dan jawaban acuan, sebelum dan sesudah sistemnya diubah.	daftar pertanyaan uji yang jawabannya dicek setiap kali prompt diubah	AIE, MLE
Guardrail (AI)	Pagar pengaman yang membatasi apa yang boleh dijawab atau dilakukan AI.	chatbot menolak membuka data pribadi nasabah walau diminta	AIE, DG

9 • Data governance, privasi, MLOps (1/3)

Istilah	Artinya	Contoh	Peran
Data governance	Aturan tentang data: siapa pemiliknya, apa definisinya, siapa boleh melihat, dan cara menjaga kualitasnya.	hanya tim kredit yang boleh melihat gaji nasabah, dan definisi 'nasabah aktif' cuma ada satu	DG, DE, AE
Data quality	Seberapa layak data dipakai: lengkap, akurat, tidak dobel, konsisten, dan tepat waktu.	aturan: setiap transaksi wajib punya tanggal dan kode toko, dicek setiap hari	DG, DE, AE
Single source of truth	Satu sumber angka resmi yang dipakai semua tim, supaya tidak ada dua versi angka untuk hal yang sama.	tim keuangan dan operasional akhirnya memakai satu tabel omzet yang sama	DG, AE, DE, BI
SOP data	Prosedur baku tertulis untuk mengumpulkan, menyimpan, dan memakai data, supaya tidak tercecer.	setiap cabang wajib mengirim laporan stok dengan format yang sama	DG
Data steward	Orang yang merawat data sehari-hari: definisinya, kualitasnya, dan aturan pemakaiannya.	memastikan definisi 'pelanggan aktif' tertulis dan dipakai sama oleh semua tim	DG
Metadata	Data tentang data: nama kolom, artinya, pemiliknya, kapan terakhir di-update, dan dari mana asalnya.	keterangan bahwa tabel penjualan di-update setiap pagi oleh tim DE	DG, DE, AE

9 • Data governance, privasi, MLOps (2/3)

Istilah	Artinya	Contoh	Peran
Data catalog	Daftar terpusat semua data di perusahaan beserta metadatanya, supaya orang tahu data apa yang ada dan di mana.	mencari 'retur' di Collibra, DataHub, atau Purview, lalu menemukan tabel dan pemiliknya	DG, AE
Data lineage	Jejak asal-usul data: dari sumber mana, lewat langkah apa saja, sampai jadi angka di dashboard.	menelusuri angka omzet di dashboard balik ke tabel kasir aslinya	DG, AE, DE
Hak akses	Aturan siapa boleh melihat atau mengubah data apa, biasanya per peran.	hanya tim kredit yang boleh membuka kolom gaji nasabah	DG, DE
Data pribadi	Data tentang orang yang bisa dikenali, langsung atau lewat gabungan dengan data lain.	nama, nomor HP, alamat, foto KTP	Semua
PII	Personally identifiable information: data yang bisa menunjuk ke satu orang tertentu.	nama lengkap, NIK, nomor HP, email, alamat rumah	Semua
UU PDP	UU Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi, aturan tentang cara data pribadi dikumpulkan, dipakai, dan dijaga.	file daftar pelanggan berisi nomor HP tidak di-upload ke chatbot AI	Semua

9 • Data governance, privasi, MLOps (3/3)

Istilah	Artinya	Contoh	Peran
Anonimisasi	Data diolah sampai orangnya tidak bisa dikenali lagi, termasuk lewat gabungan dengan data lain.	laporan hanya berisi jumlah pesanan per kota, tanpa nama pembeli	DC, DE, DA
Masking	Menyamarkan sebagian isi data sensitif supaya data tetap bisa dipakai tanpa membuka identitas.	nomor HP hanya tampil beberapa digit awal, sisanya diganti huruf x	DC, DE
Kebocoran data	Data perusahaan atau data pribadi terbaca pihak yang tidak berhak, termasuk saat ditempel ke layanan AI publik.	menempel daftar nasabah ke chatbot gratis untuk minta dianalisiskan	Semua
MLOps	Cara kerja untuk merawat model machine learning di production: deploy, monitoring, dan training ulang secara teratur.	model churn otomatis di-training ulang tiap bulan dan dimonitor akurasi	MLE
Model monitoring	Mengecek terus apakah tebakan model tetap bagus setelah dipakai, karena kondisi nyata bisa berubah.	akurasi model penipuan dimonitor mingguan, dan alert dikirim kalau turun	MLE, DS
Data drift	Data baru yang masuk ke model mulai berbeda dari data training-nya, sehingga tebakannya memburuk.	pola belanja berubah saat pandemi, model prediksi permintaan jadi meleset	MLE, DS

10 • Karier dan melamar (1/5)

Istilah	Artinya	Contoh	Peran
Career switch	Pindah karier dari bidang lain ke bidang baru, misalnya dari pemasaran atau keuangan ke data.	staf administrasi yang belajar SQL lalu melamar posisi DA junior	Semua
Latar non-IT	Latar pendidikan atau kerja di luar teknologi informasi. Tetap bisa masuk ke data, dengan bekal pengetahuan bidang asalnya.	lulusan pertanian yang menganalisis pergerakan harga komoditas	Semua
Fresh graduate	Lulusan baru yang belum punya pengalaman kerja penuh waktu.	lowongan DA yang mencantumkan 'fresh graduate dipersilakan melamar'	Semua
Magang	Bekerja sementara untuk belajar di perusahaan, biasanya saat kuliah atau baru lulus, bisa berlanjut jadi kerja tetap.	magang di tim data, lalu ditawari posisi junior setelah lulus	Semua
Entry level dan junior	Jenjang awal karier. Pekerjaannya dibimbing, dan pengalaman panjang belum dituntut.	lowongan Junior Data Analyst yang menerima lulusan baru	Semua
Lowongan dan requirement	Info lowongan kerja beserta daftar syaratnya: pendidikan, pengalaman, alat, dan skill.	lowongan DA yang kini juga menyebut dasar ETL di daftar syaratnya	Semua

10 • Karier dan melamar (2/5)

Istilah	Artinya	Contoh	Peran
Hard skill dan soft skill	Kemampuan teknis seperti SQL disebut hard skill; komunikasi, bertanya, dan bekerja sama termasuk soft skill.	menulis query itu hard skill; menggali kebutuhan user itu soft skill	Semua
Beginner, intermediate, advanced	Tingkat penguasaan skill: dasar, menengah, dan lanjut.	SQL beginner: SELECT dan WHERE; intermediate: JOIN, CTE, dan subquery	Semua
Learning curve	Seberapa cepat seseorang menguasai hal baru. Konsep yang kuat membuat belajar alat berikutnya lebih cepat.	sudah paham BigQuery, jadi belajar Redshift terasa lebih cepat	Semua
Hands-on	Belajar dengan langsung mengerjakan, tidak hanya membaca atau menonton.	belajar Power BI sambil membuat dashboard dari data harga pangan sungguhan	Semua
Bootcamp	Kursus intensif beberapa minggu sampai bulan untuk menyiapkan peserta masuk kerja di bidang tertentu.	kursus intensif data science sebelum mulai melamar kerja	Semua
Portfolio	Kumpulan contoh kerja yang bisa dilihat recruiter, sebagai bukti kita sanggup mengerjakan tugas nyata.	dashboard dari data publik, repo pipeline, dan tulisan singkat berisi temuan	Semua

10 • Karier dan melamar (3/5)

Istilah	Artinya	Contoh	Peran
End-to-end	Mengerjakan sesuatu dari awal sampai akhir. Untuk portfolio DE: dari sumber data sampai tabel siap dipakai analis.	data dari API dimuat ke PostgreSQL, dipindah ke ClickHouse, lalu dirapikan dengan dbt	Semua
GitHub	Situs untuk menyimpan dan membagikan repo. Bagi orang data, sering jadi etalase portfolio.	link github.com/nama-anda/pipeline-harga ditulis di CV	Semua
LinkedIn	Jejaring sosial profesional untuk mencari kerja, membangun relasi, dan memamerkan portfolio.	menulis posting singkat tentang temuan dari dashboard portfolio	Semua
Networking	Membangun relasi dengan orang di bidang yang sama untuk saling belajar dan berbagi peluang.	berkenalan dengan praktisi data di acara komunitas, lalu tetap berkabar	Semua
Komunitas data	Kelompok orang yang belajar dan berbagi soal data, sering juga berbagi info lowongan.	grup diskusi yang rutin membahas studi kasus dan membagikan lowongan	Semua
Recruiter dan HR	Recruiter mencari dan menyaring kandidat. HR (human resources) mengurus rekrutmen dan kepegawaian.	recruiter mengirim DM di LinkedIn setelah melihat portfolio di profil kandidat	Semua

10 • Karier dan melamar (4/5)

Istilah	Artinya	Contoh	Peran
CV	Ringkasan pendidikan, pengalaman, dan skill untuk melamar kerja. Bagi orang data, sertakan link portfolio.	CV ringkas dengan link GitHub dan dashboard	Semua
ATS	Applicant tracking system: software untuk menerima dan menyaring lamaran. CV dibaca sistem ini dulu, baru recruiter.	portal lowongan tempat pelamar meng-upload CV dan mengisi formulir lamaran	Semua
Screening	Tahap penyaringan awal lamaran, dari CV atau telepon singkat, sebelum kandidat diundang tes atau interview.	recruiter menelepon sebentar untuk menanyakan ekspektasi gaji dan kapan bisa mulai	Semua
Take-home test	Tugas analisis yang dikerjakan di rumah dengan deadline, lalu dikirim ke perusahaan.	diberi satu file penjualan dan beberapa hari untuk menulis temuan dan rekomendasi	Semua
Live coding	Menulis query atau kode langsung di depan interviewer, biasanya tanpa bantuan AI.	diminta menulis query pelanggan dengan belanja terbesar dalam waktu singkat	Semua
Case interview	Interview berisi satu masalah bisnis singkat, dan kandidat diminta menjelaskan cara memecahkannya.	'Penjualan aplikasi turun bulan ini. Apa yang Anda cek pertama?'	Semua

10 • Karier dan melamar (5/5)

Istilah	Artinya	Contoh	Peran
STAR	Situation, task, action, result: urutan bercerita untuk pertanyaan behavioral dari pengalaman nyata.	situasi data berantakan, tugas merapikan, tindakan yang diambil, hasilnya	Semua
Freelance	Bekerja lepas per proyek untuk klien, tanpa ikatan sebagai karyawan tetap. Pesaing di platform global seperti Upwork datang dari banyak negara.	membuat dashboard penjualan untuk toko online selama beberapa minggu	Semua
Remote	Bekerja dari jarak jauh, tidak di kantor. Bisa untuk perusahaan dalam atau luar negeri.	analisis di Indonesia bekerja untuk perusahaan luar negeri lewat meeting online	Semua
Startup	Perusahaan rintisan berbasis teknologi; data biasanya jadi bahan keputusan sehari-hari.	aplikasi transportasi online yang punya tim DE, DA, dan DS terpisah	Semua
FMCG	Fast-moving consumer goods: produk sehari-hari yang cepat laku, seperti makanan, minuman, dan sabun.	menganalisis penjualan mi instan per toko per minggu	Semua
Konsultan data	Orang atau perusahaan yang membantu klien membereskan dan memakai datanya, biasanya per proyek.	membantu perusahaan yang datanya tercecer di banyak Excel menyusun data warehouse	Semua



Bagian 11

Penutup

Dari mana angka dan cerita di handout ini, dan apa batasnya.



Sumber dan batasan



Data dari snapshot lowongan satu hari, tanpa LinkedIn

538 lowongan dari 9 situs publik, 3 Oktober 2026. Bukan total lowongan Indonesia. n di bawah 20: baca sebagai arah.



Cerita praktisi adalah pengalaman pribadi

Narasumber dan host sesi DA (1 Okt) dan DE (2 Okt 2026).



Penilaian penyusun belum diuji data

Pohon jalur, label ramah pemula, sebagian roadmap.



Status alat dicek 3 Oktober 2026

Harga dan paket gratis bisa berubah. Cek halaman resmi.

Peta Dunia Data

3 Oktober 2026

Disusun untuk peserta **Ngobrol Karir Data & AI** 2026.

