

PROPOSAL TEKNIS VENDOR

Pengembangan Solusi Data Modelling & Reporting Terintegrasi Kluster Pelayanan dengan Teknologi Big Data

Modernisasi data mart Pelayanan menjadi arsitektur *lakehouse* medallion (Bronze–Silver–Gold) berbasis Hadoop, Spark, Apache Iceberg, dan Airflow/Talend — lengkap dengan CDC, column-level security, kualitas data berlapis, metadata & lineage, serta data mart tematik berbasis entitas peserta.

3+1

Bulan dev + maintenance

~42

Tabel sumber di-ingest

3 Layer

Model konseptual–logikal–fisikal

8+

Tenaga ahli inti

Ringkasan Eksekutif

BPJS Kesehatan mengelola salah satu basis data jaminan kesehatan terbesar di dunia — lebih dari 270 juta peserta dengan ratusan juta transaksi pelayanan, klaim, dan kunjungan setiap tahun. Kluster **Pelayanan** adalah jantung analitik organisasi, namun data mart eksisting menghadapi keterbatasan struktural pada skalabilitas, kualitas data, dan kecepatan pelaporan.

Proposal ini menjawab langsung kebutuhan dalam Kerangka Acuan Kerja (KAK/TOR) untuk *Pengadaan Jasa Pengembangan Solusi Data Modelling dan Reporting Terintegrasi Kluster Pelayanan dengan Teknologi Big Data*. Venturo Pro mengusulkan transformasi data mart Pelayanan menjadi arsitektur **lakehouse medallion** modern yang mendukung transaksi ACID, evolusi skema, pemisahan compute/storage, serta ingesti batch dan streaming — sepenuhnya berjalan di atas kluster Hadoop eksisting BPJS Kesehatan.

-65%

Waktu refresh data mart

Dari batch harian penuh ke incremental CDC + Spark, target latensi ≤ 30 menit (ilustratif).

≥95%

Skor kualitas data Gold

DQ gate otomatis di setiap layer berdasarkan 6 dimensi DMBOK.

~14 bln

Payback period

Estimasi ilustratif dari efisiensi compute, jam analisis, dan penurunan rework.

Tiga Pilar Solusi

- 1. Arsitektur Lakehouse Medallion.** Landing → Bronze → Silver → Gold di atas Apache Iceberg (ACID, schema evolution, time-travel) dengan format kolumnar Parquet/ORC, partisi berbasis timestamp, dan kompaksi file kecil.
- 2. Pipeline ELT Modern + CDC.** Extract-Load-Transform dengan data mentah langsung ke landing, CDC untuk minim dampak ke sistem sumber, orkestrasi Talend/Airflow, transformasi Spark, query via Trino/Hive, sajian Superset.

3. Governance, Quality & Security by Design. Kualitas data otomatis di setiap layer, metadata 3-tipe (teknis/bisnis/operasional) + lineage, proteksi data tingkat kolom (Voltage/CLS), serta data mart tematik berbasis entitas peserta.

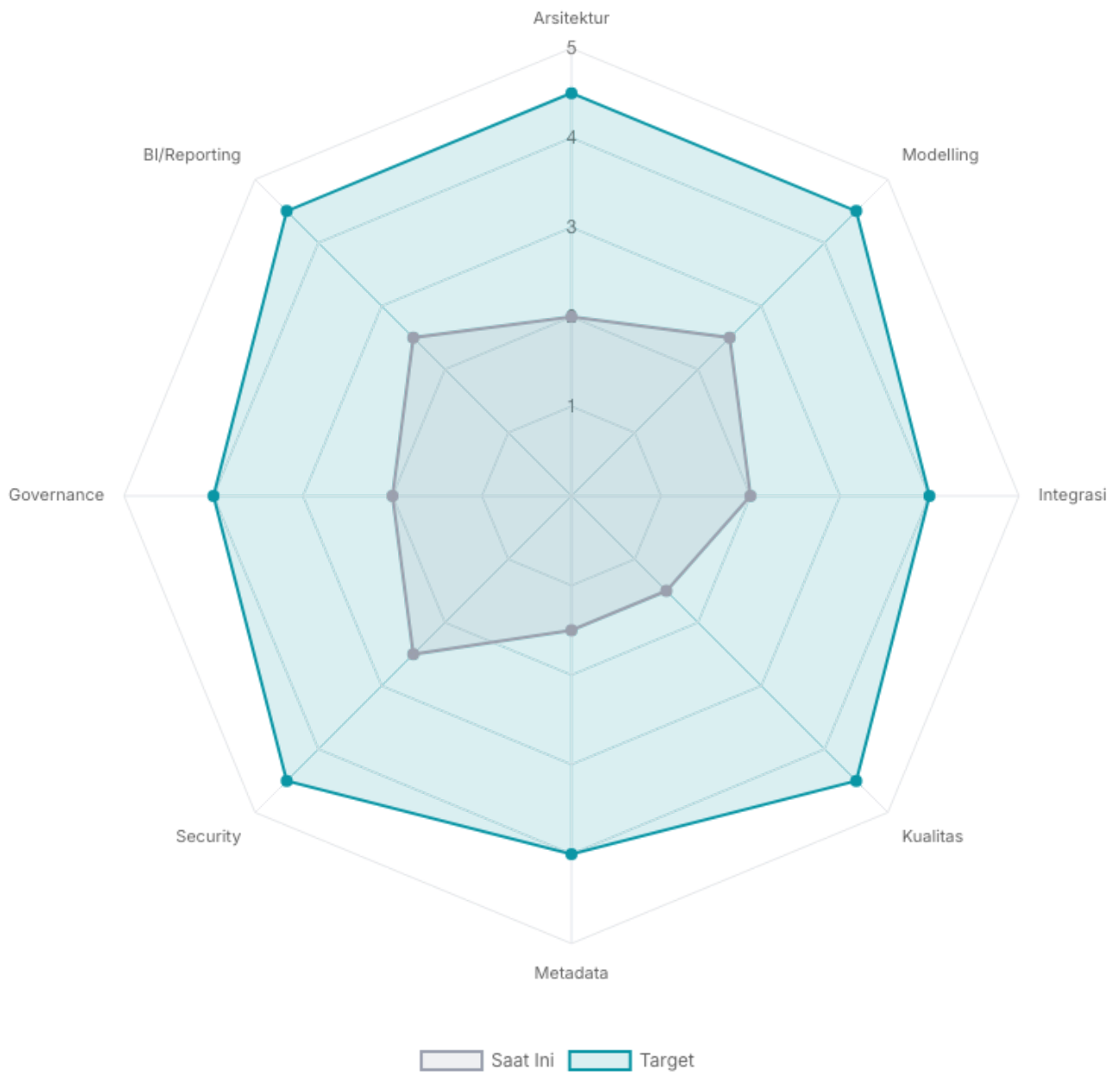
Rekomendasi. Venturo Pro merekomendasikan **GO** atas inisiatif ini. Studi kelayakan menunjukkan kelayakan teknis (memanfaatkan infrastruktur eksisting), kelayakan finansial (ROI positif < 2 tahun), serta kepatuhan penuh terhadap UU PDP No. 27/2022 dan praktik tata kelola DAMA-DMBOK / COBIT / ISO 27001.

2 Penilaian Strategis & Maturitas Data

Penilaian maturitas dilakukan terhadap delapan domain kapabilitas data, mengacu pada DAMA-DMBOK dan model maturitas data analytics. Skala 1 (Initial/Ad-hoc) hingga 5 (Optimized). Angka bersifat ilustratif berdasarkan pola umum organisasi pembayar kesehatan skala besar dan akan dikalibrasi pada fase Evaluasi Sistem Eksisting.

Domain Kapabilitas	Skor Saat Ini	Target	Kesenjangan & Catatan
Arsitektur Data	2.0 — Repeatable	4.5	Data warehouse/data lake terpisah; belum lakehouse ACID. Target: medallion + Iceberg.
Data Modelling & Design	2.5	4.5	Model fisik ada, dokumentasi konseptual/logikal parsial. Target: 3-layer terdokumentasi penuh.
Data Integration (ELT/ETL)	2.0	4.0	Batch penuh harian, dampak ke sumber. Target: ELT + CDC incremental.
Data Quality	1.5 — Initial	4.5	DQ reaktif/manual. Target: DQ gate otomatis 6 dimensi di setiap layer.
Metadata & Lineage	1.5	4.0	Lineage tidak terotomasi. Target: katalog metadata + lineage source-to-target.
Data Security & Privacy	2.5	4.5	Enkripsi at-rest ada; CLS/masking belum granular. Target: CLS/RLS/tokenization + KMS.
Data Governance	2.0	4.0	Kepemilikan data tidak formal. Target: RACI + stewardship + standar penamaan.
BI & Analytics / Reporting	2.5	4.5	Pelaporan lambat, banyak ekstrak manual. Target: semantic layer + self-service Superset.

Profil Maturitas: Saat Ini vs Target



Rata-rata maturitas saat ini $\approx 2.1 / 5.0$, target pasca-proyek $\approx 4.3 / 5.0$. Lonjakan terbesar pada Data Quality dan Metadata & Lineage — dua domain yang paling lemah dan paling berdampak pada keandalan pelaporan.

Pembandingan terhadap penyelenggara jaminan/layanan kesehatan kelas dunia. Tujuannya menempatkan ambisi arsitektur BPJS Kesehatan dalam konteks praktik terbaik global pada platform data analitik kesehatan.

Organisasi	Platform Data	Pendekatan Arsitektur	Praktik Unggulan yang Relevan
NHS England (UK)	Secure Data Environment / Federated Data Platform	Lakehouse + federated, Trusted Research Environment	Akses data terkontrol, lineage menyeluruh, privacy-by-design.
Kaiser Permanente (US)	EDW + cloud lakehouse	Integrated care data, clinical + claims unified	Single source of truth pasien lintas care + claims.
Singapore Health (SingHealth/MOH)	National Electronic Health Record + analytics	Hub-and-spoke, strong data governance	Standar interoperabilitas & tata kelola data nasional.
Medicare (Australia / Services Australia)	Enterprise data warehouse + analytics	Centralised claims analytics, fraud detection	Deteksi fraud/abuse berbasis analitik klaim skala nasional.
UnitedHealthcare / Optum (US)	Optum data & AI platform	Cloud lakehouse + advanced analytics/ML	Data mart tematik, real-time risk scoring, ML ops.
BPJS Kesehatan (target)	Hadoop lakehouse (Iceberg/Spark/Trino)	Medallion + data mart tematik peserta	On-prem ACID lakehouse, CDC, CLS, self-service reporting.

Posisi BPJS Kesehatan. Dengan menargetkan lakehouse medallion ber-ACID di atas Hadoop on-premise, BPJS Kesehatan menyetarakan fondasi arsitekturnya dengan peer global, sambil menjaga kedaulatan data nasional (data tetap on-premise) — keunggulan yang tidak dimiliki banyak peer berbasis cloud publik.

Inventarisasi permasalahan dikelompokkan ke dalam tujuh kategori sesuai ruang lingkup TOR: Arsitektur, Pipeline/Integrasi, Kualitas Data, Model Data, Metadata/Lineage, Keamanan, serta Pelaporan/Operasional. Severitas: **Kritis** **Tinggi** **Sedang** **Rendah**.

#	Permasalahan	Akar Penyebab	Dampak	Severitas	Rekomendasi
A. Arsitektur Data					
1	Data warehouse & data lake terpisah, tidak ada lapisan ACID	Arsitektur warehouse klasik tanpa table format transaksional	Inkonsistensi data, sulit update/delete, no time-travel	Kritis	Adopsi Apache Iceberg untuk ACID + schema evolution
2	Belum ada arsitektur berlapis (medallion) yang formal	Transformasi monolitik tanpa pemisahan zona	Sulit debug, reprocessing mahal, kualitas tak terlokalisir	Tinggi	Terapkan Landing/Bronze/Silver/Gold
3	Banyak file kecil (small files problem) di data lake	Ingesti tanpa kompaksi otomatis	NameNode pressure, query lambat di Hadoop	Tinggi	Kompaksi terjadwal + sizing partisi optimal
4	Strategi partisi tidak konsisten / absen	Tidak ada konvensi partisi berbasis timestamp	Full scan, biaya compute tinggi	Tinggi	Partisi by event/load date + hidden partitioning Iceberg
5	Compute & storage tidak terpisah	Coupling pada arsitektur lama	Skalabilitas terbatas, kontensi resource	Sedang	Pisahkan via Iceberg + Spark/Trino on YARN/K8s
6	Format file campuran (CSV/row-based)	Tidak ada standar format kolumnar	I/O boros, kompresi rendah	Sedang	Standarisasi Parquet/ORC kolumnar
B. Pipeline & Integrasi Data					
7	Full-load batch harian membebani sistem sumber	Tidak ada CDC/incremental	Beban OLTP tinggi, jendela batch melebar	Kritis	Implementasi CDC + incremental load
8	Paradigma ETL (transform sebelum load) kaku	Transformasi di tengah jalur	Re-ingest mahal saat logika berubah	Tinggi	Beralih ke ELT: raw ke landing, transform di lakehouse
9	Tidak ada penanganan streaming	Pipeline batch-only	Tak bisa near-real-time reporting	Sedang	Arsitektur hybrid batch + streaming (Kafka/Spark)

#	Permasalahan	Akar Penyebab	Dampak	Severitas	Rekomendasi
10	Error handling & retry minim	Tidak ada logging/alerting terstandar	Kegagalan senyap, kehilangan/korupsi data	Tinggi	Retry otomatis, dead-letter, alerting proaktif
11	Penjadwalan manual / cron tersebar	Tidak ada orkestrator terpusat	Dependensi tak terkelola, sulit audit	Tinggi	Orkestrasi Airflow/Talend dengan DAG
12	Konfigurasi koneksi DB hard-coded per job	Tidak ada context global	Sulit maintenance, risiko bocor kredensial	Sedang	Context global Talend + secrets manager
13	Tidak ada uji performa ingestion/transform	Belum ada baseline kapasitas	SLA pelaporan tak terjamin saat data tumbuh	Sedang	Performance test kapasitas saat ini + proyeksi
14	Pemrosesan tidak paralel / tidak teroptimasi	Job sekuensial, tanpa tuning Spark	Durasi siklus panjang	Sedang	Parallelisme + tuning partisi/shuffle Spark
C. Kualitas Data					
15	Tidak ada DQ check otomatis dalam pipeline	QC reaktif/manual pasca-load	Data buruk lolos ke pelaporan	Kritis	DQ gate otomatis tiap layer (DMBOK 6 dimensi)
16	Duplikasi data peserta/klaim	Tidak ada dedup & MDM	Over/under-count metrik	Tinggi	Dedup rule + golden record entitas peserta
17	Nilai null/invalid pada field kritikal	Validasi input lemah di hulu	Agregasi keliru, integritas rujukan rusak	Tinggi	Completeness & validity rules + quarantine
18	Inkonsistensi kode (diagnosa/faskes/wilayah)	Reference data tidak terstandar	Join gagal, laporan tak komparabel	Tinggi	Master/reference data + conformity check
19	Data tidak timely (telat refresh)	Jendela batch panjang	Keputusan berbasis data usang	Sedang	Incremental + SLA timeliness termonitor
20	Tidak ada penanganan late-arriving data	Asumsi data lengkap saat load	Restatement metrik historis	Sedang	Late-arriving handling + SCD pada dimensi
D. Model Data					
21	Model konseptual/logikal tidak terdokumentasi	Hanya ada skema fisik	Ketergantungan tribal knowledge	Tinggi	Model 3-layer konseptual-logikal-fisikal
22	Skema datamart sub-optimal (over/under normalized)	Desain tanpa pola dimensional jelas	Query lambat / redundansi	Tinggi	Star schema + denormalisasi strategis

#	Permasalahan	Akar Penyebab	Dampak	Severitas	Rekomendasi
23	Tidak ada konvensi penamaan konsisten	Penamaan ad-hoc lintas tim	Kebingungan, error join	Sedang	Naming convention + data dictionary
24	Belum ada data mart tematik per entitas peserta	Mart berorientasi proses, bukan entitas	Sulit analitik 360° peserta	Tinggi	Thematic mart berbasis entitas peserta
25	Model tidak future-ready (sulit tambah sumber)	Skema rigid	Rework mahal saat ekspansi	Sedang	Schema evolution Iceberg + desain ekstensibel
E. Metadata & Lineage					
26	Lineage source-to-target tidak ada	Tidak ada katalog/lineage tool	Impact analysis sulit, audit lemah	Tinggi	Lineage otomatis (Hive Metastore/OpenLineage)
27	Metadata bisnis (definisi elemen) absen	Tidak ada business glossary	Interpretasi metrik berbeda antar unit	Sedang	Business glossary + kamus data
28	Metadata operasional tidak tercatat	Tidak ada logging run/volume	Tak bisa optimasi waktu proses	Sedang	Catat durasi siklus & volume per unit waktu
F. Keamanan & Privasi					
29	Tidak ada proteksi tingkat kolom (CLS)	Kontrol akses hanya level tabel	Risiko paparan PII/data sensitif	Kritis	Column-level security + Voltage tokenization
30	PII tidak di-masking untuk pengguna analitik	Tidak ada dynamic masking	Non-compliance UU PDP	Tinggi	Dynamic masking + role-based RLS
31	Manajemen kredensial lemah	Akses tersebar, tak terpusat	Risiko kebocoran kredensial	Tinggi	KMS + secrets vault + akses jam kerja
32	Audit trail akses data tidak lengkap	Logging akses minim	Sulit forensik & akuntabilitas	Sedang	Audit logging + Ranger policy
G. Pelaporan & Operasional					
33	Pelaporan lambat & banyak ekstrak manual (Excel)	Tidak ada semantic layer/self-service	Beban tim data, inkonsistensi angka	Tinggi	Semantic layer + Superset self-service
34	Tidak ada dashboard monitoring pipeline	Observability minim	MTTR tinggi saat insiden	Tinggi	Dashboard monitoring + alerting end-to-end
35	Tidak ada metrik latensi/throughput/error rate	Tak ada observability metrics	Sulit jaga SLA	Sedang	Metrik latency/traffic/error termonitor

#	Permasalahan	Akar Penyebab	Dampak	Severitas	Rekomendasi
36	Query analitik bersaing dengan beban operasional	Tidak ada isolasi workload	Kinerja query tak stabil	Sedang	Workload isolation via Trino + resource group
37	Tidak ada validasi integritas referensial antar-stage	Cek RI tak otomatis	Orphan records di mart	Sedang	RI check tiap tahap ELT/ELT
38	Dokumentasi teknis tidak terpelihara	Tidak ada standar dokumentasi	Keberlanjutan terancam	Sedang	Dokumentasi model/pipeline/waktu proses lengkap
39	Ketergantungan vendor/individu (bus factor)	Knowledge tidak ditransfer	Risiko operasional tinggi	Tinggi	Knowledge transfer terstruktur + materi referensi
40	Tidak ada baseline biaya compute/storage	Tak ada cost observability	Inefisiensi anggaran TI	Rendah	Cost monitoring + optimasi partisi/kompaksi

Daftar di atas merupakan himpunan representatif (40 isu inti); inventarisasi lengkap akan difinalisasi pada fase Evaluasi Sistem Eksisting (bulan ke-1) dan dilaporkan dalam Dokumen Evaluasi Sistem Data Eksisting.

5.1 Kelayakan Bisnis (CBA, ROI, Payback)

Analisis biaya-manfaat berikut bersifat **ilustratif** (angka indikatif, bukan penawaran final) untuk menunjukkan logika kelayakan finansial. Nilai aktual ditetapkan dalam RAB resmi.

Komponen	Tahun 0 (Investasi)	Manfaat Tahunan (ilustratif)	Keterangan
Pengembangan solusi (jasa 3+1 bln)	Rp X (per RAB)	—	Biaya langsung personil + non-personil
Efisiensi compute & storage	—	+ Rp 0,9 M	Incremental load, kompaksi, kolumnar
Penghematan jam analis (less manual ekstrak)	—	+ Rp 1,4 M	Self-service + semantic layer
Penurunan rework akibat data buruk	—	+ Rp 0,8 M	DQ gate otomatis
Percepatan pengambilan keputusan	—	+ Rp 1,1 M	Reporting lebih cepat & akurat
Total manfaat tahunan	—	+ Rp 4,2 M	Ilustratif

~138%

ROI 3 tahun (ilustratif)

~14 bln

Payback period

Positif

NPV (disc. 10%)

5.2 Kelayakan Teknis

Komponen	Kesiapan	Catatan
Kluster Hadoop (HDFS/YARN)	Siap	Tersedia; jadi fondasi storage & compute on-premise
Apache Spark	Siap	Engine transformasi (per TOR)
Apache Iceberg	Perlu setup	Table format ACID di atas HDFS — diimplementasi proyek
Airflow / Talend	Perlu konfigurasi	Orkestrasi & CDC/integrasi (per TOR)
Hive Metastore / Trino	Perlu setup	Katalog & query engine SQL
Superset	Perlu setup	Visualisasi & self-service BI
Voltage (data protection)	Perlu integrasi	Tokenization/proteksi kolom (per TOR)
Kafka (streaming)	Opsional	Untuk kebutuhan streaming/CDC stream

Verdict teknis: LAYAK. Mayoritas fondasi (Hadoop, Spark) telah tersedia; solusi dirancang menyesuaikan

kapabilitas infrastruktur eksisting BPJS Kesehatan, meminimalkan capex baru.

5.3 Kelayakan Operasional

Pekerjaan dilaksanakan **on-site** di Kantor Pusat BPJS Kesehatan untuk akses data, dengan akses kredensial pada hari/jam kerja sesuai ketentuan operasional; pekerjaan non-data dapat dilakukan online. Fase pemeliharaan 1 bulan + knowledge transfer memastikan tim internal mampu mengoperasikan, memelihara, dan menangani gangguan pipeline secara mandiri.

5.4 Kelayakan Keamanan

Solusi menerapkan defense-in-depth: enkripsi at-rest & in-transit, column-level security, dynamic masking, tokenization (Voltage), RLS berbasis peran, KMS untuk pengelolaan kunci, dan audit logging penuh — selaras ISO 27001.

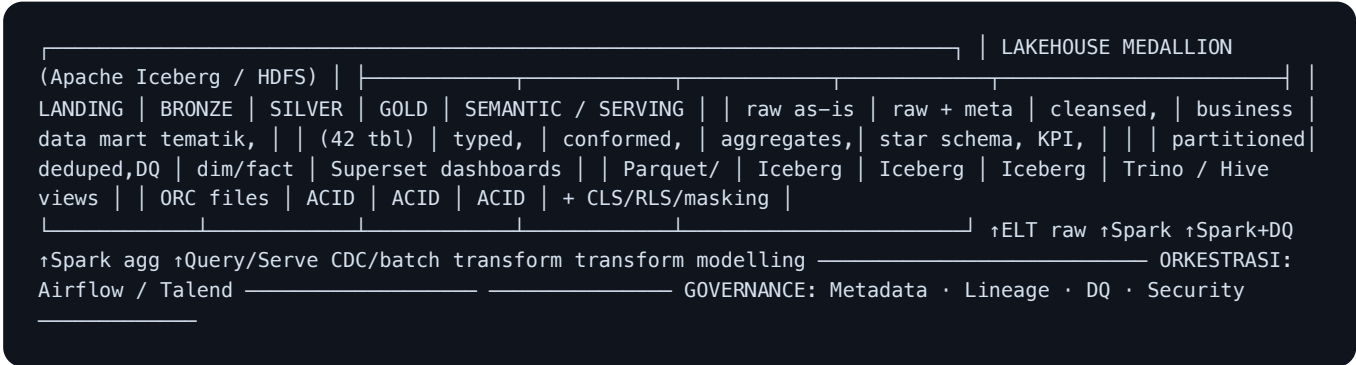
5.5 Kelayakan Regulasi — UU PDP No. 27/2022

Prinsip UU PDP	Penerapan dalam Solusi
Pembatasan tujuan & minimisasi data	Hanya kolom relevan masuk Gold mart; PII di-tokenize
Keamanan pemrosesan	CLS/RLS/masking/enkripsi + audit trail
Akuntabilitas	Lineage + metadata + RACI governance
Hak subjek data	Mampu telusur & kelola data peserta via lineage
Kedaulatan data	Seluruh data tetap on-premise di pusat BPJS Kesehatan

6

Arsitektur Target

Arsitektur target menggabungkan prinsip **Lakehouse** (unifikasi data lake + warehouse), **Medallion** (zona berlapis kualitas progresif), elemen **Data Mesh** (kepemilikan domain via data mart tematik), dan **Data Fabric** (metadata aktif + lineage). Seluruhnya berjalan di atas Hadoop on-premise.



Definisi Zona Medallion

Zona	Tujuan	Transformasi	Format / Tabel
Landing	Penampungan data mentah as-is dari sumber, tanpa transformasi	Tidak ada (raw)	Parquet/ORC files di HDFS
Bronze	Data mentah ter-registrasi: typed, di-partisi, ditambah metadata teknis (load ts, source)	Type casting, partisi, audit cols	Iceberg (ACID)
Silver	Data bersih & terkonform: dedup, validasi, standarisasi kode, integritas referensial	Cleansing, DQ gate, conform, SCD	Iceberg (ACID)
Gold	Data siap bisnis: fact & dimension, agregasi, metrik	Dimensional modelling, aggregation	Iceberg star schema
Semantic / Serving	Data mart tematik per entitas peserta, view semantik, dashboard	Business views, KPI, security policy	Trino/Hive views + Superset

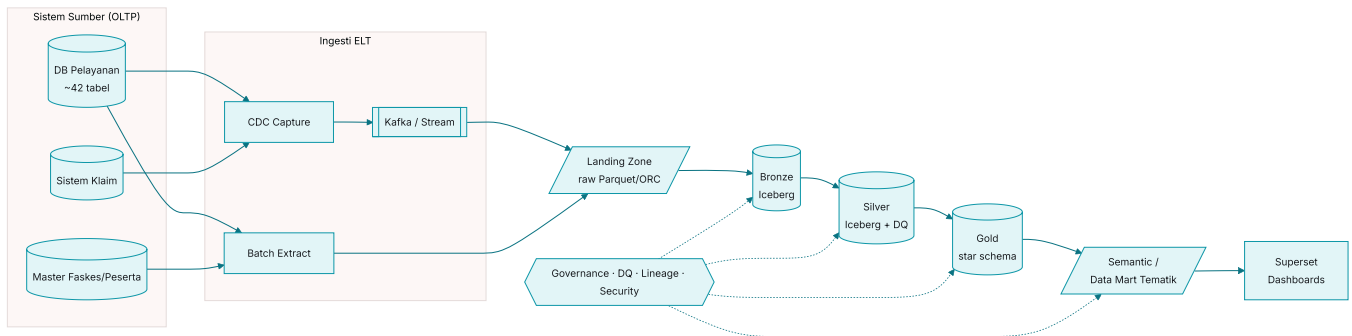
Tumpukan Teknologi

Storage: HDFS (Hadoop)	Table format: Apache Iceberg
Compute: Apache Spark	Orkestrasi: Airflow / Talend
CDC / Integrasi: Talend + CDC	Streaming: Kafka (opsional)
Katalog: Hive Metastore	Query engine: Trino / Hive
BI / Dashboard: Apache Superset	Format file: Parquet / ORC

Proteksi data: Voltage (CLS/token)

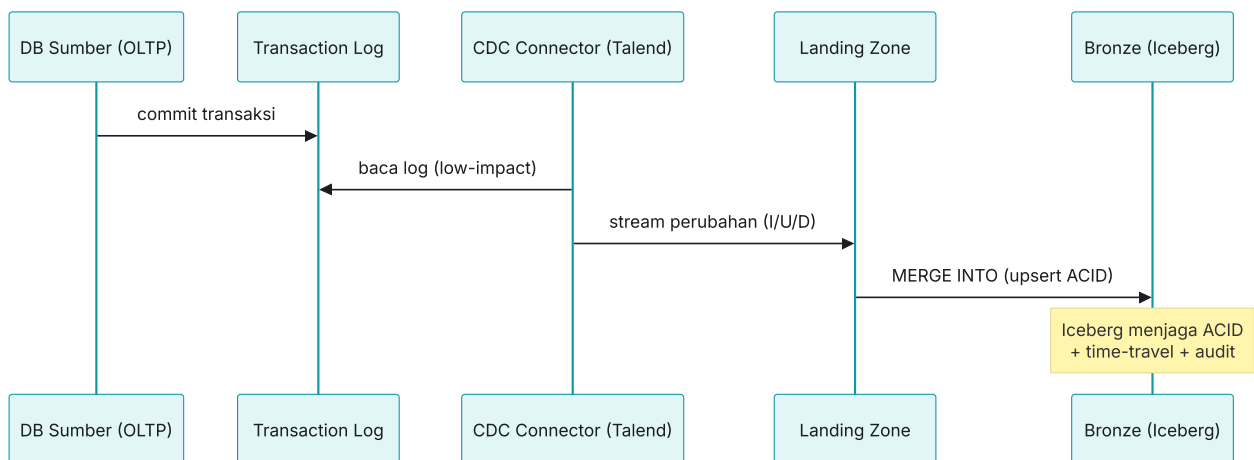
Security policy: Ranger + KMS

Diagram alur end-to-end dari sistem sumber hingga dashboard, dengan jalur CDC untuk perubahan inkremental dan jalur batch untuk muatan periodik.



Mekanisme Change Data Capture (CDC)

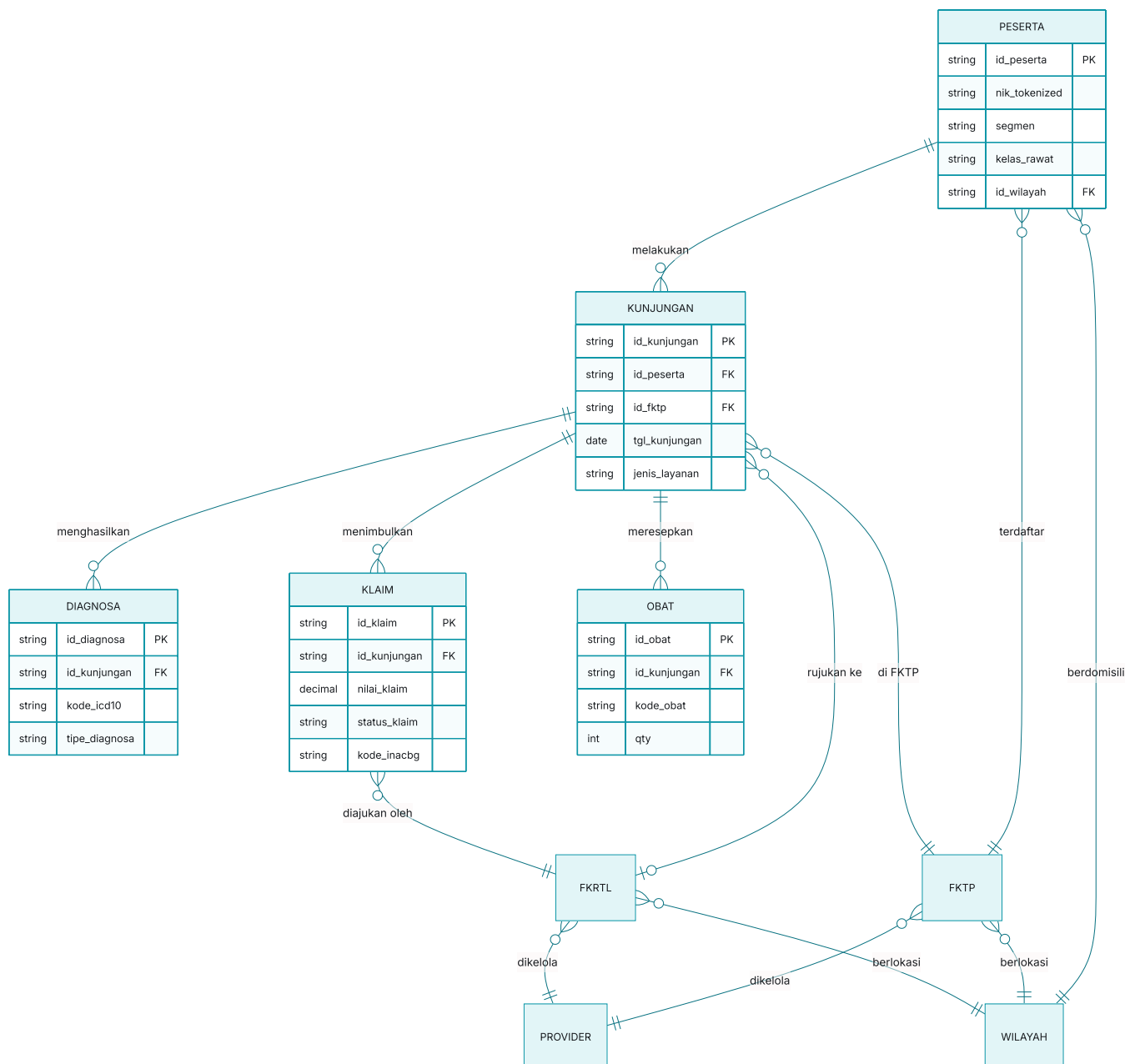
CDC menangkap hanya perubahan (insert/update/delete) dari sistem sumber, sehingga ekstraksi tidak membebani OLTP secara signifikan — memenuhi syarat TOR "tanpa dampak signifikan pada kinerja sistem sumber".



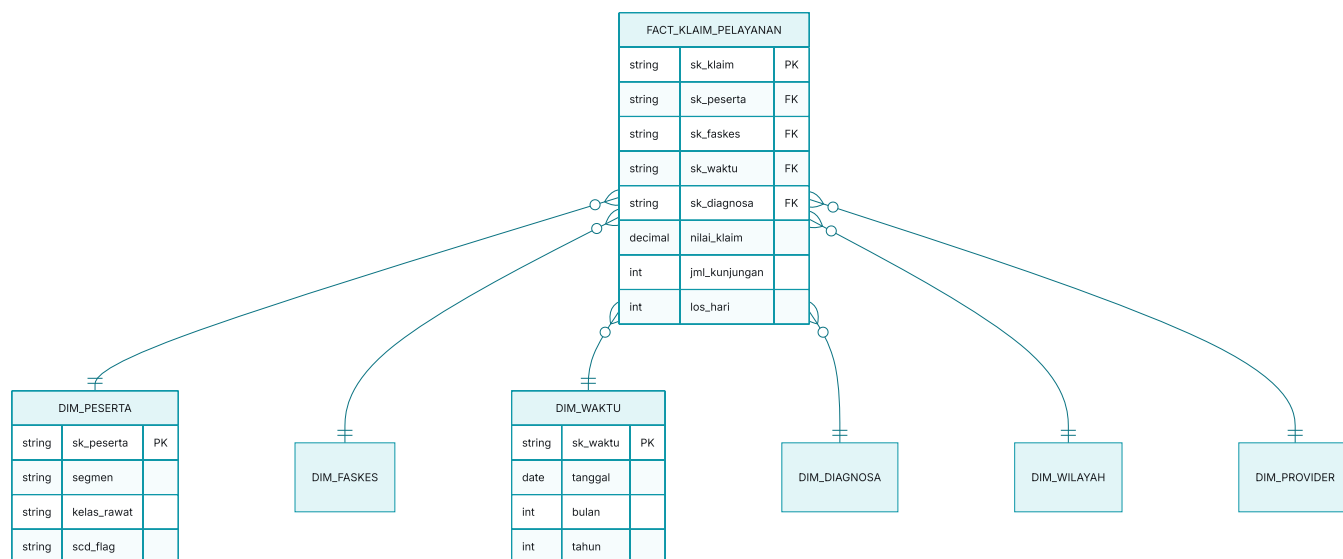
Sesuai TOR, model data disusun dalam tiga lapisan: **konseptual** (entitas bisnis & relasi), **logikal** (atribut, kunci, normalisasi), dan **fisikal** (tipe data, partisi, indeks pada Iceberg).

8.1 Model Konseptual

Entitas inti kluster Pelayanan: **Peserta, FKTP, FKRTL, Kunjungan, Diagnosa, Klaim, Obat, Provider, Wilayah.**



8.2 Model Logikal — Star Schema (Gold)



8.3 Model Fisikal

Aspek	Keputusan Desain
Table format	Apache Iceberg (ACID, schema evolution, hidden partitioning)
Format file	Parquet (default) / ORC, kompresi Snappy/ZSTD
Partisi	Berdasarkan timestamp (tgl_kunjungan / load_date), bucketing pada high-cardinality keys
SCD	SCD Type 2 pada DIM_PESERTA, DIM_FASKES untuk historisasi
Kompaksi	Rewrite/compaction terjadwal untuk small-file management
Naming convention	Prefix layer + domain + entitas (mis. slv_pelayanan_kunjungan)

8.4 Data Mart Tematik Berbasis Entitas Peserta

Sesuai TOR, dikembangkan data mart tematik yang menjadikan **peserta** sebagai pusat (peserta-360°): menyatukan profil, riwayat kunjungan, diagnosa, klaim, dan utilisasi obat untuk analitik dan pelaporan yang konsisten dan lakehouse-ready.

Kerangka tata kelola mengacu pada **DAMA-DMBOK** (knowledge areas), **COBIT 2019** (governance & management objectives TI), dan **ISO 27001** (keamanan informasi).

Matriks RACI

Aktivitas	BPJS Data Owner	Data Steward	Venturo PM	Data Engineer	QC
Definisi kebijakan data	A	R	C	I	I
Desain arsitektur & model	A	C	R	R	C
Pengembangan pipeline ELT	I	C	A	R	C
Aturan kualitas data	C	A	C	R	R
Klasifikasi & keamanan kolom	A	R	C	R	I
Validasi & UAT	A	C	R	C	R
Knowledge transfer	I	C	A	R	C

R = Responsible, A = Accountable, C = Consulted, I = Informed

10 Kualitas Data (DMBOK)

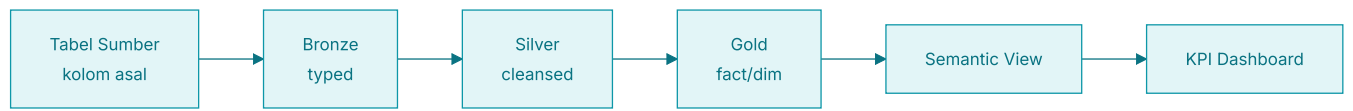
Pemeriksaan kualitas data otomatis terintegrasi dalam pipeline pada **setiap layer** (Landing→Gold), berdasarkan enam dimensi kualitas DMBOK. Data yang gagal divalidasi dikarantina (quarantine) dan dilaporkan.

Dimensi	Definisi	Contoh Aturan	Layer
Completeness	Kelengkapan nilai wajib	id_peserta IS NOT NULL; rasio null < 1%	Bronze→Silver
Accuracy	Kebenaran terhadap referensi	kode ICD-10 valid; nilai_klaim ≥ 0	Silver
Consistency	Keselaran lintas tabel/waktu	status_klaim konsisten dgn kunjungan	Silver→Gold
Uniqueness	Tanpa duplikasi	PK unik; dedup peserta/klaim	Silver
Validity	Sesuai format/domain	tgl ≤ today; format NIK; range cek	Bronze→Silver
Timeliness	Kesegaran data	SLA refresh ≤ 30 menit; freshness check	End-to-end

DQ Gate. Setiap transformasi Spark menyertakan langkah validasi: lolos → promote ke layer berikutnya; gagal → quarantine + alert. Metrik DQ (pass rate per dimensi) tersaji di dashboard monitoring.

11 Metadata & Lineage

Tiga tipe metadata dikelola sesuai TOR: **teknis** (struktur & tipe data), **bisnis** (definisi elemen data), dan **operasional** (data lineage, durasi siklus, volume per unit waktu).



Lineage source-to-target mendukung *impact analysis* (apa yang terpengaruh bila kolom sumber berubah) dan audit kepatuhan. Katalog metadata terintegrasi dengan Hive Metastore dan didokumentasikan dalam pemetaan Sumber-ke-Target.

Tipe Metadata	Isi	Sumber/Tool
Teknis	Struktur tabel, tipe data, partisi, skema	Hive Metastore / Iceberg catalog
Bisnis	Definisi elemen, business glossary, KPI	Data dictionary + glossary
Operasional	Lineage, run duration, volume/unit waktu, freshness	Airflow logs + OpenLineage

Proteksi data tingkat kolom (sesuai TOR) sebagai inti, didukung lapisan keamanan defense-in-depth.

Kontrol	Mekanisme	Penerapan
Column-Level Security (CLS)	Policy per kolom sensitif	Ranger + Voltage; PII hanya untuk peran berwenang
Row-Level Security (RLS)	Filter baris berbasis peran/wilayah	Trino/Hive policy
Dynamic Masking	Mask saat query oleh peran non-privileged	NIK → xxxx-xxxx
Tokenization	Substitusi nilai sensitif dgn token	Voltage (per TOR)
Encryption	At-rest (HDFS) & in-transit (TLS)	HDFS encryption zones
Key Management (KMS)	Rotasi & kustodi kunci terpusat	Hadoop KMS / KMS enterprise
Access Control	RBAC + akses jam kerja	Kerberos + Ranger; kredensial on hari/jam kerja
Audit Logging	Catat seluruh akses data	Ranger audit + SIEM

13

Monitoring & Observability

Dashboard pemantauan pipeline end-to-end dan sistem peringatan proaktif (per TOR). Memantau kinerja pipeline, mengidentifikasi masalah, dan memastikan alur data benar.

Metrik	Deskripsi	Ambang Alert (ilustratif)
Latency	Durasi siklus pipeline end-to-end	> 45 menit
Throughput / Traffic	Volume data per unit waktu	Deviasi > 30% dari baseline
Error / Failure rate	Frekuensi kegagalan job	> 2% job gagal
DQ pass rate	Persentase lolos validasi per dimensi	< 95%
Freshness	Selisih waktu data terbaru vs sekarang	> SLA refresh
Resource utilization	CPU/memory/HDFS	> 85%

Alerting proaktif. Notifikasi otomatis (email/chat) saat ambang terlampaui, dengan runbook troubleshooting untuk mempercepat MTTR selama fase pemeliharaan.

Lebih dari 25 use case analitik dikelompokkan ke dalam lima tema. Setiap use case memetakan Objektif, KPI, Sumber, dan Visual.

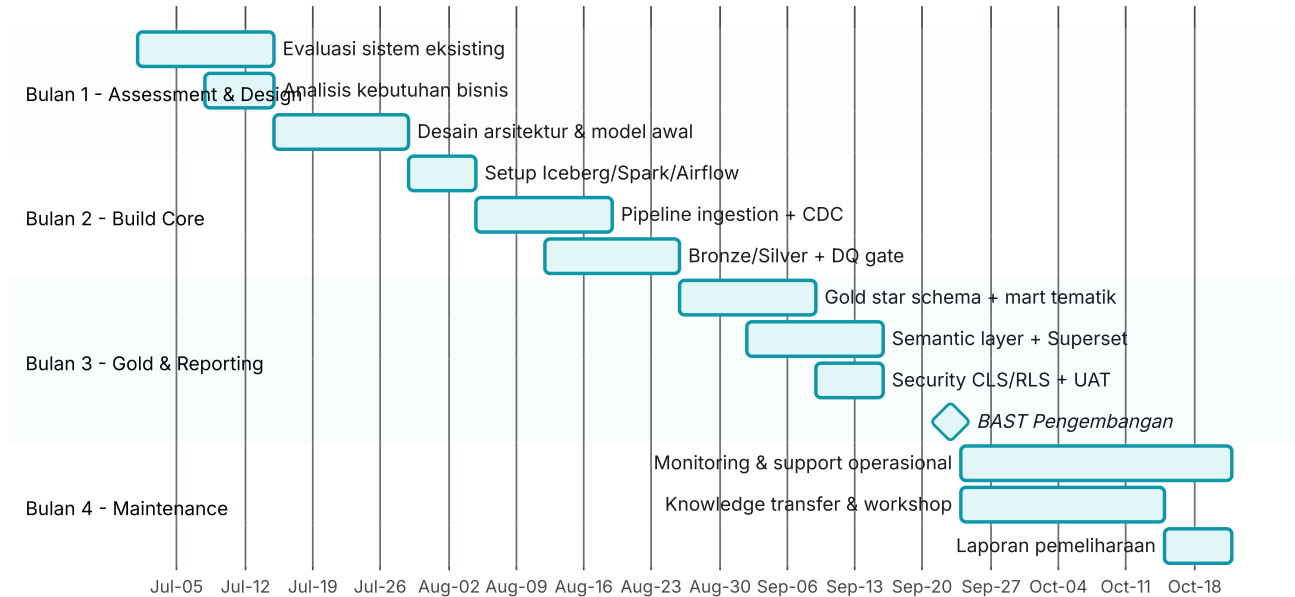
#	Use Case	Objektif	KPI Utama	Sumber (Gold)	Visual
A. Utilisasi Pelayanan					
1	Volume kunjungan FKTP	Pantau beban FKTP	Jml kunjungan / hari	fact_kunjungan	Time series
2	Rasio rujukan FKTP→FKRTL	Efektivitas gatekeeping	% rujukan	fact_kunjungan	Gauge + trend
3	Length of Stay (LOS) rawat inap	Efisiensi RS	Rerata LOS	fact_klaim	Box/bar
4	Tingkat readmisi	Mutu layanan	% readmisi 30 hari	fact_kunjungan	KPI card
5	Utilisasi per kelas rawat	Profil pemanfaatan	Kunjungan/kelas	dim_peserta	Stacked bar
B. Klaim & Keuangan					
6	Total nilai klaim	Kontrol biaya	Rp klaim	fact_klaim	KPI + trend
7	Rerata biaya per kasus (INA-CBG)	Benchmark tarif	Rp/kasus	fact_klaim	Bar by CBG
8	Status & aging klaim	Percepat verifikasi	% pending/aging	fact_klaim	Funnel
9	Klaim per faskes	Profil provider	Rp/faskes	dim_faskes	Ranking table
10	Deteksi anomali klaim (fraud)	Cegah fraud/abuse	Skor anomali	fact_klaim	Scatter/alert
C. Profil Peserta (360°)					
11	Demografi peserta aktif	Segmentasi	Jml peserta/segmen	dim_peserta	Donut/map
12	Riwayat layanan peserta	Care continuity	Touchpoints/peserta	mart_peserta360	Timeline
13	Beban penyakit kronis	Manajemen kronis (PRB)	Prevalensi	fact_diagnosa	Heatmap
14	Peserta high-cost	Manajemen risiko	Top-N biaya	mart_peserta360	Pareto
15	Tingkat keaktifan iuran vs utilisasi	Sustainability	Rasio klaim/iuran	mart_peserta360	Scatter
D. Epidemiologi & Diagnosa					
16	Distribusi diagnosa (ICD-10)	Surveilans penyakit	Top diagnosa	fact_diagnosa	Bar/treemap

#	Use Case	Objektif	KPI Utama	Sumber (Gold)	Visual
17	Tren penyakit musiman	Antisipasi lonjakan	Tren kasus	fact_diagnosa	Time series
18	Peta sebaran penyakit per wilayah	Targeting intervensi	Kasus/wilayah	dim_wilayah	Choropleth
19	Pola persesepan obat	Rasionalitas obat	Top obat	fact_obat	Bar
E. Provider & Wilayah					
20	Kinerja FKTP (kapitasi)	Pembayaran berbasis kinerja	KBK score	dim_faskes	Scorecard
21	Ketersediaan & sebaran faskes	Pemerataan akses	Rasio faskes/peserta	dim_faskes	Map
22	Beban layanan per wilayah	Alokasi sumber daya	Kunjungan/wilayah	dim_wilayah	Choropleth
23	Provider outlier biaya	Audit provider	Deviasi biaya	dim_provider	Scatter
24	SLA verifikasi klaim per RS	Mutu administrasi	Hari verifikasi	fact_klaim	Bar
25	Cakupan layanan (coverage)	Pemerataan akses	% peserta terlayani	mart_peserta360	KPI map
26	Monitoring kualitas data mart	Operasional data	DQ pass rate	dq_metrics	Dashboard ops

Rencana Pelaksanaan (Delivery Plan)

Timeline 3 bulan pengembangan + 1 bulan pemeliharaan, sesuai TOR. Knowledge transfer dilaksanakan pada masa pemeliharaan.

Timeline Proyek (16 minggu)



Fase	Minggu	Deliverable Kunci
Assessment & Design	1-4	Dokumen Evaluasi Sistem Eksisting; draft desain model & pipeline
Build Core	5-8	Pipeline ingestion + CDC; Bronze/Silver + DQ; Dokumen Desain Pipeline
Gold & Reporting	9-12	Gold mart + mart tematik; Superset; CLS/RLS; UAT; BAST; Dokumen Desain Model Data
Hypercare / Maintenance	13-16	Monitoring; troubleshooting; Knowledge Transfer; Laporan Pemeliharaan

Rencana Sumber Daya (Tim)

Komposisi tim mengikuti kualifikasi TOR (8 personel inti), diperkuat peran advisory tambahan untuk menjamin kualitas arsitektur, keamanan, dan tata kelola.

Peran	Pendidikan/Sertifikasi	Pengalaman	Jml	Tanggung Jawab
Project Manager / Team Leader	Min. S2 Komputer/Informatika/Manajemen/Teknik atau terkait TI	Min. 5 thn sebagai PM bidang data	1	Perencanaan, koordinasi, pelaksanaan proyek
Business Analyst (Muda)	Min. S1 terkait TI	Min. 3 thn bidang data	1	Evaluasi sistem eksisting & analisis kebutuhan bisnis
Data Designer (Muda)	Min. S1 terkait TI	Min. 3 thn bidang data	1	Perancangan arsitektur & struktur data
Data Engineer (Muda)	Min. S1 terkait TI	Min. 3 thn; ≥1 menguasai CDC, ≥1 menguasai Talend	3	Pengembangan job ELT/ETL & model data
Quality Control	Min. S1 terkait TI	Min. 3 thn QC	1	Pemeriksaan data & performa pipeline
Administrasi	Min. D3	Min. 1 thn administrasi	1	Dokumentasi pekerjaan
Peran Advisory Tambahan (nilai tambah Venturo Pro)				
Solution / Security Architect	S2/sertifikasi arsitektur & keamanan	Senior	1 (part-time)	Reviu arsitektur lakehouse & desain keamanan CLS/KMS
Data Governance Specialist	Sertifikasi DAMA/CDMP	Senior	1 (part-time)	Metadata, lineage, kebijakan tata kelola & DQ framework

Sesuai RAB TOR: PM 32 hari; BA, Data Designer, QC, Admin masing-masing 4 bulan; Data Engineer 3 orang × 3 bulan + 1 orang × 1 bulan (maintenance); laporan bulanan 4×.

Daftar Risiko (Risk Register)

#	Risiko	Prob.	Dampak	Mitigasi	Owner
1	Akses data terbatas (on-site, jam kerja) memperlambat dev	Tinggi	Tinggi	Jadwalkan kerja on-site efisien; siapkan data sample/sandbox; pekerjaan non-data online	PM
2	Kualitas data sumber lebih buruk dari perkiraan	Tinggi	Tinggi	Profiling dini di fase evaluasi; quarantine + remediation plan	QC
3	Jumlah/skema 42 tabel berubah pasca-evaluasi	Sedang	Sedang	Schema evolution Iceberg; desain pipeline parametrik	Data Engineer
4	Beban CDC berdampak ke OLTP sumber	Sedang	Tinggi	Log-based CDC low-impact; throttling; uji performa	Data Engineer
5	Kapasitas kluster Hadoop tidak memadai	Sedang	Tinggi	Capacity test + proyeksi; tuning partisi/kompaksi	Solution Architect
6	Integrasi Voltage/CLS kompleks	Sedang	Sedang	PoC dini; koordinasi tim security BPJS	Security Architect
7	Timeline 3 bulan ketat	Tinggi	Tinggi	Scope per milestone; paralelisasi; reuse template	PM
8	Ketergantungan keputusan/approval BPJS	Sedang	Sedang	Cadence governance mingguan; RACI jelas	PM
9	Knowledge transfer tidak efektif	Rendah	Sedang	Workshop + pendampingan + materi referensi terstruktur	Data Governance Specialist
10	Resistensi pengguna terhadap tool baru (Superset)	Rendah	Rendah	Pelatihan + dashboard intuitif + champion internal	BA
11	Scope creep permintaan tambahan	Sedang	Sedang	Change control + baseline scope per TOR	PM
12	Risiko keamanan/kebocoran data selama dev	Rendah	Kritis	Akses terkontrol, audit, NDA, no data extract	Security Architect

Kriteria Sukses	Metrik	Target (ilustratif)
Arsitektur lakehouse operasional	Zona medallion + Iceberg ACID berjalan	100% layer aktif
Pipeline ELT + CDC berjalan	Job terjadwal sukses	≥ 98% success rate
Kualitas data Gold	DQ pass rate 6 dimensi	≥ 95%
Latensi refresh	Durasi siklus end-to-end	≤ 30 menit (incremental)
Dampak ke sistem sumber	Overhead OLTP saat ekstraksi	Tidak signifikan
Keamanan kolom	PII terproteksi (CLS/masking)	100% kolom sensitif
Dokumentasi teknis	Model, pipeline, kamus data, lineage	Lengkap & diserahkan
Knowledge transfer	Workshop + materi + absensi	Tim internal mandiri
Adopsi pelaporan	Dashboard self-service dipakai	≥ target unit pengguna

Keahlian Lakehouse Modern

Pengalaman pada Iceberg, Spark, Airflow/Talend, Trino, dan Superset di lingkungan Hadoop on-premise skala besar.

Governance & Security First

Praktik DAMA-DMBOK, COBIT, ISO 27001, dan kepatuhan UU PDP melekat dalam metodologi, bukan tambahan.

Eksekusi Tepat Scope

Metodologi terstruktur per milestone yang mengeksekusi seluruh ruang lingkup TOR dalam 3+1 bulan.

Keberlanjutan & Alih Pengetahuan

Dokumentasi menyeluruh + knowledge transfer interaktif memastikan kemandirian tim internal BPJS Kesehatan.

Rekomendasi Eksekutif

Venturo Pro merekomendasikan untuk **melanjutkan (GO)** inisiatif transformasi data mart Pelayanan menjadi arsitektur lakehouse medallion. Inisiatif ini layak secara teknis, finansial, operasional, dan regulasi; memanfaatkan infrastruktur eksisting; dan memberikan fondasi data yang skalabel, aman, dan tepercaya untuk mendukung pelaporan serta analitik organisasi jangka panjang.

Venturo

EXPERT PROGRAMMERS

Research & Development Office

Jl Rinjani No 26 Oro Oro Dowo Kec. Klojen
Kota Malang Jawa Timur 65119

Proposal Teknis — Pengembangan Solusi Data Modelling dan Reporting
Terintegrasi Kluster Pelayanan dengan Teknologi Big Data untuk BPJS
Kesehatan.

Email **hello@venturo.id**

WhatsApp **0851-2804-3814**

Instagram **venturo.pro**

TikTok **venturo.pro**

Dokumen Rahasia · Versi 1.0 · Juni 2026

Venturo Pro © 2026 — The Biggest Programmer Resource in Malang, Jawa Timur.