

Memilih layanan AWS analitik



Memilih layanan AWS analitik: AWS Panduan Keputusan

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Merek dagang dan tampilan dagang Amazon tidak boleh digunakan sehubungan dengan produk atau layanan apa pun yang bukan milik Amazon, dengan cara apa pun yang dapat menyebabkan kebingungan di antara pelanggan, atau dengan cara apa pun yang merendahkan atau mendiskreditkan Amazon. Semua merek dagang lain yang tidak dimiliki oleh Amazon merupakan hak milik masing-masing pemiliknya, yang mungkin atau mungkin tidak terafiliasi, terkait dengan, atau disponsori oleh Amazon.

Table of Contents

Panduan keputusan	1
Pengantar	1
Memahami	2
Pertimbangkan	6
Pilih	15
Gunakan	19
Jelajahi	28
Riwayat dokumen	30
.....	xxxi

Memilih layanan AWS analitik

Mengambil langkah pertama

Tujuan	Bantu menentukan layanan AWS analitik mana yang paling cocok untuk organisasi Anda.
Terakhir diperbarui	September 24, 2025
Layanan yang tercakup	<ul style="list-style-type: none">• Amazon Athena• AWS Clean Rooms• Amazon Data Firehose• Amazon DataZone• Amazon EMR• AWS Glue• Amazon Kinesis Data Streams• Layanan Dikelola Amazon untuk Apache Flink• Amazon Managed Streaming for Apache Kafka• Alur Kerja Terkelola Amazon untuk Apache Airflow• OpenSearch Layanan Amazon• Quick• Amazon Redshift• Amazon S3• Amazon SageMaker

Pengantar

Data adalah dasar bagi bisnis modern. Orang dan aplikasi perlu mengakses dan menganalisis data dengan aman, yang berasal dari sumber baru dan beragam. Volume data juga terus meningkat, yang

dapat menyebabkan organisasi kesulitan menangkap, menyimpan, dan menganalisis semua data yang diperlukan.

Memenuhi tantangan ini berarti membangun arsitektur data modern yang memecah semua silo data Anda untuk analitik dan wawasan - termasuk data pihak ketiga - dan membuatnya dapat diakses oleh semua orang di organisasi, di satu tempat, dengan tata kelola end-to-end. Juga semakin penting untuk menghubungkan sistem analitik dan pembelajaran mesin (ML) Anda untuk mengaktifkan analitik prediktif.

Panduan keputusan ini membantu Anda mengajukan pertanyaan yang tepat untuk membangun arsitektur data modern Anda pada AWS layanan. Ini menjelaskan cara memecah silo data Anda (dengan menghubungkan data lake dan gudang data Anda), silo sistem Anda (dengan menghubungkan ML dan analitik), dan silo orang Anda (dengan meletakkan data di tangan semua orang di organisasi Anda).

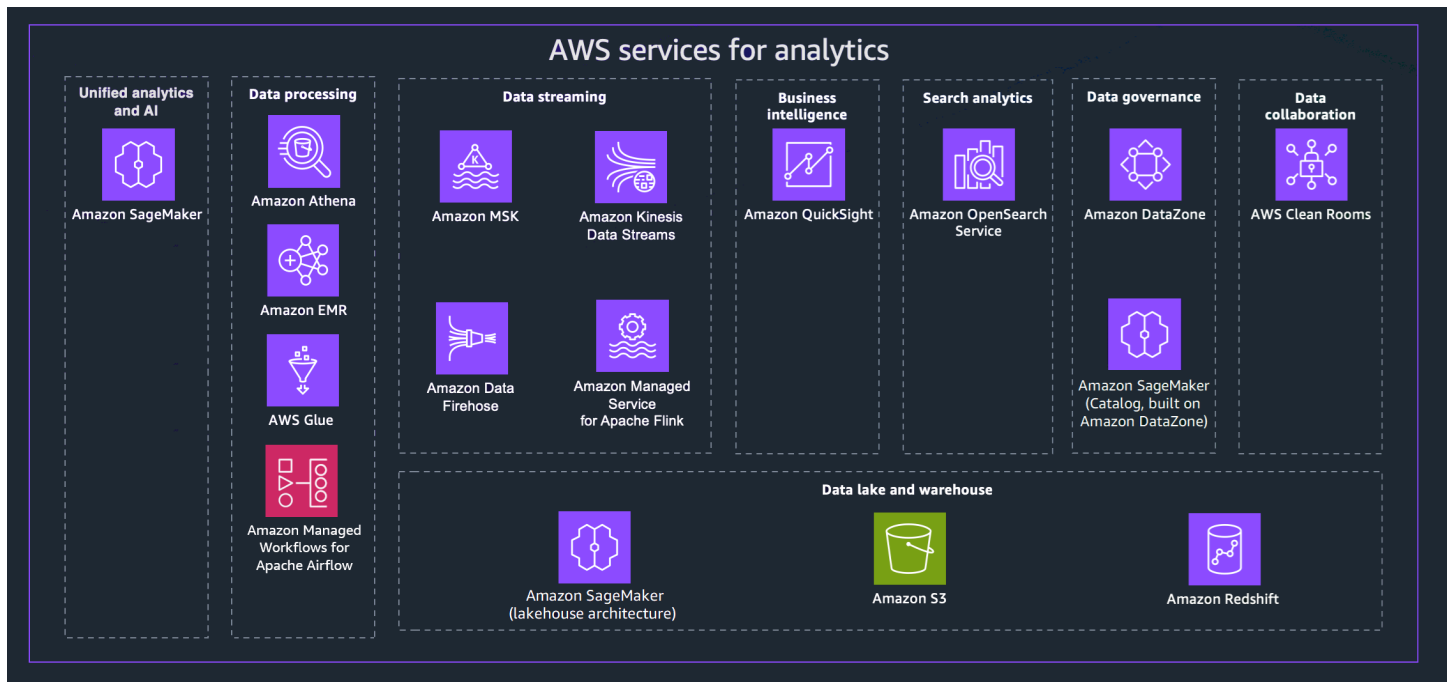
[Latihan delapan menit ini berasal dari presentasi satu jam oleh Sirish Chandrasekaran dan Rick Sears di re:Invent 2024. Ini memberikan gambaran umum tentang bagaimana perusahaan fiksi, Maxdome, menggunakan SageMaker Unified Studio AI dan analitik — bagian dari generasi Amazon berikutnya — untuk membuka wawasan data. SageMaker](#)

Memahami layanan AWS analitik

Strategi data modern dibangun dengan seperangkat blok bangunan teknologi yang membantu Anda mengelola, mengakses, menganalisis, dan bertindak berdasarkan data. Ini juga memberi Anda beberapa opsi untuk terhubung ke sumber data. Strategi data modern harus memberdayakan tim Anda untuk:

- Gunakan alat atau teknik pilihan Anda
- Gunakan kecerdasan buatan (AI) untuk membantu menemukan jawaban atas pertanyaan spesifik tentang data Anda
- Kelola siapa yang memiliki akses ke data dengan kontrol keamanan dan tata kelola data yang tepat
- Memecah silo data untuk memberikan yang terbaik dari kedua data lake dan penyimpanan data yang dibuat khusus
- Menyimpan sejumlah data, dengan biaya rendah, dan dalam format data berbasis standar terbuka
- Hubungkan data lake, gudang data, database operasional, aplikasi, dan sumber data federasi Anda ke dalam keseluruhan yang koheren

AWS menawarkan berbagai layanan untuk membantu Anda mencapai strategi data modern. Diagram berikut menggambarkan AWS layanan untuk analitik yang dicakup oleh panduan ini. Tab yang mengikuti memberikan detail tambahan.



Unified analytics and AI

Generasi [Amazon](#) berikutnya SageMaker menggabungkan kemampuan pembelajaran AWS mesin (ML) dan analitik yang diadopsi secara luas untuk memberikan pengalaman terintegrasi untuk analitik dan AI, menyediakan akses terpadu ke semua data Anda. Menggunakan [Amazon SageMaker Unified Studio](#), Anda dapat berkolaborasi dan membangun lebih cepat dengan AWS alat yang sudah dikenal untuk pengembangan model, pengembangan aplikasi AI generatif, pemrosesan data, dan analitik SQL, semuanya dipercepat oleh Amazon Q Developer, asisten AI generatif kami untuk pengembangan perangkat lunak. Akses data Anda dari data lake, gudang data, atau sumber pihak ketiga dan federasi, dengan tata kelola bawaan untuk memenuhi persyaratan keamanan perusahaan.

Data processing

- [Amazon Athena](#) membantu Anda menganalisis data tidak terstruktur, semi-terstruktur, dan terstruktur yang disimpan di Amazon S3. Contohnya termasuk format data CSV, JSON, atau kolumnar seperti Apache Parquet dan Apache ORC. Anda dapat menggunakan Athena untuk menjalankan kueri ad-hoc menggunakan ANSI SQL, tanpa perlu mengumpulkan atau memuat data ke Athena. [Athena terintegrasi dengan Quick, AWS Glue Data Catalog, dan layanan lainnya.](#) [AWS](#) Anda juga dapat menganalisis data dalam skala besar dengan [Trino](#), tanpa perlu

mengelola infrastruktur, dan membangun analitik real-time menggunakan Apache Flink dan Apache Spark.

- [Amazon EMR](#) adalah platform cluster terkelola yang menyederhanakan menjalankan kerangka kerja data besar, seperti Apache Hadoop dan Apache Spark, untuk memproses dan menganalisis sejumlah besar data AWS . Dengan menggunakan kerangka kerja ini dan proyek sumber terbuka terkait, Anda dapat memproses data untuk tujuan analitik dan beban kerja intelijen bisnis. Amazon EMR juga memungkinkan Anda mengubah dan memindahkan sejumlah besar data ke dalam dan keluar dari penyimpanan AWS data dan database lainnya, seperti Amazon S3.
- Dengan [AWS Glue](#), Anda dapat menemukan dan terhubung ke lebih dari 100 sumber data yang beragam dan mengelola data Anda dalam katalog data terpusat. Anda dapat membuat, menjalankan, dan memantau pipeline ETL secara visual untuk memuat data ke dalam data lake Anda. Selain itu, Anda dapat segera mencari dan menanyakan data katalog menggunakan Athena, Amazon EMR, dan Amazon Redshift Spectrum.
- [Amazon Managed Workflows for Apache](#) Airflow (MWAA) adalah implementasi Apache Airflow yang dikelola sepenuhnya yang memudahkan pembuatan, jadwal, dan pemantauan alur kerja data di cloud. MWAA secara otomatis menskalakan kapasitas alur kerja untuk memenuhi kebutuhan Anda dan terintegrasi dengan layanan keamanan. AWS Anda dapat menggunakan MWAA untuk mengatur alur kerja di seluruh layanan analitik Anda, termasuk pemrosesan data, pekerjaan ETL, dan pipeline pembelajaran mesin.

Data streaming

- Dengan [Amazon Managed Streaming for Apache](#) Kafka (Amazon MSK), Anda dapat membangun dan menjalankan aplikasi yang menggunakan Apache Kafka untuk memproses data streaming. Amazon MSK menyediakan operasi bidang kontrol, seperti untuk membuat, memperbarui, dan menghapus cluster. Ini memungkinkan Anda menggunakan operasi data-plane Apache Kafka, seperti untuk memproduksi dan mengkonsumsi data.
- Dengan [Amazon Kinesis Data](#) Streams, Anda dapat mengumpulkan dan memproses aliran besar catatan data secara real time. Jenis data yang digunakan dapat mencakup data log infrastruktur TI, log aplikasi, media sosial, umpan data pasar, dan data clickstream web.
- [Amazon Data Firehose](#) adalah layanan yang dikelola sepenuhnya untuk mengirimkan data streaming real-time ke tujuan seperti Amazon S3, Amazon Redshift, OpenSearch Amazon Service, Splunk, dan Apache Iceberg Tables. Anda juga dapat mengirim data ke titik akhir HTTP kustom atau titik akhir HTTP yang dimiliki oleh penyedia layanan pihak ketiga yang

didukung, termasuk Datadog, Dynatrace, MongoDB, New Relic, LogicMonitor Coralogix, dan Elastic.

- Dengan [Amazon Managed Service untuk Apache Flink](#), Anda dapat menggunakan Java, Scala, Python, atau SQL untuk memproses dan menganalisis data streaming. Anda dapat membuat dan menjalankan kode terhadap sumber streaming dan sumber statis untuk melakukan analitik deret waktu, memberi umpan dasbor waktu nyata, dan metrik.

Business intelligence

[Quick](#) memberi pengambil keputusan kesempatan untuk mengeksplorasi dan menafsirkan informasi dalam lingkungan visual yang interaktif. Dalam satu dasbor data, Quick dapat menyertakan AWS data, data pihak ketiga, data besar, data spreadsheet, data SaaS, data B2B, dan banyak lagi. Dengan Quick Q, Anda dapat menggunakan bahasa alami untuk mengajukan pertanyaan tentang data Anda dan menerima tanggapan. Misalnya, “Apa kategori terlaris di California?”

Search analytics

[OpenSearch Layanan Amazon menyediakan](#) semua sumber daya untuk OpenSearch klaster Anda dan meluncurkannya. Ini juga secara otomatis mendeteksi dan mengganti node OpenSearch Layanan yang gagal, mengurangi overhead yang terkait dengan infrastruktur yang dikelola sendiri. Anda dapat menggunakan kueri langsung OpenSearch Layanan untuk menganalisis data di Amazon S3 dan layanan lainnya AWS .

Data governance

Dengan [Amazon DataZone](#), Anda dapat mengelola dan mengatur akses ke data dengan menggunakan kontrol berbutir halus. Kontrol ini membantu memastikan akses dengan tingkat hak istimewa dan konteks yang tepat. Amazon DataZone menyederhanakan arsitektur Anda dengan mengintegrasikan layanan manajemen data, termasuk Amazon Redshift, Athena, AWS Glue Quick,, sumber lokal, dan sumber pihak ketiga.

Data collaboration

[AWS Clean Rooms](#) adalah ruang kerja kolaborasi yang aman di mana Anda dapat menganalisis kumpulan data kolektif tanpa memberikan akses ke data mentah. Anda dapat berkolaborasi dengan perusahaan lain dengan memilih mitra yang ingin Anda ajak berkolaborasi, memilih kumpulan data mereka, dan mengonfigurasi kontrol peningkatan privasi untuk mitra tersebut. Saat Anda menjalankan kueri, AWS Clean Rooms membaca data dari lokasi asli data tersebut dan

menerapkan aturan analisis bawaan untuk membantu Anda mempertahankan kendali atas data tersebut.

Data lake and data warehouse

- [Amazon generasi berikutnya](#) sepenuhnya SageMaker kompatibel dengan Apache Iceberg, memungkinkan Anda menyatukan data di seluruh Amazon Simple Storage Service (Amazon S3) data lake dan gudang data Amazon Redshift. Hal ini memungkinkan pembuatan analitik dan aplikasi AI dan pembelajaran mesin (ML) pada satu salinan data. Melalui integrasi Zero-ETL, Anda dapat melakukan streaming data dari sumber operasional dalam waktu dekat, menjalankan kueri gabungan di berbagai sumber, dan mengakses data dengan alat yang kompatibel dengan Apache Iceberg. Anda dapat mengamankan data Anda dengan menentukan izin berbutir halus yang diberlakukan di semua analitik dan alat dan mesin MLmu.
- [Amazon S3](#) dapat menyimpan dan melindungi hampir semua jumlah dan jenis data, yang dapat Anda gunakan untuk fondasi danau data Anda. Amazon S3 menyediakan fitur manajemen sehingga Anda dapat mengoptimalkan, mengatur, dan mengonfigurasi akses ke data Anda untuk memenuhi persyaratan bisnis, organisasi, dan kepatuhan spesifik Anda. Tabel Amazon S3 menyediakan penyimpanan S3 yang dioptimalkan untuk beban kerja analitik. Menggunakan pernyataan SQL standar, Anda dapat menanyakan tabel Anda dengan mesin kueri yang mendukung Iceberg, seperti Athena, Amazon Redshift, dan Apache Spark.
- [Amazon Redshift adalah layanan](#) gudang data skala petabyte yang dikelola sepenuhnya. Amazon Redshift dapat dihubungkan ke data lakehouse di SageMaker Amazon, memungkinkan Anda menggunakan kemampuan analitik SQL yang kuat pada data terpadu Anda di seluruh gudang data Amazon Redshift dan danau data Amazon S3. Anda juga dapat menggunakan Amazon Q di Amazon Redshift, yang menyederhanakan penulisan SQL melalui bahasa alami.

Pertimbangkan kriteria untuk layanan AWS analitik

Ada banyak alasan untuk membangun analisis data AWS. Anda mungkin perlu mendukung greenfield atau proyek percontohan sebagai langkah pertama dalam perjalanan migrasi cloud Anda. Atau, Anda mungkin memigrasikan beban kerja yang ada dengan gangguan sesedikit mungkin. Apa pun tujuan Anda, pertimbangan berikut dapat berguna dalam menentukan pilihan Anda.

Assess data sources and data types

Menganalisis sumber data dan tipe data yang tersedia untuk mendapatkan pemahaman yang komprehensif tentang keragaman data, frekuensi, dan kualitas. Memahami setiap tantangan potensial dalam memproses dan menganalisis data. Analisis ini sangat penting karena:

- Sumber data beragam dan berasal dari berbagai sistem, aplikasi, perangkat, dan platform eksternal.
- Sumber data memiliki struktur, format, dan frekuensi pembaruan data yang unik. Menganalisis sumber-sumber ini membantu dalam mengidentifikasi metode dan teknologi pengumpulan data yang sesuai.
- Menganalisis tipe data, seperti data terstruktur, semi-terstruktur, dan tidak terstruktur menentukan pendekatan pemrosesan dan penyimpanan data yang tepat.
- Menganalisis sumber dan tipe data memfasilitasi penilaian kualitas data, membantu Anda mengantisipasi potensi masalah kualitas data — nilai yang hilang, inkonsistensi, atau ketidakakuratan.

Data processing requirements

Menentukan persyaratan pemrosesan data untuk bagaimana data dicerna, diubah, dibersihkan, dan disiapkan untuk analisis. Pertimbangan utama meliputi:

- Transformasi data: Tentukan transformasi spesifik yang diperlukan untuk membuat data mentah cocok untuk analisis. Ini melibatkan tugas-tugas seperti agregasi data, normalisasi, penyaringan, dan pengayaan.
- Pembersihan data: Menilai kualitas data dan menentukan proses untuk menangani data yang hilang, tidak akurat, atau tidak konsisten. Menerapkan teknik pembersihan data untuk memastikan data berkualitas tinggi untuk wawasan yang andal.
- Frekuensi pemrosesan: Tentukan apakah pemrosesan real-time, mendekati waktu nyata, atau batch diperlukan berdasarkan kebutuhan analitis. Pemrosesan waktu nyata memungkinkan wawasan langsung, sementara pemrosesan batch mungkin cukup untuk analisis berkala.
- Skalabilitas dan throughput: Mengevaluasi persyaratan skalabilitas untuk menangani volume data, kecepatan pemrosesan, dan jumlah permintaan data bersamaan. Pastikan bahwa pendekatan pemrosesan yang dipilih dapat mengakomodasi pertumbuhan masa depan.

- **Latensi:** Pertimbangkan latensi yang dapat diterima untuk pemrosesan data dan waktu yang dibutuhkan dari konsumsi data hingga hasil analisis. Ini sangat penting untuk analitik real-time atau sensitif waktu.

Storage requirements

Tentukan kebutuhan penyimpanan dengan menentukan bagaimana dan di mana data disimpan di seluruh pipeline analitik. Pertimbangan penting meliputi:

- **Volume data:** Menilai jumlah data yang dihasilkan dan dikumpulkan, dan memperkirakan pertumbuhan data masa depan untuk merencanakan kapasitas penyimpanan yang memadai.
- **Retensi data:** Tentukan durasi penyimpanan data untuk analisis historis atau tujuan kepatuhan. Tentukan kebijakan penyimpanan data yang sesuai.
- **Pola akses data:** Memahami bagaimana data akan diakses dan ditanyakan untuk memilih solusi penyimpanan yang paling sesuai. Pertimbangkan operasi baca dan tulis, frekuensi akses data, dan lokalitas data.
- **Keamanan data:** Prioritaskan keamanan data dengan mengevaluasi opsi enkripsi, kontrol akses, dan mekanisme perlindungan data untuk melindungi informasi sensitif.
- **Optimalisasi biaya:** Optimalkan biaya penyimpanan dengan memilih solusi penyimpanan yang paling hemat biaya berdasarkan pola dan penggunaan akses data.
- **Integrasi dengan layanan analitik:** Pastikan integrasi yang mulus antara solusi penyimpanan yang dipilih dan alat pemrosesan data dan analitik dalam pipeline.

Types of data

Saat memutuskan layanan analitik untuk pengumpulan dan penyerapan data, pertimbangkan berbagai jenis data yang relevan dengan kebutuhan dan tujuan organisasi Anda. Jenis data umum yang mungkin perlu Anda pertimbangkan meliputi:

- **Data transaksional:** Mencakup informasi tentang interaksi atau transaksi individu, seperti pembelian pelanggan, transaksi keuangan, pesanan online, dan log aktivitas pengguna.
- **Data berbasis file:** Mengacu pada data terstruktur atau tidak terstruktur yang disimpan dalam file, seperti file log, spreadsheet, dokumen, gambar, file audio, dan file video. Layanan Analytics harus mendukung konsumsi format file yang berbeda.

- **Data peristiwa:** Menangkap kejadian atau insiden yang signifikan, seperti tindakan pengguna, peristiwa sistem, peristiwa mesin, atau peristiwa bisnis. Peristiwa dapat mencakup data apa pun yang tiba dalam kecepatan tinggi yang ditangkap untuk pemrosesan onstream atau hilir.

Operational considerations

Tanggung jawab operasional dibagi antara Anda, dan AWS, dengan pembagian tanggung jawab yang bervariasi di berbagai tingkat modernisasi. Anda memiliki pilihan untuk mengelola sendiri infrastruktur analitik Anda AWS atau memanfaatkan berbagai layanan analitik tanpa server untuk mempelajari beban manajemen infrastruktur Anda.

Opsi yang dikelola sendiri memberi pengguna kontrol yang lebih besar atas infrastruktur dan konfigurasi, tetapi mereka membutuhkan lebih banyak upaya operasional.

Opsi tanpa server mengabstraksikan sebagian besar beban operasional, menyediakan skalabilitas otomatis, ketersediaan tinggi, dan fitur keamanan yang kuat, memungkinkan pengguna untuk lebih fokus pada membangun solusi analitis dan mendorong wawasan daripada mengelola infrastruktur dan tugas operasional. Pertimbangkan manfaat dari solusi analitik tanpa server ini:

- **Abstraksi infrastruktur:** Layanan tanpa server manajemen infrastruktur abstrak, membebaskan pengguna dari tugas penyediaan, penskalaan, dan pemeliharaan. AWS menangani aspek operasional ini, mengurangi overhead manajemen.
- **Penskalaan dan kinerja otomatis:** Layanan tanpa server secara otomatis menskalakan sumber daya berdasarkan tuntutan beban kerja, memastikan kinerja optimal tanpa intervensi manual.
- **Ketersediaan tinggi dan pemulihan bencana:** AWS menyediakan ketersediaan tinggi untuk layanan tanpa server. AWS mengelola redundansi data, replikasi, dan pemulihan bencana untuk meningkatkan ketersediaan dan keandalan data.
- **Keamanan dan kepatuhan:** AWS mengelola langkah-langkah keamanan, enkripsi data, dan kepatuhan untuk layanan tanpa server, mematuhi standar industri dan praktik terbaik.
- **Pemantauan dan pencatatan:** AWS menawarkan kemampuan pemantauan, pencatatan, dan peringatan bawaan untuk layanan tanpa server. Pengguna dapat mengakses metrik dan log terperinci melalui Amazon CloudWatch.

Type of workload

Saat membangun saluran analitik modern, menentukan jenis beban kerja yang akan didukung sangat penting untuk memenuhi kebutuhan analitis yang berbeda secara efektif. Poin keputusan utama yang perlu dipertimbangkan untuk setiap jenis beban kerja meliputi:

Beban kerja Batch

- Volume dan frekuensi data: Pemrosesan batch cocok untuk volume data yang besar dengan pembaruan berkala.
- Latensi data: Pemrosesan batch mungkin menimbulkan beberapa keterlambatan dalam memberikan wawasan dibandingkan dengan pemrosesan waktu nyata.

Analisis interaktif

- Kompleksitas kueri data: Analisis interaktif membutuhkan respons latensi rendah untuk umpan balik cepat.
- Visualisasi data: Mengevaluasi kebutuhan alat visualisasi data interaktif untuk memungkinkan pengguna bisnis mengeksplorasi data secara visual.

Streaming beban kerja

- Kecepatan dan volume data: Beban kerja streaming memerlukan pemrosesan waktu nyata untuk menangani data berkecepatan tinggi.
- Data windowing: Tentukan jendela data dan agregasi berbasis waktu untuk streaming data guna mengekstrak wawasan yang relevan.

Type of analysis needed

Tentukan dengan jelas tujuan bisnis dan wawasan yang ingin Anda peroleh dari analitik. Berbagai jenis analitik melayani tujuan yang berbeda. Contoh:

- Analisis deskriptif sangat ideal untuk mendapatkan gambaran historis
- Analisis diagnostik membantu memahami alasan di balik peristiwa masa lalu
- Analisis prediktif memperkirakan hasil masa depan
- Analisis preskriptif memberikan rekomendasi untuk tindakan optimal

Cocokkan tujuan bisnis Anda dengan jenis analitik yang relevan. Berikut adalah beberapa kriteria keputusan utama untuk membantu Anda memilih jenis analitik yang tepat:

- **Ketersediaan dan kualitas data:** Analisis deskriptif dan diagnostik bergantung pada data historis, sementara analitik prediktif dan preskriptif memerlukan data historis yang cukup dan data berkualitas tinggi untuk membangun model yang akurat.
- **Volume dan kompleksitas data:** Analisis prediktif dan preskriptif memerlukan pemrosesan data yang substansif dan sumber daya komputasi. Pastikan infrastruktur dan alat Anda dapat menangani volume dan kompleksitas data.
- **Kompleksitas keputusan:** Jika keputusan melibatkan banyak variabel, kendala, dan tujuan, analitik preskriptif mungkin lebih cocok untuk memandu tindakan yang optimal.
- **Toleransi risiko:** Analisis preskriptif dapat memberikan rekomendasi, tetapi disertai dengan ketidakpastian terkait. Pastikan bahwa pengambil keputusan memahami risiko yang terkait dengan output analitik.

Evaluate scalability and performance

Menilai skalabilitas dan kebutuhan kinerja arsitektur. Desain harus menangani peningkatan volume data, permintaan pengguna, dan beban kerja analitis. Faktor keputusan utama yang perlu dipertimbangkan meliputi:

- **Volume dan pertumbuhan data:** Menilai volume data saat ini dan mengantisipasi pertumbuhan di masa depan.
- **Kecepatan data dan persyaratan waktu nyata:** Tentukan apakah data perlu diproses dan dianalisis secara real-time atau mendekati waktu nyata.
- **Kompleksitas pemrosesan data:** Menganalisis kompleksitas tugas pemrosesan dan analisis data Anda. Untuk tugas komputasi intensif, layanan seperti Amazon EMR menyediakan lingkungan yang terukur dan terkelola untuk pemrosesan data besar.
- **Konkurensi dan beban pengguna:** Pertimbangkan jumlah pengguna bersamaan dan tingkat beban pengguna pada sistem.
- **Kemampuan penskalaan otomatis:** Pertimbangkan layanan yang menawarkan kemampuan auto-scaling, yang memungkinkan sumber daya naik atau turun secara otomatis berdasarkan permintaan. Ini memastikan pemanfaatan sumber daya yang efisien dan optimalisasi biaya.
- **Distribusi geografis:** Pertimbangkan layanan dengan replikasi global dan akses data latensi rendah jika arsitektur data Anda perlu didistribusikan di beberapa wilayah atau lokasi.

- Trade-off biaya-kinerja: Seimbangkan kebutuhan kinerja dengan pertimbangan biaya. Layanan dengan kinerja tinggi mungkin datang dengan biaya yang lebih tinggi.
- Perjanjian tingkat layanan (SLAs): Periksa AWS layanan yang SLAs disediakan untuk memastikan mereka memenuhi skalabilitas dan ekspektasi kinerja Anda.

Data governance

Tata kelola data adalah serangkaian proses, kebijakan, dan kontrol yang perlu Anda terapkan untuk memastikan manajemen, kualitas, keamanan, dan kepatuhan aset data Anda yang efektif. Poin keputusan utama yang perlu dipertimbangkan meliputi:

- Kebijakan penyimpanan data: Menentukan kebijakan penyimpanan data berdasarkan persyaratan peraturan dan kebutuhan bisnis dan menetapkan proses untuk pembuangan data yang aman ketika tidak lagi diperlukan.
- Jejak audit dan pencatatan: Tentukan mekanisme pencatatan dan audit untuk memantau akses dan penggunaan data. Menerapkan jejak audit komprehensif untuk melacak perubahan data, upaya akses, dan aktivitas pengguna untuk kepatuhan dan pemantauan keamanan.
- Persyaratan kepatuhan: Memahami peraturan kepatuhan data spesifik industri dan geografis yang berlaku untuk organisasi Anda. Pastikan arsitektur data selaras dengan peraturan dan pedoman ini.
- Klasifikasi data: Klasifikasi data berdasarkan sensitivitasnya dan tentukan kontrol keamanan yang sesuai untuk setiap kelas data.
- Pemulihan bencana dan kelangsungan bisnis: Merencanakan pemulihan bencana dan kelangsungan bisnis untuk memastikan ketersediaan dan ketahanan data jika terjadi kejadian tak terduga atau kegagalan sistem.
- Berbagi data pihak ketiga: Jika berbagi data dengan entitas pihak ketiga, terapkan protokol dan perjanjian berbagi data yang aman untuk melindungi kerahasiaan data dan mencegah penyalahgunaan data.

Security

Keamanan data dalam pipeline analitik melibatkan perlindungan data di setiap tahap pipeline untuk memastikan kerahasiaan, integritas, dan ketersediaannya. Poin keputusan utama yang perlu dipertimbangkan meliputi:

- Kontrol akses dan otorisasi: Menerapkan protokol otentikasi dan otorisasi yang kuat untuk memastikan bahwa hanya pengguna yang berwenang yang dapat mengakses sumber daya data tertentu.
- Enkripsi data: Pilih metode enkripsi yang sesuai untuk data yang disimpan dalam database, data lake, dan selama pergerakan data antara berbagai komponen arsitektur.
- Penyembunyian dan anonimisasi data: Pertimbangkan perlunya penyembunyian data atau anonimisasi untuk melindungi data sensitif, seperti PII atau data bisnis sensitif, sambil memungkinkan proses analitis tertentu berlanjut.
- Integrasi data aman: Menetapkan praktik integrasi data yang aman untuk memastikan bahwa data mengalir dengan aman di antara berbagai komponen arsitektur, menghindari kebocoran data atau akses tidak sah selama pergerakan data.
- Isolasi jaringan: Pertimbangkan layanan yang mendukung [Titik Akhir VPC Amazon](#) untuk menghindari mengekspos sumber daya ke internet publik.

Plan for integration and data flows

Tentukan titik integrasi dan aliran data antara berbagai komponen pipa analitik untuk memastikan aliran data dan interoperabilitas yang mulus. Poin keputusan utama yang perlu dipertimbangkan meliputi:

- Integrasi sumber data: Mengidentifikasi sumber data dari mana data akan dikumpulkan, seperti database, aplikasi, file, atau eksternal APIs. Tentukan metode konsumsi data (batch, real-time, berbasis peristiwa) untuk membawa data ke dalam pipeline secara efisien dan dengan latensi minimal.
- Transformasi data: Menentukan transformasi yang diperlukan untuk menyiapkan data untuk analisis. Tentukan alat dan proses untuk membersihkan, mengumpulkan, menormalkan, atau memperkaya data saat bergerak melalui pipa.
- Arsitektur pergerakan data: Pilih arsitektur yang sesuai untuk pergerakan data antar komponen pipeline. Pertimbangkan pemrosesan batch, pemrosesan aliran, atau kombinasi keduanya berdasarkan persyaratan waktu nyata dan volume data.
- Replikasi dan sinkronisasi data: Tentukan mekanisme replikasi dan sinkronisasi data untuk menyimpan data up-to-date di semua komponen. Pertimbangkan solusi replikasi waktu nyata atau sinkronisasi data berkala tergantung pada persyaratan kesegaran data.

- **Kualitas dan validasi data:** Menerapkan pemeriksaan kualitas data dan langkah-langkah validasi untuk memastikan integritas data saat bergerak melalui pipeline. Tentukan tindakan yang akan diambil ketika data gagal validasi, seperti peringatan atau penanganan kesalahan.
- **Keamanan dan enkripsi data:** Tentukan bagaimana data akan diamankan selama transit dan saat istirahat. Tentukan metode enkripsi untuk melindungi data sensitif di seluruh pipeline, dengan mempertimbangkan tingkat keamanan yang diperlukan berdasarkan sensitivitas data.
- **Skalabilitas dan ketahanan:** Pastikan bahwa desain aliran data memungkinkan skalabilitas horizontal dan dapat menangani peningkatan volume data dan lalu lintas.

Architect for cost optimization

Membangun saluran analitik Anda AWS memberikan berbagai peluang pengoptimalan biaya. Untuk memastikan efisiensi biaya, pertimbangkan strategi berikut:

- **Ukuran dan pemilihan sumber daya:** Ukuran sumber daya Anda dengan tepat berdasarkan persyaratan beban kerja aktual. Pilih AWS layanan dan jenis instans yang sesuai dengan kebutuhan kinerja beban kerja sambil menghindari penyediaan berlebihan.
- **Penskalaan otomatis:** Terapkan auto-scaling untuk layanan yang mengalami berbagai beban kerja. Penskalaan otomatis secara dinamis menyesuaikan jumlah instance berdasarkan permintaan, mengurangi biaya selama periode lalu lintas rendah.
- **Instans Spot:** Gunakan Instans Spot Amazon EC2 untuk beban kerja yang tidak kritis dan toleran terhadap kesalahan. Instans Spot dapat secara signifikan mengurangi biaya dibandingkan dengan instans sesuai permintaan.
- **Instans Cadangan:** Pertimbangkan untuk membeli Instans AWS Cadangan untuk mencapai penghematan biaya yang signifikan dibandingkan harga sesuai permintaan untuk beban kerja yang stabil dengan penggunaan yang dapat diprediksi.
- **Tingkat penyimpanan data:** Optimalkan biaya penyimpanan data dengan menggunakan kelas penyimpanan yang berbeda berdasarkan frekuensi akses data.
- **Kebijakan siklus hidup data:** Menetapkan kebijakan siklus hidup data untuk memindahkan atau menghapus data secara otomatis berdasarkan usia dan pola penggunaannya. Ini membantu mengelola biaya penyimpanan dan menjaga penyimpanan data selaras dengan nilainya.

Pilih layanan AWS analitik

Sekarang setelah Anda mengetahui kriteria untuk mengevaluasi kebutuhan analitik Anda, Anda siap untuk memilih layanan AWS analitik mana yang tepat untuk kebutuhan organisasi Anda. Tabel berikut menyelaraskan rangkaian layanan dengan kemampuan umum dan tujuan bisnis.

Kategori	Untuk apa itu dioptimalkan?	Layanan
Pengolahan data	<p>Analitik terpadu dan AI</p> <p>Analisis dan pengembangan AI</p> <p>Dioptimalkan untuk menggunakan lingkungan pengembangan tunggal, Amazon SageMaker Unified Studio, untuk mengakses data, analitik, dan kemampuan AI.</p>	Amazon SageMaker
	<p>Analitik interaktif</p> <p>Dioptimalkan untuk melakukan analisis dan eksplorasi data real-time, yang memungkinkan pengguna untuk melakukan kueri dan memvisualisasikan data secara interaktif.</p>	Amazon Athena
	<p>Pemrosesan data besar</p> <p>Dioptimalkan untuk memproses, memindahkan, dan mengubah data dalam jumlah besar.</p>	Amazon EMR
	<p>Katalog data</p> <p>Dioptimalkan untuk memberikan informasi rinci</p>	AWS Glue

Kategori	Untuk apa itu dioptimalkan?	Layanan
	<p>tentang data yang tersedia, struktur, karakteristik, dan hubungannya.</p> <p>Orkestrasi alur kerja</p> <p>Dioptimalkan untuk membuat, menjadwalkan, dan memantau alur kerja data menggunakan Apache Airflow untuk mengoordinasikan proses analitik dan pekerjaan ETL.</p>	<p>Amazon MWAA</p>
Streaming data	<p>Apache Kafka memproses data streaming</p> <p>Dioptimalkan untuk menggunakan operasi pesawat data Apache Kafka dan menjalankan versi open source Apache Kafka.</p> <p>Pemrosesan waktu nyata</p> <p>Dioptimalkan untuk pengambilan dan agregasi data yang cepat dan berkelanjutan, termasuk data log infrastruktur TI, log aplikasi, media sosial, umpan data pasar, dan data clickstream web.</p>	<p>Amazon MSK</p> <p>Amazon Kinesis Data Streams</p>

Kategori	Untuk apa itu dioptimalkan?	Layanan
	<p>Pengiriman data streaming waktu nyata</p> <p>Dioptimalkan untuk mengirimkan data streaming real-time ke tujuan seperti Amazon S3, Amazon Redshift, Service, Splunk OpenSearch , Apache Iceberg Tables, dan titik akhir HTTP kustom atau titik akhir HTTP yang dimiliki oleh penyedia layanan pihak ketiga yang didukung.</p>	<p>Amazon Data Firehose</p>
	<p>Membangun aplikasi Apache Flink</p> <p>Dioptimalkan untuk menggunakan Java, Scala, Python, atau SQL untuk memproses dan menganalisis data streaming.</p>	<p>Layanan Dikelola Amazon untuk Apache Flink</p>
Kecerdasan bisnis	<p>Dasbor dan visualisasi</p> <p>Dioptimalkan untuk mewakili kumpulan data yang kompleks secara visual, dan menyediakan kueri bahasa alami dari data Anda.</p>	<p>Quick</p>
Analisis pencarian	<p>OpenSearch Cluster terkelola</p> <p>Dioptimalkan untuk analisis log, pemantauan aplikasi real-time, dan analisis clickstream.</p>	<p>OpenSearch Layanan Amazon</p>

Kategori	Untuk apa itu dioptimalkan?	Layanan
Tata kelola data	<p>Mengelola akses data</p> <p>Dioptimalkan untuk mengatur manajemen yang tepat, ketersediaan, kegunaan, integritas, dan keamanan data sepanjang siklus hidupnya.</p>	<p>Amazon DataZone</p>
Kolaborasi data	<p>Mengamankan data kamar bersih</p> <p>Dioptimalkan untuk berkolaborasi dengan perusahaan lain tanpa berbagi data dasar mentah.</p>	<p>AWS Clean Rooms</p>
Danau data dan gudang	<p>Akses terpadu ke danau data dan gudang data</p> <p>Dibangun di atas arsitektur lakehouse untuk mengoptimalkan untuk menyatukan akses data di seluruh danau data Amazon S3, gudang data Amazon Redshift, database operasional, dan sumber data pihak ketiga dan federasi.</p>	<p>Amazon SageMaker</p>
	<p>Penyimpanan objek untuk danau data</p> <p>Dioptimalkan untuk menyediakan fondasi data lake dengan skalabilitas yang hampir tidak terbatas dan daya tahan tinggi.</p>	<p>Amazon S3</p>

Kategori	Untuk apa itu dioptimalkan?	Layanan
	<p>Pergudangan data</p> <p>Dioptimalkan untuk menyimpan, mengatur, dan mengambil volume besar data terstruktur dan terkadang semi-terstruktur dari berbagai sumber dalam suatu organisasi.</p>	Amazon Redshift

Gunakan layanan AWS analitik

Anda sekarang harus memiliki pemahaman yang jelas tentang tujuan bisnis Anda, dan volume dan kecepatan data yang akan Anda konsumsi dan analisis untuk mulai membangun jalur data Anda.

Untuk mengeksplorasi cara menggunakan dan mempelajari lebih lanjut tentang masing-masing layanan yang tersedia—kami telah menyediakan jalur untuk mengeksplorasi cara kerja masing-masing layanan. Bagian berikut menyediakan tautan ke dokumentasi mendalam, tutorial langsung, dan sumber daya untuk membantu Anda memulai dari penggunaan dasar hingga penyelaman mendalam yang lebih canggih.

Amazon Athena

- Memulai dengan Amazon Athena

Pelajari cara menggunakan Amazon Athena untuk menanyakan data dan membuat tabel berdasarkan data sampel yang disimpan di Amazon S3, menanyakan tabel, dan memeriksa hasil kueri.

[Memulai dengan tutorial](#)

- Mulai dengan Apache Spark di Athena

Gunakan pengalaman notebook yang disederhanakan di konsol Athena untuk mengembangkan aplikasi Apache Spark menggunakan notebook Python atau Athena. APIs

[Memulai dengan tutorial](#)

- Katalog dan atur pertanyaan federasi Athena dengan arsitektur lakehouse Amazon SageMaker

Pelajari cara menyambungkan, mengatur, dan menjalankan kueri gabungan pada data yang disimpan di Amazon Redshift, DynamoDB, dan Snowflake melalui data lakehouse di Amazon SageMaker

[Baca blognya](#)

- Menganalisis data di Amazon S3 menggunakan Athena

Jelajahi cara menggunakan Athena pada log dari Elastic Load Balancers, yang dihasilkan sebagai file teks dalam format yang telah ditentukan sebelumnya. Kami menunjukkan cara membuat tabel, mempartisi data dalam format yang digunakan oleh Athena, mengubahnya menjadi Parquet, dan membandingkan kinerja kueri.

[Baca posting blog](#)

AWS Clean Rooms

- Menyiapkan AWS Clean Rooms

Pelajari cara mengatur AWS Clean Rooms di AWS account Anda.

[Baca panduannya](#)

- Buka wawasan data di seluruh kumpulan data multi-pihak menggunakan Resolusi AWS Entitas tanpa membagikan data yang mendasarinya AWS Clean Rooms

Pelajari cara menggunakan persiapan dan pencocokan untuk membantu meningkatkan pencocokan data dengan kolaborator.

[Baca posting blog](#)

- Bagaimana privasi diferensial membantu membuka wawasan tanpa mengungkapkan data di tingkat individu

Pelajari cara Privasi AWS Clean Rooms Diferensial menyederhanakan penerapan privasi diferensial dan membantu melindungi privasi pengguna Anda.

[Baca blognya](#)

Amazon Data Firehose

- Tutorial: Membuat aliran Firehose dari konsol

Pelajari cara menggunakan Konsol Manajemen AWS atau AWS SDK untuk membuat aliran Firehose ke tujuan yang Anda pilih.

[Baca panduannya](#)

- Mengirim data ke aliran Firehose

Pelajari cara menggunakan sumber data yang berbeda untuk mengirim data ke aliran Firehose Anda.

[Baca panduannya](#)

- Mengubah data sumber di Firehose

Pelajari cara menjalankan fungsi Lambda Anda untuk mengubah data sumber yang masuk dan mengirimkan data yang diubah ke tujuan.

[Baca panduannya](#)

Amazon DataZone

- Memulai dengan Amazon DataZone

Pelajari cara membuat domain DataZone root Amazon, mendapatkan URL portal data, menelusuri DataZone alur kerja dasar Amazon untuk produsen data dan konsumen data.

[Memulai dengan tutorial](#)

- Mengumumkan ketersediaan umum garis keturunan data di generasi Amazon dan Amazon berikutnya SageMaker DataZone

Pelajari cara Amazon DataZone menggunakan penangkapan garis keturunan otomatis untuk fokus mengumpulkan dan memetakan informasi silsilah secara otomatis dari dan Amazon Redshift. AWS Glue

[Baca blognya](#)

Amazon EMR

- Memulai dengan Amazon EMR

Pelajari cara meluncurkan cluster sampel menggunakan Spark, dan cara menjalankan PySpark skrip sederhana yang disimpan di bucket Amazon S3.

[Memulai dengan tutorial](#)

- Memulai dengan Amazon EMR di Amazon EKS

Kami menunjukkan kepada Anda cara memulai menggunakan Amazon EMR di Amazon EKS dengan menerapkan aplikasi Spark pada cluster virtual.

[Jelajahi panduannya](#)

- Memulai dengan EMR Tanpa Server

Jelajahi bagaimana Amazon EMR Serverless menyediakan lingkungan runtime tanpa server yang menyederhanakan pengoperasian aplikasi analitik yang menggunakan kerangka kerja open source terbaru.

[Memulai dengan tutorial](#)

AWS Glue

- Memulai dengan AWS Glue DataBrew

Pelajari cara membuat DataBrew proyek pertama Anda. Anda memuat kumpulan data sampel, menjalankan transformasi pada kumpulan data tersebut, membuat resep untuk menangkap transformasi tersebut, dan menjalankan pekerjaan untuk menulis data yang diubah ke Amazon S3.

[Memulai dengan tutorial](#)

- Mengubah data dengan AWS Glue DataBrew

Pelajari tentang AWS Glue DataBrew, alat persiapan data visual yang memudahkan analis data dan ilmuwan data untuk membersihkan dan menormalkan data guna mempersiapkannya untuk analitik dan pembelajaran mesin. Pelajari cara membuat proses ETL menggunakan AWS Glue DataBrew

[Memulai dengan lab](#)

- [AWS Glue DataBrew hari perendaman](#)

Jelajahi cara menggunakan AWS Glue DataBrew untuk membersihkan dan menormalkan data untuk analitik dan pembelajaran mesin.

[Memulai dengan Workshop](#)

- [Memulai dengan AWS Glue Data Catalog](#)

Pelajari cara membuat yang pertama AWS Glue Data Catalog, yang menggunakan bucket Amazon S3 sebagai sumber data Anda.

[Memulai dengan tutorial](#)

- [Katalog data dan crawler di AWS Glue](#)

Temukan bagaimana Anda dapat menggunakan informasi dalam Katalog Data untuk membuat dan memantau pekerjaan ETL Anda.

[Jelajahi panduannya](#)

Amazon Kinesis Data Streams

- [Memulai tutorial untuk Amazon Kinesis Data Streams](#)

Pelajari cara memproses dan menganalisis data stok waktu nyata.

[Memulai dengan tutorial](#)

- [Pola arsitektur untuk analitik real-time menggunakan Amazon Kinesis Data Streams, bagian 1](#)

Pelajari tentang pola arsitektur umum dari dua kasus penggunaan: analisis data deret waktu dan layanan mikro berbasis peristiwa.

[Baca blognya](#)

- [Pola Arsitektur untuk analitik real-time menggunakan Amazon Kinesis Data Streams, bagian 2](#)

Pelajari tentang aplikasi AI dengan Kinesis Data Streams dalam tiga skenario: intelijen bisnis generatif real-time, sistem rekomendasi real-time, dan streaming dan inferensi data Internet of Things.

[Baca blognya](#)

Amazon Managed Service for Apache Flink

- Apa itu Amazon Managed Service untuk Apache Flink?

Memahami konsep dasar Amazon Managed Service untuk Apache Flink.

[Jelajahi panduannya](#)

- Layanan Dikelola Amazon untuk Workshop Apache Flink

Dalam lokakarya ini, Anda akan mempelajari cara menyebarkan, mengoperasikan, dan menskalakan aplikasi Flink dengan Amazon Managed Service untuk Apache Flink.

[Hadiri lokakarya virtual](#)

Amazon MSK

- Memulai dengan Amazon MSK

Pelajari cara membuat kluster MSK Amazon, memproduksi dan mengonsumsi data, serta memantau kesehatan kluster Anda menggunakan metrik.

[Memulai dengan panduan](#)

- Lokakarya MSK Amazon

Ikuti lebih dalam lokakarya MSK Amazon langsung ini.

[Memulai dengan Workshop](#)

Amazon MWAA

- Memulai dengan Amazon MWAA

Pelajari cara membuat lingkungan MWAA pertama Anda, mengunggah DAG ke Amazon S3, dan menjalankan alur kerja pertama Anda.

[Memulai dengan tutorial](#)

- Membangun jaringan data dengan Amazon MWAA

Pelajari cara membuat pipeline end-to-end data yang mengatur AWS layanan analitik lain seperti Glue, EMR, dan Redshift. [Posting blog ini mengeksplorasi pendekatan yang](#)

disederhanakan dan digerakkan oleh konfigurasi untuk mengatur pekerjaan inti dbt menggunakan MWAA dan Cosmos, dengan pekerjaan yang menjalankan transformasi di Amazon Redshift.

[Baca posting blog](#)

- Lokakarya Amazon MWAA

Jelajahi latihan langsung untuk mempelajari cara menerapkan, mengonfigurasi, dan menggunakan Amazon MWAA untuk orkestrasi alur kerja data.

[Memulai dengan Workshop](#)

- Praktik terbaik untuk Amazon MWAA

Pelajari pola arsitektur dan praktik terbaik untuk menggunakan Amazon MWAA dalam alur kerja analitik Anda.

[Baca panduannya](#)

OpenSearch Service

- Memulai dengan OpenSearch Layanan

Pelajari cara menggunakan OpenSearch Layanan Amazon untuk membuat dan mengonfigurasi domain pengujian.

[Memulai dengan tutorial](#)

- Memvisualisasikan panggilan dukungan pelanggan dengan OpenSearch Layanan dan Dasbor OpenSearch

Temukan panduan lengkap tentang situasi berikut: bisnis menerima sejumlah panggilan dukungan pelanggan dan ingin menganalisisnya. Apa subjek dari setiap panggilan? Berapa banyak yang positif? Berapa banyak yang negatif? Bagaimana manajer dapat mencari atau meninjau transkrip panggilan ini?

[Memulai dengan tutorial](#)

- Memulai dengan lokakarya Amazon OpenSearch Tanpa Server

Pelajari cara menyiapkan domain Amazon OpenSearch Tanpa Server baru di konsol. AWS Jelajahi berbagai jenis kueri penelusuran yang tersedia, dan rancang visualisasi yang menarik,

dan pelajari cara mengamankan domain dan dokumen berdasarkan hak istimewa pengguna yang ditetapkan.

[Memulai dengan Workshop](#)

- Basis Data Vektor yang Dioptimalkan Biaya: Pengantar teknik kuantisasi OpenSearch Layanan Amazon

Pelajari bagaimana OpenSearch Layanan mendukung teknik kuantisasi skalar dan produk untuk mengoptimalkan penggunaan memori dan mengurangi biaya operasional.

[Baca posting blog](#)

Quick

- Memulai dengan analisis data cepat

Pelajari cara membuat analisis pertama Anda. Gunakan data sampel untuk membuat analisis sederhana atau lebih maju. Atau Anda dapat terhubung ke data Anda sendiri untuk membuat analisis.

[Jelajahi panduannya](#)

- Visualisasi dengan Cepat

Temukan sisi teknis intelijen bisnis (BI) dan visualisasi data dengan AWS. Pelajari cara menyematkan dasbor ke aplikasi dan situs web, serta mengelola akses dan izin dengan aman.

[Memulai dengan kursus](#)

- Lokakarya cepat

Mulailah perjalanan cepat Anda dengan lokakarya

[Memulai dengan lokakarya](#)

Amazon Redshift

- Memulai dengan Amazon Redshift Tanpa Server

Pahami alur dasar Amazon Redshift Serverless untuk membuat sumber daya tanpa server, terhubung ke Amazon Redshift Tanpa Server, memuat data sampel, dan kemudian menjalankan kueri pada data.

[Jelajahi panduannya](#)

- Lokakarya menyelam dalam Amazon Redshift

Jelajahi serangkaian latihan yang membantu pengguna mulai menggunakan platform Amazon Redshift.

[Memulai dengan Workshop](#)

Amazon S3

- Memulai dengan Amazon S3

Pelajari cara membuat DataBrew proyek pertama Anda. Anda memuat kumpulan data sampel, menjalankan transformasi pada kumpulan data tersebut, membuat resep untuk menangkap transformasi tersebut, dan menjalankan pekerjaan untuk menulis data yang diubah ke Amazon S3.

[Memulai dengan panduan](#)

Amazon SageMaker

- Memulai dengan SageMaker

Pelajari cara membuat proyek, menambahkan anggota, dan menggunakan contoh JupyterLab buku catatan untuk mulai membangun.

[Baca panduannya](#)

- Memperkenalkan Amazon generasi berikutnya SageMaker: Pusat untuk semua data, analitik, dan AI Anda

Pelajari cara memulai pemrosesan data, pengembangan model, dan pengembangan aplikasi AI generatif.

[Baca blognya](#)

- Apa itu SageMaker Unified Studio?

Pelajari tentang kemampuan SageMaker Unified Studio dan cara mengaksesnya saat Anda menggunakan Amazon SageMaker.

[Baca panduannya](#)

- Memulai dengan arsitektur rumah danau Amazon SageMaker

Pelajari cara membuat proyek dan menelusuri, mengunggah, dan menanyakan data untuk kasus penggunaan bisnis Anda di Amazon SageMaker.

[Baca panduannya](#)

- Koneksi data dalam arsitektur lakehouse Amazon SageMaker

Pelajari bagaimana arsitektur lakehouse menyediakan pendekatan terpadu untuk mengelola koneksi data di seluruh AWS layanan dan aplikasi perusahaan.

[Baca panduannya](#)

- Katalog dan atur pertanyaan federasi Athena dengan arsitektur rumah danau SageMaker

Pelajari cara menyambungkan, mengatur, dan menjalankan kueri gabungan pada data yang disimpan di Amazon Redshift, DynamoDB, dan Snowflake untuk proyek Amazon Anda. SageMaker

[Baca blognya](#)

Jelajahi cara menggunakan layanan AWS analitik

Editable architecture diagrams

Diagram arsitektur referensi

Jelajahi diagram arsitektur untuk membantu Anda mengembangkan, menskalakan, dan menguji solusi analitik Anda. AWS

[Jelajahi arsitektur referensi analitik](#)

Ready-to-use code

Solusi unggulan

AWS Solusi

Analisis yang Dapat Diskalakan Menggunakan Apache Druid di AWS

Terapkan kode AWS yang dibangun untuk membantu Anda mengatur, mengoperasikan, dan mengelola Apache Druid, lingkungan hosting yang hemat biaya AWS, sangat tersedia, tangguh, dan toleran terhadap kesalahan.

[Jelajahi solusi ini](#)

Jelajahi solusi yang telah dikonfigurasi sebelumnya dan dapat diterapkan serta panduan implementasinya, yang dibuat oleh AWS

[Jelajahi semua AWS solusi keamanan, identitas, dan tata kelola](#)

Documentation

Whitepaper Analytics

Jelajahi whitepaper untuk wawasan lebih lanjut dan praktik terbaik dalam memilih, menerapkan, dan menggunakan layanan analitik yang paling sesuai dengan organisasi Anda.

[Jelajahi whitepaper analitik](#)

AWS Blog Data Besar

Jelajahi posting blog yang membahas kasus penggunaan data besar tertentu.

[Jelajahi blog AWS Big Data](#)

Riwayat dokumen

Tabel berikut menjelaskan perubahan penting pada panduan keputusan ini. Untuk pemberitahuan tentang pembaruan panduan ini, Anda dapat berlangganan umpan RSS.

Perubahan	Deskripsi	Tanggal
RE: menciptakan pembaruan	Tautan yang diperbarui di seluruh panduan keputusan dan menambahkan Alur Kerja Terkelola Amazon untuk Apache Airflow.	September 24, 2025
RE: menciptakan pembaruan	Referensi yang diperbarui ke Amazon SageMaker, Amazon SageMaker Unified Studio (tidak lagi pratinjau), dan Amazon SageMaker Lakehouse di seluruh panduan keputusan.	September 9, 2025
RE: menciptakan pembaruan	Menambahkan SageMaker AI Unified Studio dan AWS Clean Rooms. Dokumen yang diperbarui secara keseluruhan dengan fitur dan kemampuan AI baru.	Februari 20, 2025
Publikasi awal	Panduan pertama kali diterbitkan.	17 November 2023

Terjemahan disediakan oleh mesin penerjemah. Jika konten terjemahan yang diberikan bertentangan dengan versi bahasa Inggris aslinya, utamakan versi bahasa Inggris.