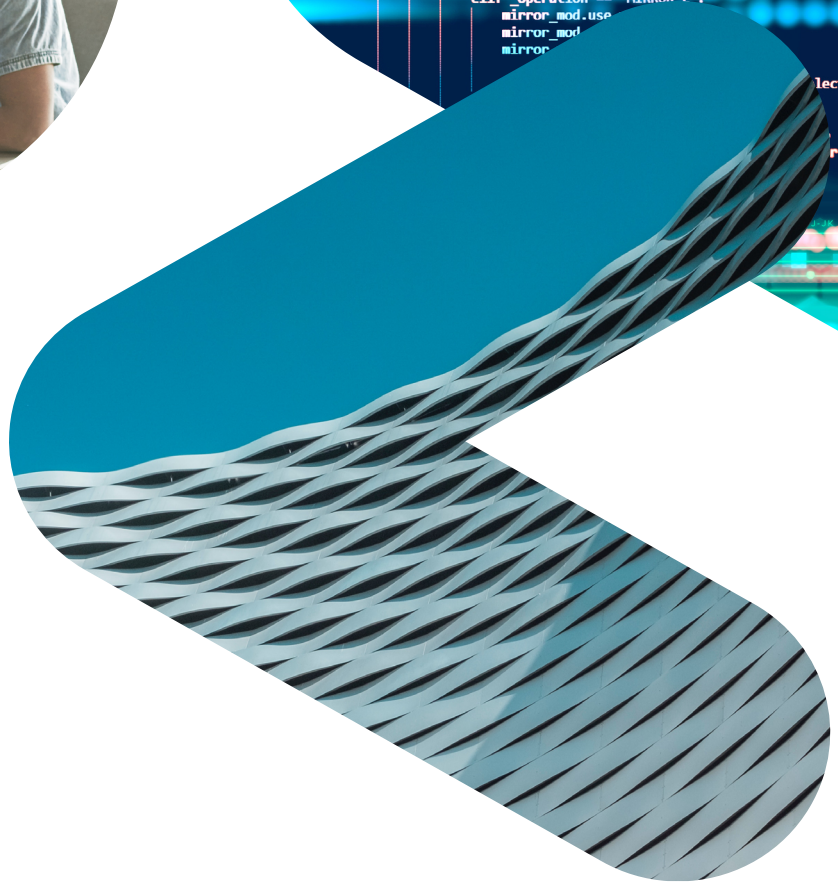


snowflake®

THE MODERN DATA STREAMING PIPELINE

Top analytical streaming reference architectures and use cases across 7 industries

TABLE OF CONTENTS

THE STATE OF DATA STREAMING

Today's fast-paced world demands timely insights and decisions, which is driving the importance of streaming data. While traditional batch processing systems store and process data in groups, streaming data refers to data that is continuously generated from a variety of sources.

Data streaming — the process of continuously capturing and processing data at a certain speed — can encompass a broad range of latency depending on industry, business and use case needs. It may be evaluated in terms of seconds, minutes or even hours. The sources of this data are proliferating, thanks to the surge in connected devices. Across various industries, streaming data and analytics play a crucial role in making better-informed decisions, leading to better outcomes, faster.

Meanwhile, enterprises are collecting more data than ever before. And the volume, variety and velocity of data means that most of that data originates and is stored in the cloud. Petabytes of data are available, but that data goes stale in storage solutions outside traditional data warehouses because these legacy systems can't efficiently handle new, semi-structured types of data. Or, more data is just too expensive to store in traditional systems that weren't built for the cloud. These systems, built when far less data was collected and far fewer users needed data access, still struggle under their legacy architecture and technology.

And even though Hadoop and other NoSQL systems have made it possible to ingest varying data types, getting the best analytics from that data is nearly impossible. These systems have left enterprises without the right foundation to turn their growing hunger for data into real business insight.

In this ebook, we examine the challenges enterprises have faced turning data into transformative business insights. You'll also learn about the technologies that are now available to enterprises to keep up with the analytics revolution. With a modern data platform, you and your organization can be champions.

For example:

- **Manufacturers** use streaming sensor readings from IoT devices to optimize maintenance or the supply chain.
- **Retailers** use streaming website analytics to personalize the customer experience.
- **Media companies** use streaming viewer statistics to plan and create more engaging content.

However, the process of data streaming presents challenges. For data to be valuable to the business, it needs to be ingested from diverse sources with low latency. This can be costly due to the complexity and resource requirements involved in analyzing large volumes of disparate data in near real time. Organizations must strike a balance between latency and cost while navigating this tension.

Leaders across industries are addressing these challenges and preparing for future business requirements with a simplified solution that combines streaming and batch processing pipelines in a single architecture.

Snowflake engaged with customers across diverse sectors to explore their most common streaming use cases, and examine their architecture choices for optimizing performance and efficiency. This ebook highlights how leading organizations leverage Snowflake and collaborate with partners such as Amazon Web Services (AWS) to centralize their data in a single platform — achieving streamlined, cost-effective streaming success.

STREAMING PATTERNS AND ARCHITECTURES ACROSS INDUSTRIES

The concepts of "real time" and "streaming" can vary significantly based on the industry context. While real-time data emphasizes minimizing latency to receive information almost instantly, streaming refers to the continuous ingestion and processing of data as it arrives from various sources. Not all streaming use cases require real-time speed to provide value. For example, retailers who stream data and get an updated view of inventory every 10 minutes can effectively make informed decisions about restocking, distribution and order fulfillment.

Understanding this distinction helps businesses tailor solutions to specific industry needs and better manage the trade-off between minimal delays and high costs. For many use cases, data streaming is a continuum that balances latency and cost in a way that yields optimized ROI, helping your streaming architecture deliver business value.

It's important to note that in this ebook, the focus is on analytical use cases that utilize streaming data to uncover trends and generate decision-making insights — not on transactional use cases that rely on real-time data for immediate operational adjustments, such as instantly identifying and addressing equipment failures.

Snowflake, with our partners and ecosystem, has helped hundreds of customers implement these solutions across industries. Taking an industry lens, this ebook provides recommended reference architecture based on our observations of these high-demand use cases. Let's explore these analytical use cases in the following pages.

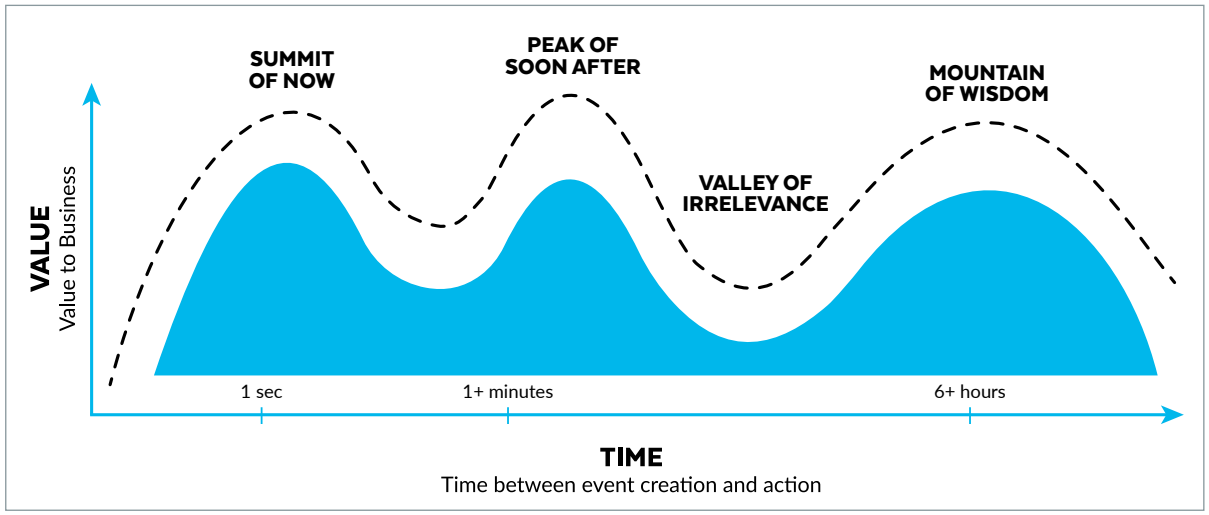


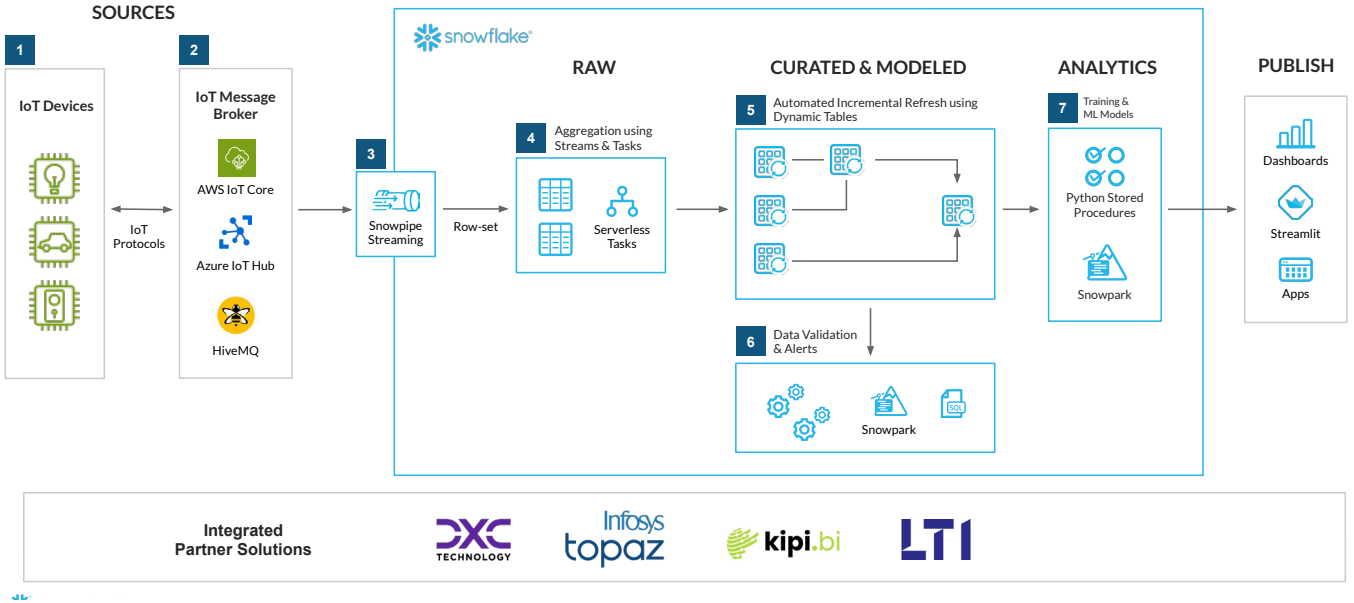
FIGURE 1: FIGURE ≠ INSTANTANEOUS

MANUFACTURING

Data streaming helps manufacturing companies ingest critical data from across the value chain, such as sensor readings from production equipment, inventory levels, supplier performance metrics and customer demand patterns. This continuous flow of data enables timely decision-making, process optimization and proactive quality control.

IoT analytics overview

IoT data opens new opportunities for manufacturers, helping them generate new revenue streams with user behavior data, reduce operating costs by detecting potential issues and improve product quality through equipment performance data. Streaming high volumes of IoT data at low latency helps accelerate these decisions for faster outcomes.

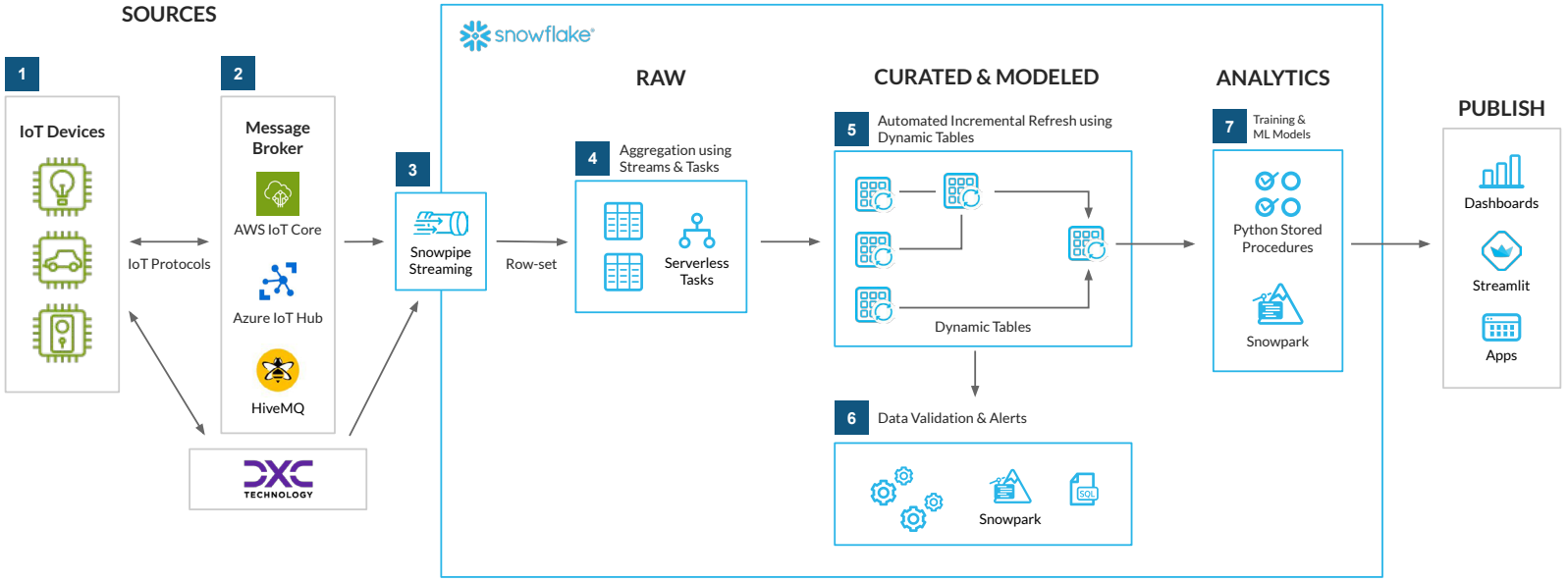


1. Smart devices, sensors and other IoT devices generate continuous and streaming data.
2. Due to frequently unreliable internet connectivity, IoT devices communicate using IoT protocols such as MQTT and an IoT message broker. The message broker uses a publish-and-subscribe mechanism to interact with other services, which subscribe to specific topics within the broker to access device data.
3. A streaming service can be used to ingest and buffer real-time device data together with Snowpipe Streaming for row-set data to ensure reliable ingestion and delivery to a staging table in Snowflake.
4. For strong orchestration needs, Snowflake's Streams & Tasks features automate the workflows required to aggregate incoming data. You can optionally use Dynamic Tables (in public preview) for aggregating and materializing the aggregated results.
5. Use Dynamic Tables to automate incremental processing for continuous data transformation.
6. Snowpark can be used to further enrich and validate data.
7. Data then is used for transformation with business logic and/or machine learning training with Snowpark.

Partners such as Infosys, LTI, DXC and Kipi offer solutions integrated with Snowflake's streaming capabilities. Visit page 24 for more details.

Connected products

Manufacturing connected products — internet-connected devices like IoT sensors on industrial machinery, smart consumer goods, and connected vehicles and fleets — generate petabytes of data on a daily basis. To generate value out of this information, manufacturers need a streaming architecture that continuously aggregates and summarizes data from multiple sources in near real time. With timely connected product data, manufacturers can generate new lines of business, optimize operations, increase production and enhance the customer experience.

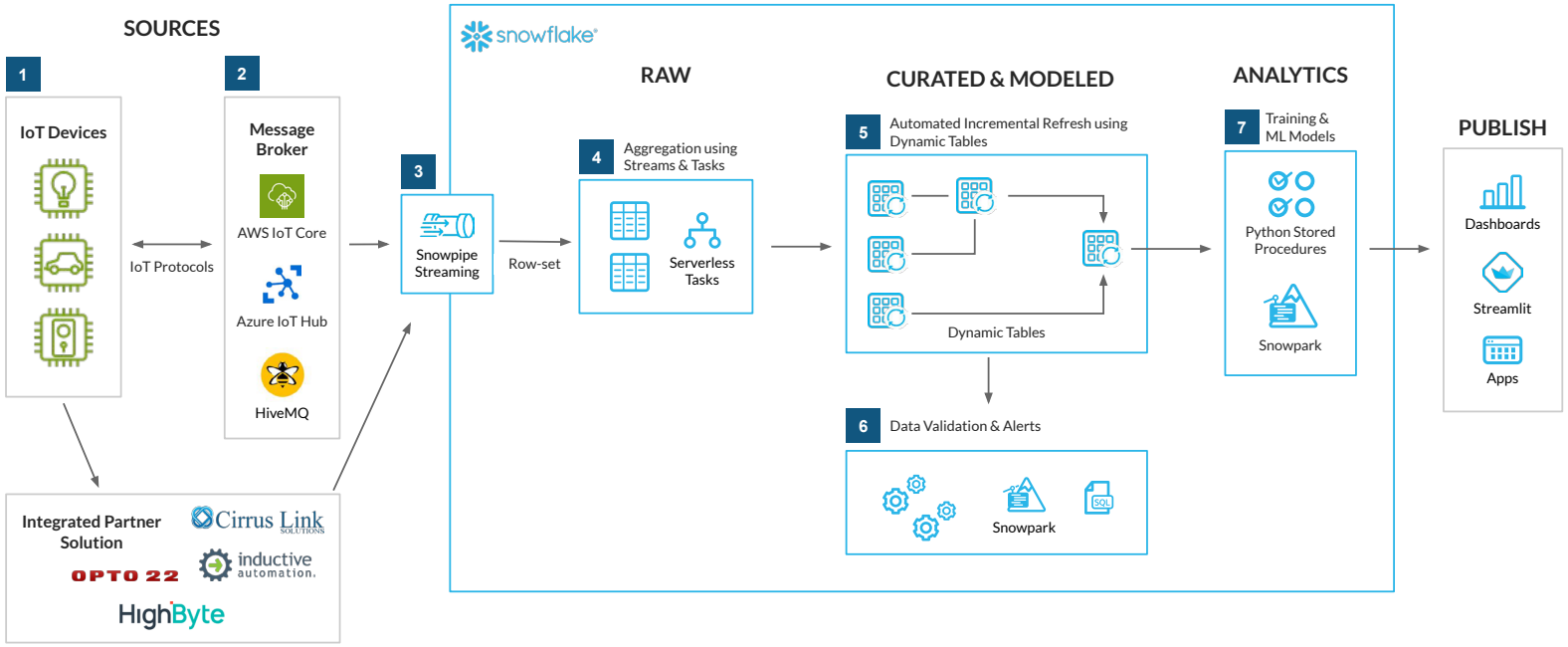


1. Smart devices, sensors and other IoT devices generate continuous and streaming data.
2. Due to frequently unreliable internet connectivity, IoT devices communicate using IoT protocols such as MQTT and an IoT message broker. The message broker uses a publish-and-subscribe mechanism to interact with other services, which subscribe to specific topics within the broker to access device data.
3. A streaming service can be used to ingest and buffer real-time device data together with Snowpipe Streaming for row-set data to ensure reliable ingestion and delivery to a staging table in Snowflake.
4. For strong orchestration needs, Snowflake's Streams & Tasks features automate the workflows required to aggregate incoming data. You can optionally use Dynamic Tables for aggregating and materializing the aggregated results.
5. Use Dynamic Tables to automate incremental processing for continuous data transformation.
6. Snowpark can be used to further enrich and validate data.
7. Data then is used for transformation with business logic and/or machine learning training with Snowpark.

Partners such as DXC offer solutions integrated with Snowflake's streaming capabilities. Visit page 24 for more details.

Shop floor OT optimization

Operational Technology (OT) data provides insights into factory operations, product quality and machine performance. By streaming near real-time OT data and converging it with IT data, manufacturers can identify operational inefficiencies and fix issues before they turn into critical problems like machine failure, helping manufacturers save costs.



1. Smart devices, sensors and other IoT devices generate continuous and streaming data.
2. Due to frequently unreliable internet connectivity, IoT devices communicate using IoT protocols such as MQTT and an IoT message broker. The message broker uses a publish and subscribe mechanism to interact with other services, which subscribe to specific topics within the broker to access device data.
3. A streaming service can be used to ingest and buffer real-time device data together with Snowpipe Streaming for row-set data to ensure reliable ingestion and delivery to a staging table in Snowflake.
4. For strong orchestration needs, Snowflake's Streams & Tasks features automate the workflows required to aggregate incoming data. You can optionally use Dynamic Tables for aggregating and materializes the aggregated results.
5. Use Dynamic Tables to automate incremental processing for continuous data transformation.
6. Snowpark can be used to further enrich and validate data.
7. Data then is used for transformation with business logic and/or machine learning training with Snowpark.

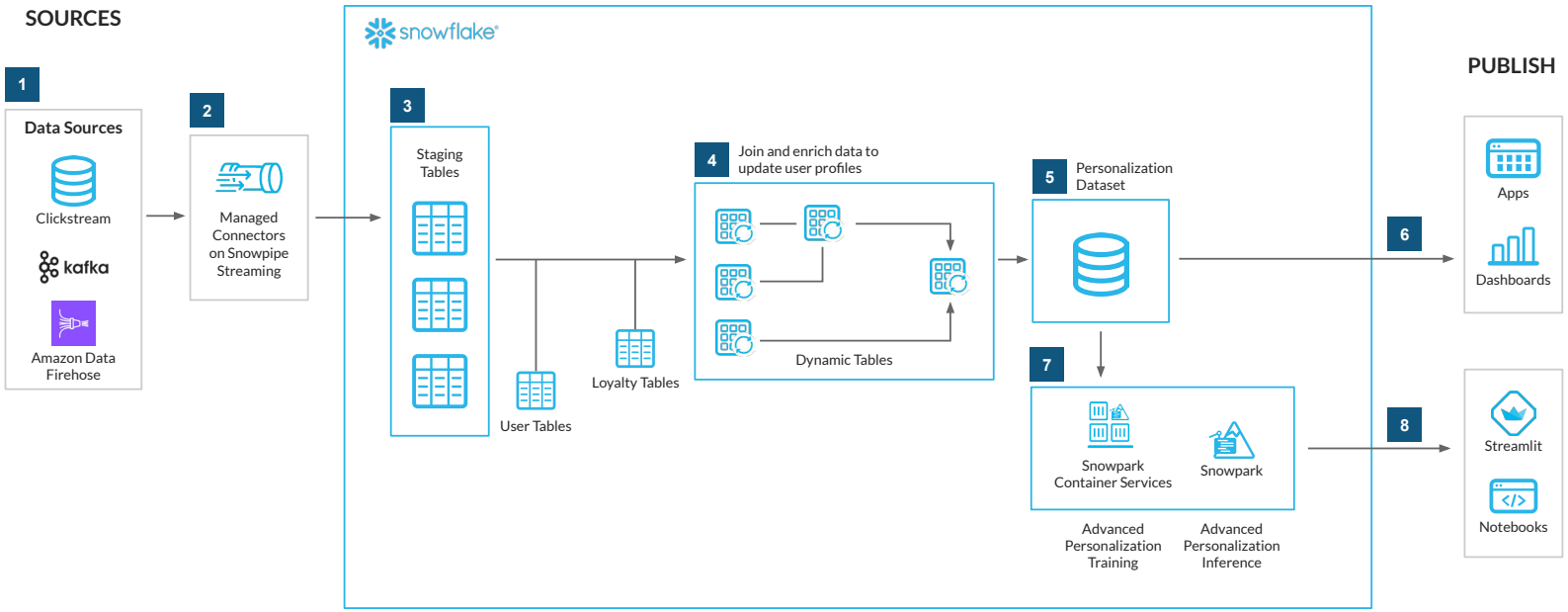
Partners such as Cirrus Link, Opto22, Inductive Automation and HighByte offer solutions integrated with Snowflake's streaming capabilities. Visit page 24 for more details.

RETAIL AND CONSUMER PACKAGED GOODS

Data streaming in retail involves ingesting and processing data from online transactions, inventory systems, customer interactions, point-of-sale terminals, website analytics and other sources. With this data, retailers can make timely decisions, personalize marketing and improve efficiency in their supply chain.

Advanced personalization

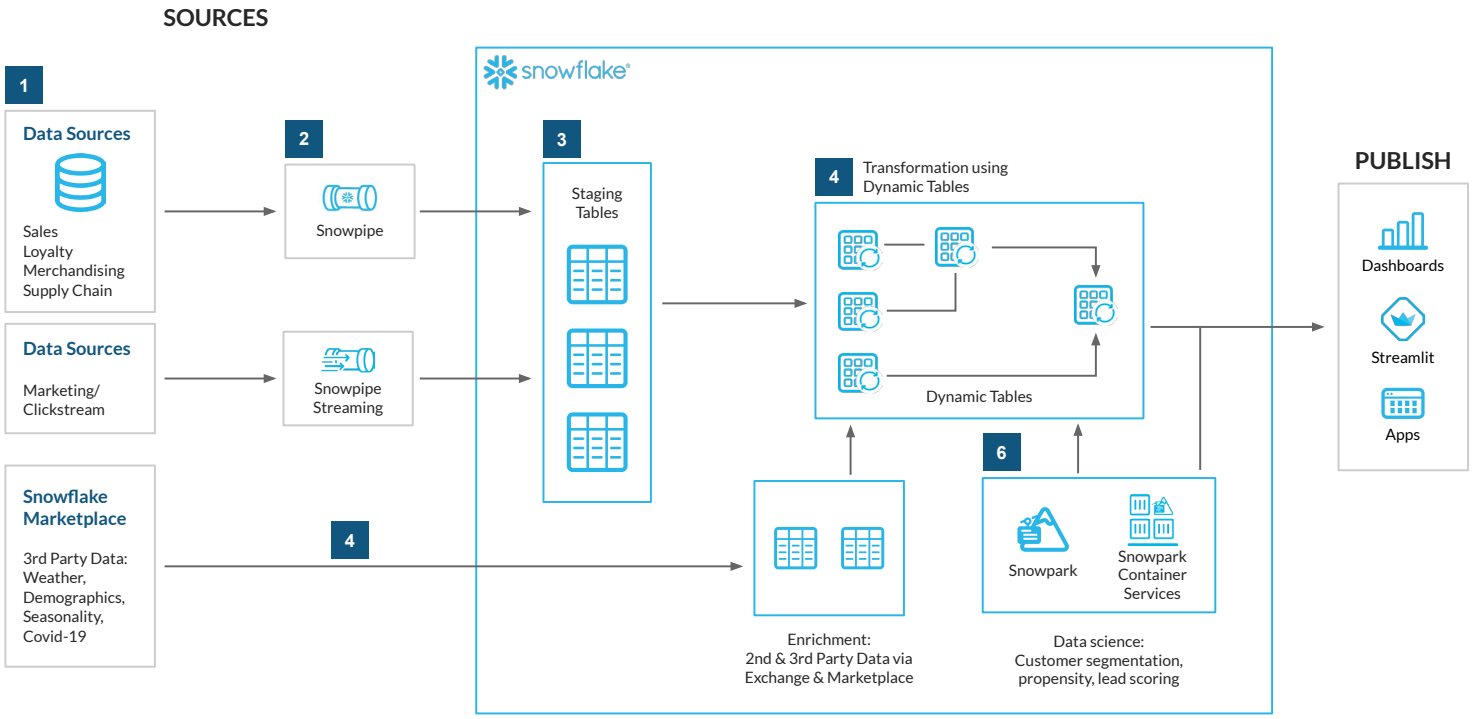
Helping boost sales and customer satisfaction, a tailored shopping experience is critical to business. To build an effective personalization strategy, retailers need near real-time customer behavior data to inform quick decisions around targeted promotions. The need to run multiple reports and optimize personalized customer experience at scale requires data engineering teams to have near real-time data ready for faster insights.



1. Clickstream data from websites and apps is written to Kafka topics or Amazon Data Firehose.
2. Ingest clickstream data as row-sets using managed connectors on Snowflake — Kafka and Amazon Data Firehose connector powered by Snowpipe Streaming.
3. Data lands in Snowflake in staging tables with low latency (~1 second).
4. Aggregate, normalize and enrich click events, by joining against user tables, loyalty tables etc. using a DAG of Dynamic Tables.
5. Consumable personalization dataset for presentation layer can be Dynamic Table itself.
6. Power the personalization experience on websites and apps using the final personalization dataset.
7. Train final advanced personalization model using Snowpark Container Services. Serves advanced personalization model as a Snowflake UDF using the ONNX runtime to accelerate inference.
8. Data scientist can interact with notebook and view application via Streamlit.

End-to-end inventory management

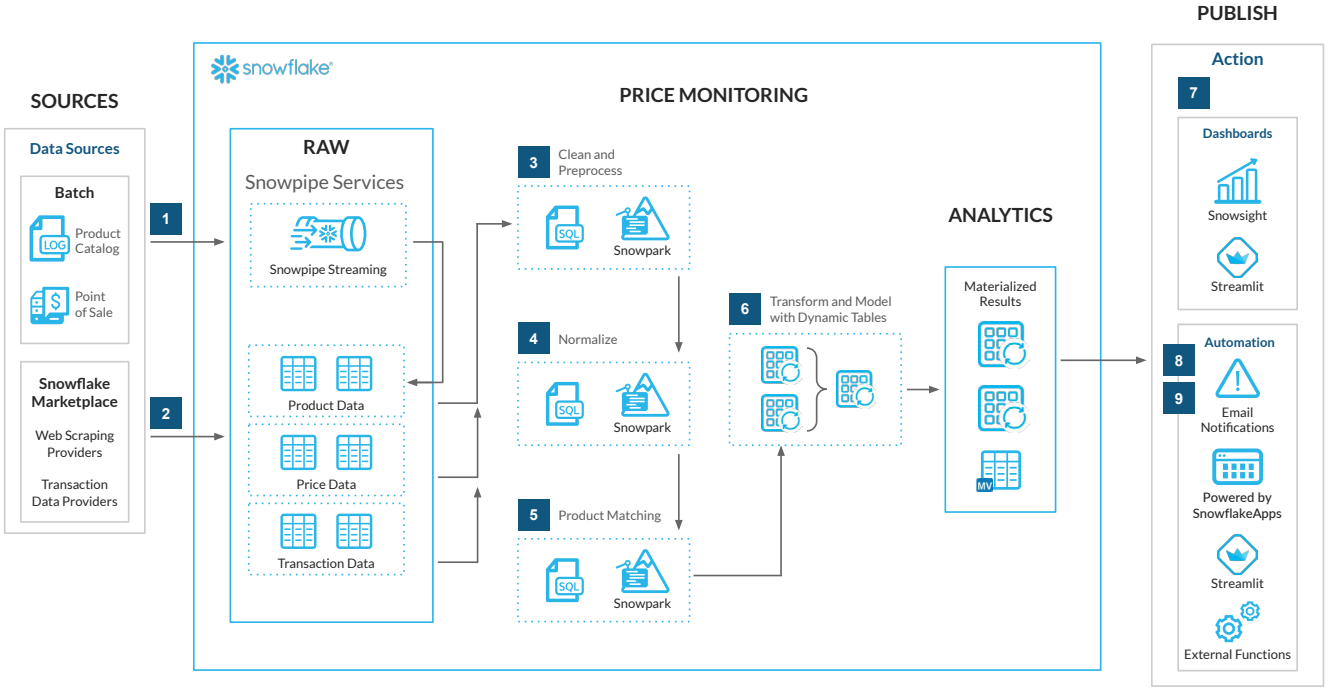
On-Time, In-Full (OTIF) — which measures if the shipment arrived on time with products that match the purchase order — is top of mind for all retailers. Optimizing OTIF is important for both consumer goods companies and retailers to avoid unexpected penalty fees. With near real-time insight into SKU-level inventory, fulfillment, channels, and customer and product data, retailers can forecast demand more accurately and provide customers with the most up-to-date inventory information for a smoother shopping experience.



1. Load sales, loyalty, supply chain data.
2. Snowpipe enables file ingestion and Snowpipe Streaming for row-set ingestion.
3. Data lands in staging tables in Snowflake.
4. Get or purchase third-party data from Snowflake Marketplace.
5. Transform and model the data ensure it is structured in a consistent format using Dynamic Tables. Dynamic Tables can join staging tables with 2nd and 3rd party data via marketplace to create enriched datasets.
6. Use Snowpark and Snowpark Container Services for AI/ML processing: customer segmentation, recommendations, propensity, lead scoring, etc.
7. Create native apps, Streamlit apps and connect to your favorite dashboards.

Retail price monitoring

To monitor and optimize pricing strategies, retailers need near real-time insight into the product and price a customer is looking for. By streaming live data, such as product detail, price and quantity, across different channels and competitors, retailers can act faster, experiment and identify the optimal price to maximize revenue.



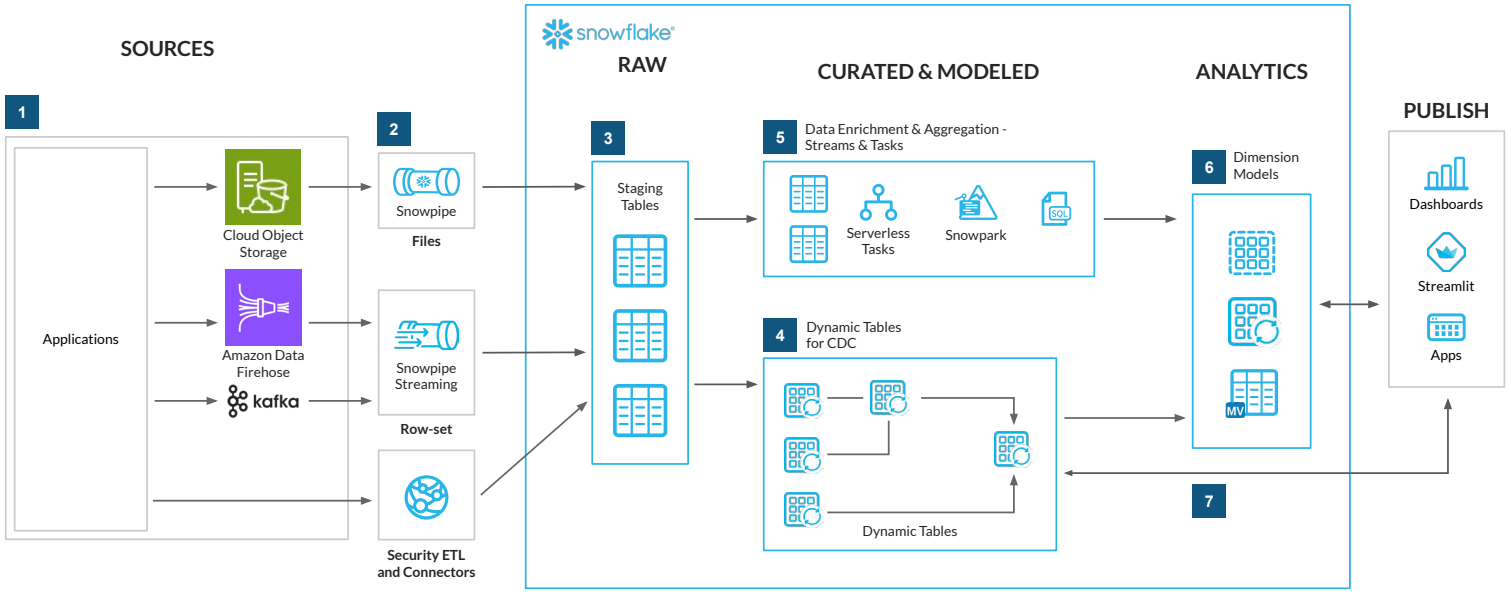
1. Load 1st party Product Catalog and Point of Sale data using Snowpipe for file ingestion and Snowpipe Streaming for row-set ingestion.
2. Create a share to get access to live 3rd party web scraping and product transaction datasets.
3. Clean the data to remove any inconsistencies, errors or missing values using Snowpark and Python code; use Dynamic Tables for SQL processing.
4. Normalize the data to adjust for different currencies, taxes or promotional discounts.
5. Identify products across multiple channels using unique identifiers like SKU, UPC, EAN or ASIN. Enhance matching accuracy by applying NLP machine learning techniques to product attributes, such as names and descriptions, when direct identifier matches are unavailable.
6. Transform and model the data to ensure it is structured in a consistent format using Dynamic Tables.
7. Create dashboards or reports to track pricing trends, compare channels, and identify anomalies or opportunities reading from materialized views or Dynamic Tables for optimum performance.
8. Establish automation rules to take action on pricing insights.
9. Implement an automated dynamic pricing system that adjusts your prices based on real-time market conditions, competitor actions and other relevant factors.

CYBERSECURITY

Staying out of the headlines and avoiding the wrath of the C-suite are table stakes for cybersecurity teams. Streaming analytics help these teams get ahead of the latest cyber threats and protect their organizations. Such insights are the backbone for security teams that have made the shift from a reactive security stance to a proactive one. Low-latency streaming helps them detect and respond to threats almost as quickly as they emerge — minimizing damage, containing threats quickly and keeping CISOs smiling.

Enterprise security

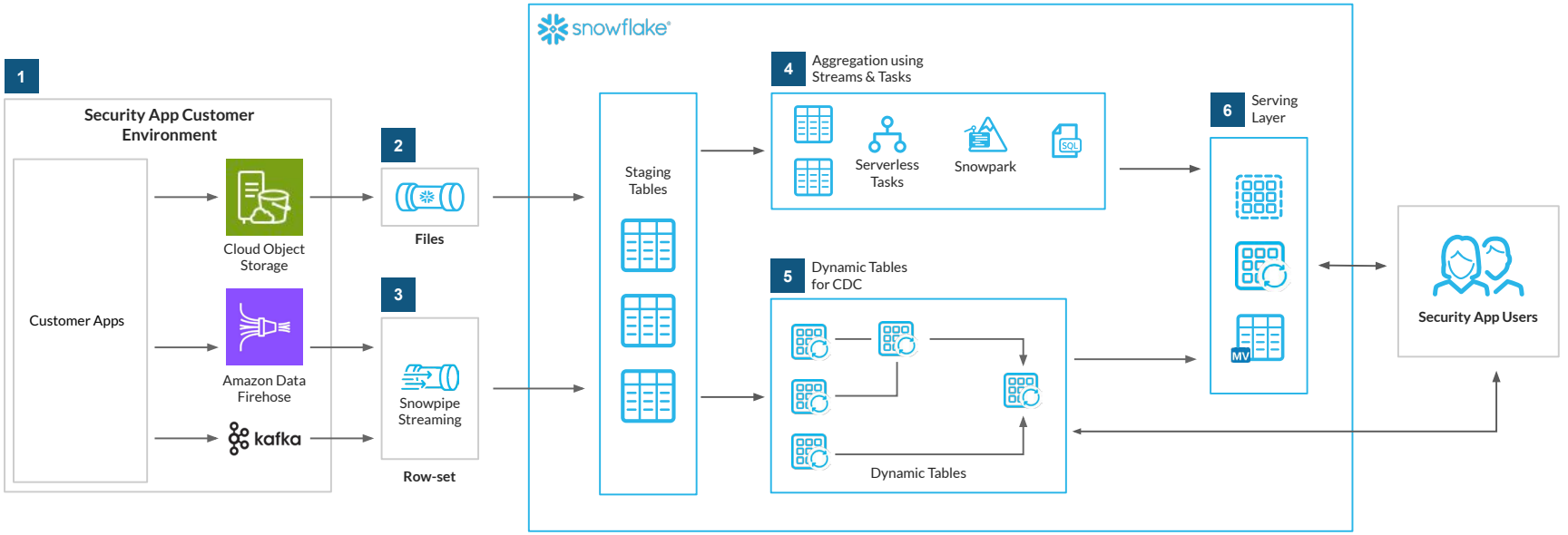
Financial services organizations, for example, can improve their security posture by streaming on-premises security logs — then merging and transforming raw logs, and joining them with existing views to create curated data for dashboards. The organizations can then use those dashboards to query and analyze activities to help find threat actors and act before they cause damage.



1. Data from different applications is loaded through streaming services like Amazon Data Firehose and Kafka. Files can be kept in cloud object storage.
2. For files, Snowpipe handles file auto ingestion into a staging table in Snowflake. Use Snowpipe streaming for row-set streaming data at lower frequency. Or use a 3rd party security ETL and connectors directly to stream data into Snowflake.
3. Land data in staging tables.
4. Use Streams & Tasks where strong orchestration is needed, Dynamic Tables for automated incremental refreshes, and Snowpark for advanced transformation using Python and DataFrame APIs.
5. Aggregate, normalize and enrich with automated incremental refreshers using Dynamic Tables. Transform and model the data to ensure it is structured in a consistent format.
6. Dimension models or Views are created, optimized to power user-facing dashboards.
7. Or feed from Dynamic Tables directly to dashboards, apps or Streamlit.

Security apps

Cyber app providers need timely data to remove silos and reduce complexity by quickly analyzing the most up-to-date logs, assets and configurations in one place. Streaming live data helps them collect and process data from a wide variety of sources, with diverse format, schema and fidelity to provide security analytics, conduct threat hunting and incident response, and more.



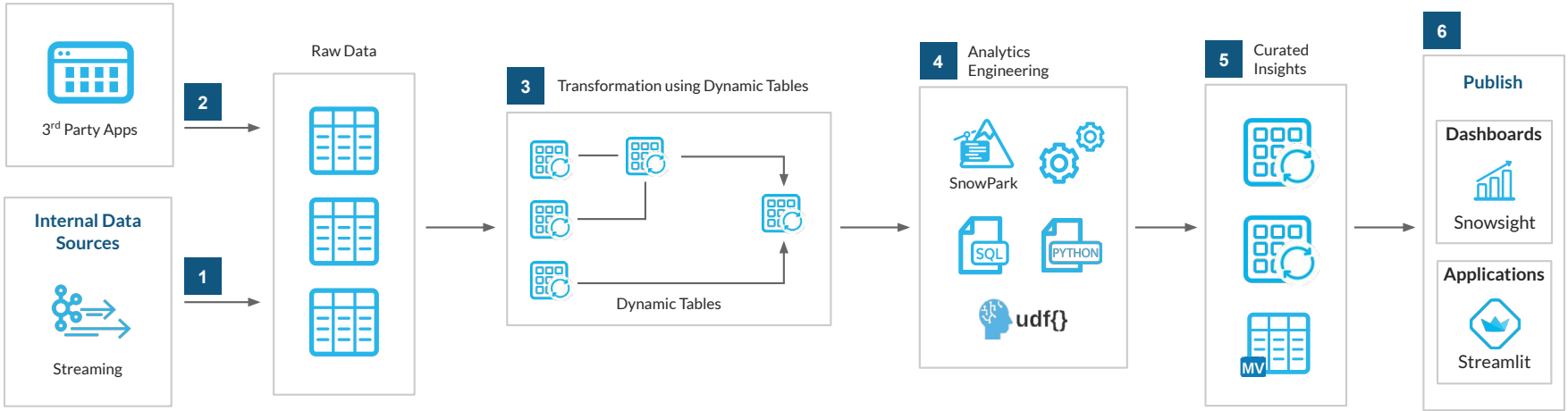
1. Data from security apps of customers' environment is loaded through streaming services like Amazon Data Firehose or Kafka. Files can be kept in cloud object storage.
2. For files, Snowpipe handles file auto ingestion into a staging table in Snowflake.
3. Use Snowpipe streaming for row-set streaming data at lower frequency, like Kafka and Amazon Data Firehose.
4. Use Streams & Tasks where strong orchestration is needed, Dynamic Tables for automated incremental refreshes, and Snowpark for advanced transformation using Python and DataFrame APIs.
5. Aggregate, normalize and enrich with automated incremental refreshers using Dynamic Tables. Transform and model the data to ensure it is structured in a consistent format.
6. Dimension models or Views are created and optimized to power user-facing dashboards.
7. Or feed from Dynamic Tables directly to dashboards, apps or Streamlit.

HEALTHCARE AND LIFE SCIENCES

Healthcare and life sciences organizations use data to provide highly personalized patient care and proactive treatment. Streaming analytics enable them to continuously capture and analyze data to help improve patient care, shorten hospital stays and provide better patient support or self-service capabilities. They also use these insights for predictive analytics and near real-time monitoring of IoT devices.

Medical IoT device ingestion

In analytical use cases that mostly collect health data that is not mission-critical, healthcare providers use medical IoT data to monitor patients in near real time, track adherence to medications and manage chronic conditions remotely. Streaming high volumes of this IoT data at low latency helps provide healthcare teams with more accurate and timely information that can lead to faster decision-making and better patient care.



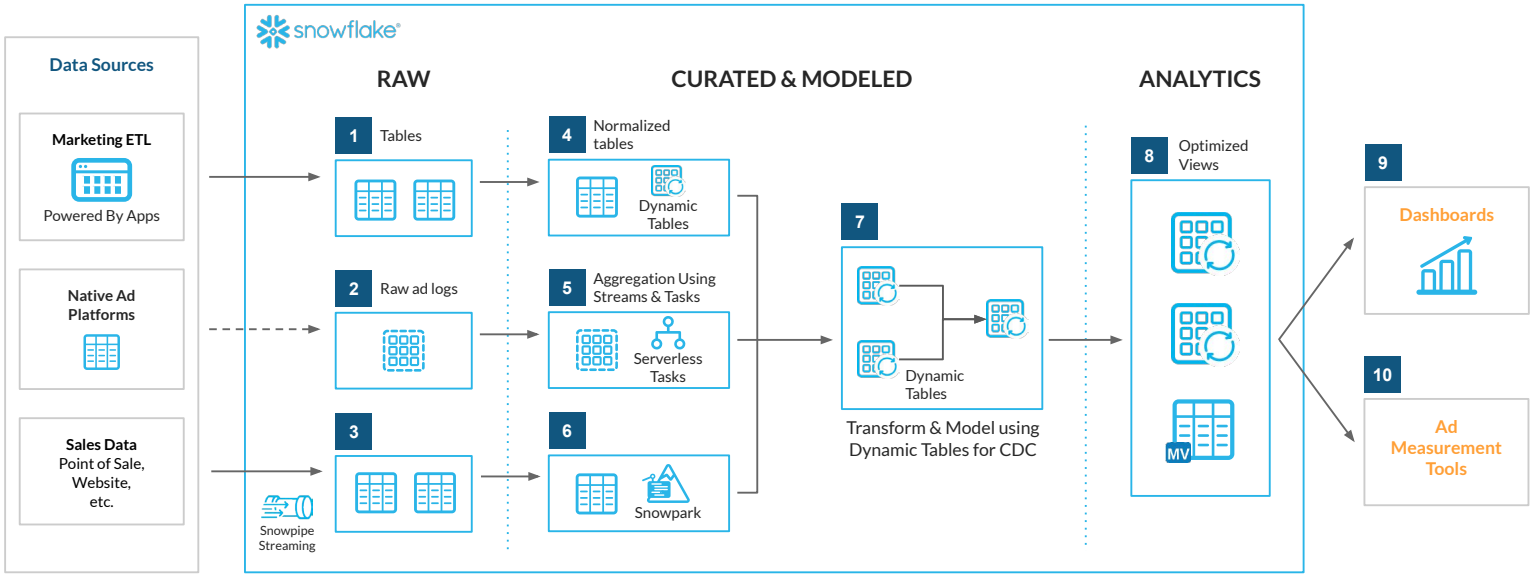
1. Source data from IOT-enabled devices are streamed into MQTT broker. The streaming Ingest service listens to those streams and pushes them into Snowflake.
2. Data from 3rd party systems like CRM are ingested into the raw layer.
3. Once the data lands into the raw base tables, Dynamic Tables reads the incremental records and processes them by merging them with 3rd party data in micro batches.
4. Snowpark Python and stored procedures perform analytic aggregations and generate metrics.
5. Dynamic Tables and Materialized Views to support analytics are created to support the frontend analytics.
6. Dashboards and applications are hosted in Snowflake or with partners to drive actions from medical devices.

ADVERTISING, MEDIA, AND ENTERTAINMENT

Organizations in the advertising, media and entertainment (AME) industry often work with very low-latency use cases, such as ad bidding and media buying. While analytical pipelines often focus on different use cases that don't require extreme low latency, AME companies still need to ingest and process near real-time data for analytical goals like campaign intelligence. Streaming data like user interactions, viewing history, watch time, social media shares, and ad click-through rates and conversions helps companies create personalized recommendations, evaluate the success of content and optimize ad campaign effectiveness.

Campaign intelligence

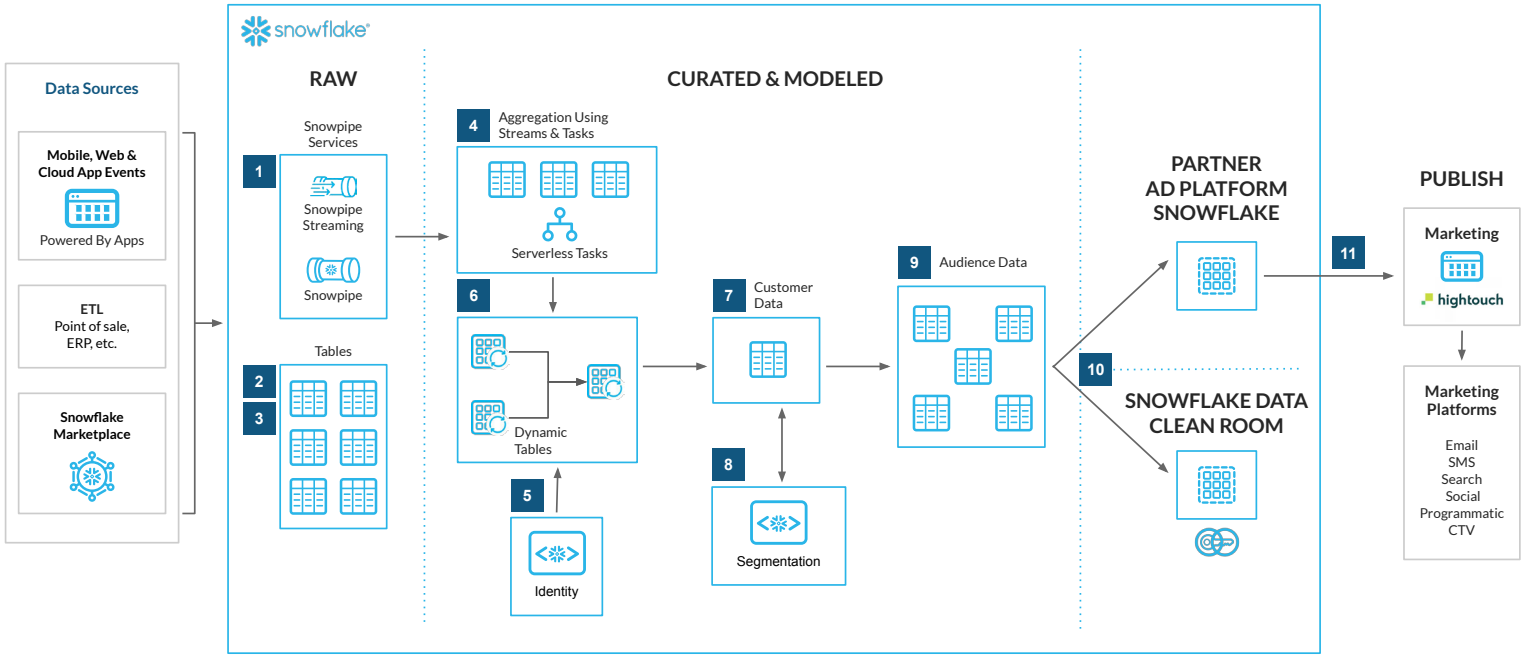
AME organizations gather, analyze and apply data to optimize their advertising and marketing campaigns. By streaming near real-time data to understand audience behavior, preferences and trends, these companies can create more targeted, personalized campaigns while optimizing content dynamically and enhancing campaign attribution.



1. Data from Marketing ETL tools are loaded into staging tables.
2. Data from Native Ad platforms are added via secure data sharing.
3. Other data, such as sales data, is ingested as normal, either streaming or in batches. Use Snowpipe Streaming for row-set streaming data.
4. Marketing data is normalized, so all platforms can be viewed together using Dynamic Tables.
5. Data from native ad platforms is aggregated, as necessary, and is normalized to match (flattened, deduplicated, cleansed etc). Use tasks for serverless with external orchestration, or Dynamic Tables for incremental processing.
6. Sales data is also processed and normalized, to match the cadence of the advertising data.
7. Use Dynamic Tables for change data capture. Transform and model the data to ensure it's structured in a consistent format.
8. Dimension models or Views are created and optimized to power user-facing dashboards.
9. The preferred BI tool is used to read from the tables or views for user-facing dashboards.
10. Tools for ad measurement, such as MMM, can be used from the normalized data, simplifying the process.

Composable Customer Data Platform (CDP)

A composable CDP provides a flexible approach to managing customer data across various channels and touchpoints. With the increased control and higher data granularity that a composable CDP provides, AME companies can stream near real-time data directly from their central data repository for segmentation and activation to any channel.

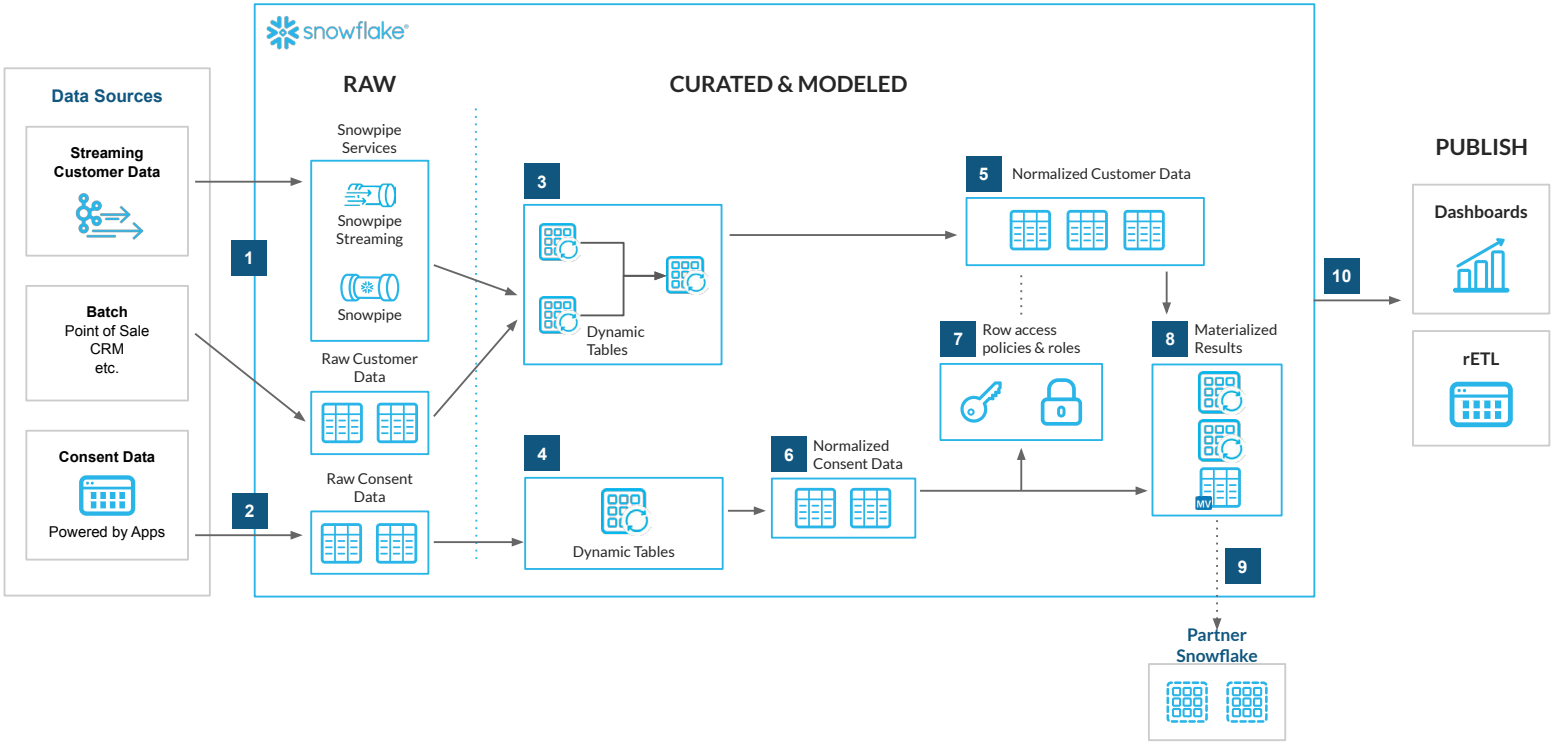


1. Data from websites and mobile apps are ingested using partner tools, either streaming or in batch. Use Snowpipe Streaming for row-set streaming data ingestion.
2. That data is combined with other data loaded via ETL.
3. Marketplace data is added to enrich customer profiles.
4. Aggregate, normalize and enrich click events, Use Streams & Tasks where strong orchestration is needed.
5. Native Identity is leveraged to combine cross-channel profiles.
6. Use Dynamic Tables for change data capture. Transform and model the data to ensure it's structured in a consistent format.
7. This data is then combined into a golden customer record.
8. Segmentation is done on top of this data, breaking customers into groups by similarity.
9. Audience data is created from this segmented data.
10. Ad platforms in the Snowflake ecosystem receive a share or use a DCR to activate.
11. Reverse ETL tools, such as HighTouch, can be leveraged with external ad platforms.

Partners such as Hightouch offer solutions integrated with Snowflake's streaming capabilities. Visit page 24 for more details.

Consent management

Consent management enables organizations to collect customer data in adherence with users' consent preferences and in compliance with relevant regulations. Organizations use this, for example, to manage and document customer opt-ins and opt-outs for various marketing communications. Data streaming helps AME companies manage user consent preferences across channels in near real time, helping them deliver the right communication to the right people.



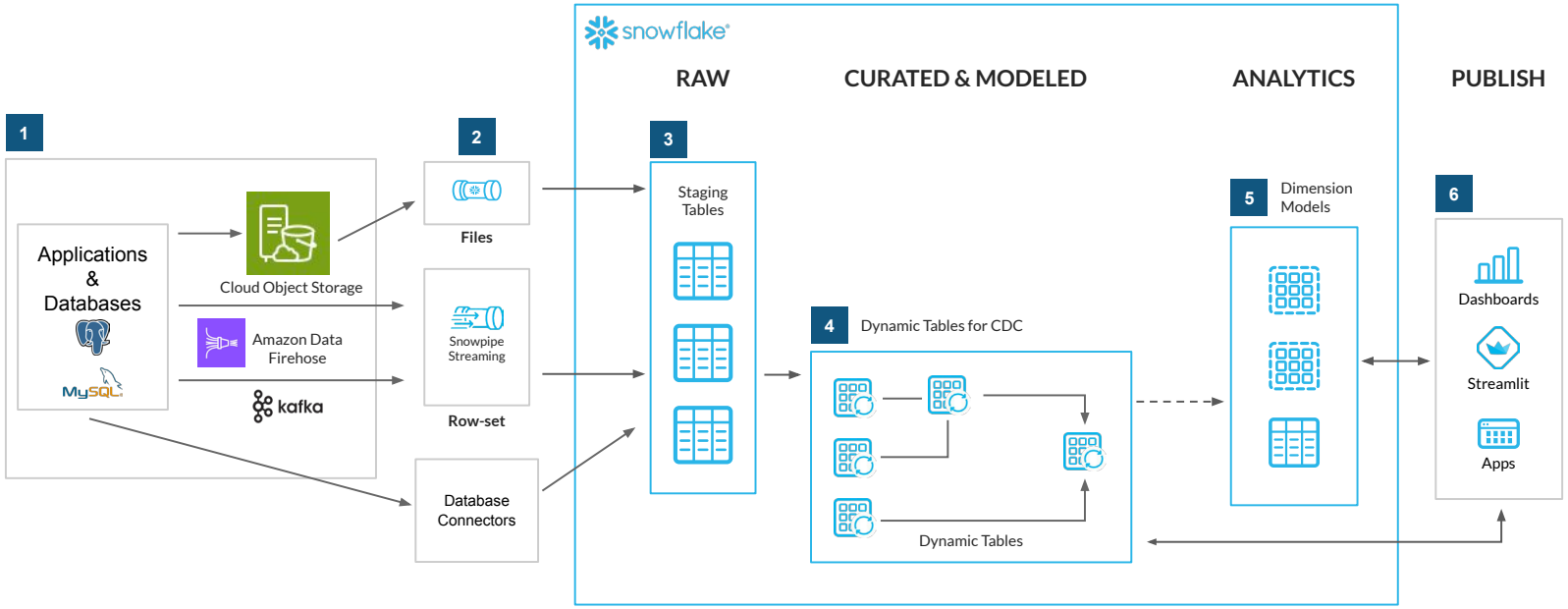
1. Customer data is loaded in batches, or streamed using Snowpipe Streaming (for row-set) and Snowpipe (for files).
2. Consent data from partner apps is loaded.
3. Customer data is deduplicated and aggregated using Dynamic Tables.
4. Consent data is similarly processed.
5. Processed customer data is normalized.
6. Consent data is also stored normalized.
7. Consent data is used to create Row Access Policies and, matching roles are created to govern customer data.
8. Consent data and customer data are combined into views with the current account.
9. Views are used to share governed data with partners.
10. Roles are used internally to control use of data on dashboards and for activation.

FINANCIAL SERVICES

Streaming helps financial services institutions ingest vital data from sources, such as stock market feeds, credit card transactions, economic indicators and customer transactions — and then use this data to assess and mitigate risk, personalize customer experiences, and improve products and services. The use cases in this book focus on how financial service organizations can cost-effectively ingest and process streaming analytics in near real time.

CDC for customer and payment analysis

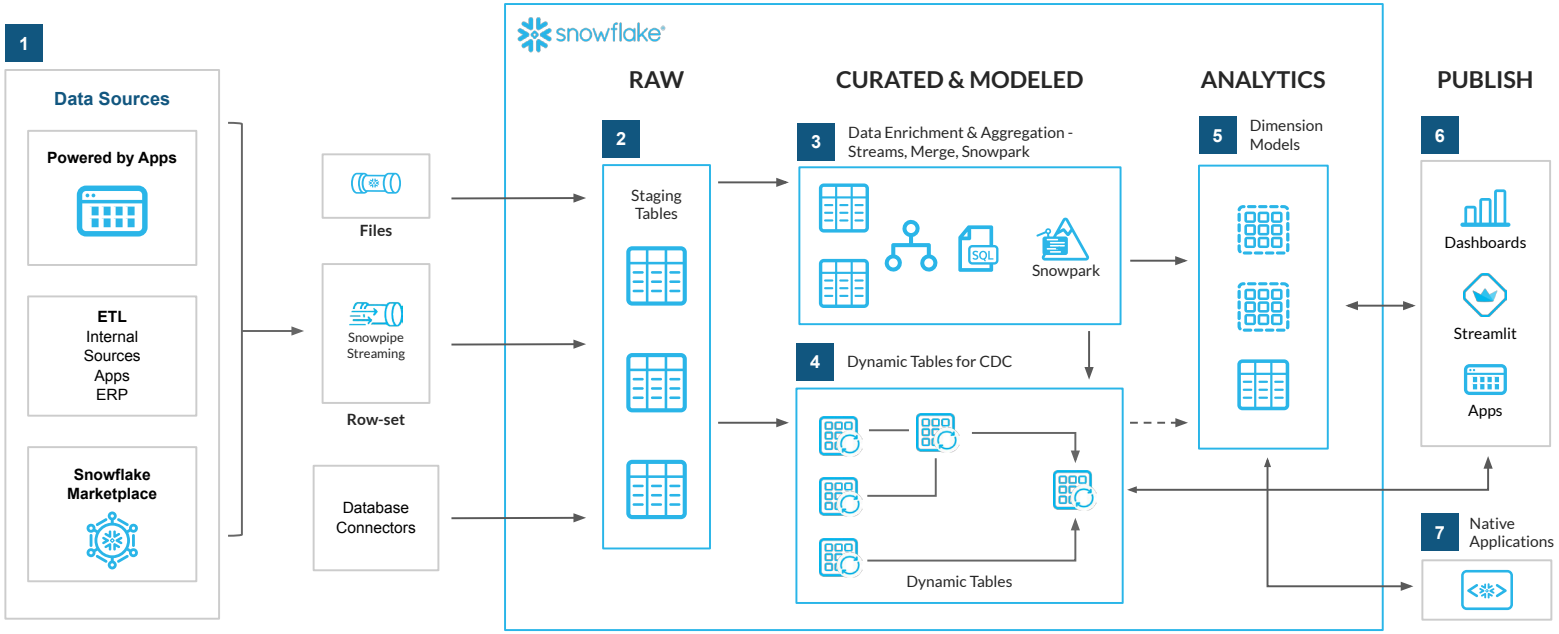
Change data capture (CDC) for customer and payment analysis involves capturing and processing near real-time changes to customer and payment data to gain insights, detect anomalies and improve decision-making processes. Data streaming is instrumental in implementing CDC, providing continuous data ingestion and analysis capabilities that help financial services institutions obtain accurate customer and payment analyses.



1. Customer data from applications is loaded in by batch, or streamed using Snowpipe Streaming (for row-set) and Snowpipe (for files).
2. Payment data from OLTP database is streamed from Snowflake native database connector.
3. Customer and payment data is loaded into raw staging tables along with auxiliary datasets.
4. Customer data is deduplicated and aggregated using Dynamic Tables. Payment data is uploaded and merged to ensure latest CDC records are reflected.
5. Dimensional tables and views are normalized and created as durable audit records and optimized for querying.
6. Dashboards and applications are hosted in Snowflake for analysis and internal operations and can be shared with partner institutions.

Regulatory reporting

Financial services organizations need to curate data to meet regulatory, risk and compliance mandates. Data streaming facilitates regulatory compliance by providing a continuous audit trail of traceable customer and payment data changes. This enables organizations to continuously monitor and report on activities to help improve reporting accuracy, foster greater transparency and address applicable regulatory requirements.



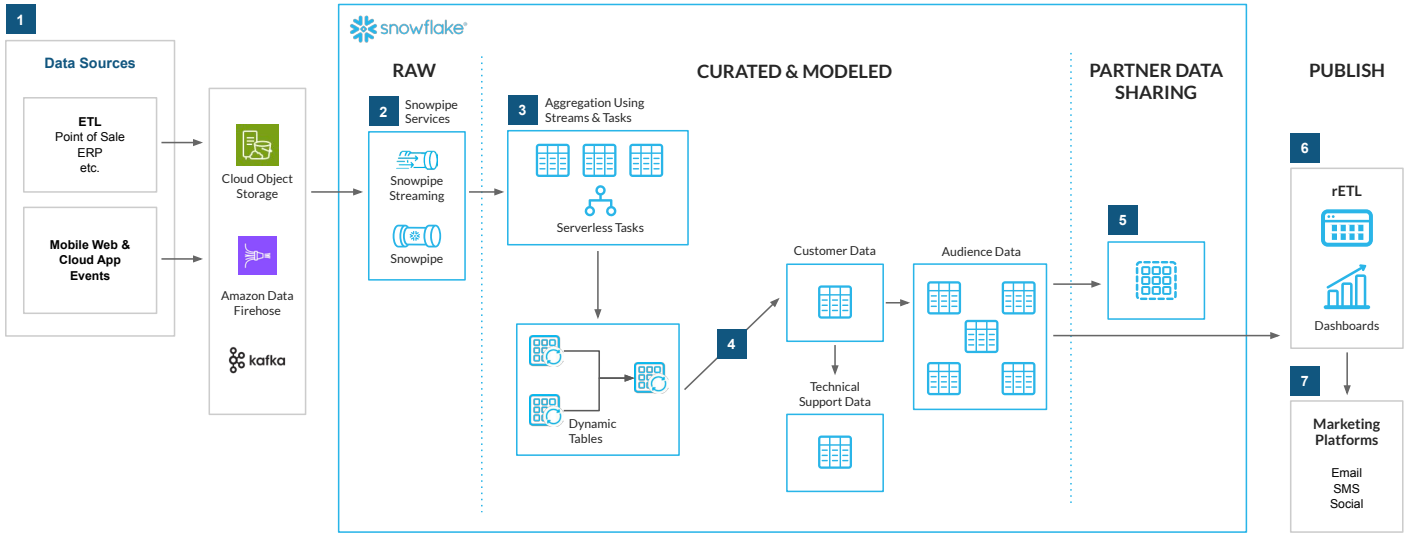
1. Acquire datasets from internal sources and third-parties using batches, streams, native connectors and direct shared data.
2. Store data at scale in any format. Securely share governed data between producer and consumer using object tagging, masking and transparent cloud region replication.
3. Integrate data from disparate sources using only required compute and language of choice (SQL, Python).
4. Automate integration of incremental and historical data with Dynamic Tables.
5. Provide regulation-specific model views for consumption by specific markets, regulators and reports using Data Masking. Run reports in parallel without contention, paying only for use.
6. Publish reports for global and market-specific regulatory bodies and drive internal reporting dashboards as built-ins (Snowsight) or integrated with additional BI dashboards.
7. Run applications with guided data input sharing and reporting enrichment.

TELECOMMUNICATIONS

Streaming in telecom involves ingesting data from network performance metrics, customer interaction logs, billing and usage records, and location data to enable efficient network operations, enhance customer experiences, and help with targeted marketing and network planning. While telecom companies have high demand for very low-latency use cases, this book focuses on analytical use cases, which often stream data in seconds and minutes, versus at sub-second speeds.

New subscriber management and onboarding

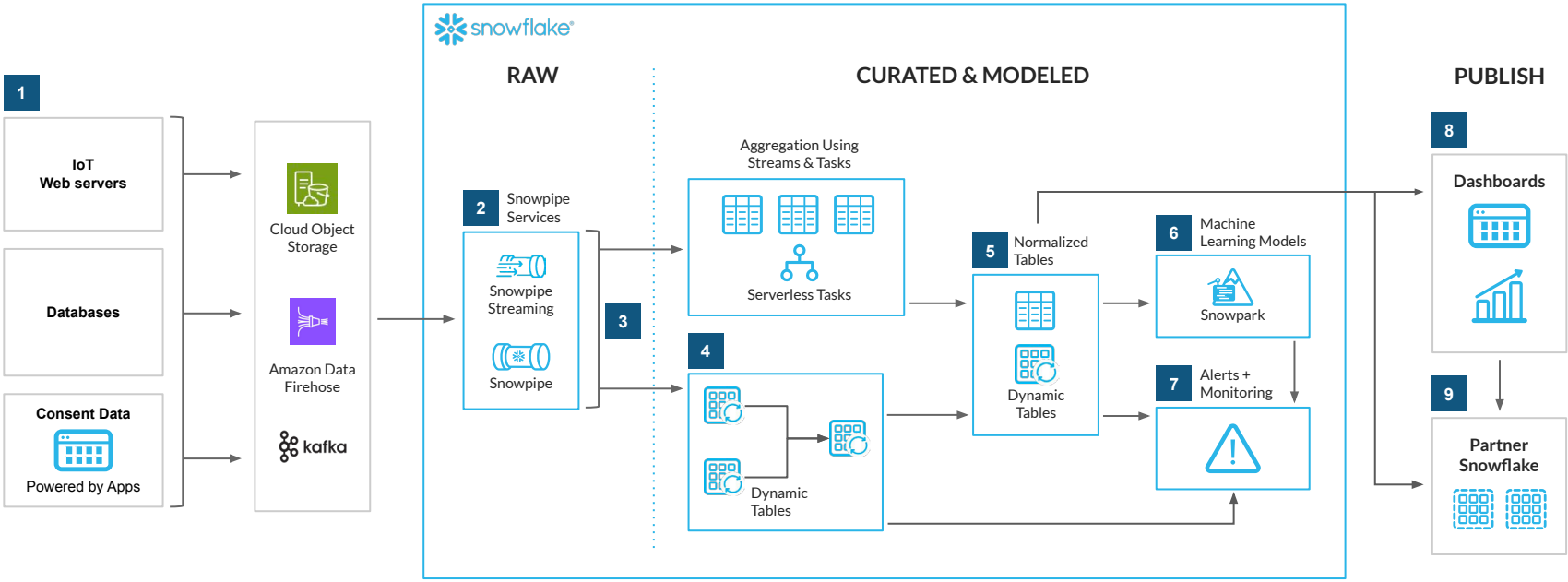
Subscriber management and onboarding refers to the process of acquiring new customers, managing their accounts and facilitating their seamless integration into the telecom network and service. Executing this process efficiently is important for telecom providers to deliver a positive customer experience and maximize customer lifetime value. Data streaming helps telcos capture and process new subscriber data in near real time to help provision accounts faster, accelerate onboarding and deliver more personalized offerings.



1. Data arrives from mobile and web apps where customers may be trying to place an order or troubleshooting, and streamed through Kafka.
2. Customer information from other offline systems from stores, technician trucks, and other field equipment are also loaded via Kafka or sent to files on object storage depending on source.
3. The Snowflake Kafka Connector ingests data using Snowpipe Streaming into a raw table while Snowpipe automatically loads files from object storage into raw tables.
4. The raw tables are deduplicated and cleaned using Streams and Tasks.
5. The cleaned tables from Snowpipe Streaming and Snowpipe are joined with other datasets and transformed using Dynamic Tables to aggregate and filter data to present tables for different business units. For example, a complete customer profile is stored into the customer data and used for accounting and billing purposes while that customer data is also joined with technical support information on any network or order issue and sent to a technical support team.
6. Aggregated customer and network information is shared with partners such as technical equipment vendors or outside sales and support staff.
7. Drive internal reporting dashboards as built-ins (Snowsight) or integrated with additional BI dashboards.
8. Run marketing campaigns with customer data and filtered customer groups.

Device analytics and usage management

Collecting, analyzing and managing analytics data from networked devices like smartphones, tablets, IoT devices and routers helps telecom companies better understand usage patterns and device performance. Streaming this data in near real time gives companies timely insight into metrics like signal strength, data usage, battery life and network latency, which helps companies identify performance issues faster and proactively address them for a more seamless user experience.



1. Data from IoT devices, apps and databases are streamed through Kafka. Use Snowpipe streaming for row-set streaming data at lower frequency, like Kafka.
2. Files can be kept in cloud object storage. For files, Snowpipe handles file auto ingestion into a staging table in Snowflake.
3. Use Streams & Tasks where strong orchestration is needed, and Dynamic Tables for automated incremental refreshes.
4. Aggregate, normalize and enrich data with automated incremental refreshers using Dynamic Tables. Transform and model the data to ensure it is structured in a consistent format.
5. Dimension models or Views are created and optimized to power user-facing dashboards.
6. Data is fed into Snowpark and Snowflake machine learning models to predict failures and for capacity management.
7. Alerts and monitoring is applied across the transformation steps and on end tables for anomaly detection and error reporting.
8. Dashboards are shared with different business units, and Streamlit apps can be created off the data in the normalized views.

MAPPING THE STREAMING ANALYTICS ECOSYSTEM

Depending on where you are in your streaming journey, you may adopt industry-specific solutions or technologies. Snowflake has an extensive partner network to help you be successful. These partners provide an array of solutions to ensure data flows quickly, smoothly and reliably in a variety of environments and architectures. Here's a look at some of the partners that make the process possible.

PARTNER SPOTLIGHT: THE AMAZON DATA FIREHOSE INTEGRATION WITH SNOWFLAKE'S SNOWPIPE STREAMING

What Amazon Data Firehose is

Providing an easy way to deliver data to Snowflake Snowpipe Streaming, Amazon Data Firehose (Firehose) is a fully managed, serverless ETL service for delivering real-time streaming data to popular data lakes, data stores and analytics tools. With Firehose, you don't need to write applications or manage resources. You configure your data producers to send data to Firehose, and it automatically delivers the data to the destination that you specified. You can also configure Firehose to transform your data before delivering it.

Key capabilities include:

- **Reliably load real-time streams into data lakes, warehouses and analytics services**
- **Create a delivery stream, select your destination and start streaming real-time data with just a few clicks**
- **Automatically provision and scale compute, memory and network resources without ongoing administration**

How it integrates with Snowflake and how customers benefit

The Snowflake Data Cloud is a global network that links businesses to their essential data and applications. Two methods exist for loading data into a Snowflake database: the file-based option using Snowpipe, and the row-based method via Snowflake's Snowpipe Streaming API. Prior to the introduction of Snowpipe Streaming, customers utilized Snowpipe for both batch and streaming workloads. The Snowpipe Streaming API lets clients write one or more rows into a Snowflake table with minimal latency and cost.

Firehose, a no-code, serverless service, can now be integrated with Snowflake's Snowpipe Streaming. This new functionality enables customers to stream data from numerous sources, including Amazon Kinesis Data Streams, into the Snowflake Data Cloud, presenting a simplified way to stream data into Snowflake at single-digit latency — and at a fraction of the cost, compared to traditional alternatives.

Setting up Firehose streams for delivery to Snowflake tables takes just a few clicks. Firehose is capable of handling gigabytes of data per minute and allows records to be analyzed in Snowflake within seconds. With no commitments or upfront investments, customers only pay for the streamed data. By leveraging Snowpipe Streaming, Firehose provides customers with a fully managed, no-code solution that is cost-effective and has lower latency. To initiate the use of Firehose, visit http://aws.amazon.com/firehose.

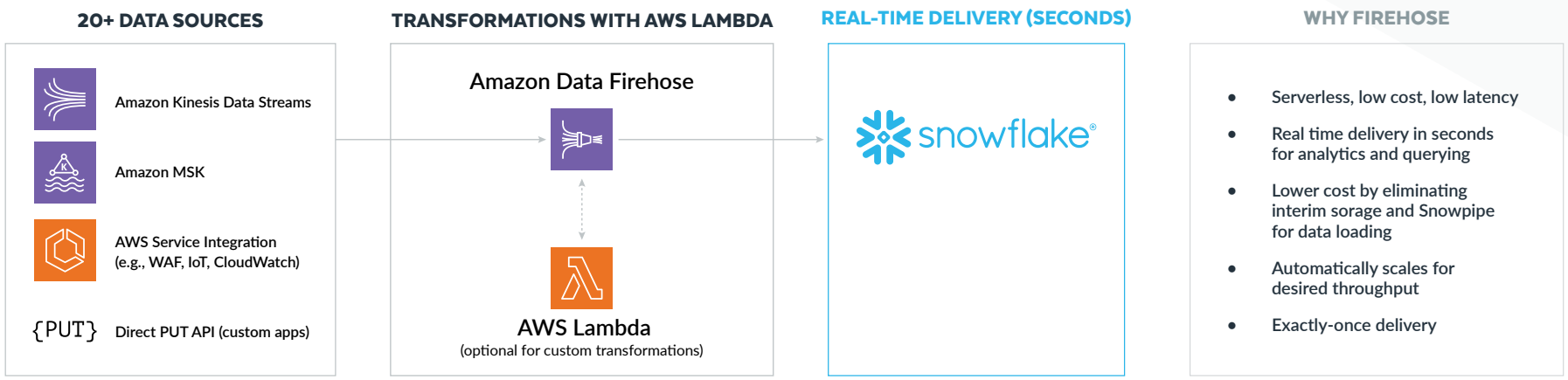


FIGURE 2: AMAZON DATA FIREHOSE—DELIVER DATA STREAMS USING SNOWPIPE STREAMING

Amazon Data Firehose and Snowflake in action

A closer look at the top three use cases across industries.

- **Healthcare monitoring**: Collecting and analyzing data from wearable devices and medical sensors to monitor patients' health conditions remotely and help with early detection of health issues, reduce hospital readmissions and provide personalized healthcare.
- **Predictive maintenance**: Analyzing near real-time data from IoT-enabled sensors on equipment to predict equipment failures and optimize maintenance leading to cost reductions.
- **Supply chain optimization**: Using IoT sensors to track the movement and condition of goods throughout the supply chain to enhance visibility, reduce errors, minimize delays and improve overall supply chain efficiency.

A ROBUST PARTNER ECOSYSTEM

Cirrus Link

Cirrus Link offers enterprise industrial internet of things (IIOT) software for a superior OT solution, enabling the enterprise to connect OT data to IT. It does this through Cirrus Link's IoT Bridge for Snowflake, which connects manufacturing and OT data to Snowflake. By using Message Queuing Telemetry Transport (MQTT) from Inductive Automation's Ignition platform in a no-code, self-learning and easy-to-implement solution, clients can transform data into valuable insights.

Coalesce

Founded in 2020, Coalesce is the only data transformation platform built for scale. Coalesce combines the speed of an intuitive graphical user interface (GUI) with the flexibility of code and the efficiency of automation, empowering its customers with increased data-team productivity and insights. As the only transformation solution uniquely built for Snowflake, Coalesce offers a way to visually build, adjust and deploy Dynamic Tables (in public preview) in Snowflake orders of magnitude faster without requiring hand-coding. Coalesce provides many "out-of-the-box" node types in addition to empowering users to create their own custom node types. This helps data teams drastically reduce the time and effort needed to build efficient data pipelines of any scale.

dbt Labs

dbt Labs is on a mission to empower data practitioners to create and disseminate organizational knowledge. With dbt Cloud, customers can simplify Dynamic Table development and management by defining transformations, managing model versions and lineage, and testing before deploying to production. This optimizes performance, efficiency and accessibility for analytics on streaming data.

DXC Technology

DXC Technology helps global companies run their mission-critical systems and operations while modernizing IT, optimizing data architectures, and ensuring security and scalability across public, private and hybrid clouds. DXC, together with Snowflake, provides a holistic data and AI solution, integrating connected data with all required data sets to enable internal and external data sharing and monetization faster by uniting data silos together in the cloud. This enables cross-functional, data-driven AI use cases leveraging fleet, customer, internal and third-party data, while also delivering improvements in product quality, predictive maintenance and data quality by analyzing and predicting anomalies based on vehicle signals.

HighByte

HighByte is an Industrial DataOps company delivering modeled, contextualized data to the cloud. HighByte Intelligence Hub offers two native connectors. The Snowflake Streaming connector uses the Snowpipe Streaming API to publish to Snowflake tables. The Snowflake SQL connector enables users to query Snowflake tables.

Hightouch

Hightouch is a complete, composable customer data platform (CDP) on Snowflake designed to collect, unify and activate customer data. Streaming Reverse ETL syncs data from Snowflake at extremely low latency and unlocks new high-speed use cases, such as those related to abandoned cart campaigns, conversion events, onboarding emails, lead routing and more.

Inductive Automation

Inductive Automation enables unlimited digital transformation with the Ignition industrial platform. Organizations use Inductive Automation's Ignition platform with Cirrus Link's IoT Bridge for Snowflake to publish OT data via MQTT, store it in Snowflake's Data Cloud with no code, and query it in Ignition for reports, analytics and applications.

Informatica

Informatica brings data and AI to life. Its Superpipe for Snowflake takes full advantage of Snowpipe streaming for customers to replicate and stream both initial and incremental data into Snowflake Cloud at a rate 3.5x faster than traditional change data capture.

Infosys

Infosys is a global leader in next-generation digital services and consulting, providing an end-to-end customer view by drawing correlations of connected equipment by various customer transactions (e.g., sales, invoices and payments) and hierarchical attributes (e.g., by service orders, sites or buildings). The solution acquires digital data in real time from IoT sources like HVAC equipment and chillers. It blends IoT data with ERP systems and applied intelligence to unleash hidden correlations and patterns to achieve a higher sales uptick and realize an increased customer engagement index. The Infosys network management system brings real-time advanced metering infrastructure (AMI) data into an enterprise data warehouse. Meter events, meter exceptions and usage data are brought into Snowflake via Snowpipe Streaming to get insights.

kipi.bi

kipi.bi is a real-time fleet-management solution that helps businesses overcome data gaps and deliver rapid insights at scale. Helping optimize route planning, minimize fuel consumption and respond dynamically to changing conditions, its solution enables the ingestion and processing of real-time data streams from IoT devices on fleet vehicles. It accomplishes this by using Snowflake Snowpipe Streaming for volume and velocity data ingestion, and Dynamic Tables (in public preview) for adaptable data schema to accommodate evolving fleet requirements. The solution leverages third-party Snowflake Marketplace data sets (like weather data and fuel costs), then analyzes data in Snowflake, making it ready for consumption and an ML workload. kipi.bi also uses Streamlit in Snowflake to create a user-friendly data app, providing an intuitive, interactive dashboard that allows fleet managers to effortlessly visualize and analyze real-time insights.

LTIMindtree

LTIMindtree is a global technology consulting and digital solutions company that enables enterprises across industries to reimagine business models, accelerate innovation and maximize growth by harnessing digital technologies. LTIMindtree's REDAR is an AI-based automated solution that powers ecommerce growth strategy with smarter and faster insights to decode market demand for a sustainable product portfolio. REDAR uses Snowpipe to process reviews and rating data in micro batches, helping triangulate market performance, product performance and portfolio through algorithmic interventions to arrive at actionable recommendations on pricing, new product opportunities and promotional strategies.

Sigma

Sigma is a cloud analytics platform that reinvents how businesses work with data. By bringing together the data team and business leaders into a single tool, Sigma provides customers with near-instant access to their cloud data warehouse, unlocking essential insights for the organization. Snowflake's Dynamic Tables feature (in public preview) presents a new, more optimized materialization in Sigma, making it a more valuable tool for analysis. Users now describe their data set in the Sigma interface and have full visibility into the definition and status of the materialization within Sigma. This allows users to work effectively with slices or aggregations of very large data sets, without needing third-party tools or becoming blocked on the data team.

Solace

Solace specializes in event-driven architectures and provides a platform for real-time data movement. The Solace PubSub+ Snowflake Streaming Connector allows organizations to ingest high throughput/low latency event streams into Snowflake via the Streaming API. By unlocking real-time data flow between Solace event streams and Snowflake, businesses can ensure efficient, reliable and scalable event-driven architectures.

Streamkap

Streamkap makes streaming ETL effortless for change data capture (CDC) from all popular databases. Streamkap's real-time CDC pipelines stream data into Snowflake through Snowpipe Streaming. It takes just seconds to enable low-latency modeling with Dynamic Tables.

Striim

Striim's Intelligent Integration unifies data across clouds, data and applications in real time. By seamlessly moving data from over 150 sources to Snowflake, Striim intelligently combines batch and streaming to provide leading data ingestion performance for data and AI initiatives on Snowflake.

ThoughtSpot

ThoughtSpot is an AI-powered analytics company that lets users find data-driven insights and take action. Using ThoughtSpot's natural language search (NLS), business users are empowered to discover data-driven insights from Snowflake Dynamic Tables (in public preview) to predict sentiment analysis from beauty product reviews in real time.

ADVANCING THE BENEFITS OF STREAMING

IT leaders are realizing significant benefits from investing in data streaming. In a recent **industry survey**, 76% of IT leaders reported increased business and IT efficiency, while 73% observed improved profits.

With low-latency data streaming from Snowpipe Streaming, IT leaders across industries can gain insights from data within seconds, hours or days. They can achieve near real-time data monitoring to help make better, faster decisions. They can mobilize streaming data to create stronger customer connections and optimize operations. And they can increase resilience cost-effectively by identifying trends and implementing proactive risk-mitigation strategies.

Check out this streaming data pipelines page on Snowflake.com to learn more.

ABOUT SNOWFLAKE

Snowflake enables every organization to mobilize their data with Snowflake's Data Cloud. Customers use the Data Cloud to unite siloed data, discover and securely share data, and execute diverse artificial intelligence (AI) / machine learning (ML) and analytic workloads. Wherever data or users live, Snowflake delivers a single data experience that spans multiple clouds and geographies. Thousands of customers across many industries, including 691 of the 2023 Forbes Global 2000 (G2K) as of January 31, 2024, use the Snowflake Data Cloud to power their businesses.

Learn more at **snowflake.com**