A Buyer's Guide to Streaming Data Integration For Google Cloud Platform

A Buyer's Guide to Streaming Data Integration For Google Cloud Platform

Table of Contents

CHAPTER 1 Setting the Stage for Digital Transformation

Throughout the history of modern data analytics, one constant driving force has been the need for accurate, comprehensive, and current data. The business world is evolving at a faster pace than ever, and the factors that create and maintain an advantage increasingly relate to who can make decisions, respond to change, serve customers, and execute on opportunities quicker than the competition.

For decades, businesses accepted the fact that producing reports involved manual labor and that the data they had to work with was days or even weeks old. Customers, partners, and colleagues were accustomed to delays in receiving reports and responses to questions.

The primary source for data is operational or transactional business systems. Converting the data into a useful form requires a process called extract/transform/load (ETL) in which data is algorithmically changed from a source format into one that the destination application can understand. ETL is most commonly used in data warehousing applications in which large amounts of data from multiple sources is converted into a common format. Another popular ETL use is in acquisition scenarios in which the acquired company's data needs to be transformed into a format that can be loaded into the acquiring company's existing data stores.

Few data scientists or database administrators would say they like ETL, yet by some estimates, data preparation can consume up to 80% of their time.[1] The process of writing transformation scripts is laborious, error-prone, and manual intervention is often still required during the data-loading stage. It is a necessary step, however, in making data useful to the organization. Clearly, any solutions that can effectively automate or obviate the need for these largely manual processes would be in high demand.

The cost and complexity of gathering data from a wide range of sources and delivering it in a usable form for real-time analytics and reporting was a pipe dream for many organizations. Legacy hardware infrastructure was ill-suited for low-latency operations, and the software needed to handle multiple data streams without losing or duplicating packets was proprietary and costly if it existed at all.

The Digital Transformation Game-Changer

Digital transformation has changed the goals and economics of business, and in the process, propelled streaming data to the forefront.

There are four main factors driving this trend:

Data-driven decision-making – The emergence of low-cost, massively scalable databases and the software to manage them is changing in the way organizations of all sizes make decisions. Gut instinct and hunches are rapidly giving way to a disciplined approach to decision-making driven by data.

Cloud computing – The modest cost savings of rentable infrastructure are overwhelmed by the business agility benefits organizations gain by having access to vast libraries of world-class software, in most cases, with a few mouse clicks. Software that would take weeks or months to install locally

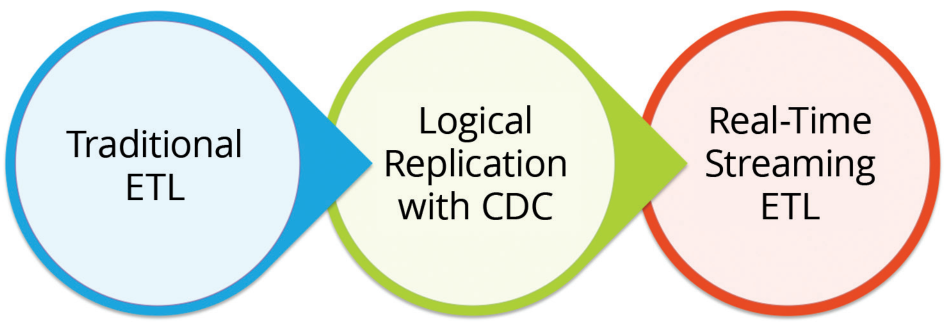
[1] Armand Ruiz, "The 80/20 Data Science Dilemma", *InfoWorld*, Sept. 26 2017 https://www.infoworld.com/article/3228245/the-80-20-data-science-dilemma.html

can be fired up in minutes, enabling organizations to experiment with new functionality and integrate advanced capabilities into their existing processes with low risk and at little cost.

Mobility – Customers and employees are now constantly connected, enabling organizations to run their businesses faster and get information to people who need it on the spot. For example, the ability to extend a customer an offer based upon the person's location or search activity can be a game-changer in marketing. Mobility also enables organizations to manage their field forces with greater efficiency, making the latest information instantly available to those who need it.

Open source software – Many software applications that were once the domain of just a few large enterprises are now available at no cost under open source licenses. This is changing the economics of software by significantly increasing the selection and power of tools that organizations can now buy. For example, stream processing applications like Kafka, Samza, Storm, and Spark Streaming are all available as free downloads and as low-cost services in the cloud.

These innovations have created pressing requirements for businesses to leverage more data more accurately and at or near the speed with which it was generated. Still stuck in batch mode, many IT organizations have struggled with these new business requirements. Data, as it turns out, is not generated in batches but rather row by row, line by line, and event by event. Traditional ETL didn't reflect this real-world reality. So business has thrown down the gauntlet: Move increasing volumes of data in real time in a format readily usable and consumable to support time-sensitive operational decision making.



Data movement technologies have advanced from traditional batch ETL to streaming integration to meet contemporary needs

Seen in this business context, it is easy to understand the fast-growing interest in streaming data integration (SDI), a.k.a. streaming ETL, which is a foundational element of enterprise modernization. Streaming integration is all about putting the infrastructure in place to collect and process data as it is generated. But processing streamed data involves much more than simply capturing information. Data needs to be moved from the source to its destination in a usable form. Kafka, for example, is a powerful and highly scalable messaging software platform for managing multiple data streams and delivering them to subscribers, which may be people, machines, and software applications. However, it does nothing to transform that data into a usable format. That's where streaming data integration comes in.

SDI software performs the transformations that normalize data into a form that can be loaded into transactional and analytical engines. It converts potentially heterogeneous input streams into a consistent format that databases understand. From there, queries can be performed with standardized business tools like the venerable SQL to enable real-time decisions by human operators or to kick off automated functions. There are even new versions of SQL emerging that allow queries against live, streaming data.

The goal is simple: Achieve a comprehensive view of your business using all data sources, including log files, databases, sensors, and messaging systems, enabling timely insight for real-time decision support.

Why Streaming Data Integration is Necessary

There are several frameworks available, some commercial and some open source, that developers can leverage to undertake data streaming and the related real-time data processing. And there is an increasing number of tools that simplify the process of programming, scaling, monitoring, and managing data streams.

That's a good thing because International Data Corp. and others have projected that in 2025 the world will create 163 zettabytes of data, about ten times as much as was created in 2016. IDC estimates that within five years' time, one-quarter of all data will be created in real time, with internet of things (IoT) devices contributing 95% of that volume.[2]

To visualize that amount of data, imagine a football field covered in 64 GB iPads. A stack totaling 163 zettabytes would reach 1.1 million miles into the air, or roughly four times the distance from the Earth to the Moon.

For organizations looking to benefit from streaming integration, the future is bright. According to Gartner, "By 2022, over 75% of data integration tools will natively support stream data integration, resulting in fewer interoperability issues between different data integration tools and low-latency processing."[3] Fewer and fewer interoperability problems will help make very low latency processing the norm rather than an exception.

What's In This eBook

This eBook intends to show how streaming integration is an integral part of the modern data architecture of any enterprise today. It will show the clear connection between vital business requirements or business drivers in general and the growing need for real-time decision-making support using real-time data from a wide variety of disparate sources.

Specifically, this eBook aspires to:

- Clearly describe what streaming integration is, comparing and contrasting it with more traditional data architectures such as traditional extract/transform/load (ETL) and batch processing;
- Show how streaming data integration (SDI) fits into existing data architectures as well as emerging new architectures;

---

[2] Sarah Wray, "IDC Predicts Ten-fold Increase in Data by 2025", *TMForum*, April 2017. https://inform.tmforum.org/data-analytics-and-ai/2017/04/idc-predicts-ten-fold-increase-data-2025/

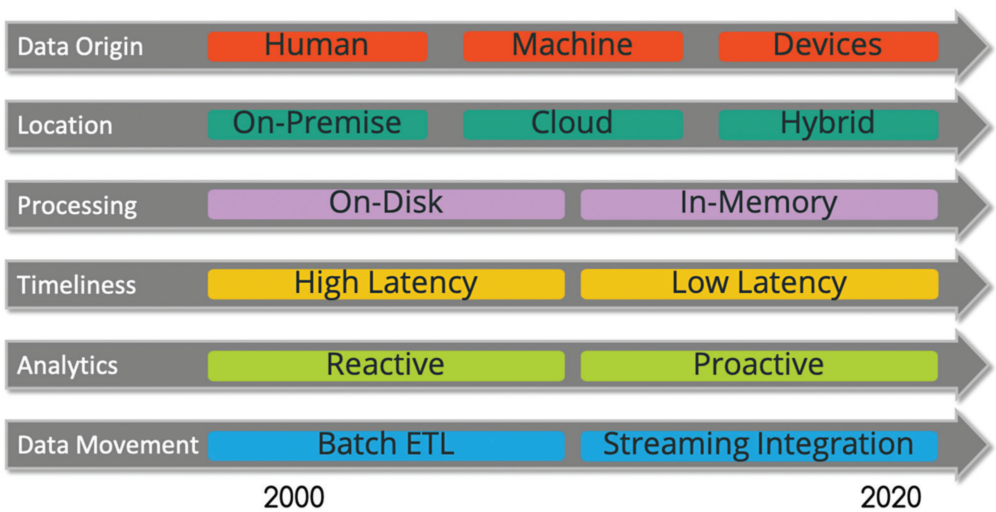
[3] Gartner: Adopt Stream Data Integration to Meet Your Real-Time Data Integration and Analytics Requirements, 15 March 2019, Ehtisham Zaidi, W. Roy Schulte, Eric Thoo

- Describe SDI's role in the context of cloud migration, hybrid cloud, multi-cloud and so on;
- Outline the central business value propositions of SDI;
- Describe what organizations and IT leaders should look for in an SDI solution, with a focus on the value and benefits of combining stream processing and SDI in one integrated platform;
- Show compelling SDI use cases that can help organizations envision their own applications; and
- Demonstrate the benefits of using the Striim SDI platform in combination with the Google Cloud Platform specifically.

CHAPTER 2 How SDI Fits Into A Modern Data Architecture

Compared to today's complex environment, the data management demands of a decade ago were simple. Most operational data was kept on premises in relational databases housed within captive data centers. Database schemas were well-defined and rarely changed. Data volumes grew at a predictable rate. The enterprise data warehouse was populated via a batch ETL process that was time-consuming but acceptable given that analytics was limited to long time horizons.

Today's data management landscape is almost chaotic by comparison. Relational data stores are now just one of an array of transactional and analytical engines that include legacy relational databases as well as multiple NoSQL options like graph, key-value stores, and schema-free document databases. Data warehouses coexist with free-form data lakes that hold unstructured content of every imaginable type. In-memory analytics engines perform decision-support calculations on transactional data in near real-time. And in many cases, an enterprise's data assets are spread across multiple public and private clouds, with data constantly moving back and forth.



Today's data requires a different approach to data management, integration, and analytics

Increasingly, this mix also includes live data cascading in from sensors, point-of-sale systems, transaction logs, newsfeeds, social media posts, and other continuously updated sources. Nearly every action involving a microprocessor generates data, including application development, website interactions, network traffic, and database logs. The value of this data has a half-life that may be measured in seconds. Capturing, interpreting, and determining whether to discard or commit this data to long-term storage is a decision measured in microseconds. And the data volumes continue to grow at unprecedented speed.

In the future, it will be impossible to store all the data we create. The only way to manage data volumes at the scale that's predicted will be to use the information in-flight while it has the greatest value and then throw most of it away. That means making decisions on the fly about what data has value and for how long.

Some people argue that massively scalable storage platforms like Hadoop make ETL obsolete because organizations can now analyze large volumes of both structured and unstructured data in the same place. However, few operational systems can tolerate that kind of unpredictability. While data analytics tools can increasingly handle a mix of data types, the need for structured data will exist far into the future.

Data architectures will increasingly be designed for split-second decision making. Data that is priceless at one point in time may be valueless five seconds later. The companies that win will be the ones that can identify and extract that value, determining what to use when, what to keep, and what to throw away.

Cloud Equation

Data analytics and cloud computing will drive much of the evolution of these next-generation architectures. There is no question that organizations are rapidly moving analytical data operations and even some transaction processing to the cloud. Forrester Research estimates that global spending on big data projects in the cloud is growing 7.5 times faster than spending for on-premises solutions.[4] Not only do cloud vendors offer speed, variety, and scale that few enterprises can match, cloud applications enable customers to pay only for the processing power they use rather than requiring them to provision and install expensive server clusters.

Streams are the next frontier of data analytics. Conventional business intelligence applications have been used to analyze historical data in batch to make educated guesses about future trends for more than 20 years. Still, such capabilities are now table stakes in the enterprise world. The organizations that will outdistance their competitors in the future will be those that make decisions on the spot for use cases ranging from offering coupons to a shopper at the checkout register to trading stocks based upon small pricing perturbations.

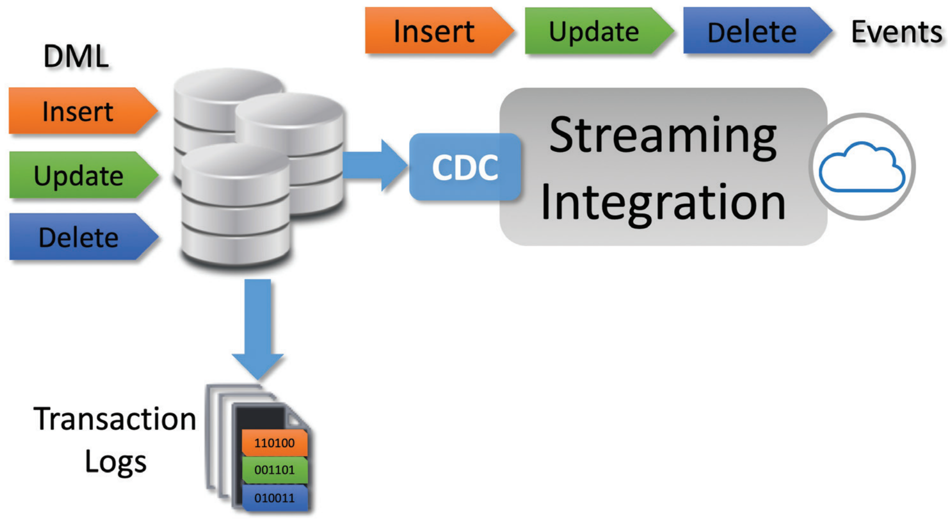
This shift to the cloud presents some practical challenges, however; particularly in the areas of latency and data integrity. To achieve optimal value in the cloud, data needs to be uploaded to cloud databases as quickly as possible after being committed to a local data store. Traditional ETL processes aren't up to the task since the lengthy procedure of loading data in batch obviates the opportunity to process it quickly. Loading full database snapshots is also impractical because snapshots require processing to pause while each new snapshot is uploaded. Few organizations today can afford even the minimal downtime that the process introduces.

The Change Data Capture Solution

Streaming data integration using change data capture (CDC) enables local processing to continue uninterrupted while data is synchronized to the cloud. In this approach, a snapshot of the operational database is taken and uploaded to the cloud without the need to take down transactional systems. While the snapshot is being uploaded, any new transactions are captured in a database log file. These changes are then applied to the cloud database sequentially. Future database updates continue to be replicated in this manner until both the local and cloud stores are in sync, after which the process continues in the background to ensure near real-time replication.

---

[4] Brian Hopkins, "Move Big Data To The Public Cloud With An Insight PaaS", *Forrester*, Aug. 8, 2017 https://go.forrester.com/blogs/insight-paas-accelerate-big-data-cloud/



Change Data Capture can turn database operations stored in transaction logs into streaming events for real-time delivery

But CDC is more than just about synchronizing copies. Traditional replication solutions of the kind often used in disaster recovery scenarios duplicate data but don't understand database concepts like update and delete. SDI solutions provide much greater sophistication because they are designed to understand database management rather than simply storage. A robust SDI platform integrates not only with databases but also with distributed streaming platforms as well as log files, IoT sensors, and messaging middleware. It does all this while ensuring data persistence and "exactly once" delivery, which is a complex and valuable capability that ensures that every message that was produced at one end of a stream is successfully committed at the other without duplication.

Not all SDI platforms can accommodate the scenario described above. Organizations that want to combine stream processing with cloud analytics should consider a solution that integrates with the most widely used streaming platforms, enterprise databases, and cloud data stores.

An SDI solution should also be distributed by design, particularly for environments with many edge devices. Data volumes in that scenario can easily overwhelm available bandwidth, causing latency problems or even data loss. Such environments require that data be processed at different points along the streaming pipeline so that only the most important information traverses the full network backbone. A distributed architecture is even more important when cloud data stores are involved and bandwidth is at a premium.

Designing for the Future

Real-time information will be an increasingly critical part of the data landscape in the future, and modern data architectures must be built from the ground up with this assumption. Few legacy database management systems were designed to accommodate streaming sources. While they excel at repetitive processing of large amounts of stored data, they are architecturally incompatible with the demands of continuous data delivery.

For example, the concept of a record or end-of-file marker doesn't exist in the streaming world because data isn't packaged in files. Rather, input event streams are a continuous, unbounded sequence of event records that are constantly in motion. SDI solutions must continually process the information on the fly, make decisions about it, and compare data-in-motion to data-at-rest, something no relational store can accommodate. For that reason, most streaming solutions work entirely in memory. Business-critical tasks like fraud detection, security monitoring, and fault detection in factory floor equipment don't tolerate the delays of ETL, so those tasks demand a new type of memory-based multithreaded analytics engine that operates at wire speed.

Legacy databases will be with us for many decades to come and a modern enterprise data architecture must take that reality into account. An SDI solution should be flexible enough to work with existing formats and schemas in relational databases and data warehouses as well as accommodate queries in well-established languages like SQL.

The drive to digital transformation will demand that analytics be at the heart of every decision a business makes. Speed will be the factor that distinguishes leaders from also-rans. As organizations put in place the infrastructure to accommodate this new reality, their data architecture decisions should proceed from the assumption that the future will be streamed.

CHAPTER 3 The Business Value of Streaming Data Integration

When many people hear the term "streaming data," the image that comes to mind is of a control panel monitoring the flow of oil through a pipeline or data feeds from sensors on a factory floor. While those are both classic examples of streaming data, the applications today are far more diverse and are expanding by the day.

Streaming data can be any data generated by a device as a result of an event. Events may be initiated by humans – such as by pressing a key or issuing a voice command – or generated by machines as a result of taking a reading, issuing an alert, or snapping a photo.

In the days before the internet became a ubiquitous backbone for connecting devices of all kinds, many events were processed in batch. For example, transactions in a retail store might be aggregated at the end of the day and uploaded to a master ledger for processing overnight. Today, those point-of-sale devices are likely to be connected to a central order processing system that captures transactions in real time and performs such housekeeping tasks as adjusting inventories, generating resupply orders, and updating e-commerce websites. Constantly connected devices thus enable retailers to be nimbler in managing their costs as well as customer expectations.

As the volume and variety of connected devices increases, so do the potential applications of streaming data and the possibilities for businesses to use it to their advantage. Among the common types of event streams today are customer orders, airline reservations, insurance claims, web clickstreams, newsfeeds, market data, email, and geolocation information. Having the ability to capture this data in real time and integrate it into business decisions is increasingly a source of competitive advantage.

When companies use real-time data for operational decision making, it improves both the bottom line and top line. First, it enables revenue growth by supporting initiatives that drive better customer experience, such as new innovative services that offer competitive differentiation. Secondly, having up-to-date information about business operations drives cost savings by improving how efficiently the business is run. Real-time operational visibility helps optimize resource use, minimize waste or errors, ensure compliance with regulations. Cybersecurity and fraud detection using real-time streaming data are great examples of reducing risks and driving costs down.

Real-time, continuous data access improves business operations, resulting in revenue growth and cost savings.

Game-Changing Technology

The emergence of powerful stream processing engines in recent years has transformed some industries fundamentally. Many conveniences that we now take for granted would be impossible without streaming data integration.

Take same-day shipping. The ability for a retailer to dispatch a product for delivery within minutes of the customer placing an order is a modern miracle enabled by stream processing. A customer's order triggers an inventory lookup to identify the best fulfillment location, generates a pick order and a request for pickup by the appropriate delivery service. A receipt is printed along with a shipping label and possibly promotional offers to be inserted in the package.

Upon pickup by the carrier, a tracking application is initiated that pinpoints the location of the package at each scan point or even tracks the progress of delivery trucks using GPS data. Customers can monitor the status of their orders along the way and may even receive a text notification upon delivery. Each stage of this process requires carefully choreographed programmatic decisions that need to be made in response to events. It would be impossible for a manual process to achieve these results economically.

Another example of the use of sophisticated stream processing is ride-sharing services. Perhaps you remember the days when finding a cab required standing on the side of the road and waving at passing traffic or placing a phone call to a dispatcher who then sent the request over two-way radio. The first option was frustrating and unpredictable, while the second was slow at best.

Modern ride-sharing services use stream processing to make the process efficient and transparent. When a user requests a ride using a mobile app, the order is streamed to a server that matches it with a continually updated database of drivers in the vicinity whose locations are tracked using streaming global positioning (GPS) data. When a driver accepts a request, the application exchanges information between the two parties in real time, calculates a pickup point and time, and continually updates both rider and driver on each other's location. Algorithms process streamed data to estimate pickup times. The estimated fare is also calculated based upon such variables as distance, vehicle type, traffic conditions, and the volume of rider requests. Fares may also be adjusted interactively if plans change.

The location of the vehicle during the trip is continually tracked using streamed GPS data and arrival times adjusted accordingly. Upon conclusion of the journey, the rider's credit card is charged, commissions are calculated, feedback requests are generated, and account information is instantly updated. Each step involves streamed data.

Both of these applications have fundamentally changed their industries. More than half of retailers now offer same-day delivery.[5] Ride-hailing is expected to be a $216.8 billion market in 2020, growing to nearly $319 billion by 2023.[6]

---

[5] BRP, 2017, The Number of Retailers Offering Same Day Delivery has Tripled in the Past Year to Meet Elevated Customer Expectations, According to a New BRP Survey [Press Release] Retrieved from http://www.prweb.com/releases/2017/09/prweb14682318.htm

[6] Source: Statista. https://www.statista.com/outlook/368/100/ride-hailing/worldwide

Streaming's Broad Business Impact

Many scenarios that weren't previously considered potential use cases for stream processing are finding new value thanks to the integration of real-time data. A primary driver of these new applications is the internet of things (IoT).

Device data is typically generated in streams. By moving it in real time to an analytical system, businesses can learn much about their equipment. Sensor data can identify machinery that is underperforming or is overloaded, thereby creating bottlenecks on the assembly line. Owners can also monitor factors like temperature, power consumption, slowdowns, and other factors that predict anomalies and failures that can have a significant cost impact. Batch data transformation is too slow to enable this kind of analysis in real time, although historical records are an essential factor in pinpointing anomalies. Real-time data integration is a necessary step in developing analytics that enables preventive maintenance.

One of the principal values of predictive analytics is in cost avoidance. Take the example of equipment used in oil drilling rigs or on gas pipelines. These assets are often located hundreds of miles from the nearest city. When equipment fails, service personnel may need to be dispatched on chartered aircraft for repairs. In the meantime, millions of dollars worth of production may be lost to downtime.

Using predictive analytics, data can be captured from sensors in the field and compared to historical records to look for warning signs that indicate equipment is about to fail. Repairs can then be made during scheduled maintenance periods without causing additional downtime or requiring costly travel and overtime.

Predictive analytics is useful in a wide variety of other scenarios as well. Retailers use predictive algorithms to prevent inventory shortages. Utilities use them to balance capacity. Stadium owners use them for crowd control. Municipalities use them to keep traffic flowing. Package delivery services use them to optimize staffing and the allocation of rolling stock. Historical records have long been used to predict the future, and the analysis of such records has been a mainstay of data warehousing for many years. What's different now is that we have the ability to match those records to live data to enable real-time decision-making. That's the business value of streaming data integration.

Stream processing of sensor data is also finding entirely new applications. For example, there are now weather services that continually monitor data streams from sensors on the ground to detect evidence of hailstorm activity and notify their customers the moment the ice begins falling. Hailstorms are a highly localized phenomenon, and pinpointing affected areas can be a hit-or-miss proposition. Armed with real-time information, insurance companies can now dispatch claims adjusters more efficiently, public works departments can begin repairs faster, and roofing contractors can get their salespeople onsite within minutes after a storm ends.

Another example in the insurance industry is the discounts some providers are offering customers who agree to install equipment in their vehicles to monitor driving behavior. Insurers can capture data streamed from these devices and reward safe drivers with premium discounts. Vehicle monitors can also be used to detect accidents and quickly dispatch emergency personnel to minimize human suffering.

Streaming data use cases don't have to be as sophisticated as these to deliver value. In many cases, the data sources are already in place, but the technology to capture and analyze data streams until recently wasn't robust enough to make it useful. Today, all it takes is some imagination and basic programming skills to infuse real-time data into existing processes. Here are just a few examples.

- An event organizer can analyze data about crowd patterns captured by smart video cameras and automatically dispatch catering, security, or support staff to avoid backups, generate revenue, and improve customer experience.
- A local retailer can monitor social media for mentions of its brand or posts that express a need for its product and respond accordingly.
- An airport hotel can watch flight reports for cancellation notifications and send discount offers to affected passengers over Twitter.
- A financial planner can automate the tracking of security prices to automatically place orders on behalf of individual customers according to each one's risk thresholds.
- An e-retailer can track transactions in real time and adjust prices and promotional offers based upon demand for certain products.

The business case for streaming data is often powerful. For example, credit card companies are battling fraud by capturing transactions in real time and comparing them to historical patterns of fraudulent activity. This tactic, combined with the use of EMV (Europay, MasterCard, Visa) has cut the volume of card-present fraud losses by more than half since 2015.[7] Streaming data is also redefining customer experience in the retail industry, where sellers constantly look for ways to improve customer experience through the use of customization, rapid delivery, and personalization, all of which require stream processing.

Streaming data can also be used in scenarios in which the impact is invisible. For example, modern intrusion detection and prevention systems used by IT cybersecurity organizations rely upon log data from networks, databases, and access points to spot breach attempts. Real-time data feeds are compared to historical records to detect variations from normal activity. Anomalous events can thus quickly be isolated, and the damage minimized.

The Role of SDI

Streaming data comes in different formats and at different intervals. It originates from a multitude of sources and may be transmitted by a variety of stream processing engines, message queues, and CDC logs from multiple vendors and open source applications. Some data may need to be routed for immediate in-memory processing, while some may be cached or stored for different periods of time defined by policies. Dropped connections may temporarily halt some streams and require data to be put in a queue for later transmission and reconciliation. Duplicate data needs to be deleted. In some cases, data may be "pulled" by the user via a query; in others, it's "pushed" without user involvement.

Data integration is an essential step toward making data useful. Take the example of geo coordinates. There are several standard ways to represent a location on earth. For example, the latitude of San Francisco is 37.773972 in Decimal Degrees, 37° 46.4383 N in Degrees Decimal Minutes, and 37° 26.298 N

---

[7] Steele, Jason. "Credit card fraud and ID theft statistics" *CreditCards.com*, October 24, 2017. https://www.creditcards.com/credit-card-news/credit-card-security-id-theft-fraud-statistics-1276.php

in Degrees Minutes Seconds. A fleet management application that requires geolocation data may be programmed to require that coordinates be expressed in a single format. If data arrives from multiple sources, there is a good chance some of it will need to be converted before it can be processed. Batch ETL is too slow to complete the transformations in time to make the data actionable, but SDI tools are up to the task.

Another common example is currency conversion. While this is a straightforward process in a batch operation, real-time currency conversion has a variety of new applications in trading, fraud prevention, tax calculation, and reconciliation. With 180 currencies in use around the world, this feature is critical for global financial organizations that need to keep a tight rein on their operations.

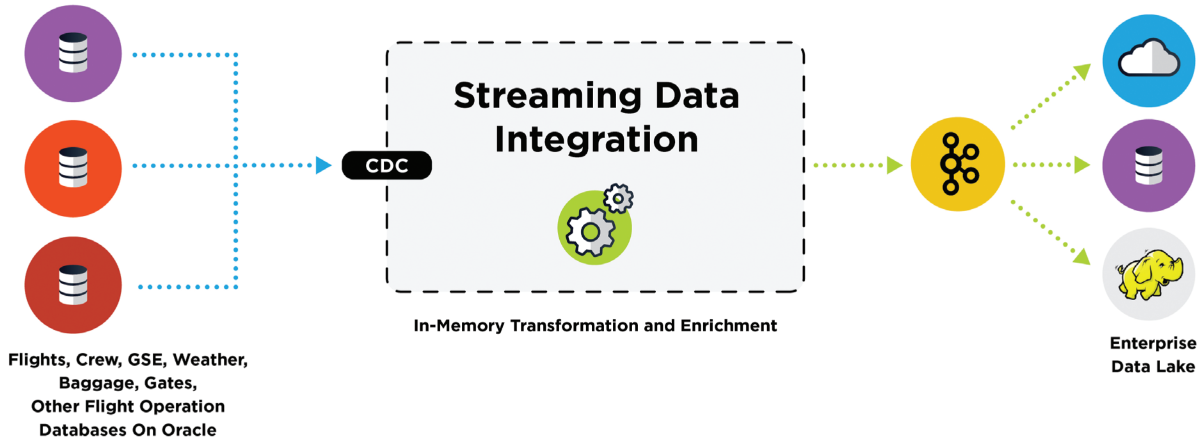
SDI in Practice

Let's look at a couple of real-world examples of SDI's business value in an enterprise context. As nearly all financial transactions have gone digital, criminals have invented new ways to launder money. The lag between the time a transaction occurs and its reconciliation with account records can leave just enough of an opening for criminals to withdraw cash advances, deposit the money into a compromised account, and transfer it to another account where it can be accessed anonymously.

One of Europe's largest retail banks is using advanced analytics and is fighting back with streaming data. It monitors ATMs for "impossible distance" transactions, a type of fraud detection that looks for transactions that occur at multiple geographically separate ATMs within an implausibly short span of time. The bank also looks for ATM withdrawals that are transferred to another account and immediately withdrawn and shut down the activity before it completes.

Another popular new use of streaming data is in real-time logistics. Transportation and delivery companies can use live data about conditions across their networks to make instantaneous decisions about routing, resource deployment, and maintenance, among other things.

One of the leading global airlines uses a combination of real-time and decision support analytics to improve operational decision-making. Its application ingests high volumes of flight, crew, and other operational data from enterprise databases using low-impact CDC. The streaming data integration solution performs in-memory transformations for data-in-motion before delivering the data in a consumable format to the airline's Kafka messaging solution in sub-seconds. The airline's existing tools then move the real-time data from Kafka to a NoSQL database, HDFS, and its enterprise data warehouse. The data lake solution powers live operational dashboards using data with less than two-minutes latency, and supports decision making for mission-critical flight operations across different airports.



Streaming data integration from different business systems using low-impact CDC enables operational intelligence for flight operations

Using live KPI dashboards, the carrier can rapidly detect events that may impact flight schedules and crew scheduling, and continuously optimizes flight operations based upon the latest updates to flights or weather forecasts. It can also track ground operations, such as de-icing, in real time to quickly determine the risk of flight delays. To improve on-time departures, flight operations teams can make immediate tactical decisions such as speeding the de-icing process by requesting another equipment, assigning a different aircraft, and reassigning passengers to other flights and gates.

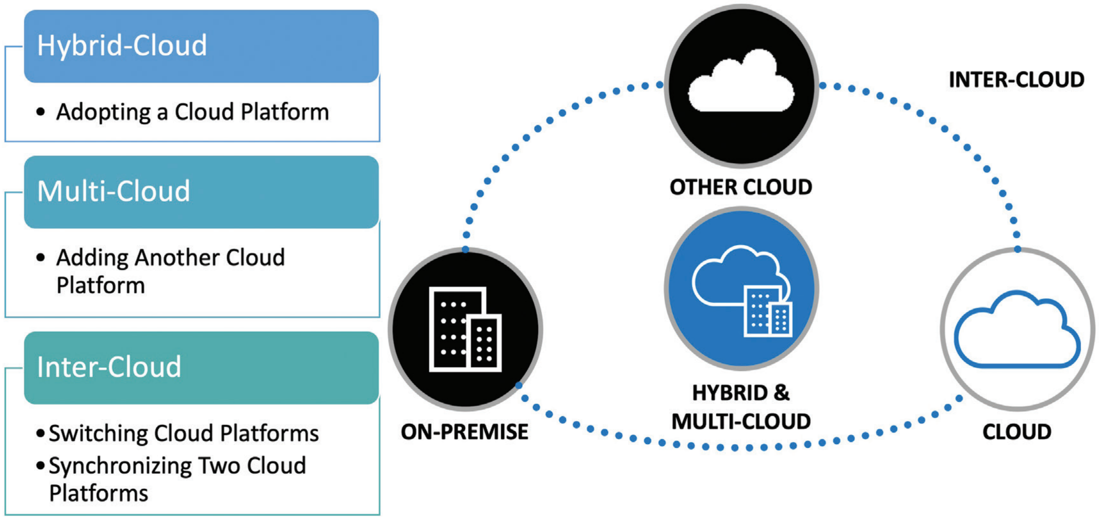
SDI is crucial to ensuring that the scenarios outlined above are possible. In the case of the bank, transaction data from ATMs needs to be compared against historical account activity and geographic data in less than a second so that fraud can be detected and shut down at the point of the transaction. Traditional ETL might enable banks to detect that fraud has occurred, but with streaming analytics they can not only stop false transactions in their tracks but look for signs that indicate the likelihood that a transaction is fraudulent before it even closes.

The airline must harmonize data from multiple sources, including weather feeds, real-time flight tracking information, and constantly shifting crew and equipment schedules so the decisions can be rendered in seconds. It would be impossible to make these decisions on a timely basis without the ability to transform and integrate data instantaneously. The more data streams that need to be combined and acted upon programmatically, the more critical SDI is to delivering business value.

SDI and the Cloud

Streaming data has taken on special business significance with the ongoing massive move to the cloud. Gartner expects that “By 2022, 75% of all databases will be deployed or migrated to a cloud platform, with only 5% ever considered for repatriation to on-premises.” [8] The business value of cloud migration is not limited to savings on software licenses, lower database administration costs, and almost limitless scalability. It also delivers rapid access customers gain to powerful tools that amplify the value of data, or that can be applied to analytics and machine learning.

[8] Gartner, 1 July 2019, Gartner Says the Future of the Database Market Is the Cloud [Press Release] Retrieved from https://www.gartner.com/en/newsroom/press-releases/2019-07-01-gartner-says-the-future-of-the-database-market-is-the



Adopting a hybrid cloud, multi-cloud or inter-cloud architecture requires continuous connectivity among on-premises and all cloud environments

Cloud migration can be a complex and risky prospect, though, particularly when business-critical, sensitive, and operational data is involved, and outages or data loss are unacceptable. In addition, some organizations must maintain a canonical record of their data on-premises for compliance purposes and can only replicate copies to the cloud.

The migration itself can be quite complex and time consuming. Enterprise database migrations typically require careful refactoring of the application code, the database and the schemas before any data is even moved. With the planning, development, and testing phases, it can take anywhere from a few weeks to several months, in some cases years – all depending on the applications and databases involved.

For organizations that want to take advantage of cloud scalability, flexibility, and economics, such lead times may frustrate or foil their migration ambitions. As noted earlier, digital transformation almost by definition mandates that organizations take advantage of the agility the cloud enables.

For all those benefits, however, few organizations are migrating wholesale to infrastructure-as-a-service. Hybrid cloud is expected to be the dominant architecture for the foreseeable future for reasons that include compliance, control, and optimization of IT assets. Nearly 60% of enterprises are using a hybrid cloud model today, and 84% have a multi-cloud strategy, according to Flexera's 2019 State of the Cloud Report.[9]

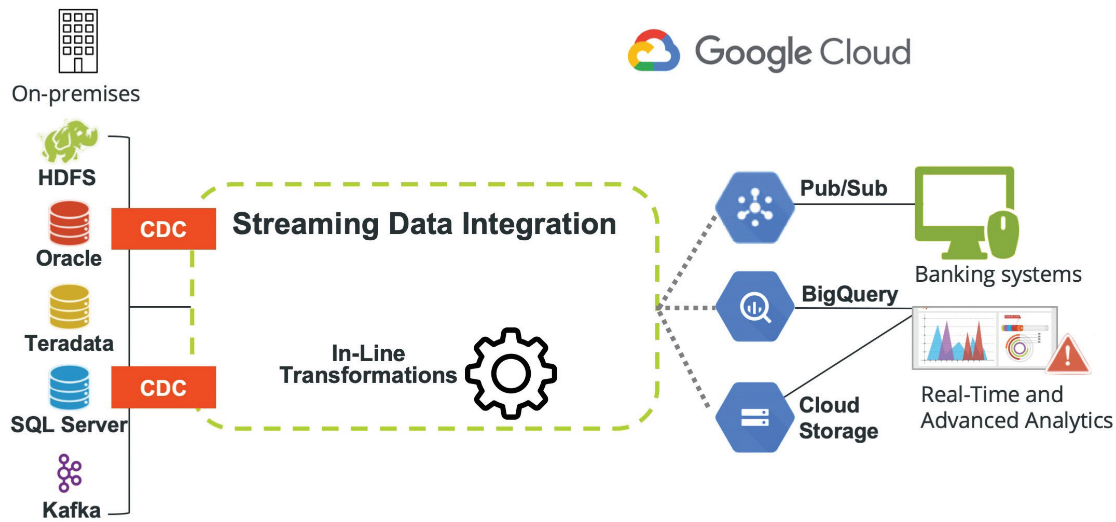
Data gravity, a term that describes the cost and difficulty of shifting large amounts of data from one platform to another, is one of the principal inhibitors of multi-cloud strategies. Egress fees charged by cloud infrastructure providers, along with the time required to upload and download large data sets over telecommunication lines, can force organizations to centralize data in one place, even if other platforms provide more useful tools.

Businesses can have the best of both worlds by using SDI and CDC to synchronize their on-premises and cloud databases, giving them the flexibility to run workloads where they make the most sense. For example, a leading financial institution uses CDC with SDI to perform a phased transition of

---

[9] Flexera. RightScale State of the Cloud Report, 2019
https://resources.flexera.com/web/media/documents/rightscale-2019-state-of-the-cloud-report-from-flexera.pdf

high-value workloads to Google Cloud for its modernization effort. The bank synchronizes its on-premises banking and data warehouse systems with multiple Google Cloud services, including Google BigQuery, Google Cloud Storage, and Google Cloud Pub/Sub. The gradual transition minimizes business disruption while enabling the much-needed modernization. Streaming integration allows the bank to continuously feed on-premises systems data to the new Google Cloud services so that it can fully leverage the modern Google Cloud Platform features for transforming how it performs daily business operations. By building next-generation banking applications on Google Cloud, it can offer differentiated services to its customers. Using real-time data in Google Cloud, it can implement time-sensitive, advanced analytics solutions for fraud detection and customer experience initiatives that move the business forward.



On-premises to cloud integration using SDI allows to offload high-value workloads to modern cloud platforms

Using SDI within the same cloud environment opens up the path to many new use cases. HomeServe, a leading home assistance provider, offers a smart water leak solution, LeakBot device. It detects hidden leaks before they cause significant damage. Leak detection can trigger a series of operational events and changes in the LeakCentral system. Visibility to each data change in the sequence of events is essential to understanding how a leak event unfolded. If batch ETL-based data snapshots were being used to ingest data, the all-important detail of what happened between snapshots would be missed.

Streaming data integration to Google BigQuery enables operational dashboards and granular data for machine learning models

HomeServe uses Striim's streaming data integration platform to deliver rich operational data to the Google BigQuery analytics environment from its MySQL database hosted in Google Cloud. With Striim continuously moving the operational data to BigQuery, it can be analyzed for operational intelligence and to provide insurance companies with a compelling business case for their service. HomeServe can perform post-operational analysis by delving into individual leaks, aggregating the flow rate of the leak, the engineer's home visit and report.

Data scientists use granular, streaming data to analyze how the device performs in the field, enabling them to optimize the service to homeowners and insurance partners.

A similar approach can be used in a multi-cloud scenario. For example, a company may want to synchronize its production data with a warehouse from one cloud provider and machine learning libraries from another. A combined SDI/CDC approach makes this possible with no interruption to production workloads.

CHAPTER 4

1 + 1 = 3: Streaming Data Integration and the Google Cloud Platform

The Google Cloud Platform offers powerful data management engines for every potential use case, ranging from transaction processing to analytics and machine learning. The offering encompasses three major cloud-native platforms.

Cloud SQL is a fully-managed database service that supports MySQL, Microsoft SQL Server, and PostgreSQL, which are three of the world's four most popular open source database management systems. Google's offering makes it easy to set up, maintain, manage, and administer relational databases on its cloud platform, enabling customers to connect to nearly any application built for those database engines. The service also automates data provisioning, storage management, and other administrative tasks.

BigQuery is a highly scalable cloud data warehouse that enables users to quickly analyze petabytes of data at high speed using SQL. BigQuery's serverless architecture minimizes operational overhead by automating resource provisioning behind the scenes, so customers pay only for the compute power they use. It is a popular core database for the development of machine learning models.

Cloud Spanner is a cloud-native scalable, enterprise-grade, globally distributed, and consistent database service that combines the benefits of relational structure with horizontal scalability. It delivers high-performance transactions and strong consistency across regions and continents with "five nines" availability and enterprise-grade security.

There are many advantages to shifting workloads from on-premises infrastructure to the Google Cloud. With the cloud's pay-as-you-go pricing model and emerging architectures like serverless computing, customers often see dramatic cost savings compared to traditional software licenses and on-premises hardware infrastructure.

Google Cloud Platform providers also offer a wide range of analytics and business intelligence tools that users can mix and match without incurring the overhead of installation, software licenses, and maintenance. Enhanced services such as Google BigQuery BI Engine, which is an in-memory analysis service for high-speed interactive visual analytics, also deliver advanced features that would be impractical or impossible to implement on-premises. Finally, Google maintains partnerships with dozens of independent software providers whose services can be provisioned with a few clicks.

Multi-Cloud and Multi-Engine

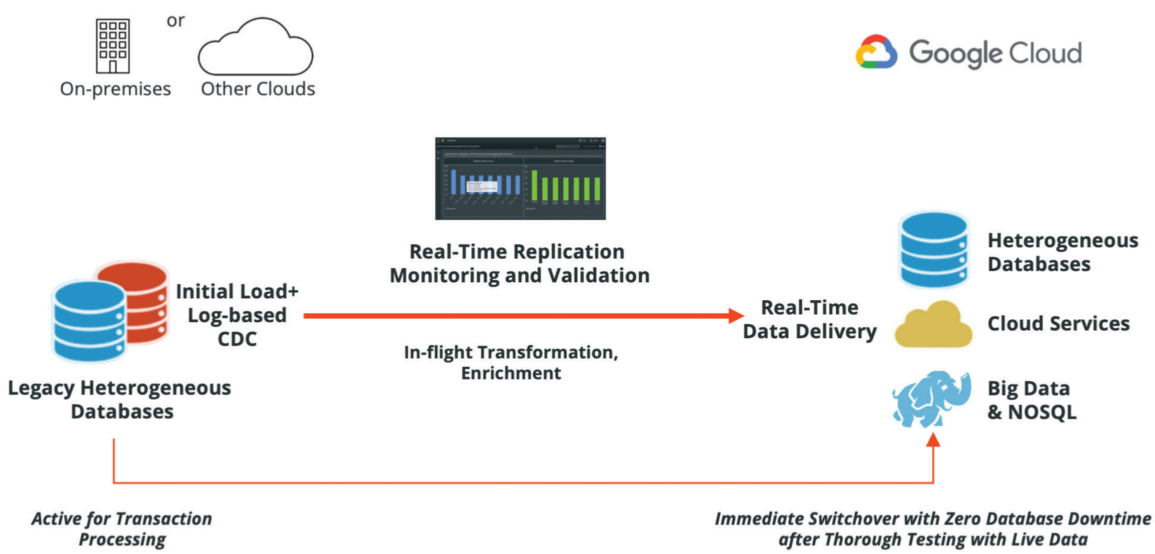
Many customers will want to use more than one database engine. For example, they may have legacy applications that require the consistency and ANSI SQL compatibility of Cloud SQL for MySQL but want to base future applications on a more flexible and scalable platform such as Cloud Spanner. Customers may also want to migrate historical data to BigQuery and combine it with real-time transactional data for predictive analytics. An SDI solution that supports multiple target environments makes setting up such a multi-faceted environment relatively straightforward.

Another popular use case is database migration. For example, customers may want to save costs by migrating from an on-premises Oracle environment to Cloud SQL Postgres or shifting from a Teradata data warehouse to BigQuery. Such migrations can involve considerable time and effort in data normalization and schema transformation, particularly when very large data warehouse volumes are involved.

As noted earlier, database migration can be a complex and time-consuming process. Using a stream processing platform that combines data integration with CDC minimizes business disruption and risks, ensures that data in the cloud is current, and protects against data loss during the conversion.

The Striim SDI platform offers both initial load and CDC from diverse sources like Oracle, SQL Server, MySQL, PostgreSQL, HPE NonStop, AWS RDS, and major data warehouses to the Google Cloud Platform. This enables cloud migration with zero database downtime and continually feeds real-time data to targets with full context for immediate use in analytic processing. Striim handles the filtering, transformation, aggregation, and enrichment of data-in-motion.

The process employs agents, which are lightweight processes that run on-premises close to the source data. Agents execute the process flow that extracts data and publishes it as events to the cloud database. Striim agents enable sub-second latency and secure connectivity from on-premises sources to the cloud.



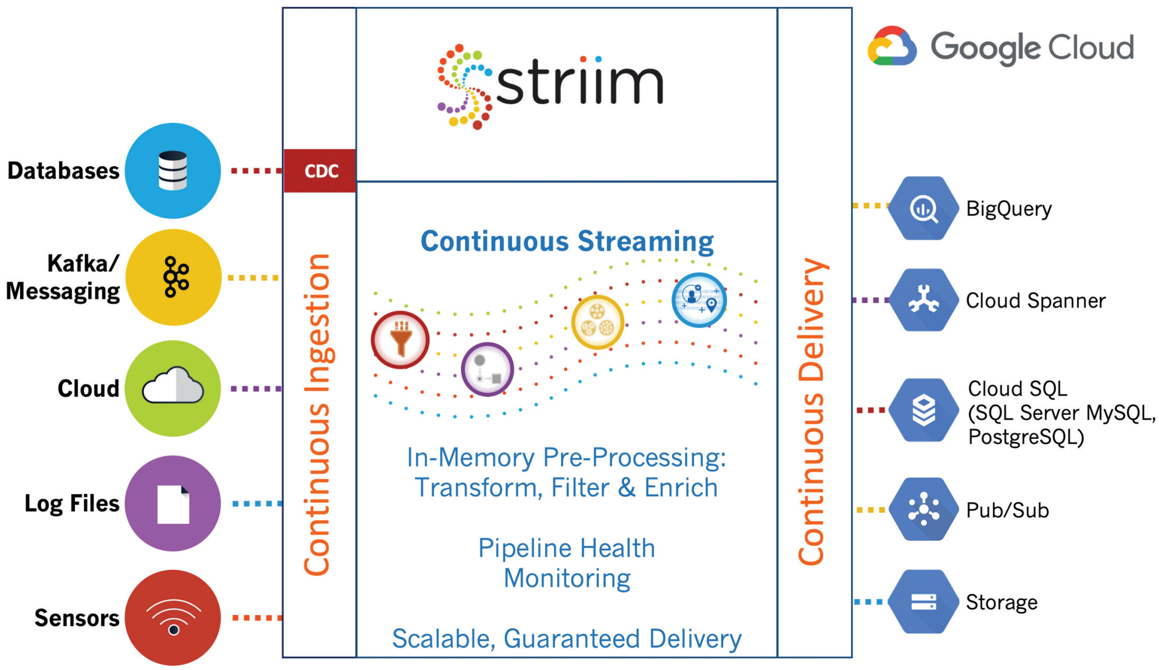
CDC enables to keep legacy systems active during the migration process to eliminate database downtime

Unlike traditional batch ETL, Striim continuously integrates data from on-premises and cloud data sources into the target environment on the Google Cloud. This can include large volumes of data from a variety of sources ranging from operational databases to data warehouses to log files and unstructured messaging data. Even very large schemas containing thousands of tables can be replicated into the Google Cloud.

Striim continuously filters, aggregates, transforms, masks, and enriches in-flight data using SQL-based queries and customer-defined operators before delivery to the cloud. Users develop integration rules via a template-based, drag-and-drop user interface. They can also create custom dashboards to visualize their data and preview source data as it streams to ensure that the necessary transformations are taking place. The result is that destination data stores like Google Cloud SQL always have up-to-date data in a usable form.

The Striim platform also enables customers to enjoy the full value of the Google Cloud databases. For example, data can be continuously migrated from sources like operational databases, data warehouses, and even Amazon Web Services storage to both Google BigQuery and Cloud Spanner – enabling both operational transactions and analytics to be performed with the most current information. Striim also integrates with Google Cloud Pub/Sub for reliable staging of real-time events and Google Cloud Dataflow streaming analytics service for immediate delivery to applications and data stores.

Real-time integration also enables customers to unlock the power of Google's market-leading line of machine learning and artificial intelligence development tools. Customers can continue to run their on-premises environments while migrating to the cloud at their own pace. They can choose to develop and deploy solutions entirely on the cloud, entirely on-premises, or develop on the cloud and deploy in a hybrid on-premises/cloud topology. At the same time, they can tap into the benefits of predictive analytics and machine learning using the most sophisticated tools available anywhere.



The Striim platform offers real-time data ingestion, stream processing, and continuous delivery to major Google Cloud services

In markets characterized by nearly constant change, organizations can no longer afford the luxury of batch processes. As mentioned earlier, leading analyst firms expect that within two years more than half of major new business systems will incorporate continuous intelligence using real-time data. In order to build these next-generation applications, the firm advises businesses to master the technology and skills of messaging, business process management, API management, stream analytics, machine learning, decision modeling, business rule management, and prescriptive analytics.

Building the Business Case for SDI

The shift from batch ETL to streaming data integration enables businesses to offload their high-value, operational workloads to the cloud because they can easily stream their existing data to cloud-based analytics solutions and new business systems in real time and in the right format. They can also rapidly build real-time analytics applications that make use of a combination of sources like Apache Kafka, NoSQL databases, and Apache Kudu analytics.

The growing power and sophistication of cloud platforms have made state-of-the-art database and analytics tools available to everyone, which is one factor fueling the cloud's 25%-plus annual growth rate. One could argue that any business that isn't taking advantage of cloud toolsets is at a competitive disadvantage today, making SDI an essential part of doing business. However, executives may want a more specific ROI case to make the SDI investment.

When building a business case, start with scenarios in which data integration is currently a hindrance to making optimal use of data, whether streaming or not. Search the internet to find statistics about average business impacts. Apply that data to your own internal metrics.

Here are some recommendations for building a business case.

Start With the Low-Hanging Fruit

The most immediate opportunities are in areas that are currently constrained due to poor data integration. Business decisions are delayed or not made at all because complete and current information isn't available. The business impact is well understood. Perhaps the lack of integration has caused delays in closing books or completing an acquisition. Business managers may have expressed frustration about data quality or timeliness. These are cases in which the need for more current information is well understood, as is the impact on the business. Decision makers are likely to be receptive to any solution that breaks the logjam.

Reduce IT Costs

In a perfect world, business decisions would be driven by opportunity, but the reality is that cost savings are more likely to get the attention of the CFO in the short term. Given the severe shortage of IT talent in today's market, people costs are a significant area of opportunity.

Look at how the data integration process is currently managed in your organization. If software engineers are spending days or weeks on ETL, it's likely the company can use their talents more productively. Legacy platforms have been slow to adopt SDI features, and according to industry analysts, many still don't connect with popular event data sources like the MQTT open messaging protocol, the Constrained Application Protocol for IoT devices, Apache Kafka, and AWS Kinesis. A modern SDI platform that enables a drag-and-drop approach to build integration rules can significantly improve developer productivity and even allows some rules definition to be delegated to business users.

If developers are spending time integrating open source solutions, consider the people's time involved. While open source software may be free to download and install, the costs of training developers and administrators to use it, integrating with operational systems, applying patches and updates, and keeping current with new versions can easily outstrip the cost of a packaged solution. The financial impact of building an enterprise-grade streaming integration platform goes far beyond license fees. A significant investment is required to integrate multiple components into a solution that is secure, scalable, and reliable.

Reduce Operational Costs

By incorporating data from sensors into operations, organizations can often see significant cost savings. For example, the American Council for an Energy-Efficient Economy has estimated that smart technologies, which adjust lighting and temperatures based upon streamed sensor data, save an average of between 8% and 18% of energy costs in midsize buildings.[10] PricewaterhouseCoopers reports that the use of predictive maintenance technology in factories – another popular streaming application – yields an average overall cost reduction of 12%, a 9% percent improvement in uptime, a 14% reduction in safety and health risks, and 20% longer asset lifetimes.[11]

An organization with a field service operation could use fleet optimization software employing live GPS feeds and sensor data to increase productivity an average of 15%, while reducing fuel costs more than 20%, according to Frost & Sullivan.[12]

Chances are there are applications of IoT that are relevant to your business.

Enable Your Cloud Strategy

Any organization that is undergoing a digital transformation must factor cloud computing into its plans. However, migrating to the cloud often takes more time and costs more money than anticipated. As noted in chapter 3, database migrations take weeks or even months, during which time IT staff members are tied up with plumbing and not working on projects that move the business forward.

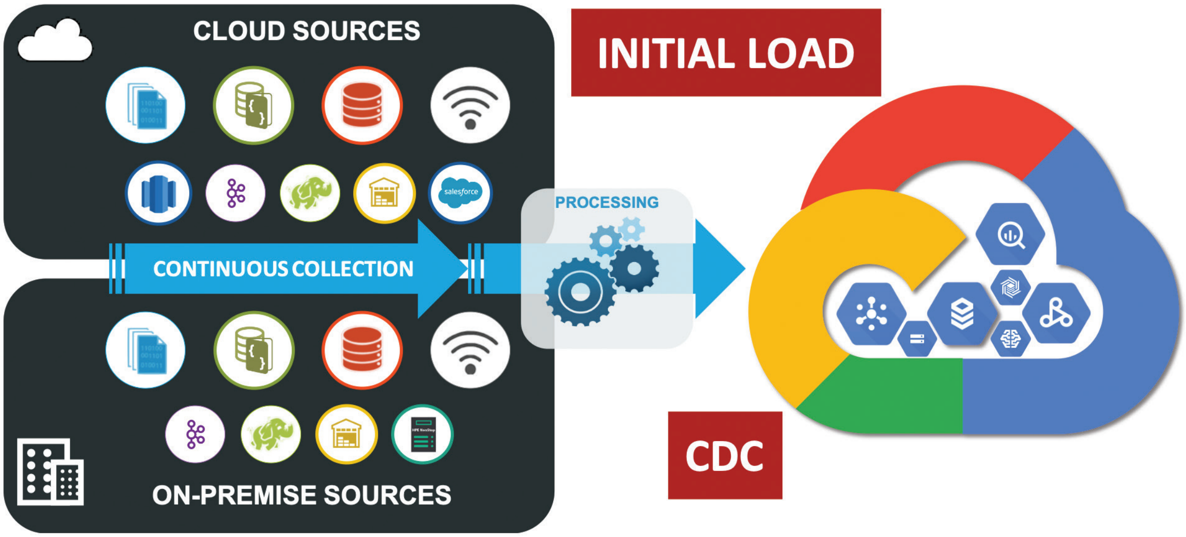
An SDI solution can enable database migrations to be completed seamlessly, with minimized risks while supporting continuous data movement and processing in a hybrid cloud environment on an ongoing basis. Compare the people costs of taking IT staff offline for a data migration effort to the cost of automating the process with a commercial SDI service.

---

[10] Perry, Christopher. "Smart Buildings Save Energy and Improve Occupant Comfort" *ACEEE*. Dec 20, 2017 https://www.aceee.org/blog/2017/12/smart-buildings-save-energy-and

[11] PwC and Mainnovation. "Predictive Maintenance 4.0 Beyond the Hype: PdM 4.0 Delivers Results" Sept. 2018 https://www.pwc.be/en/documents/20180926-pdm40-beyond-the-hype-report.pdf

[12] GeoTab Africa."Concrete Business Benefits of Fleet Management" Feb. 27, 2019. https://geotabafrica.com/business-benefits-of-fleet-management/



Streaming data integration combines initial load with CDC to enable online database migration with zero downtime

Your estimates should also factor in the cost of downtime, which Statista estimates is over $300,000 per hour on average.[13] The cost of idling production databases for the time needed to complete a migration can easily dwarf the cost of software that makes the process invisible.

Challenge Assumptions

It's tempting to assume that certain costs or constraints are baked into the business when some creative thinking can change the equation. Brainstorm ways in which access to real-time data could yield efficiencies or new business opportunities that aren't readily apparent.

Learn from early adopters and those in other industries. For example, in an e-commerce environment, recommendation engines that suggest companion products or offer discounts based upon the contents of a shopping cart have demonstrated value. McKinsey has estimated that 35% of Amazon's e-commerce sales are the result of its real-time recommendation engine. To estimate your own business value, calculate the impact that a 10% increase in average order size would have on profitability, and compare that to the cost of the software required to enable the necessary functionality.

Look for correlations that aren't obvious. For example, it's a well-known fact that the longer visitors linger on a website and the more pages they visit, the more likely they are to become customers. Real-time analytics can help spot these engaged visitors and extend offers that turn them into qualified prospects while they are still on the site. Similarly, historical analytics can help e-retailers detect signs that a visitor is about to abandon a shopping cart. These insights can be compared to live streaming data to intercept shoppers before they leave or extend an offer by email that brings them back to complete the transaction.

[13] Statistica. Average cost per hour of enterprise server downtime worldwide in 2019. https://www.statista.com/statistics/753938/worldwide-enterprise-server-hourly-downtime-cost/

 •

Consider Customer Experience

While estimating the bottom-line value of improved customer experience can be tricky, it's worth keeping in mind that research has shown that CEOs consistently rank this objective as one of their top two business priorities.

And there are ways to measure the value of improved customer experience. One metric every business should understand is customer lifetime value (CLV), or the average amount that a customer spends with the business over the lifetime of a relationship. CLV can be cross tabulated in different scenarios, such as the average amount spent by one-time customers versus those who return for a second visit or by the type of products they purchase. Other metrics include Net Promoter Score, customer acquisition costs, churn rates, and responses to satisfaction surveys. There is also ample research that demonstrates repeat customers do considerably more business over time than one-time visitors, and that the cost of acquiring new customers is many times higher than the cost of keeping an existing customer happy.

Earlier, we described the benefits that companies have realized through innovations like overnight delivery, package tracking, and instant matching of drivers and passengers in ride-sharing scenarios, all of which are made possible by streaming data. These capabilities are now available as services and can be quickly employed to deliver value without the costs of building them from scratch.

There are many other ways streaming data can be incorporated into an enhanced customer experience. For example, customers who call a support line are easily irritated by the need to repeat information to multiple customer service representatives (CSRs). Streaming data integration can match caller ID information with customer records pulled from several sources in real time to give CSRs a complete view of the customer's history and preferences to make support calls more efficient and pleasing.

Streamed data from connected devices can also identify product malfunctions in the field or interactions that indicate a customer is having difficulty using a product. The manufacturer or installer can use this information to intervene before frustration levels rise, creating a more pleasant customer experience.

Many organizations mine data on social networks such as Twitter and Instagram to look for references to their company and respond immediately with an acknowledgment or an offer. Auto repair shops can respond to tweets referencing mechanical problems and hotels can welcome visitors to their city. The evidence of business value in these cases is often anecdotal, but the investment to monitor the sources and trigger alerts is trivial.

 •

CHAPTER 6

What To Look For In An SDI Platform

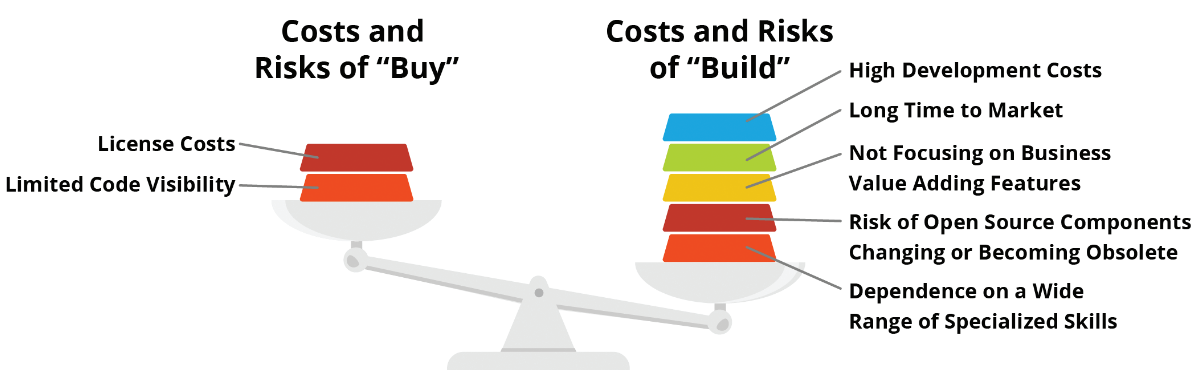
The projected rapid growth in the market for SDI frameworks and tools, combined with fast growth in data volumes, has created a bounty of choice for prospective users. It has also created confusion when it comes time to choose the right category of tools, software, and platforms. The result can be a rat's nest of solutions that simply don't work well together.

This situation has made the process of choosing strategic platforms more difficult. Because a comprehensive SDI solution is so central to data-driven organizations, making smart decisions is critical.

Myriad of Choices

Choices range from build-your-own solutions to managed services to SDI in the cloud. There are tradeoffs to each approach.

- Licensing costs can be minimized by using open source streaming frameworks. Integrating the software into an existing environment and customizing it is entirely in the user's hands. For that approach to work, organizations need internal staff expertise that can be scarce and expensive. If time to market for streaming solutions is important, this option may not be the best one.



Building with open source comes with hidden costs that outweigh the license cost savings

- Organizations that already have one or more data integration tools in-house may choose to work with existing vendors when specific integration use cases demand the combination of SDI with other integration modes, such as batch ETL. The risk is that not all legacy vendors support the full range of streaming platforms, if they offer streaming support at all. This approach can also become cumbersome if multiple vendors are involved.
- Another option is to engineer unique solutions for each data stream on a case-by-case basis while the organization figures out if it will require real-time analytics as the norm. This is a costly option, particularly if one-off integrations later need to be adapted to a commercial SDI system, but it can be a good choice for organizations that aren't ready to commit.
- A specialized commercial streaming integration platform is an option to consider if existing vendors don't provide the full required set of SDI capabilities. The risk is in introducing another variable into the data management architecture.

- Organizations might also want to consider one of the currently available multipurpose stream processing platforms, which are gaining considerable traction with organizations aiming to build and implement data management strategies that combine real-time and batch data. This approach enables organizations to quickly realize and leverage SDI even if they have little experience with the technology.

Benefits of an All-in-One SDI Platform

Combining stream processing and integration in a single platform has significant benefits for organizations that are laying the groundwork for deeper analytics and improved operational intelligence. Think of these platforms as the infrastructure that enables an ever-widening array of operations as streaming data sources proliferate.

Depending upon the platform, these multipurpose engines can undertake a variety of tasks, such as pattern detection, data filtering, multi-stream data joins, and even a broad spectrum of built-in real-time analytics. Their APIs can be used to initiate instant response to events as well to integrate with existing dashboards.

Not all of these platforms are created equal, however, which can make choosing the right one a tricky process. For example, some include analytic features, while others don't. Support for popular data sources and destinations varies, as does compatibility with a broad range of sources and transport protocols (such as TCP, UDP, HTTP, MQTT, and AMQP) and message buses such as JMS, MQ Series, and Flume.

Recommendations for Researching and Choosing Vendors

Generally speaking, if an organization's business requirements involve some combination of analytics and data integration, a multipurpose platform is a good idea owing to its ability to handle both of these functions within a single code base, user interface, and API set. This obviates the need to patch together new solutions that may not work well with others in use.

Here are some additional considerations that can help decision-makers choose the right SDI platform.

Technology:

- Look for support for continuous data integration from as many data sources as suit your needs. Platforms that over-focus on certain data sources or transport protocols may miss out when new sources and transport protocols are added, inhibiting organizational agility. Be sure the platform you choose is built to enable real-time low-impact data ingestion both from traditional relational databases as well as from non-traditional sources, including log files, IoT sensors, Hadoop, NoSQL databases, and messaging systems like Kafka.

- Ensure that any platform under consideration is fit for advanced stream processing tasks, such as transformation, aggregation, and filtering, as well as streaming analytics. It is also essential that the platform enables in-memory, SQL-based stream processing, so you can build applications that deliver the data in a consumable format with sub-second latency without requiring specialized skills.

- For business-critical use cases, the platform should support your requirements with built-in exactly-once processing, security, high-availability, and scalability using a distributed architecture.
- Look for a platform that deftly manages data flows, both end-to-end and in memory. Such a platform will be able to ensure data processing and delivery is consistent across all flows, so there are no missing data or out-of-sync situations in target systems.
- Ask if the platform is equipped with enhanced cloud integration capabilities, including the new line of cloud-native databases and data warehouses.
- Understand that cloud-focused streaming platforms carry with them the potential for cost reductions with respect to both developing and supporting streaming applications, especially those involving unusually high throughput. Furthermore, cloud-focused platforms by nature leverage the many ancillary benefits of the cloud, such as disaster recovery, self-provisioning, high availability, and continuous patching and versioning.
- Platforms should also support popular publish-and-subscribe (pub/sub) messaging platforms like Kafka. And they should enable data delivery to standard relational databases (cloud-based or on premises), Kafka data stores, and non-relational databases like NoSQL.
- The best platforms offer a fast time to market with wizards and drag-and-drop user interfaces.
- The platform should ensure continuous monitoring of data pipelines and delivery validation for zero data loss.
- For those organizations that are considering platform-as-a-service options, the provider should make its offerings available in popular cloud marketplaces. In addition to integrating with the host cloud platform, marketplace software can be provisioned quickly for evaluation purposes, thereby minimizing the time required for setup and testing.
- The more advanced SDI platforms now feature visual data discovery and analytics visualization capabilities. These allow users to search for and filter historical data as well as streaming data from multiple sources for a wide variety of applications, such as anomaly detection, without requiring extensive coding skills.

Company:

- Research the management team of the SDI vendors you consider. Ask if they have deep experience in traditional data integration approaches with which your own staff is likely to be familiar. If so, inquire about how that deep experience comes into play in leveraging real-time data.
- Check customer references. The platform vendor should have a fast-growing base of customers in your vertical industry or business function.
- The vendor should be able to demonstrate the uniqueness of its platform’s value proposition compared to other choices in the market.
- Ensure that the platform vendor clearly demonstrates a focus on continuous integration, data ingestion, and data processing.

- The platform vendor should be able to clearly explain how it adds value to any stream-processing services your organization may already be using.

With the SDI boom continually giving birth to new vendors and products, the difficulties of deciding which platforms to consider can be daunting. Focus on SDI vendors that demonstrably integrate their solutions with specific offerings from major cloud vendors. The best platforms will be those that can demonstrate the ability to continually evolve with the market and support new cloud services as they emerge.

CONCLUSION

The World is Streaming

Business happens in real time. As we have pointed out throughout the course of this eBook, the limitations of legacy systems have held organizations back from taking full advantage of the real-time nature of their interactions with customers, partners, and suppliers. These restrictions no longer need to apply. Cloud computing has made world-class data analytics tools available to everyone. The technologies needed to capture data streams and turn them into business value are now available with just a few clicks in cloud marketplaces. Organizations that attempt to do business in a real-time world with legacy data integration tools are fighting with one hand tied behind their backs.

Striim provides real-time, continuous streaming integration both on premises and in the cloud. Its unique change data capture approach to data replication and distributed architecture support database migration to the cloud with zero downtime. Advanced monitoring and alerting features complement its SQL-based stream processing and in-line analytics of real-time data. Striim supports all of the most popular source and destination platforms to enable enterprise-wide use of streaming integration.

To learn more, or to start a fully-functional, free trial of the Striim platform, please visit **www.striim.com**

To schedule a demo, please contact us at **info@striim.com**