



Processing and Analyzing Streams *of* CDRs in Real Time



An Explosion of Data, A Block to Intelligence

“The V of Big Data Velocity means both how fast data is being produced and how fast the data must be processed to meet demand.”

Gartner

Communication Service Providers (CSPs) are facing an explosion in data volume, velocity and variety, yet are struggling to transform this data into business value. IP networks and services, user devices, location-based apps, 4G and small cell networks are all generating new Big Data at unprecedented scale. The potential for Big Data is widely recognized, particularly if the data can be transformed into operational intelligence and acted upon in real-time.

We now live in a world where the consumer expects real-time responses, where network and service capacity can be configured dynamically based on demand, and where slow responses to revenue and customer experience issues are no longer tolerated. Real-time, actionable intelligence is therefore required for the next generation of customer experience, revenue management, network and service performance and fraud detection applications.



Goal:

to extract the maximum value from Big Data streams by driving real-time decisions, actions and alerts.

Real-time Operational Intelligence

Most operational decisions are still taken after the fact based on reports generated by offline systems. Existing OSS and BSS systems tend to be highly customized and tailored to specific use cases, with architectures that are not designed for the performance requirements of real-time Big Data applications. This results in high total cost of performance when faced with real-time data streams – high hardware costs, high TCO for maintenance and upgrade, high integration costs between multiple data siloes, and overall, poor Big Data performance.

In contrast to traditional Business Intelligence platforms, streaming analytics delivers real-time intelligence and alerts with near-zero latency. Arriving data streams are processed in real-time without having to store the data, and also streamed continuously into the existing systems.

Streaming analytics drives real-time optimization of the customer experience, minimizes fraud, improves service quality and network performance, and drives new real-time revenue streams. In fact, with streaming integration and real-time analysis of all their Big Data streams, service providers can now assess and correct the profitability of their network and services in real-time.

Real-time Analytics from Streaming Machine Data

The core of SQLstream's streaming data management platform is a massively scalable, distributed engine for analyzing unstructured Big Data streams using continuous standards-compliant SQL. Machine data from networks, services, log files, sensors and devices are collected and analyzed 'on-the-fly' over moving time windows, without having to store first. Applications can be built quickly and easily as pipelines of streaming SQL queries, enabling ultra-low latency analytics to be extracted from the raw data streams.

	Streaming SQL	RDBMS SQL
Query Duration	SQL queries execute continuously.	SQL queries complete and exit.
Query Scope	SQL queries operate on arriving data over time windows (from milliseconds to months).	Ad-hoc SQL queries execute over stored data in a static schema.
Parallel Processing	Streaming SQL queries can be distributed in-memory over multiple servers.	Processing executed centrally over a single in-memory or disk-based repository.

Table 1: Streaming SQL and RDBMS SQL comparison

SQLstream s-Server is the only mature, 100% SQL-compliant solution for streaming data processing, transforming streaming machine into actionable operational intelligence in real-time. Unlike traditional platforms based on RDBMS technology, a streaming data management platform executes SQL queries over live data streams without having to store the data. Table 1 describes the primary differences between streaming SQL for fast data, and traditional SQL for Business Intelligence applications.

The results is the ability to extract the maximum value from Big Data streams by driving real-time decisions, actions and alerts.

The Benefits of Streaming Analytics

The effective use of streaming analytics from real-time network, service, device and consumer data can have a significant impact on operational expenditure and customer experience – the two key business drivers for today's service providers. Most service providers now agree that real-time operations are key to delivering against their business objectives.

Table 2 highlights some of the key characteristics required to utilize Big Data in real-time, comparing existing OSS/BSS systems with streaming analytics.

	Streaming Analytics	Legacy OSS/BSS
Operational Intelligence	Real-time, sub-second responses	Offline, high latency decisions in the order of hours or days
Total Cost of Performance	Low	High
Real-time Integration	Real-time continuous integration with Continuous ETL	Expensive and often limited integration between data siloes

Table 2: the Streaming Advantage

The key benefits of streaming analytics to communication service providers:

- **Lowest Total Cost of Performance** for real-time operations. Streaming analytics scales to process millions of records per second on commodity hardware. Also, with standards-based SQL (SQL:2008 and SQL:2011 compliant) as the streaming data processing language, streaming applications can be deployed and enhanced in a fraction of the time of traditional database-oriented systems.
- **Improve operational efficiency** through effective real-time decisioning, real-time analytics and streaming integration with operational systems.
- **Improve customer satisfaction** through proactive customer service assurance and real-time service quality monitoring.
- **Eliminate bad debt** and customer dissatisfaction through real-time fraud detection and prevention.

Streaming Analytics for CDR Processing

Every voice call and IP service in a telecommunications network generates usage records. These service usage records, Call Details Records (CDR) for a voice network and IP Data Records (IPDR) for IP networks, contain information about the call or session that is used for applications such as billing, service quality monitoring and fraud detection. Data record formats are controlled by standards, but unfortunately standards vary by industry. For example mobile, cable networks and IP SIP-based networks all have different reporting capabilities and requirements.

Real-time CDR Analytics

Most service providers still process CDRs in batch mode, usually once a day, but sometimes as infrequently as weekly or even monthly. Even where certain types of CDRs are processed more frequently, this is at best hourly. The reason is in part the performance restrictions of the underlying legacy platforms, but also a belief that real-time, sub-second CDR analysis is too costly. This may have been true with traditional RDBMS-based systems, but streaming analytics platforms enable millions of CDRs per second to be collected and processed in parallel.

Stream processing for CDR analytics applications offers the lowest total cost of performance of any approach – less hardware, lower software costs, with prebuild integrations and analytics libraries, significant lower implementation and maintenance costs.

The primary applications for real-time (sub-second) streaming CDR analytics include real-time call rating, real-time fraud prevention, service quality and customer experience monitoring, and least cost routing. Streaming analytics also enables real-time reporting of customer, service and vendor profitability.

Real-time fraud prevention

Call fraud that remains undetected can get passed through to customer bills and have a detrimental impact on customer experience. Even in most cases where fraud is detected, detection tends to be the following day or even after the weekend, resulting in direct revenue loss and additional OPEX expenditure to correct. Streaming analytics enables the immediate, real-time detection of suspect call activity based on data streams such as switch logs, IP spoofing events, unusual call destinations and unusual call usage patterns.

Streaming Analytics for CDR Processing

Real-time call rating

Rating is the process of determining the charge for each call or service session. The process involves CDR collection, removal of duplicates, session reconstruction, caller ID or IP verification, cost calculation based on the rate plan, application of discounts and finally storing the records in a CDR warehousing. Although rating is a core function of any billing operation, most rating engines still operate in batch mode due to scalability, data integration and rapidly changing business requirements. Streaming CDR processing and analytics can deliver real-time rating at scale, while persisting aggregated rating information continuously through to the core billing platforms.

Customer Reporting

Access to real-time billing information has a significant positive impact on customer experience. Streaming analytics for real-time rating, performance monitoring and policy control can be used to deliver real-time visibility of service performance, costs and discounts to the customer.

Optimization of Least Cost Routing (LCR)

A routing system provides the network path for a call or session based on quality requirements, discounting models and operator profitability. For example, the selection of the lowest cost interconnect partner for roaming calls in a wireless network, or for SIP-based SBC (Session Border Controller) networks, the selection of the lowest cost IP network path given the desired quality. Operational intelligence from streaming CDR analysis can be used to update least cost routing tables dynamically, either to improve a customer's service where quality has fallen below the SLA, or to increase profitability where quality can be maintained at appropriate levels.

Call performance monitoring

Most performance monitoring tools aggregate data by 15 minute intervals, usually aggregated by port or even by network device. However, most IP-based services are bursty by nature, where problems may last for a few seconds, but be sufficient to force a quality issue or dropped call. SQLstream can provide real-time monitoring of both network performance and CDR data in order to identify issues in real-time. Streaming intelligence can also be integrated with the LCR database, policy server or bandwidth managers in order to drive dynamic updates.

Streaming Analytics for CDR Processing

Real-time profitability analysis

Service providers calculate metrics such as gross margin based on comparing the cost to the customer versus the cost of service delivery for that customer. However, margin analysis is normally an offline activity based on long-term trend data. With streaming analytics, the profitability of a service offering, a customer type, or even an individual customer can be visualized in real-time. This enables real-time decisions to improve profitability, for example, through dynamic updates of the least cost routing algorithms.

Dynamic policy management

The Policy Manager or Server is responsible for (a) the selection of quality and routes and (b) actions to be taken (often by customer type against SLAs). For example when a call drops or has quality issues, or bandwidth limits are reached during the session, dynamic updates based on streaming analytics can drive corrective action to bring the service back within its SLA. Other policy-driven actions include the proactive notification of the consumer, or the application of discounts applied during or after the session.

Standard Format for CDR Warehousing

Much of the cost of CDR processing is the conversion of many different CDR formats and structure to a common format that can be used for further analysis. Streaming data management addresses this as part of the core stream processing pipeline, transforming all CDRs to a standard format in real-time, and delivering this as a continuous stream to existing CDR warehousing platforms.

The Value Proposition for Streaming CDR Analytics

The analysis of CDR data is at the core of a service providers business. CDRs and IPDRs offer essential information for call rating, service quality, consumer location and SLA performance. Yet many service providers are struggling to accelerate their CDR processing for real-time, sub-second latency. Performance limitations of existing CDR warehouses and high TCO for existing processing platforms are a major factor.

Big Data technology and streaming analytics offer a way forward. Streaming analytics enables large volumes of CDRs from all the different types of network and equipment to be collected and analyzed in real-time with a low overall total cost of performance. This enables service providers to:

- Eliminate bad debt and customer dissatisfaction through real-time fraud detection and prevention.
- Optimize call quality and service profitability through fast and accurate least cost routing decisions.
- Reduce churn and improve retention through better call quality and proactive customer service assurance.
- Reduce network and leased facility expenditure through better negotiation of trunking, interconnects and 3rd party network costs.
- Improve overall operational efficiency through real-time performance monitoring and dynamic, automated updates of corrective actions.

Streaming Big Data analytics offers a way forward, with Big Data scalability and with the lowest cost of real-time performance.

The Anatomy of a CDR stream processing application

Stream processing applications consist of one or more concurrent analytical streaming SQL pipelines. Each step in the processing pipeline is a streaming SQL operation that selects from either a machine data source or the output of any other streaming SQL node. Streaming processing pipelines are built from multiple processing nodes, each do a SELECT FROM and INSERT INTO a streaming View or external system. Essentially therefore a streaming SQL pipeline is build from a cascade of streaming SQL Views. Real-world streaming applications often consist of many interconnected streaming pipelines, all executing in parallel, processing the data as they are piped through over sliding time windows. Figure 1 illustrates a simplified example of a streaming CDR analytics pipeline taken from the customer case study described in the next section.

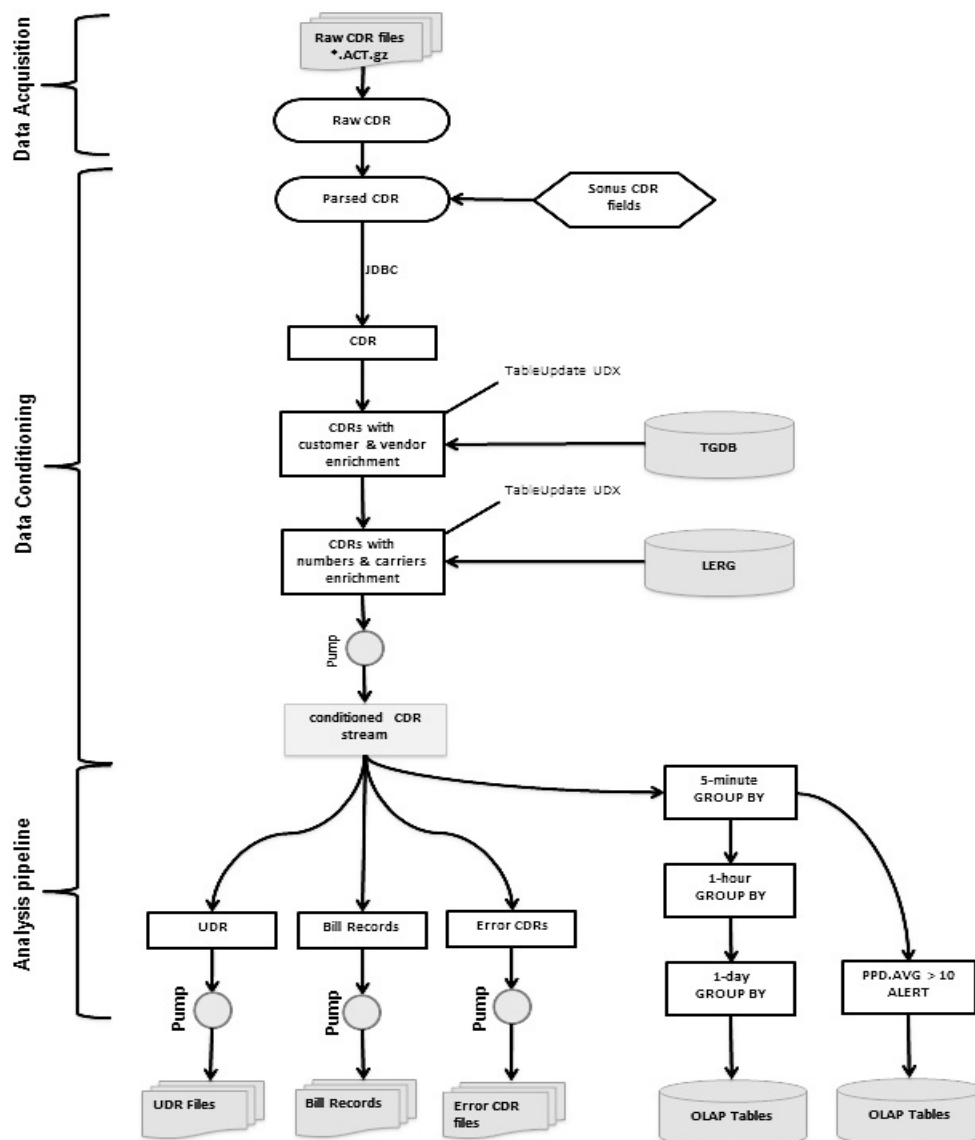


Figure 1: Example Streaming Pipeline for CDR Analytics

Data collection

The first stage in any streaming analytics pipeline is the data collection and conditioning. Different conditioning requirements may be necessary for different applications. The underlying data management architecture allows for data to be reused and directed into multiple processing pipelines without reprocessing any data.

CDR and IPDRs are captured as real-time streams of records.

Here, CDRs are available in both archived and unarchived log files. The Log Agent performs a Change Data Capture (CDC) on new records and new archived logs as soon as they are created. Each Log Agent (a lightweight Java-based agent) also performs remote filtering and conditioning, and can be programmed to deliver only the data records of interest.

Data parsing and conditioning

The core streaming analytics pipeline captures and combines all CDR streams. Duplicates are removed, CDRs combined for session reconstruction, and records transformed into a standard field format for further processing.

CDR enrichment

The streaming platform combines the processed CDR records with external data warehouse and RDBMS data as they flow through the system. Typically this consists of additional customer and vendor information. External JOINS with static, stored data can be cached in memory if required. This may be necessary if the speed of the processing pipeline exceeds the read latency of the external system.

Continuous Aggregation and Integration

CDR records can be aggregated in parallel over different time windows. For example, aggregated 5-minute, 15-minute and 60-minute streams are updated incrementally as new records arrive, and streamed continuously to existing CDR warehouses and billing systems.

Analytics and Alerts

Real-time analytics and alerts are also executed in parallel with any conditioning, filtering and aggregation. The performance of the platforms means results are updated immediately as new records arrive.

Case Study: SIP Trunking Provider

The customer provides SIP trunking services and on-demand cloud-based IP communications across markets such as enterprise networks and contact centers. The customer was struggling with data volume as their business grew, and with the excessive operational costs required to maintain multiple in-house solutions. In summary, their problems were:

- ✎ Queries on the legacy CDR data warehouse could not deliver real-time performance, often taking many hours to complete.
- ✎ Cost of ownership for the existing application was high, both hardware costs and also the cost of maintaining highly customized and bespoke analytics and system integrations.
- ✎ Lack of flexibility, existing systems too rigid and difficult to change for new CDR data sources, CDR formats and analytics requirements.

This led the company to evaluate alternative solutions. After an exhaustive search of traditional telecoms vendors and solutions, none of which offered improved business agility and lower cost of ownership, they looked to the wider Big Data market and selected SQLstream. They determined that streaming SQL analytics offered the scalability required with the lowest overall cost of performance. SQLstream offered:

- Increased operational efficiency through real-time collection, processing and integration of all CDR data on a single platform.
- Reduction in operational expenditure through the replacement of multiple legacy applications.
- Future-proof, agile platform, with low system lifecycle TCO.
- The tools and training for the customer's IT team to take responsibility for the system.

The initial priority was to deliver the Big Data scalability and performance required to process their increasing volume of raw CDR and IPDR records. With SQLstream's training and expertise, the customer's IT team deployed the SQLstream s-Server platform and began to process streaming CDR records immediately. They deployed real-time analytics and monitoring applications, and the streaming aggregations and transformations for integration with downstream applications. These existing applications included the CDR warehouse, billing platforms and operational assurance and alerting systems.

The project delivered against the customer's success criteria:

- **Performance.** The customer's initial requirement was for a minimum of 6,000 records/second on a single 4-core CPU, with scale-up capability as both data volume and analytics load increase over time. SQLstream immediately delivered more than 15000 records per second with all data processing, streaming integrations and analytics executing simultaneously.
- **Intelligent data collection.** Agent-based CDR data collection architecture, capable of filtering unnecessary records at source (and identifying new devices and the resulting log files dynamically).
- **Standard CDR data format across all record types.** Combine all CDR formats into a common Universal Data Format (UDR) and stream out into comma-delimited files for bulk loading into a backend billing system.
- Analytics, including real-time processing and rating of CDR data, and real-time exceptions and alerts based on patterns and thresholds.
- CDR aggregations and warehousing. Deliver continuously updated 5-minute, hourly and daily aggregations from real-time CDR feeds and persist to the existing CDR data warehouse.

A key requirement was the enablement of the customer's IT team to deliver and deploy solutions based on the SQLstream s-Server platform. The team undertook the SQLstream Developer Accreditation course and was able to undertake successfully the rollout of real-time streaming analytics, data collection, and continuous integrations with external systems.

Implementation

Figure 1 illustrated the simplified view of the streaming pipeline deployed in the initial project. CDR data flows through the pipeline and is transformed continuously and in real-time as new records arrive. The stream processing pipeline consists of four concurrently executing stages:

- Data acquisition, the top of the pipeline, where streams of raw CDR records are collected.
- Conditioning, where CDR records are parsed, transformed and enriched.
- Streaming analytics, the execution of customer's business logic and rules.
- Delivery, the bottom of the pipeline, where streams of conditioned data and analytics can be visualized and integrated with the existing external systems.

Real-time CDR Acquisition

CDR data is collected in real-time from log files. SQLstream's remote log agent architecture collected CDR records in real-time from the log files generated by the network equipment. Each log agent also performed parsing at source, thereby reducing the overall bandwidth requirements for data backhaul by filtering out unwanted fields.

CDR Conditioning

Conditioning transforms the source data into a form required to perform the necessary analytics. Data is parsed, cleansed, transformed and joined with relevant external data (enrichment). Parsing was also dependent on the message type (i.e. START, STOP, INTERMEDIATE, and ATTEMPT).

Streaming Enrichment

The streaming CDR records are enhanced with customer, telephone number and vendor information from existing systems. External lookups are executed in real-time against two external databases using user-defined operations (UDXes). External data can be cached if necessary should the streaming pipeline execute exceed the lookup latency.

Real-time Analytics and Streaming Aggregation

CDRs are aggregated by message type and over 5-minute, hour, and daily time periods in parallel, calculating metrics and related alerts that include:

- Answer-to-Seizure Ratio
- Post Dial Delay
- Minutes of use
- Average Call Hold Time in minutes.

Error handling

All CDR records that have flagged an exception are streamed to an error handler for later analysis. Errors include failed enrichment lookups (missing data for example), unmatched records and corrupt data formats. Errors can be investigated, repaired and replayed if required.

Continuous Integration

CDRs are processed into several aggregated output streams. Output streams are updated continuously, with data being delivered to log files, or written directly into external systems through the available APIs. Output streams include:

- Universal Data Records (UDRs) – a common record format for all CDR records, including enrichment data, written to an existing data warehouse for longer term trend analysis.
- Billing records, based on the UDR format, streamed continuously to existing billing platforms. Billing records also contain additional attributes used to measure latency.
- Aggregated UDR streams for continuous update to the data warehouse's OLAP cubes.
- CDR enrichment errors for further analysis and replay.
- Real-time alerts.



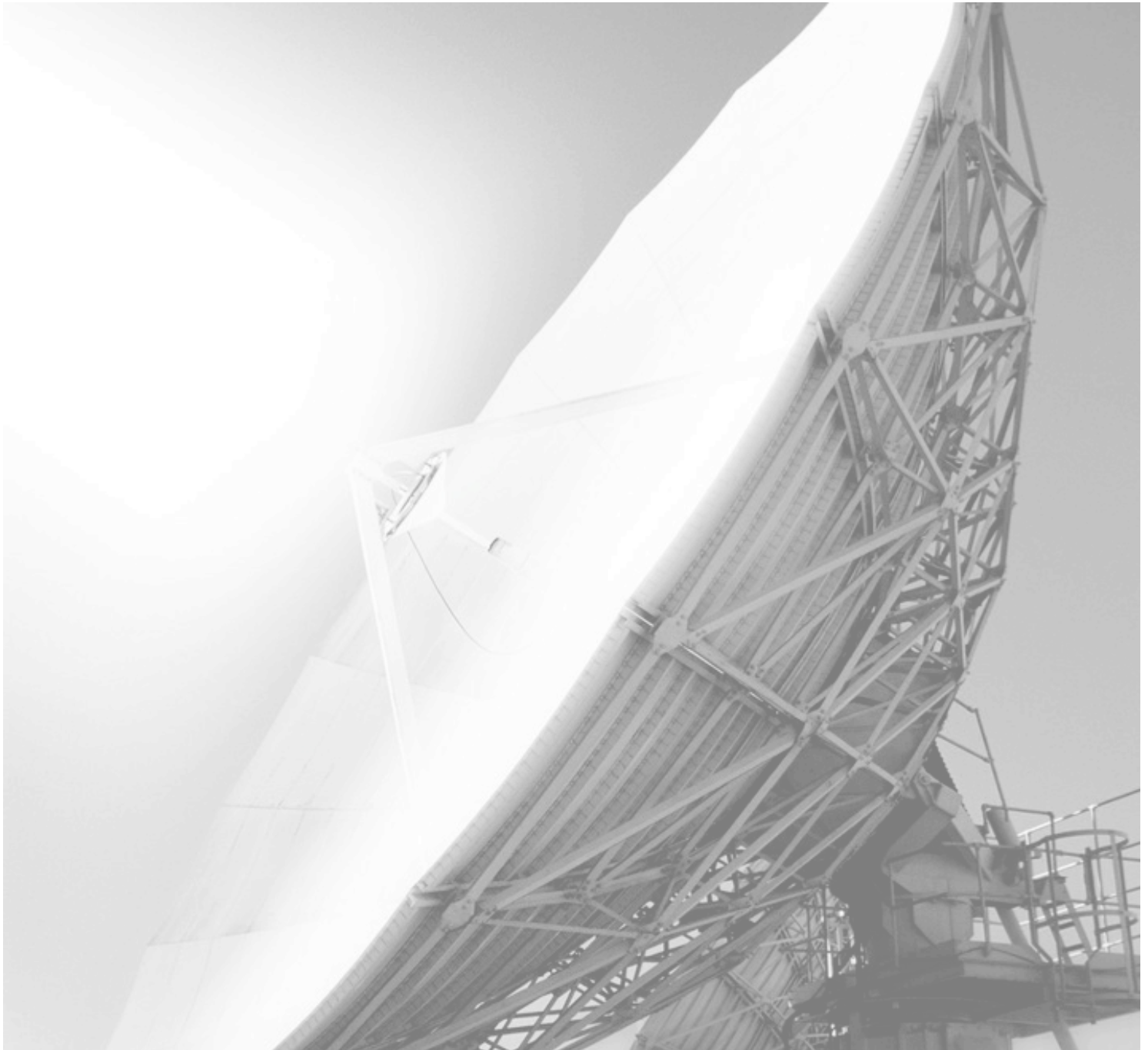
Conclusions

Streaming analytics represents a new dimension in Big Data management for communication service providers. Struggling with multiple data siloes, bespoke legacy applications with high TCO, and system performance that is unable to scale for today's real-time Big Data requirements, service providers are looking to the wider world of Big Data technology for the solution.

Streaming analytics offers a way forward. Streaming data management captures and processes real-time data in memory, without having to persist the data first. Many of the traditional cost and system integration issues are eliminated, and real-time decisioning based on sub-second latency is guaranteed for the most demanding of Big Data problems.

For real-time CDR processing, SQLstream's streaming data management platform offers applications such as real-time fraud prevention, real-time rating, real-time service profitability analysis and dynamic updates for least cost routing tables.

For the service provider, this means reduced operational expenditure, improved operational efficiency and improved customer satisfaction. SQLstream also offers the lowest Total Cost of Performance for real-time operations of any technology or platform. Streaming analytics scales to process millions of records per second on commodity hardware, and with standards-compliant SQL as the streaming data processing language, streaming applications can be deployed and enhanced in a fraction of the time of traditional database-oriented systems.



SQLstream, Inc.

1540 Market Street
San Francisco, CA,
94102
www.sqlstream.com

SQLstream delivers the only true standards-based SQL stream processing platform for real-time data wrangling, analytics, and action. Global enterprises use SQLstream Blaze to build streaming applications in minutes to keep operations running at optimal efficiency, protect systems from security threats, and support real-time customer engagement. With SQLstream Blaze, companies can go from raw data to actionable insights easily, continuously, and in real time, and compete effectively in the fast-changing business world. In 2016, AWS announced that Amazon Kinesis Analytics is based in part on technology components licensed from SQLstream.

