

Leveraging mainframe data for modern analytics

Confluent and Attunity provide seamless mainframe integration and query offload for modern, distributed analytics environments via Apache Kafka

Large enterprises, government agencies, and many other organizations have long relied on mainframe computers to deliver core systems managing some of their most valuable and sensitive data. Mainframes remain important to business today and are likely to continue to be critical infrastructure for many organizations for the foreseeable future. As important as these systems are, however, they can be difficult to integrate into today's data-driven, analytics-focused business processes as well as the environments that support them.

Most businesses looking to integrate mainframes into a broader analytical environment take a brute force approach, querying directly into the mainframe system to access the data they need. This approach can be costly in environments where billing is based on how many MIPS (millions of instructions per second) the system uses, because each new query eats up more instructions, adding to that month's bill. Additionally, a significant amount of tuning and optimization is typically required to get a data warehouse sitting on a mainframe to support the kind of broad, deep, and fast analysis that businesses need today.

One solution for a business facing these challenges is to offload the data from the mainframe to another environment such as Apache® Hadoop™ or an MPP database. Mainframe offload is an effective approach for some business environments, but it involves an expensive and time-consuming process to implement. And more often than not, the result of all that effort is just a new data silo. Businesses need to produce results from multiple data types and sources in real time. Doing so requires a new kind of solution—one that keeps mainframe data current and available to the broader ecosystem without all the environmental complexity and without breaking the bank.



Attunity unlocks mainframe data

Attunity Replicate is a software solution that provides an efficient and cost-effective way to bring mainframe data into a modern analytics environment with its change data capture (CDC) technology for DB2, VSAM, and IMS. Attunity Replicate provides low-latency and low-impact data integration for mainframe databases. With Attunity Replicate, you can extract and source mainframe data efficiently in real-time, and deliver those changes to a data warehouse or another enterprise analytics environment. The software replicates data continuously with low latency, providing a real-time alternative to a “lift and shift” approach to moving data

Attunity Replicate enables change record filtering and supports a wide variety of data sources and platforms. It supports transactional delivery for replicating into relational databases, batch-optimized delivery for data warehouses, message-oriented delivery for application and streaming based integration platforms, and change audit trails to facilitate other forms of integration. The software can be easily installed and configured using an intuitive, wizard-based GUI – so no extra coding is required.

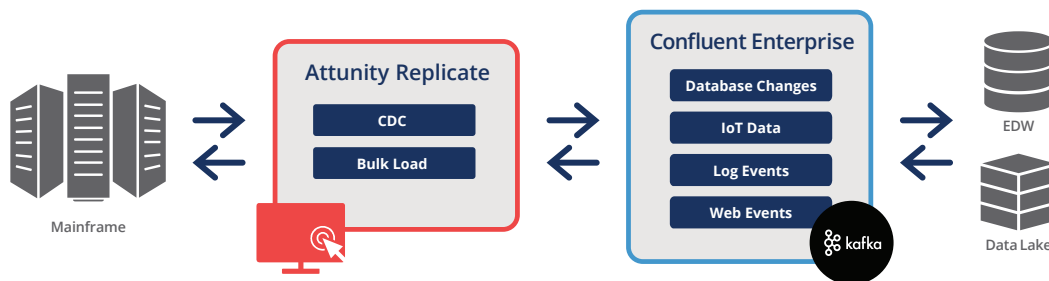
Because it replicates data continuously, Attunity Replicate eliminates the need to move data in periodic batches, keeping that data in sync and providing real-time access to the data in other platforms. Attunity Replicate provides log-based CDC for DB2 (leveraging the DB2 journals) as well as for VSAM and IMS, rather than performing repeated brute-force queries into the data. This provides a unique, non-invasive approach to capturing changes which does not incur the hefty price tag in MIPS described above. Attunity Replicate enables event-driven, stream-based processing via Apache Kafka.

Kafka brings mainframe data to life

Clients can leverage Kafka’s modern distributed architecture to move mainframe data in real time. Unlike using a Hadoop or database offload, which will first store and index data so that it can be accessed by querying, Kafka captures data for processing in flight. Although it may look like a messaging system — with producers publishing messages that are available to consumers in milliseconds — Kafka works more like a distributed database for your mainframe and other data.

When you write a message to Kafka, it is replicated to multiple servers and committed to disk. Kafka provides a core commit log for the entire distributed environment, persisting the data so that any number of systems can consume it at any time. Kafka’s consumers perform the database read function while its producers — in this case, your mainframe environment — perform the database write function. Once the data is published, there is no need to incur the cost in time and MIPS associated with going back to the mainframe.

Kafka’s unique architecture enables its fast performance. A single Kafka broker can handle hundreds of megabytes of reads and writes per second from thousands of clients. Each broker can persist terabytes of messages without impacting performance.



Confluent makes Kafka enterprise-ready

The Confluent Platform extends Apache Kafka to deliver the full promise of a modern streaming platform. Confluent makes deploying and managing a Kafka environment easy, reliable, secure, and auditable – bringing your mainframe data into a modern analytics environment while providing key services to make the open-source Kafka software truly enterprise-ready:

- **Confluent Control Center** - provides a comprehensive monitoring and management system for Kafka. It enables users to track both the flow of data through the cluster as well as important details regarding resource consumption.
- **Auto Data Balancing** - optimize resource utilization and reliability. The rack-aware rebalance algorithm optimizes for disk space utilization. Partition movements are executed with minimal impact to your traffic.
- **Multi-Datacenter Replication** - use to replicate Kafka for a failover data center (active/passive), synchronize two active data centers (active/active), or replicate to a global data center.
- **Schema Registry** - the Confluent Schema Registry provides a central repository for the data syntax of each Kafka topic. Applications can easily leverage the service to maintain consistent formats within their data pipelines.
- **REST Proxy** - the Confluent REST proxy provides access to Kafka from any network connected devices that can connect with HTTP.
- **Connectors** - the Confluent Platform includes several pre-built connectors based on the Kafka Connect framework, enabling users to deploy data pipelines between basic JDBC and HDFS data systems with a few simple configuration options.
- **Clients** - supplementing the pure Java clients available in Kafka, Confluent provides native clients for other popular languages: C/C++, Python, and Go. These clients support the full range of Kafka producer/consumer API's.

Confluent also provides 24/7 enterprise support for critical bug fixes, indemnification, training, and operational advice to customers.

It's a new day for the mainframe

With Attunity Replicate, you can unlock your mainframe data without incurring the complexity and expense that come with sending ongoing queries into the mainframe database. With Apache Kafka, you can deliver that data in real-time to the most demanding analytics environments. And with Confluent, you can ensure that your analytics environment includes the broadest possible range of data sources and destinations, while ensuring true enterprise-grade functionality. The combined solution puts mainframe data right at the heart of your modern, distributed analytics environment.