



www.informatica.com

2100 Seaport Blvd
Redwood City, CA 94063, USA
Tel: +1 (800) 653 3871

Informatica Stream Processing

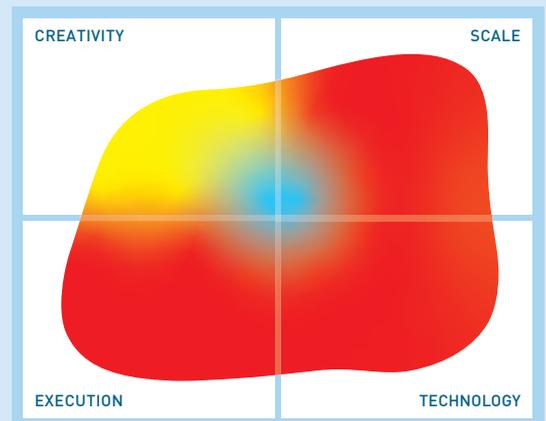
“ Informatica Cloud Mass Ingestion allowed us to generate hundreds of mappings in a very short time. It’s a straightforward, secure bridge from source to target, which is exactly what we need. We don’t require a VPN in order to maintain data security. ”
University of New Orleans

The company

Informatica is a private software company that covers a broad range of data management capabilities. It has more than 10,000 customers, operates world-wide and across all industries, and has revenues in excess of \$1.3 bn.

What is it?

Informatica offers stream processing as part of Intelligent Data Management Cloud (IDMC), a broad, cloud-ready, and well-integrated data management platform. IDMC also includes a number of other services, a unified metadata and AI layer (CLAIRE), and over 10,000 metadata-aware connectors that cover all three major public clouds (among other things). In addition, Informatica has



The image in this Mutable Quadrant is derived from 13 high level metrics, the more the image covers a section the better. Execution metrics relate to the company, Technology to the product, Creativity to both technical and business innovation and Scale covers the potential business and market impact.

What does it do?

Informatica’s solution for stream processing consists of several different products and services that combine within the singular platform of IDMC. The backbone of this solution consists of three services: Cloud Mass Ingestion, High Performance Messaging for Distributed Systems, and Data Engineering Streaming. Respectively, these provide streaming ingestion, high-speed messaging, and stream processing. Other Informatica offerings, such as Cloud Data Integration and Enterprise Data Catalog, can then add to this core. Taken as a whole, IDMC lets you ingest streaming data and move it to wherever it needs to be, while processing, transforming, and governing it as necessary.

More specifically, Cloud Mass Ingestion provides format-agnostic data movement and mass data ingestion, including file transfer, CDC (Change Data Capture) and exactly-once database replication. It also offers mass streaming ingestion from a variety of sources, complete with real time monitoring, alerting, and lifecycle management.

High Performance Messaging for Distributed Systems is what it says on the tin: a performant messaging system boasting “ultra-low latency”, targeted at distributed systems. In addition, it provides high resiliency and guaranteed message delivery. Alternately, you could use Kafka, or another messaging service, via the connectivity options

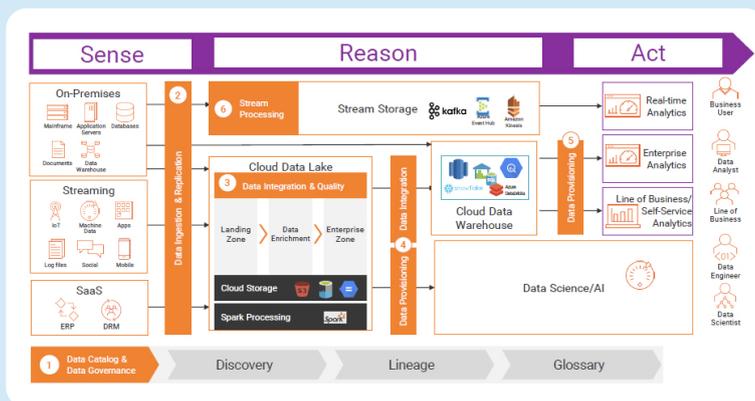


Figure 1 – Reference architecture for IDMC

partnered with Datumize in order to maximise its compatibility with IoT. The platform also offers comprehensive cloud capabilities and is available on a consumption-based pricing model.

IDMC provides a single architecture (and user experience) for data ingestion – whether via streaming, batch, or whatever else – that leads into solutions for stream processing and data integration (see Figure 1). Features relevant to streaming include real-time ingestion, mass ingestion, automated handling of schema and data drift, and a Kappa messaging architecture.

Analytics and modelling	★★★★
Development	★★★★★
Architecture	★★★★
Deployment	★★★★★

Connectivity	★★★★★
Integration	★★★★★
Self-service	★★★★★
Non-analytics streaming functionality	★★★★★

“Informatica Cloud Mass Ingestion is so easy to use that it saves us 90 percent of the ETL effort. I can just open a browser and access it anytime, anywhere.”
University of New Orleans

that Informatica provides. For instance, Cloud Data Integration can gather and load in batch data from Kafka directly, with an “at least once” delivery guarantee.

Enterprise Data Catalog can also scan Kafka deployments in order to extract relevant metadata (message structure, for instance).

Finally, Data Engineering Streaming is a continuous event processing engine built on Spark Streaming that is designed to handle big data. It supports both batch and streaming data, and it features out of the box connectivity to various messaging sources, as well as no-code visual development (shown in **Figure 2**). As part of the latter, it provides hundreds of prebuilt transformation functions, connectors and parsers. You can also pipe in your own code, or build your own functions and whatnot using Informatica’s business rules builder. Essentially, it allows you to enrich your streaming data in real time. This could mean improving data quality, masking sensitive data, aggregation, or what have you.

pipeline for data science and machine learning which helps you to apply machine learning and AI models to your streaming flows.

In addition, Informatica is keenly aware of the need to govern streaming data, and the company’s suite of data governance products are available for this purpose. Data cataloging, preparation, discovery, lineage, and visualization are all available, and have been designed to promote self-service and collaboration. Security features, like masking, authentication and access control, are also available, and real-time job monitoring, analytics, and visualisation are provided via the Operational Insights service.

Why should you care?

Informatica provides a high-quality, comprehensive stream processing solution that is positioned as just one part of a much larger, and broader, integrated data platform. Moreover, it is highly compatible with the cloud, and includes native cloud ingestion; it provides a broad range of connectivity, exemplified by the sheer quantity of connectors and scanners provided; and it offers a unified user experience, regardless of whether you’re deploying in-cloud or on-prem, which data sources you’re using, whether you’re using it for streaming or batch processing, and so on. The latter is particularly important, in that it allows you to abstract out much of the underlying complexity of stream processing, thus enabling your business users to work that much more efficiently and effectively.

We are also impressed by IDMC’s ability to transform streaming data in real-time, especially regarding its substantial number of built-in transformations. This is an area where Informatica’s breadth of capability really shines, by allowing you to combine high-end data quality and masking with stream processing. Moreover, Informatica’s data governance and security solutions combine with streaming in much the same way, with similar benefits.

The Bottom Line

Informatica Data Management Cloud offers highly capable and well-integrated stream processing as part of its overall data management functionality.

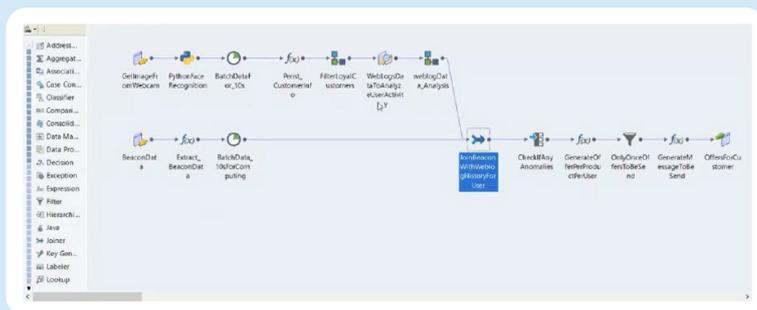


Figure 2 – Informatica Data Engineering Streaming pipeline

IDMC also supports Spark Structured Streaming, which can be important if you want to aggregate streaming data based on event time (not processing time) and hence reorder data that has arrived out of order before delivering it to your data target. It also supports Confluent Schema Registry, which can be used to parse Kafka messages, retrieve message structure, and handle schemas as they change and grow.

Moreover, you can use CLAIRE to augment your solution with machine learning. For example, to automatically detect and generate schemas. This is particularly beneficial in that it allows you to discover and rectify data (and schema) drift. Informatica also provides a ready-made integration

FOR FURTHER INFORMATION AND RESEARCH [CLICK HERE](#)