Streaming ingest and intelligent store

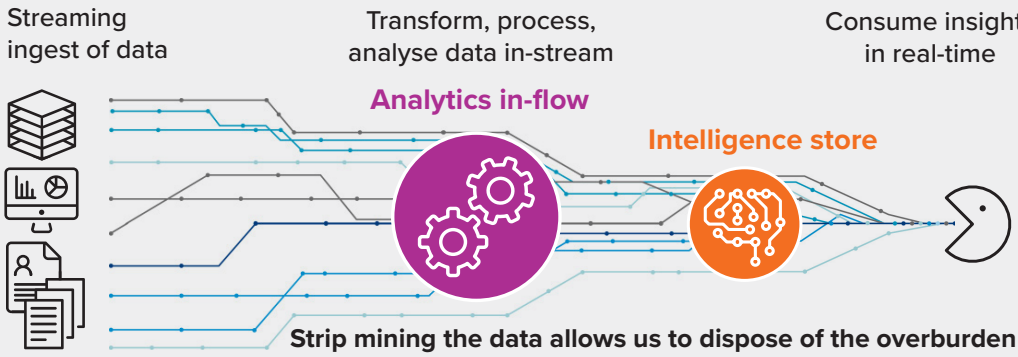
Unleash the full potential of your data by leveraging GFT's streaming ingest and intelligent store

In recent years the 'data lake' has provided a flexible storage and archiving environment that has helped minimise overheads; with data analysis strategies such as schema on-read and simplified read / write activity, helping to improving productivity and performance.

However, as the data explosion has accelerated these lakes have become unkempt muddy swamps, impacting their transparency through inescapable neglect.

The rate of introduction of new data sources is outpacing improvements in price and size of storage; inevitably in the future, we will be unable to store everything without significant cost and effort.

We need a new approach for generating insights from our data; one that focuses on the value in the data stream, ensures mission-critical data remains a priority, whilst utilising a 'strip-mining' approach for dealing with the huge overburden of data.



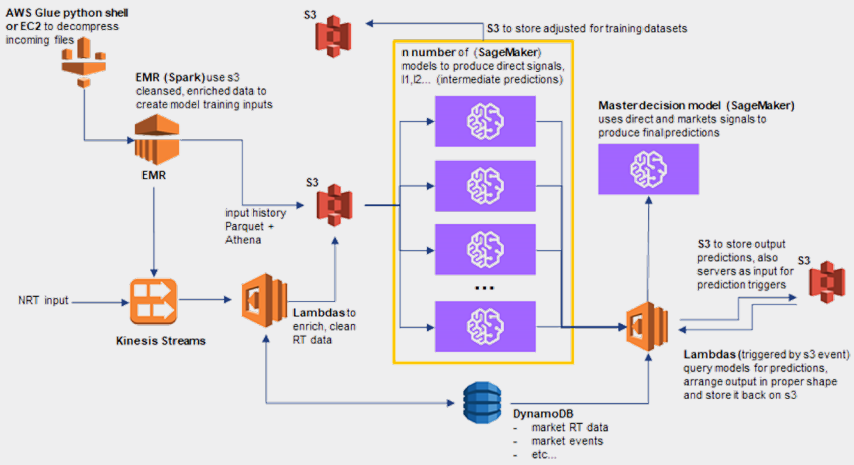
1. **Injecting intelligence into business services**: Stream processing enables our clients to develop the capability to deliver models that create a competitive advantage to the business; the best-run companies run their business on models. Advances in Artificial Intelligence / machine learning techniques paired with the power of AWS allows us to leverage the improved compute, memory and network resources.

2. **Streaming ingest and analytics in-flow**: The stream ingests both structured and unstructured data through real-time data pipelines. Once in the pipeline we can use advanced techniques to clean, transform and analyse in near real-time the data and meta-data. Modern data catalogues support an API enabled meta-data repository across large disparate data sources, enabling the presentation of a unified model to the enterprise.

3. **Real-time consumption and the intelligent store**: The in-flow analytics generate focused end-point consumable data sets ready for real-time consumption; we capture this via an intelligent store (IS) that is readily available and optimised for easy access or made available in real-time in the data pipeline. The IS provides the ability to audit model decisions and find the right data with ease, mitigating data swamp issues related to running a schema-less architecture.

4. **Smart archiving strategy**: As the number of unstructured data sources and the associated volume explode, it would be unwise to try to keep all data in an already muddy data lake. The new data strategy ensures that the archive collects all mission critical data for both business operations and regulatory requirements. We do not discard structured raw data, rather we focus purely on extracting the useful data from the unstructured sources.

AWS native sample machine learning pipeline



Benefits

- An easily discoverable and well governed data source for end consumers
- Accountable models: IS supports the audit of any model decisions, provides an audit trail and reproducible results
- Opens up lineage and self-service across usable data
- Utilise modern ML frameworks such as SageMaker to support model execution and maintenance in production, freeing up data science resources to focus on business priorities
- Better models – data scientists do not make bad assumptions on unlabeled raw data and do not waste time on stale data.

Partner reference customers / solutions

GFT have built a data lake that works for both capital markets and retail bank businesses, by taking a federated approach to the build and deployment of solutions. Our solution provides lineage visualisation integrated with a lineage store and the rollout of a data catalogue.

The system consumes all structured and unstructured incoming data via streaming and batch processes. We have created a common metadata model, and a code and release generator that enables 'self-service' for all actors within the system.

The design and build of an operational processing platform was based on Hadoop, Spark, Akka and Kafka. Centralised global operational processing of 'liquidity and capital reporting' is managed by a single processing platform.

The business was acquired by a new parent company, thereby taking the surveillance platform product to a wider market, creating more challenging performance requirements. The solution allowed a standardised ability to on-board new customers.

The GFT difference

Big enough to deliver – small enough to care

Straight-forward decision-making and a 'hands-on' mentality

A passion for innovation

Our commitment to delivery

A focus on sustainability, rather than short-term profits

About GFT

GFT Technologies SE is a global technology partner focused on digital transformation in the financial sector. Founded in 1987, we have around 5,500 experts in 12 countries.

› gft.com

GFT Technologies SE r Schelmenwasenstr. 34, 70567 Stuttgart, Germany r T. +49 711 620 420 r www.gft.com r info@gft.com