



gathr

www.gathr.one

720 University Avenue,
Suite 130 Los Gatos, CA 95032
USA
Email: contact@gathr.one

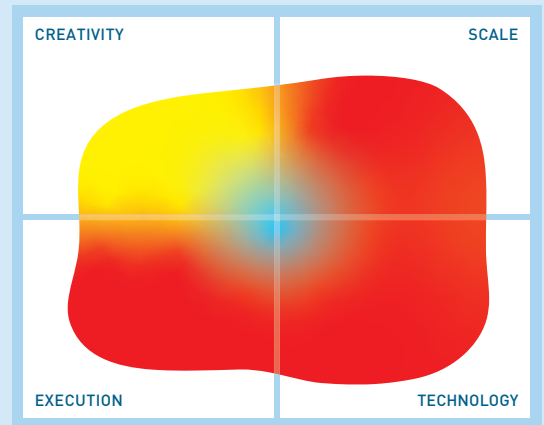
Gathr

The company

Gathr is an Impetus product, with Impetus itself a privately held software services, solutions and products company. Its focus is in data warehousing, big data, and streaming analytics, it was founded in 1991, and it is currently based in Los Gatos, USA. It also has several development centres in India.

The company has more than 1,700 employees spread across the globe. Its business partners include AWS, Microsoft, Google Cloud, Snowflake, Databricks, IBM, and Cloudera. In particular, Kyvos Insights is another Impetus company whose SQL on Hadoop engine can interoperate with Gathr.

“ [Gathr] helped us replace 1.5M lines of code through its visual interface. ”
Communication Data Analytics



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.

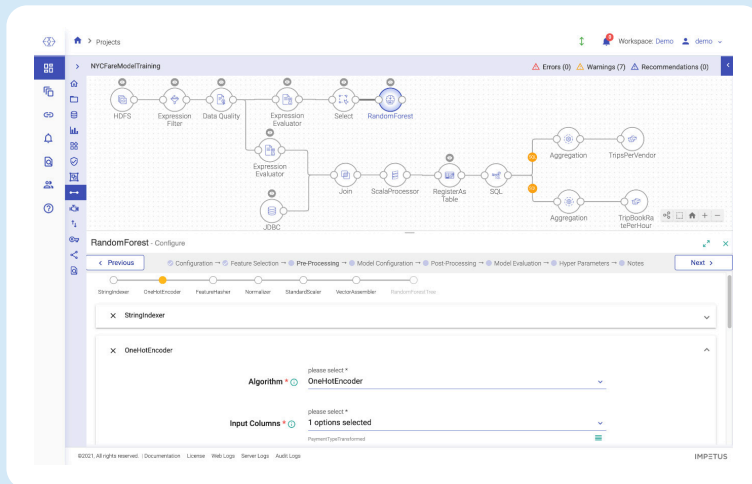


Figure 1 – Model training pipeline in Gathr

What is it?

Gathr (formerly StreamAnalytix) is positioned as a self-service-oriented data pipeline platform, and more specifically, as an all-in-one platform for creating, managing, and deploying data pipelines. This includes data processing, ingestion, cleansing, transformation, blending, loading, preparation, integration, visualization, analytics, and more. Its analytics capabilities are particularly rich, and can be performed in real-time on streaming, batch, and micro-batch data. This is no surprise, considering the product’s origin as a dedicated streaming

analytics offering. At the same time, the breadth of capability available explains the recent rebrand: ‘StreamAnalytix’ is too narrow a descriptor for what the platform has become.

Gathr’s streaming analytics capability is itself built on top of several popular and mature open-source technologies. Rather than trying to replace, say, Kafka, Impetus sells Gathr’s streaming analytics on the promise that it will save you the trouble of combining these technologies into a single offering yourself, while at the same time providing enterprise level scalability and support. It is available as a SaaS, on-premises, in-cloud, or as part of a hybrid solution. It includes multi-tenancy support, and is compatible with the “big three” cloud providers as well as Databricks.

Gathr is currently available in three editions: Gathr Unlimited is the full, but self-managed, package; Gathr Pro is the pay-per-use SaaS equivalent; and Gathr Free is a SaaS version with limited functionality (it is largely restricted to data ingestion). Note that at time of writing, Gathr Pro is significantly more limited in functionality than Gathr Unlimited (though not as much as Gathr Free). However, we are told that this will change in relatively short order.

Analytics and modelling	★★★★★	Connectivity	★★★★★
Development	★★★★★	Integration	★★★★★
Architecture	★★★★★	Self-service	★★★★★
Deployment	★★★★★	Non-analytics streaming functionality	★★★★★

“ [Gathr] helped us build and make our use-case production ready in 2 weeks that would otherwise take us 3 months. ”
Broadcast Media

What does it do?

Gathr enables you to assemble, debug, and deploy data pipelines and applications using a low/no-code, drag-and-drop visual interface, shown in **Figure 1**. It will also present an automated, dynamic view of the schema and outputs of your pipeline as you build it: a useful sanity-checking measure. A variety of pre-built components (including more than three hundred ETL functions) are provided for your pipelines, and you can modify these or even add your own if necessary. Moreover, Gathr pipelines have a self-healing capability, and will hence automatically recover from failures whenever possible. Pipeline and data set management are also provided, complete with version control (either using the platform itself or outsourcing to Git) and data profiling.

The platform also features extensive support for machine learning. This includes model training, building, scoring, and deployment, as part of both real-time and batch processes and – in the latter case – via either the platform itself or a REST API. In fact, **Figure 1** displays a model training pipeline. Gathr also supports a variety of machine learning algorithms, as well as SparkML, H2O, TensorFlow, R and Python. PMML (Predictive Model Mark-up Language) is also supported, providing model portability. A variety of model management techniques are also supported, including A/B and Champion/Challenger. It also offers hot-swapping and model versioning.

Aside from all this, the product also boasts data lineage, metadata management, data monitoring, alerting, and CDC capabilities (the latter on MySQL, MSSQL, PostgreSQL, Oracle and MongoDB), as well as over thirty connectors to third-party and open-source products. It supports Apache Storm, Spark and Flink as stream processing engines, but uses the same front-end for all three. In addition, you use Gathr apps in exactly the same way regardless of the underlying streaming engine. This makes moving from one engine to another a relatively easy task, since you won't need to redesign your apps or get used to

“ [Gathr] helps us identify anomalies in our incoming data in real-time. It helps us scan ~10k applications for any abnormal user behavior. ”
Banking and Finserv

a new interface. Impetus also offers a migration utility to help you move from various competing platforms to Gathr.

Why should you care?

Gathr has a number of strengths. Its breadth is perhaps the most apparent: it now stretches far beyond just streaming analytics. This both increases its overall value and applicability as a product, and makes it easier to integrate your streaming analytics with many of your other data integration and engineering needs, since they will be part of the same platform. It's also easy to use, thanks to its low/no-code visual pipelines and consistent user interface, and provides impressive support for machine learning.

What's more, by leveraging open-source technologies as part of an enterprise-grade solution, Gathr provides many of the advantages of open-source development (such as community-driven development, future-proofing and whatnot) without its most major drawbacks (such as lack of enterprise support, and the need to assemble your solution yourself, both of which are taken care of for you by Gathr).

Gathr Pro (the premium SaaS version of Gathr) is particularly interesting, or rather, it will be when its full intended functionality is available. At that point, it would give you a fully-managed, cloud-deployable solution that can start small, with a low cost and footprint, then scale up with you as you grow, creating a very “hands-off” data pipeline experience (and for streaming analytics in particular). Moreover, even without considering Gathr Pro, the scalability and cloud aspects of this already exist as part of Gathr Unlimited.

The Bottom Line

Gathr offers a wide-ranging data pipeline solution that is notably well-positioned to address streaming analytics. It combines the strengths of open source with the reliability and support of an enterprise solution, in the cloud, and at scale, while also offering significant ease of use, integration, and SaaS capabilities, among other things. In short, it is certainly worthy of your consideration.

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