

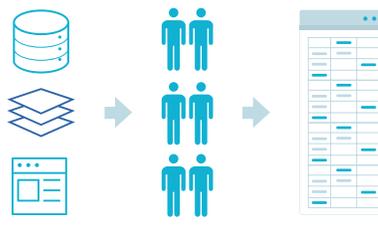
# Tamr on Google Cloud Platform Enhances Time To Value And Ease Of Data Preparation

Business analysts and application developers are under increasing pressure to generate clean, comprehensive datasets for analysis to drive business value.

Without high quality, unified data, downstream analytics are ineffective and, in most cases, time is of the essence when generating insight. Markets and opportunities move very quickly and businesses need the capability to respond even faster with insights that matter. A failure to do so results, at a minimum, in slower growth and foregone profit.

Unfortunately, the current operating environment of most enterprises isn't architected to meet these requirements for a variety of reasons, including:

- + Amount of manual effort required to integrate sources: Integrating disparate data sources into a unified, clean dataset requires significant manual labor. Programmers are needed to interrogate individual datasets, map them to target schemas, and eliminate duplicate records. This slows down the process of acquiring needed datasets and creates a large backlog for IT.
- + Previous work is rarely reused: Whether it's transformations written by IT or similar projects created by colleagues, completed work is rarely reused in today's environments. This inability to discover and leverage existing work hinders time to value and forces analysts to recreate existing assets.



Tamr on Google Cloud Platform brings scalable data preparation to the cloud and empowers business analysts to build and publish new data sets for analytics, with minimal coding and less IT support required.

These realities put organizations in danger of missing opportunities and being less competitive in their respective markets. Enterprises need a solution that can quickly integrate data sources in a more self-service way while leveraging work already done across the company.

Tamr on Google Cloud Platform brings scalable data preparation to the cloud and empowers business analysts to build and publish new data sets for analytics, with minimal coding and limited IT support required. The solution leverages the data preparation capabilities of Tamr in addition to the robust, elastic processing capabilities of Google Cloud Platform, bringing together two worlds that make analytics much quicker and easier for both the business analyst and IT.

Tamr's offering integrates with Google Cloud Platform, specifically Google Cloud Dataflow, to perform all functionality that analysts require to ingest data of their choice, prepare it, then publish the resulting dataset to a downstream tool or system of choice for advanced analytics.

Specifically, in the integration workflow:

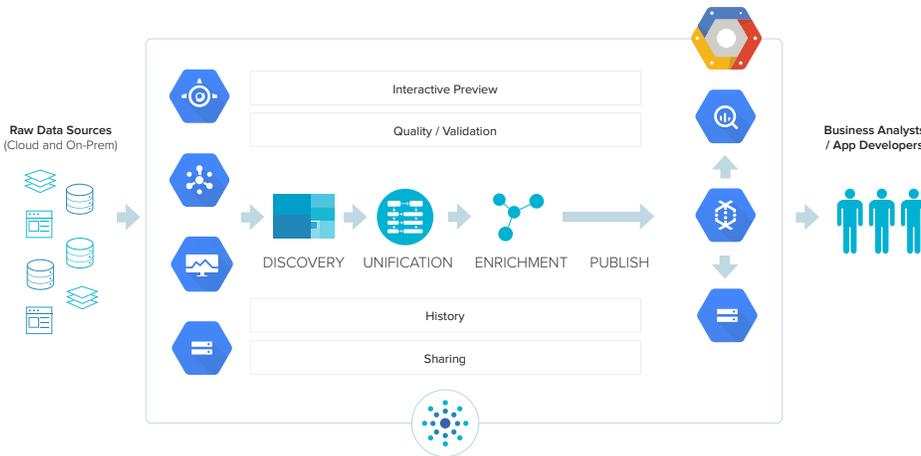
- + Users ingest data from a variety of sources, regardless of format, location or volume, including data captured in offerings such as Google Cloud Storage.
- + Tamr will perform data preparation operations (through machine learning and expert sourcing) to develop a clean, unified dataset for consumption.
- + The user will then publish the resulting dataset, using Google Cloud Dataflow, to destinations such as BigQuery storage, where it'll be used downstream for analytics. Google Cloud Dataflow is a simple,

flexible, and powerful tool that can be used to perform data processing tasks of any size.

- + After initially publishing a project, Tamr can continue to enrich future versions with new and updated data and with new information from experts.

Within the data preparation phase, Tamr focuses on empowering users to quickly discover datasets of interest, unify and clean them, then publish the result to a target of choice. The solution enables users to overcome key hurdles in quickly generating key datasets of interest in a much more self-service manner. Moreover, with a WYSIWYG interface, analysts can use Tamr to build these individualized datasets with minimal coding required

Tamr's specific functionality is as follows:



- + Data Discovery: Users first need to discover datasets before they can be used. Tamr on Google Cloud Platform allows users to discover existing, raw datasets that have been cataloged by Tamr as well as leverage datasets that colleagues or IT have already curated. This sharing component of Tamr's offering minimizes rework and drastically improves time to value.
- + Unification and Enrichment: Once relevant data is ingested, Tamr will significantly automate the unification and enrichment of user datasets via its unique blend of machine learning and expert sourcing. This will rapidly improve time to value in creating the core unified dataset that many business analysts seek.
- + Interactive Preview: Giving users the ability to see the result of their work interactively is a core component of the offering. Users can send a sample of their data to one of many preview tools, such as BI dashboards, to assess the level of quality of their dataset. If certain data is missing or of poor quality, users can return to Tamr for adjustments.
- + Quality: Most unified datasets still have issues with quality and, unfortunately, this is typically a heavily manual process to correct with involvement from central IT, who also needs to spend time re-correcting the same problems. With Tamr on Google Cloud Platform, analysts can reuse transformations (written by IT or other analysts) on new datasets through push-button functionality. This capability enables IT or analysts to write transformations once and use them many times, driving ease of use and productivity.
- + Publish: Using the power of Google Cloud Dataflow, Tamr will execute a job on the user's behalf to move the needed data from raw sources, conduct the required transformations, and publish the resulting dataset to a target store, such as BigQuery. Given Tamr is leveraging Google Cloud Dataflow, the processing is very scalable and rapid.

Tamr on Google Cloud Platform will allow users such as business analysts to overcome the biggest technical hurdles involved with generating datasets needed to quickly make informed business decisions. The offering significantly automates the integration of data sources and reduces the need to scale with only human resources, speeding up the integration process and easing IT's backlog. Moreover, it allows business users across the organization, as well as IT, to collaborate and reuse work already performed, drastically enhancing time to value and reducing duplicate effort.

Tamr focuses on empowering users to quickly discover datasets of interest, unify and clean them, then publish the result to a target of choice.

Tamr on Google Cloud Platform will allow users such as business analysts to overcome the biggest technical hurdles involved with generating datasets needed to quickly make business decisions.