

Technology Use Case Study

Aurum

Making Data Simple

As companies grow, collecting and organizing data becomes an essential task in day-to-day business operations. However, in the world of modern business dealings, the sheer volume of data can be an ever-present challenge. Companies today have various data sources, people working on them, and pipelines of things to be analyzed. Good data processing is crucial to the survival of any businesses, but how to manage the information overload?

Raul Castro Fernandez, PhD researcher at MIT CSAIL and founder of data discovery tool Aurum, explains, “**One of the biggest problems that many companies face is that even though they now appreciate that data is valuable, they have a lot of internal questions that they could solve with data. But it will be hard to understand which data will support your question.** Hence, the point of Aurum is that it will serve as a tool, so it makes it easy for you to detect which data might contain the answer.”

Aurum seeks to revolutionize the industry and heighten business performance by extracting data more efficiently and enabling employees to put their time toward other tasks.

What is Aurum?

Aurum is a system designed to tackle data discovery problems. It introduces a new discovery algebra called the Source Retrieval Query Language (SRQL) that lets users directly search for relevant data sources through a set of primitives that expose the relative insights of the underlying data.

Aurum solves two problems in a business setting: i) merging data that is stored across multiple storage systems, from databases to data lakes; ii) answering complex questions by accessing thousands of data spreads from various data sources.

At the start of any potential inquiry, most of a company’s energy will be spent figuring out which data will address the question. Fernandez states that he often hears about how companies “struggle to even find the data, and this is why they are so interested because it’s a way for them to shorten their time.”

Aurum is designed to not only find the most relevant data set, but also provides metadata that can help with the preparation and processing of future data. One such example of this is Aurum’s ability to merge data stored across multiple storage systems, from databases to data lakes. Aurum can connect to multiple sources within an organization and create a model of everything that’s available.

For instance, a user on the system might be searching for data related to “X.” If the user has the correct API’s or functions needed to search for “X,” Aurum can build a set of APIs or a set of functions that the user can see.

Fernandez explains, “Let’s say that you have data that could be more useful if you could join data with a different department. Would merging the data provide value or become fatal? Aurum will help you determine whether merging data is possible and what steps are required to execute the process.”

(continued)

For more information about CSAIL Alliances industry engagements, please visit:

cap.csail.mit.edu

What is Aurum? *(continued)*

Aurum is also useful in recognizing duplicate data sets, a common plague of industry environments. In every industry where data is important, it's common for employees or researchers to 1. copy data, 2. make some changes, 3. copy that data over again, 4. repeat. This creates several different copies that confuse records and reduce the quality of data over time. Aurum can help simplify these situations, determining which data sets are copies of each other so that users can remove or reconcile them as needed.

Connecting with CSAIL Alliances

In recent news, Aurum has partnered with industries like Koser Bank, BASF, Dell EMC, BT, and Merck to track key performances of the product. These industry connections wouldn't have been possible without CSAIL Alliances.

Fernandez says, **“With CSAIL Alliances, I was provided with many opportunities to showcase my work to Alliance members. I would deliver a presentation during an Alliances event and connect with interested company representatives.** In speaking with the members after my presentation, we would determine if Aurum could be useful for their specific data challenge. If Aurum could help with their problem, I would send them a demo after the event. We specifically built this Aurum prototype that can be deployed so that even if I cannot access the data from the company, I can just send them the prototype and they can spend about 20 minutes setting it up and see if it brings value to the company... the prototype was useful for companies trying to determine if Aurum would be helpful in their organization.”

With the Alliance member connections, Fernandez was able to track the impact and value

“It is a value proposition working with big companies,” says Fernandez. “Aurum is able to help them determine the best way to find the answers they are looking for and I can learn the many approaches they use to try to discover the data. It is mutually beneficial.”

Aurum promises to take data discovery to the next level, helping industries locate, extract, and analyze tremendous amounts of data. The promise of both better information and better time management is sure to be a major draw. After the strong start Aurum got through its connection with CSAIL Alliances, they're ready to build a wider community and help companies everywhere make data work for them.