tetrascience

Streamlining synthetic biology at Ginkgo Bioworks

CASE STUDY

Ginkgo Bioworks is pioneering new ways to design biology. The company's innovative platform enables scientists to edit DNA to program cells, similar to how software developers program computers. The new biological designs created by Ginkgo can serve a wide range of objectives, from identifying new therapeutics for pharmaceutical companies to helping agriculture companies better protect crops from pests.

To engineer new organisms at scale, Ginkgo has constructed foundries—biological factories through which teams design, build, and test thousands of organisms. Results are delivered to customers and incorporated into a rapidly growing Ginkgo "codebase"—a library of cells, enzymes, and genetic programs that help jumpstart the next projects. These unique capabilities strongly differentiate Ginkgo from the rest of the industry.

For cell design testing, Ginkgo uses a diverse collection of scientific instruments—such as mass spectrometers, plate readers, thermal cyclers, liquid handlers, and more. Collecting instrument data and preparing data for downstream applications often require software engineers to spend excessive time creating home-grown tools.

"We've invested a significant amount of resources collecting and parsing data from instruments," says Jamie Cho, head of software at Ginkgo. "We want to reduce the overall cognitive load of our team so we can focus more on science and less on file format details."

Meanwhile, Ginkgo's data team wants to increase the value of data generated by instruments. If Ginkgo can capture more context surrounding data production along with raw files, it can use that contextual information to optimize instrument utilization, conduct predictive maintenance, and generate other key insights about foundries.

Ultimately, streamlining data operations and contextualizing data could have a huge impact on the company and its customers. Ginkgo could scale its business more efficiently, speed the delivery of new biology designs, and capitalize on emerging technologies to more easily engineer biology.

Simplifying data collection and preparation

The Ginkgo software and data teams began searching for an external platform to automate the collection and parsing of data from scientific instruments—and they discovered TetraScience. "We have a company mandate to buy more and build fewer home-grown solutions," says Cho. "With TetraScience, we saw a platform that could help us reduce burdens on our software engineers."

The Tetra Scientific Data Cloud can automate the collection and parsing of raw data from numerous instruments. The Ginkgo team began its TetraScience engagement by onboarding mass spectrometers (or "omics" instruments) and subsequently adding plate readers. In all, more than 120 instruments have been onboarded to the TetraScience platform.

GINKGO BIOWORKS

Challenges:

- Refocus software engineers on high-value work by reducing the need to develop home-grown tools
- Optimize lab operations and enhance efficiency by enabling analysis of instrument data

Solution:

The Tetra Scientific Data Cloud™ simplifies data collection and parsing from scientific instruments while helping engineer data for analytics applications and AI.

Outcomes:

- Automated data collection and parsing for 120+ scientific instruments
- Replaced home-grown solutions with an enterprise-quality solution that provides a higher level of scalability, robustness, and operational excellence
- Gained access to functionality previously unavailable with in-house solutions
- Refocused software engineers on higher-value tasks
- Streamlined onboarding of scientific instruments by removing complexity and reducing 60% of steps
- Enhanced data contextualization, preparing data for analytics applications and ultimately AI

To collect instrument data, the Ginkgo team uses the Tetra File-Log Agent—a high-speed, instrument-agnostic agent that detects changes in file-based outputs from instruments and automatically uploads data into the Tetra Scientific Data Cloud. "The File-Log Agent is much more robust than our home-grown data mover," says Michael Khalandovsky, senior software engineering manager at Ginkgo. "We no longer have issues with data completeness—it's an important benefit for us."

> ***"We have a company mandate to buy more and build fewer home-grown solutions. With TetraScience, we saw a platform that could help us reduce burdens on our software engineers."***
>
> —Jamie Cho
> Head of Software, Ginkgo Bioworks

The Tetra Scientific Data Cloud uses a Tetra Intermediate Data Schema (IDS) to map vendor-specific information to vendor-agnostic information. The Ginkgo team then adds another layer of information onto that format to support the company's specific data management requirements.

"In the past, we went to extremes to parse data from certain instruments," says Cho. "For us, the ability to automatically parse data is probably the single most important capability of the TetraScience platform."

Reducing the need for home-grown parsers

Implementing the Tetra Scientific Data Cloud has eliminated the painstaking work of developing and maintaining custom parsers for many instruments. "In some cases, it felt like you had to be a .NET hacker to piece everything together," says Christina Hawkes, senior software engineer at Ginkgo. "The resulting parser was brittle and difficult to work on. We are very happy to stop using our home-grown parsers for those instruments."

So far, the Ginkgo software team has eliminated home-grown parsers for approximately 50 percent of the scientific instruments used for testing. "With the File-Log Agent, we're looking forward to completely retiring some home-grown file-moving tools for many instruments," says Cho.

The Tetra Scientific Data Cloud has consequently reduced the need for certain specialized instrument and programming knowledge within the software team—and enabled software engineers to reallocate their time. "Our software engineers no longer have to understand the details of all the instruments we're handling," says Cho. "We can focus them on higher levels of value delivery, like work that directly affects the user experience. That's a huge win for Ginkgo."

Streamlining instrument onboarding

For Ginkgo's IT team, the Tetra File-Log Agent API has reduced the time to onboard scientific instruments. "We can push a configuration to the PC agents via the API, simplifying and standardizing the steps that our IT team has to follow," says Hawkes. "They just install the Tetra File-Log Agent and that's it."

The faster onboarding experience is particularly helpful as Ginkgo continues to expand its business. Ginkgo acquired multiple organizations and doubled its headcount in 2022 alone. The IT team can scale the company's collection of instruments in a fraction of the time as before, allowing scientists to get to work fast.

Speeding scientific workflows

By automatically pushing data to downstream locations, the Tetra Scientific Data Cloud eliminates complicated steps for exporting instrument files—and that enables scientists to complete runs faster. “We’ve significantly simplified workflows,” says Cho. “We’ve slimmed the BioTek Gen5 operating manual from 40 pages to 10, which saves scientists approximately 60 percent of their time on runs.”

Saving time on testing ultimately helps the company deliver new designs to customers faster. “One of our company priorities is enhancing efficiency,” says Cho. “Working with TetraScience, we are increasing efficiency overall while accelerating biology engineering.”

Generating new lab insights

Using the Tetra Scientific Data Cloud, the Ginkgo data team is investigating the possibility of adding richer metadata to the raw instrument data, which in turn will help the team generate new insights about the foundry and its instruments. “It’s possible to understand when data acquisition started for a particular run, what method was used, and what actions a scientist may have taken before starting the run,” says Nicholas Flores, test engineer on the data team at Ginkgo. “As a result, we can pinpoint ways to improve foundry performance.”

By analyzing metadata about instrument usage over time, the team is looking forward to being able to optimize utilization and anticipate maintenance needs. “At the moment, we don’t have a good way to gauge when our omics instruments are being used, for how long, and by whom,” says Denyse Dodd, senior product manager at Ginkgo. “As we capture more instrument data using the TetraScience platform, we’re eager to see how we can monitor instrument utilization and improve predictive maintenance.”

The Tetra Scientific Data Cloud will enable Ginkgo to conduct those analyses and other advanced data analytics functions in the Snowflake cloud. TetraScience’s Snowflake integration will enable data to flow automatically from the Tetra Scientific Data Cloud into Snowflake, unlocking the power of visualization via dashboards. “Exporting data to Snowflake will be a big benefit to us,” says Cho. “We’ll be able to eliminate even more home-grown code and start to capitalize on a range of Snowflake capabilities.”

Envisioning an AI future

Looking ahead, Ginkgo is committed to tapping into the vast potential for using artificial intelligence (AI) and machine learning (ML) to engineer biology and improve biosecurity. To make the most of these technologies, the company’s leadership understands that “data is queen”: Powering AI applications requires large volumes of high-quality, AI-ready data.

“Consistent parsing and ingestion of data is a powerful enabler for AI. We are glad to have TetraScience as our partner for large-scale onboarding of new lab data sources,” says Dmitriy Ryaboy, Ginkgo’s VP of AI Enablement. “We are looking forward to an ongoing collaboration with TetraScience as we make progress on the AI journey.”

About Ginkgo Bioworks

GINKGO BIOWORKS

Ginkgo Bioworks uses the most advanced technology on the planet—biology—to grow better products. With a leading platform for cell programming, the company provides flexible, end-to-end services that solve challenges for organizations in food, agriculture, pharmaceutical, industrial chemical, and specialty chemical markets.