CIO
APPLICATIONS

ENTERPRISE SEARCH
EDITION

Diskover Data

A Simplistic Approach to Data Curation

Chris Park,
Founder & CTO
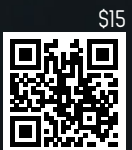
Diskover Data

A Simplistic Approach to Data Curation

> " We are truly committed and passionate about growing our platform to its fullest potential. We are aiming for the top!

Chris Park,
Founder & CTO

By Joe Phillip

Each day, companies gather, process, and analyze an influx of data from multiple sources ranging from product reviews and social media posts to stocks and transactional data. While some of the business data is organized, a large chunk of it—nearly 80 - 90 percent—is still unstructured; and these numbers continue to grow at an alarming rate per year.

Although this set of data isn't well-organized or easy to access, it is abundant in indispensable insights regarding what makes customers tick, how to create engaging experiences, and how people perceive a brand. It helps companies evaluate their customer support quality and align all their lines of business, operations, and objectives accordingly.

"Unstructured data is sitting everywhere, untagged and unidentified. Interestingly, this data can be the game-changer in achieving business success when integrated into the workflow," states Chris Park, Founder and CTO, Diskover Data. "What businesses lack is the ability to leverage the massive amounts of unstructured data to make informed business decisions."

Enter Diskover Data! Diskover Data empowers clients with a next-generation Data curation platform to manage data and customize reporting for their specific needs efficiently, helping them make the right data-related decisions about time, resources, and purchases. By helping clients create a global index in an open-source non-proprietary format. along with immediate and relevant information, the company enables clients to maintain the data independence of their index.

But delve deeper, and you will see there is so much more to Diskover Data! To truly understand what this company brings to the table, you have to know the story behind its genesis and the people who conceptualized it!

The Confluence of People, Process, and Technology

The initial idea behind forming the company spurred from a solution to problems that Park faced during his decade-long stint
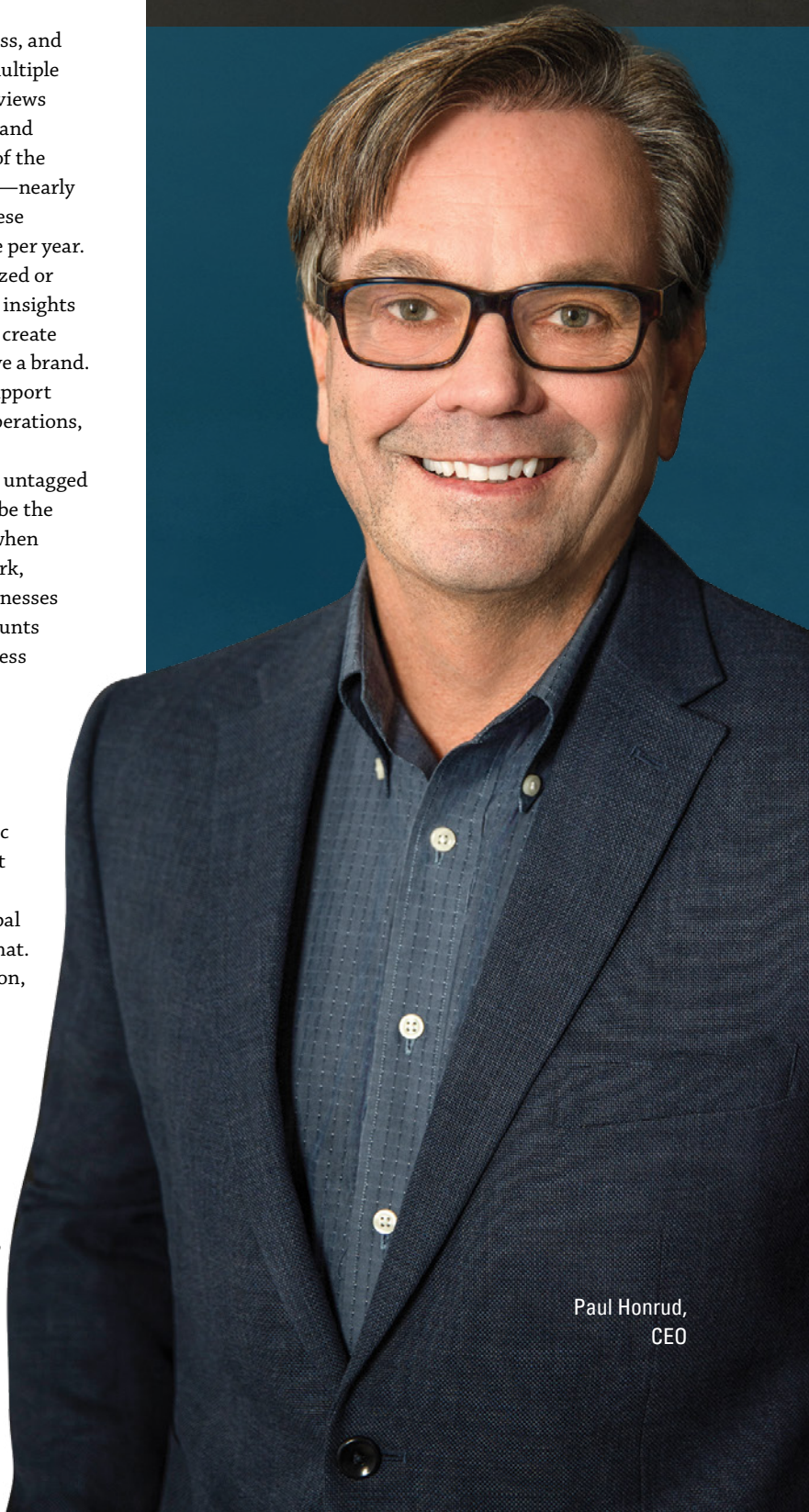
Paul Honrud,
CEO

as a System Engineer in the Media and Entertainment (M&E) industry. Working with large VFX and animation studios like Digital Domain, Zoic Studios, and Bardel Entertainment, he often found himself struggling to deal with large amounts of unstructured data and make sense of it. And this was a time before the enterprise world was totally new to the "data" game! Given that there wasn't a single platform capable of handling it, the data was getting out of hand and costlier to store—with no good visibility. The existing software solutions were not open-source, affordable, web-based, and scalable for large amounts of data. So, he left his career as a System Engineer in 2016 with one goal in mind—building a solution to help M&E and other industries deal with their massive data issue.

"I built the solution as an open-source data hub platform, to be open and extensible. Elasticsearch provides an open-source back-end foundation which indices can be visualized not only with Diskover-web but also other visualization tools like Kibana," says Park. "In the early stages of developing Diskover I was living in Kyoto, Japan. I would bike ride along the Kamo-gawa river for inspiration and new ideas and this was an influential factor in the early architecture and design of the software. The beauty and culture of Japan helped me to see new ways to store and visualize data." After four years of the software—operating under the name Shirosaidev—being a community open-source project, Park released an enterprise paid version to offer customers more features and better support—setting the foundation to the robust Diskover Data platform we know of today.

A year later, Park partnered with industry-veterans like Paul Honrud and Marianne Zuhorski to establish and expand the business. At the time, Honrud had already etched his name in the enterprise data world by conceptualizing and founding DataFrameworks. He brings more than 27 years of experience in data management and storage, along with a wealth of knowledge in growing technology-based businesses.

It is this passion for data management that eventually led him to know of Park. "Throughout my career, I've been abreast of innovative technologies that could make an impact in the data management and storage world. And when I came across Park's platform, I knew the value it could bring to businesses. It was perfectly adapted and architected to deliver the performance and scale customers need," says Honrud.

Marianne, who joined Honrud in late 2016, has been in the tech industry for over 25 years and has held various roles ranging from technical to executive-level sales and strategic alliances. She excels at creating synergistic partnerships to deliver revenue acceleration, as well as

Marianne Zuhorski, CRO

figuring out the most successful route to market for early-stage companies. Marianne's last role was as VP of Strategic Alliances for DataFrameworks until the revenue trajectory caught Dell's attention and resulted in an acquisition to them in Aug 2018. And the rest as they say, is history!

Working in liaison, Park, Honrud, and Zuhorski, who currently serve as Diskover Data's CTO, CEO, and CRO, have successfully established Diskover Data as a leader in the data management market, rapidly scaling the capabilities of their open source platform.

Solving the "Search Problem"

At its core, the Diskover Data platform is a customizable, cost-effective data curation platform that helps businesses organize, index, and track their unstructured data. It is designed to make people and workflows more efficient through empowerment, automation, and anomaly detection. Customers are welcome to opt for a trial period before subscribing to a plan, to see whether the solution is a perfect fit for their business or not. The software

is cost-effective and is available in four levels of subscription: Community, Essential, Professional, and Enterprise. "Also, with a robust and user-friendly UI, the Diskover Data platform empowers organizations to efficiently manage and visualize their data. It allows them to make right data-backed business decisions about time, resources, and purchases," says Honrud.

The Diskover Data platform is designed to enable users to effectively manage unstructured data with the simplicity of search. It is an open-source, scale-out file system indexer that leverages Elasticsearch to manage and index data across heterogeneous storage systems. With regards to business information, everything ultimately becomes a search problem. Businesses already have mountains of unstructured data, but what they lack is information. How do they make sense of the massive amounts of unstructured data to make informed business decisions? Information on their data enables better decisions. These decision makers need to be enabled with desired information all with the simplicity of search.

Unlike structured data, unstructured data often lacks business context required to provide information needed to answer difficult questions from a business context. By integrating search with workflow, the search index becomes empowered with business context.

Now once decisions makers are empowered, a few more requirements come into play!

Business context is often derived via additional metadata that becomes "searchable," and the open-source Elasticsearch platform is well suited to enrich indexes with additional metadata. The information and index need to be complete, timely, extensible, open, yet serve the needs of multiple stakeholders or use cases. Let's explore each bullet in more detail.

• Complete Index: Data is often fragmented across many geo-locations, cloud storage, and individual storage volumes. Diskover Data's platform is designed to index multiple fragmented data locations.
• Timely Index: The indexes within the platform need to be up to date and current, the company has customers indexing some of their data volumes every 30 minutes.
• Open Index: The index might be used as the inventory of intellectual property assets for the life of use or reuse. An open-source platform provides a

> " As an open-source file system indexer that leverages Elasticsearch to manage and index data across heterogeneous storage systems, the Diskover Data platform allows users to search and organize their files more effectively

great foundation for long-term asset preservation as vendor proprietary indexes may not stand the test of time.
• Extensible Index: Index enrichment might be specific to a particular infrastructure component or may pull enrichment information from a component being decommissioned.
• The combination of open-source with Elasticsearch scale out architecture provide ability for customers to harvest context from obscure locations if desired.
• Multiple stakeholders: With unstructured data fragmented across multiple volumes, some business users are simply trying to find an asset/file, then using additional metadata to ensure it is the desired file, for example, image resolution. Storage admins might simply be looking for all the files over three years old because the actual storage volume is 98 percent full. Operations managers might be looking for all files for a completed project, etc., or files that have met the service level retention agreements required by their external customers.
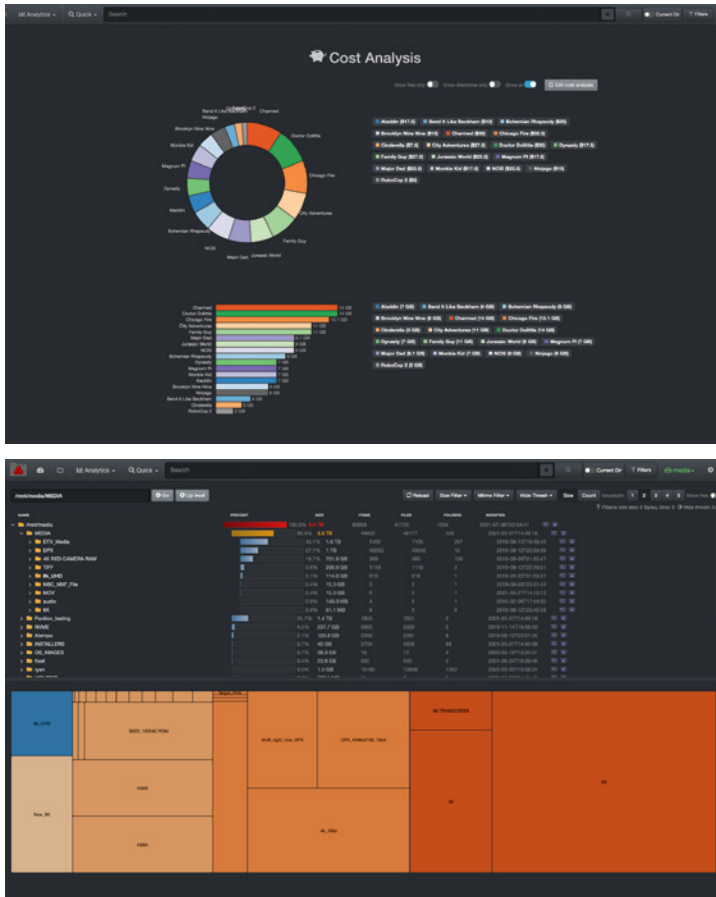
When a search is either integrated or becomes a key part of the workflow, the solution becomes mission-critical rapidly. Customers are at an inflection point and trying to decide what data to move to the cloud. But again, figuring out what data belongs where and when to support workflow optimization is another challenge. The additional business context within the Diskover Data platform brings increased accuracy to these data decisions. "When workflows leave a data management footprint, we use the initial index as a view into challenges that lurk within and where mitigation and maintenance are required. The action plan is based on the data workflow challenges illuminated by the baseline index which then require prioritization in the context of the customer and impacts," says Park.

The Role of Community Building

In addition to industry best practices around road maps and strategic planning, Diskover Data offers an open-source product called "Community Edition." which allows customers to engage with the product at no expense to them. "The best way to get more and more people onboard a good innovation is by building it in an open-source community. It allows the product to scale, meaning more people can experience the innovation, regardless of their vertical. Also, Diskover Data encourages transparency and trust alongside helping organizations become agile and collaborative in their data innovation. "We are not a startup that is building technology in search of a customer. We have existing customers that we have been

working with for a long time that are informing our product design and decisions. There is nothing on the market today that is as scalable and open source as the Diskover Data platform," prides Park.

On all counts, the platform is as unique as it is robust! Yet, the team behind the product improves it continually. They know once you provide customers with a solution to one problem or requirement, then the next problem that lies beyond the solved problem becomes visible. As such, they have dedicated nearly a third of the company to solely focus on innovation. The team has decades of experience with customers across verticals dealing with workflows that rely on unstructured data. The team takes requirements and innovation input from customers and incorporates it into product planning. They turn the proposed solutions very quickly and get them back out to the customers who requested them for validation.

Internally, Diskover Data is a highly collaborative organization. "This is key to our success! Very few customers have identical workflows or infrastructure, however the underlying challenges with managing and curating unstructured data are often very similar. We find our internal collaboration with our partners and customers leads to the best outcomes. Since Diskover Data embraced virtual and geographically remote employees, COVID had no impact on our culture," prides Park. "In this regard, Marianne's role is very strategic as we need support, blessing, and certification of all on-premise and cloud platform vendors. In addition, we are fortunate to have two strategic OEM agreements in place thanks to her efforts."

For The Data, We Are Yet to See

As file counts and locations continue to increase, Diskover Data envisions to pioneer innovative solutions that can make the workflows more automated and efficient. There are great opportunities to innovate in terms of anomaly detection, machine learning, and the visualization of areas of interest with regards to better mapping the mass of unstructured data to workflows. The company is currently working on the upcoming version of the software, which is expected to be completed in the next three to six months, ensuring that it is stabilized and ready to use for everyone.

"We are exploring the power of ML in the detection of cold data and other anomalies. This would enhance our solution a notch higher to provide better services to our clients. We are truly committed and passionate about growing our platform to its fullest potential. We are aiming for the top!," concludes Park. **CA**

Learn more about Diskover at
www.diskoverdata.com

CIO
APPLICATIONS

Diskover Data

TOP ENTERPRISE SEARCH *SOLUTION PROVIDERS* *2021*

Awarded by
CIO APPLICATIONS

Diskover Data

TOP ENTERPRISE SEARCH *SOLUTION PROVIDERS* *2021*

Awarded by
CIO APPLICATIONS

The annual listing of 10 companies that are at the forefront of providing Enterprise Search solutions and transforming businesses