



DevOps Veterans Bring Their Vision for High Performing Software Teams to Machine Learning and Data Science

Dotscience debuts DevOps for ML platform, empowering data science teams to achieve reproducibility, accountability, collaboration and continuous delivery of AI models

LONDON and SAN FRANCISCO – July 30, 2019 — DevOps has transformed the way software engineers deliver applications by making it possible to collaborate, test and deliver software continuously. Today, [Dotscience](#), the pioneer in DevOps for machine learning (ML), emerged from stealth to signal the rise of a new paradigm where ML engineering should be just as easy, fast and safe as modern software engineering when using DevOps techniques. For data science and ML organizations to achieve this DevOps for ML nirvana, the right tooling and processes need to be in place such as run tracking and collaboration, automated and full provenance (a complete record of all the steps taken to create an AI model) of AI model deployments and model health tracking throughout the AI lifecycle.

"Artificial Intelligence has the potential to reinvent the global economy, but as a discipline it's the Wild West out there," said Luke Marsden, founder and CEO at Dotscience. "We've seen damaging levels of chaos and pain in efforts to operationalize AI due to insufficient tooling and ad-hoc processes. The lessons learned from DevOps sorely need to be applied to ML."

History Repeats Itself: AI and Data Science Today is Like Software Engineering was in the 1990s

In the 1990s, software engineering work was split across development, testing and operations silos. Developers would work on a feature until it was done, often finding out too late that somebody else had been working on another part of the code that clashed with theirs. Without version control and continuous integration, software engineering was difficult. The advent of DevOps in the late 2000s was and continues to be transformative for software development. In fact, [Forrester declared 2018 to be the year of enterprise DevOps](#) with data confirming that 50% of organizations were implementing DevOps and that the movement had reached its "escape velocity." Forrester also "emphasized the importance of a collaborative and experimental culture in order to develop, drive and sustain DevOps success."

"Version control and the workflows that it enables now allow software teams to iterate quickly because they can easily reproduce application code and collaborate with each other," explained Marsden. "However, because ML is fundamentally different to software development, data science and AI teams today are stuck where software development was in the late 1990s. We are fixing that by creating tooling which respects the unique ways that working with data, code and models together is different to working with just code. This 'DevOps' approach to ML provides a fundamentally better and more collaborative work environment for data engineers, data scientists and AI teams."

The Disjointed State of AI Development Today

Reproducibility and productivity are inextricably linked. It is difficult to be productive when different team members cannot reproduce each other's work. In normal software development it is enough to version the code and configuration of an application and teams have seen dramatic increases in productivity working this way. In ML reproducibility, and therefore collaboration, is more difficult because putting the code in version control isn't enough.

“Collaboration around ML projects is harder than in normal software engineering because teams need a way to track not just the versions of their code, but also the *runs* of their code which tie together input data with code versions, model versions and the corresponding hyperparameters and metrics,” said Mark Coleman, VP of Product and Marketing at Dotscience. “While some of the largest and most engineering-inclined companies have invested in creating proprietary tooling to solve this problem, many companies don't have the necessary ability or budget and instead turn to manual processes that are both inefficient and risky. These cumbersome processes are often opaque and discourage collaboration, creating knowledge silos with teams, increasing key person risk and significantly diminishing team performance.”

In addition to enabling efficient collaboration, accurately tracking the ML model development process through run tracking means that the full provenance of a given model's creation is recorded. This aids debugging and can be invaluable if businesses must defend a model's actions to auditors, customers or in court—a key requirement for any AI application that is making life-changing decisions in production.

Dotscience's market research report, “[The State of Development and Operations of AI Applications](#),” [also announced](#) today, found that the top three challenges respondents experienced with AI workloads are duplicating work (33.2%), rewriting a model after a team member leaves (27.8%) and difficulty justifying value (27%). The report examines the AI maturity of businesses by how they are deploying AI today and investigates the need for accountability and collaboration when building, deploying and iterating on AI. The study also found that 52.4% of respondents track provenance manually and 26.6% do not track but think it is important. When provenance is tracked manually this usually means that teams are using spreadsheets with no access controls to record how their models were created which is both risky and cumbersome.

It is Now Possible to Achieve DevOps for ML, Immediately

[In a separate press release today](#), Dotscience launched its software platform for collaborative, end-to-end ML data and model management, enabling ML and data science teams to achieve reproducibility, accountability, collaboration and continuous delivery across the AI model lifecycle. Dotscience allows ML and data science teams to simplify, accelerate and control every stage of the AI model lifecycle by solving for critical issues when developing AI applications.



Dotscience delivers the following features to make AI projects faster and less risky, ML teams happier and more productive and helps track data, code, models and metrics throughout the AI model lifecycle, delivering the simplest and fastest way to achieve DevOps for ML:

- Concurrent collaboration across developer and operations teams
- Version control of the model creation process
- Automated tooling to maintain the provenance record in real time
- The ability to explore and optimize hyperparameters when training a model
- Tracked workflows that allow users to work with the open source tools they love and build better models by staying focused on the ML

Additional Resources

- Visit dotscience.com
- Follow [@getdotscience](#) on Twitter and [LinkedIn](#)
- Read the Dotscience [blog](#)
- Dotscience's [State of Development and Operations of AI Applications 2019 Report](#)

About Dotscience

Dotscience, the pioneer in DevOps for machine learning (ML), brings DevOps principles followed by high-performing software teams to ML and data science. The Dotscience software platform for collaborative, end-to-end ML lifecycle management empowers ML and data science teams in industries including fintech, autonomous vehicles, healthcare and consultancies to achieve reproducibility, accountability, collaboration and continuous delivery across the AI model lifecycle. Founded in 2017 by container storage veteran Luke Marsden, Dotscience is headquartered in the UK with offices in the US. Its mission is to accelerate and unlock the true value of data and analytics assets through AI.

#####

Media Contact:

Jocelynn Stidham
Bhava Communications for Dotscience
(703) 863-1277
dotscience@bhavacom.com