

Key Insights from the NYC Event

THRIVE: A Conversation With Data Leaders

Align Data Science with Business Goals,
Hiring & Retaining Your Data Talent,
Centralized vs. Embedded ML Teams



Irina Ashurova

Sr. Director of Data Development @
Pitney Bowes



Shane Murray

Former SVP of Data and Insights @
The New York Times



Justin Norman

Former VP of Data Science @ Yelp
Marine, ex-Fitbit, ex-Cloudera, ex-Cisco



vectice



MONTE CARLO

July 2022

Thank you for joining in New York City

Exclusive to VPs & Directors in Data Science & Machine Learning



Speakers:



Irina Ashurova

Sr. Director of Data Development @ Pitney Bowes

Leading the BI team responsible for analyzing data and business requirements and providing reports and data visualization. Responsible for data architecture, ETL processes to populate data warehouses from different sources.



Shane Murray

Former SVP of Data and Insights @ The New York Times

Formerly leading the entire data organization at The New York Times. Responsible for the Data and Insights Team with a heavy focus on the data platform.



Justin Norman

Former VP of Data Science @ Yelp
ex-Cloudera, ex-Fitbit, ex-Cisco, Marine

Experience leading 5 data science teams at Yelp, the corporate Data Product Management team, the Yelp DS/ML experimentation platform (Bunsen) and the Core Machine Learning Platform.



Irina Ashurova:

On Aligning Data Science Teams With the Business

Irina: Pitney Bowes is a big organization that includes multiple lines of business. As a whole, we have around 80 people reporting to the SVP of Data and Analytics. I lead the data division responsible for data development. My team consists of 16 people, partially distributed in the US, India and Canada.

Pitney Bowes used to have one centralized team that stored our data on a data lake. This made sense at the time because leaders needed global insight into how the business was doing. But while a brilliant data engineering team ran the operation, they were limited in their knowledge of our business needs.

We broke the siloes of the data teams by using data governance and data mesh to create a top-down approach to define our data domains coming from business needs.

We now consider business departments (data scientists, BI, and analytics) as the data consumers. When companies come to them with requests and tasks, my team delivers data to their departments based on the data mesh paradigm.



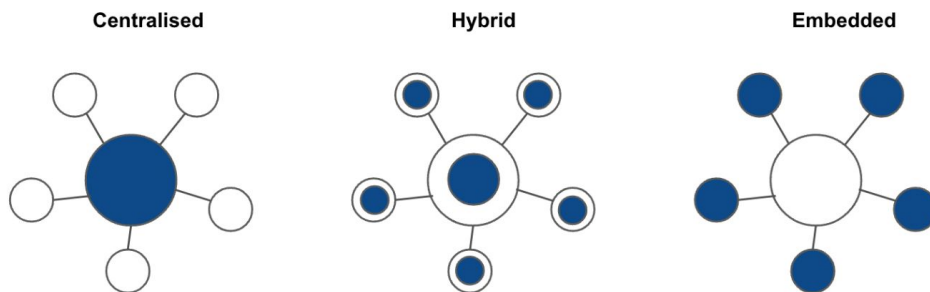
Irina Ashurova:

On Centralized vs. Embedded ML Teams

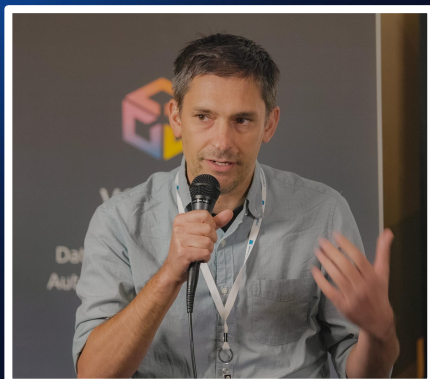
Irina: We assign domains to cross-functional teams led by members of my team. These members are the principal owners, and their task is to understand which resources are needed for a job and assemble their domain experts.

Embedding your data teams is a process that constantly evolves as new data products are required to meet business needs.

To succeed, you need to keep alignment between the business side and the data team.



Teams are quite small, consisting of four to five people with different maturity levels depending on the urgency of a task. I try to create challenging tasks for my team. When people are challenged, they will feel inspired to keep going!



Shane Murray



Justin Norman

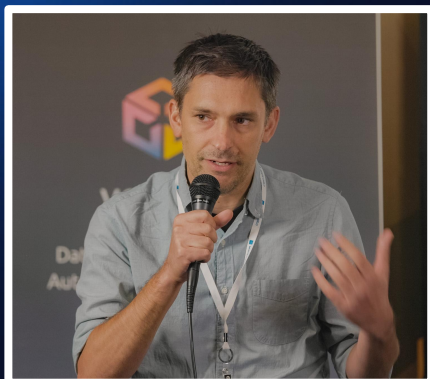
On Centralized vs. Embedded ML Teams

Shane: Until around 2018, we were a centralized group with about 50 people working very closely between engineering, analytics, and data science. Once we matured, we moved our analytics teams further from the central data team.

We made data systems and tooling accessible in a central data platform so the team could start working cross-functionally with our product teams. This was only possible after we centralized our team first, and required both strong leadership and cooperation from the data side and product side.

Companies need to centrally build and mature their data team first before distributing it into the organization: “centralize and then decentralize”. - Shane Murray

Justin: Some teams benefit more from a centralized structure than others. If you're on a product analytics team, you need to know who can answer questions, approve decisions or delegate resources. If you're on a machine learning team that builds and deploys real models, you need an embedded structure that allows you to remain compatible with the engineering process.



Shane Murray



Justin Norman

Talent Acquisition And Retention in ML

Shane: Once we managed to hire early talent, it gave us a momentum boost as we could hire more A-players that wanted to join a meaningful team. When hiring, focus on people who can complement or add to your team skill set.

Managers should value leadership skills over individual contributor skills. Junior managers should focus less on what they're doing individually, and more on managing and improving their star team.

Providing autonomy for teams will boost employee morale and help with talent retention. - Shane Murray

Internal talent development is almost always the priority. Only when you're facing a new domain, you might need to bring in outside domain experts.

Justin: If you're onboarding talent from outside, you must do due diligence before starting the hiring process. This includes defining the role clearly and how it fits into the team, think about career trajectories, and design a smooth onboarding procedure to reduce training time.

During the first 3 months, people develop an opinion of their new job and colleagues. It's critical to guide new employees in that timeframe to align their technical skills and desired career prospects with your organization's goals.



Shane Murray



Justin Norman

Facing Challenges With Remote Work

Justin: It was a very stressful time for us during lockdown. I had joined Yelp's executive team a couple of months prior and we struggled leading the business remotely. Acquiring talent became a challenge because we couldn't connect with people on a personal level about career paths, assistance and mentorship.

As data scientists working primarily on computers, we took our systems of collaboration for granted because we used to work next to each other with direct lines of communication.

During the pandemic, we had to develop systems and tools to improve our remote collaboration process. - Justin Norman

What helped us was building the concept of collaboration into project and program management that had previously happened organically. This helped a lot, especially when recruiting people from outside the country.


Shane: The biggest challenge when we went remote (especially for new juniors) was the energy around the mission and the impact of their work. Finding ways to collaborate and grow in-person was critical for us, so we shifted into a hybrid model once we started going back to the office.



Vectice auto-documents your AI team's work by capturing important ML assets and augmenting them with available metadata (regardless of platform distribution).


Vectice helps enterprise data science teams automatically to manage projects and build compliant models that can be deployed with a long-term runtime.

Vectice speeds up new iterations of existing models and reduces onboarding time for new talent. This allows your AI/ML team to measure and analyze existing team efforts.

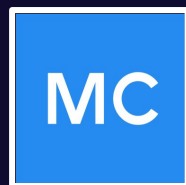
 Share project knowledge

 Up to date documentation

 Track assets, versions & lineages

 Showcase your team's impact

[Learn more about Vectice](#)



Monte Carlo

As businesses increasingly rely on data to power digital products and drive better decision-making, it's mission-critical that this data is accurate and reliable.

Monte Carlo's Data Observability Platform automatically monitors and alerts for data issues across your warehouses, lakes, ETL, and business intelligence tools.

Monte Carlo provides automatic, field-level lineage, root cause analysis, and centralized data cataloging to better understand the accessibility, location, health, and ownership of their data.

[Learn more about Monte Carlo](#)