Rise of the Industrial Data Scientist

By Peter Reynolds

Keywords

Artificial Intelligence (AI), Industrial IoT, AIoT, IT/OT Convergence, Aspen Technology, Pharma, Biotech, Industrial Data Scientist

Summary

Executive panel led by ARC Advisory Group's Peter Reynolds, Principal Analyst and Heiko Claussen, Senior Vice President for Artificial Intelligence at Aspen Technology, David Leitham, Senior Vice President and General Manager at Aspen Technology and Jose Valls, Microsoft CTO of Process Manufacturing and Energy addressed the growing role of industrial data scientists in key industries. While the industrial sector is undergoing a transformation driven by AI and the Internet of Things (IoT), simultaneously the workforce is shifting as traditional domain experts are replaced with tech savvy workers who bring a new level of operational expertise. Enter the industrial data scientist, a new breed of data analyst with access to more industrial data than ever before and the advanced technology to translate that information into actionable intelligence.

Key factors influencing the rise of the industrial data scientist include:

- Organizations are unable to realize the full value of Industrial AI due to poor industrial data quality and management, internal silos, and a lack of collaboration among relevant teams.

- The self-sufficiency that an industrial data scientist brings to the table helps address innovation and solve problems with greater agility and scalability.

- AspenTech's Industrial AI solution and cohesive digital reference architecture, bring data science capabilities and domain expertise together.

While the industrial sector is undergoing a transformation driven by AI and the Internet of Things (IoT), simultaneously the workforce is shifting as traditional domain experts are replaced with tech savvy workers who bring a new level of operational expertise.

ARC
Advisory Group

The Current State of Industrial AI in the Process Industries

Research from ARC Advisory Group in 2021 on the Convergence of AI and IoT (AIoT) report and The State of AI Research by research specialist Vanson Bourne accurately describe the current state of AI in the process industries. Both reports highlight the need to improve collaboration and reduce complexity and break down organizational silos between data science and domain expertise.

The term "artificial intelligence of things or AIoT" is used to describe the confluence of AI and Industrial IoT (IIoT) technological forces. AIoT is built for industrial companies looking for better ways to connect their evolving workforce to data-driven decision tools and digitally augment work and business processes. However, leveraging AI requires data science capability, which adds additional complexity to an already complex environment.

While engineering roles are skilled in analyzing large amounts of data, setting up and creating production-grade machine learning environments is not easily accomplished. Therefore, unlocking the value of industrial data through AI requires a hybrid approach.

The paradigm of Industrial AI is to deliver measurable business outcomes for capital-intensive industries. Industrial organizations don't need to be sold on the value of Industrial AI, but rather the challenge is in realizing it. The Vanson Bourne research found here surveyed over 200 IT and Operations decision-makers across industries, providing key insights into the current state of Industrial AI adoption. The study revealed the core challenges that inhibit organizations from realizing the full value of Industrial AI are poor industrial data quality and management, internal silos, a lack of collaboration among relevant teams, and a clear strategy around Industrial AI.

What Is an Industrial Data Scientist?

The traditional data scientist's role combines computer science, statistics, and mathematics. Industrial data scientists' core mission is to build more comprehensive, performant and sustainable AI/ML models that are fit-for-purpose, domain-specific and address focused, real-world use cases. They analyze, process, and model data; and have competency and knowledge of pre-processing, types of models, deployment concepts like Machine Learning Operations (MLOps), aspects of hardware deployments, or cloud and

edge deployments. The data scientist focuses more on the algorithmic parts and the toolchain improvements.

On the other hand, the industrial data scientist is a unique combination of domain knowledge with an understanding of applying AI aspects and identifying opportunities and problem solving. Equipped with the best AI tools which have been democratized, the industrial data scientist is not dependent on other organizations to analyze data and determine outcomes.

The self-sufficiency that an industrial data scientist brings to the table helps address innovation and solve problems with greater agility and scalability.

While the industrial data scientists maintain a certain level of data science acumen, they can efficiently collaborate with data scientists because they can articulate and speak the language of the data science application or product. The self-sufficiency that an industrial data scientist brings to the table helps address innovation and solve problems with greater agility and scalability. The essence of an industrial data scientist is domain expertise, combined with a robust toolchain or set of packaged programming tools to solve challenging industrial problems, such as predicting future conditions or events using industrial data and AI.

How Industrial AI Is Being Used to Resolve Challenges

Industrial AI offers a broad spectrum of use cases driven by industrial data, with predictive, prescriptive maintenance at the forefront to reduce or eliminate equipment downtime. However, the global pandemic has accelerated the industry's desire to digitalize, especially in the pharma and biotech industries. And according to David Leitham, "we've seen great efficacy in predicting device failures with great specificity which continues to drive that towards zero unplanned downtime and eliminating loss batches, which are both expensive and disruptive to the full supply chain." Advanced demand modelling, working in conjunction with planning, scheduling, and utilizing big data to anticipate shifts in and proactively adjust for demand for different therapeutics, has become increasingly important because therapeutics are becoming more targeted.

Beyond pharma and biotech in the chemical industry, it's common to have dedicated models for equipment and leverage a hybrid modelling approach. Hybrid modelling combines the first principle knowledge with experience and new insights from data. Industrial AI also helps improve model

sustainment at the edge by continuously using data to update and train in process conditions that are otherwise difficult to model, such as aging equipment. Using historical data already collected, Industrial AI can automatically build schedules or automate processes or find root causes of failures in equipment or inability to meet a daily or weekly schedule.

How AspenTech Is Bridging the Gap Between Domain Expertise and AI

AspenTech's broad portfolio of performance engineering, production optimization, asset performance management, value chain optimization and interconnected hybrid-modelling applications help chemical engineers, operations, and other engineering disciplines to collaborate and drive higher value. With AspenTech's hybrid modelling approach, the software provides a combination of first principle modelling workflows supplied by a chemical engineer, with the inclusion of data science workflows, like pre-processing, model training, model and algorithm selection. Engineers can easily collaborate to create hybrid models and jointly bring them into the data science toolchains.

Looking to one example of an industrial data scientist workflow, when flowsheet models don't exist for custom operational assets, instead of building these from scratch, data can be manually collected that best represents a broad range of operations for this equipment. The industrial data scientist in the organization would "prune the data" to ensure data quality. This pre-processing step checks for missing sensors and then builds an adequate or fitting model using AspenTech's model builder solution for this particular use case. This hybrid model then is enriched with physical constraints to enforce mass balances or other criteria specific to the use case. Once imported into flowsheet simulators, the chemical engineer and industrial data scientist can jointly optimize the model. This approach brings the best of both worlds together in one offering and the advantage of using the data from the field to inform/update the process model.

Conclusion

With more than 40 years of experience and focus on the industrial manufacturing space, each aspect of the AspenTech solution is tuned to the customers' and user persona and requirements with domain expertise. Each of these solutions sits on a cohesive digital reference architecture, which

helps bring all those capabilities together. Bridging the gaps and expertise allows each expert to contribute where they add value and have the comfort level with an application and interface tuned to them holistically; these solutions come together to solve broader problems.

Industrial companies will continue to look for better ways to connect their evolving workforce to data-driven decision tools and digitally augment work and business processes. However, leveraging AI requires data science capability, which adds additional complexity to an already complex environment.
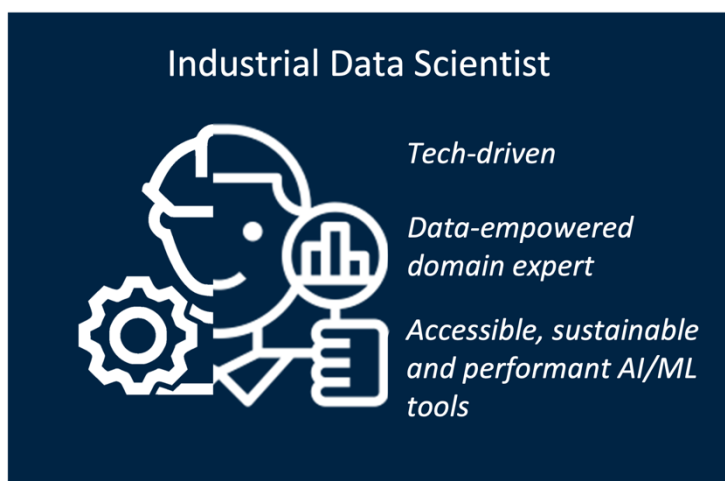
Attributes of the Industrial Data Scientist

Building organizational competency around data science is a high priority for industrial manufacturers. The investment in industrial data scientist roles and building a level of data science acumen is justified as they can efficiently collaborate with data scientists. The Industrial Data Scientist is a new breed of tech-driven, data-empowered domain experts with access to more industrial data than ever before, as well as the accessible AI/ML and analytics tools needed to translate that information into actionable intelligence across the enterprise. Many data scientists in the industrial sector today come to the job with a background in chemical, petroleum or industrial engineering and not computer science or software engineering.

Industrial Data Scientists focus on solving real-world problems in the field. They draw on their domain experience to incorporate domain knowledge into data science projects – a level of expertise that traditional data scientists don't naturally carry. ARC Advisory Group believes existing roles such as

the Advanced Process Control (APC) engineer are great examples of complementary skills and areas to focus in building internal competency. Key factors in ensuring a successful Industrial Data Science competency program include:

- Simplification of the computing AI/ML infrastructure.
- Simplify AI/ML deployment.
- Incorporate domain expertise collaboration techniques.
- Consider existing organizational capabilities in Advanced Process Control and Modelling as a starting point.

For further information or to provide feedback on this article, please contact your account manager or the author at preynolds@arcweb.com. ARC Views are published and copyrighted by ARC Advisory Group. The information is proprietary to ARC and no part of it may be reproduced without prior permission from ARC.