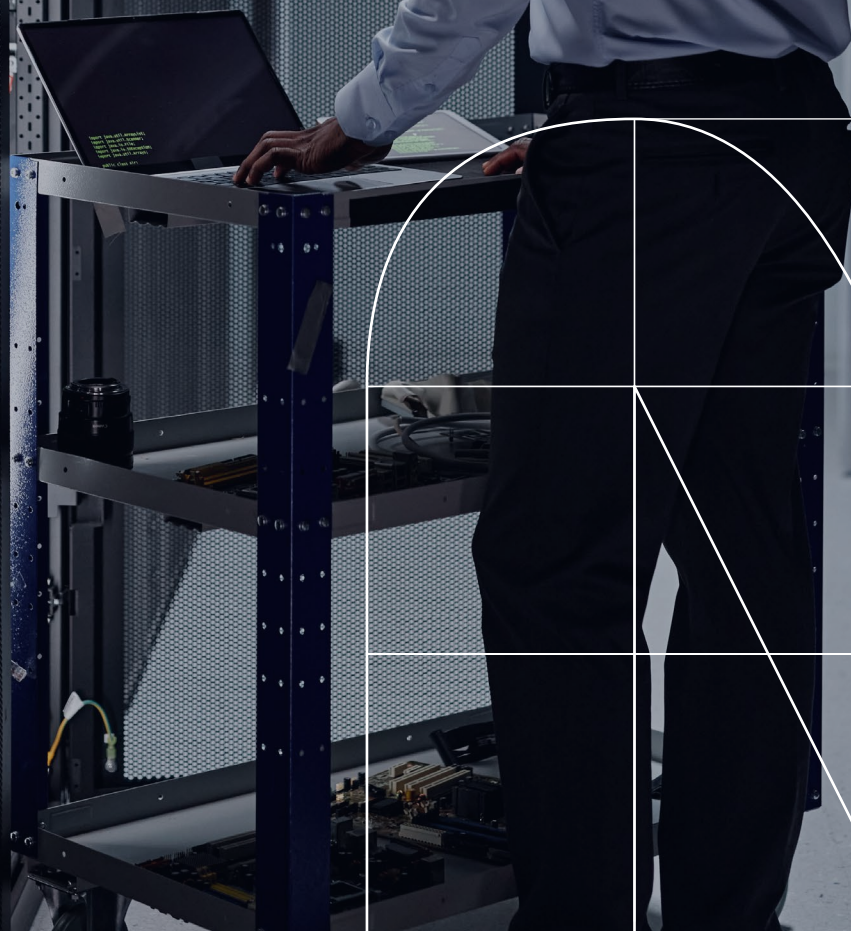NTT DATA

The NTT DATA virtual data scientist approach

Sustainable data science: solutions for meticulous results

Unlock the business potential of AI/ML with access to advisory services, expert consultations and the development of data science-related projects on a subscription-based commercial model.

Contents

A new approach to data science

In the rapidly evolving field of artificial intelligence (AI) and machine learning (ML) (AI/ML), businesses are confronted with both immense possibilities and a maze of complexities, which can prove overwhelming.

The many diverse AI/ML technologies, methodologies and frameworks available make it difficult to choose the right approach. Limited standardization, interoperability issues and the constant influx of new tools contribute to confusion and highlight the complexities that hinder seamless adoption.

In addition, technological advancements often outpace businesses' ability to keep up, leading to hasty adoptions without considering long-term implications or alignment with organizational goals.

In the face of this complexity and confusion, there is a critical need for solutions that align with the dynamic nature of data science.

The traditional, static models of software development fall short in addressing the inherent fluidity of AI/ML projects. What businesses need is not just a tool but a transformative approach – one that evolves with the ever-changing demands of data science, enabling them to adopt these transformative solutions and benefit from data-driven innovation.

Enter virtual data scientist: an evolutionary paradigm designed to simplify the complexities of the current AI/ML landscape.

Virtual data scientist acknowledges the challenges businesses face and offers a dynamic, tailored and navigable path forward.

In this paper, we explore the intricacies of virtual data scientist to understand how it mitigates the challenges presented by the current state of AI/ML and how it can propel businesses into a future where data science is both seamless and agile.

The imperative of AI/ML for businesses

There are compelling reasons behind the push for businesses to embrace AI/ML. Innovation, efficiency and strategic decision-making are some of the transformative benefits that these technologies bring to the forefront.

In a rapidly evolving digital landscape, AI/ML is not so much an option or a technological investment as a strategic imperative for businesses aiming to thrive and lead. From driving innovation to enhancing decision-making and optimizing operations, the benefits of this transformative technology are vast and multifaceted. Businesses that recognize and act on this imperative are the ones poised to redefine the possibilities of success in the digital age.

Driving innovation

At the heart of the AI/ML revolution lies the promise of unparalleled innovation. These technologies empower businesses to explore uncharted territories, uncover novel insights and devise solutions that were once considered unattainable. AI/ML enables organizations to foster a culture of continuous experimentation, push the boundaries of what's possible and stay at the forefront of their respective industries.

Enhancing decision-making

The ability to make informed, data-driven decisions is a cornerstone of competitive advantage. Fueled by vast amounts of data, AI/ML algorithms provide a level of predictive and prescriptive analytics that traditional methods cannot match. By leveraging these technologies, businesses gain a strategic edge, enabling them to foresee trends, identify opportunities and navigate challenges with unprecedented precision.

Optimizing operational efficiency

AI/ML excel in automating and optimizing processes, freeing up valuable human resources for more strategic and creative endeavors. From predictive maintenance in manufacturing to dynamic pricing strategies in ecommerce, these technologies can streamline operations, reduce costs and enhance overall efficiency across diverse industries.

Personalizing customer experiences

In an era where customer experience is paramount, AI/ML empower businesses to deliver personalized and tailored interactions. From recommendation engines in ecommerce to predictive customer service, these technologies enable businesses to understand individual preferences, anticipate needs, and create a customer-centric approach that fosters loyalty and satisfaction.

Unleashing the power of big data

The proliferation of data is both a challenge and an opportunity for businesses. AI/ML provide the tools to extract actionable insights from vast datasets, transforming raw information into valuable knowledge. The ability to harness the power of big data equips businesses with a competitive advantage, enabling them to make strategic decisions based on a comprehensive understanding of their operational landscape.

Staying competitive in the digital era

AI/ML are foundational pillars of digital transformation. Businesses that embrace these technologies position themselves as leaders in the digital era, staying ahead of competitors and establishing themselves as innovative, adaptive and future-ready entities.

"

The ability to harness the power of big data equips businesses with a competitive advantage, enabling them to make strategic decisions based on a comprehensive understanding of their operational landscape.

Current landscape: two states of data science adoption

State 1: Indefinite postponement

In this state, businesses find themselves in a perpetual loop of indecision, indefinitely postponing their adoption of AI/ML.

Given that technology and business change constantly, the illusion of being "completely ready" traps organizations in a cycle of missed opportunities and innovation stagnation. The fear of being unprepared becomes a permanent roadblock to embracing the transformative power of data science.

By delaying adoption, businesses hinder their agility and risk falling behind competitors who strategically leverage advanced technologies. This leads to stagnation in an environment where anticipating trends and personalizing solutions define success.

State 2: Impulsive adoption

Businesses in this state move too fast, hastily embracing data science and AI without careful consideration.

This impulsive approach involves a rush to implement cutting-edge technologies, often without a clear understanding of business objectives or alignment with the data science process.

The deployment of unproven or immature technologies may result in technical setbacks and disillusionment within the workforce. Moreover, the lack of transparent communication during rapid adoption may erode customer trust, raising concerns about privacy and ethical considerations. The premature pursuit of innovation without a strategic roadmap can lead to inefficient resource allocation and, paradoxically, unrealized expectations, undermining the intended benefits of the initial leap.

Indefinite postponement	Impulsive adoption
Falling behind competitors	Innovation without strategic direction
Stagnation and innovation lag	Risk of unproven technologies
Lost market share	Expectations not met
Diminished customer satisfaction	Inefficient resource allocation
Inability to anticipate market trends	Imbalanced return on investment

Striking the right balance

To avoid the pitfalls of perpetual indecision and premature adoption, organizations need to strike the right balance between the two: acknowledging the need for readiness while embracing the dynamic nature of technology.

A strategic and adaptive adoption of data science avoids the extremes of perpetual postponement or impulsive implementation, enabling businesses to harness the full potential of data science without succumbing to the risks associated with indecision or haste.

Addressing the pitfalls of current data science approaches

The success of data science projects is often hindered by adopting methodologies that are not inherently dynamic and exploratory. The shortcomings of the prevalent software development approach and pitfalls in current data science approaches underscore the need for a paradigm shift.

The rigidities of software development methodologies do not align with the iterative, dynamic nature of data science. Organizations must recognize these challenges and transition to adaptive approaches that embrace the evolving journey of data science, fostering innovation, efficiency and continuous value delivery.

Using a software development approach: a mismatch for data science

Data science projects are often subjected to the constraints of the traditional software development approach. The framework of this approach does not fit the iterative, exploratory nature of data science. This mismatch leads to a myriad of challenges which hinder the adaptability and innovation crucial for success.

Preemptively scoped projects: inaccurate timelines

The (common) tendency to preemptively scope data science projects often results in inaccurate timelines and costs, as the unpredictable nature of data science requires flexibility and adaptability. Fixed project scopes struggle to accommodate the evolving nature of insights discovery, leading to delays and frustration.

Budget overruns: the cost of unpreparedness

The software development mindset, which focuses on predefined requirements, can lead to substantial upfront investments in data science projects. However, without aligning with the evolving needs of dynamic data-driven requirements, the costs can escalate rapidly. The expense of retrofitting or adjusting strategies midway through a project can be significant, impacting the overall feasibility of data science endeavors.

Lack of data-driven insights: missing the essence

One critical pitfall is the failure to generate truly data-driven insights. The traditional software development approach tends to prioritize predefined requirements over the iterative exploration and experimentation that data science demands. This oversight can result in projects that lack the depth of understanding derived from harnessing the full potential of available data.

Nondynamic approaches: inability to adjust

The inherent dynamism of data science projects clashes with nondynamic methodologies. The inability to adjust dynamically based on evolving solutions, emerging insights, or changing business requirements stifles the potential for innovation. The rigidity of such approaches hampers adaptability, a crucial element in the rapidly changing landscape of data science.

Lack of continuous value flow

A notable shortfall in the current ways of working is the absence of a continuous value flow. The traditional software development model often results in projects that do not seamlessly integrate with broader organizational goals or provide ongoing value. This discontinuity disrupts the innovation flow, hindering the realization of the full potential of data science.

Embracing the scientific essence: the data science process unveiled

The timeless principles of scientific inquiry are the foundation of data science. Against the backdrop of prevalent industry trends and fast-paced IT landscape, the importance of adhering to the scientific method in data science becomes increasingly clear.

Defining data science

Data science, in its truest form, is not merely the application of cutting-edge technology or the deployment of the latest trends. It is, at its core, an iterative and ongoing process of deriving insights from data through a systematic approach of hypothesis formulation, testing and refinement.

"

The overarching goal is not to deliver a software project but to glean meaningful information from observation.

The data science process

The data science process unfolds in a series of four interconnected steps:

Hypothesis formulation

Initiate the process by formulating hypotheses based on the questions at hand. These hypotheses serve as educated guesses about the relationships within the data.

Implications calculation

Calculate the implications of the formulated hypotheses, predicting outcomes and patterns that can be expected if the hypotheses hold true.

Comparison to observation

Compare the calculated implications with real-world observations or experimental data. This step serves as the litmus test, determining the accuracy and validity of the hypotheses.

Refinement or denial

Based on the comparison, refine the hypotheses if they align with observations or deny them if inconsistencies arise.

This iterative cycle continues until robust insights are derived.

Why traditional project styles don't work for data science initiatives

The traditional project styles prevalent in the IT industry, often driven by delivery-oriented methodologies, pose a challenge when applied to data science initiatives.

The data science process requires a fluid, hypothesis-driven approach that contrasts sharply with the defined and entrenched ways of work characteristic of traditional software development life cycles (SDLCs). While delivering a software project is a success in an IT sense, it falls short of embodying the essence of data science, which demands an ongoing execution of the scientific method for information extraction.

The importance of following the data science process

Following the data science process ensures that it's the data which drives your initiatives, rather than technological trends. The iterative nature of this process accommodates the inherent unpredictability and complexity of data science projects.

By remaining faithful to the scientific method, organizations can uncover profound insights, make informed decisions and avoid the trap of prioritizing deliverables over the genuine pursuit of knowledge from data observations.

In essence, the advantage of the data science process lies in its ability to transform data into actionable intelligence through a scientific lens, thereby transcending the confines of traditional project styles.

Introducing the evolutionary catalyst: virtual data scientist

Against this backdrop of dynamic challenges and relentless innovation, virtual data scientist (VDS) presents a solution that transcends conventional boundaries and redefines the way organizations approach data science.

Virtual data scientist is a model we use for data science projects. It's a way of work that enables us to develop sustainable data science solutions by ensuring that the data science methodology is meticulously followed.

The virtual data scientist approach is a distillation of the lean agile methodology, which aims to offer a continuous flow of value. Virtual data scientist offers organizations access to advisory services, expert consultations and the development of data science-related projects on a subscription-based commercial model.

This approach prioritizes continuous value delivery throughout the data science lifecycle, driven by regular feedback, improvements and dedicated customer support. It ensures that clients benefit from a personalized experience, tailored to their specific business environment and requirements.

Our Virtual Data Scientist solution

Our approach balances the evolving demands of data science with adherence to standardized methodologies, guaranteeing both flexibility and reliability. And, by consistently ensuring the right expertise is available at each project stage, we guarantee optimal resource utilization.

| Client stakeholders | Senior data scientist | Skilled resources | One in / One out | Virtual resources | |

Our subscription-based commercial model gives clients access to data science expertise and solutions in a way that is flexible and ensures a continuous flow of value.

Benefits to clients

With our approach, clients benefit from:

- **Low upfront costs:** A subscription-based model gives clients room to save costs by paying a regular, low subscription fee as opposed to a huge outlay of capital at the commencement of a project.

- **Reduced risk:** Clients are exposed to minimal risk as they may cancel their subscription if they are not satisfied with the services.

- **Continuous flow of value:** This prioritizes ongoing value at every step of the data science lifecycle. Regular feedback loops, regular improvements and ongoing support mitigate the risks of rigid scope.

- **Sustainable solutions:** Our data science process relies heavily on the guidance of data and metrics to track the effectiveness of our solutions.

- **Customized solutions:** We understand that every business is unique, so the discovery, extraction and reshaping of relevant raw data is specifically tailored to your environment and requirements.

How we deliver

Virtual Data Scientist Seat

The Virtual Data Scientist Seat is occupied by different individuals at different stages of the project, based on the need for a particular skill set at the time. The objective is to "fill the seat" with the most qualified individual for each task, creating a highly efficient and adaptable virtual data scientist environment.

This flexible approach to staffing the Virtual Data Scientist Seat gives our clients access to the full spectrum of expertise and talent in our team, and ensures that the right person with the right skills is available to address the unique demands of the project as it evolves over time.

Project Lead

A Senior Data Scientist is assigned as the Project Lead to oversee the project at a high level and take on various essential responsibilities, including:

- **Strategic direction:** Defining the strategic vision for the project, setting clear objectives and ensuring alignment with the client's overarching goals while adhering to data science processes

- **Resource allocation:** Efficiently allocating project resources, including people and technology, to maximize productivity

- **Quality assurance:** Stringent quality control measures are crucial to ensure accuracy and compliance with industry standards

- **Stakeholder communication:** Establishing and maintaining clear lines of communication with the client and NTT DATA stakeholders, keeping them informed of project progress and addressing any concerns or questions

- **Risk mitigation:** Identifying potential risks and challenges that may arise during the project and developing mitigation strategies to ensure successful project outcomes

- **Weekly meetings:** Organizing and leading weekly meetings with the project team and stakeholders to provide updates, discuss challenges and make necessary adjustments to project strategies

- **Reporting:** Generating comprehensive reports that offer insights into project progress, key performance indicators and milestones achieved, ensuring transparency and accountability

Features of our approach

Affordability

Virtual data scientist operates on the principle that transformative data science solutions should be accessible to organizations of all sizes. Affordability is not merely a feature but a fundamental tenet, ensuring that the power of data-driven insights is democratized across the business landscape, and businesses of any size can afford this approach.

Dynamic flexibility

Virtual data scientist is designed to be flexible. It embraces the iterative and evolving aspects of the data science process, giving organizations the agility to adapt to changing requirements and emerging insights seamlessly.

Aligned to the data science process

Rather than imposing a rigid framework, our virtual data scientist approach integrates seamlessly with the iterative journey of hypothesis formulation, testing and refinement. This alignment ensures that the essence of data science is preserved, driving insights from the data itself.

Constant stakeholder engagement

Recognizing the integral role of stakeholders in the success of data science initiatives, we foster constant engagement. Stakeholders, with their domain expertise, contribute throughout the data science process, enriching the outcomes and ensuring that the solutions align with organizational objectives.

Ability to pivot

In the face of evolving challenges and opportunities, virtual data scientist empowers organizations with the ability to pivot, allowing them to navigate the intricacies of data science projects by making informed adjustments based on real-time data and market dynamics.

A sustainable solution

Virtual data scientist is not a transient solution but a sustainable ecosystem that adapts to the changing needs of data science. It ensures that organizations can rely on a continuous and enduring source of data-driven value.

Enable data-driven decision-making with virtual data scientist

Virtual data scientist is a paradigm shift in the approach to data science.

By aligning with the scientific method and prioritizing continuous value delivery, it transcends the limitations of traditional project styles. Its affordability, dynamic flexibility, alignment with the data science process, constant stakeholder engagement, ability to pivot and sustainability make it the ideal choice for organizations seeking to navigate the complexities of data science in a transformative and efficient manner.

Virtual data scientist guides organizations toward a future where data-driven decision-making is not just a goal but part of their DNA.

Contact us

We welcome any inquiries about this document and its contents.

Please contact:

Darshil Ramchander (MEA)
Senior Data Scientist
+27 11 575 7998
Darshil.ramchander@global.ntt

Chris Wigget (MEA)
Head: Data & Analytics
+27 11 575 7080
chris.wiggett@global.ntt

List of abbreviations

AI	artificial intelligence
ML	machine learning
VDS	virtual data scientist
IT	information technology
SDLC	software development life cycle
VDC	virtual data center
MEA	Middle East and Africa

Learn more about NTT DATA
services.global.ntt

NTT DATA