run:ai

Ready for Success with AI?

A Benchmarking Model for AI Infrastructure Maturity

Introduction

In the spring of 2020, as the novel Coronavirus spread widely around the world, a research university in London made headlines for their Natural Language Processing research that helped NHS clinicians understand a wider range of Covid symptoms than they had seen previously. Their research saved critical time by getting patients the right treatments, early enough to make an impact on patient outcomes.

One day all enterprises will use AI, and specifically Machine and Deep Learning (ML / DL), to quickly respond to new realities and market changes. However, today, few companies have the ability to use DL and ML to problem solve so quickly.

Why?

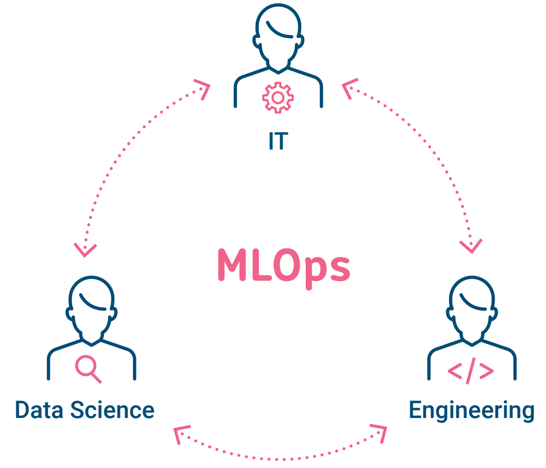
Many enterprises are experimenting with AI but have not yet learned how to complete complex machine or deep learning (ML/ DL) projects and bring them to production. Developing DL models involves considerable IT and infrastructure skills. Huge data sets have to be aggregated, stored, moved, protected, and managed. Training and testing models require high levels of compute capacity and performance and moving models to production requires complex infrastructure setup and cumbersome model optimizations. These practices fit into a new discipline that is often referred to as 'MLOps'. Machine Learning Operations (MLOps) is defined as the practice of operationalizing and managing the lifecycle of Machine Learning development. When MLOps is working well inside an organization, their AI initiatives make it to market quickly.

AI success is rooted in seamless collaboration between data science researchers, IT teams that provision resources, and software engineers who implement the models in a production environment. **The observations presented in this ebook are based on interviews with dozens of enterprises with deep learning teams across many industries and locations.** We have used their collective wisdom to present a model that can help enterprises benchmark themselves – providing insight by using the experience of mature companies to help guide their journey into advanced Deep Learning.

What is MLOps?

MLOps arose from the understanding that in the ML development lifecycle there is a gap between what Data Science needs and what traditional IT processes should provide.

- Data Scientists are researchers. In an efficient system, they should focus on building ML models without needing to also manage the infrastructure required to run machine learning and deep learning. Concepts such as containers, drivers, and operating systems ideally should be as 'invisible' in their routine work.
- Traditional IT should focus on systems, as per their usual routine, without having to manually allocate resources.



The AI Infrastructure Maturity Model

In talking to many enterprises, we found it beneficial to measure the maturity level of AI (and specifically Deep Learning) by studying two things:

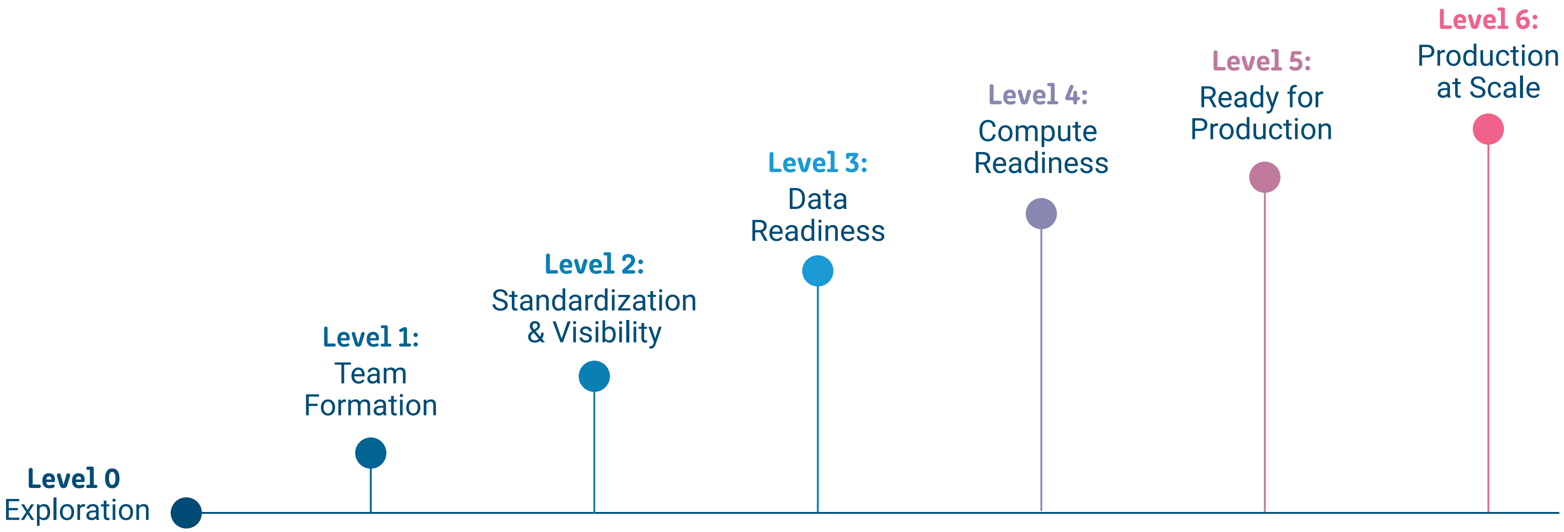
- Their readiness for fast delivery of AI models
- Their ability to scale AI capabilities

Benchmarking along these criteria follows an organization from the early days of ML and DL where ML work is done by single individuals, to advanced-stage organizations that are extremely proficient in using ML and gaining the most value from their data.

The following maturity model most closely relates to organizations working on DL and using specialized hardware accelerators such as GPUs for building and training models. Note that real life includes many more components of AI that influence maturity at the various stages. **For the purpose of clarity (and brevity), this e-book will not cover all areas of the MLOps lifecycle, but instead focus on optimization of the specific areas where data science interacts with IT, the AI Infrastructure stack.**

As in many organizational transformations, deep learning initiatives evolve across a maturity model: a pathway or journey that takes an enterprise from an initial stage through stepwise optimizations to a highly mature transformative level. At each stage in the AI infrastructure maturity model, the organization benefits from a higher degree of DL productivity that propels the company towards business-critical objectives.

Our Maturity Model starts with ML exploration and continues through standardizations and optimizations to a level of data science that consistently generates transformative business value. We invite enterprises to use this model to benchmark their current AI Infrastructure readiness and to map their next steps on the path to a transformational level of MLOps maturity, particularly as it relates to mastering the many infrastructure challenges they encounter when implementing AI.



The AI Infrastructure Maturity Model Levels

The model is comprised of six levels that encompass the milestones achieved as data science and IT work more closely together. Each stage brings the organization closer to maximizing the business value of its machine/deep learning efforts.

run:ai

Level 0: Exploration

As companies begin their AI journey, they initially want to see if advanced AI capabilities such as deep learning can help to capture real value for their particular business. In the exploration phase, we often see a small data science team, using whatever data and compute resources are readily available, carrying out a few proof-of-concept projects.

If at some point the company chooses to embrace data science as a strategic initiative, it must start to create an ecosystem in which data science projects can consistently and productively move from research to production.

> “
>
> ***I see our company as ready for DL, but the experience so far has been a bit of a jungle, without clear processes.***
>
> IT Leader at Fortune 2000 Company
>
> ”

Level 1: Team Formation

TL; DR

- Organization forms a small research team
- Compute resources are acquired and allocated to researchers statically
- Manual scripts are used for preparing data and running models
- Datasets and scripts are stored on local machines

In Level 1, the organization forms a small team of researchers that is tasked with research around specific business goals.

Team members in a smaller research organization often use containers to enable standardization of tools and workflows, using scripts that are stored and maintained in a centralized repository. Though there can be considerable value in utilizing cloud servers for ML and DL projects, many of the companies we interviewed are utilizing bare metal GPU for their initial projects. IT provisions a workstation or two and the researchers can begin to experiment. This helps companies control costs and determine if a bigger investment is relevant.

IT Support at Level 1

As smaller research teams typically share infrastructure like GPU in small silos, scheduling machine usage is handled manually among the team members themselves, often in spreadsheets. IT is often absent from data science plans at this stage. One disadvantage of the lack of support from IT is that both scheduling and data set management are done by each data scientist individually. Data is typically stored locally, and has to be moved manually from machine to machine.

Maturity Score: Low

In this early maturity level, the research and IT teams still have limited visibility into each other's priorities, workflows, or needs. Resources are managed ad-hoc. Resources are managed on an ad-hoc basis and resource sharing is limited and inefficient. **Although an increasing number of models are being explored, the machine/deep learning processes are mostly manual.**

Level 2: Standardization & Visibility

TL; DR

- IT starts establishing organizational control
- Policies begin to be formed around data access
- Monitoring tools are installed and dashboards are created
- Container images are standardized and centralized

At level 2 monitoring tools are installed and dashboards are created to provide a clear picture of how teams operate and how GPUs are utilized within the organization. Visibility also requires that some standardization of data science workflows begins.

IT Support at Level 2

With greater visibility, IT supports data science by creating policies around user access, and more importantly, setting policies for prioritization of jobs based on business initiatives. Policy-based rules are established at this phase, requiring IT and DS teams to work together to set those rules.

As the research team becomes more productive and more compute and storage resources are required, IT looks to establish organizational controls. Policies are formed and enforced around data access, container images are standardized and centralized, and IT best practices are applied to hardware consolidation and procurement, software installations, use of operating systems, drivers, and so on.

There are many examples of standardization – and in this phase best practices begin to emerge. For example, container standardization includes having different Docker images for Jupyter notebooks, Pycharm, and more, as well as customized images that support Tensorflow and Pytorch. Docker images should support common python ML libraries like Panda, scikit-learn, pillow, and more. Images shouldn't include datasets or scripts, which should instead come from a mounted storage volume.

Maturity Score: Growing

Growing visibility is the **enabler for informed decisions on how resources can be better optimized and helps determine if new infrastructure investments are required.** Being able to see and track users, to view resource utilization by user/group/job over time, the wait/runtime of jobs, under- and over-utilization of resources, yields greater efficiency. Visibility enables greater control for admins, like setting policies for job priorities, container standardization, and best practices for working with docker and data science frameworks and libraries.

The Run:AI Monitoring and Cluster Management tool greatly improves productivity by giving IT leaders a holistic view of GPU infrastructure utilization, usage patterns, workload wait times, and costs.

> “*RunAI's system provides a clean interface for our researchers to submit and manage jobs and a comprehensive job dashboard with advanced analytics that allows us to monitor system use and user needs. All these capabilities ensure we can make optimal use of all GPU resources, train larger models, and minimize research downtime.*
>
> Chief Scientist at a Large University Research Center ”

Level 3: Data Readiness

TL; DR

- Automating data pipelines begins
- High-performance storage system is procured to centralize datasets
- Datasets are tracked and versioned

As data science initiatives progress, perhaps some begin to develop specific goals for production-worthy projects and the research teams begin to grow in size. Scope of their projects increase as well. More and more data are being accumulated and need to be organized and labeled in order to be useful for research initiatives.

In level 3, data science teams focus on building automated pipelines for ingesting new data, cleaning and transforming data and creating and storing datasets for experimentation. As the number and size of datasets increase, the organizations in this phase often take the next step and procure a high-performance storage system that makes the data accessible to all machines. Changes in the centralized datasets are tracked and versioned.

IT Support at Level 3

Centralization of the data infrastructure gives IT far greater visibility into and control of data utilization. Granular data access controls determine which users and groups can access which data sets and when. Backup and recovery processes are implemented to protect data assets against corruption, loss, or accidental erasures.

Maturity Score: Medium

As IT is brought in to help centralize data, speed of experimentation improves considerably, with monitoring tools and dashboards providing a clear picture over time of how data is accessed by users and groups.

Level 4: Compute Readiness

TL; DR

- Team members share GPU resources
- Resource allocation via spreadsheet is replaced with automation
- Pool of GPUs are shared between teams according to priorities and pre-defined fairness rules
- Scheduling enables prioritization and policy-setting
- Unattended workloads can be paused and resumed

Now that data is being managed across a centralized infrastructure, the organization's machine/deep learning initiatives can start to scale.
At this stage, additional data scientists are hired and perhaps multiple teams are working in parallel on diverse projects. Researchers are running many training jobs concurrently, occasionally using multiple GPUs for a single training session. They require centralized and dynamic consumption of GPU resources so that training workloads can be allocated automatically and run unattended. The resource-scheduling spreadsheet is replaced with a scheduler, built in coordination with IT, or by using orchestration software, like Run:AI, built for this purpose.

IT Support at Level 4

As data science initiatives scale, high-end GPU servers are procured as a centralized hardware pool to dynamically support the fluctuating demands for GPU compute power. Researchers easily provision resources for code development or training models, with little or no IT intervention. The GPU pool is shared among teams and users according to business priorities and predefined policies. Teams are able to launch batch training jobs that start and stop automatically.

Maturity Score: Medium/High

Centralizing compute resources and applying advanced scheduling mechanisms provide companies at this level with capabilities like batch scheduling for unattended training. Networking and communication can be automated between machines using a cluster orchestrator. Smart queuing ensures that allocation of resources can be automated and optimized. **The tools are in place to optimize resource utilization and align resource allocation policies with business priorities.**

Run:AI's platform helps customers in this phase by offering a Kubernetes-based scheduler to optimize resource allocation and utilization. Run:AI pools compute resources and then applies advanced scheduling to dynamically set policies and orchestrate jobs. IT gains full control over GPU utilization across machines, clusters, and sites, while data scientists gain easy access to compute when and how they need it.

> “
> ***Our training workloads require and made use of containerization, but lacked the hardware resource optimization, prioritization, and job management capabilities we needed. Run:AI enabled teams to fairly share resources based on pre-set policies, and to confer resource elasticity for when additional capacity is available.***
>
> Data Science Team Lead, Fortune 500 Company
> ”

Level 5: Ready for Production

TL; DR

- Researchers easily tune model hyperparameters, running distributed training across multiple GPU
- Build, train and inference workloads all dynamically get the resources they need
- Models are optimized for - and launched in - production environments often organized by a new team, MLOps

In this maturity level, the research teams are starting to get good results and solve real business problems. As models are tested and prepared for production, the training and tuning tasks become longer and more frequent.
Data science tools for experiment tracking, results visualization, and model versioning are established, allowing researchers to spin pre-trained models, reproduce previous experiments, share models and results with other team members, and more. Researchers start experimenting with advanced methodologies such as: training jobs that are automatically distributed across multiple GPUs and machines in order to shorten training times. They regularly tune model hyperparameters by spawning swarms of jobs to test various parameter values. Automated systems can be integrated to run hyperparameter optimization processes with minimal human intervention.

In this level reliance on a new team begins to emerge – Machine Learning Operations, or MLOps. MLOps engineers define, design and build deployment workflows. They work with IT to make sure compute resources are being dedicated for models running in production, and with software engineers to optimize and prepare models for production. They focus efforts on building workflows for validating and testing models before production, and manage many important issues around DL model properties such as cost, accuracy, and bias.

IT Support at Level 5

At this phase the Machine Learning Operations and IT teams work together to support accelerated data science initiatives. They configure networks to enable distributed training and establish and enforce workload allocation policies. IT also helps set up production environments, including defining and establishing workflows around security, creating visibility and logging tools. Teams at this phase optimize utilization of GPU effectively, and DL workflows from building models, to training, to inference and production run efficiently. For example, interactive sessions for model building and debugging are directed to low-end servers, long training sessions to high-end servers, and inference sessions are conducted on dedicated servers, perhaps even using small fractions of GPUs enabling many inference jobs to run concurrently. IT infrastructure maps to the needs of the research teams at all stages, whether they are building models, training them or completing inference and readying models for production.

Level 5: Ready for Production

Maturity Score: High

IT has clear visibility into production environments and can distinguish among different types of ML workloads, such as build, train and inference, in order to optimize resource allocation. In this level, automation of resource allocation enables scaled training, hyperparameter optimization and sharing trials and results across teams. In addition, companies at this level of maturity often have a team, known as MLOps, who take responsibility for much of the lifecycle challenges of getting models into production, becoming the bridge between data science, IT and engineering.

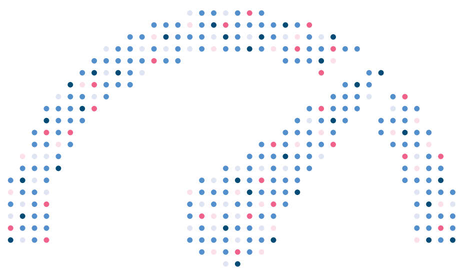
At this stage, Run:AI customers use our platform to understand how they are utilizing and managing resources across teams, clusters and nodes. Run:AI helps to run distributed training – with a seamless workflow for submitting jobs, scheduling and managing the lifecycle of the job across nodes, viewing results and logs, etc. The platform also enables hyperparameter tuning – with one command for running tens of jobs in parallel, managing and executing these jobs efficiently, and quickly. Finally, by distinguishing between build, train and inference workloads, Run:AI helps IT and data science teams define policies around their priorities, hosting machines, availability, and more.

“

We could go down the route of classic cluster job management technologies with all its limitations or bet on the scalability and robustness of newer technology, with Run:AI. It's all about ease-of-use and managing workloads while ensuring resources are appropriately managed, freed up, and available when needed. We are on a journey to change how healthcare is provided, and we will bring Run:AI along in this journey.

Lead Researcher – Healthcare organization

”



Level 6: Production at Scale

TL; DR

- Virtualization helps achieve full abstraction between DS and IT
- Teams can use multiple GPU or fractions of GPU for full optimization of compute resources
- Automation of pipelines for transferring models from research to production is in place
- Feedback loops on performance of inference models in production yield continuous model improvement

In the final stage of our maturity model, the companies we see successfully scaling and deploying models in a repeatable way have actually achieved complete abstraction between data science and IT. Perhaps surprisingly, as more and more models are being released into production and model optimization is taking place at scale, processes reach maximum efficiency so that data science and IT teams no longer need regular interaction, because their processes are automated and predictable.

MLOps engineers at companies in this maturity level are typically working on establishing automated pipelines for transferring models from research to production. Trained models are versioned and stored in a centralized location and are accessed by serving modules. Much work is done to optimize model serving, ensuring SLA requirements are met continuously and efficiently. Virtualization can help here to improve cost efficiency of expensive resources like GPUs. Monitoring tools to track performance metrics like model accuracy become important to detect drifting so that model predictions do not deteriorate over time. In addition, automated pipelines are built for retraining models in production, executed periodically or upon detection of model drifting.

IT Support at Level 6

In this stage, IT and data science teams work seamlessly because at each stage of the data science workflow - building models, training and inference - they have automated workflows and have optimized the allocation and utilization of compute and storage resources.

Level 6: Production at Scale

Maturity Score: Nirvana

Collaboration between the research and operations teams is becoming increasingly automated as models are productized, with strong feedback loops regarding the performance of inference models during testing and after deployment into production. These feedback loops continuously improve the machine/deep learning models so that they are closely aligned with business needs. With empowered IT and research teams working together in an automated way, and with full visibility across systems and workflows, **the organization is now reaping maximum business value and ROI from its machine/deep learning initiatives.**

Beginning from the initial phases, where teams are being formed and data is centralized, through full abstraction of hardware, Run:AI helps our customers identify and solve challenges of bringing AI to production. Run:AI's virtualization and acceleration platform enables companies to:

- Run training workloads on as many GPUs / resources as those jobs need, by automatically accessing idle GPU resources.
- Set policies and priorities for training jobs, so that jobs get the resources they require based on business goals and objectives.
- Pool resources to more efficiently utilize resources for build, train and inference jobs – as each of those phases of DL needs different compute and memory resources.
- Run inference tasks and lightweight training workloads in parallel on fractional GPUs to optimize efficiency and cost.
- Move models from research to production through automated CI/CD pipelines.
- Monitor production models in real time, with the results fed into an automatic retraining system.

What is AI Virtualization?

Traditional computing uses virtualization to share a single physical resource between multiple workloads. For AI, virtualization can enable acceleration of a single workload to take as many GPU resources as the workload needs, or to use a fraction of a GPU when fewer resources are needed. Cluster utilization can be optimized as elasticity is built into resources allocated to large training workloads and inference workloads can easily run on less compute resources using a fraction of a GPU.

Conclusion

The research institution we mentioned above - whose projects have direct impact on positive patient outcomes for Covid-19 - is a Run:AI customer. We've seen first-hand how data science has the potential to dramatically uplift business performance and differentiate companies in today's highly competitive markets. But capturing true business value from data science advanced data-based approaches requires a paradigm shift in how companies integrate their data science research teams into IT, engineering, and production ecosystems. A company's IT infrastructure stack must have the right tools at each stage of the process if ML-based applications are to consistently meet their business objectives.
For a company to truly reap the game-changing benefits of data science, its data scientists must be able to work hand in glove with all stakeholders across the company: from their internal business customers who specify and oversee the requirements, to the IT operations team that facilitates their infrastructure needs and the development team that builds and maintains ML-based production environments.

Run:AI's AI/ML virtualization platform is an important tool in the AI infrastructure stack. Focusing on deep learning neural network models that are particularly compute-intensive, Run:AI creates a pool of shared GPU and other compute resources that are provisioned dynamically to meet the needs of jobs in process. By abstracting workloads from the underlying infrastructure, organizations speed their AI maturity, allowing data scientists to focus on models, while letting IT teams gain control and real-time visibility of compute resources across multiple sites, both on-premises and in the cloud.

See for yourself how Run:AI can accelerate your company's journey to full AI infrastructure maturity and bring your data science initiatives to production quickly and effectively.

0 Exploration

One or two data scientists explore the business value of deep learning.

1 Team Formation

Researchers are hired, the team is assigned business goals. Static allocation of resources and manual processes are the norm.

2 Standardization and Visibility

IT starts establishing control and forms policies around data access & container standardization. Monitoring tools and dashboards are created.

3 Data Readiness

Automating data pipelines begins. High-performance storage is procured. Datasets are tracked and versioned.

4 Compute Readiness

Team members share resources in automated way. GPUs are pooled according to policies and priorities. Scheduling workloads begins and unattended workloads can be paused and resumed.

5 Ready for Production

Researchers tune hyperparameters, run distributed training across multiple GPU. Build, train and inference workloads all get the resources they need. MLOps teams form.

6 Production at Scale

Virtualization helps achieve full abstraction between DS and IT. Teams can use multiple GPU or fractions of GPU for full optimization of compute. Automation of pipelines for transferring models from research to production is in place. Feedback loops yield continuous model improvement.