

GETTING YOUR DATA AND APPLICATIONS READY FOR PRECISION MEDICINE

By Dr Frank Lee, Global Healthcare & Life Sciences Industry Leader at IBM Systems

THE CHALLENGE:

Biomedical research institutes and healthcare providers are dealing with an enormous growth of data, mainly unstructured, that is flowing from many sources – fast. This data needs to be captured, labeled, cleaned, stored, managed, analyzed and archived. The disparate file types generated create silos that impede data access, drive down efficiency, drive up costs, and slow times to insight.

The volume and complexity of data also drives the adoption of modern analytical frameworks such as big data (Hadoop and Spark) and AI (machine learning and deep learning) and applying it for the thousands of research and business applications (e.g. genomics, bioinformatics, imaging, translational and clinical). Supporting big data demands for rapidly evolving frameworks and workloads along with the collaborative nature of biomedical research requires comprehensive storage and compute capabilities.

Because of these challenges, it's imperative for the infrastructure and underlying IT architecture to become agile, data-driven and application-optimized – in short – becoming ready to advance precision medicine.

THE SOLUTION:

The IBM high performance data and AI (HPDA) architecture for healthcare and life sciences is built on IBM's history of delivering best practices in high-performance computing (HPC). In fact, the basic HPDA framework was used to construct Summit and Sierra – currently two of the world's most powerful supercomputers designed for data and AI. The architecture

is designed to help life science organizations easily scale and expand compute and storage resources independently as demand grows, to ensure maximum performance and business continuity. It supports the wide range of development frameworks and applications required for industry innovation with optimized hardware as a foundation – without unnecessary re-investments in technology.

The HPDA reference architecture is based on software-defined infrastructure (SDI) solutions that offer advanced policy-driven data and resource management capabilities. Currently it can support major computing paradigms such as traditional HPC, data analytics, machine learning and deep learning.



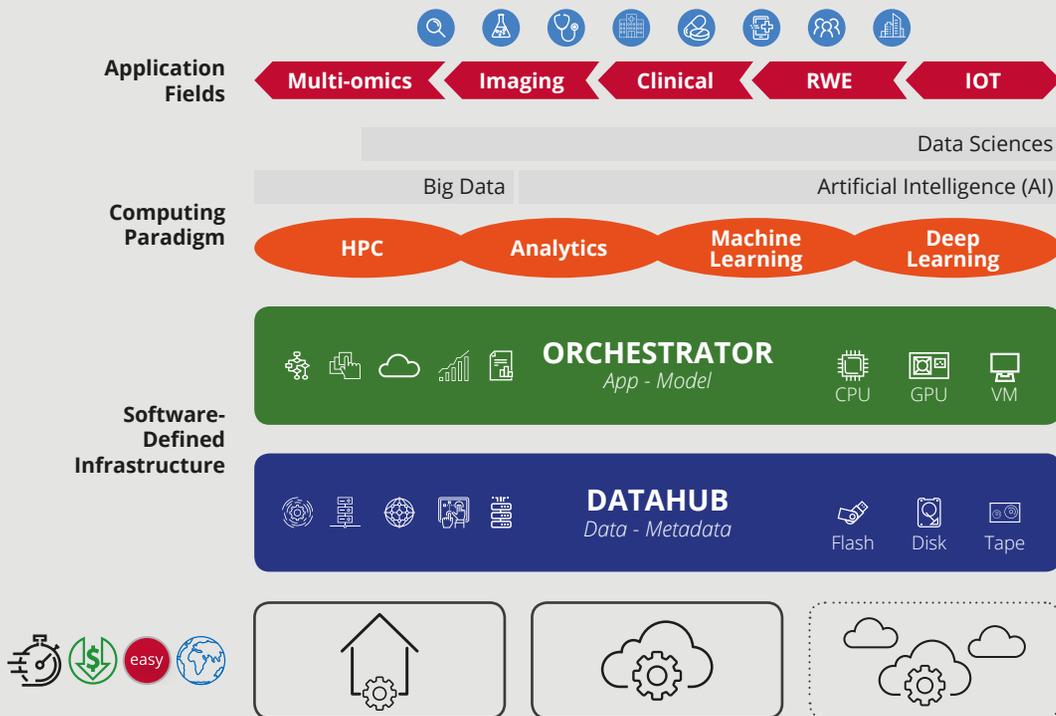
THE SECOND LAYER IS THE “ORCHESTRATOR” WHICH BRINGS EFFICIENT SCALABLE COMPUTATIONAL CAPABILITIES BASED ON A SHARED INFRASTRUCTURE TO SCHEDULE MILLIONS OF JOBS AND DEPLOY POLICY-DRIVEN RESOURCE MANAGEMENT WITH CRITICAL FUNCTIONS LIKE PARALLEL COMPUTING AND PIPELINING FOR FASTER TIME TO INSIGHTS AND BETTER OUTCOMES.”



These capabilities then become the infrastructure and informatics foundation for developing and deploying applications for fields such as genomics, imaging, clinical, real-world evidence and Internet-of-things. The HPDA architecture can be implemented on-premise in a local data center and off-premise in a private or public cloud. IBM has also demonstrated advanced use case and platforms that can be deployed as hybrid cloud.

The architecture has a “Datahub” layer designed to manage the ocean of unstructured data that is siloed in disparate systems using advanced tiering functions, peering, and cataloging. The advanced capabilities allow the data to be captured very rapidly, stored safely, accessed easily and be shared globally in the most secured and regulation-compliant way wherever and whenever the data is needed.

HPDA for Healthcare & Life Sciences - Reference Architecture





The second layer is the “Orchestrator” which brings efficient scalable computational capabilities based on a shared infrastructure to schedule millions of jobs and deploy policy-driven resource management with critical functions like parallel computing and pipelining for faster time to insights and better outcomes.

The capabilities of the datahub and orchestrator that were designed as two separated abstraction layers can be extended all the way to the cloud. They work together moving data to balance and control the dispatching of workloads avoiding bottlenecks that may cause jobs to run slower.

The logic behind this reference architecture was to free workloads from hardware constraints: assign the optimal resources needed (CPU, GPU, FPGA, VM) and address unpredictable workload requirements. This architecture is portable to different infrastructure providers, deployable to different hardware technologies and allows the workloads to become reusable through validated and hardened platforms.

 THIS TRUE DATA-DRIVEN, CLOUD-READY, AI-CAPABLE SOLUTION IS BASED ON DEEP INDUSTRY EXPERIENCE AND CONSTANT FEEDBACK FROM LEADING ORGANIZATIONS THAT ARE AT THE FOREFRONT OF PRECISION MEDICINE.”

This true data-driven, cloud-ready, AI-capable solution is based on deep industry experience and constant feedback from leading organizations that are at the forefront of precision medicine.

THE GAIN:

Users and infrastructure providers are achieving valuable results and significant benefits from the HPDA solution.

The key values for users:

- **Ease-of-use:** self-service App Center with a graphical user interface based on advanced catalog and search engines that allows users to manage the data in real-time with maximum flexibility.
- **High-performance:** cloud-scale data management and multicloud workload orchestration allows users to place data where it makes sense and provision the required environment for peak demand periods in the cloud, dynamically and automatically, for as long as needed, to maximize performance.
- **Low cost:** policy-based data management that can reduce storage costs up to 90% by automatically migrating file and object medical data to the optimal storage tier.
- **Global collaboration:** allows multi-tenant access and data sharing that spans across storage systems and geographic locations enabling many research initiatives around the globe with a common reference architecture to establish strategic partnerships and collaborate.

The key values for infrastructure providers:

- **Easy to install:** a blueprint that compiles best practices and enables IT architects to quickly deploy an end-to-end solution architecture that is designed specifically to match different use cases and requirements.
- **Fully tested:** IT architecture based on a solid roadmap of future-ready proven infrastructure that can easily be integrated into the existing environment protecting already made investments, especially the hardware purchase and cloud services.
- **Global Industry Ecosystem:** wide ecosystem to align with the latest technologies for hybrid multicloud, big data analytics and AI to optimize data for cost, compliance and performance expected by end users.

For more information: read the [Infographic](#) and watch the [Video](#).