# Breakthrough research with Databiology and IBM Spectrum Computing

*Scalable omics informatics and information management on IBM software defined infrastructure*

## Highlights

- Scalable pan-omics information management and orchestration delivered as a service on high performance computing (HPC) cloud environment from IBM® Spectrum Computing and SoftLayer® including IBM Aspera® software and IBM POWER® servers

- Central hub to manage data and to orchestrate activities using the IBM Spectrum Computing software defined infrastructure (SDI) solutions and based on the IBM Reference Architecture for Genomics

- Omics support for genomics, transcriptomics, proteomics, and metabolomics, and others

Life science and healthcare companies today find the management and processing of omics big data increasingly complex and resource intensive. This is due to the use of diverse sets of technologies such as scientific instruments, analysis software, data formats and IT infrastructure paradigms. Vast quantities of interdependent data are generated at exponential rates. Taking advantage of this volume and variety is critical to success for enterprises and institutions.

Databiology for Enterprise (DBE) is a secure information management and process orchestration solution for omics data and applications in the cloud and on-premises datacenters. DBE connects data, analytical engines, and reporting tools for research and clinical workflows. DBE is integrated with SDI solutions from IBM Spectrum Computing to distribute data-heavy computing across hybrid infrastructures and geographies. This integration optimizes use of scarce computational resources to deliver faster time-to-completion and discovery.

### Accelerating "Instrument to Insight" workflows in life sciences

Databiology for Enterprise (DBE) software offers high performance omics instrument to insights (I2I). DBE offers configure, command and collaborate role-based functionality, pan-omics information management capabilities and I2I process orchestration, using a powerful front-end and API. These features allow researchers to configure processes and compute assets, command applications and analysis work, and to collaborate with

colleagues and partners—all with greater ease. DBE solutions are also scalable, to protect your investment as demands and requirements change.

To facilitate faster insights, Databiology has integrated offerings with key IBM solutions and tools. Customers can expect faster time to results for next generation software, translational analytics and personalized medicine research. DBE Integrated offerings with IBM support proprietary and commercial applications including Databiology.

- Improved collaboration and sharing across domains and geographies
- Reduced hardware sprawl and storage costs
- Faster IT response and support of new models and technologies

IBM expertise adds unique, tailored capabilities for the integrated DBE offerings. The IBM Reference Architecture for Genomics and IBM Spectrum Computing SDI solutions including IBM® Spectrum LSF and IBM Spectrum Scale™, provide optimized workload, resource, and data lifecycle management in the cloud, on or off premises and in hybrid models. IBM SDI solutions offer high performance, low latency systems management solutions and services to pool technical computing resources, manage them efficiently across multiple groups and get the most out of IT investments.

- IBM Spectrum LSF is a powerful workload management platform for demanding, distributed HPC environments.
- Spectrum Scale is a proven, scalable, high performance data and file management solution based on IBM General Parallel File System (GPFS™) technology. Spectrum Scale provides world-class storage management with extreme scalability, flash accelerated performance, and automatic policy-based storage tiering from flash through disk to tape.

Additional IBM solutions used in the DBE integrated offerings which help to draw insights and make better decisions include:

- Aspera FASP® (fast, adaptive and secure protocol) helps to secure and speed transport of large omics datasets across the globe.
- HPC cloud offerings delivered by SoftLayer support dynamic cloud bursting for compute resources. These solutions enable high I/O intensive production applications to help improve high performance computing in cloud environments.
- IBM POWER servers are built on a flexible, open platform with the first processor designed for big data workloads. POWER servers combine computing power, memory bandwidth and I/O in ways that are easier to consume and manage, and provide high resiliency, availability and security features.

The IBM Reference Architecture for Genomics can help gain efficiency, scalability and performance—in the cloud and on or off premises—by enabling a common shared services platform that spans genomics research, translational medicine, and personalized medicine. The IBM Reference Architecture for Genomics is:

- Data centric: providing researchers and clinicians with data management capabilities with scalability and ease-of-administration
- Software defined: including software-based abstraction layers for computation, storage and cloud ensuring portability and uniform access across domains
- Application ready: supports plug-and-play for key applications with data management, version control, workload manage-ment, workflow orchestration, and monitoring
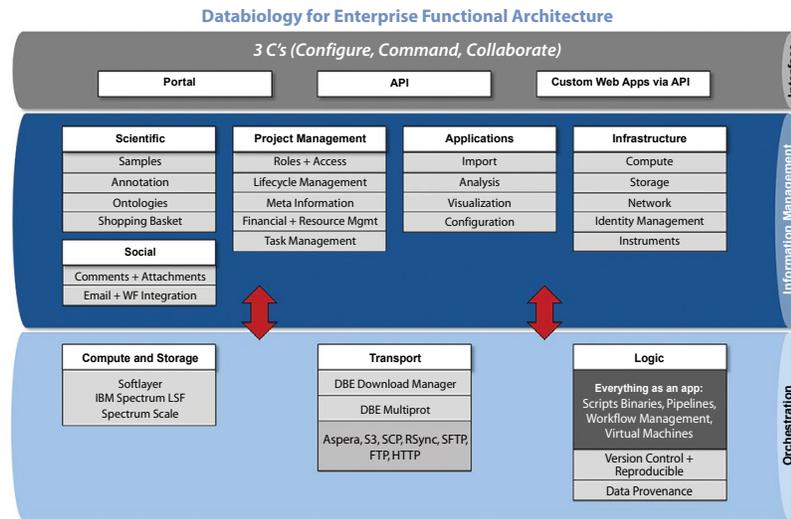
**Databiology for Enterprise Functional Architecture**

**3 C's (Configure, Command, Collaborate)**

| Portal | API | Custom Web Apps via API |

*Interface*

**Scientific**
- Samples
- Annotation
- Ontologies
- Shopping Basket

**Social**
- Comments + Attachments
- Email + WF Integration

**Project Management**
- Roles + Access
- Lifecycle Management
- Meta Information
- Financial + Resource Mgmt
- Task Management

**Applications**
- Import
- Analysis
- Visualization
- Configuration

**Infrastructure**
- Compute
- Storage
- Network
- Identity Management
- Instruments

*Information Management*

**Compute and Storage**
- Softlayer
- IBM Spectrum LSF
- Spectrum Scale

**Transport**
- DBE Download Manager
- DBE Multiprot
- Aspera, S3, SCP, RSync, SFTP, FTP, HTTP

**Logic**
- **Everything as an app:** Scripts Binaries, Pipelines, Workflow Management, Virtual Machines
- Version Control + Reproducible
- Data Provenance

*Orchestration*

*Figure 1.* The Databiology and IBM solution enables better efficiency, scalability, and performance.

With DBE integrated with IBM solutions, researchers and clinicians can be confident that their analytics and data are secure, scalable and computed at benchmark setting pace. DBE and IBM integrated solutions can help companies achieve faster time to market for new products:

- Instantly scale the latest diagnostics pipelines and deploy them anywhere in the world.
- Build customized web fronts end to present users with the latest functionality, the way they like to see it.

The integrated offerings provide features such as functional optimization, so that users can do their jobs. For instance, role based approaches abstract the IT details from scientists and clinicians so they can easily build and run complex I2Is, retrieve information in context and manage projects. Resource management is simplified with a central hub. Through the hub,

control information management for pan-omics data and ontologies, configuration for all versions of all applications and assignment of all infrastructure resources.

In addition, compliance requirements are supported with full data provenance and reproducible science for automated data tracking and built-in ability to rerun any process with its version of data, software and parameters. Additional benefits include automated discovery and resource optimization:

- Foundation for automated discovery in omics—begin tracking relationships between the work of different science teams, the impact of changes in reference data sets and annotations on your existing continuum of research and analysis
- Better view on consumption of resources and optimized utilization—increased automation frees up IT specialists for value adding activities, allows best use of your storage and compute assets, reduces the movement of data and decreases the support burden

## Why Databiology?

Databiology is a privately held global life sciences software and services company based in Oxford and San Francisco, at the forefront of the adoption of omics informatics and information management. Databiology is enabling scientific breakthroughs through sustainable and comprehensive use and re-use of information assets through unique enterprise information management solutions to the global life sciences industry. For more details visit: http://www.databiology.com/

## Why IBM?

IBM Spectrum Computing offers a comprehensive portfolio of software defined infrastructure solutions designed to help your organization deliver IT services in the most efficient way possible, optimizing resource utilization to speed time to results and reduce costs. These offerings help maximize the potential of your infrastructure to accelerate your analytics, high-performance computing (HPC), Apache Hadoop, Spark and cloud-native applications at any scale, extract insight from your data and get higher-quality products to market faster.

Whether deployed in a data center or on the cloud, IBM Spectrum Computing solutions fuel product development, critical business decisions and breakthrough insights in financial services, manufacturing, digital media, oil and gas, life sciences, government, research and education. From designing Formula One race cars to credit risk analysis, organizations in a wide variety of industries are using IBM Spectrum Computing as a foundation for software defined infrastructure solutions for big data, analytics, HPC and cloud to improve business results.

## For more information

To learn more about Databiology and IBM solutions, contact your IBM representative or IBM Business Partner, or visit: **ibm.com**/spectrumcomputing

Additionally, IBM Global Financing provides numerous payment options to help you acquire the technology you need to grow your business. We provide full lifecycle management of IT products and services, from acquisition to disposition. For more information, visit: **ibm.com**/financing

DCS03065-USEN-01