Hitachi Data Systems

HITACHI
Inspire the Next

Checklist for Big Data Infrastructure Readiness

Leaders want the power of big data analytics to fulfill its promise of business insight and competitive advantage. Whether big data is from businesses, people or machines, its volume, velocity and variety make big demands on the IT infrastructure. The right infrastructure can make or break your big data project. With this checklist you can gauge the readiness of your infrastructure for big data.

❑ DISTRIBUTED COMPUTING

Horizontal scalability and massively parallel processing allow multiple concurrent big data analytics. Distributed computing supports data-intensive operations like "map and reduce" for Hadoop clusters.

Big Data Advantage - Computational power, speed and scale to process, manage and manipulate immense amounts of data.

❑ VIRTUALIZED INFRASTRUCTURE

Server, storage and network virtualization with application programming interfaces (APIs) leverage functionality to expertly handle big data at each layer. For example, virtualized infrastructure provides workload agility and mobility in the server layer, data performance and persistence in the storage layer, and connectivity in the network layer. Storage virtualization provisions storage in efficient, consolidated pools with the right mix of performance and capacity for big data on a petabyte scale.

Big Data Advantage - Increased efficiency, flexibility, resource use, and consistent data presentation to analytics applications.

❑ CONVERGED INFRASTRUCTURE

Server, storage, network and software may be pre-integrated with or tuned for specific big data applications like in-memory analytics and optimized for big data workloads, such as Hadoop-based or similar NoSQL distributed frameworks. Consider converged infrastructures that are pre-integrated with virtualization technologies. Also consider pre-integrations with automated management features and interfaces to external big data sources, like the Internet or cloud (RESTful API).

Big Data Advantage - Faster time to big data analysis, reduced complexity and lower infrastructure operating costs.

❑ OBJECT STORE

With tremendous scalability, and the ability to store and manage billions of objects, object stores incorporate system (application-supplied) and custom (user-supplied) metadata. Metadata searches take only 1/10th of the compute cycles that full data searches do.

Big Data Advantage - Massive scalability and ready search for unstructured data, a key requirement for big data analytics.

❑ HIGH AVAILABILITY

Physical device redundancy and caching methods, combined with data replication to a remote site, are proven methods for high availability. Fault-tolerant, distributed file systems help assure uninterrupted service.

Big Data Advantage - Continuous data management and analytics operations.

❑ DATA PROTECTION

Protect data at the physical storage layer with RAID-6 or other device-appropriate erasure-coding algorithms. Further protect data with a point-in-time data snapshots and replication to a remote location. A backup management application can orchestrate these and ensure recovery point and recovery time objectives are met.

Big Data Advantage - Data recoverability from operational failures and disaster.

With a robust virtualization ecosystem, highly resilient architectures, and highly evolved data management, Hitachi Data Systems has the right infrastructure for your big data platforms and applications, including analytics engines like Hadoop. Hitachi Consulting can identify big data business opportunities, and evaluate and architect big data solutions.