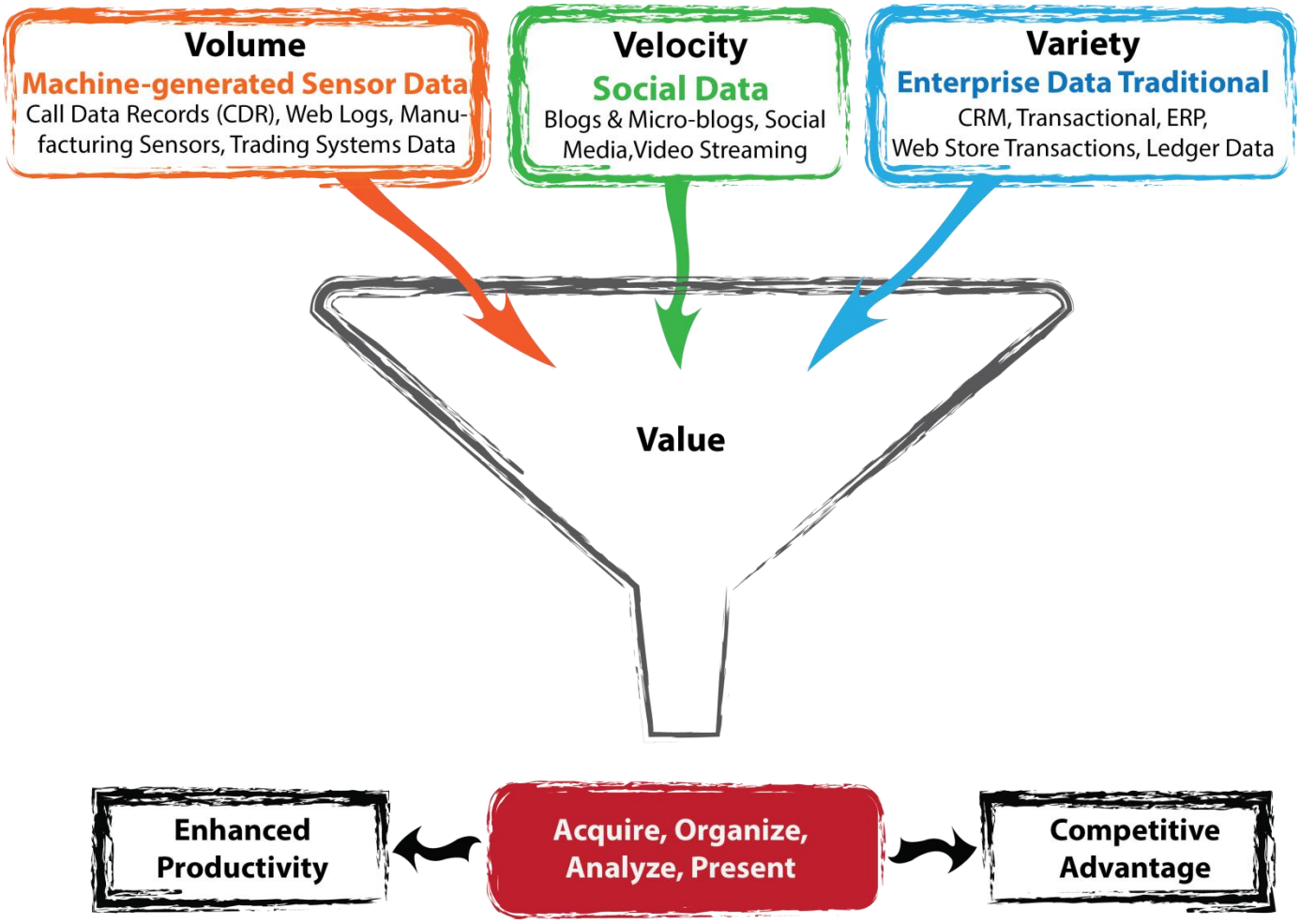
BIG DATA IS MESSY

PARTNER WITH SCALABLE

SCALABLE SYSTEMS

HADOOP SOLUTION

SCALABLE SYSTEMS



WHAT IS BIG DATA ?

Each day human beings create 2.5 quintillion bytes of data. In the last two years alone over 90% of the data on the planet has been created and there is no sign that this will change, in fact, data creation is increasing. The reason for the tremendous explosion in data is that there are so many sources such as; sensors used to collect atmospheric data, data from posts on social media sites, digital and motion picture data, data generated from daily transaction records and cell phone and GPS data, just to name a few. All of this data is called Big Data and it encompasses three dimensions: Volume, Velocity, Variety.

To derive value from Big Data, organizations need to restructure their thinking. With data growing so rapidly and the rise of unstructured data accounting for 90% of the data today, organizations need to look beyond the legacy and exclusive frameworks that place severe limitations on managing Big Data efficiently and productively.

HOW TO MANAGE BIG DATA ?

Businesses across the globe are facing the same cumbersome problem; an ever growing amount of data combined with a limited IT infrastructure to manage it. Big Data is substantially more than just a large volume of data collecting within the organization, it is now the signature of most commercial enterprises and raw unstructured data is the standard fare. Disregarding Big Data is no longer a choice. Organizations that are unable to manage their data will be overpowered by it. Ironically, as organizations access to ever increasing amounts of knowledge has increased dramatically, the rate that an organization can process this gold mine of data has decreased. Extracting derivative value from data is what enables an organization to enhance productivity and competitive advantage. Today the technology exists to efficiently store, manage and analyze virtually unlimited amounts of data and that technology is called Hadoop.

What is Hadoop?

Apache Hadoop is 100% open source, and pioneered a fundamentally new way of storing and processing data. Instead of relying on expensive, proprietary hardware and different systems to store and process data, Hadoop enables distributed parallel processing of huge amounts of data across inexpensive, industry-standard servers that both store and process the data, and can scale without limits. With Hadoop, no data is too big. And in today's hyper-connected world where more and more data is being created every day, Hadoop's breakthrough advantages mean that businesses and organizations can now find value in data that was recently considered useless.

But what exactly is Hadoop, and what makes it so special? In it's basic form, it is a way of storing enormous data sets across distributed clusters of servers and then running "distributed" analysis applications in each cluster. It's designed to be robust, in that the Big Data applications will continue to run even when failures occur in individual servers or clusters. It's also designed to be efficient, because it doesn't require the applications to shuttle huge volumes of data across the network. It has two main parts; a data processing framework (MapReduce) and a distributed file system (HDFS) for data storage. These are the components that are at the heart of Hadoop and really make things happen.

Hadoop MapReduce

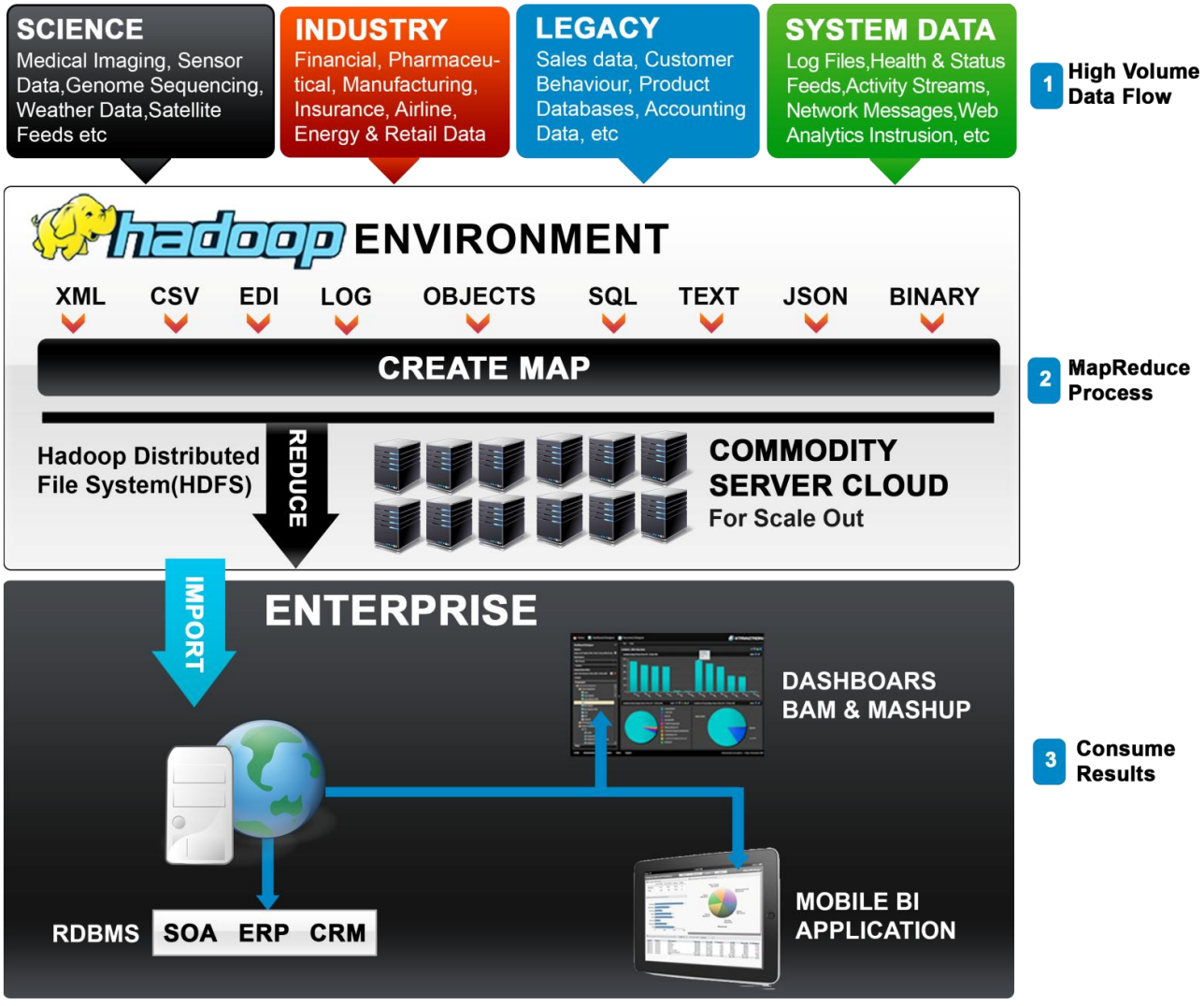
MapReduce is a Java based data processing framework tool that is used to work with the data itself. It's the tool that actually gets data processed. But Hadoop is not really a database. It stores data and data can be pulled out of it, but there are no queries involved, SQL or otherwise. Hadoop is more of a data warehousing system so it needs a system like MapReduce to actually process the data. MapReduce runs as a series of jobs, with each job essentially a separate Java application that goes out into the data and starts pulling out information as needed. Using MapReduce instead of a query gives data seekers a lot of power and flexibility, but also adds a level of complexity.

Hadoop Distributed File System (HDFS)

HDFS is like a big storage container of the Hadoop system, Data is stored in the container until there is a need to do something with it. This could include performing an analysis on it within Hadoop or capturing and exporting a set of data to another tool and performing the analysis there.

HOW HADOOP IS INTEGRATED IN ENTERPRISE ?

The majority of enterprises are typically conservative and pragmatic and are primarily interested in reducing the cost of the environment while squeezing more value out of their data. Getting real answers to help figure out how out how Hadoop fits into the existing environment alongside data warehouses and other legacy systems is important. So low-cost and scalability are key requirements of any unstructured data management platform in order for it to add significant value to the enterprise.



GUIDE FOR SCALABLE HADOOP IMPLEMENTATION:

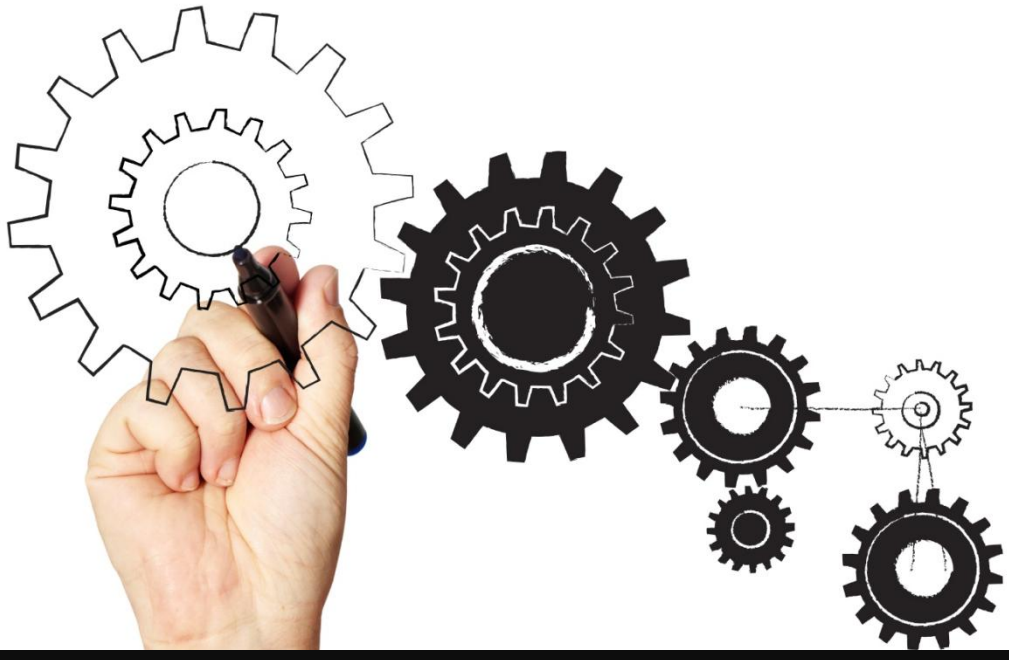
This guideline for Hadoop implementation will help you create a vision for your organization that goes beyond data sitting in databases to building rapidly on deep insights, creating competitive advantage and becoming a truly data driven organization. The guideline utilizes a four step approach that will effectively identify the correct implementation strategy for your organization and unlock the potential the value of your data.

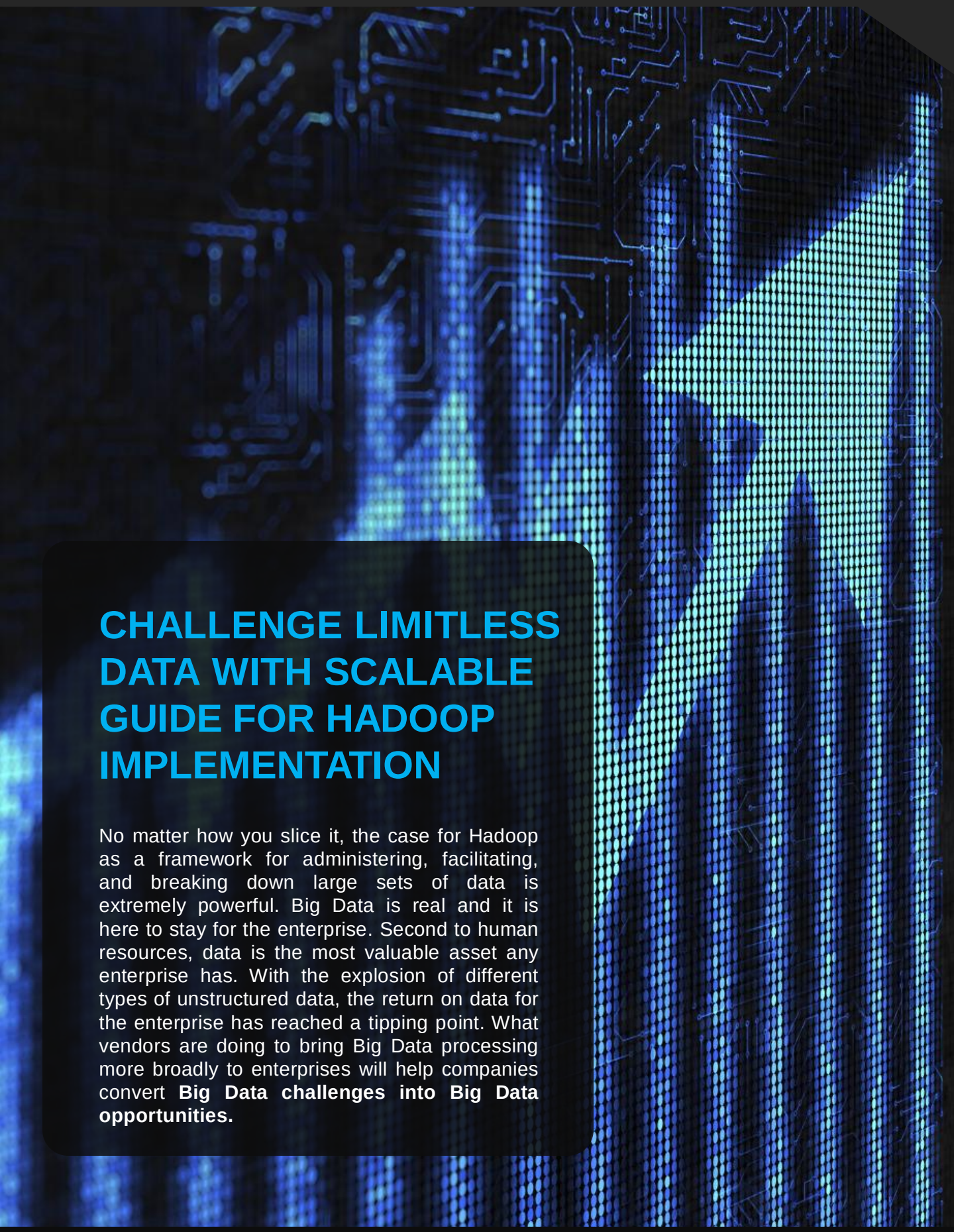
1. Conduct evaluations: Identify potential utilization of Hadoop in the organization by identifying the structure, workload, accessibility, rate of change and location of data that is used by the organization.

2. Leverage best practices: Review current best practices for operating and working with the components that are needed to build Hadoop clusters by identifying potential opportunities, risks and alternatives.

3. Recommend configuration and design suggestions: A successful Hadoop implementation should take into consideration specific organization requirements. Hadoop provides supplemental services that will guaranty a productive Hadoop deployment and integration into the environment. These service include support, training, cluster management and data management services.

4. Perform a gap analysis to identify risks: By performing a gap analysis up front, any potential risks to a successful Hadoop implementation will be identified and accounted for before undertaking the implementation.



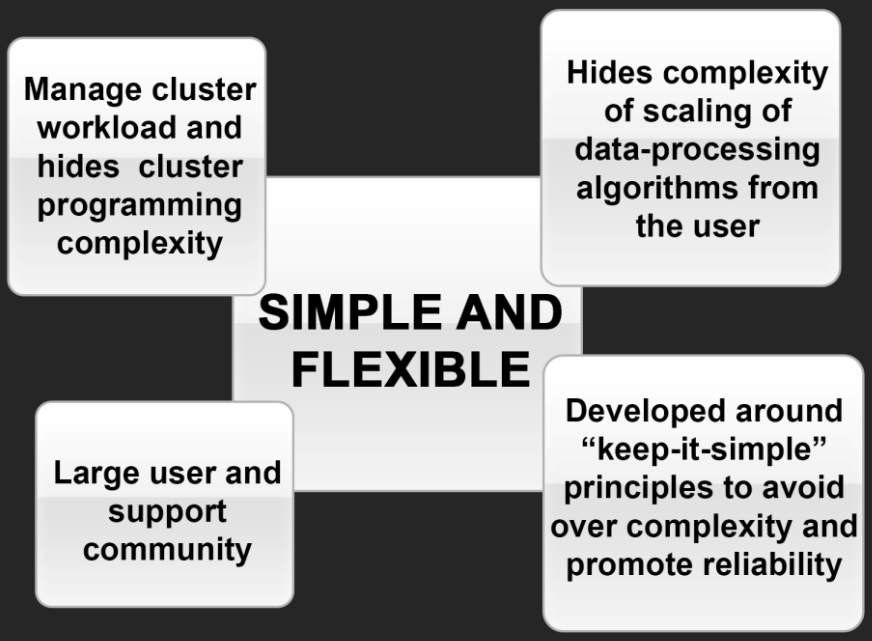


CHALLENGE LIMITLESS DATA WITH SCALABLE GUIDE FOR HADOOP IMPLEMENTATION

No matter how you slice it, the case for Hadoop as a framework for administering, facilitating, and breaking down large sets of data is extremely powerful. Big Data is real and it is here to stay for the enterprise. Second to human resources, data is the most valuable asset any enterprise has. With the explosion of different types of unstructured data, the return on data for the enterprise has reached a tipping point. What vendors are doing to bring Big Data processing more broadly to enterprises will help companies convert **Big Data challenges into Big Data opportunities.**

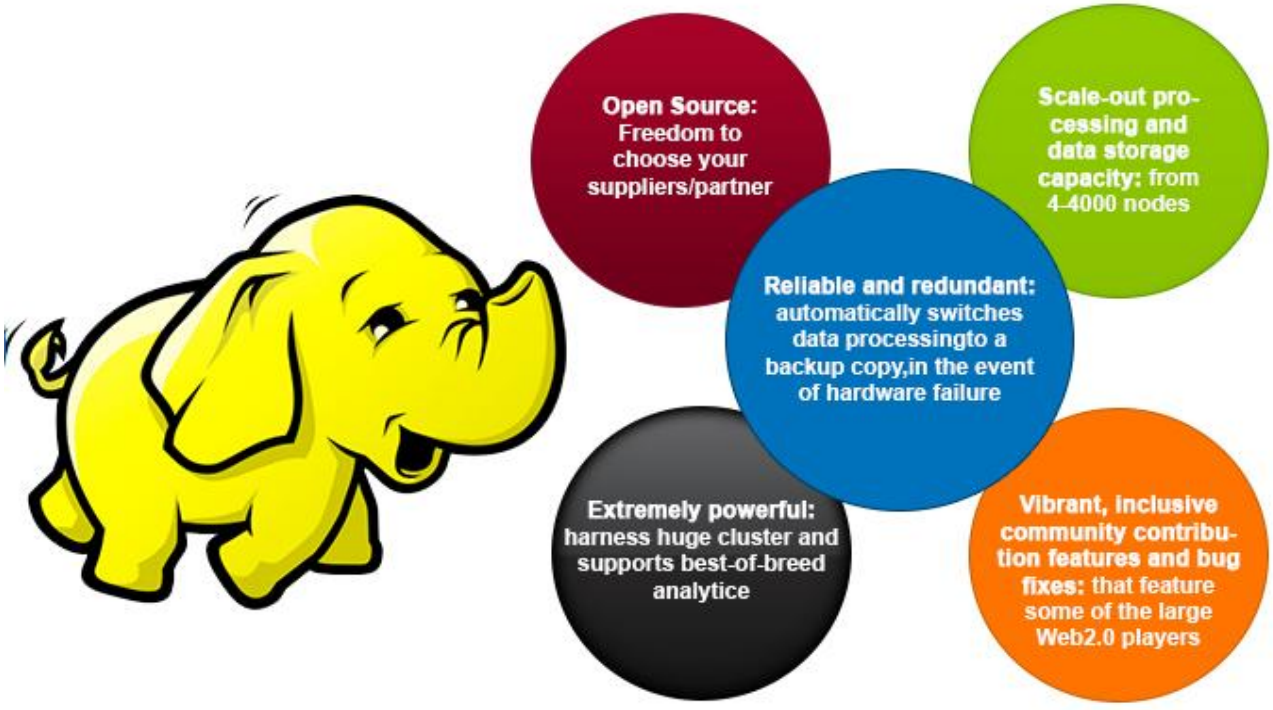
WORK WITH OUR SPECIALIST TO ARRANGE YOUR OWN GUIDE

Scalable Systems Services for Hadoop will help you plan, strategize, and manage Apache Hadoop for large scale data transformation, dissection, processing and analysis. To learn more about the services that are available, visit http://www.scalable-systems/hadoop



visit http://www.scalable-systems/hadoop

KEY FEATURES OF HADOOP



HOW SCALABLE SYSTEMS CAN HELP YOU?

In the world of big data, 500 terabytes is merely Tiny Data, but a 1% error in 500 terabytes is 5 million megabytes. Big data platforms must operate and process data at a scale that leaves little room for error.

Big data clusters must be built for speed, scale, and efficiency.

Scalable Systems helps organizations to understand this new approach to data management, where to start, and how to measure the short and long term benefits. If your organization is considering HADOOP, partner with Scalable Systems for the following services:

- Big Data Vision, Business Strategy & Implementation
- Big Data Roadmap Creation
- Big Data Business Case Development and Proof of Concept
- Design, Development, implementation, support and maintenance of Hadoop platform
- Planning, Installation, configuration, maintenance, monitoring, Backup / recovery, Performance tuning of Hadoop Cluster
- Integrating existing data integration and business intelligence platform with Hadoop
- Applying Map Reduce patterns to Big data, streamlining HDFS for big data of Hadoop environment, Integrating R and Hadoop for statistics
- Implementing Predictive Analytics using Mahout
- Development and implementation of related Hadoop technologies such as Flume, Oozie, Sqoop, Hbase, Avro, Snappy, MySQL, Hive, Ping, Crunch, R, RHIPE, RHadoop

ABOUT SCALABLE SYSTEMS

Scalable Systems is a global Information technology, consulting and outsourcing company. We help our clients to unleash their potential by using tomorrow's technology today. Besides saving cost by modernizing existing business processes our clients partner with us on 'What's Next' leaving the competition behind.

We take a holistic approach to solve industry challenges by focusing both on Technology and vertical. Our focus on technology innovation and collaboration with leading technology companies of the world helps us to be ahead of the curve. Our focus on verticals helps us to understand industry specific challenges. Our value lies in our ability linking the best technology solution to a complex business challenges in the most cost-effective and innovative way.

Headquartered in New Jersey we are a Minority Certified company by National Minority Supplier Development Council and State of New Jersey with operation in USA, Europe and Asia.

Scalable Systems
Email: info@scalable-systems.com
Web: www.scalable-systems.com

This edition published June 2013