

# Microsoft Big Data

Solution Brief

**Contents**

Introduction ..... 2

The Microsoft Big Data Solution ..... 3

Key Benefits ..... 3

    Immersive Insight, Wherever You Are ..... 3

    Connecting with the World's Data ..... 3

    Any Data, Any Size, Anywhere ..... 4

Additional Information ..... 4

## Key Customer Challenges

- **Making sense of the explosion of data:** Organizations need the right tools to make sense of the overwhelming amount of data generated by declining hardware costs and complex data sources.
- **Understanding a growing variety of data:** Organizations need to analyze both relational and non-relational data. Over 85 percent of data captured is unstructured.
- **Enabling real-time analysis of data:** New data sources—such as social media sites like Twitter, Facebook, and LinkedIn—are producing unprecedented volumes of data in real time, which cannot be analyzed effectively with simple batch processing.
- **Achieving simplified deployment and management:** Organizations need a streamlined deployment and setup experience that simplifies the complexity of Apache Hadoop. Ideally they would prefer to have fewer installation files that package the required Hadoop-related projects instead of making the choice themselves.

## Introduction

Today, organizations are struggling to gain business insight from the unprecedented volume of data they are capturing. This includes vast amounts of unstructured data such as files, images, videos, blogs, clickstreams, and geo-spatial data. For organizations, the main challenge is to learn how to effectively process both structured and unstructured data without the burden of setting up complex distributed storage and compute clusters.

Organizations are looking for an effective way to combine internal and external data and services. They want to mine data from social media sites like Twitter, Facebook, and LinkedIn. They also want to make more timely decisions based on the data they capture. To achieve this, organizations need to analyze their data in real time instead of simply relying on batch processing.

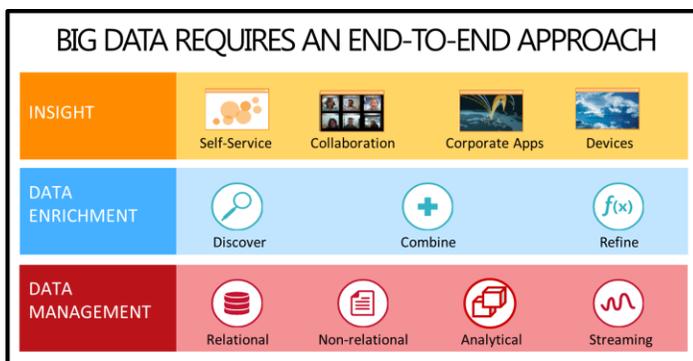
New technologies, such as Hadoop, have emerged to offer customers the opportunity to store and analyze petabytes of unstructured data inexpensively. In addition, organizations can connect to data from hundreds of trusted data providers—including demographic data, environment data, financial data, retail and sports data, and social media data—combining it with their personal data through self-service tools like Microsoft PowerPivot. Today, various vendors provide Hadoop deployments, but most of them operate in a silo outside the scope of central IT and are not yet enterprise-ready.

Microsoft has been doing Big Data long before it was popular. For example, Microsoft Bing analyzes over 100 petabytes of data to deliver high-quality search results. Organizations can use the Microsoft Big Data solution to unleash actionable insights from a broad and diverse range of data through familiar tools like Microsoft Office and Microsoft SharePoint. It combines the simplicity of Windows with the power and reliability of the Hortonworks Data Platform (HDP) to deliver new insights from all their data. It also enables customers to uncover new value by connecting to the world's data and services.

## The Microsoft Big Data Solution

Microsoft's vision is to enable all users to gain actionable insights from virtually any data, including insights previously hidden in unstructured data. To achieve this, Microsoft has a comprehensive Big Data strategy that offers:

- **A modern data management layer** that supports all data types—structured, semi-structured, and unstructured data at rest or in motion.
- **An enrichment layer** that enhances your data through discovery, combining with the world's data and by refining with advanced analytics.
- **An Insights layer** that provides insights to all users through familiar tools like Office.



To help accelerate the adoption of its Big Data solution in the enterprise, Microsoft will offer Hadoop both as a cloud-based service on the Microsoft Windows Azure platform and as an on-premises distribution on Microsoft Windows Server.

**HDInsight** is Microsoft's new Hadoop-based service, built on the Hortonworks Data Platform (HDP) that offers 100% compatibility with Apache Hadoop. Windows Azure HDInsight Service runs in the cloud, while Microsoft HDInsight Server runs on Windows Server. HDInsight will enable customers to gain business insights from structured and unstructured data of virtually any size and activate new types of data irrespective of its location. Rich insights from Hadoop can be combined seamlessly with the Microsoft Business Intelligence (BI) platform to give customers the ability to enrich their models with publicly available data and services using familiar tools like Office and SharePoint.

Microsoft's Big Data solutions also provide the simplicity and manageability of Microsoft Windows for Hadoop, through easy deployment and integration with

System Center. Through Windows Azure HDInsight Service, Microsoft offers elasticity with its Big Data solutions in the cloud.

## Key Benefits

- **Immersive insight, wherever you are**, from any data with familiar Office and BI tools.
- **Connecting with the world's data** to unlock hidden patterns through a combination of internal and publicly available data and services, including social media sites.
- **Any data, any size, anywhere** through a modern data management platform that supports any data, with the simplicity of Windows, and the elastic scalability of the cloud.

## Immersive Insight, Wherever You Are

Microsoft's Big Data solution unlocks insights from all types of data using familiar Microsoft Office and BI tools. Specifically, Microsoft's solution will enable customers to:

- **Analyze Hadoop data with familiar tools:** Microsoft enables analysts and business users to interact with and gain valuable insight from Hadoop functions from the very familiar Microsoft Excel interface with the Hive add-in for Excel
- **Get immersive insights from any data:** Organizations can use familiar BI tools like Microsoft SQL Server Analysis Services (SSAS), PowerPivot, and Power View through the Hive Open Database Connectivity (ODBC) Driver to analyze unstructured data in Hadoop. Organizations can also enable self-service BI on relational data using PowerPivot and Power View in SQL Server 2012.
- **Drive insights through simplified programming:** Microsoft simplifies programming on Hadoop through integration with .NET and new JavaScript libraries. Developers can use the new JavaScript libraries to easily write MapReduce programs in JavaScript, and then deploy their JavaScript code from a simple browser.

## Connecting with the World's Data

Microsoft's Big Data solution can enable breakthrough discoveries by combining data and models with publicly available data and services, including social media sites like Twitter, Facebook, and LinkedIn. This enables

customers to uncover hidden patterns, using the applications and mining algorithms on Windows Azure Marketplace.

- **Discover the right data:** Microsoft Big Data solution offers unique tools to facilitate discovery of data both within and outside an organization. An Azure Lab, codenamed "Data Explorer," enables customers to discover relevant datasets through automatic recommendations. Another lab codenamed "Data Hub" enables an organization to create a private data market to facilitate discovery and sharing of data and analytical models. The Azure Marketplace DataMarket enables discovery and sharing outside the firewall and with 3rd party data sources.
- **Combine with the world's data:** The Azure Marketplace empowers customers to connect to data, smart mining algorithms and people outside their firewalls. Windows Azure Marketplace offers hundreds of datasets from trusted third party providers.
- **Refine with external data:** Microsoft Big Data solution enables customers to convert their raw data into credible consistent data with enterprise information management tools. It also enables enrichment through advanced analytics: Microsoft provides out-of-the-box data mining algorithms with SQL Server Analysis Services. Microsoft Big Data solution also supports commonly used 3rd party tools and frameworks such as Mahout. Finally, through Hadoop streaming, it supports bespoke mining algorithms written in C++, C#, Python, Ruby, and Pearl.

## Any Data, Any Size, Anywhere

Microsoft enables customers to seamlessly store and process data of all types, including structured, unstructured, and real-time data, through a modern data management platform. It provides the simplicity of Windows on Hadoop, extends data warehouses with Hadoop, and offers the elastic scalability of the cloud to Big Data.

- **Enterprise-ready Hadoop with HDInsight:** Microsoft's HDInsight is an Enterprise-ready Hadoop service based on the HDP which offers the most reliable, innovative and trusted distribution available for Windows Server and Windows Azure. Integration with Active Directory enables IT to

secure their Hadoop clusters using enterprise-based security policies. Integration with Microsoft System Center allows IT to easily manage their Hadoop clusters and effectively meet SLAs.

- **Bringing the simplicity and manageability of Windows to Hadoop:** Smart packaging from Microsoft enables simple and straightforward installation of your Hadoop clusters. Accelerate the deployment with the cloud by deploying a Hadoop cluster on Windows Azure in just 10 minutes. The integration with Apache Ambari in HDP and System Center, simplifies the provisioning, monitoring and management of Hadoop clusters.
- **Seamlessly extend your data warehouse with HDInsight:** Hadoop connectors for SQL Server and Parallel Data Warehouse appliance enable easy integration of Hadoop with Microsoft Enterprise Data Warehouses and BI solutions. In addition, you can also integrate Hadoop with your relational Data Warehouses using HCatalog.
- **Seamless scale and elasticity of the cloud:** Microsoft offers two options for deploying Hadoop on Windows Server—in a cloud-based environment or on-premises. Windows Azure HDInsight Service is a cloud-based service that offers elastic peta-scale analytics on Microsoft's cloud platform. It also offers seamless migration to Microsoft HDInsight Server that runs on premise.
- **Open Big Data platform:** Through HDP, HDInsight is 100% compatible with Apache Hadoop. Microsoft has already submitted proposals to Apache, including new JavaScript libraries for Hadoop (developed by Microsoft), as well as the Hive ODBC Driver.

## Additional Information

For more information on the Microsoft Big Data solution, go to [www.microsoft.com/bigdata](http://www.microsoft.com/bigdata).

Published October 19, 2012