

# Data Lake

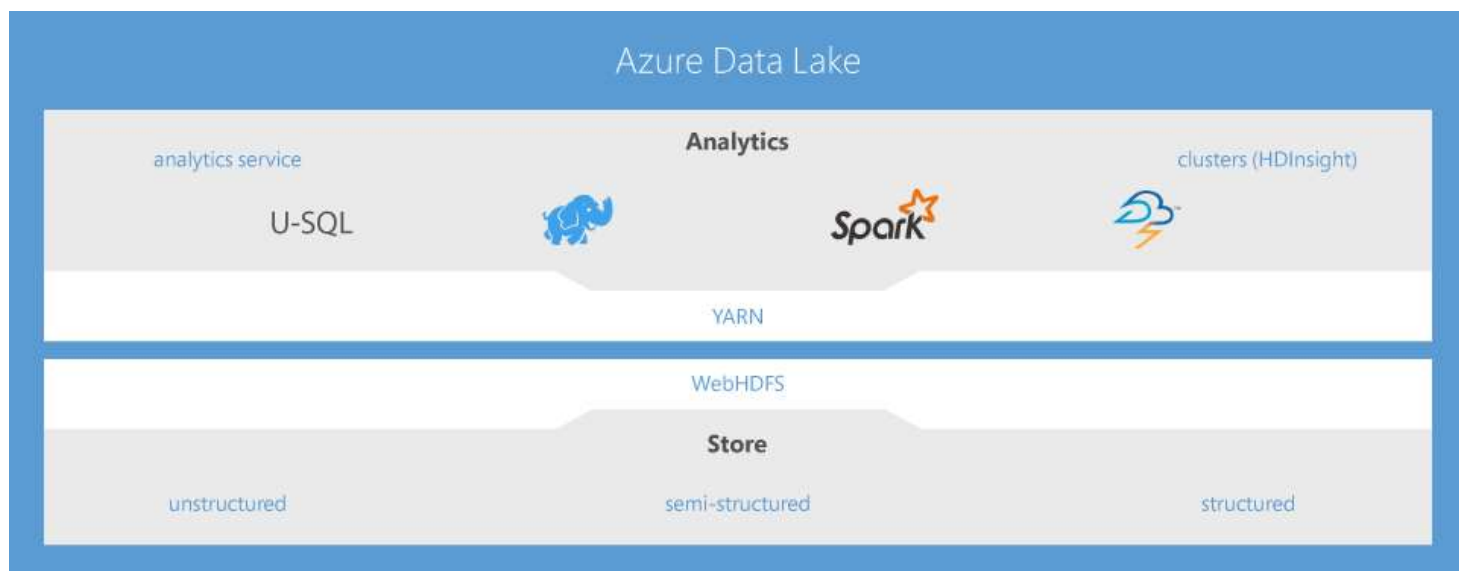
A no-limits data lake to power intelligent action

- ✓ Store and analyze petabyte-size files and trillions of objects
- ✓ Develop massively parallel programs with simplicity
- ✓ Debug and optimize your big data programs with ease
- ✓ Enterprise-grade security, auditing, and support
- ✓ Start in seconds, scale instantly, pay per job
- ✓ Built on YARN, designed for the cloud

Try it now [\(en-us/free/\)](#)

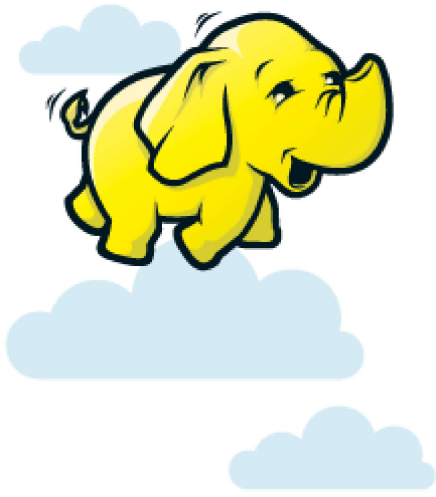
Azure Data Lake includes all the capabilities required to make it easy for developers, data scientists, and analysts to store data of any size, shape, and speed, and do all types of processing and analytics across platforms and languages. It removes the complexities of ingesting and storing all of your data while making it faster to get up and running with batch, streaming, and interactive analytics. Azure Data Lake works with existing IT investments for identity, management, and security for simplified data management and governance. It also integrates seamlessly with operational stores and data warehouses so you can extend current data applications. We've drawn on the experience of working with enterprise customers and running some of the largest scale processing and analytics in the world.

for Microsoft businesses like Office 365, Xbox Live, Azure, Windows, Bing, and Skype. Azure Data Lake solves many of the productivity and scalability challenges that prevent you from maximizing the value of your data assets with a service that's ready to meet your current and future business needs.



## Data Lake Analytics—a no-limits analytics job service to power intelligent action

The first cloud analytics service where you can easily develop and run massively parallel data transformation and processing programs in U-SQL, R, Python, and .Net over petabytes of data. With no infrastructure to manage, process data on demand, scale instantly, and only pay per job. [Learn more \(/en-us/products/data-lake-analytics/\)](https://en-us/products/data-lake-analytics/).



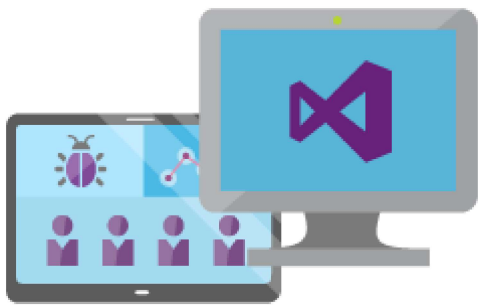
## HDInsight—cloud Apache Spark and Hadoop® service for the enterprise

HDInsight is the only fully managed Cloud Hadoop offering that provides optimized open source analytic clusters for Spark, Hive, Map Reduce, HBase, Storm, Kafka, and R-Server backed by a 99.9% SLA. Each of these Big Data technologies as well as ISV applications are easily deployable as managed clusters, with enterprise level security and monitoring. [Learn more \(/en-us/products/hdinsight/\)](https://en-us/products/hdinsight/).



# Data Lake Store—a no-limits data lake that powers big data analytics

The first cloud data lake for enterprises that is secure, massively scalable and built to the open HDFS standard. With no limits to the size of data and the ability to run massively parallel analytics, you can now unlock value from all your unstructured, semi-structured and structured data. [Learn more \(/en-us/services/data-lake-store/\)](#).



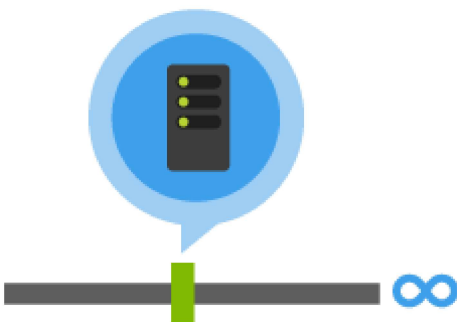
## Develop, debug, and optimize big data programs with ease

Finding the right tools to design and tune your big data queries can be difficult. Data Lake makes it easy through deep integration with Visual Studio, Eclipse, and IntelliJ, so that you can use familiar tools to run, debug, and tune your code. Visualizations of your U-SQL, Apache Spark, Apache Hive, and Apache Storm jobs let you see how your code runs at scale and identify performance bottlenecks and cost optimizations, making it easier to tune your queries. Our execution environment actively analyzes your programs as they run and offers recommendations to improve performance and reduce cost. Data engineers, DBAs, and data architects can use existing skills, like SQL, Apache Hadoop, Apache Spark, R, Python, Java, and .NET, to become productive on day one.



## Integrates seamlessly with your existing IT investments

One of the top challenges of big data is integration with existing IT investments. Data Lake is a key part of Cortana Intelligence (<https://www.microsoft.com/en-us/cloud-platform/cortana-intelligence-suite>), meaning that it works with Azure Synapse Analytics, Power BI, and Data Factory for a complete cloud big data and advanced analytics platform that helps you with everything from data preparation to doing interactive analytics on large-scale datasets. Data Lake Analytics gives you power to act on all your data with optimized data virtualization of your relational sources such as Azure SQL Server on virtual machines, Azure SQL Database, and Azure Synapse Analytics. Queries are automatically optimized by moving processing close to the source data, without data movement, thereby maximizing performance and minimizing latency. Finally, because Data Lake is in Azure, you can connect to any data generated by applications or ingested by devices in Internet of Things (IoT) scenarios.



## Store and analyze petabyte-size files and trillions of objects

Data Lake was architected from the ground up for cloud scale and performance. With Azure Data Lake Store your organization can analyze all of its data in a single place with no artificial constraints. Your Data Lake Store can store trillions of files where a single file can be greater than a petabyte in size which is 200x larger than other cloud stores. This means that you don't have to rewrite code as you increase or decrease the size of the data stored or the amount of compute being spun up. This lets you focus on your business logic only and not on how you process and store large datasets. Data Lake also takes away the complexities normally associated with big data in the cloud, ensuring that it can meet your current and future business needs.



## Affordable and cost effective

Data Lake is a cost-effective solution to run big data workloads. You can choose between on-demand clusters or a pay-per-job model when data is processed. In both cases no hardware, licenses, or service specific support agreements are required. The system scales up or down with your business needs, meaning that you never pay for more than you need. It also lets you independently scale storage and compute, enabling more economic flexibility than traditional big data solutions. Finally, it minimizes the need to hire specialized operations teams typically associated with running a big data infrastructure. Data Lake minimizes your costs while maximizing the return on your data investment. A recent study showed HDInsight delivering 63% lower TCO than deploying Hadoop on premises over five years.



## Enterprise grade security, auditing, and support

Data Lake is fully managed and supported by Microsoft, backed by an enterprise-grade SLA and support. With 24/7 customer support, you can contact us to address any challenges that you face with your entire big data solution. Our team monitors your deployment so that you don't have to, guaranteeing that it will run continuously. Data Lake protects your data assets and extends your on-premises security and governance controls to the cloud easily. Data is always encrypted; in motion using SSL, and at rest using service or user-managed HSM-backed keys in Azure Key Vault. Capabilities such as single sign-on (SSO), multi-factor authentication, and seamless management of millions of identities is built-in through Azure Active Directory. You can authorize users and groups with fine-grained POSIX-based ACLs for all data in the Store enabling role-based access controls. Finally, you can meet security and regulatory compliance needs by auditing every access or configuration change to the system.

**Build Data Lake solutions using these powerful solutions**



## HDInsight

(/en-us/products/hdinsight/)

Provision cloud Hadoop, Spark, R Server, HBase, and Storm clusters



## Data Lake Analytics

(/en-us/products/data-lake-analytics/)

Distributed analytics service that makes big data easy



## Azure Data Lake Storage

(/en-us/products/storage/data-lake-storage/)

Scalable, secure data lake for high-performance analytics

Apache Hadoop (<https://hadoop.apache.org/>).<sup>®</sup> and associated open source project names are trademarks of the Apache Software Foundation (<https://apache.org/>).