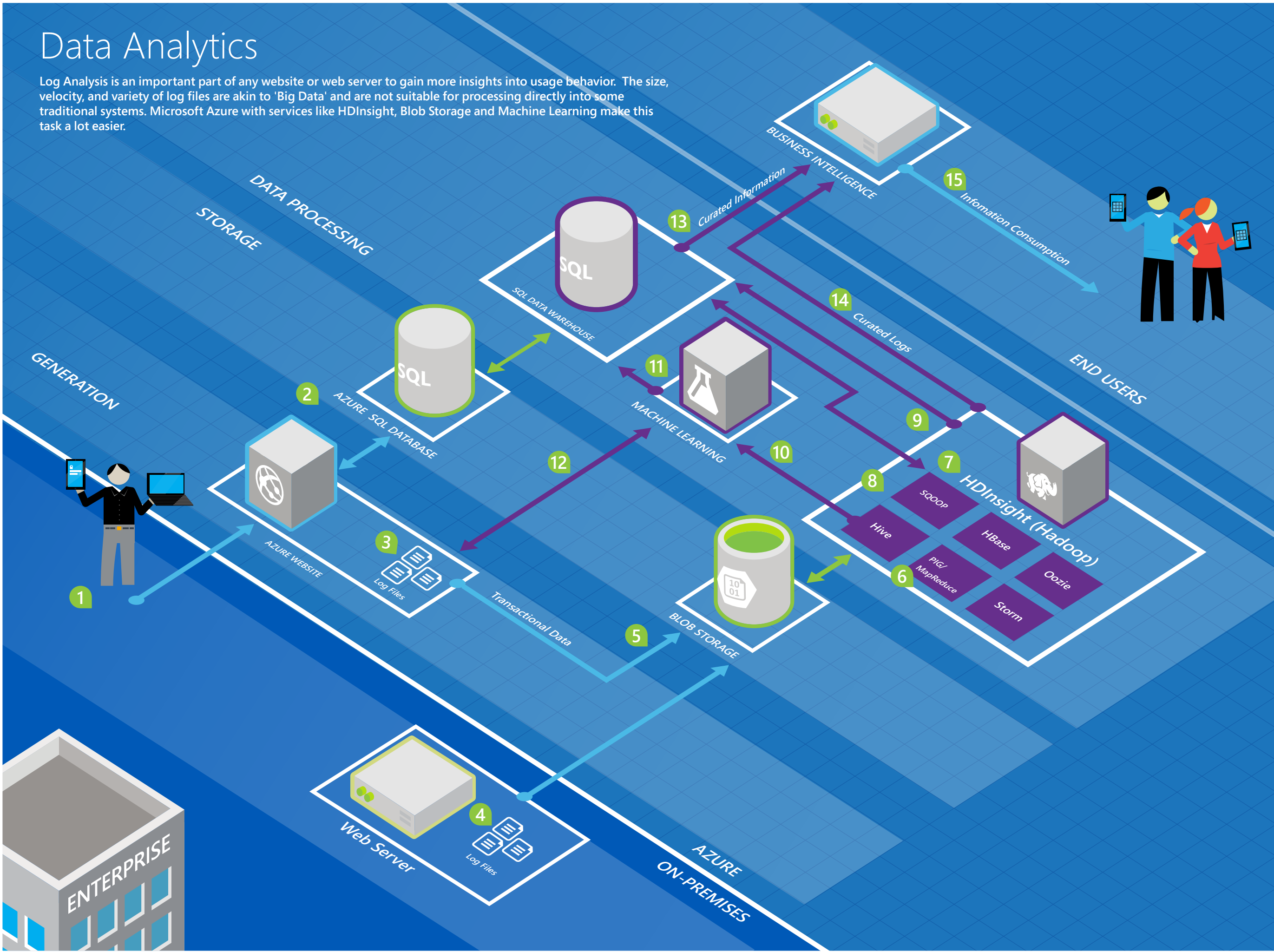


# Data Analytics

Log Analysis is an important part of any website or web server to gain more insights into usage behavior. The size, velocity, and variety of log files are akin to 'Big Data' and are not suitable for processing directly into some traditional systems. Microsoft Azure with services like HDInsight, Blob Storage and Machine Learning make this task a lot easier.



- 1 A user accesses an Azure website, leaving behind a trail of data to mine and information to extract.
- 2 The Azure website has underlying structured data hosted on Azure SQL Database.
- 3 The Azure website generates clickstream data and usage logs.
- 4 The users' actions on cloud or on-premises servers generate logs (ex. server, security, or audit).
- 5 This unstructured / semi-structured data is stored in Blob storage for access from various systems.
- 6 HDInsight processes unstructured/semi-structured data using Hadoop tools like Hive, Pig, and Mahout. It can also orchestrate the data movement using Sqoop and Oozie. For real-time capabilities, you can use Storm and HBase.
- 7 HDInsight supports PowerShell to automate tasks like creating and deleting clusters, running MapReduce programs, running Hive commands, and more.
- 8 SQOOP interacts with structured data stores like Azure SQL Database to import or export data in and out of the HDInsight (Hadoop) clusters.
- 9 Data processed in HDInsight is fed into existing data warehouses via common ETL channels to enrich the information contained in them.
- 10 Azure Machine Learning (ML) accesses processed data in Hive Tables to discover and predict future trends, and/or assist with security/performance risks.
- 11 ML feeds prediction information into the existing data warehouse.
- 12 ML also feeds prediction information to the Azure Website directly for immediate customer centric action
- 13 The Data Warehouse sources data into BI systems like Power BI, SQL Server Analysis Services, and/or custom apps to create data models.
- 14 HDInsight sends processed and curated log information to the BI system.
- 15 The BI systems publish information models into Dashboards and reports using tools like Power View, Power Map, SQL Server Reporting Services, and SharePoint BI for enduser consumption.