Solution Prerequisites

Azure PowerShell (SDK)

<https://www.microsoft.com/web/handlers/webpi.ashx/getinstaller/WindowsAzurePowershellGet.3f.3f.3fnew.appids>

Azure Automation Account with RunAs Account

<https://docs.microsoft.com/en-us/azure/automation/automation-sec-configure-azure-runas-account>

AzureRM PowerShell Modules for Azure Automation (latest is preferred)

<https://www.powershellgallery.com/packages/AzureRM/3.4.0>

Azure Data Lake (HDInsight) Tools for Visual Studio

<https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-visual-studio-tools-get-started>

Most recent SQL Server Data Tools

<https://msdn.microsoft.com/en-us/library/hh500335(v=vs.103).aspx>

Most recent SQL Server Management studio

<https://docs.microsoft.com/en-us/sql/ssms/download-sql-server-management-studio-ssms>

ODBC Driver 13.1 for SQL Server

<https://www.microsoft.com/en-us/download/details.aspx?id=53339>

Command Line Utilities 13.1 for SQL Server

<https://www.microsoft.com/en-us/download/details.aspx?id=53591>

Windows PowerShell

* Command-line shell designed for system administrators
* Built on top of .Net Framework
* Accepts and returns .Net objects
* Tasks performed via cmdlets
* PowerShell ISE

Azure PowerShell

* Set of cmdlets for working with Azure
* Installed on its own or via Azure SDK
* Component teams responsible for cmdlets for their component/service in Modules
* AzureRM module example

Azure Automation

* Allows you to automate tasks that are:
  + Repeated
  + Long-running
  + Error-prone
* Encapsulates logic in Runbooks
  + PowerShell
  + PowerShell Workflow
  + Graphical
* Uses Assets
  + Variables
  + Credentials
  + Schedules
  + Modules
  + Certificates
  + Connections
* Installing Module AzureRM

Azure Storage – BLOB

* Cloud-based storage for object data
  + Text files
  + Binary files
  + Documents
  + Media files
  + Logs
* Objects reside in Storage Containers
* Scalable, highly available
  + Various levels of redundancy
  + LRS = Locally Redundant Storage
    - 3 copies of storage exist within data center
* Security via Shared Access Signature
  + private access keys

Apache Hadoop

* Apache Hadoop
  + Distributed computing storage, analysis
* Control Nodes
  + Orchestrate assigning work to Worker nodes
* Worker Nodes
  + Do the work and return results to Control Nodes
* Map Reduce
  + Legacy software framework for Hadoop
  + Allows batch processing of data sets in parallel

Hive

* Data warehouse built on Hadoop
* Uses SQL-like language called HiveQL
  + Abstraction on top of Map Reduce
* Create tables over the non-tabular data in the cluster

Hadoop on HDInsight

* Hortonworks Data Platform on Azure
* Choose the number and size of nodes, Microsoft configures it for you
* Rather than HDFS on nodes, uses Blob Storage or Data Lake Storage

Azure SQL Database

Azure SQL Data Warehouse

Polybase

Azure Analysis Services

Power BI