Hadoop in Windows Server using PolyBase

Requirements

Required Software: (Hadoop/ Hadoop bin files/Hadoop Configuration files/Java)

https://drive.google.com/open?id=1e6pYQxZrr1JQaAR5ywOWNxkEYjQKKpBL

SQL Server 2017 Developer Edition https://www.microsoft.com/en-us/sql-server/sql-server-downloads

SQL Server Edition **Standard, Web, Express with Advanced Services, Express** does not support head node they require head node

Enterprise Edition only can be used for head node and other can be Scale out with multiple compute nodes.

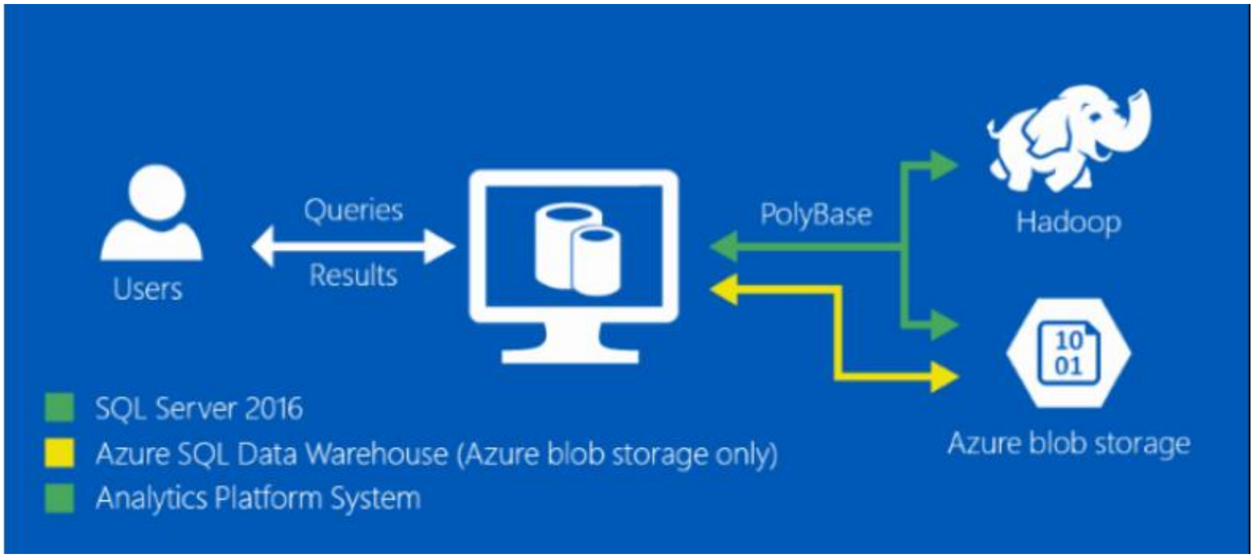
Before we start Some of the difference between Hadoop and SQL

HADOOP	SQL
Schema on Read Approach	**Schema on Write Approach**
When we write the data on Hadoop, has distributed file system so brings all the files without any rules. Then when we dictate the data we apply rules to the code that reads the data rather of preconfiguring the data ahead of time.	In Schema on Write while coping the data from database A to B the data is first checked before inserting to the database (checked the datatypes of the data and the data size) if not meet the requirement reject the data
Data Storage Process Hadoop	**Data Storage Process in SQL**
Data is stored in Compressed file of either text or others datatype. In the moment files or data stored are replicated in multiple nodes in Hadoop distributed filing system	Data is stored in a Logical form in Related Table and Columns

Installation of Java

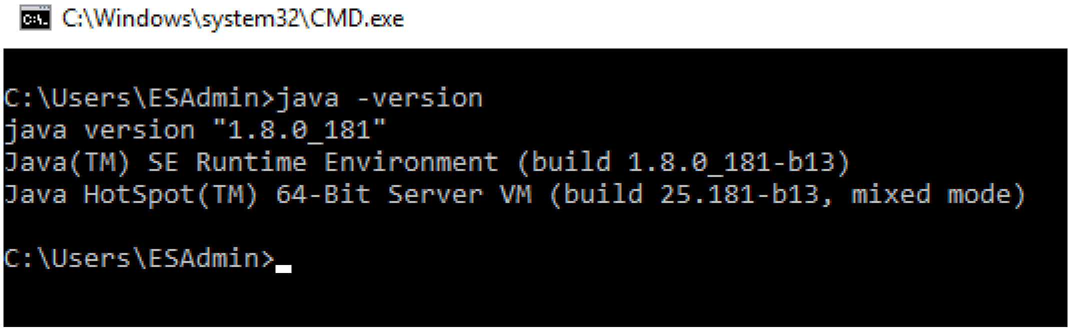
Before installing Hadoop File Distributed System (HDFS) and **PolyBase** feature in SQL Server in 2016 or later requires Java. We need to first install Java in our system.

PolyBase enables your SQL Server 2016 and later instance to process Transact-SQL queries that read data from Hadoop. The same query can also access relational tables in your SQL Server. **PolyBase** enables the same query to also join the data from Hadoop and SQL Server. **PolyBase** is a mediator between SQL Server and Hadoop.


Users
Queries
Results
PolyBase
Hadoop
10
01
Azure blob storage
🟩 SQL Server 2016
🟨 Azure SQL Data Warehouse (Azure blob storage only)
🟩 Analytics Platform System

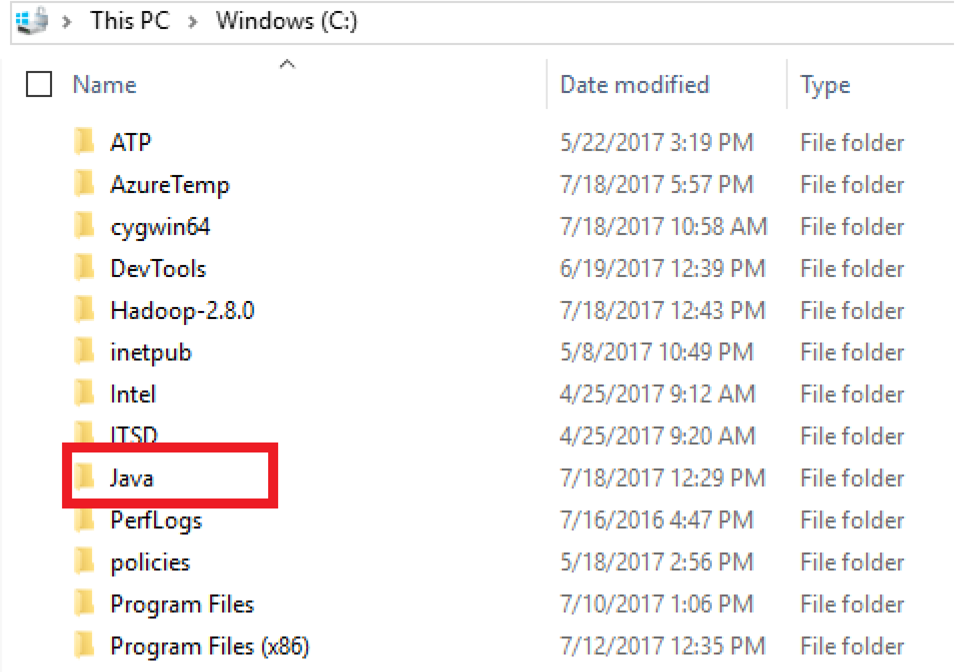
Set up

1. Check either Java 1.8.0 is already installed on your system or not, use **"Java -version"** to check.



```
C:\Users\ESAdmin>java -version
java version "1.8.0_181"
Java(TM) SE Runtime Environment (build 1.8.0_181-b13)
Java HotSpot(TM) 64-Bit Server VM (build 25.181-b13, mixed mode)

C:\Users\ESAdmin>_
```

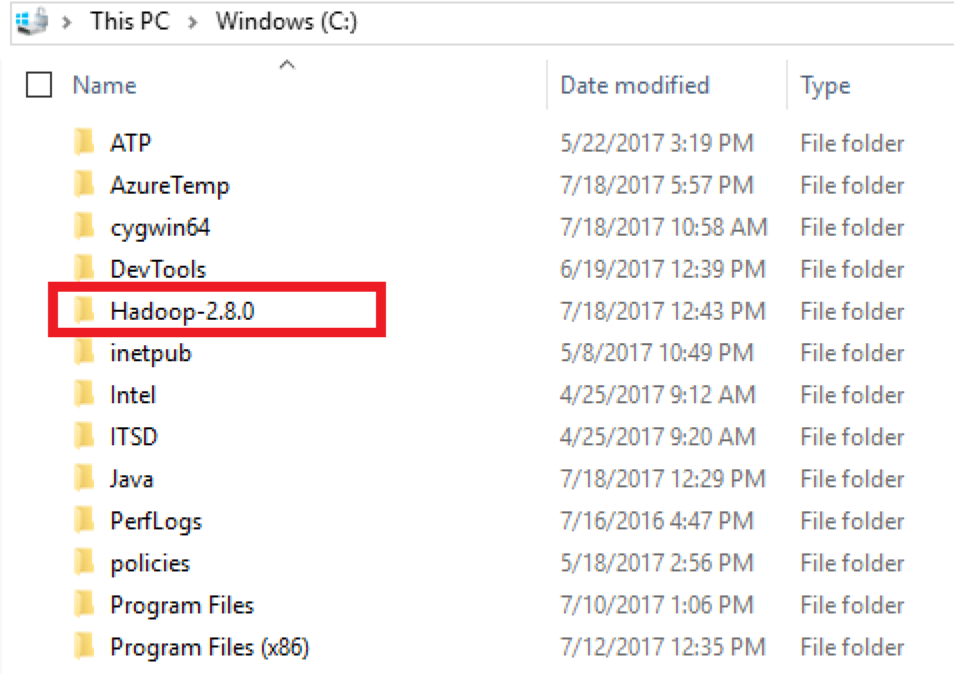
2. If Java is not installed on your system then first install java under **"C:\JAVA"**



This PC › Windows (C:)

□ Name	Date modified	Type
ATP	5/22/2017 3:19 PM	File folder
AzureTemp	7/18/2017 5:57 PM	File folder
cygwin64	7/18/2017 10:58 AM	File folder
DevTools	6/19/2017 12:39 PM	File folder
Hadoop-2.8.0	7/18/2017 12:43 PM	File folder
inetpub	5/8/2017 10:49 PM	File folder
Intel	4/25/2017 9:12 AM	File folder
ITSD	4/25/2017 9:20 AM	File folder
Java	7/18/2017 12:29 PM	File folder
PerfLogs	7/16/2016 4:47 PM	File folder
policies	5/18/2017 2:56 PM	File folder
Program Files	7/10/2017 1:06 PM	File folder
Program Files (x86)	7/12/2017 12:35 PM	File folder

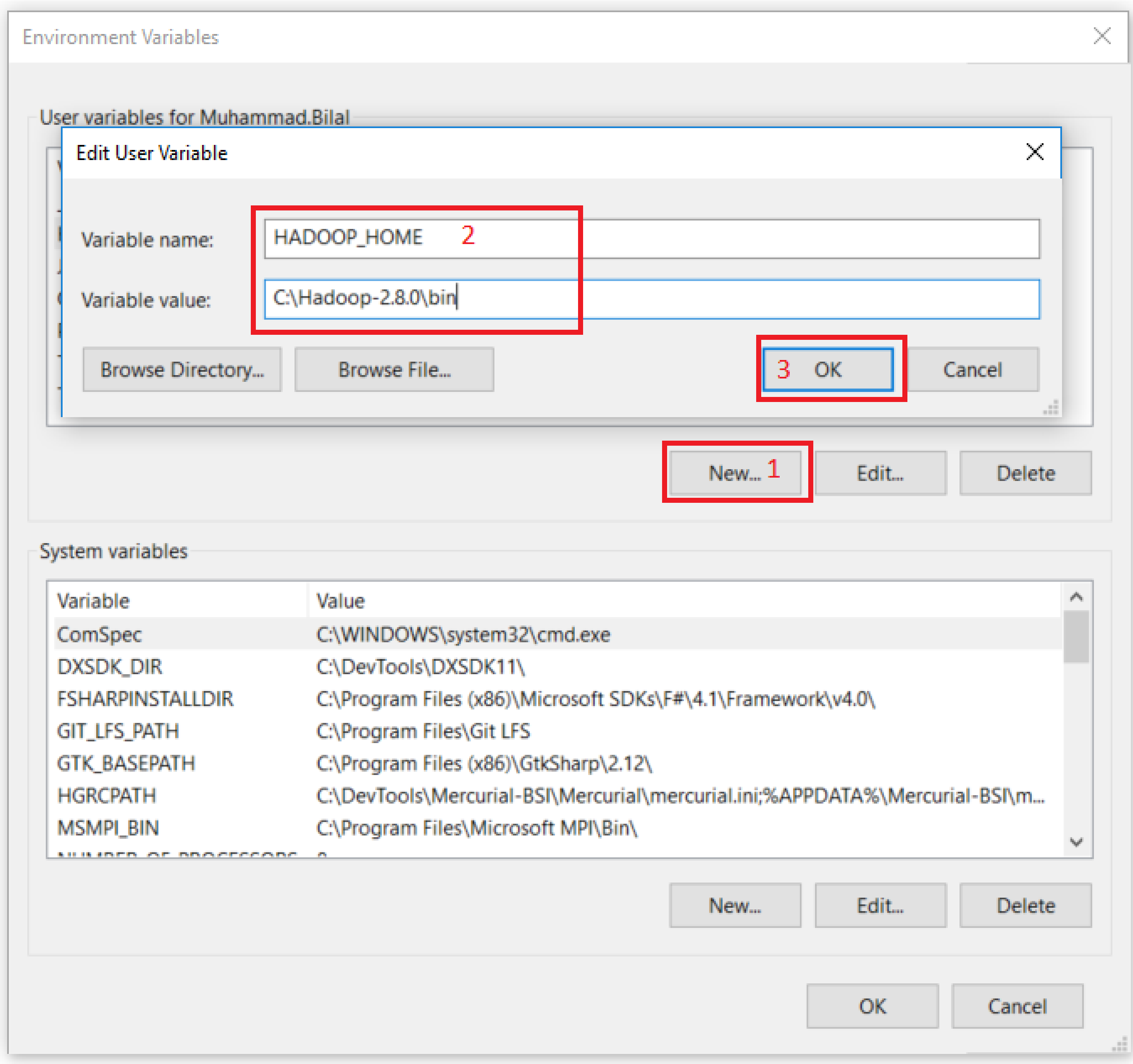
3. Extract file Hadoop 2.8.0.tar.gz or Hadoop-2.8.0.zip and place under **"C:\Hadoop-2.8.0"**.



This PC › Windows (C:)

□ Name	Date modified	Type
ATP	5/22/2017 3:19 PM	File folder
AzureTemp	7/18/2017 5:57 PM	File folder
cygwin64	7/18/2017 10:58 AM	File folder
DevTools	6/19/2017 12:39 PM	File folder
Hadoop-2.8.0	7/18/2017 12:43 PM	File folder
inetpub	5/8/2017 10:49 PM	File folder
Intel	4/25/2017 9:12 AM	File folder
ITSD	4/25/2017 9:20 AM	File folder
Java	7/18/2017 12:29 PM	File folder
PerfLogs	7/16/2016 4:47 PM	File folder
policies	5/18/2017 2:56 PM	File folder
Program Files	7/10/2017 1:06 PM	File folder
Program Files (x86)	7/12/2017 12:35 PM	File folder

4. Set the path HADOOP_HOME Environment variable on windows 10(see Step 1,2,3 and 4 below).



Environment Variables

User variables for Muhammad.Bilal

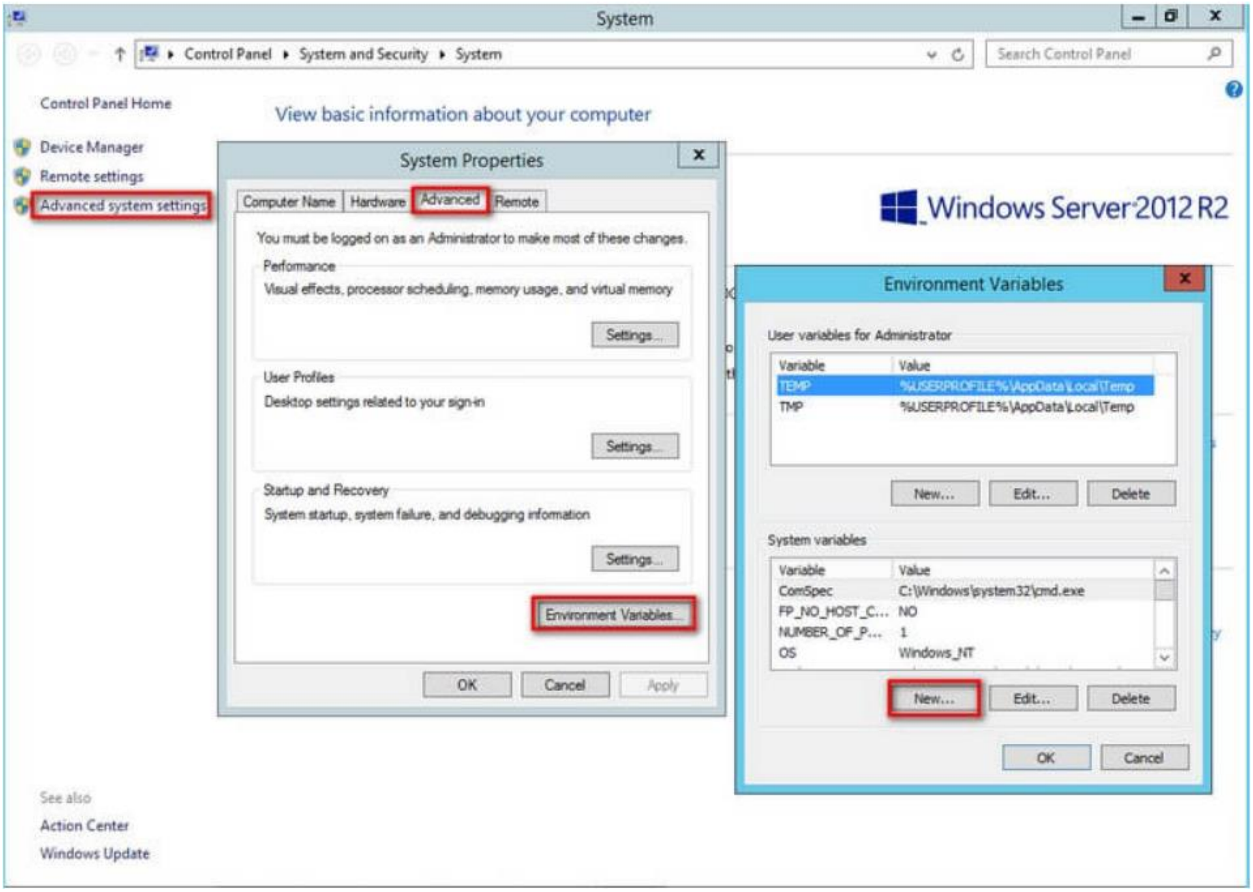
Edit User Variable

Variable name: HADOOP_HOME 2

Variable value: C:\Hadoop-2.8.0\bin

System variables

Variable	Value
ComSpec	C:\WINDOWS\system32\cmd.exe
DXSDK_DIR	C:\DevTools\DXSDK11\
FSHARPINSTALLDIR	C:\Program Files (x86)\Microsoft SDKs\F#\4.1\Framework\v4.0\
GIT_LFS_PATH	C:\Program Files\Git LFS
GTK_BASEPATH	C:\Program Files (x86)\GtkSharp\2.12\
HGRCPATH	C:\DevTools\Mercurial-BSI\Mercurial\mercurial.ini;%APPDATA%\Mercurial-BSI\m...
MSMPI_BIN	C:\Program Files\Microsoft MPI\Bin\



You must be logged on as an Administrator to make most of these changes.

Visual effects, processor scheduling, memory usage, and virtual memory

Desktop settings related to your sign-in

System startup, system failure, and debugging information

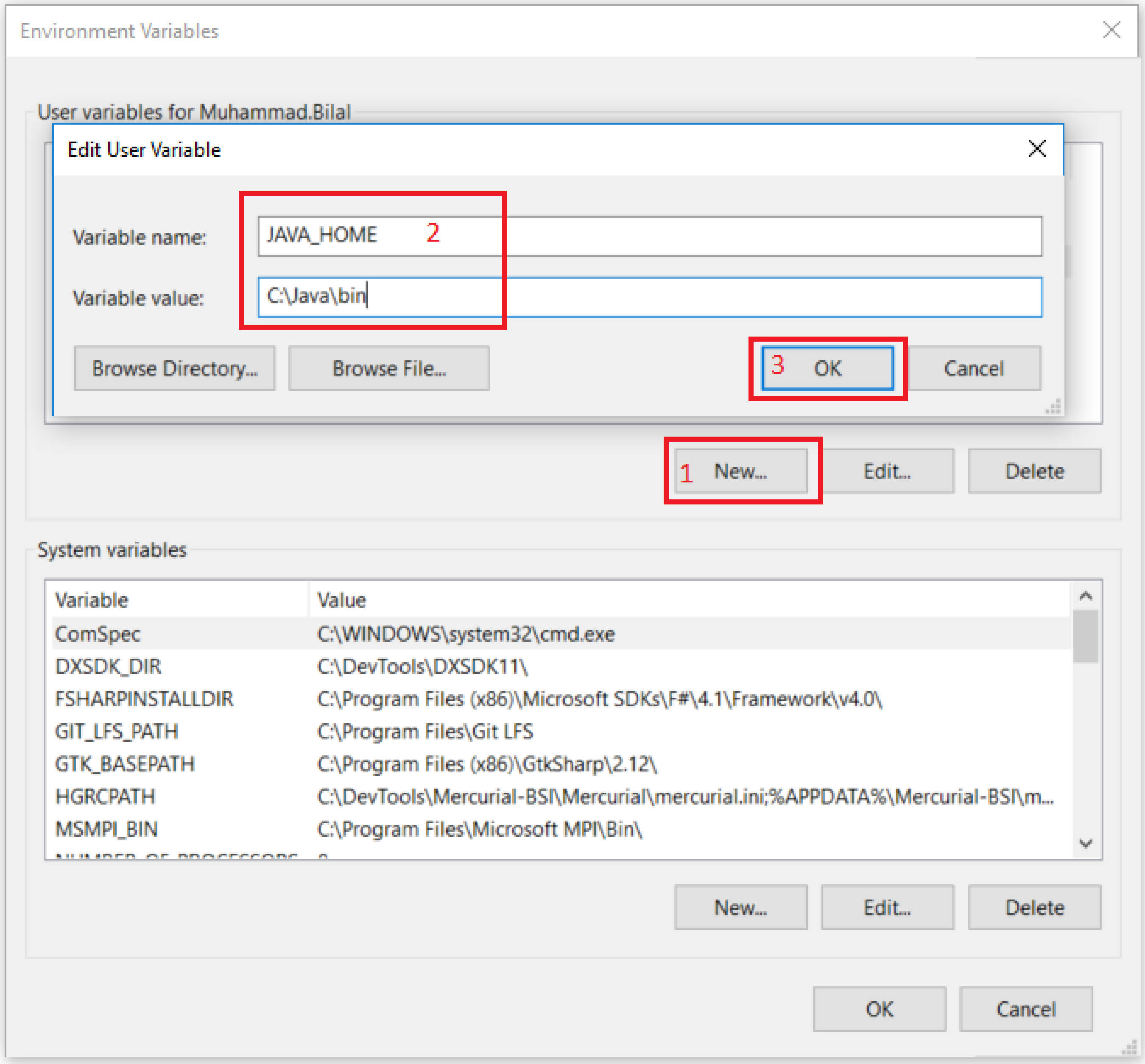
User variables for Administrator

Variable	Value
TEMP	%USERPROFILE%\AppData\Local\Temp
TMP	%USERPROFILE%\AppData\Local\Temp

System variables

Variable	Value
ComSpec	C:\Windows\system32\cmd.exe
FP_NO_HOST_C...	NO
NUMBER_OF_P...	1
OS	Windows_NT

5. Set the path JAVA_HOME Environment variable on windows 10(see Step 1,2,3 and 4 below).

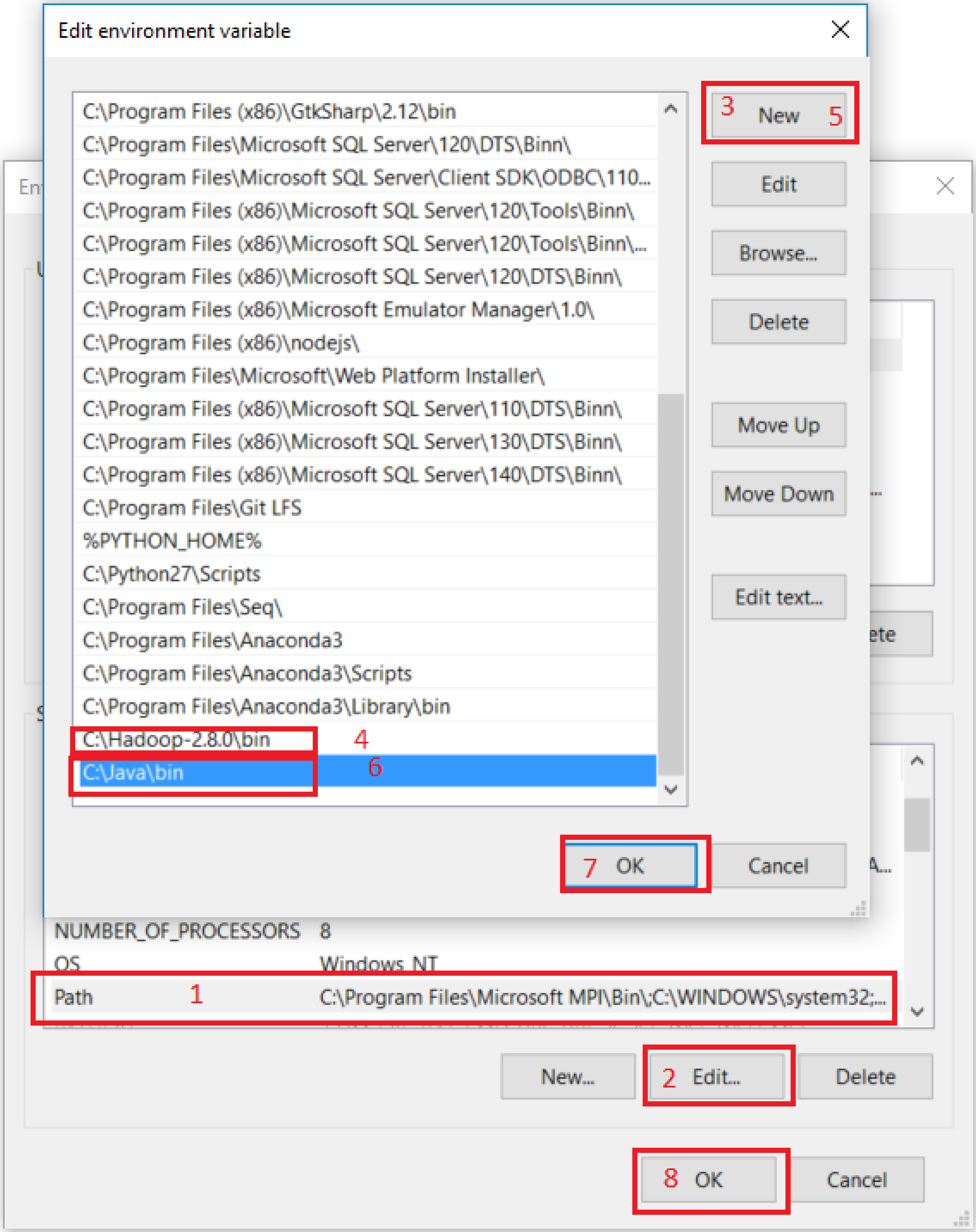


Variable name: JAVA_HOME 2

Variable value: C:\Java\bin

Variable	Value
ComSpec	C:\WINDOWS\system32\cmd.exe
DXSDK_DIR	C:\DevTools\DXSDK11\
FSHARPINSTALLDIR	C:\Program Files (x86)\Microsoft SDKs\F#\4.1\Framework\v4.0\
GIT_LFS_PATH	C:\Program Files\Git LFS
GTK_BASEPATH	C:\Program Files (x86)\GtkSharp\2.12\
HGRCPATH	C:\DevTools\Mercurial-BSI\Mercurial\mercurial.ini;%APPDATA%\Mercurial-BSI\m...
MSMPI_BIN	C:\Program Files\Microsoft MPI\Bin\
NUMBER_OF_PROCESSORS	8

6. Next, we set the Hadoop bin directory path and JAVA bin directory path.



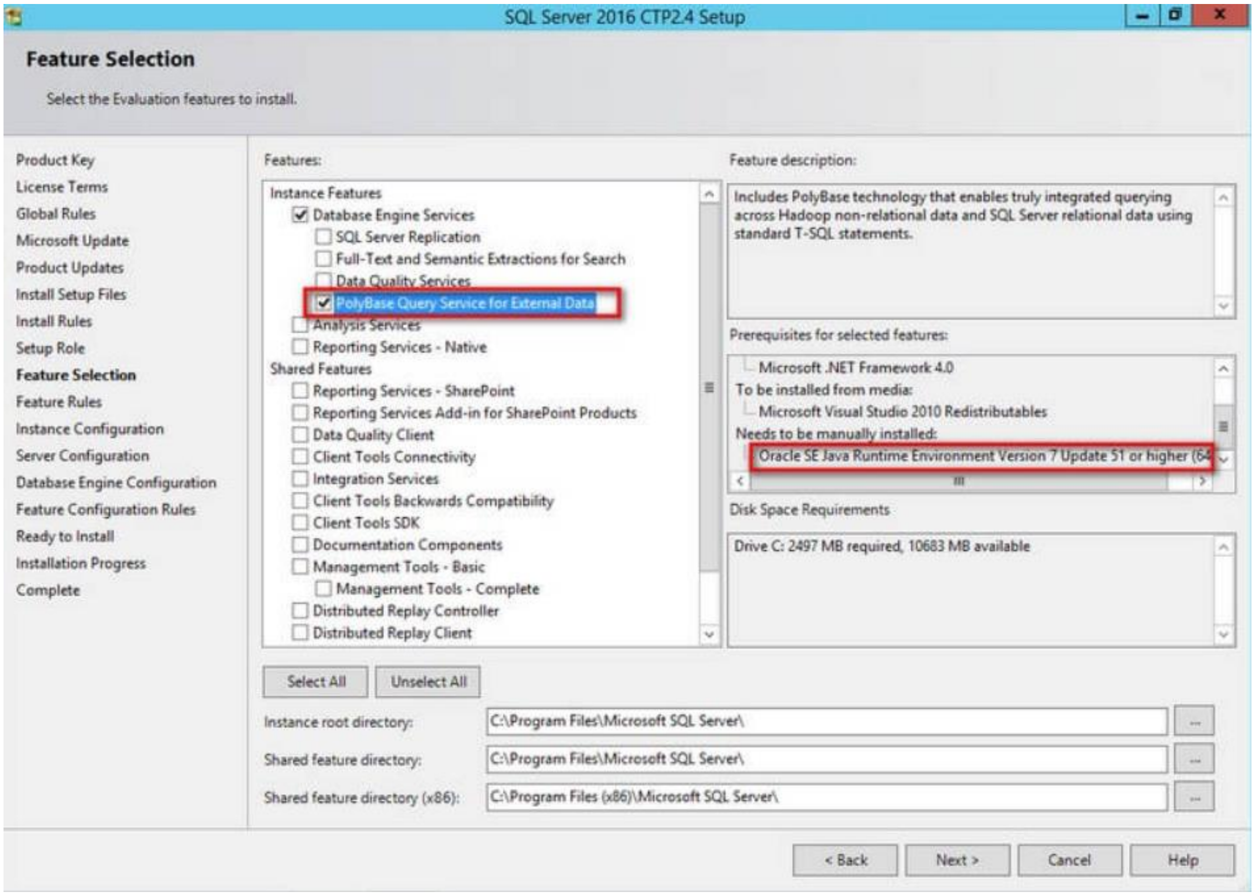
Edit environment variable

C:\Program Files (x86)\GtkSharp\2.12\bin
C:\Program Files\Microsoft SQL Server\120\DTS\Binn\
C:\Program Files\Microsoft SQL Server\Client SDK\ODBC\110...
C:\Program Files (x86)\Microsoft SQL Server\120\Tools\Binn\
C:\Program Files (x86)\Microsoft SQL Server\120\Tools\Binn\...
C:\Program Files (x86)\Microsoft SQL Server\120\DTS\Binn\
C:\Program Files (x86)\Microsoft Emulator Manager\1.0\
C:\Program Files (x86)\nodejs\
C:\Program Files\Microsoft\Web Platform Installer\
C:\Program Files (x86)\Microsoft SQL Server\110\DTS\Binn\
C:\Program Files (x86)\Microsoft SQL Server\130\DTS\Binn\
C:\Program Files (x86)\Microsoft SQL Server\140\DTS\Binn\
C:\Program Files\Git LFS
%PYTHON_HOME%
C:\Python27\Scripts
C:\Program Files\Seq\
C:\Program Files\Anaconda3
C:\Program Files\Anaconda3\Scripts
C:\Program Files\Anaconda3\Library\bin
C:\Hadoop-2.8.0\bin 4
C:\Java\bin 6

NUMBER_OF_PROCESSORS	8
OS	Windows_NT
Path 1	C:\Program Files\Microsoft MPI\Bin\;C:\WINDOWS\system32;...

Install **Polybase** SQL Server 2016 or Later

Since **Polybase** is now part of SQL Server, we can use the SQL Server 2016 installation media to do the installation. And because it was designed to interact with Hadoop, we will need to install the Oracle Java SE Runtime Environment (JRE) 7.51 (x64) or higher prior to running the SQL Server 2016 installation media.



SQL Server 2016 CTP2.4 Setup

Feature Selection

Select the Evaluation features to install.

Product Key
License Terms
Global Rules
Microsoft Update
Product Updates
Install Setup Files
Install Rules
Setup Role
Feature Selection
Feature Rules
Instance Configuration
Server Configuration
Database Engine Configuration
Feature Configuration Rules
Ready to Install
Installation Progress
Complete

Features:

Instance Features
- ☑ Database Engine Services
  - □ SQL Server Replication
  - □ Full-Text and Semantic Extractions for Search
  - □ Data Quality Services
  - ☑ PolyBase Query Service for External Data
- □ Analysis Services
- □ Reporting Services - Native

Shared Features
- □ Reporting Services - SharePoint
- □ Reporting Services Add-in for SharePoint Products
- □ Data Quality Client
- □ Client Tools Connectivity
- □ Integration Services
- □ Client Tools Backwards Compatibility
- □ Client Tools SDK
- □ Documentation Components
- □ Management Tools - Basic
  - □ Management Tools - Complete
- □ Distributed Replay Controller
- □ Distributed Replay Client

Feature description:

Includes PolyBase technology that enables truly integrated querying across Hadoop non-relational data and SQL Server relational data using standard T-SQL statements.

Prerequisites for selected features:

Microsoft .NET Framework 4.0
To be installed from media:
Microsoft Visual Studio 2010 Redistributables
Needs to be manually installed:
Oracle SE Java Runtime Environment Version 7 Update 51 or higher (64

Disk Space Requirements

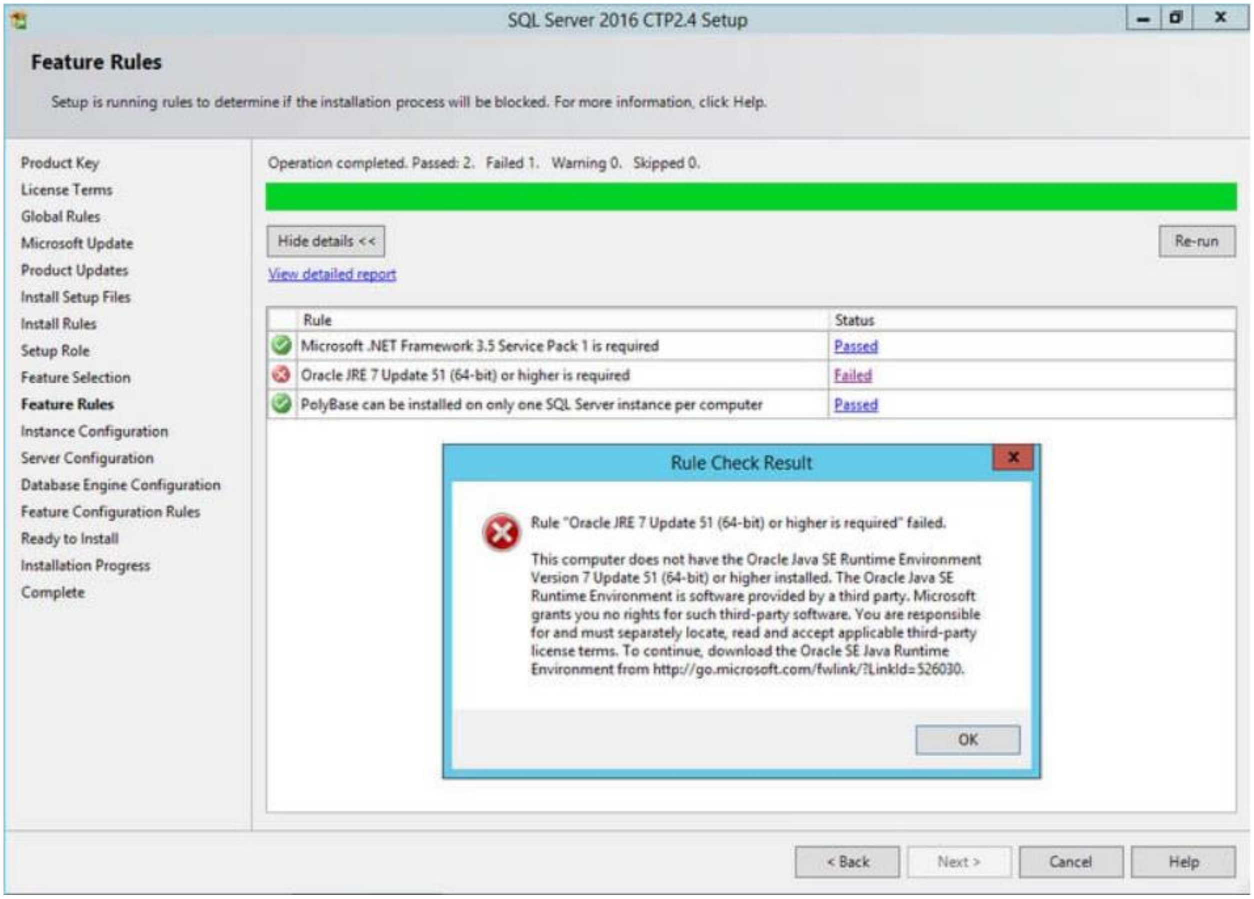
Drive C: 2497 MB required, 10683 MB available

Instance root directory: C:\Program Files\Microsoft SQL Server\

Shared feature directory: C:\Program Files\Microsoft SQL Server\

Shared feature directory (x86): C:\Program Files (x86)\Microsoft SQL Server\

If You don’t have JRE installed then the installation will fail and screen as below appears.



Feature Rules

Setup is running rules to determine if the installation process will be blocked. For more information, click Help.

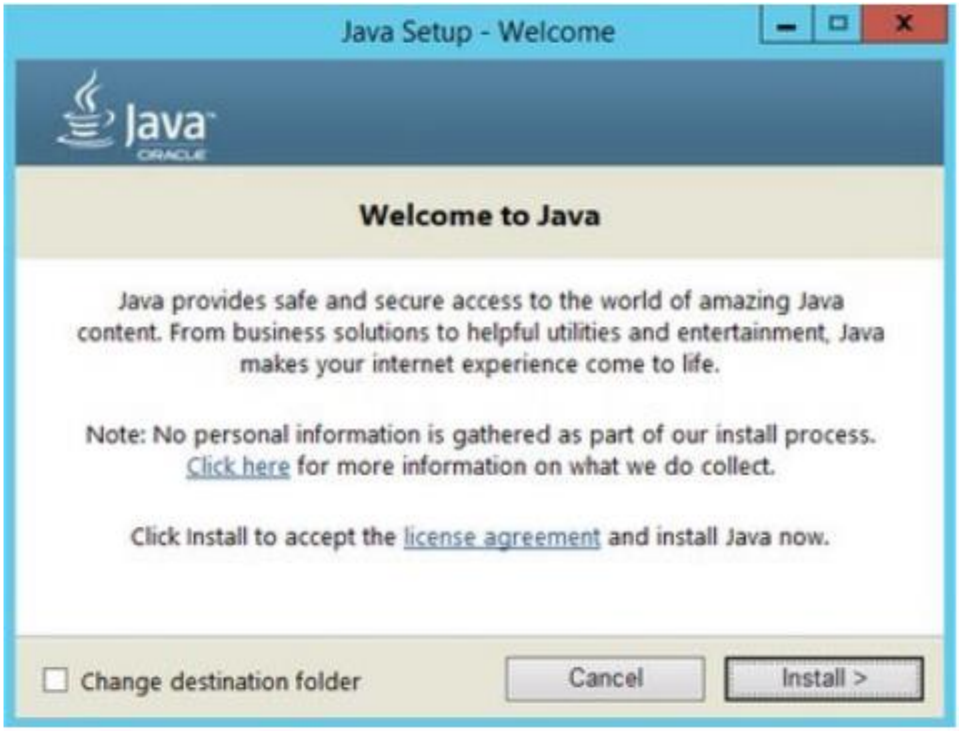
Operation completed. Passed: 2. Failed 1. Warning 0. Skipped 0.

	Rule	Status
✓	Microsoft .NET Framework 3.5 Service Pack 1 is required	Passed
☒	Oracle JRE 7 Update 51 (64-bit) or higher is required	Failed
✓	PolyBase can be installed on only one SQL Server instance per computer	Passed

Rule "Oracle JRE 7 Update 51 (64-bit) or higher is required" failed.

This computer does not have the Oracle Java SE Runtime Environment Version 7 Update 51 (64-bit) or higher installed. The Oracle Java SE Runtime Environment is software provided by a third party. Microsoft grants you no rights for such third-party software. You are responsible for and must separately locate, read and accept applicable third-party license terms. To continue, download the Oracle SE Java Runtime Environment from http://go.microsoft.com/fwlink/?LinkId=526030.

Install JRE and set the **Environment Variable** after installation which is mention in **Set Up** steps above.



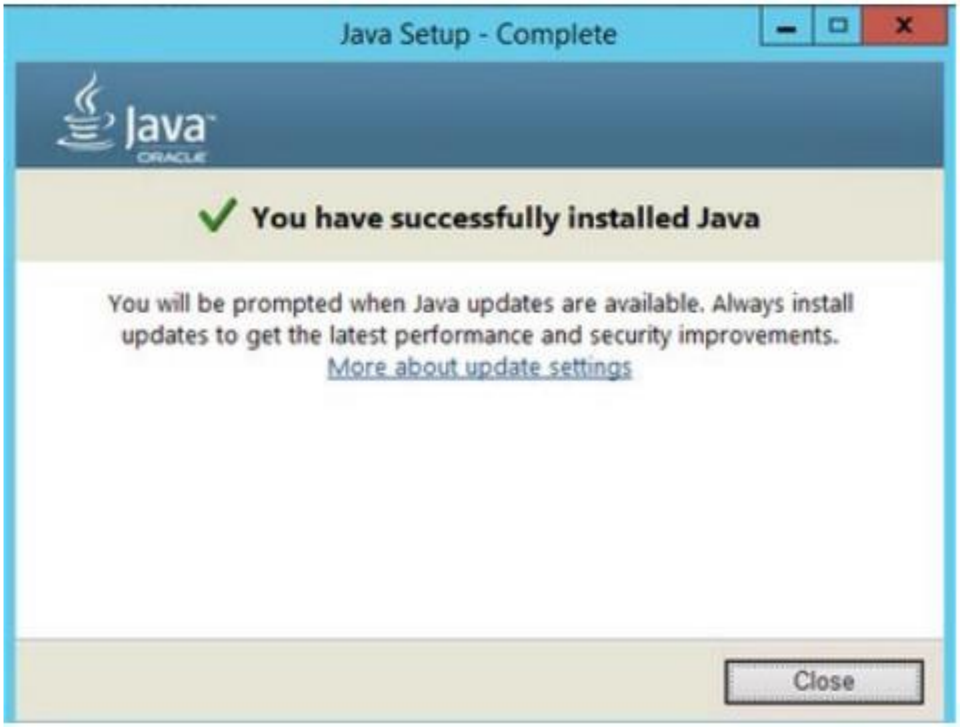
Java Setup - Welcome

Welcome to Java

Java provides safe and secure access to the world of amazing Java content. From business solutions to helpful utilities and entertainment, Java makes your internet experience come to life.

Note: No personal information is gathered as part of our install process. Click here for more information on what we do collect.

Click Install to accept the license agreement and install Java now.

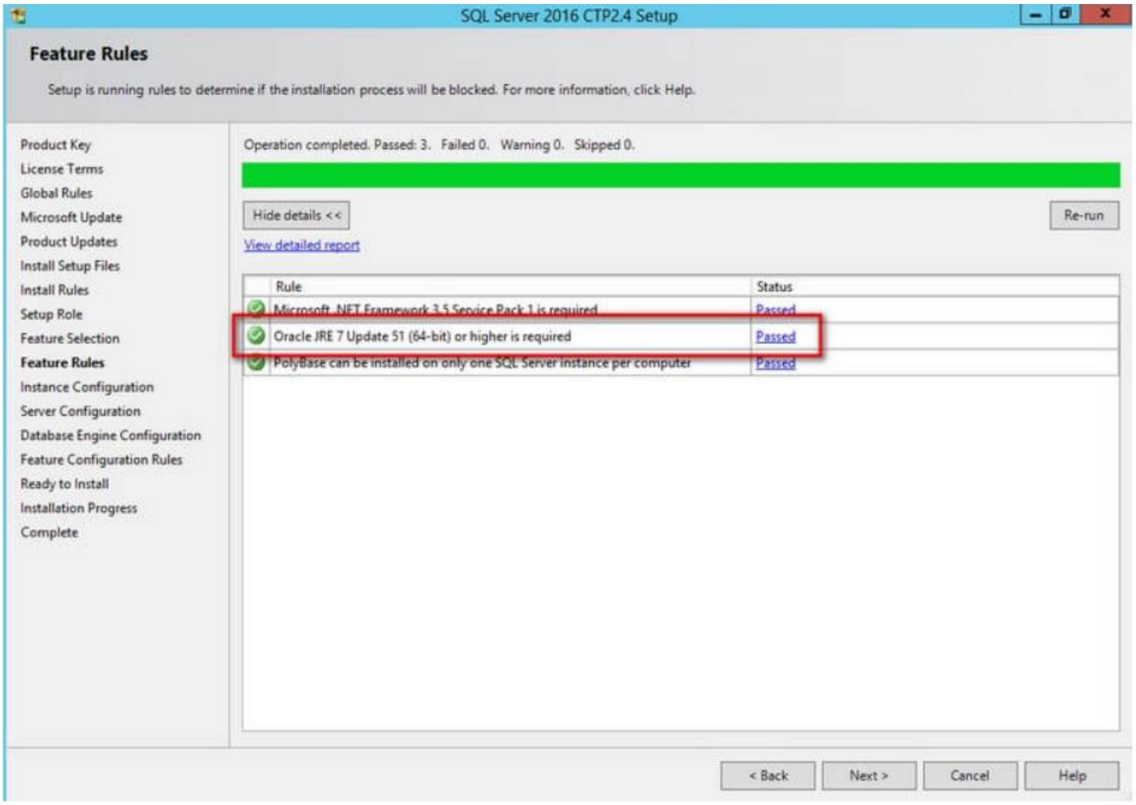


Java Setup - Complete

✓ You have successfully installed Java

You will be prompted when Java updates are available. Always install updates to get the latest performance and security improvements. More about update settings

And Continue the installation of Polybase and completed successfully.

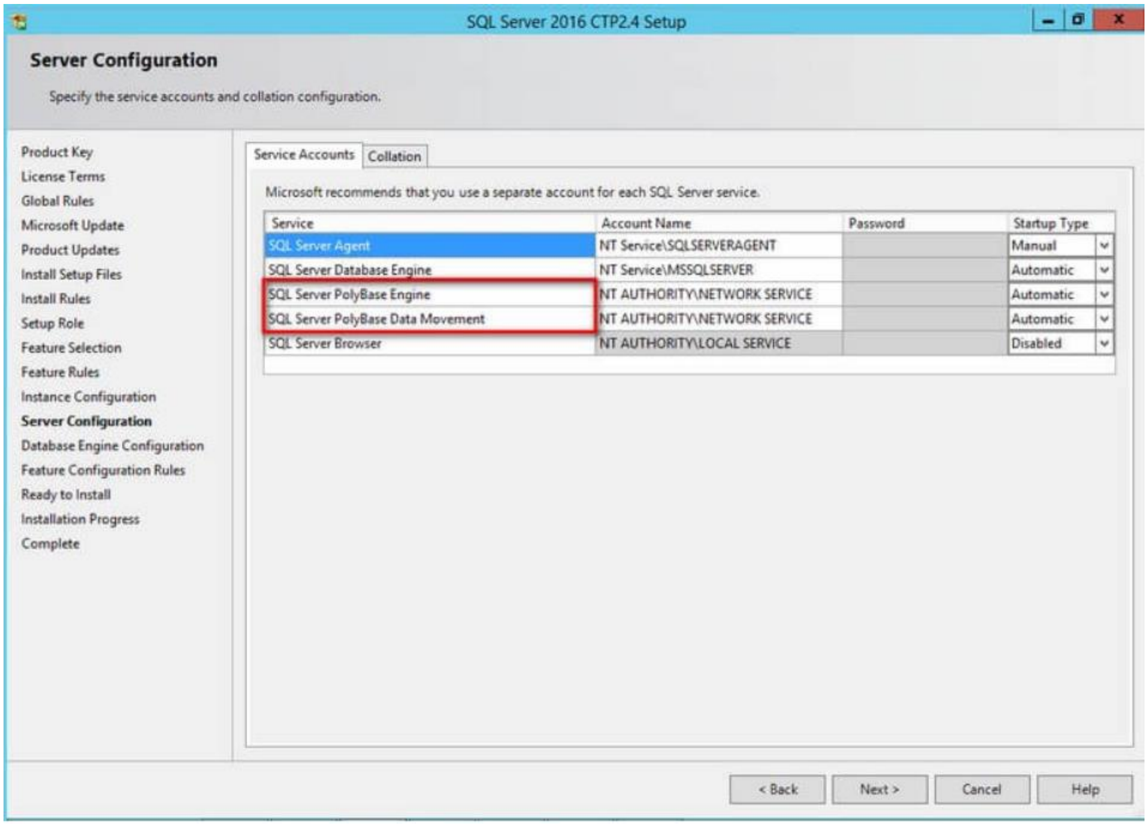


Feature Rules

Setup is running rules to determine if the installation process will be blocked. For more information, click Help.

Operation completed. Passed: 3. Failed 0. Warning 0. Skipped 0.

Rule	Status
Microsoft .NET Framework 3.5 Service Pack 1 is required	Passed
Oracle JRE 7 Update 51 (64-bit) or higher is required	Passed
PolyBase can be installed on only one SQL Server instance per computer	Passed



Server Configuration

Specify the service accounts and collation configuration.

Microsoft recommends that you use a separate account for each SQL Server service.

Service	Account Name	Password	Startup Type
SQL Server Agent	NT Service\SQLSERVERAGENT		Manual
SQL Server Database Engine	NT Service\MSSQLSERVER		Automatic
SQL Server PolyBase Engine	NT AUTHORITY\NETWORK SERVICE		Automatic
SQL Server PolyBase Data Movement	NT AUTHORITY\NETWORK SERVICE		Automatic
SQL Server Browser	NT AUTHORITY\LOCAL SERVICE		Disabled

Configuration

1. Edit file **C:/Hadoop-2.8.0/etc/hadoop/core-site.xml**, paste below xml paragraph and save this file.

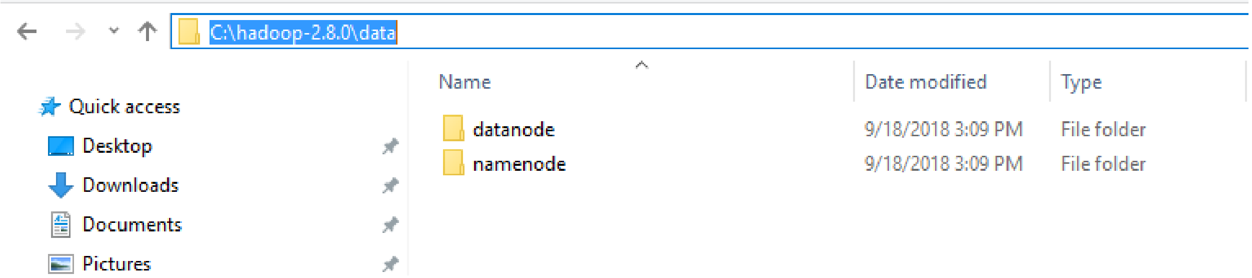
```
<configuration>
   <property>
       <name>fs.defaultFS</name>
       <value>hdfs://localhost:9000</value>
   </property>
</configuration>
```

2. Rename "mapred-site.xml.template" to "mapred-site.xml" and edit this file **C:/Hadoop-2.8.0/etc/hadoop/mapred-site.xml**, paste below xml paragraph and save this file.

```
<configuration>
   <property>
       <name>mapreduce.framework.name</name>
       <value>yarn</value>
   </property>
</configuration>
```

3. Create folder **"data"** under **"C:\Hadoop-2.8.0"**

- Create folder **"datanode"** under **"C:\Hadoop-2.8.0\data"**
- Create folder **"namenode"** under **"C:\Hadoop-2.8.0\data"**



C:\hadoop-2.8.0\data

Name	Date modified	Type
datanode	9/18/2018 3:09 PM	File folder
namenode	9/18/2018 3:09 PM	File folder

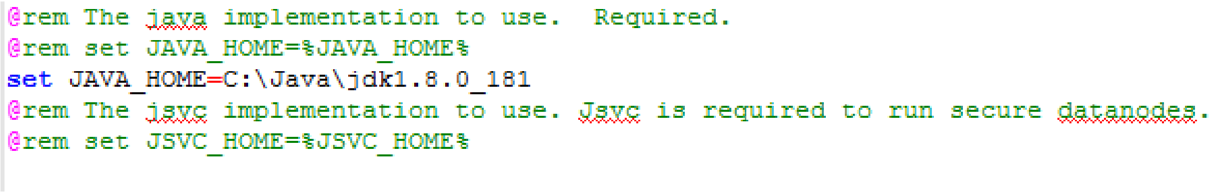
4. Edit file **C:\Hadoop-2.8.0/etc/hadoop/hdfs-site.xml**, paste below xml paragraph and save this file.

```
<configuration>
   <property>
       <name>dfs.replication</name>
       <value>1</value>
   </property>
   <property>
       <name>dfs.namenode.name.dir</name>
       <value>C:\hadoop-2.8.0\data\namenode</value>
   </property>
   <property>
       <name>dfs.datanode.data.dir</name>
       <value>C:\hadoop-2.8.0\data\datanode</value>
   </property>
</configuration>
```

5. Edit file **C:/Hadoop-2.8.0/etc/hadoop/yarn-site.xml**, paste below xml paragraph and save this file.

```
<configuration>
   <property>
        <name>yarn.nodemanager.aux-services</name>
        <value>mapreduce_shuffle</value>
   </property>
   <property>
        <name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>
        <value>org.apache.hadoop.mapred.ShuffleHandler</value>
   </property>
</configuration>
```

6. Edit file **C:/Hadoop-2.8.0/etc/hadoop/hadoop-env.cmd** by closing the command line**"JAVA_HOME=%JAVA_HOME%"** instead of set **"JAVA_HOME=C:\Java"** (On C:\java this is path to file jdk.18.0)



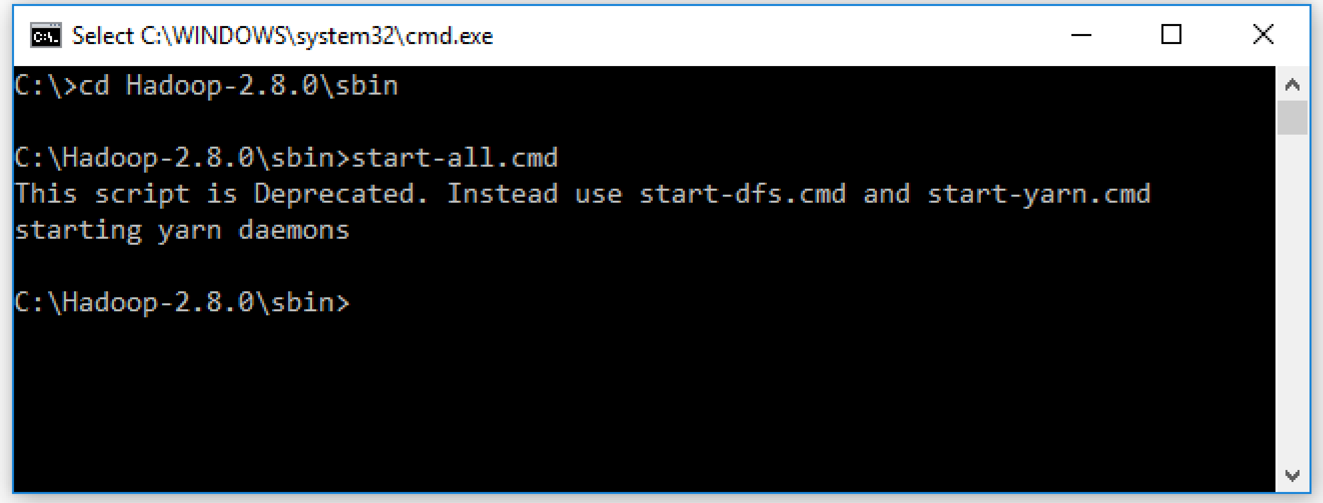
```
@rem The java implementation to use.  Required.
@rem set JAVA_HOME=%JAVA_HOME%
set JAVA_HOME=C:\Java\jdk1.8.0_181
@rem The jsvc implementation to use. Jsvc is required to run secure datanodes.
@rem set JSVC_HOME=%JSVC_HOME%
```

Hadoop Configuration

1. Download file Hadoop Configuration.zip (Link: https://drive.google.com/open?id=1e6pYQxZrr1JQaAR5ywOWNxkEYjQKKpBL)
2. Delete file bin on C:\Hadoop-2.8.0\bin, replaced by file bin on file just download (from Hadoop Configuration.zip).
3. Open cmd and typing command **"hdfs namenode –format"**. You will see

Testing

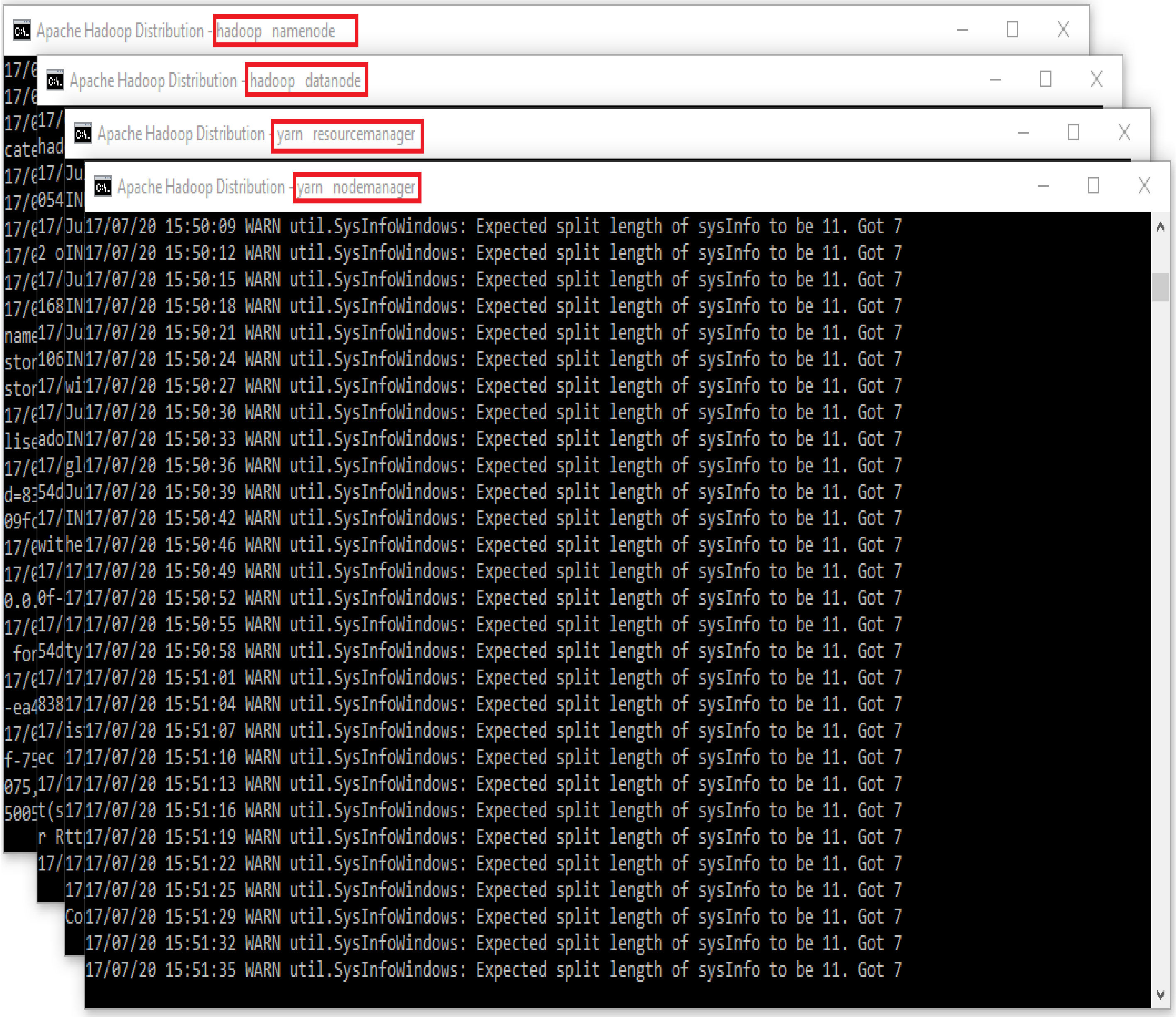
1. Open cmd and change directory to "C:\Hadoop-2.8.0\sbin" and type **"start-all.cmd"** to start
apache.



```
C:\>cd Hadoop-2.8.0\sbin

C:\Hadoop-2.8.0\sbin>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons

C:\Hadoop-2.8.0\sbin>
```

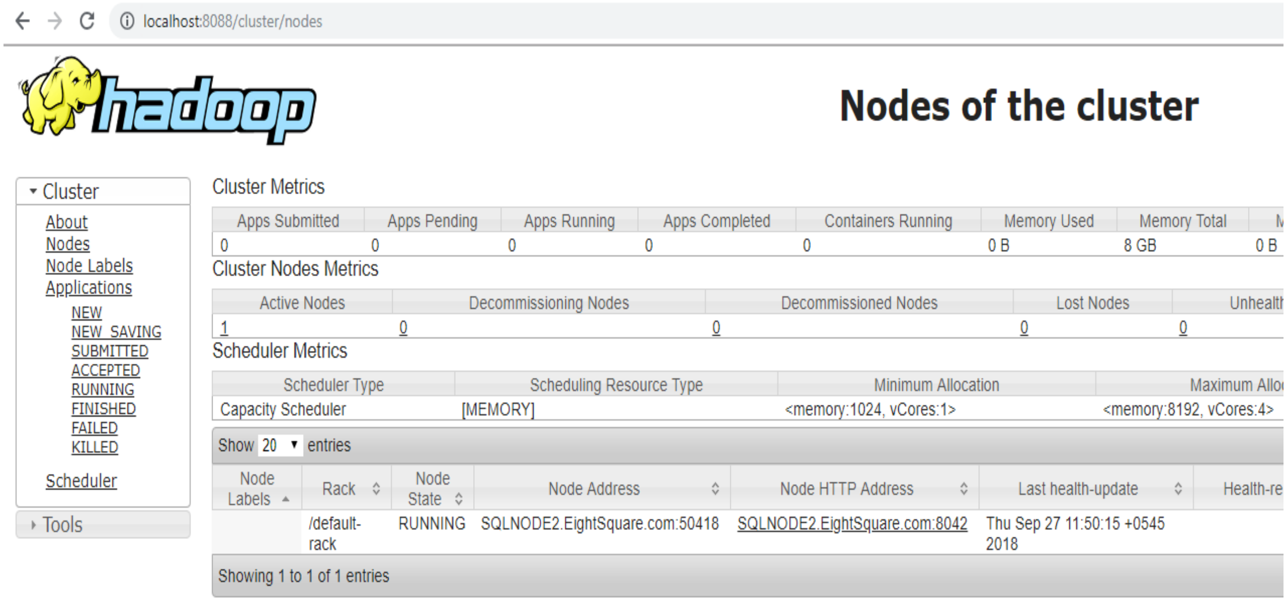
2. Make sure these apps are running

- Hadoop Namenode
- Hadoop datanode
- YARN Resource Manager
- YARN Node Manager


Apache Hadoop Distribution - hadoop namenode
Apache Hadoop Distribution - hadoop datanode
Apache Hadoop Distribution - yarn resourcemanager
Apache Hadoop Distribution - yarn nodemanager

```
17/07/20 15:50:09 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:50:12 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:50:15 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:50:18 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:50:21 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:50:24 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:50:27 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:50:30 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:50:33 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:50:36 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:50:39 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:50:42 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:50:46 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:50:49 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:50:52 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:50:55 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:50:58 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:51:01 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:51:04 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:51:07 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:51:10 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:51:13 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:51:16 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:51:19 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:51:22 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:51:25 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:51:29 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:51:32 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
17/07/20 15:51:35 WARN util.SysInfoWindows: Expected split length of sysInfo to be 11. Got 7
```

Open: http://localhost:8088



Nodes of the cluster

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	M
0	0	0	0	0	0 B	8 GB	0 B

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealth
1	0	0	0	0

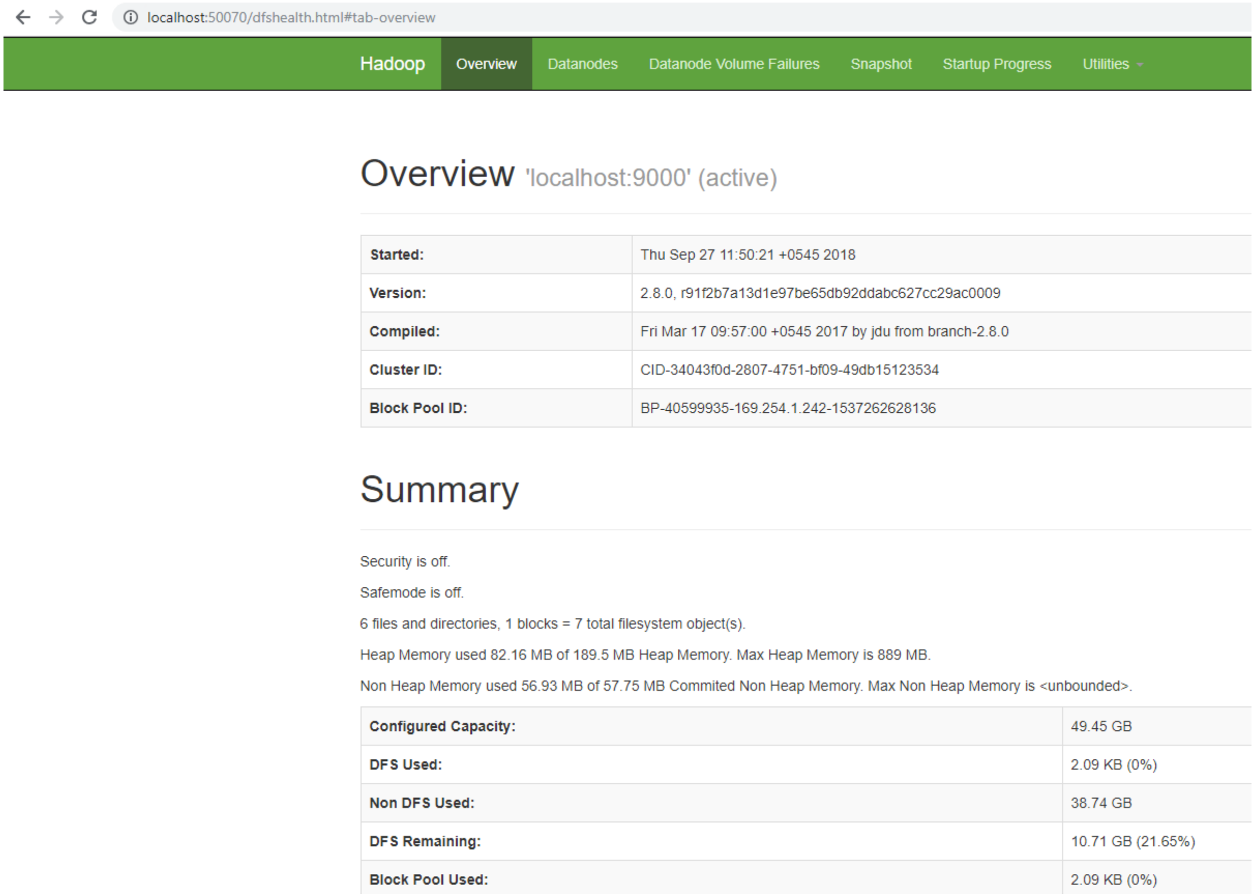
Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allo
Capacity Scheduler	[MEMORY]	<memory:1024, vCores:1>	<memory:8192, vCores:4>

Node Labels	Rack	Node State	Node Address	Node HTTP Address	Last health-update	Health-re
	/default-rack	RUNNING	SQLNODE2.EightSquare.com:50418	SQLNODE2.EightSquare.com:8042	Thu Sep 27 11:50:15 +0545 2018	

Showing 1 to 1 of 1 entries

Open: http://localhost:50070



Overview 'localhost:9000' (active)

Started:	Thu Sep 27 11:50:21 +0545 2018
Version:	2.8.0, r91f2b7a13d1e97be65db92ddabc627cc29ac0009
Compiled:	Fri Mar 17 09:57:00 +0545 2017 by jdu from branch-2.8.0
Cluster ID:	CID-34043f0d-2807-4751-bf09-49db15123534
Block Pool ID:	BP-40599935-169.254.1.242-1537262628136

Summary

Security is off.

Safemode is off.

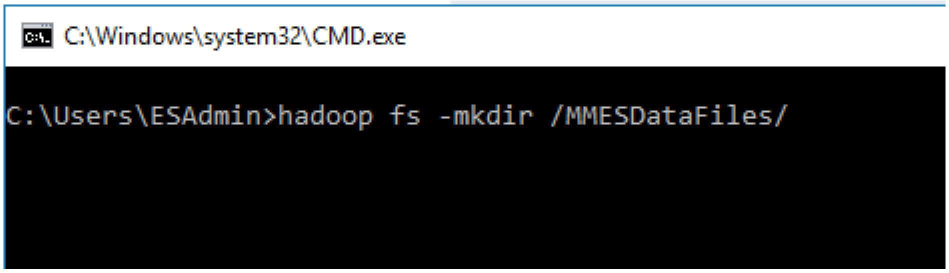
6 files and directories, 1 blocks = 7 total filesystem object(s).

Heap Memory used 82.16 MB of 189.5 MB Heap Memory. Max Heap Memory is 889 MB.

Non Heap Memory used 56.93 MB of 57.75 MB Commited Non Heap Memory. Max Non Heap Memory is <unbounded>.

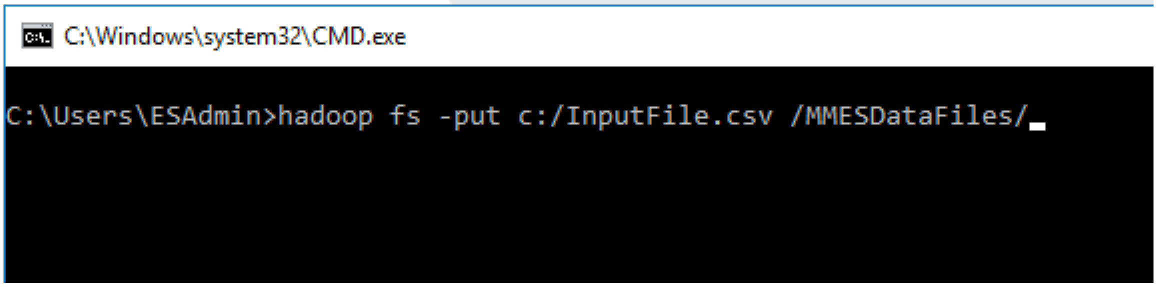
Configured Capacity:	49.45 GB
DFS Used:	2.09 KB (0%)
Non DFS Used:	38.74 GB
DFS Remaining:	10.71 GB (21.65%)
Block Pool Used:	2.09 KB (0%)

Create Directory in Hadoop



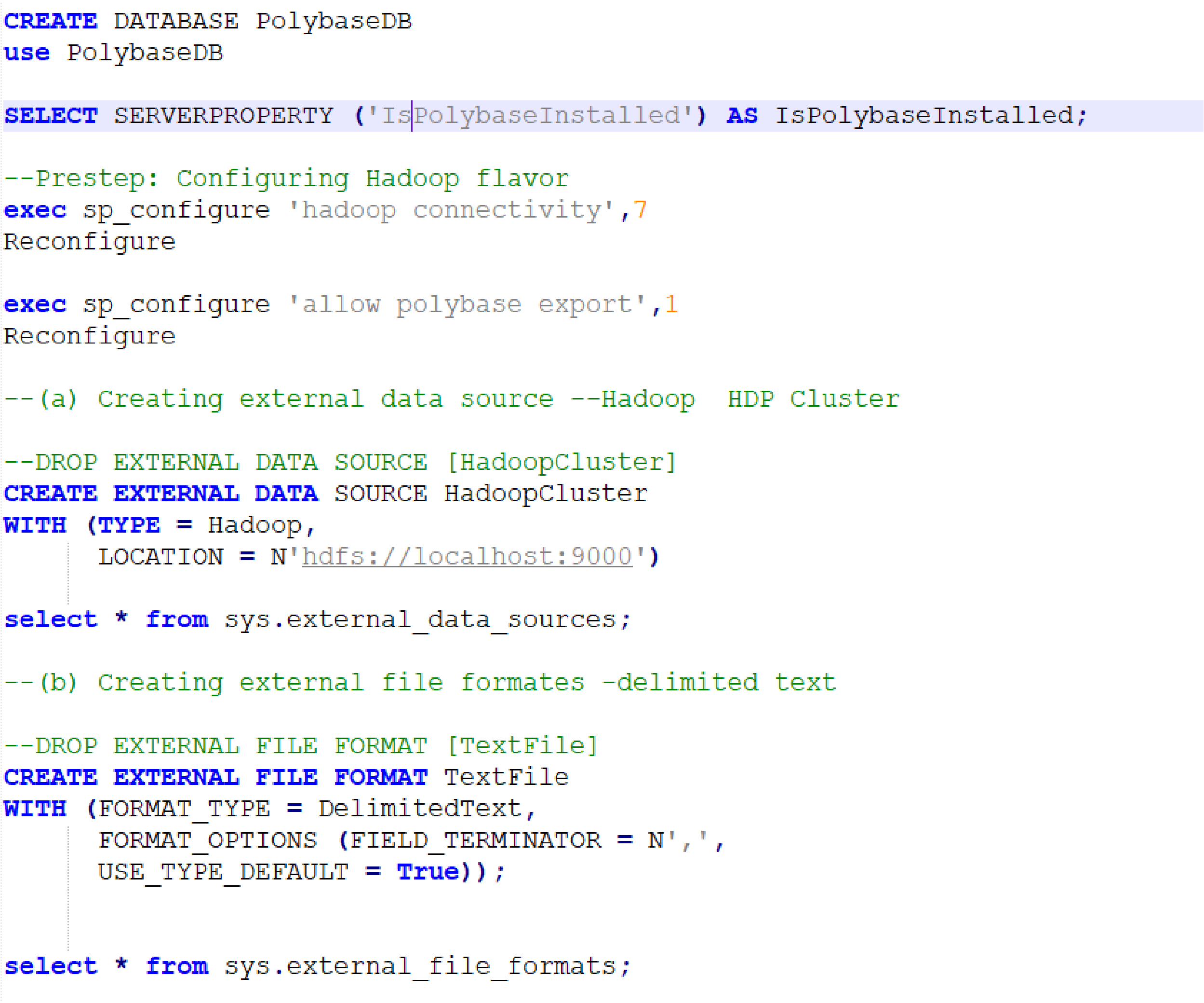
```
C:\Users\ESAdmin>hadoop fs -mkdir /MMESDataFiles/
```

Import file to MMESDataFiles directory

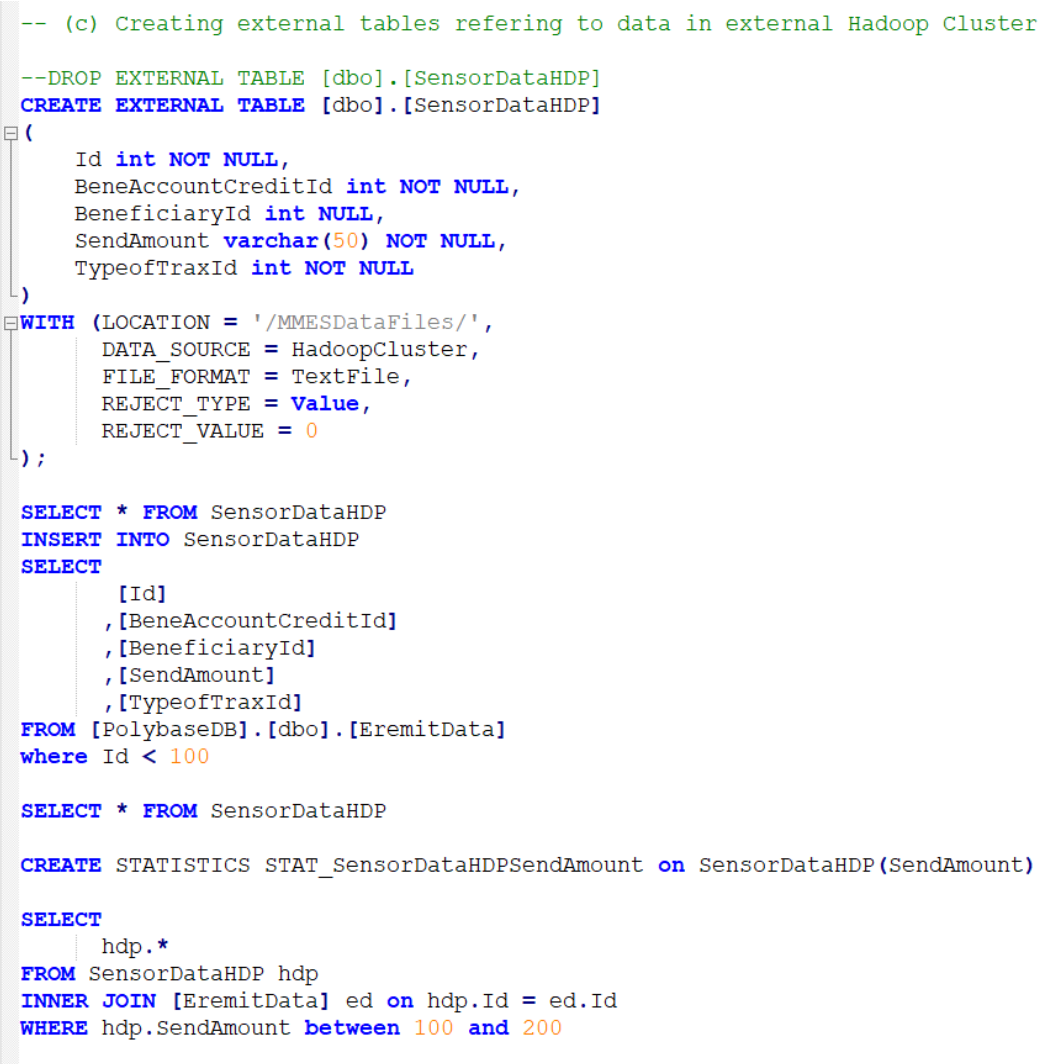


```
C:\Users\ESAdmin>hadoop fs -put c:/InputFile.csv /MMESDataFiles/_
```

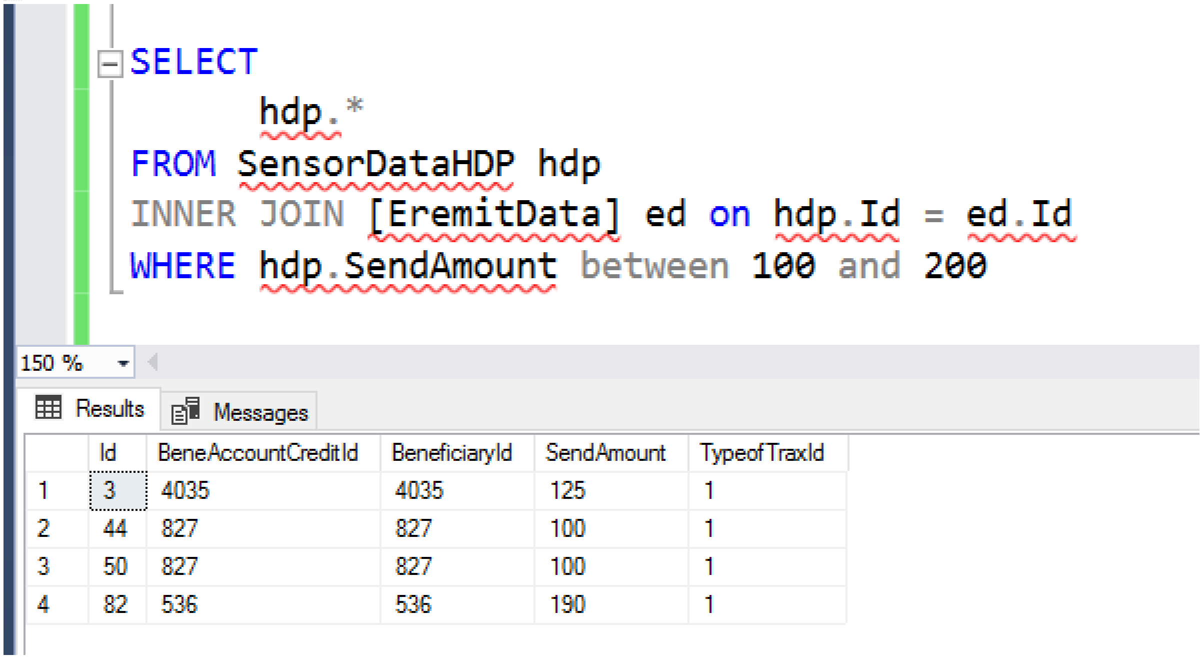
Simple procedure to configure database for **PolyBase**, Hadoop Connectivity,
And Create External Data Source, Create External File Format and Create External Table also create Statistics on external table for query optimization.



```
CREATE DATABASE PolybaseDB
use PolybaseDB

SELECT SERVERPROPERTY ('IsPolybaseInstalled') AS IsPolybaseInstalled;

--Prestep: Configuring Hadoop flavor
exec sp_configure 'hadoop connectivity',7
Reconfigure

exec sp_configure 'allow polybase export',1
Reconfigure

--(a) Creating external data source --Hadoop  HDP Cluster

--DROP EXTERNAL DATA SOURCE [HadoopCluster]
CREATE EXTERNAL DATA SOURCE HadoopCluster
WITH (TYPE = Hadoop,
      LOCATION = N'hdfs://localhost:9000')

select * from sys.external_data_sources;

--(b) Creating external file formates -delimited text

--DROP EXTERNAL FILE FORMAT [TextFile]
CREATE EXTERNAL FILE FORMAT TextFile
WITH (FORMAT_TYPE = DelimitedText,
      FORMAT_OPTIONS (FIELD_TERMINATOR = N',',
      USE_TYPE_DEFAULT = True));


select * from sys.external_file_formats;
```



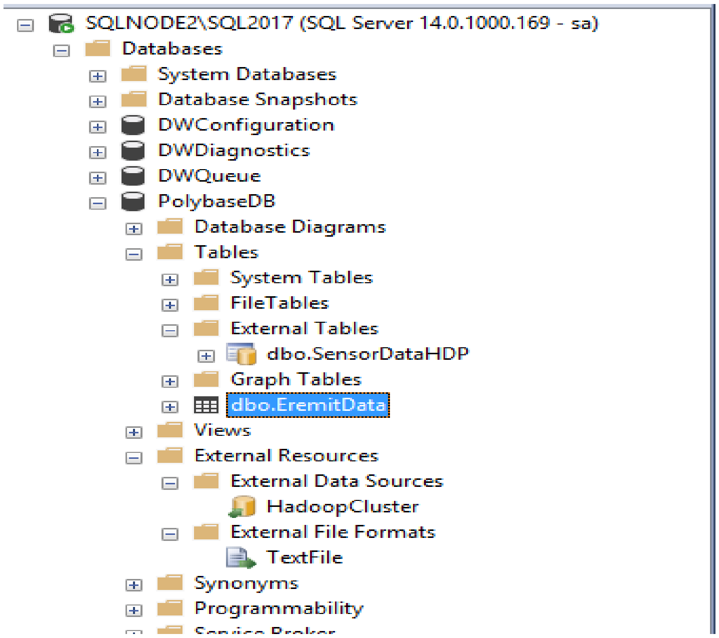
```
-- (c) Creating external tables refering to data in external Hadoop Cluster

--DROP EXTERNAL TABLE [dbo].[SensorDataHDP]
CREATE EXTERNAL TABLE [dbo].[SensorDataHDP]
(
    Id int NOT NULL,
    BeneAccountCreditId int NOT NULL,
    BeneficiaryId int NULL,
    SendAmount varchar(50) NOT NULL,
    TypeofTraxId int NOT NULL
)
WITH (LOCATION = '/MMESDataFiles/',
      DATA_SOURCE = HadoopCluster,
      FILE_FORMAT = TextFile,
      REJECT_TYPE = Value,
      REJECT_VALUE = 0
);

SELECT * FROM SensorDataHDP
INSERT INTO SensorDataHDP
SELECT
       [Id]
      ,[BeneAccountCreditId]
      ,[BeneficiaryId]
      ,[SendAmount]
      ,[TypeofTraxId]
FROM [PolybaseDB].[dbo].[EremitData]
where Id < 100

SELECT * FROM SensorDataHDP

CREATE STATISTICS STAT_SensorDataHDPSendAmount on SensorDataHDP(SendAmount)

SELECT
      hdp.*
FROM SensorDataHDP hdp
INNER JOIN [EremitData] ed on hdp.Id = ed.Id
WHERE hdp.SendAmount between 100 and 200
```



```
SELECT
      hdp.*
FROM SensorDataHDP hdp
INNER JOIN [EremitData] ed on hdp.Id = ed.Id
WHERE hdp.SendAmount between 100 and 200
```

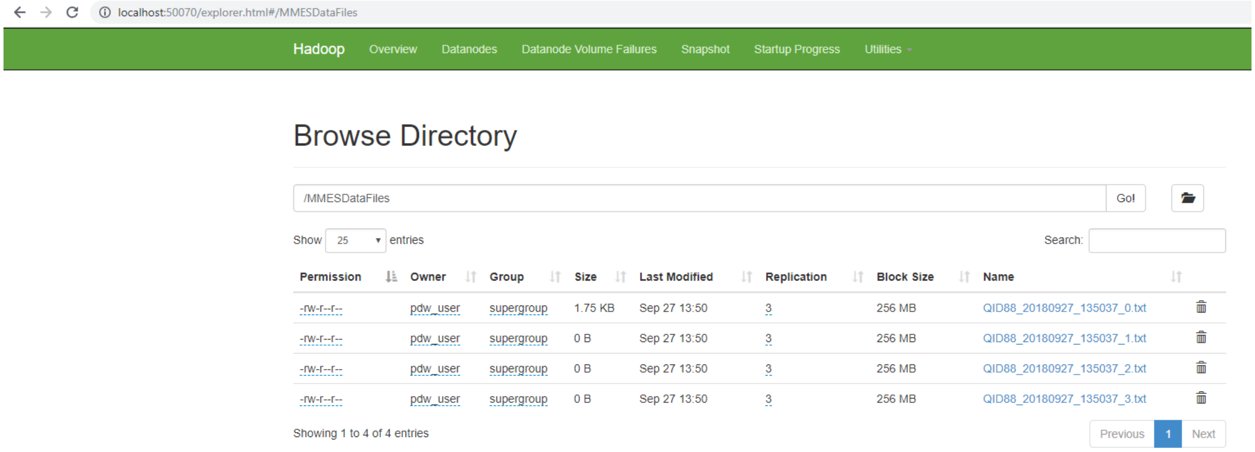
	Id	BeneAccountCreditId	BeneficiaryId	SendAmount	TypeofTraxId
1	3	4035	4035	125	1
2	44	827	827	100	1
3	50	827	827	100	1
4	82	536	536	190	1

Structure of External Data Source, External File Format and External Table in SSMS.



```
SQLNODE2\SQL2017 (SQL Server 14.0.1000.169 - sa)
    Databases
        System Databases
        Database Snapshots
        DWConfiguration
        DWDiagnostics
        DWQueue
        PolybaseDB
            Database Diagrams
            Tables
                System Tables
                FileTables
                External Tables
                    dbo.SensorDataHDP
                Graph Tables
                dbo.EremitData
            Views
            External Resources
                External Data Sources
                    HadoopCluster
                External File Formats
                    TextFile
            Synonyms
            Programmability
            Service Broker
```

File Resides in HDFS



localhost:50070/explorer.html#/MMESDataFiles

Browse Directory

/MMESDataFiles Go!

Show 25 entries Search:

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	pdw_user	supergroup	1.75 KB	Sep 27 13:50	3	256 MB	QID88_20180927_135037_0.txt
-rw-r--r--	pdw_user	supergroup	0 B	Sep 27 13:50	3	256 MB	QID88_20180927_135037_1.txt
-rw-r--r--	pdw_user	supergroup	0 B	Sep 27 13:50	3	256 MB	QID88_20180927_135037_2.txt
-rw-r--r--	pdw_user	supergroup	0 B	Sep 27 13:50	3	256 MB	QID88_20180927_135037_3.txt

Showing 1 to 4 of 4 entries

Previous 1 Next