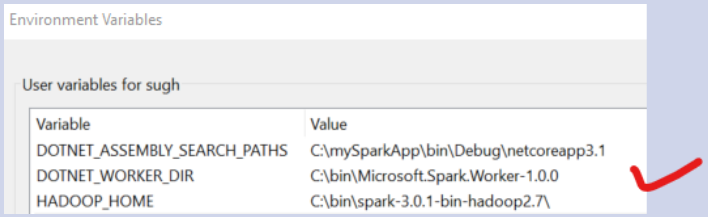
.NET FOR APACHE SPARK - TROUBLESHOOTING

Prepared by: Subhasish Ghosh, Sept2021

	Issue(s)	Solution(s)
1	.NET SDK fails to install.	Use .NET SDK for 64-bit. If you're using Windows 11, use .NET SDK 64-bit as well. If you're using older versions of Windows, see link for more information; additional components are needed for running code.
2	Locating Apache Spark fails after installing Spark on Windows.	You need to set following commands to set environment variables. setx HADOOP_HOME C:\bin\spark-3.0.1-bin-hadoop2.7\ setx SPARK_HOME C:\bin\spark-3.0.1-bin-hadoop2.7\ setx /M PATH "%PATH%;%HADOOP_HOME%;%SPARK_HOME%\bin" (change paths as on your system in actual)
3	In above line whilst setting environment variables, the 3rd command fails.	You need to open command prompt in "Administrator" mode. Then execute: setx /M PATH "%PATH%;%HADOOP_HOME%;%SPARK_HOME%\bin"
4	Firing spark-submit fails with 'command not found' or recognized.	Open a new command prompt and do, echo %SPARK_HOME% You need to set env variable for this, OR alternatively, use: > %SPARK_HOME%\bin\spark-submit
5	.NET apps fail to locate .NET for Apache Spark.	Open a new command prompt and execute: setx DOTNET_WORKER_DIR "C:\bin\Microsoft.Spark.Worker-1.0.0"
6	Executing 'dotnet build' fails with Errors.	Open the source code file either in Editor/VStudio and check syntax/code.
7	After successful build, when you fire spark-submit, it fails because it fails to get all "related dependent DLL files for it to run". In other words, you see Assembly Errors and spark-submit exits.	You need to set the environment variable: "DOTNET_ASSEMBLY_SEARCH_PATHS" to variable value of wherever on your local system you are building your app. E.g., on my Windows11 laptop, it is set as: DOTNET_ASSEMBLY_SEARCH_PATHS = C:\mySparkApp\bin\Debug\netcoreapp3.1 Setting this tells spark-submit to find all related DLLs needed to run this App (change as per your folder hierarchy). Additionally, the order is: DOTNET_ASSEMBLY_SEARCH_PATHS, then, DOTNET_WORKER_DIR, and then, HADOOP_HOME.  After setting this variable, best practice is to quit the Command prompt, open a fresh one, and also run "dotnet build" again. And then fire the spark-submit command.

Environment Variables

User variables for sugh

Variable	Value
DOTNET_ASSEMBLY_SEARCH_PATHS	C:\mySparkApp\bin\Debug\netcoreapp3.1
DOTNET_WORKER_DIR	C:\bin\Microsoft.Spark.Worker-1.0.0
HADOOP_HOME	C:\bin\spark-3.0.1-bin-hadoop2.7\