When do you use Apache Spark pools in Azure Synapse Analytics?

With the variety of Apache Spark data services that are available on Azure, the following table outlines where Azure Synapse Analytics Apache Spark pools fit in the ecosystem.

	Apache Spark	Azure HDInsight	Azure Databricks	Apache Spark for Azure Synapse Analytics
What	An open-source, memory-optimized system for managing big data workloads.	Microsoft implementation of opensource Apache Spark managed within the realms of Azure.	An advanced analytics managed Apache Spark-as-a-Service solution.	Embedded Apache Spark capability within Azure Synapse Analytics residing on the same platform that contains data warehouses and data integration capabilities, as well as integrating with other Azure services.
When	When you want to benefits of Apache Spark for big data processing and/or data science work without the Service Level Agreements (SLA's) of a provider.	When you want the benefits of OSS Spark with the SLA of a provider.	When you need an end-to-end data engineering and data science solution and management platform.	When organizations without existing Apache Spark implementations want to fire up an Apache Spark cluster to meet data engineering needs without the overhead of the other Apache Spark platforms listed.
Who	Open-source Professionals.	Open-source Professionals who require SLA's and Microsoft Data Platform experts.	Data Engineers and Data Scientists working on big data projects every day.	Data Engineers, Data Scientists, Data Platform experts and Data Analysts.
Why	To overcome the limitations of symmetric multiprocessing (SMP) systems imposed on big data workloads.	To take advantage of the OSS Big Data Analytics platform with SLA's in place to ensure business continuity.	To create and manage an end-to-end big data/data science project using one platform.	To scale efficiently with Apache Spark clusters within a one stop shop analytical platform to meet your needs.

The different data services on Azure that have implementations of Apache Spark as mentioned in the table are explained below:

Apache Spark pools in Azure Synapse Analytics:

Apache Spark pools in Azure Synapse Analytics has Apache Spark capabilities embedded. For organizations that don't have existing Apache Spark implementations yet, Apache Spark pools in Azure Synapse Analytics provide the functionality to spin up an Apache Spark cluster to meet data engineering needs without the overhead of the other Apache Spark platforms.

Data engineers, data scientist, data platform experts, and data analyst can come together within Azure Synapse Analytics where the Apache Spark cluster is running to quickly collaborate on various analytical solutions.

Apache Spark for Azure Synapse:

Apache Spark is an open-source, memory-optimized system for managing big data workloads, It is used when you require a Spark engine for big data processing or data science, and you don't mind that there is no service level agreement provided to keep the services running.

Usually it is of interest of open-source professionals. The reason for Apache Spark is to overcome the limitations of what was known as SMP systems for big data workloads.

Azure HDInsight (HDI):

HDI is an implementation by Microsoft of open-source Apache Spark, managed on the Azure Platform. You can use HDI for an Apache Spark environment when you are aware of the benefits of Apache Spark in its OSS form, but you want a Service Level Agreement (SLA).

This implementation is usually of interest to open-source professionals who required an SLA, and data platform experts experienced with Microsoft products and services.

Azure Databricks:

Azure Databricks is a managed Apache Spark-as-a-Service propriety solution that provides an end-to-end data engineering/data science platform. Azure Databricks is of interest for many data engineers and data scientists working on big data projects today. It provides the platform in which you can create and manage the big data/data science projects all on one platform.

These services are not mutually exclusive. It is common to find customers who use a combination of these technologies working together.