

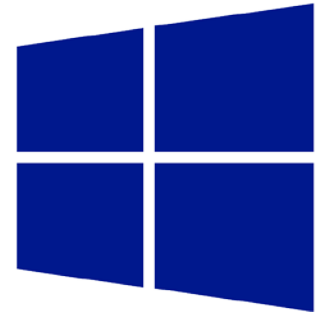
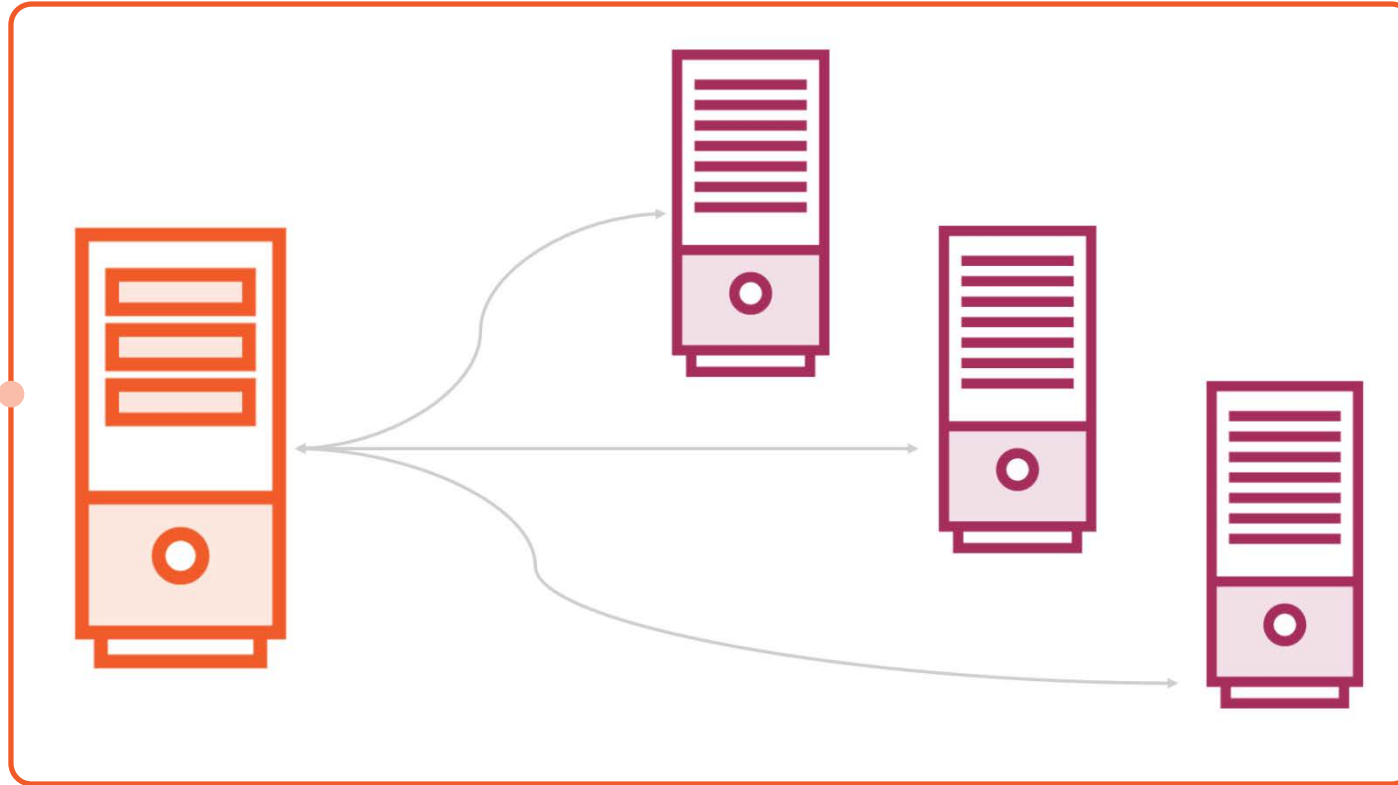
Working with Hadoop in .NET

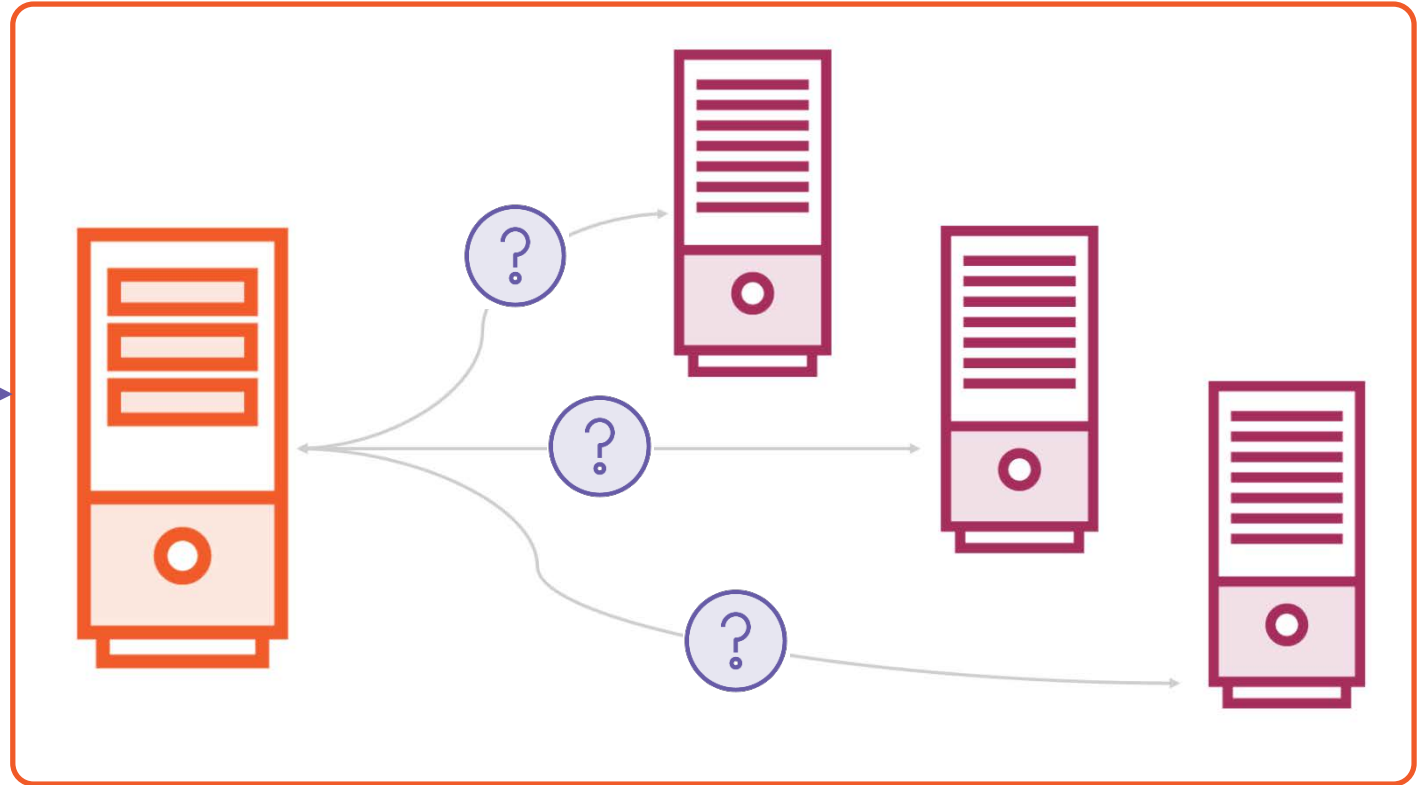
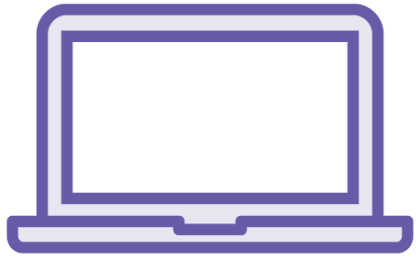


Elton Stoneman

SOFTWARE ARCHITECT

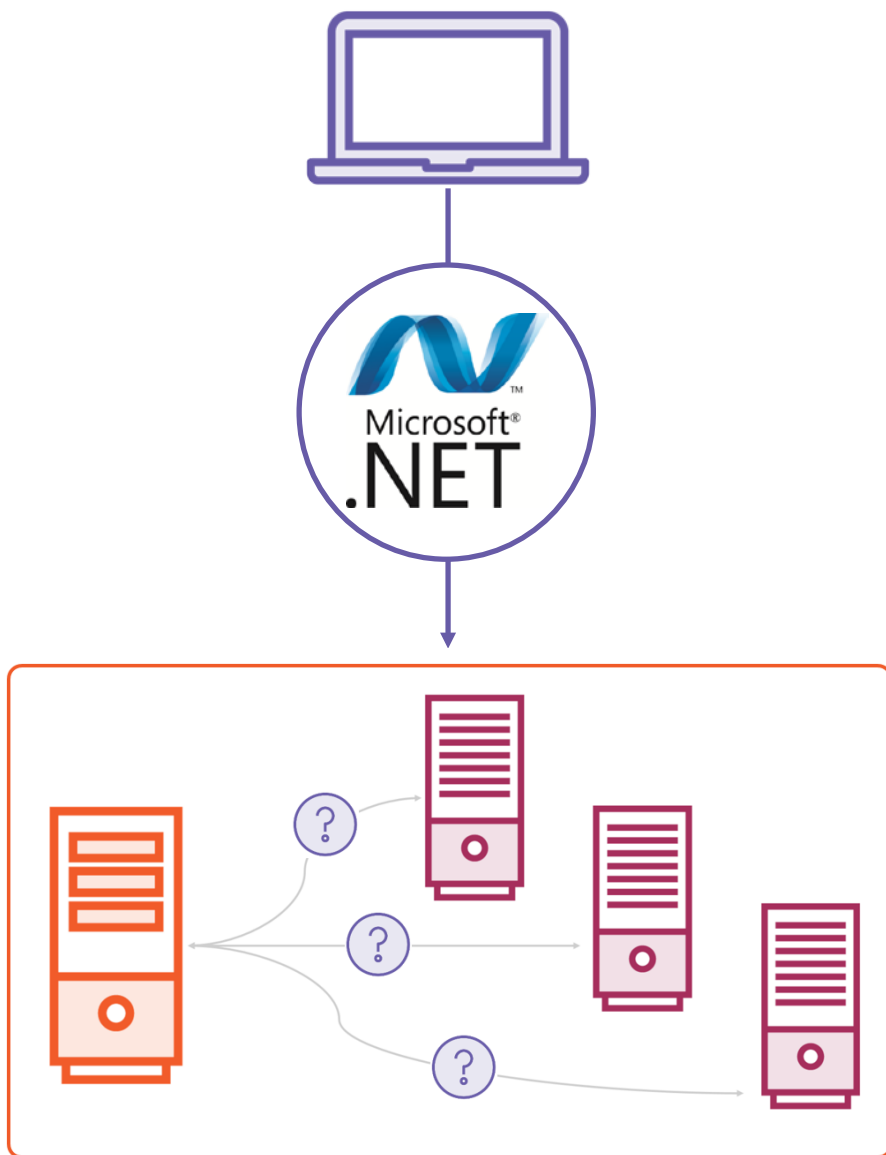
@EltonStoneman blog.sixeyed.com





Hadoop Streaming

Invoke external process



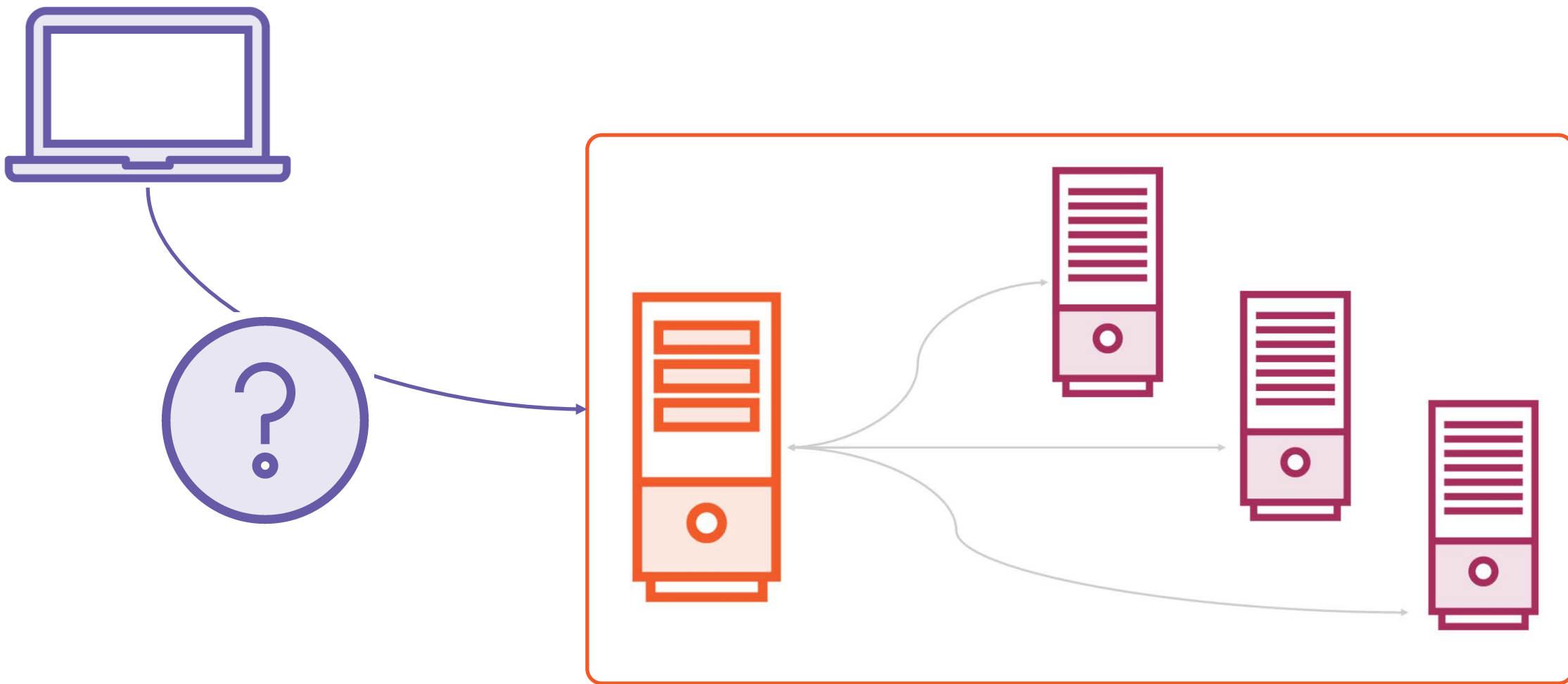
HADOOP STREAMING

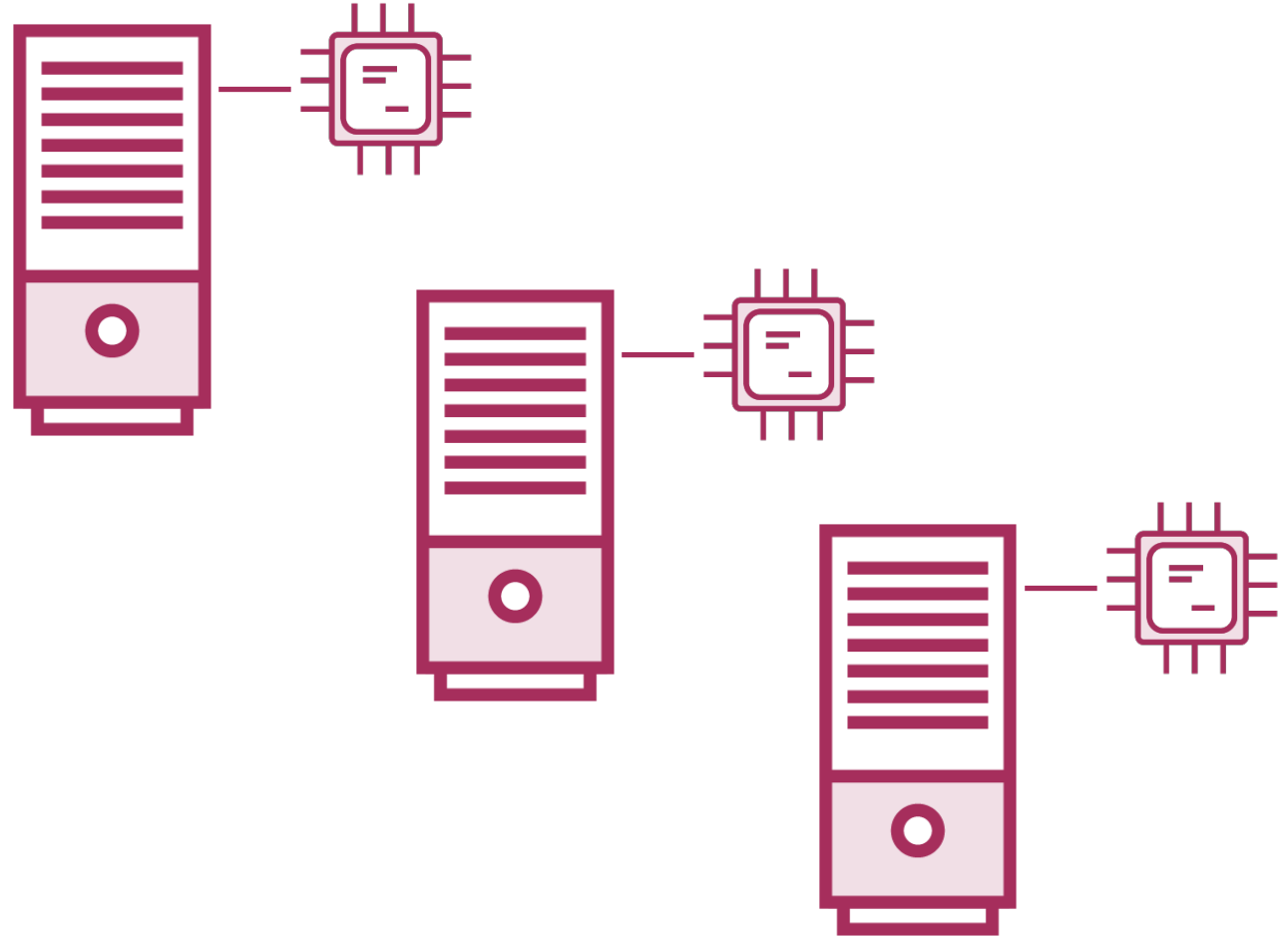
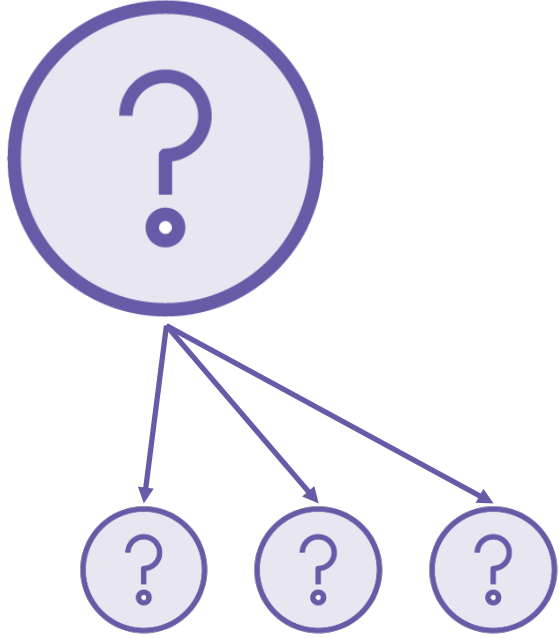
- .NET Framework console apps
- Microsoft Hadoop SDK for .NET
- .NET Core console apps

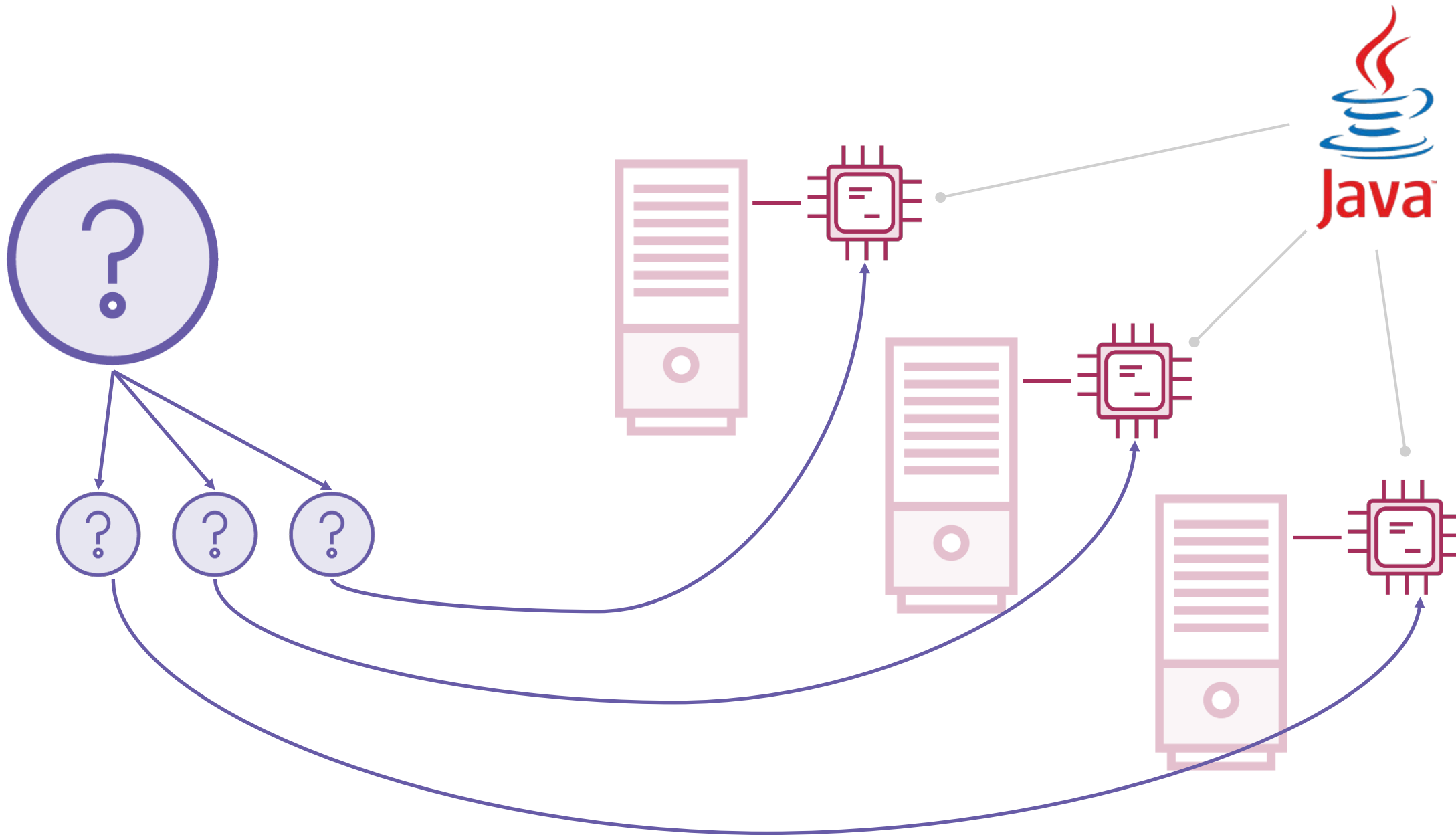


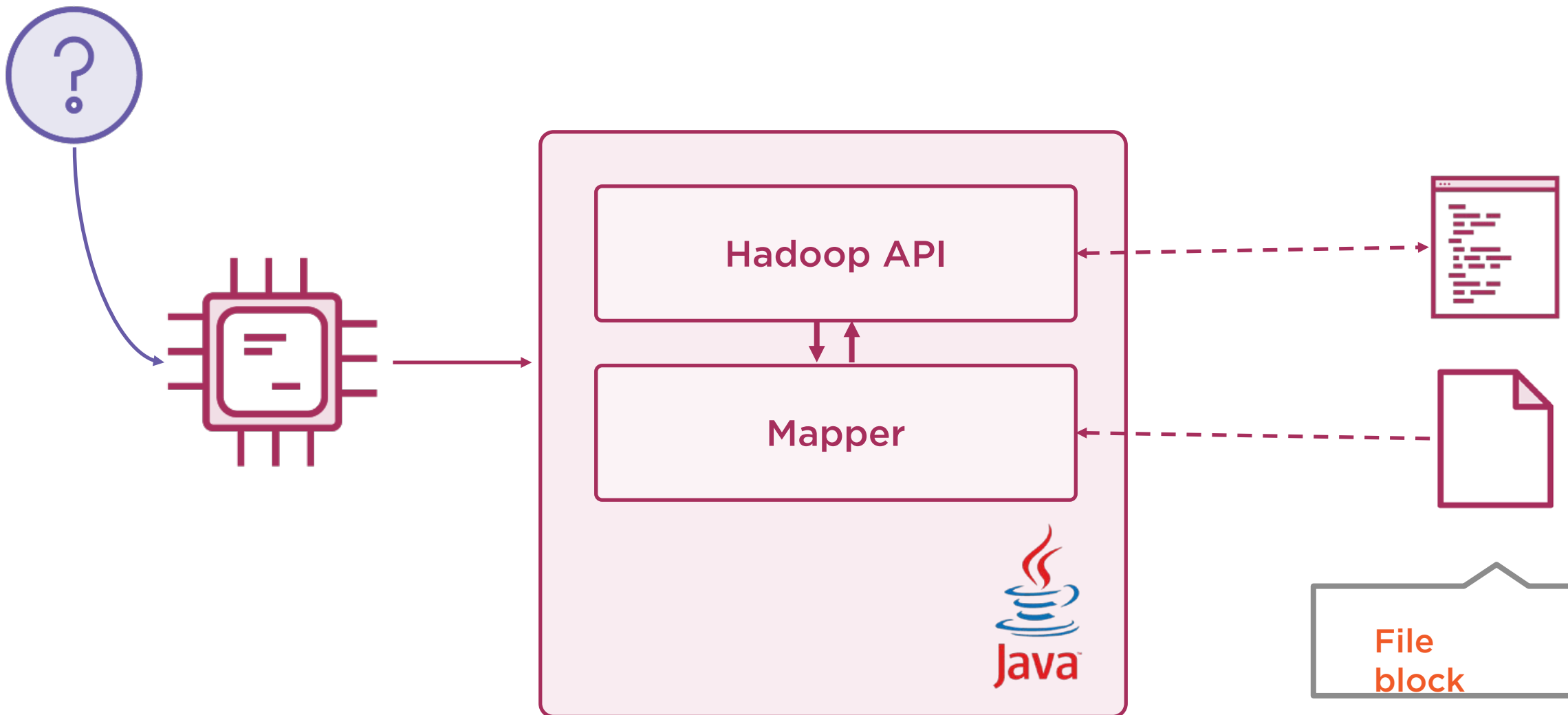
.NET MAPREDUCE POC

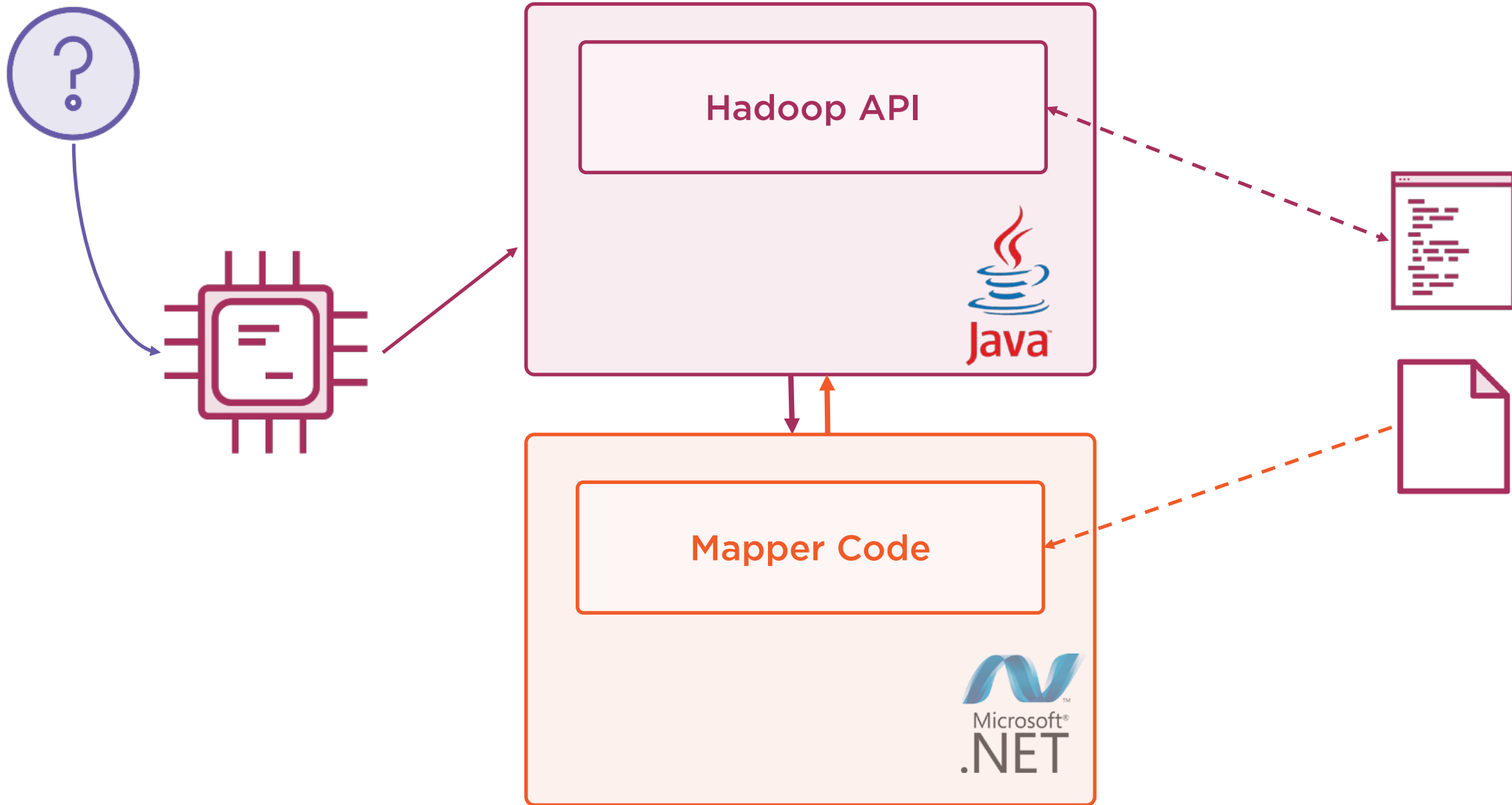
- How does .NET **communicate** with Hadoop?
- How do you ship **dependencies** for apps?
- Is the technology **supported**?
- Is the MapReduce job **portable**?

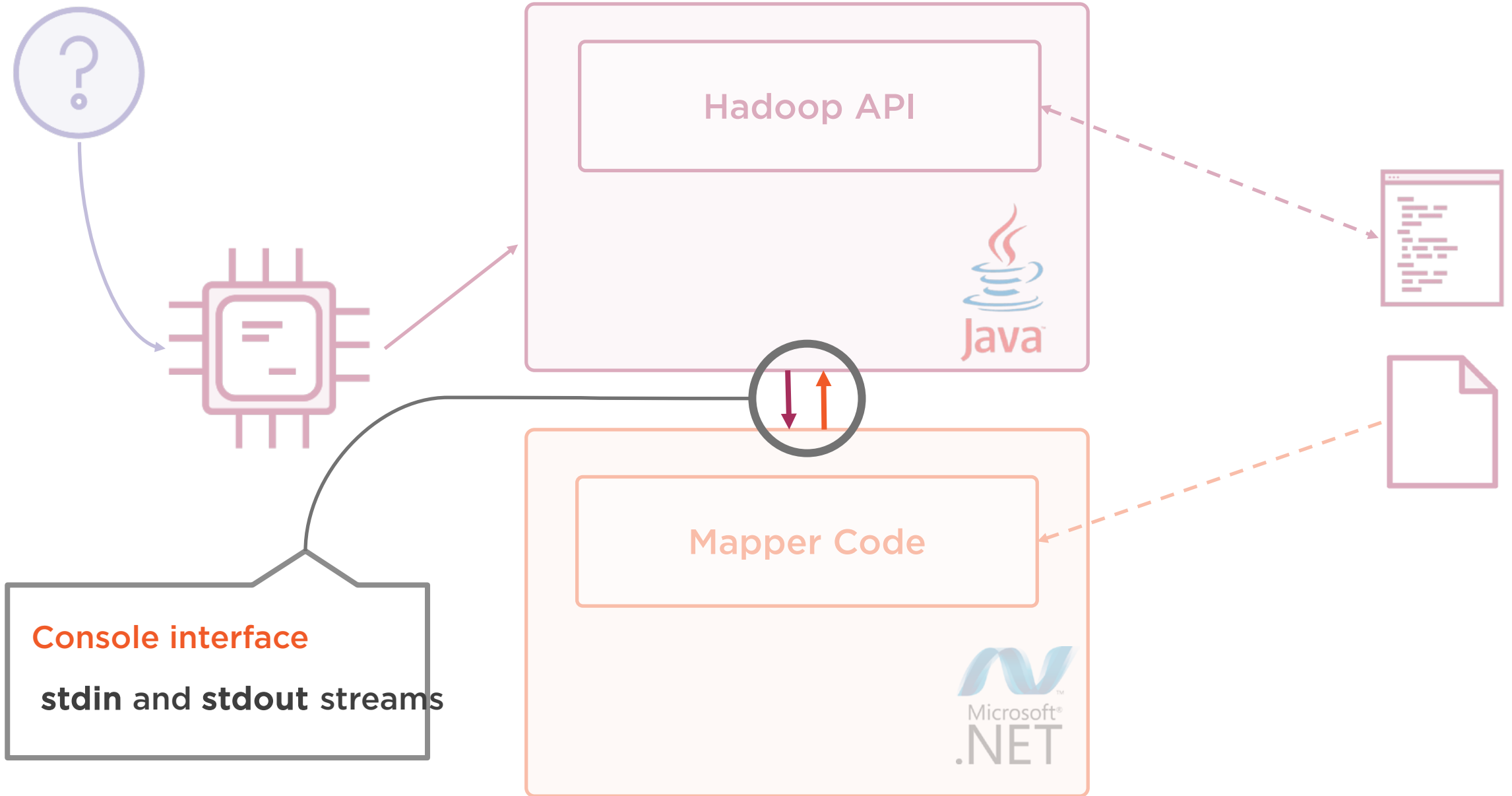




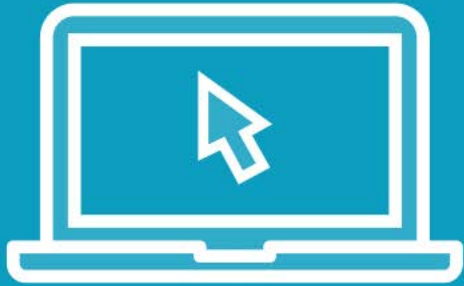






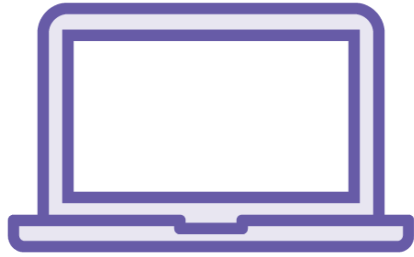


Demo

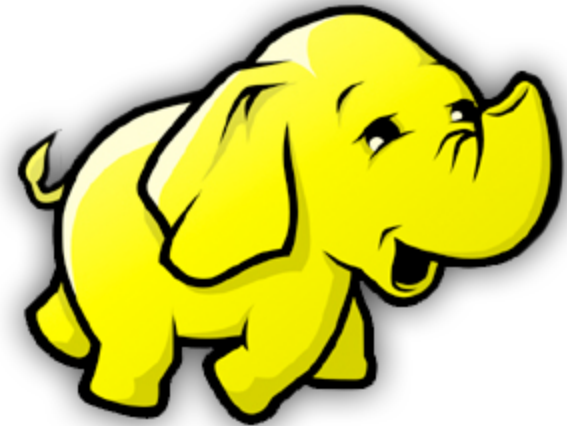


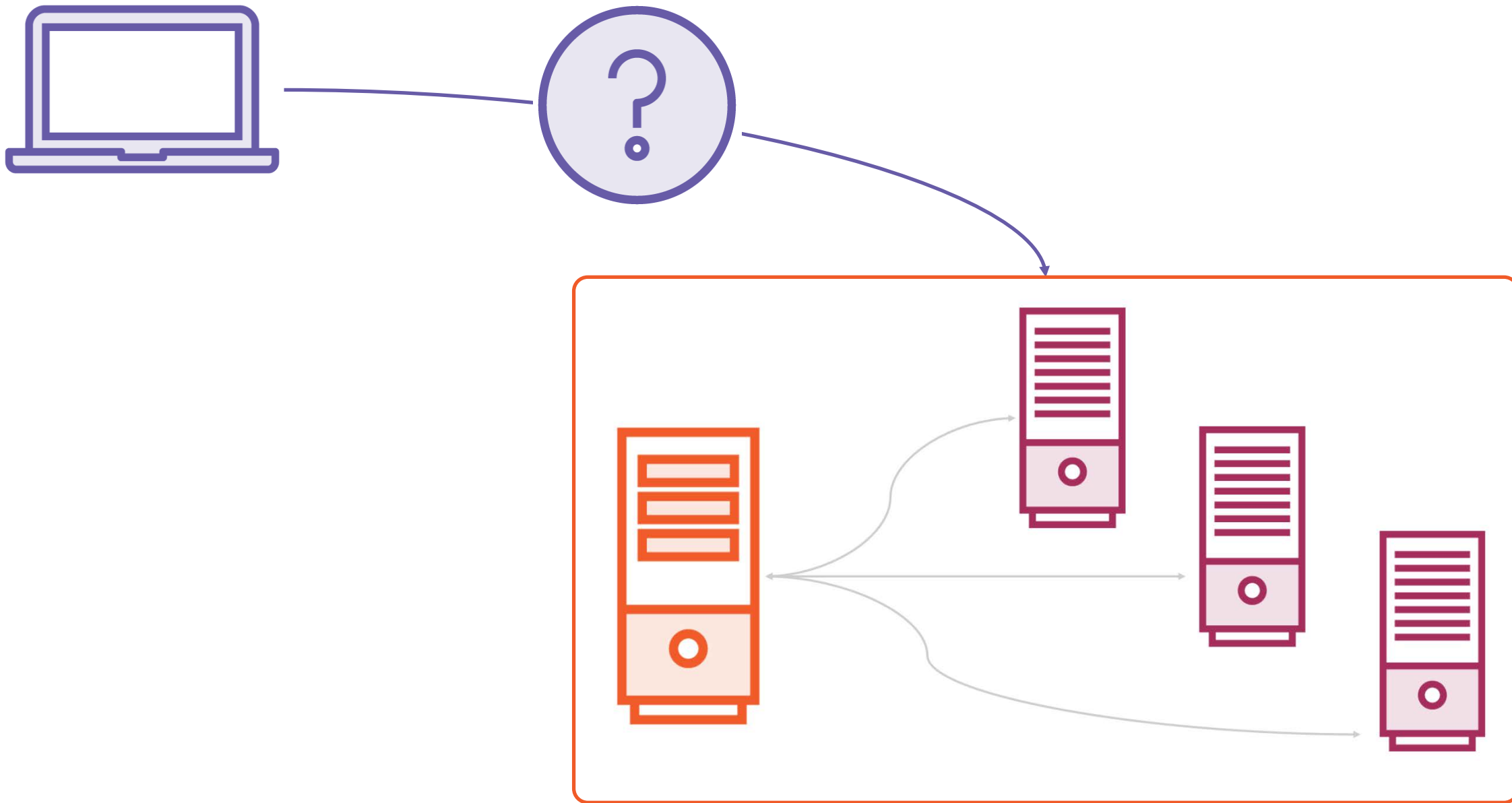
Hadoop Streaming with .NET

- .NET Framework
- Mapper console app
- Reducer console app

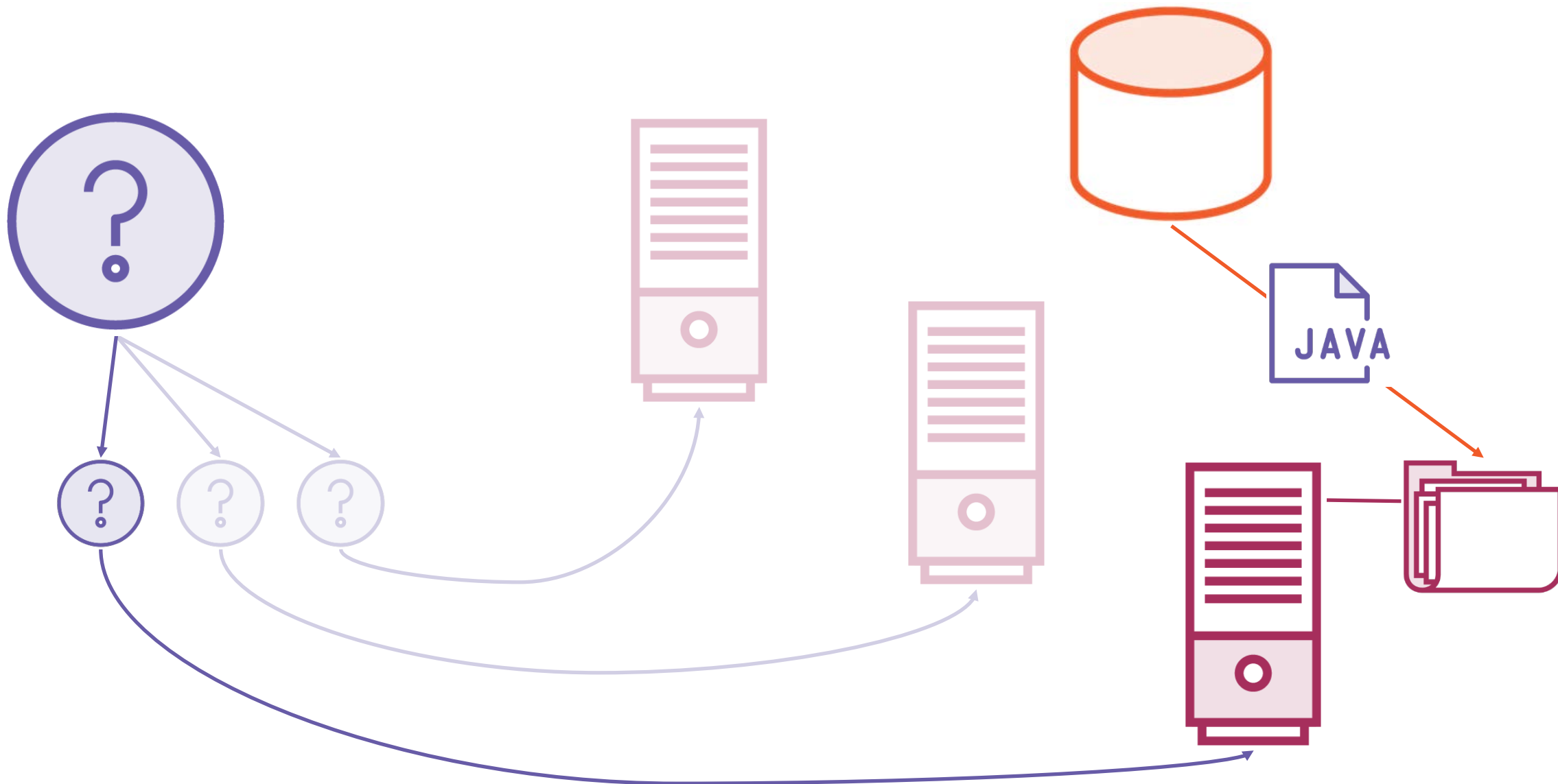


```
hadoop jar /my.jar
```

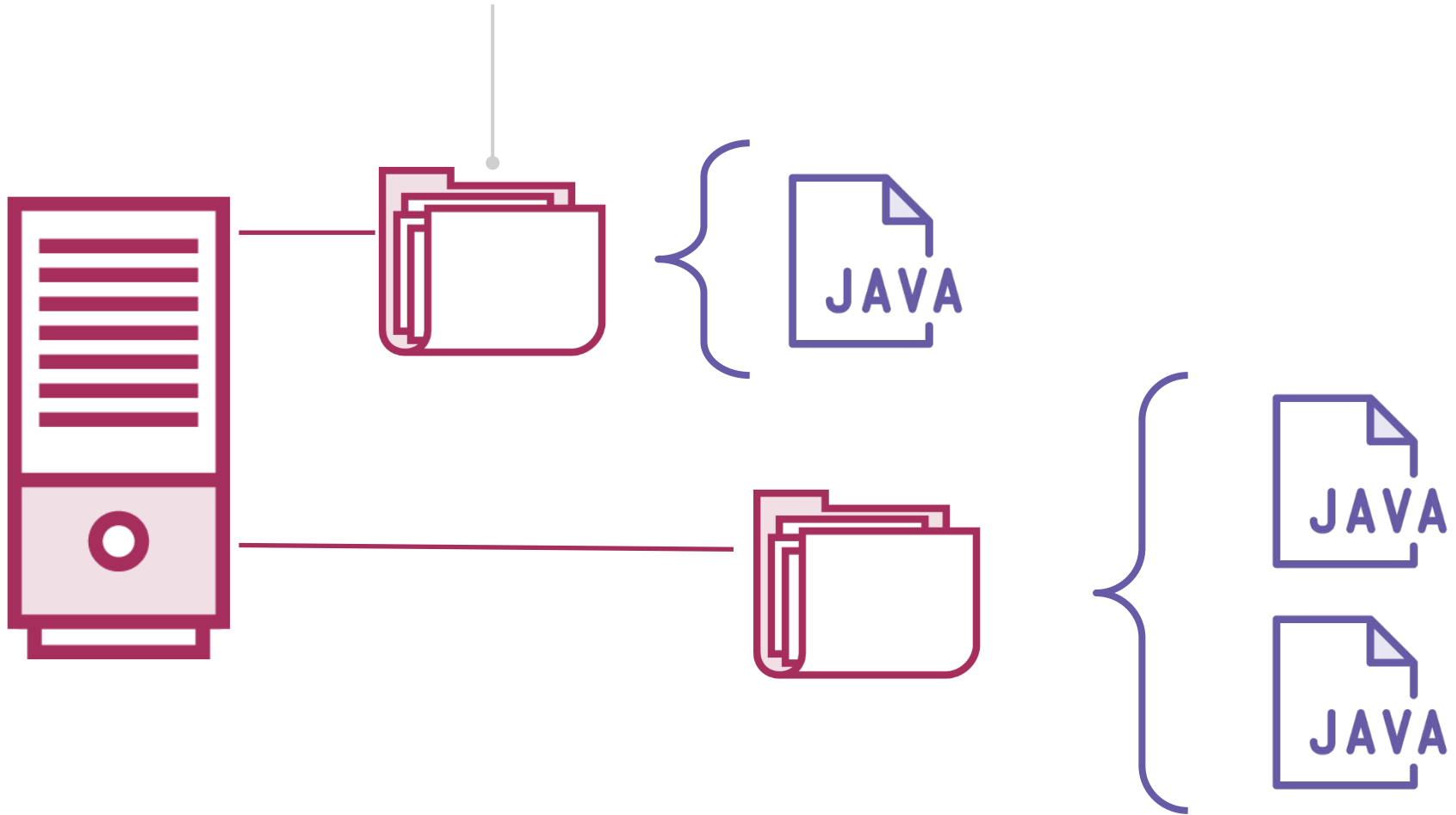


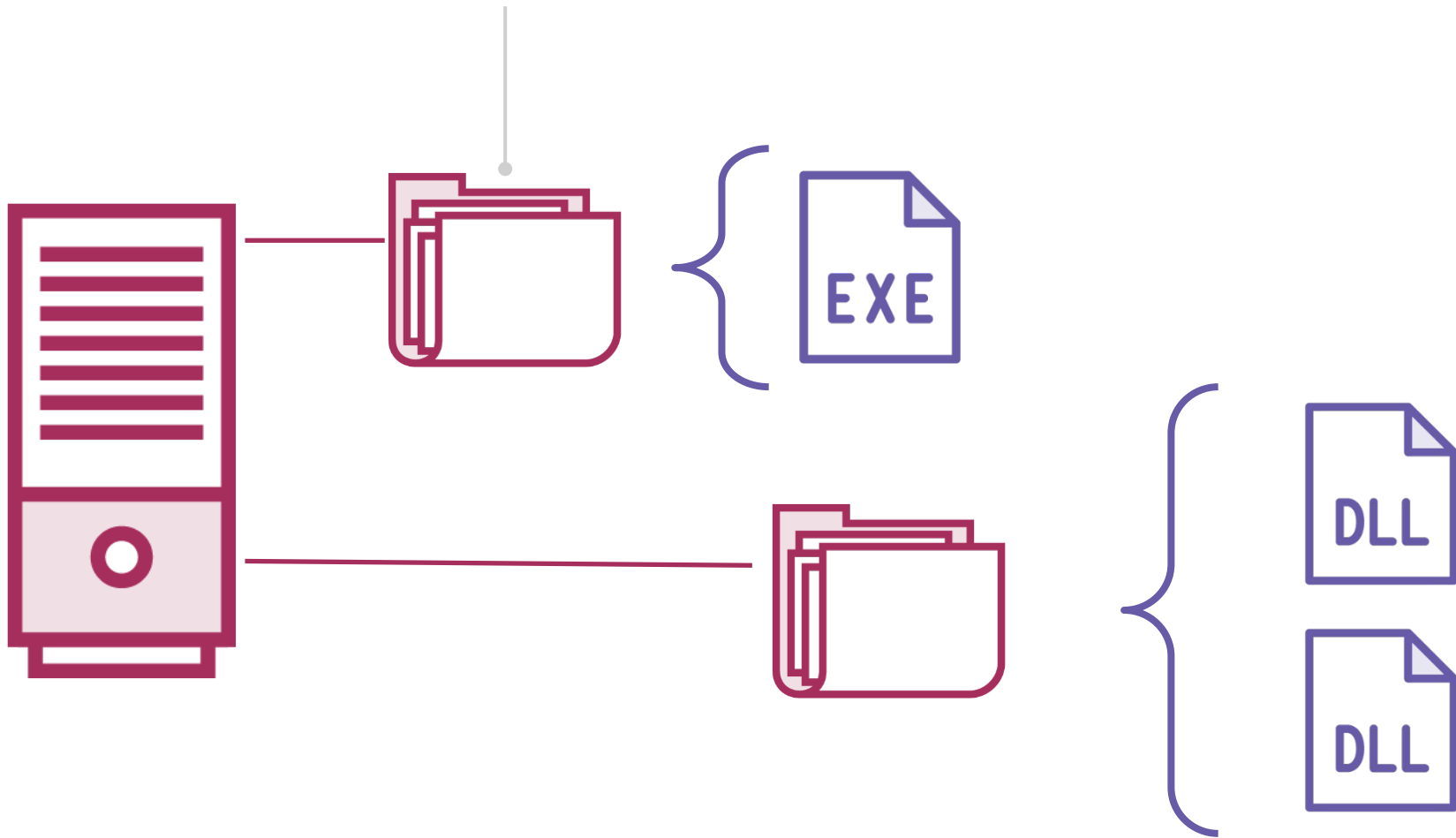


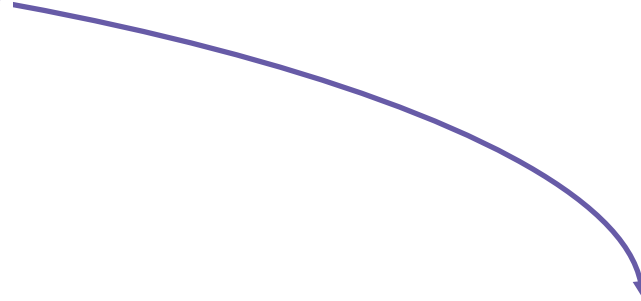
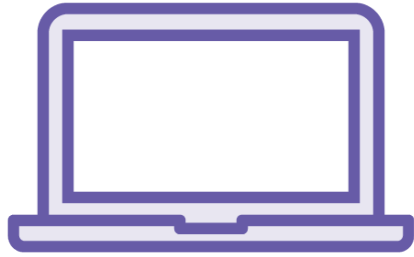




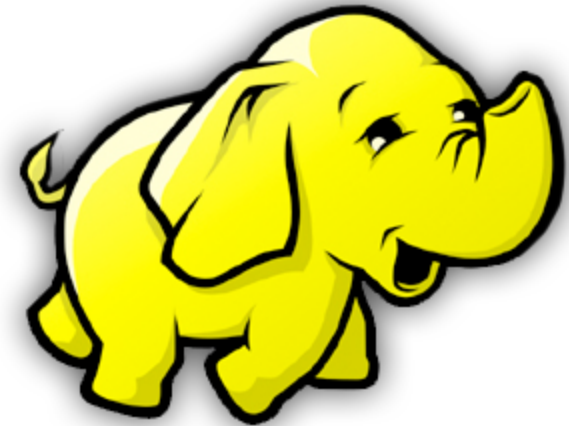








`hadoop jar streaming.jar`



hadoop jar

streaming-x.y.z.jar

-files ...

-mapper ...

-reducer ...

-input ...

-output ...

◀ **Submit a job**

◀ **Streaming JAR**

◀ **Files to ship**

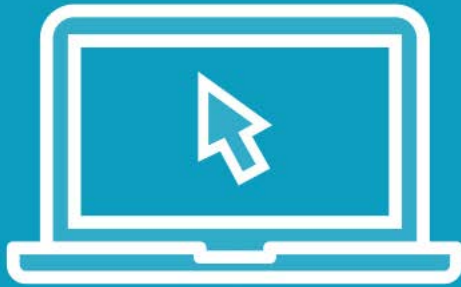
◀ **Mapper command**

◀ **Reducer command**

◀ **Input location**

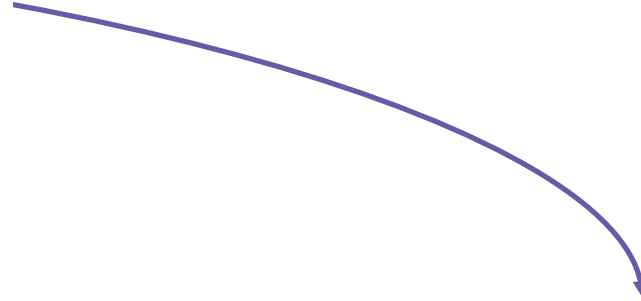
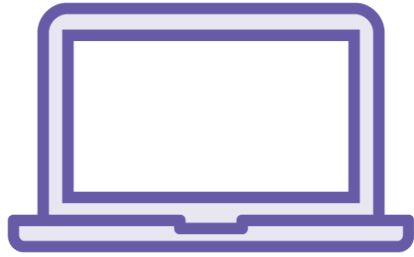
◀ **Output location**

Demo

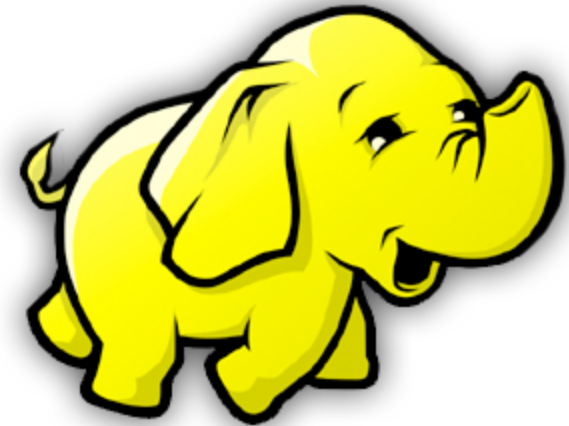


Hadoop Streaming

- Submit job
- Ship dependencies
- .NET mapper and reducer



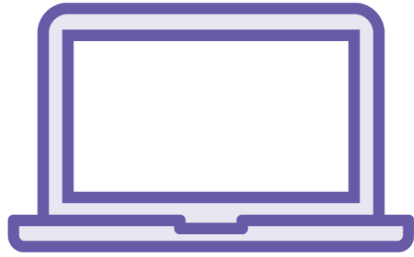
`hadoop jar streaming.jar`



.NET Framework

Mapper console app

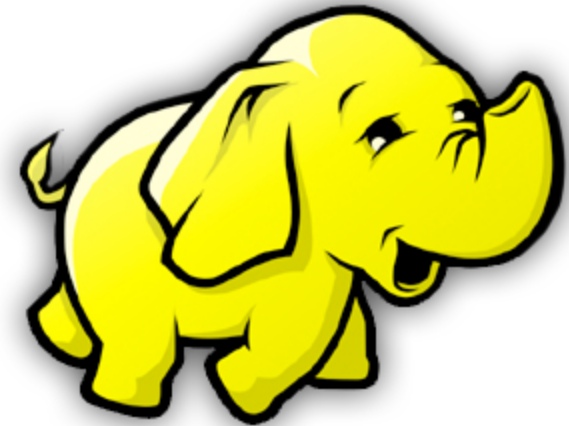
Reducer console app

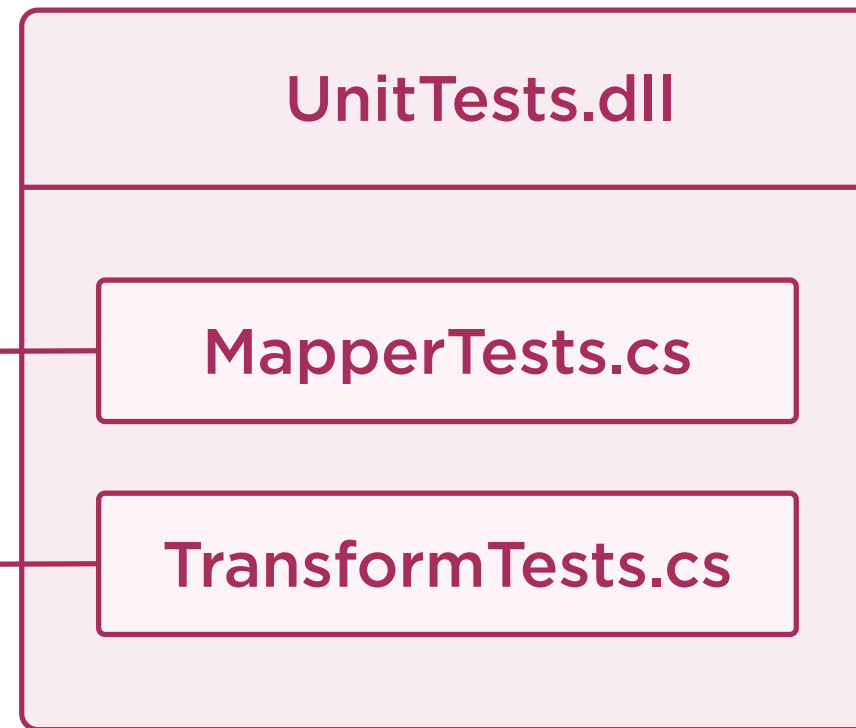
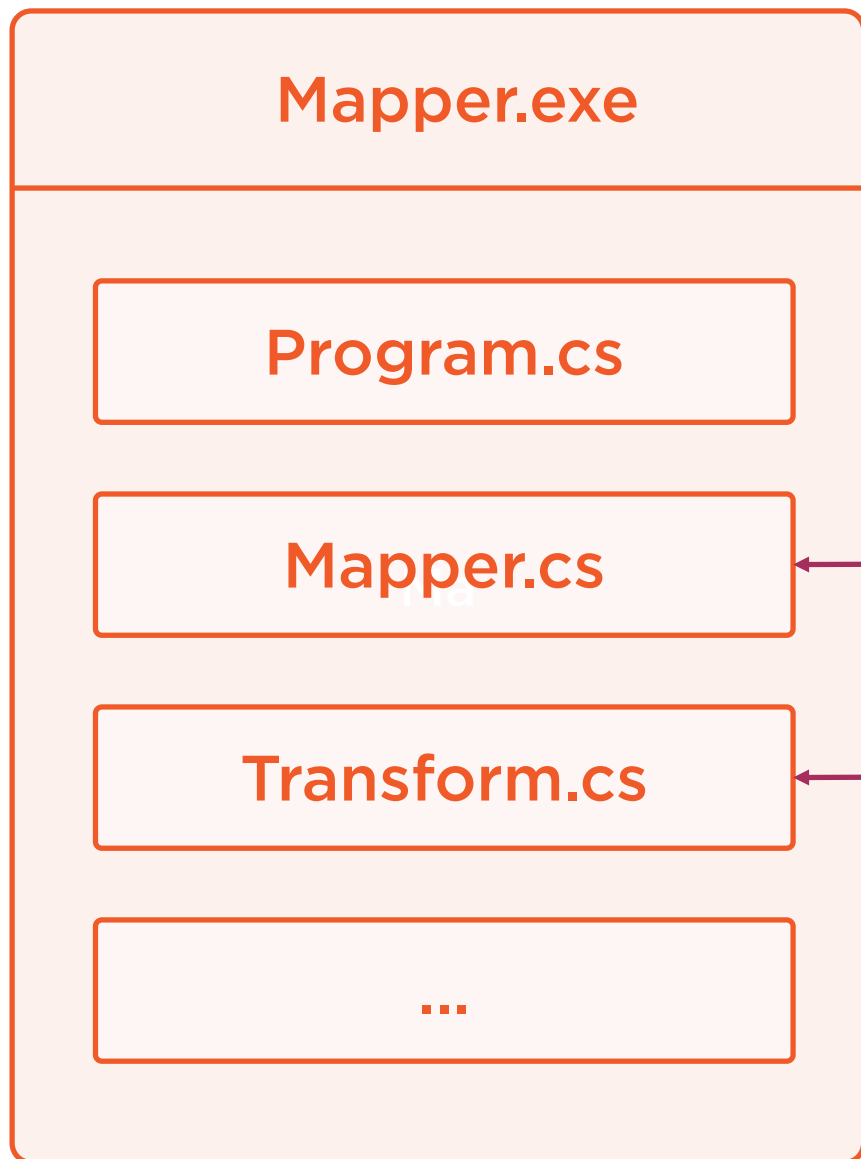


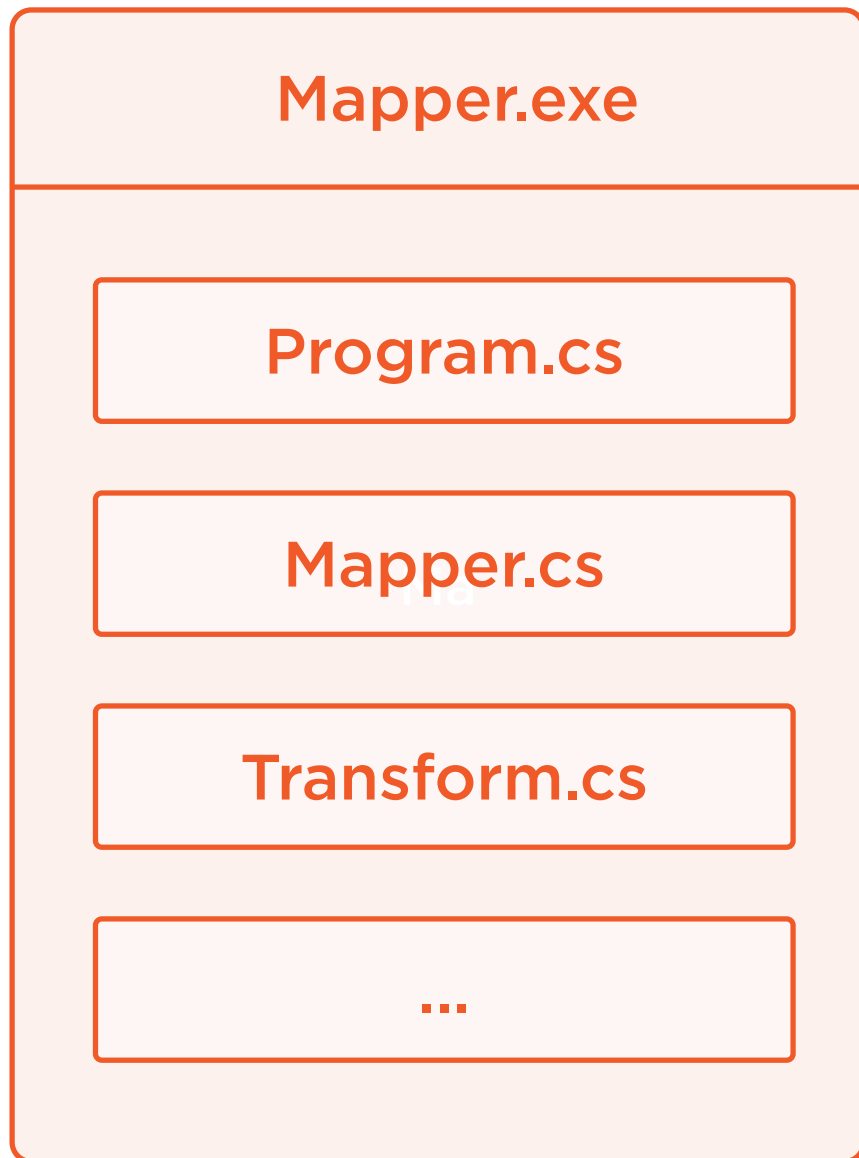
Ship dependencies

Executables & binaries

Supporting files

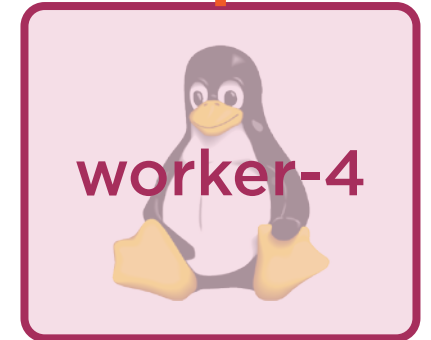
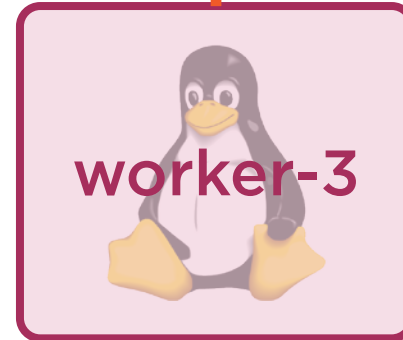
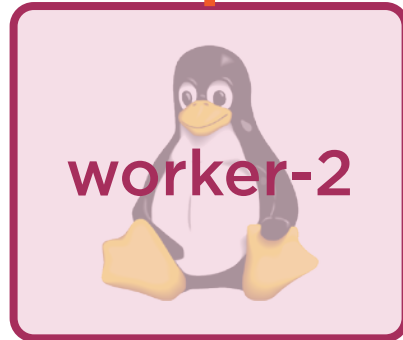
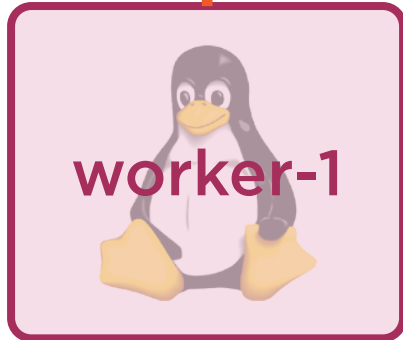
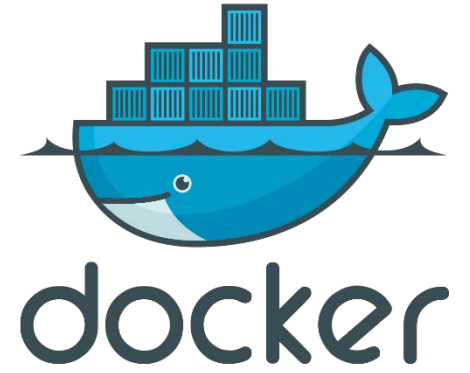


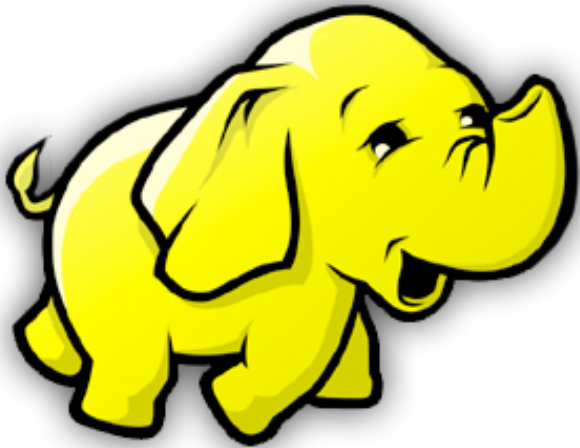




```
C:\dotnet> mapper.exe
```







stdin



stdout



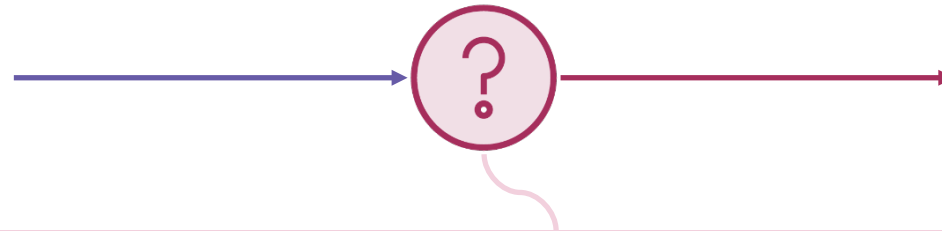
stderr

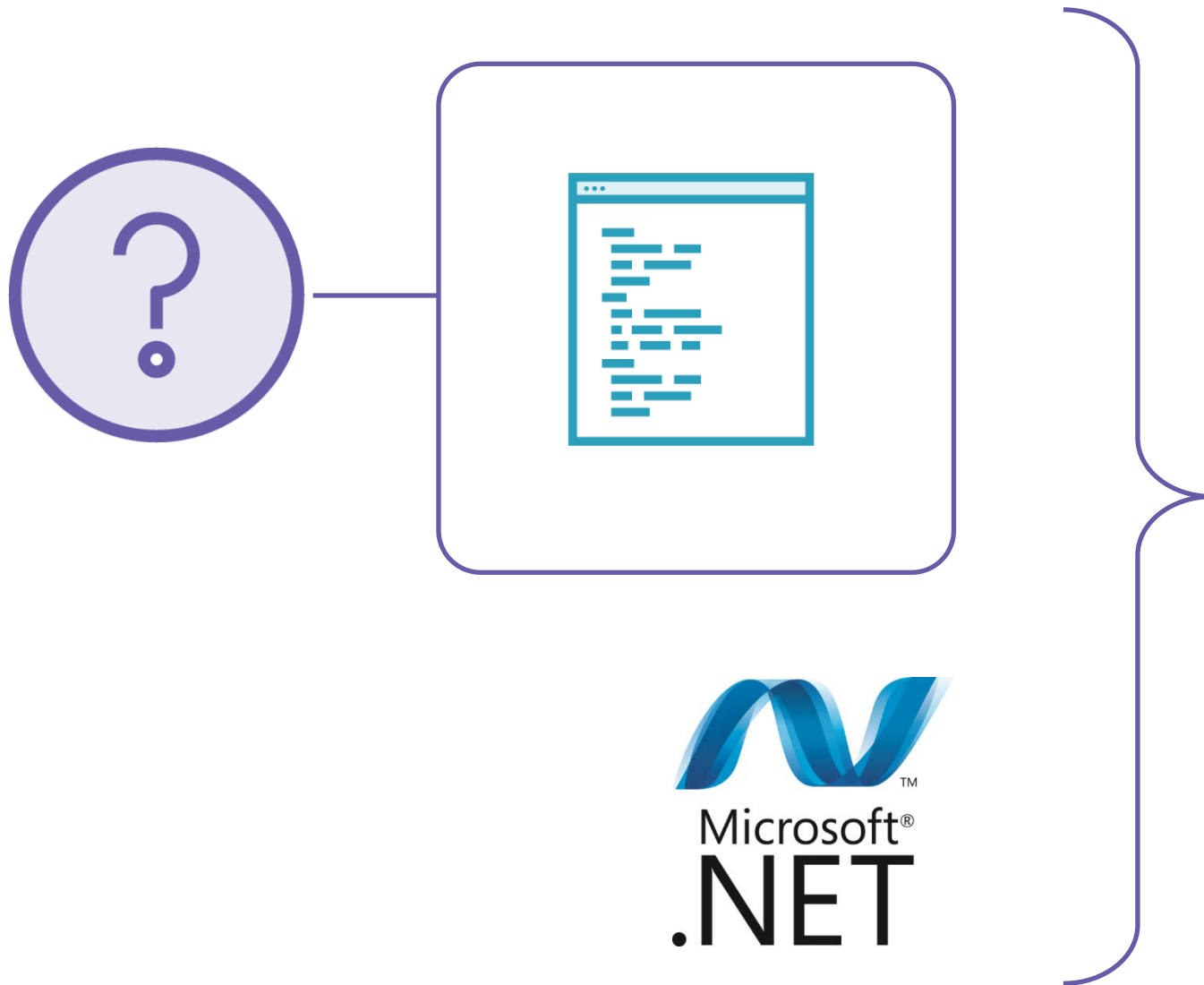


```
C:\dotnet> mapper.exe
```

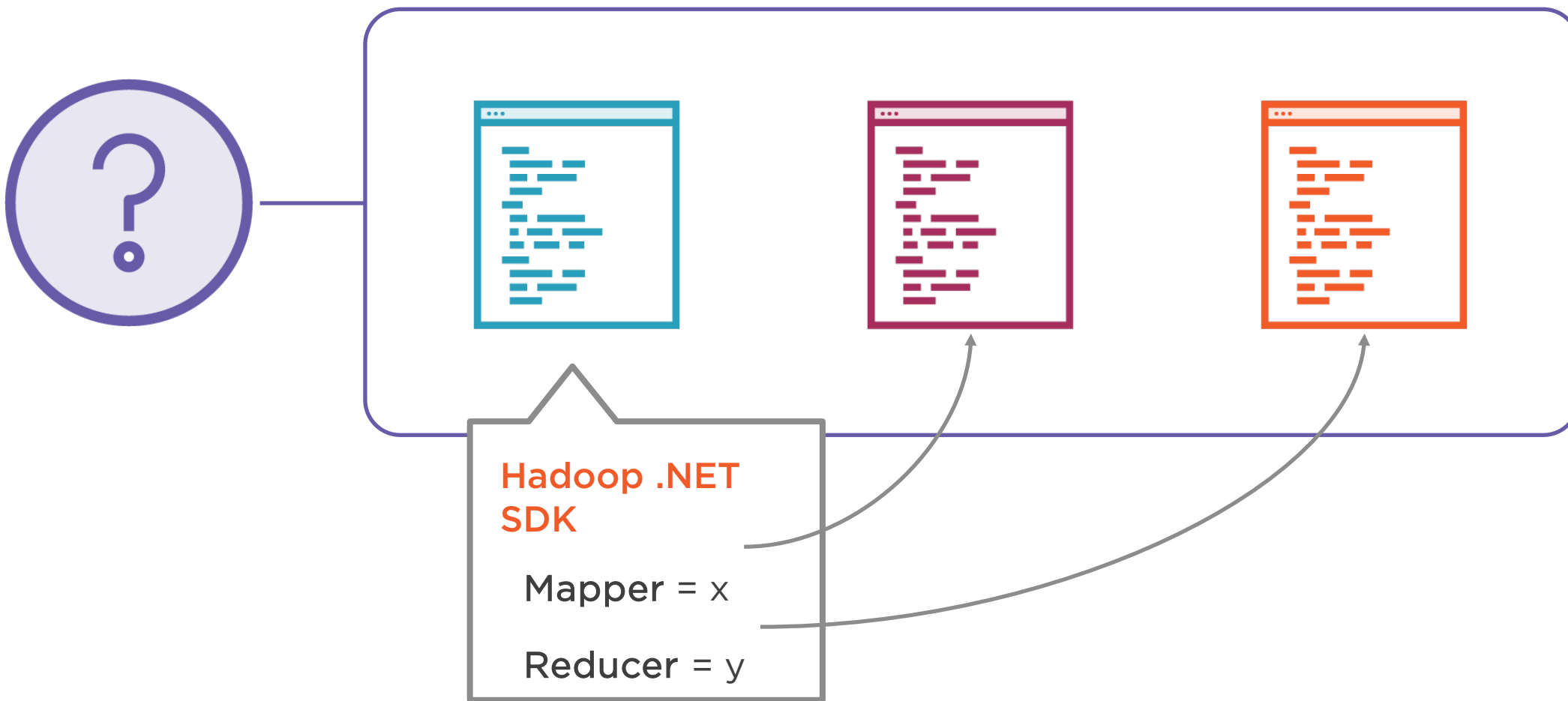
Mapper

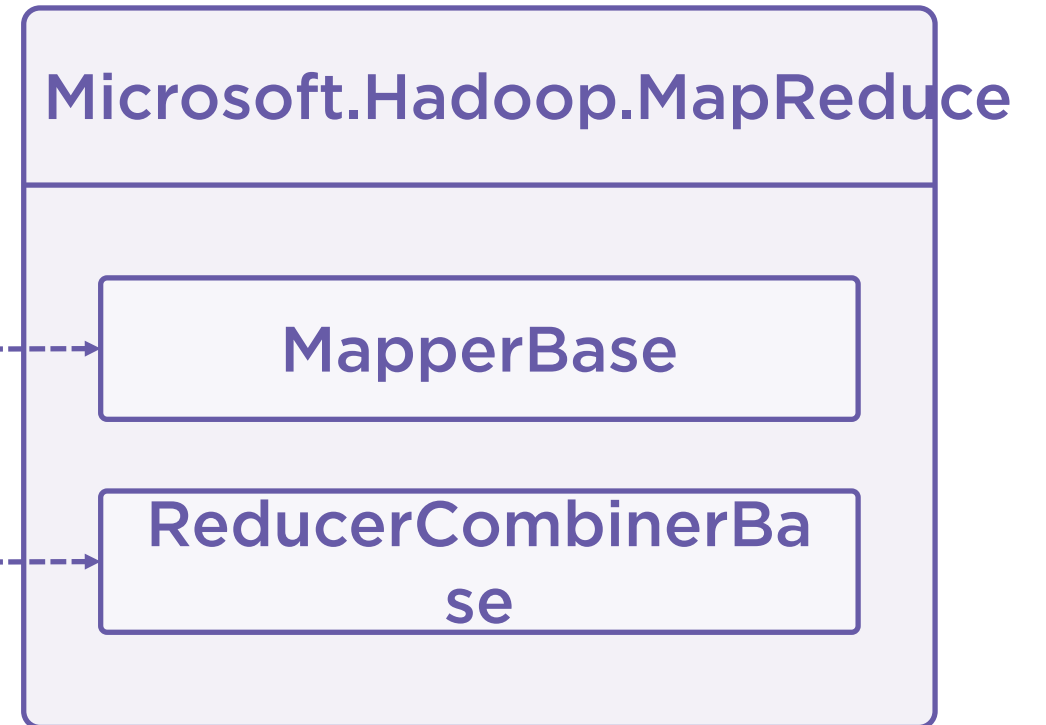
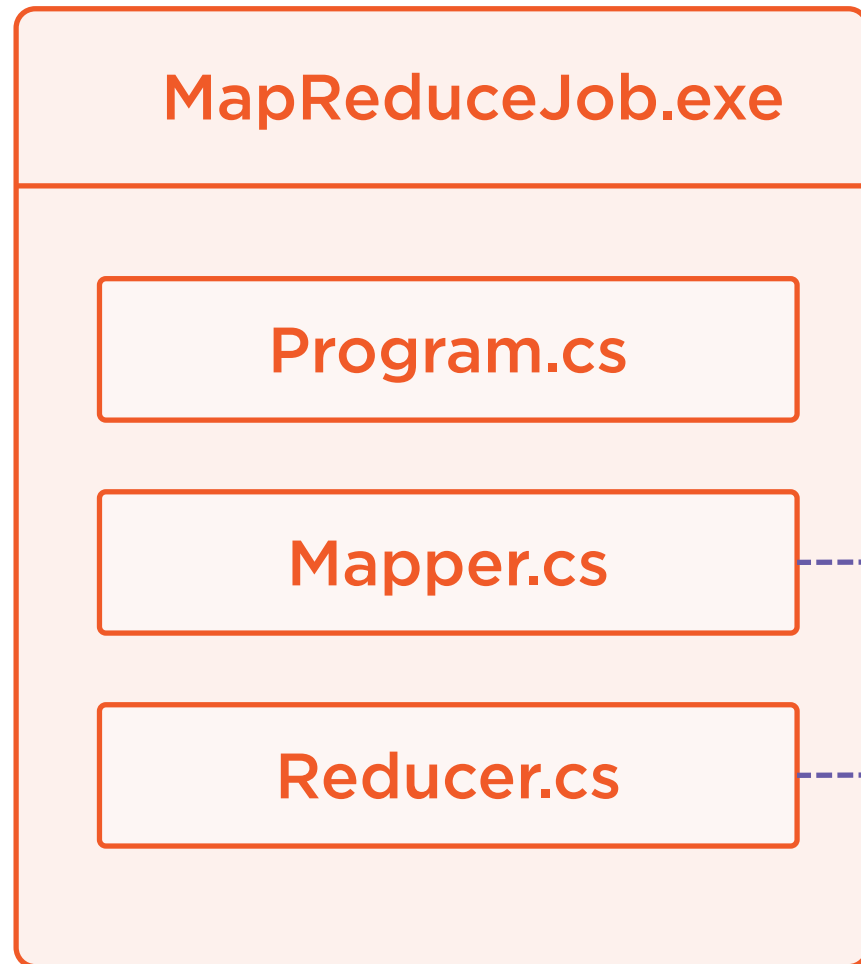
Data extraction code



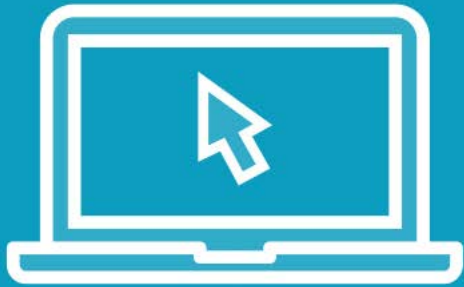


**Microsoft Hadoop
.NET SDK**



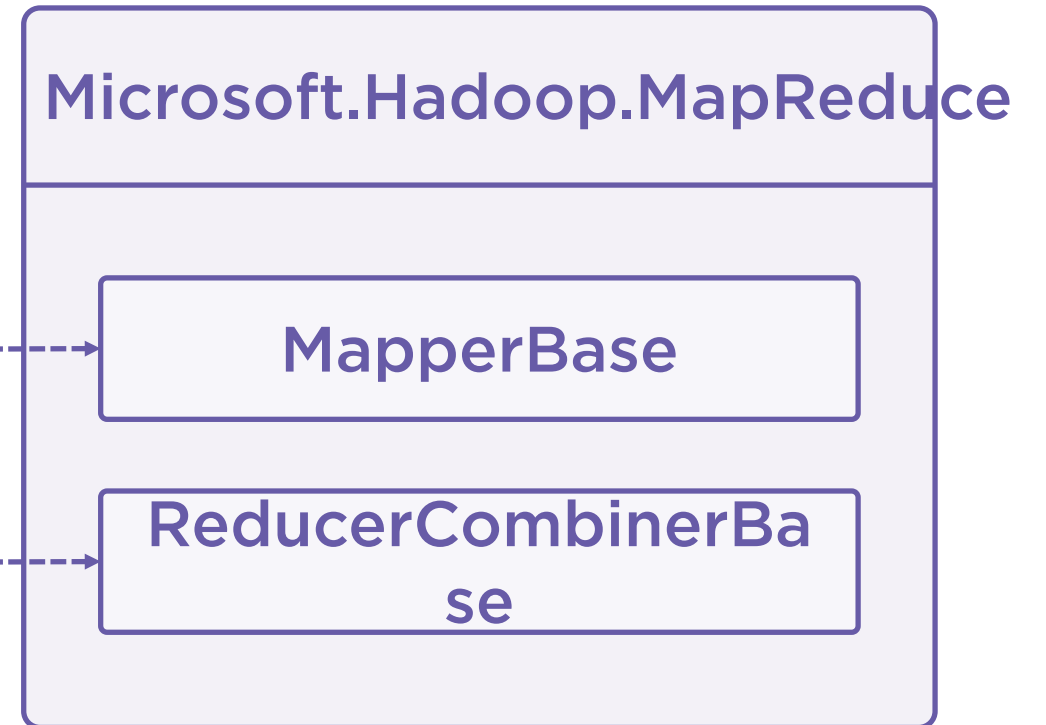
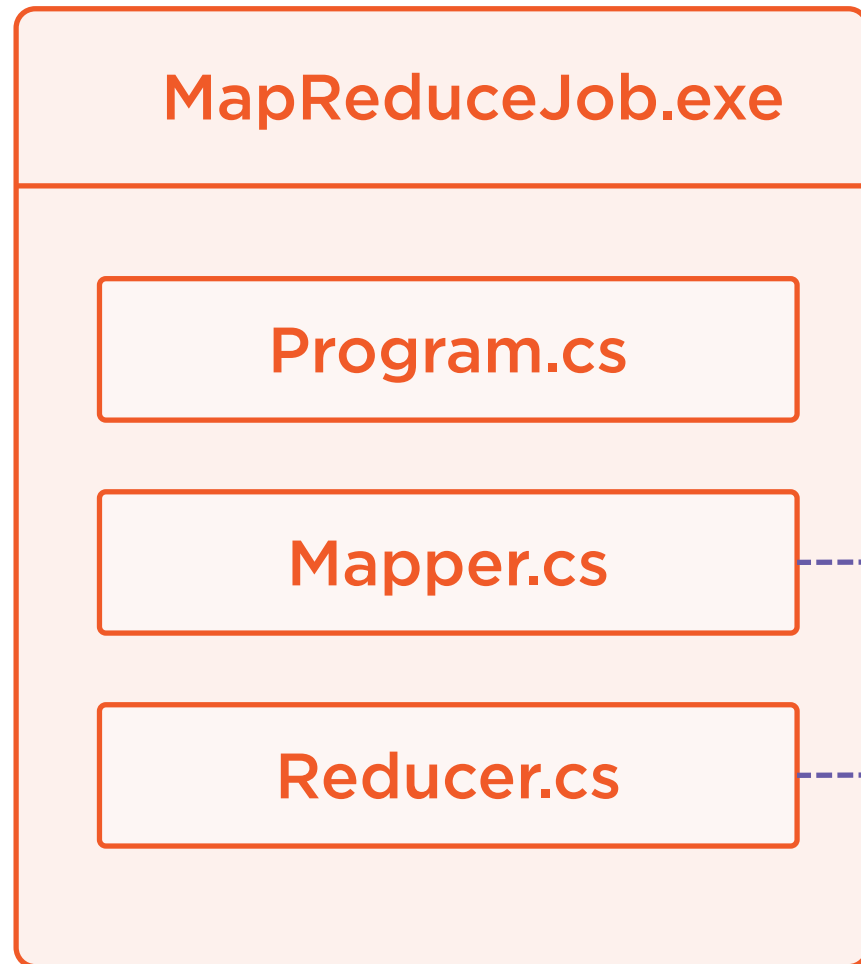


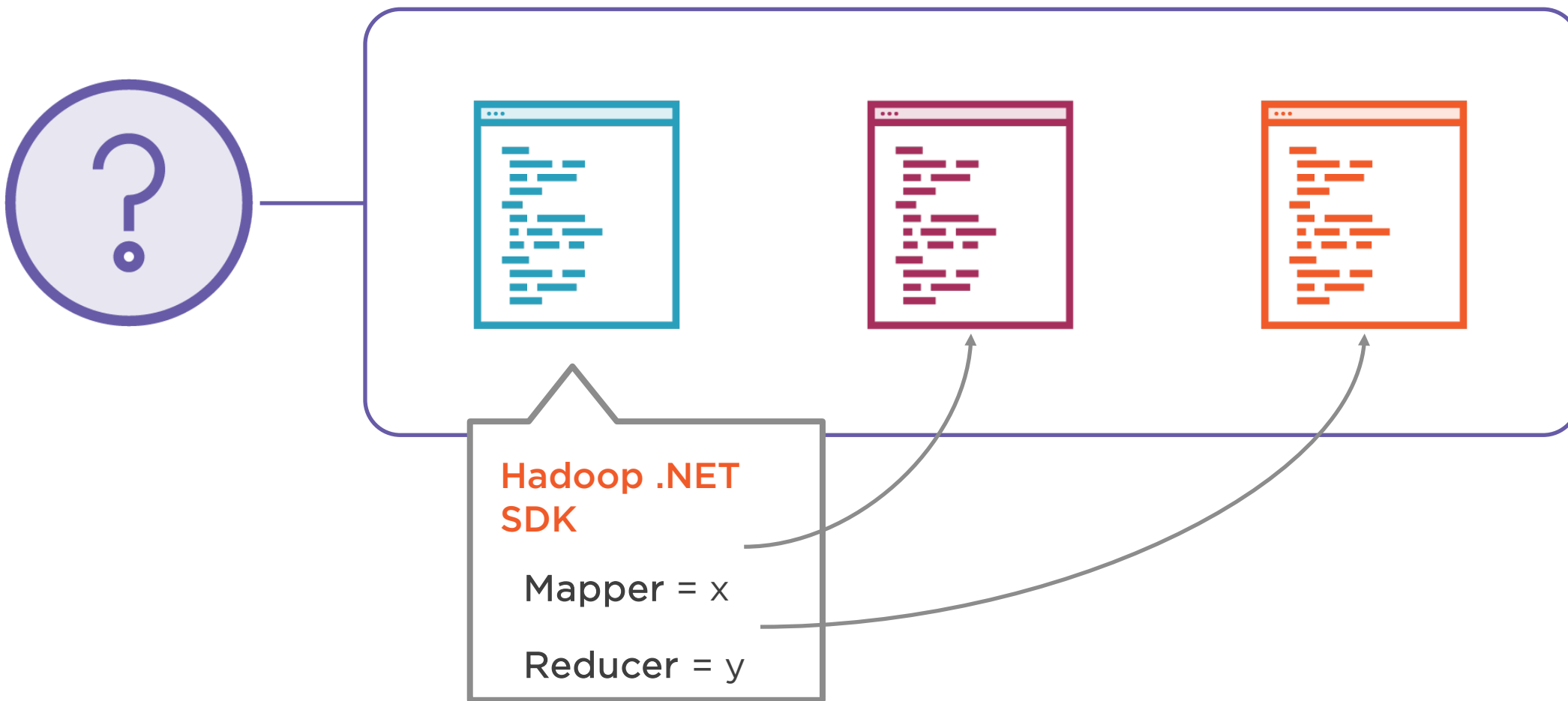
Demo

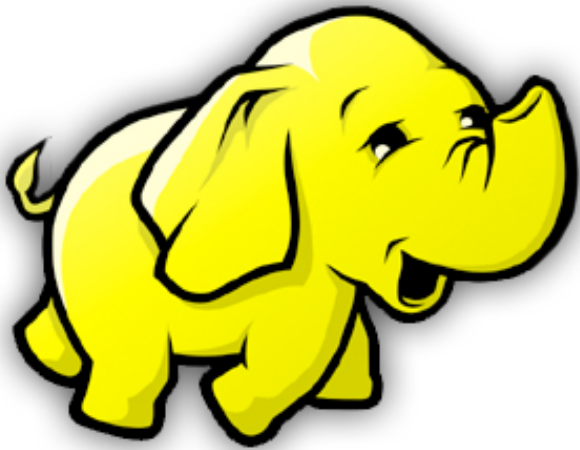


Microsoft Hadoop SDK for .NET

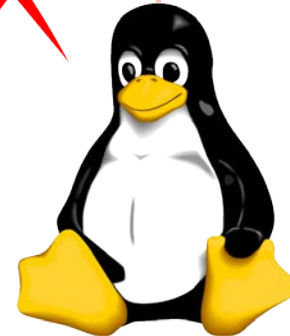
- MapReduce console app
- Mapper and reducer classes
- Submit to Hadoop





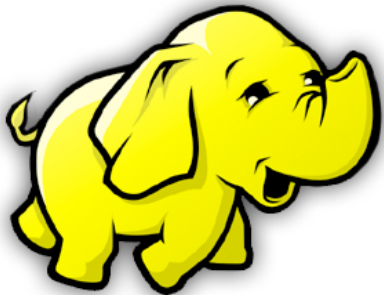


```
> Microsoft.Hadoop.MapDriver.exe
```





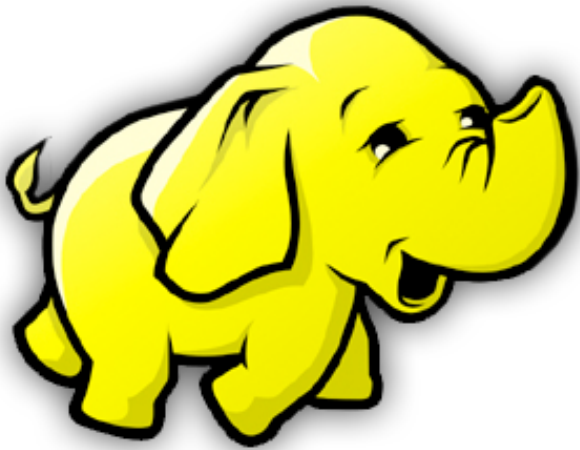
Microsoft®
.NET



MICROSOFT HADOOP SDK FOR .NET

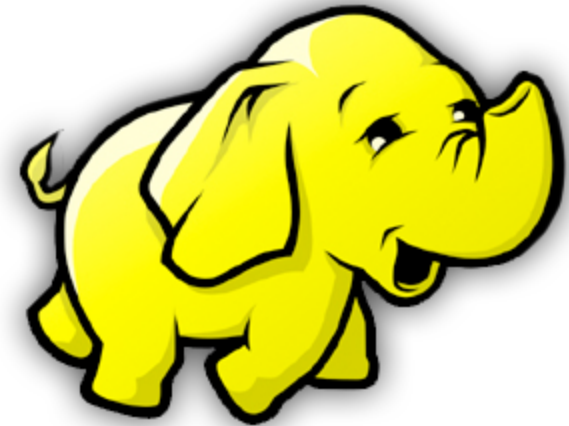
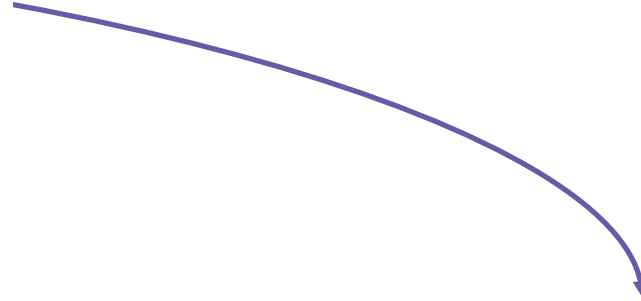
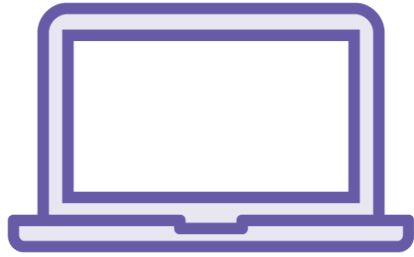
- CodePlex and NuGet
- Status: incubating
- Last release: 2015

<https://is.gd/EAOWin>



```
> Microsoft.Hadoop.MapDriver.exe
```



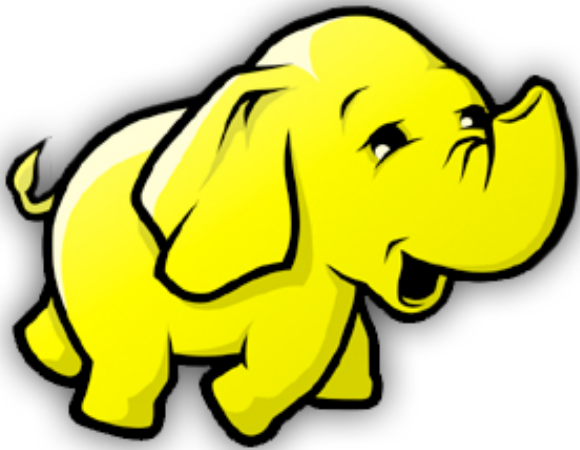


```
hadoop jar streaming.jar
```

.NET Core

Mapper console app

Reducer console app



```
> dotnet mapper.dll
```

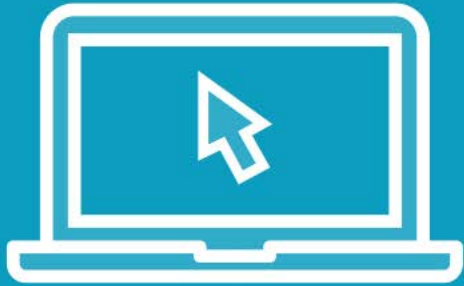




.NET CORE

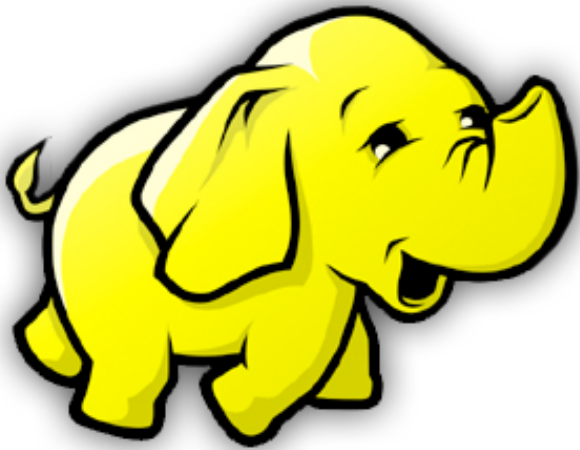
- Open source, cross platform .NET
- RTM in 2016
- Subset of full .NET
- DLLs not compatible

Demo



MapReduce with .NET Core

- Mapper console app
- Reducer console app
- Submit to Hadoop



```
> dotnet mapper.dll
```



.NET Framework

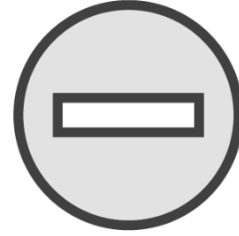
Console Apps



Full .NET

All of NuGet + your code

Long-term support



Not portable

Restricted streaming API

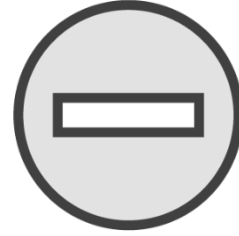
.NET Framework Hadoop SDK



Java-like API

Full .NET

All of NuGet



Windows only

Limited lifespan?

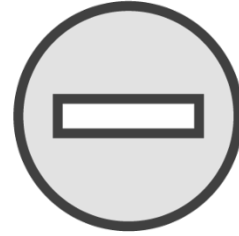
.NET Core Console Apps



**Portable across
platforms**

Common libraries

available



Limited ecosystem

Restricted streaming API



.NET Framework Console Apps

- Good on-premise
- Existing skills



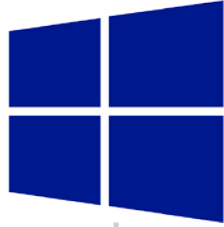
.NET Framework Hadoop SDK

- Risk of deprecation
- Too little benefit

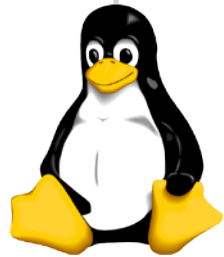


.NET Core Console Apps

- Most flexible
- Cheaper cloud

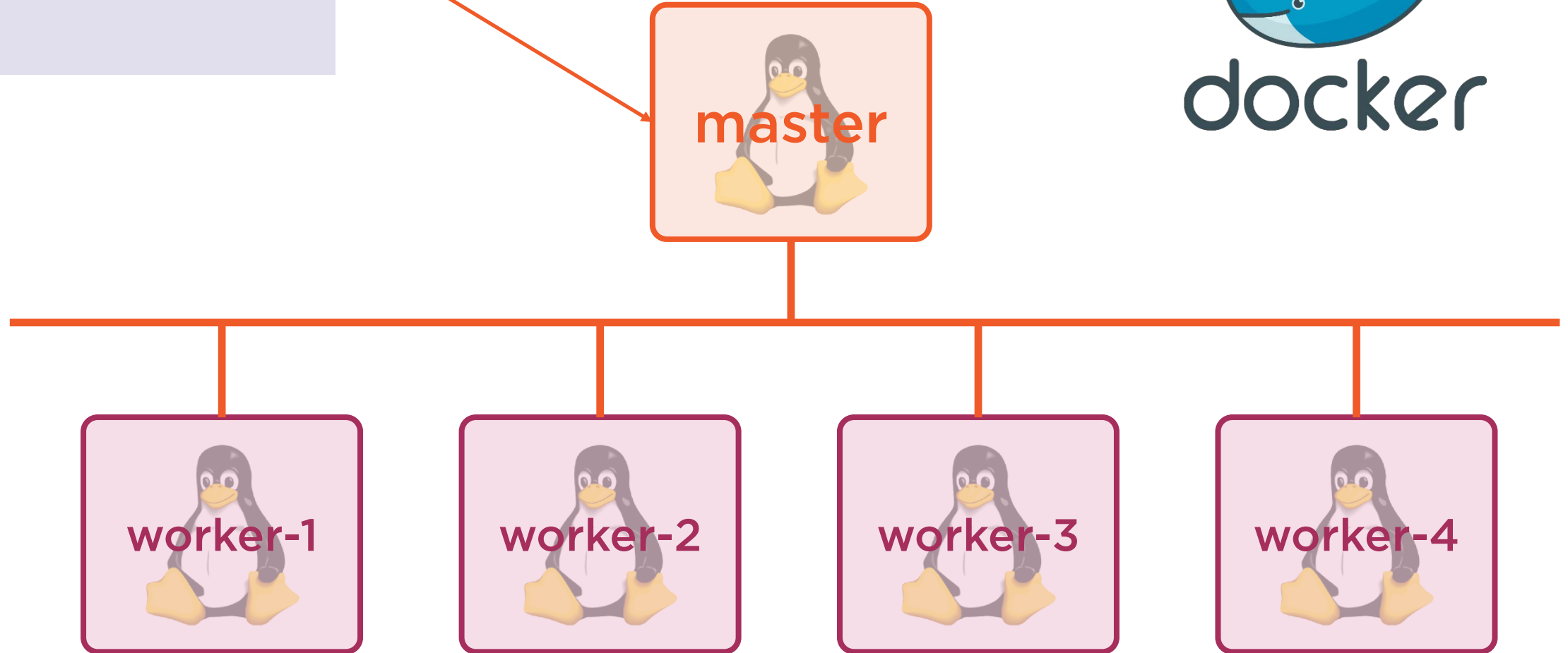
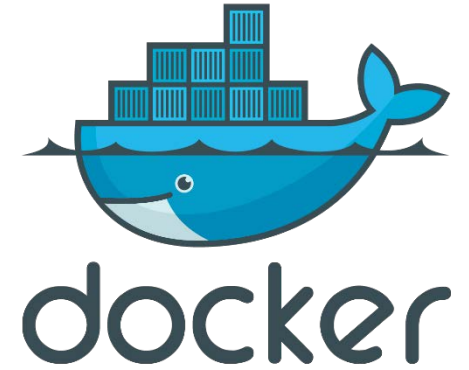


Microsoft Azure



<i>CPU Cores</i>	<i>RAM</i>	<i>Cost</i>
1	3.5GB	\$100
1	3.5GB	\$60
2	4GB	\$40

.NET Core Console Apps





.NET Framework Console Apps

- Good on-premise
- Existing skills



.NET Framework Hadoop SDK

- Risk of deprecation
- Too little benefit

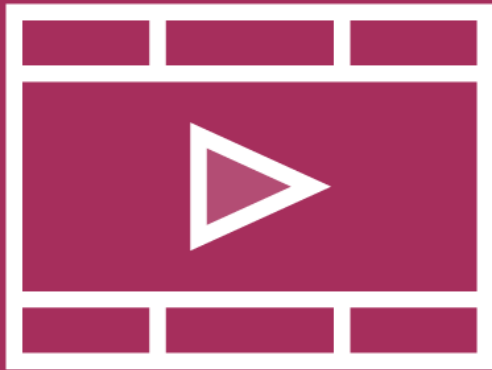


.NET Core Console Apps

- **Most flexible**
- **Cheaper cloud**



Coming Next



Querying Data with MapReduce

- Combiners & multiple reducers
- Hadoop distributed cache
- Failure handling & counters