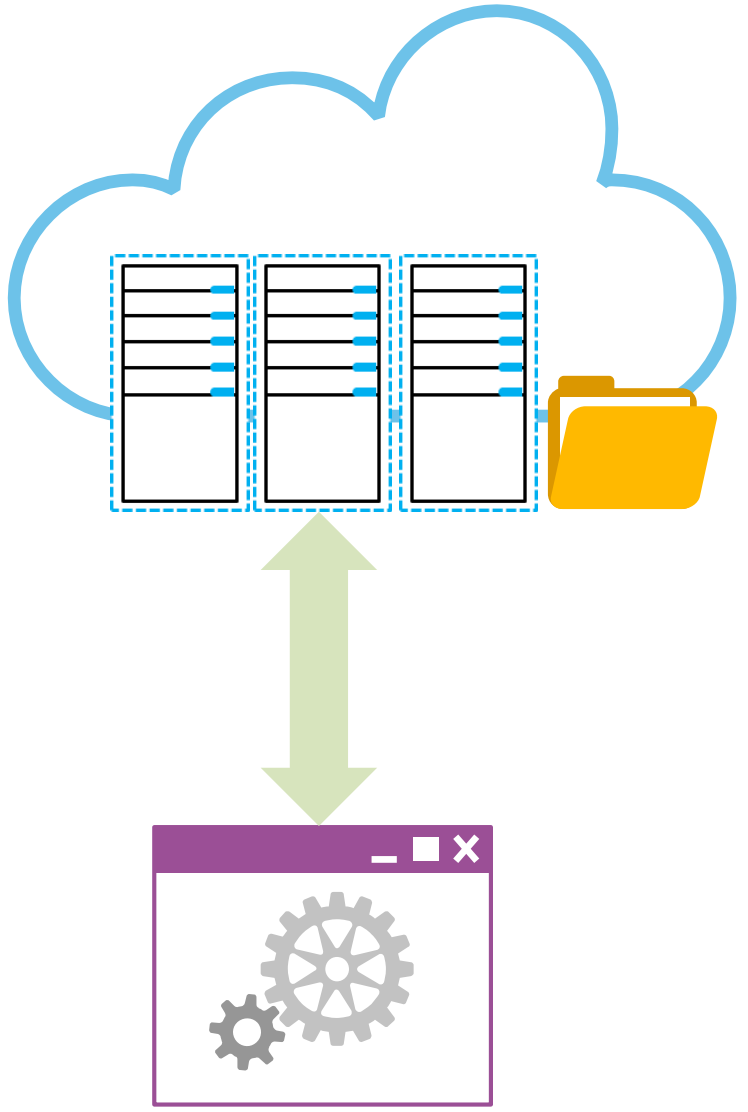
Building HDInsight Solutions with .NET

Microsoft

What can I do with .NET and HDinsight?

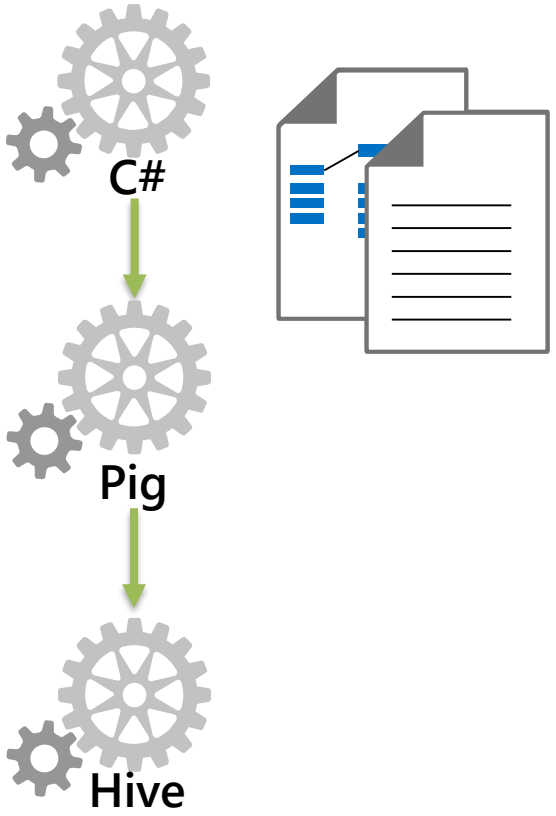
- Serialize source data using Avro
- Upload files to Azure storage
- Invoke Hadoop jobs

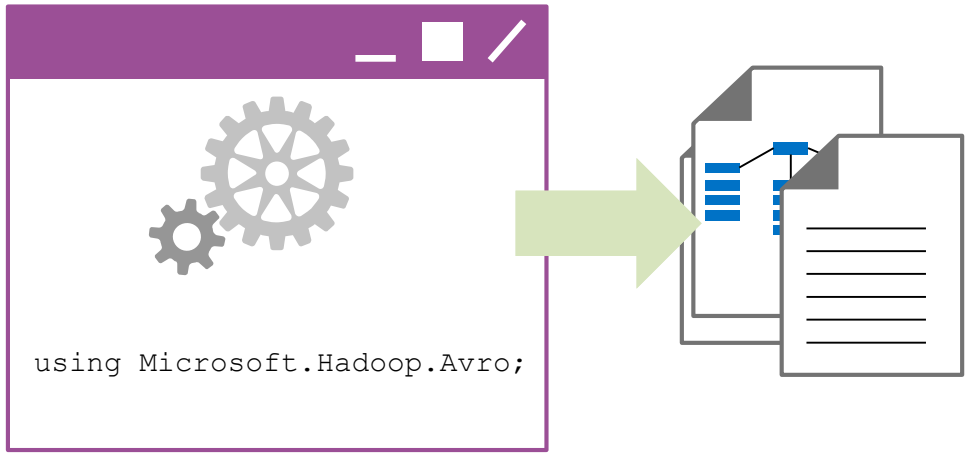


What is Avro?

Apache Avro is a splittable serialization and data interchange format

- Based on JSON - language-agnostic
- Serializes schema and data
- Supports compression

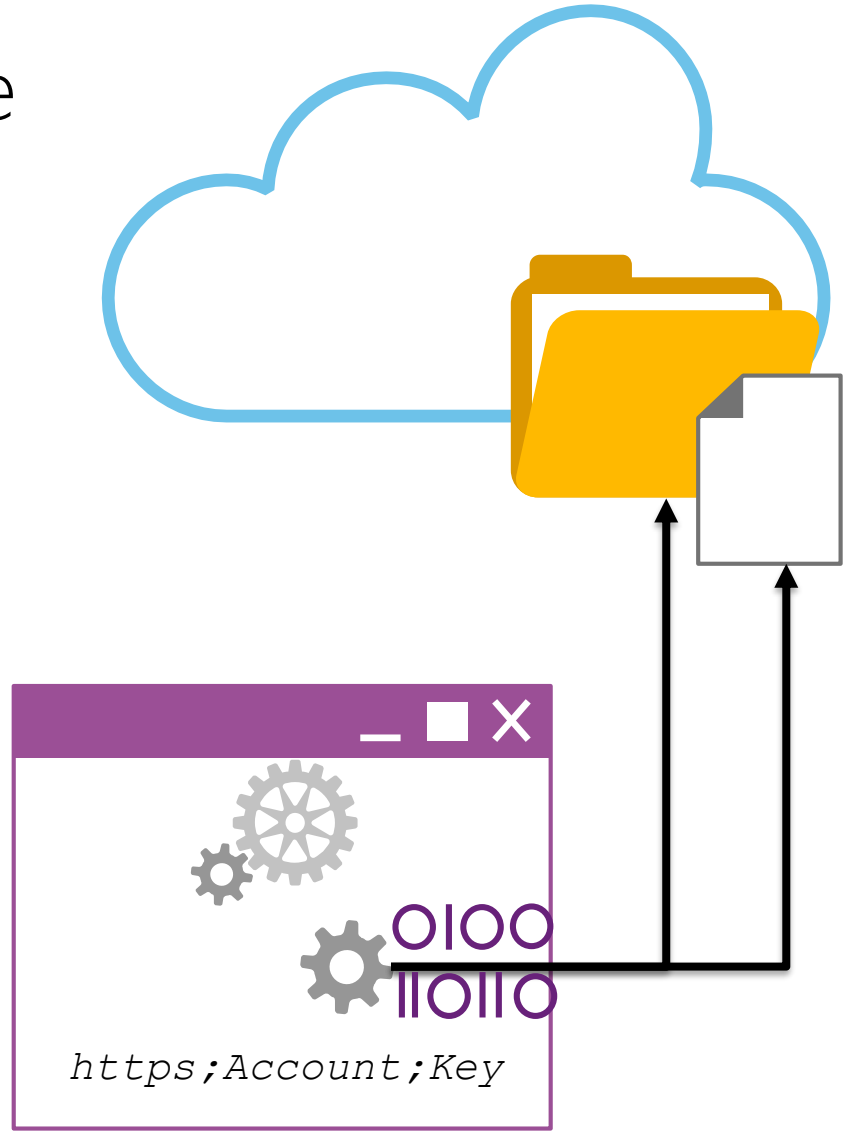




1. Import the **Microsoft Azure HDInsight Avro NuGet Package**
2. Use Avro classes to serialize a stream
   - Use **AvroSerializer** to serialize data-only
     - Use reflection to serialize .NET objects as data-only
     - Serialize other data in a generic JSON record schema
   - Use **AvroContainer** to serialize schema and data
     - Use reflection to serialize .NET objects and schema
     - Serialize other data with a generic JSON record schema

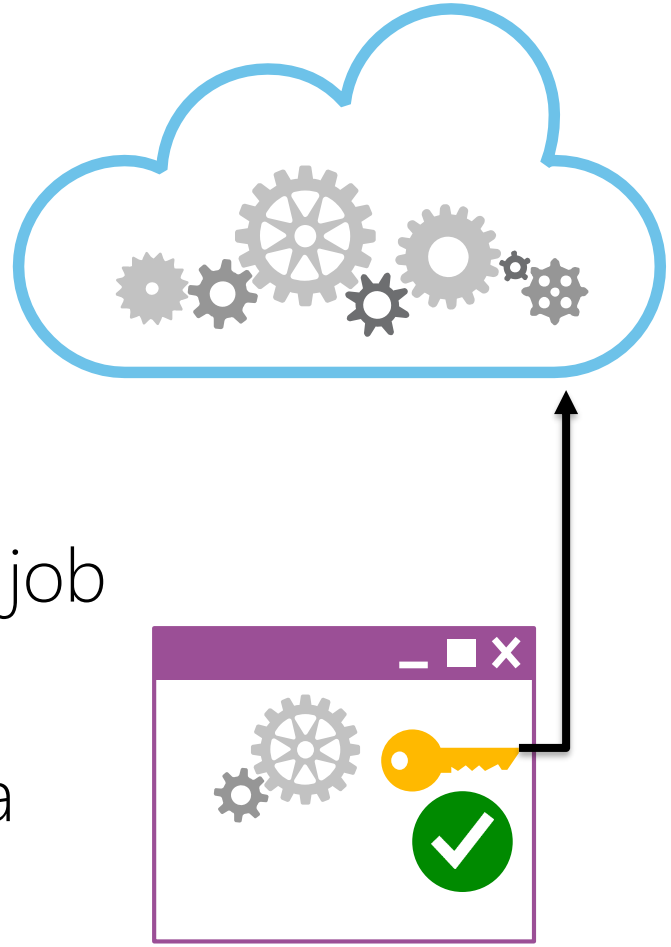
How do I write code that uses Azure storage?

1. Import Azure Storage NuGet package
2. Create a connection string for your storage account
3. Create a **CloudBlobClient** object
4. Create a **CloudBlobContainer** object that references your container
5. Create a **CloudBlockBlob** object that references a blob
6. Read or write a stream to/from the blob



How do I submit Hadoop jobs?

1. Import HDInsight Management NuGet package
   - Includes Hadoop Client
2. Create a ***Credential** object for your cluster
   - Certificate
   - Basic Authentication
3. Create a ***JobCreateParameters** object to define the job
   - MapReduce, Streaming, Pig, Hive, etc.
4. Use the **JobSubmissionClientFactory** class to create a client from your credentials
5. Use the client to create the job based on your parameters
6. Use the client to get the job ID and check status until complete



Microsoft

©2014 Microsoft Corporation. All rights reserved. Microsoft, Windows, Office, Azure, System Center, Dynamics and other product names are or may be registered trademarks and/or trademarks in the U.S. and/or other countries. The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation. MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.