

# End To End Azure Data Engineer Video Tutorial

1. *Python*
2. *Databricks*
3. *Spark*
4. *Pyspark*
5. *Spark SQL*
6. *Delta Lake*
7. *Azure Data Factory*
8. *Azure Synapse DW (Dedicated SQL POOL)*



- ✓ Python Tutorial 1 : Introduction To Python  
<https://www.youtube.com/watch?v=jtvo6jR4rMQ>
- ✓ Python Tutorial 2 : Introduction To Python Variables , String format function  
<https://www.youtube.com/watch?v=yl3uoL050gw>
- ✓ Python Tutorial 3 : Python Data types  
[https://www.youtube.com/watch?v=iorae42\\_LNU](https://www.youtube.com/watch?v=iorae42_LNU)
- ✓ Python Tutorial 4 : Python If conditions and For loops, while loops  
<https://www.youtube.com/watch?v=Y3t-UZ9wOH8>
- ✓ Python Tutorial 5 : Python Functions and Lambda function  
<https://www.youtube.com/watch?v=tnXP9sNvwjU>
- ✓ Python Tutorial 6 : Python List and Tuple , Python Collections , Python Data structures  
<https://www.youtube.com/watch?v=Pze5xxQMJCc>
- ✓ Python Tutorial 7 : Python sets and Dictionary

<https://www.youtube.com/watch?v=Phteapl02lQ>

✓ Python Tutorial 8 : Exception Handling

<https://www.youtube.com/watch?v=uQNamWDLc8c>

✓ Python Tutorial 9 : Python File I/O , how to create and read file in Python

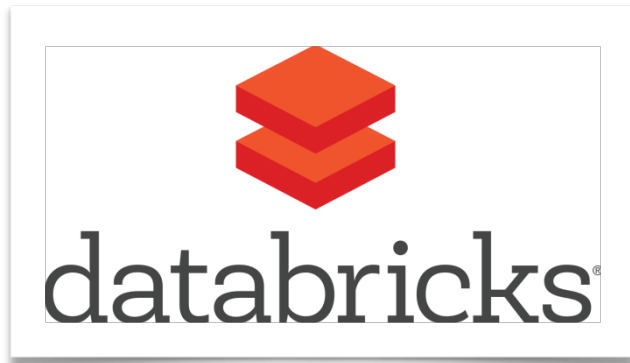
<https://www.youtube.com/watch?v=qHJjwNdLhio>

✓ Python Tutorial 10 : Python Common Functions and list comprehensions

<https://www.youtube.com/watch?v=7951aTuXWEk>

✓ Python Tutorial 11 : Python Regular Expressions

<https://www.youtube.com/watch?v=Qbrbh6o5cfw>



✓ Databricks Tutorial 1 : Introduction To Azure Databricks

<https://www.youtube.com/watch?v=2otrn2mvISQ>

✓ Databricks Tutorial 2 : Azure Databricks Concepts

<https://www.youtube.com/watch?v=BDy5VEOtmNg>

- ✓ Databricks Tutorial 3 : How to use databricks free  
<https://www.youtube.com/watch?v=iRmV9z0mIVs>
- ✓ Databricks Tutorial 4 : databricks Workspace resources  
<https://www.youtube.com/watch?v=3OqBSi8wMEs>
- ✓ Databricks Tutorial 5: How to create Notebook  
<https://www.youtube.com/watch?v=2Iuy69RMXpE>
- ✓ databricks tutorial 6 : databricks notebooks new features  
<https://www.youtube.com/watch?v=-o6N8qbDrk8>
- ✓ Databricks Tutorial 7: How to Read Json Files in Pyspark  
<https://www.youtube.com/watch?v=eTAYSZ2NIDk>
- ✓ Databricks Tutorial 8: Read xml files in Pyspark,  
<https://www.youtube.com/watch?v=IhKXfgj24MU>
- ✓ Databricks Tutorial 9: Reading excel files pyspark  
<https://www.youtube.com/watch?v=bcRqEoEMW0k>
- ✓ Databricks Tutorial 10: How to read a url file in pyspark  
<https://www.youtube.com/watch?v=783Ihm2eIKw>
- ✓ Databricks Tutorial 11 : Custom Logging in Databricks pyspark  
<https://www.youtube.com/watch?v=Snlp6p8px5Q>
- ✓ Databricks Tutorial 12 : SQL Cache, spark cache, spark persistence  
<https://www.youtube.com/watch?v=tmnYSmdrTrg>
- ✓ Databricks Tutorial 13: DBUTILS commands in databricks  
<https://www.youtube.com/watch?v=qR18LMfzULQ>
- ✓ Databricks Tutorial 14 : Databricks Variables, Widget Types

<https://www.youtube.com/watch?v=cJlfla-7-X0>

✓ Databricks Tutorial 15: Databricks workspace limits,

<https://www.youtube.com/watch?v=CRBVwmnL2Jk>

✓ Databricks Tutorial 16 : How to call one notebook from another notebook in Databricks

<https://www.youtube.com/watch?v=B1DyJScg0-k>

✓ Databricks Tutorial 17 : What is Single Node Cluster in Databricks, How to use single node cluster

<https://www.youtube.com/watch?v=tbiW8hOZY9M>

✓ Databricks Tutorial 18 :What are the types of Databricks Cluster Types and Difference.

<https://www.youtube.com/watch?v=bSKTgwzssAk>

✓ Databricks Tutorial 19 How To Create Job in Databricks, Job Scheduling In Databricks

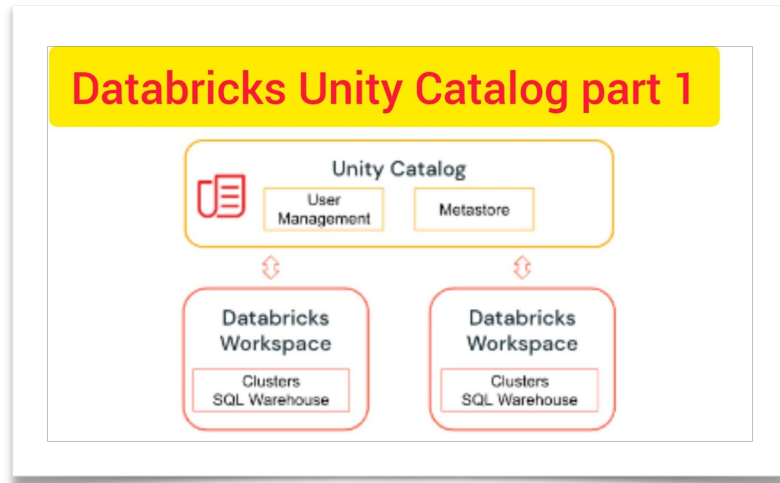
<https://www.youtube.com/watch?v=8e5vkoOblxo>

✓ Databricks Tutorial 20 Azure Data Engineering Architecture ADF + Databricks

<https://www.youtube.com/watch?v=y49l4rXKSXc>

✓ Databricks Tutorial 21 Databricks interview questions and answers

<https://www.youtube.com/watch?v=KH-rTNJSojk>



✓ Databricks Unity Catalog part1 what is databricks unity catalog?

<https://www.youtube.com/watch?v=Y-mL8LdCveA>

✓ Databricks Unity Catalog part2 How to Configure databricks unity catalog

<https://www.youtube.com/watch?v=x2RTEfdY2Bg>

✓ Databricks Unity Catalog part3 How to create storage credentials and external location

<https://www.youtube.com/watch?v=7BDhpsoyoWQ>



✓ What is scaling? Vertical vs horizontal scaling?

<https://www.youtube.com/watch?v=wg7MYTbsSC8&t=0s>

✓ what is VCPU? CPU - Core - Threads

<https://www.youtube.com/watch?v=4FQIA5FJERA>

✓ Spark Architecture Part 1: Spark Vs Hadoop MR spark vs mapreduce

<https://www.youtube.com/watch?v=Ir3U0HmnfVI&t=0s>

✓ Spark Architecture Part2 : master slave architecture , Single node cluster , multi node cluster

[https://www.youtube.com/watch?v=\\_DpT8q5auqk&t=0s](https://www.youtube.com/watch?v=_DpT8q5auqk&t=0s)

✓ Spark Architecture Part 3 : sparksession vs sparkcontext

<https://www.youtube.com/watch?v=v2roktqTixI&t=0s>

✓ Spark Architecture Part 4 : Spark job to stage and stage to task spark job spark stages

<https://www.youtube.com/watch?v=Ycwhmg6vlyc&t=0s>

✓ Spark Architecture Part 5 : Spark narrow & wide transformations

<https://www.youtube.com/watch?v=uNMkRjoGkKA&t=0s>

✓ Spark Architecture Part 6 : pyspark word count program example

<https://www.youtube.com/watch?v=3Nf3WcWwb88&t=0s>





✓ Pyspark Scenarios 1: How to create partition by month and year in pyspark

<https://www.youtube.com/watch?v=HU29qHboPN4&t=0s>

✓ pyspark scenarios 2 : how to read variable number of columns data in pyspark dataframe

<https://www.youtube.com/watch?v=R7PEwQzqYmY&t=0s>

✓ Pyspark Scenarios 3 : how to skip first few rows from data file in pyspark

<https://www.youtube.com/watch?v=4eFaWM6m-wk&t=0s>

✓ Pyspark Scenarios 4 : how to remove duplicate rows in pyspark dataframe

<https://www.youtube.com/watch?v=xw4a9qbOh-Q&t=0s>

✓ Pyspark Scenarios 5 : how read all files from nested folder in pySpark dataframe

<https://www.youtube.com/watch?v=7jxFfeQHpQ&t=0s>

✓ Pyspark Scenarios 6 How to Get no of rows from each file in pyspark dataframe

<https://www.youtube.com/watch?v=wp2KgEy0pTo&t=0s>

✓ Pyspark Scenarios 7 : how to get no of rows at each partition in pyspark dataframe

<https://www.youtube.com/watch?v=uNT08FneU4E&t=0s>

✓ Pyspark Scenarios 8: How to add Sequence generated surrogate key as a column in dataframe.

<https://www.youtube.com/watch?v=WsU7jX3KUVm&t=0s>

✓ Pyspark Scenarios 9 : How to get Individual column wise null records count

<https://www.youtube.com/watch?v=2bmH3zemRe0&t=0s>

✓ Pyspark Scenarios 10: Why we should not use crc32 for Surrogate Keys Generation?

<https://www.youtube.com/watch?v=fg6zwaYdneU&t=0s>

✓ Pyspark Scenarios 11 : how to handle double delimiter or multi delimiters in pyspark

<https://www.youtube.com/watch?v=J2Fb2lAt5Eo&t=0s>

✓ Pyspark Scenarios 12 : how to get 53 week number years in pyspark extract 53rd week number in spark

<https://www.youtube.com/watch?v=VpYcbPRsasc&t=0s>

✓ Pyspark Scenarios 13 : how to handle complex json data file in pyspark

[https://www.youtube.com/watch?v=aBNQzWV\\_UmE&t=0s](https://www.youtube.com/watch?v=aBNQzWV_UmE&t=0s)

✓ Pyspark Scenarios 14 : How to implement Multiprocessing in Azure Databricks

<https://www.youtube.com/watch?v=OQeRPh04mz4&t=0s>

✓ Pyspark Scenarios 15 : how to take table ddl backup in databricks

<https://www.youtube.com/watch?v=yukhCLUo1Qk&t=0s>

✓ Pyspark Scenarios 16: Convert pyspark string to date format issue dd-mm-yy old format

<https://www.youtube.com/watch?v=F64rlowo4IU&t=0s>

✓ Pyspark Scenarios 17 : How to handle duplicate column errors in delta table

<https://www.youtube.com/watch?v=61BhN7GPtU8&t=0s>

✓ Pyspark Scenarios 18 : How to Handle Bad Data in pyspark dataframe using pyspark schema

<https://www.youtube.com/watch?v=yKueGqJAgwM&t=0s>

✓ Pyspark Scenarios 19 : difference between OrderBy Sort and sortWithinPartitions Transformations

<https://www.youtube.com/watch?v=cr8bcpvC8Hk&t=0s>

✓ Pyspark Scenarios 20 : difference between coalesce and repartition in pyspark

<https://www.youtube.com/watch?v=9tRyWZvdUMM&t=0s>

✓ Pyspark Scenarios 21 : Dynamically processing complex json file in pyspark

<https://www.youtube.com/watch?v=qfJb45SusMo&t=0s>

✓ Pyspark Scenarios 22 : How To create data files based on the number of rows in PySpark

<https://www.youtube.com/watch?v=O1SpqoFirxc&t=0s>

✓ Pyspark Scenarios 23 : How do I select a column name with spaces in PySpark?

<https://www.youtube.com/watch?v=bErn1bHAyqw&t=83s>

## DELTA LAKE



✓ introduction To Delta Lake : What is Delta Lake, What is Azure Databricks Delta lake

<https://www.youtube.com/watch?v=FpxkiGPFyfM&t=0s>

✓ Delta Lake tutorial 2 : How to Create delta lake table By path

<https://www.youtube.com/watch?v=cbobqI3ZGuA>

✓ Delta Lake Tutorial 3: How To Create delta lake table in metastore, delta lake table

<https://www.youtube.com/watch?v=mFYF15luV2w>

✓ delta lake tutorial 4 : what is schema enforcement in delta lake

<https://www.youtube.com/watch?v=MnLFAjQz3Yg>

✓ delta lake tutorial 5 : Delta Lake Timetravel and Audit Log

<https://www.youtube.com/watch?v=fitjhpy--P4>

✓ delta lake tutorial 6 : What is Delta Lake Vacuum

<https://www.youtube.com/watch?v=YQjTQ0sgE1k>

✓ delta lake tutorial 7 : What is Delta Lake Restore

<https://www.youtube.com/watch?v=IK8VFuj9qKM&t=78s>

✓ delta lake tutorial 8 : What is Delta Lake Optimize

<https://www.youtube.com/watch?v=P1aTktA-Xi8>

## Azure Services



- ✓ Spark SQL 1 : I am going to start spark sql sessions as series.  
<https://www.youtube.com/watch?v=TA5SYL0wzFE&t=4s>
- ✓ Spark SQL 2 : How to register databricks community edition  
<https://www.youtube.com/watch?v=Jp0Wu5t2OBA&t=0s>
- ✓ Spark SQL 3: What is DataWarehouse? OLTP and OLAP?  
<https://www.youtube.com/watch?v=kFyh8hGFy6E&t=0s>
- ✓ Spark SQL 4 : how to create database in databricks?  
[https://www.youtube.com/watch?v=Lg\\_xjyXOFFc&t=0s](https://www.youtube.com/watch?v=Lg_xjyXOFFc&t=0s)
- ✓ Spark SQL 5 : databricks file system dbfs.  
<https://www.youtube.com/watch?v=dcrf81rujm0&t=0s>
- ✓ Spark SQL 6 : Spark SQL Table , Difference between Managed table and External table  
<https://www.youtube.com/watch?v=7o7P6JQYvKI&t=0s>
- ✓ Spark SQL 7 : Spark SQL Views, Different types of views, Temporary views

<https://www.youtube.com/watch?v=gp0BwkNuZNc&t=0s>

✓ Spark SQL 8 : Spark SQL Constraints NOT NULL and CHECK Constraints

<https://www.youtube.com/watch?v=DKlnHX4M88U&t=0s>

✓ Spark SQL 9: Spark SQL DML Operations INSERT Statement

<https://www.youtube.com/watch?v=Doy19FxBqQQ&t=0s>

✓ Spark SQL 10 : Difference Between Truncate and Delete #sparksql #deltalake #pyspark

<https://www.youtube.com/watch?v=YezAhvK6pgY&t=0s>

✓ Spark SQL 11 : Spark SQL Delta Delete Operations and delete and restore

<https://www.youtube.com/watch?v=14GuUhzkb9A&t=0s>

✓ Spark SQL 12 : Spark SQL Delta Update Operations #sparksql #deltalake #pyspark

<https://www.youtube.com/watch?v=QDCbCnIYeiY&t=0s>

✓ Spark SQL 13: Spark SQL Delta Restore Operations #sparksql #deltalake #pyspark

<https://www.youtube.com/watch?v=qUfn2GS630Y&t=0s>

✓ Spark SQL 14: What is slowly changing dimension #SCD #sparksql #deltalake #pyspark

[https://www.youtube.com/watch?v=\\_pcltsd7iKM&t=0s](https://www.youtube.com/watch?v=_pcltsd7iKM&t=0s)

✓ Spark SQL 15: What is SCD Type 0 and SCD Type 1 #SCD #sparksql #deltalake #pyspark

<https://www.youtube.com/watch?v=LTeejDtR6a8&t=0s>

✓ Spark SQL 16: What is slowly changing dimension Type 2 and Type 3 #SCD #sparksql

<https://www.youtube.com/watch?v=7kjrRveFNro&t=0s>

✓ Spark SQL 17: SQL Select Statement column alias and table alias #select #sparksql

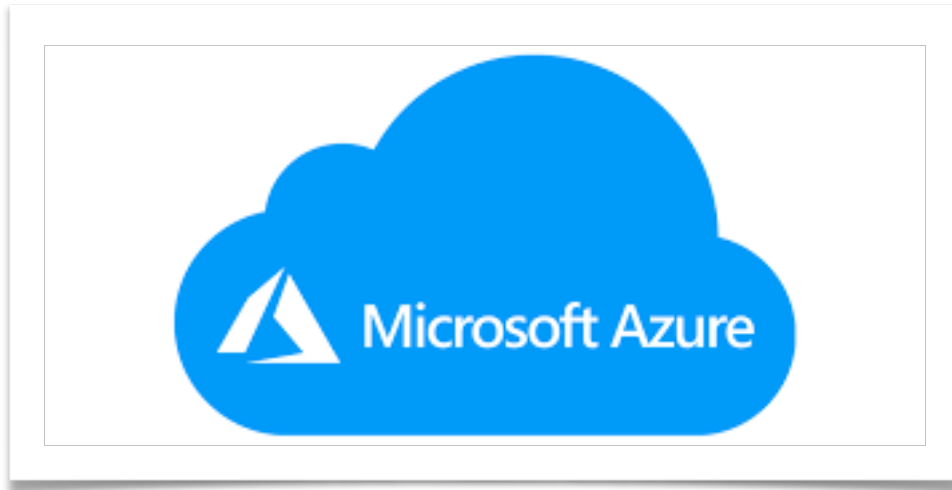
<https://www.youtube.com/watch?v=ue5nG6MYXpA&t=0s>

✓ Spark SQL 18: SQL Order By and Sort By Clauses #select #sparksql

<https://www.youtube.com/watch?v=uaRg-sCysYs&t=0s>

- ✓ Spark SQL 19 : SQL Join Types inner join full join semi #select #sparksql  
<https://www.youtube.com/watch?v=QmHp6Bbp1VA&t=0s>
- ✓ Spark SQL 20 : SQL Set Operators #sparksql #sqlsetoperators #sql #union  
<https://www.youtube.com/watch?v=sHwmfDmQei8&t=0s>
- ✓ Spark SQL 21: spark SQL group by clause and having clause  
<https://www.youtube.com/watch?v=TU1VH6B4Xaw&t=0s>
- ✓ Spark SQL 22: Advanced Group By Grouping Sets Cube Rollup  
<https://www.youtube.com/watch?v=nj6bNsRJTzk&t=0s>
- ✓ Spark SQL 23 : Window functions Rank and dense rank  
<https://www.youtube.com/watch?v=Te2d7eUw51E&t=0s>





✓ Cloud Fundamentals part 1: what is datalake.

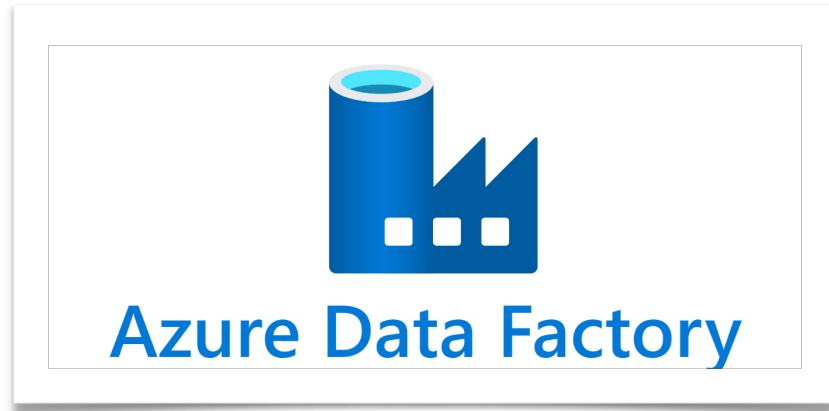
<https://www.youtube.com/watch?v=HJKu9fPLBBo>

✓ azure fundamentals part 2 : What is data center , type of cloud services

<https://www.youtube.com/watch?v=oLaP898XO8g&t=0s>

✓ azure fundamentals part 3 : types of cloud service models #cloudservices

<https://www.youtube.com/watch?v=33t4ojvqlpk>



✓ Introduction to azure data factory tutorial part 1

<https://www.youtube.com/watch?v=oPNWFPsmOK4>

✓ Azure data factory 2 : adf components

<https://www.youtube.com/watch?v=p5cF3QXGtqU&t=0s>

✓ Azure Data Factory 3 : Adls gen1 to gen2 migration

<https://www.youtube.com/watch?v=XE5jPzds474&t=0s>

✓ Azure Data Factory 4 : How to Copy Multiple files from Gen 1 to Gen2

[https://www.youtube.com/watch?v=YS\\_0tr-hwjQ&t=0s](https://www.youtube.com/watch?v=YS_0tr-hwjQ&t=0s)18

✓ Azure Data Factory 5 : self hosted integration runtime configuration

<https://www.youtube.com/watch?v=D3vPuG9JTwI&t=0s>

✓ Azure Data Factory 6 : How to migrate Onpremises Oracle data to azure datalake gen2

<https://www.youtube.com/watch?v=h1LynU-T7OE&t=0s>

✓ Azure Data Factory 7 : What is Binary File Dataset and when we will use this?

<https://www.youtube.com/watch?v=yw5tSZc3oD4&t=0s>

✓ Azure Data Factory 8 : how to split one file into multiple files using git-scm

<https://www.youtube.com/watch?v=zto8JnspkQE>

✓ Azure Data Factory 9 : what is linked self hosted integrated runtime

<https://www.youtube.com/watch?v=KcLRWX5CSXA>

✓ Azure Data Factory 10 : How do i resolve this error self-hosted IR for file linked service

<https://www.youtube.com/watch?v=fv4R74YxOSs>

✓ Azure Data Factory 11 : how to migrate onprem files to adls gen2

<https://www.youtube.com/watch?v=GfnJ26lQOHA>



✓ Azure Synapse Tutorial 1 : what is MPP architecture? What is MPP Warehouse

<https://www.youtube.com/watch?v=rdFm0Fvp4r4>

✓ Azure Synapse Tutorial 2 : Azure Synapse DW introduction

<https://www.youtube.com/watch?v=5190fPv0MnU>

✓ Azure Synapse Tutorial 3 : azure synapse Distribution types

<https://www.youtube.com/watch?v=Zsw58YeKy0Q>

✓ Azure Synapse Tutorial 4 : what is hash distribution in synapse

<https://www.youtube.com/watch?v=CgdJQ7anxgU>

✓ Azure Synapse Tutorial 5 : What is round robin partition

<https://www.youtube.com/watch?v=oTjAWBAoE4s>

✓ Azure Synapse Tutorial 6 : What is replicate distribution or Broadcast Distribution in azure synapse

[https://www.youtube.com/watch?v=u\\_XYWDH3i7I](https://www.youtube.com/watch?v=u_XYWDH3i7I)

✓ Azure Synapse Tutorial 7 : Azure Synapse dwh interview questions and answers

<https://www.youtube.com/watch?v=EMmJ8XmkqBo>

✓ Azure Synapse Tutorial 8 : Dedicated sql pool resource classes

<https://www.youtube.com/watch?v=DghCPpZ5vkQ>

✓ Difference between Synapses Analytics and Synapses Dedicated SQL pool (Azure SQL DWH)

[https://www.youtube.com/watch?v=MD\\_TztbAgX8](https://www.youtube.com/watch?v=MD_TztbAgX8)