1. SSIS Learning Curve : 20 hrs : 15 mins

|  |  |  |  |
| --- | --- | --- | --- |
| Sr. No. | Name | Length | Status |
| 1 | Introduction to SSIS | Create your first SSIS package | 24:49 |  |
| 2 | SSIS Tutorial for beginners | 16:16 |  |
| 3 | Install Visual Studio 2019 Community | SSIS 2019 Installation | Integration services 2019 | 10:03 |  |
| 4 | Install SQL Server 2019 with custom option | Install SQL Server 2019 | 7:04 |  |
| 5 | Install SQL Server 2019 | 5:07 |  |
| 6 | Import csv file to sql server using SSIS | Load CSV File in SSIS | 6:12 |  |
| 7 | Load Fixed Width File in SQL Server - SSIS | Import Fixed width file in SSIS | 6:57 |  |
| 8 | Load Excel File in SQL Server - SSIS | Import excel file in SQL Server using SSIS | 4:56 |  |
| 9 | Overwrite data to Excel file in SSIS | Export data to excel file using SSIS | 10:08 |  |
| 10 | Load XML File in SQL Server - SSIS | Import XML file in sql server using SSIS | 5:57 |  |
| 11 | Copy data from One SQL Sever Instance to another - SSIS | 5:57 |  |
| 12 | Export data from SQL to CSV File - SSIS | Export data to csv file in SSIS | 4:07 |  |
| 13 | Export data from SQL to Fixed Width File - SSIS | export data to fixed file in SSIS | 5:34 |  |
| 14 | Export data from Sql Server to Excel using SSIS | 3:22 |  |
| 15 | Execute SQL Task in SSIS | Use variables in Execute SQL task in SSIS | 16:38 |  |
| 16 | File System Task in SSIS | How to use File System Task in SSIS | 4:05 |  |
| 17 | Zip and UnZip files in SSIS | 6:39 |  |
| 18 | Script Task in SSIS | Using script task in SSIS | 8:31 |  |
| 19 | Debugging Script Task in SSIS | 12:18 |  |
| 20 | Sequence container in SSIS example | 6:14 |  |
| 21 | For loop Container in SSIS | 13:28 |  |
| 22 | Foreach Loop Container in SSIS Example | 11:56 |  |
| 23 | Foreach loop ado enumerator in SSIS example | 10:53 |  |
| 24 | Foreach loop Item Enumerator in SSIS Example | 7:50 |  |
| 25 | Load multiple excel sheets using SSIS | 6:33 |  |
| 26 | Aggregate Transformation in SSIS | 7:25 |  |
| 27 | Row Count Transformation in SSIS | RowCount Transformation in SSIS | 9:47 |  |
| 28 | Data Conversion Transformation in SSIS | 6:20 |  |
| 29 | Character Map Transformation in SSIS | 5:35 |  |
| 30 | Copy Column Transformation in SSIS | 4:21 |  |
| 31 | Derived Column Transformation in SSIS | 8:21 |  |
| 32 | Multicast Transformation in SSIS | 7:36 |  |
| 33 | Conditional Split Transformation in SSIS | 7:01 |  |
| 34 | Union All Transformation in SSIS | 4:26 |  |
| 35 | Merge Transformation in SSIS | 7:05 |  |
| 36 | Merge Join Transformation in SSIS | 7:56 |  |
| 37 | Script Component Transformation in SSIS | 13:04 |  |
| 38 | OLE DB Command Transformation in SSIS | OLEDB Command Transformation | 7:35 |  |
| 39 | Lookup Transformation in SSIS | 8:04 |  |
| 40 | Fuzzy Lookup Transformation in SSIS | 13:22 |  |
| 41 | Fuzzy Grouping Transformation in SSIS | 9:11 |  |
| 42 | Slowly Changing Dimension in SSIS | 13:58 |  |
| 43 | Error handling in SSIS | How to handle errors in SSIS | 16:34 |  |
| 44 | Execute Package Task in SSIS | 9:52 |  |
| 45 | Implement Logging to SQL Server in SSIS | 5:39 |  |
| 46 | Implement Logging to Flat File in SSIS | 5:50 |  |
| 47 | Send Email in SSIS | 5:14 |  |
| 48 | Send HTML Email in SSIS | 9:35 |  |
| 49 | Pass value from Parent Package to Child Package in SSIS | 6:54 |  |
| 50 | Execute SSIS Package from command line | Execute SSIS Package from batch file | 12:33 |  |
| 51 | Import Export Wizard in SQL Server | Import Export Wizard | 9:04 |  |
| 52 | Execute package utility in SSIS | how to execute ssis package from dtexecui | 5:55 |  |
| 53 | SSIS Package deployment step by step | Deploy SSIS package | Deploy SSIS project to ssis catalog | 10:18 |  |
| 54 | Converting from Project Deployment to Package Deployment | 7:11 |  |
| 55 | XML Configuration in SSIS | 10:33 |  |
| 56 | SQL Server Configurations in SSIS | SSIS configuration sql server | 9:16 |  |
| 57 | Checkpoints in SSIS with example | 11:58 |  |
| 58 | 58 Load Multiple Excel Sheets with different schema | 16:41 |  |
| 59 | Breakpoints in SSIS | how to use breakpoints in ssis | 20:13 |  |
| 60 | Data Viewer in SSIS | SSIS Data Viewer | 7:26 |  |
| 61 | Schedule SSIS package in SQL Server agent | create sql job to run ssis package | 10:17 |  |
| 62 | Schedule an SSIS package using Windows Task Scheduler | 8:30 |  |
| 63 | Insert update differential data in SSIS | 21:37 |  |
| 64 | Create SQL Job from SSIS Catalog Package | 7:21 |  |
| 65 | How do I stop Package Execution in SSIS catalog | 12:27 |  |
| 66 | Precedence Constraints in SSIS | 5:43 |  |
| 67 | SSIS Transactions | Sequence Container in SSIS | 8:21 |  |
| 68 | Project Deployment in SSIS and Create environment | 17:20 |  |
| 69 | How do I debug an SSIS package? | 8:26 |  |
| 70 | Foreach loop item enumerator in ssis example | 7:38 |  |
| 71 | What are expressions in SSIS ? How to write expressions ? | 18:10 |  |
| 72 | How to write expressions in SSIS ? | 18:57 |  |
| 73 | How to get FileName of all different types of Files in a Folder and write to a SQL table ? | 6:08 |  |
| 74 | How to unzip multiple zip files in SSIS ? | 5:42 |  |
| 75 | How to delete duplicate records in SSIS | 4:13 |  |
| 76 | What are variables and Parameters in SSIS | 10:48 |  |
| 77 | How to execute multiple child packages from parent package | 16:08 |  |
| 78 | Synchronous vs Asynchronous Transformations in SSIS | 2:18 |  |
| 79 | Delay Validation in SSIS | 4:04 |  |
| 80 | Load current date file into SQL server table using SSIS | 10:46 |  |
| 81 | Audit Transformation in SSIS | 4:16 |  |
| 82 | Send Mail Task in SSIS | Send Notifications using Send Mail Task in SSIS | 4:43 |  |
| 83 | Pivot in SSIS | Pivot Transformation in SSIS | 7:31 |  |
| 84 | How to use sql authentication connection in SSIS | Make the connection dynamic using variables | 8:01 |  |
| 85 | How to use Windows authentication and sql authentication in a single connection in SSIS | 14:09 |  |
| 86 | How to Execute SSIS packages in parallel | Execute Tasks in Parallel in SSIS | 7:14 |  |
| 87 | How to import dbf files into SQL Server using SSIS | 5:38 |  |
| 88 | How to read data from Stored Procedure and insert into SQL table using SSIS | 4:10 |  |
| 89 | How to read data from Stored Procedure and pass input parameter from a variable using SSIS | 12:49 |  |
| 90 | How to pass the SSIS variable value from SQL Agent Job | 9:38 |  |
| 91 | How to skip rows in flat file in SSIS | How to skip rows in CSV file in SSIS | 5:15 |  |
| 92 | How to Execute SQL Queries which are stored on disk using SSIS | 4:01 |  |
| 93 | How to convert Y and N to 1 and 0 using SSIS expressions | 12:18 |  |
| 94 | How to move files from one folder to another using SSIS | File System Task in SSIS | 7:14 |  |
| 95 | How to import a csv file into SQL Server using bulk insert SQL query | 4:13 |  |
| 96 | Deploy SSIS package using manifest file | Deploy SSIS package to File System | 11:40 |  |
| 97 | How to fix Connecting to Integration Services Access is denied error | 4:15 |  |
| 98 | how to deploy SSIS package to SQL server | how to deploy SSIS package to msdb | 7:51 |  |
| 99 | How to export data to Pipe delimited format using SSIS | 6:36 |  |
| 100 | Export data from multiple databases to CSV files using SSIS | 9:45 |  |
| 101 | export data to csv in sql server | sql server data export to csv using bcp | 9:05 |  |
| 102 | bulk insert task in ssis | 4:19 |  |
| 103 | Web Service task in SSIS | How to use Web service task in SSIS | 7:53 |  |
| 104 | consume web service from C# | 8:01 |  |
| 105 | how to use global temp table in ssis | how to use temp table in ssis | 7:01 |  |
| 106 | Execute SSIS package from Powershell | 6:15 |  |
| 107 | How to load json file in SSIS | Import json file to SQL Server using SSIS | 11:00 |  |
| 108 | How to load data from CSV file that has extra delimiter | delimiter for column "X" is not found | 16:14 |  |
| 109 | Event Handlers in SSIS | How to use Event Handlers in SSIS | 12:30 |  |
| 110 | How to check if File exists in SSIS | 13:08 |  |
| 111 | Invalid character value for cast specification in SSIS | 8:02 |  |
| 112 | The value violated the integrity constraints for the column | 5:39 |  |
| 113 | Removing special characters from data in SSIS | 7:37 |  |
| 114 | Convert data to proper case in SSIS | How do you convert a string to a proper case? | 8:18 |  |
| 115 | How to skip rows to load from excel SSIS | 8:40 |  |
| 116 | Execute Process Task in SSIS | 4:56 |  |
| 117 | Execute SQL Task in SSIS result set | 14:09 |  |
| 118 | Merge join transformation in SSIS with example | 11:29 |  |
| 119 | how can we load that special char rows into error table | 11:31 |  |
| 120 | How do you implement SCD Type 2 in SSIS Using Lookup | 24:49 |  |
| 121 | Replace null values with blank in SSIS | 6:44 |  |
| 122 | How do you write a case statement in a derived column in SSIS | 7:36 |  |
| 123 | Replace column value in SSIS | Replace a string with new string in SSIS | 7:14 |  |
| 124 | data conversion failed while converting column status value 4 in SSIS | 7:00 |  |
| 125 | Send email on SSIS package failure | Send Error description from SSIS package | 9:06 |  |
| 126 | What are third party components in SSIS | 7:06 |  |
| 127 | How to import a csv file into SQL server using C# or SSIS | 5:16 |  |
| 128 | How to load dimension table in sql using ssis | 24:46 |  |
| 129 | How do you load a fact table | 32:14 |  |
| 130 | FTP Task in SSIS | 8:29 |  |
| 131 | How do you upload multiple files to ftp server in SSIS | 5:26 |  |
| 132 | How to download files from sftp server using SSIS or C# | 9:23 |  |
| 133 | How to upload files to sftp server using SSIS or C# | 7:49 |  |
| 134 | How to import data from xml to sql server using C# | 6:47 |  |
| 135 | Percentage Sampling Transformation in SSIS | 7:47 |  |
| 136 | Row Sampling Transformation in SSIS | 5:33 |  |
| 137 | Export data to csv file with header using bcp command | 7:12 |  |
| 138 | How to create dynamic connection manager in ssis | 9:26 |  |
| 139 | What are different flat file formats in ssis | 14:53 |  |
| 140 | How to encrypt password in ssis package | How to save password in sql connection | 6:59 |  |
| 141 | Cache transformation in ssis | 6:08 |  |
| 142 | How to use lookup transformation with cache file in ssis | 9:51 |  |
| 143 | How to import multiple excel files into sql server using ssis | 18:42 |  |
| 144 | How to handle errors in script task in ssis | 11:23 |  |
| 145 | How do I redirect error rows to another table in ssis | Error handling at data flow in SSIS | 13:18 |  |
| 146 | How to import multiple xml files to sql server using ssis | 9:06 |  |
| 147 | How to store File Path into sql server table using ssis | 9:55 |  |
| 148 | How to do incremental load to mysql from sql server using ssis | 27:15 |  |
| 149 | data from excel file is getting inserted as float in sql in ssis | 7:47 |  |
| 150 | How to move files to month wise folders in ssis | 14:13 |  |
| 151 | How to merge data from multiple flat files using ssis | 9:31 |  |
| 152 | The connection string components cannot contain unquoted semicolons in SSIS | 3:56 |  |

1. Azure Data Factory Learning Curve : 21 hrs: 23 mins

|  |  |  |  |
| --- | --- | --- | --- |
| Sr. No. | Name | Length | Status |
| 1 | Introduction to Azure Data Factory | 7:50 |  |
| 2 | Top level Concepts in Azure Data Factory | 11:44 |  |
| 3 | Create your First Azure Data Factory | 17:52 |  |
| 4 | Different ways to work with Azure Data Factory | 5:26 |  |
| 5 | Pipelines and Activities in Azure Data Factory | 9:27 |  |
| 6 | Linked Services and Datasets in Azure Data Factory | 9:05 |  |
| 7 | Triggers in Azure Data Factory | 4:42 |  |
| 8 | Schedule Trigger in Azure Data Factory | 7:42 |  |
| 9 | Tumbling Window Trigger in Azure Data Factory | 11:53 |  |
| 10 | Tumbling Window Trigger Dependency in Azure Data Factory | 15:17 |  |
| 11 | Event based Triggers in Azure Data Factory | 12:21 |  |
| 12 | Integration runtime in Azure Data Factory | 15:43 |  |
| 13 | Azure Integration runtime in Azure Data Factory | 10:33 |  |
| 14 | Self Hosted Integration runtime in Azure Data Factory | 3:28 |  |
| 15 | Setting up Self Hosted Integration runtime in Azure Data Factory | 14:08 |  |
| 16 | Shared Self Hosted Integration runtime in Azure Data Factory | 10:28 |  |
| 17 | Parameterize Linked Services in Azure Data Factory | 12:58 |  |
| 18 | Parameterize Datasets in Azure Data Factory | 11:41 |  |
| 19 | Parameterize Pipelines in Azure Data Factory | 11:32 |  |
| 20 | System Variables in Azure Data Factory | 12:49 |  |
| 21 | Connectors Overview in Azure Data Factory | 10:09 |  |
| 22 | Supported File Formats in Azure Data Factory | 7:14 |  |
| 23 | Copy Data Activity in Azure Data Factory | 16:47 |  |
| 24 | Copy Data Activity in Azure data factory Continuation | 11:03 |  |
| 25 | Monitor Copy Data Activity in Azure Data Factory | 7:04 |  |
| 26 | Delete Activity in Azure Data Factory | 11:59 |  |
| 27 | Variables in Azure Data Factory | 9:14 |  |
| 28 | Set Variable Activity in Azure Data Factory | 6:05 |  |
| 29 | Append Variable Activity in Azure Data Factory | 6:33 |  |
| 30 | User Properties in Azure Data Factory | 5:56 |  |
| 31 | Execute Pipeline Activity in Azure Data Factory | 6:35 |  |
| 32 | Filter Activity in Azure Data Factory | 6:35 |  |
| 33 | ForEach Activity in Azure Data Factory | 12:05 |  |
| 34 | Get Metadata Activity in Azure Data Factory | 12:46 |  |
| 35 | If Condition Activity in Azure Data Factory | 9:01 |  |
| 36 | Wait Activity in Azure Data Factory | 4:36 |  |
| 37 | Until Activity in Azure Data Factory | 9:34 |  |
| 38 | Web Activity in Azure Data Factory | 16:14 |  |
| 39 | WebHook Activity in Azure Data Factory | 12:59 |  |
| 40 | Switch Activity in Azure Data Factory | 9:23 |  |
| 41 | Validation Activity in Azure Data Factory | 9:35 |  |
| 42 | Lookup Activity in Azure Data Factory | 19:17 |  |
| 43 | Transform Data Activities Overview in Azure Data Factory | 7:51 |  |
| 44 | Stored Procedure Activity in Azure Data Factory | 8:24 |  |
| 45 | Data flow in Azure data factory | 7:34 |  |
| 46 | Mapping Data Flow in Azure Data Factory | 17:45 |  |
| 47 | Data Flow Activity in Azure Data Factory | 9:51 |  |
| 48 | Mapping Data Flow Debug Mode in Azure Data Factory | 14:43 |  |
| 49 | Filter Transformation in Mapping Data Flow in Azure Data Factory | 14:44 |  |
| 50 | Aggregate Transformation in Mapping Data Flow in Azure Data Factory | 13:54 |  |
| 51 | JOIN Transformation in Mapping Data Flow in Azure Data Factory | 12:15 |  |
| 52 | Conditional Split Transformation in Mapping Data Flow in Azure data factory | 15:19 |  |
| 53 | Derived Column Transformation in Mapping Data Flow in Azure Data Factory | 12:18 |  |
| 54 | Exists Transformation in Mapping Data Flow in Azure Data Factory | 13:29 |  |
| 55 | Union Transformation in Mapping Data Flow in Azure Data Factory | 12:10 |  |
| 56 | Lookup Transformation in Mapping Data Flow in Azure Data Factory | 11:17 |  |
| 57 | Sort Transformation in Mapping Data Flow in Azure Data Factory | 5:45 |  |
| 58 | New Branch in Mapping Data Flow in Azure Data Factory | 17:53 |  |
| 59 | Select Transformation in Mapping Data Flow in Azure Data Factory | 11:38 |  |
| 60 | Pivot Transformation in Mapping Data Flow in Azure Data Factory | 16:13 |  |
| 61 | Unpivot Transformation in Mapping Data Flow in Azure Data Factory | 12:29 |  |
| 62 | Surrogate Key Transformation in Mapping Data Flow in Azure Data Factory | 11:30 |  |
| 63 | Window Transformation in Mapping Data Flow in Azure Data Factory | 18:21 |  |
| 64 | Alter Row Transformation in Mapping Data Flow in Azure Data Factory | 29:05 |  |
| 65 | . Flatten Transformation in Mapping Data Flow in Azure Data Factory | 12:02 |  |
| 66 | Parameterize Mapping Data Flow in Azure Data Factory | 12:30 |  |
| 67 | Validate Schema in Mapping Data Flow in Azure Data Factory | 10:46 |  |
| 68 | Schema Drift in Mapping Data Flow in Azure Data Factory | 12:08 |  |
| 69 | Wrangling Data Flow Overview in Azure Data Factory | 10:13 |  |
| 70 | Merge Queries in Wrangling Data Flow in Azure Data Factory | 10:09 |  |
| 71 | Group By in Wrangling Data Flow in Azure Data Factory | 10:56 |  |
| 72 | Different Author Modes in Azure Data Factory | 10:06 |  |
| 73 | Setup GitHub Code Repository for Azure Data Factory | 34:54 |  |
| 74 | Setup Azure DevOps Git Code Repository for Azure Data Factory | 30:57 |  |
| 75 | Use Azure Key Vault Secrets in Azure Data Factory | 19:30 |  |
| 76 | Continuous integration and deployment in Azure Data Factory | 51:53 |  |
| 77 | How to read JSON output of one Activity in to another Activity in Azure Data Factory | 15:20 |  |
| 78 | Annotations in Azure Data Factory | 9:10 |  |
| 79 | Templates Overview in Azure Data Factory | 8:17 |  |
| 80 | Global Parameters in Azure Data Factory | 15:13 |  |
| 81 | Rank Transformation in Mapping Data Flow in Azure Data Factory | 8:03 |  |
| 82 | Cache Sink and Cached lookup in Mapping Data Flow in Azure Data Factory | 25:30 |  |
| 83 | Session log in Copy Activity | Log Copied File names in Copy Activity in Azure Data Factory | 18:54 |  |
| 84 | Write Cache Sink to Activity Output in Azure Data factory | 13:30 |  |
| 85 | Parse Transformation in Mapping Data Flow in Azure Data Factory | 13:58 |  |
| 86 | Fail Activity in Azure Data Factory | 10:17 |  |
| 87 | Inline Dataset in Azure Data Factory | 8:53 |  |
| 88 | Stringify transformation in Mapping Data Flow in Azure Data Factory | 10:30 |  |
| 89 | Assert Transformation in Mapping Data Flows in Azure Data Factory | 16:54 |  |
| 90 | Flowlets in Mapping data flow in Azure Data Factory | 21:44 |  |
| 91 | Script Activity in Azure Data Factory or Azure Synapse | 15:32 |  |
| 92 | User defined Functions in Mapping data flows in Azure Data Factory | 9:14 |  |
| 93 | Fuzzy Joins Using mapping data flows in Azure Data Factory or Synapse Pipeline | 9:33 |  |
| 94 | Parameterize Linked Services using Advanced section in Azure Data Factory | 13:12 |  |
| 95 | Cast Transformation in Mapping data flows in Azure Data Factory or Azure Synapse Analytics | 10:10 |  |
| 96 | Extract Data from table of website page using Azure Data Factory | 11:41 |  |
| 97 | Per Pipeline Billing View for Azure Data factory | 5:00 |  |
| 98 | Time To Live(TTL) Setting in Azure IR to reduce cluster spin up time for dataflows | 10:55 |  |
| 99 | Create Alert rules in Azure Data factory for Pipeline or activity Failures | 11:15 |  |
| 100 | Pipeline return value in Set variable in Azure Data Factory & Azure Synapse Analytics | 11:32 |  |
| 101 | Copy activity - Pagination rules - When API response have URL for next page | 17:19 |  |
| 102 | Copy activity - Pagination rules - Variables in QueryParameters in Azure Data Factory | 13:28 |  |
| 103 | CDC(change data capture) for SQL Source in Mapping data flows in Azure Data Factory or Synapse | 27:01 |  |
| 104 | CDC (change data capture) Resource in Azure Data Factory | 24:44 |  |
| 105 | Deactivate an Activity in Azure Data Factory | 9:08 |  |
| 106 | Managed Virtual Integration Runtime in Azure Data Factory | 17:37 |  |
| 107 | TTL in Managed Vnet IR in Azure Data Factory or Azure Synapse Analytics | 14:42 |  |

1. Azure Databricks : 20 hrs

|  |  |  |  |
| --- | --- | --- | --- |
| Sr. No. | Name | Length | Status |
| 1 | Databricks: Spark Architecture & Internal Working Mechanism | 41:34 |  |
| 2 | Databricks | PySpark: RDD, Dataframe and Dataset | 12:41 |  |
| 3 | Databricks | PySpark: Transformation and Action | 16:15 |  |
| 4 | On-Heap vs Off-Heap| Databricks | Spark | Interview Question | Performance Tuning | 11:56 |  |
| 5 | Databricks | Pyspark: Cluster Deployment | 15:08 |  |
| 6 | Databricks | Pyspark| Spark Reader: Read CSV File | 17:06 |  |
| 7 | Databricks | Pyspark: Filter Condition | 14:28 |  |
| 8 | Databricks | Pyspark: Add, Rename and Drop Columns | 12:03 |  |
| 9 | Databricks | PySpark Join Types | 14:28 |  |
| 10 | Databricks | Pyspark: Utility Commands - DBUtils | 28:43 |  |
| 11 | Databricks | Pyspark: Explode Function | 15:24 |  |
| 12 | Databricks | Pyspark: Case Function (When.Otherwise ) | 13:34 |  |
| 13 | Databricks | Pyspark: Union & UnionAll | 10:25 |  |
| 14 | Databricks | Pyspark: Pivot & Unpivot | 11:33 |  |
| 15 | Databricks| Spark | Pyspark | Read Json| Flatten Json | 9:35 |  |
| 16 | Databricks | Spark | Pyspark | Bad Records Handling | Permissive;DropMalformed;FailFast | 7:24 |  |
| 17 | Databricks & Pyspark: Azure Data Lake Storage Integration with Databricks | 14:43 |  |
| 18 | Databricks & Pyspark: Ingest Data from Azure SQL Database | 12:08 |  |
| 19 | Databricks & Pyspark: Real Time ETL Pipeline Azure SQL to ADLS | 17:04 |  |
| 20 | Databricks & Pyspark: Azure Key Vault Integration | 9:32 |  |
| 21 | Databricks| Spark Streaming | 18:12 |  |
| 22 | Databricks| Spark | Performance Optimization | Repartition vs Coalesce | 21:11 |  |
| 23 | Databricks | Spark | Cache vs Persist | Interview Question | Performance Tuning | 18:56 |  |
| 24 | Databricks| Spark | Interview Questions| Catalyst Optimizer | 19:42 |  |
| 25 | Databricks | Spark | Broadcast Variable| Interview Question | Performance Tuning | 13:33 |  |
| 26 | Databricks | Spark | Adaptive Query Execution| Interview Question | Performance Tuning | 17:14 |  |
| 27 | Databricks Pyspark: Handling Null - Part1 | 10:05 |  |
| 28 | Databricks| Pyspark| Handling Null Part 2 | 14:17 |  |
| 29 | Databricks | Spark | Pyspark | UDF | 10:08 |  |
| 30 | Databricks - Spark: Data Skew Optimization | 15:03 |  |
| 31 | Databricks & Spark: Interview Question - Shuffle Partition | 5:52 |  |
| 32 | Databricks: Autoscaling | Optimized Autoscaling | 6:00 |  |
| 33 | Databricks | Pyspark: Dataframe Checkpoint | 8:14 |  |
| 34 | Databricks | Pyspark | Interview Question | Compression Methods: Snappy vs Gzip | 10:30 |  |
| 35 | Databricks | Spark | Pyspark Functions| Split | 10:41 |  |
| 36 | Databricks | Spark | Pyspark Functions| Arrays\_zip | 17:14 |  |
| 37 | Databricks | Spark | Pyspark Functions| Part 2 : Array\_Intersect | 5:11 |  |
| 38 | Databricks | Spark | Pyspark Functions| Part 3 : Array\_Except | 4:48 |  |
| 39 | Databricks | Spark | Pyspark Functions| Part 4 : Array\_Sort | 4:12 |  |
| 40 | Databricks | Spark | Python Functions| Join | 7:24 |  |
| 41 | Databricks | Spark | Pyspark | PartitionBy | 13:09 |  |
| 42 | Databricks | Spark | Pyspark | Number of Records per Partition in Dataframe | 5:53 |  |
| 43 | Databricks | Spark | Pyspark | Null Count of Each Column in Dataframe | 2:59 |  |
| 44 | Databricks - Pyspark: Find Top or Bottom N Rows per Group | 9:08 |  |
| 45 | Databricks & Spark: Interview Question(Scenario Based) - How many spark jobs get created? | 6:01 |  |
| 46 | Databricks | Pyspark: Greatest vs Least vs Max vs Min | 6:56 |  |
| 47 | Databricks | Pyspark | Delta Lake: Introduction to Delta Lake | 10:27 |  |
| 48 | Databricks| Pyspark| Delta Lake Architecture: Internal Working Mechanism | 30:13 |  |
| 49 | Databricks| Pyspark| Delta Lake: Solution Architecture | 7:53 |  |
| 50 | Databricks | Delta Lake| Pyspark: Create Delta Table Using Various Methods | 11:56 |  |
| 51 | Databricks| Pyspark| Delta Lake: Delta Table Instance | 11:29 |  |
| 52 | Databricks| Pyspark | Delta Lake: Different Approaches to Insert Data Into Delta Table | 10:47 |  |
| 53 | Databricks| Pyspark| Delta Lake: Different Approaches to Delete Data from Delta Table | 8:44 |  |
| 54 | Databricks | Pyspark | Delta Lake : Update Delta Table | 7:49 |  |
| 55 | Databricks Pyspark:Slowly Changing Dimension|SCD Type1| Merge using Pyspark and Spark SQL | 11:30 |  |
| 56 | Databricks & Pyspark: Delta Lake Audit Log Table with Operation Metrics | 17:12 |  |
| 57 | Databricks | Pyspark | Delta Lake : Slowly Changing Dimension (SCD Type2) | 20:03 |  |
| 58 | Databricks | Pyspark | Delta Lake: Time Travel | 8:47 |  |
| 59 | Databricks | Pyspark| Delta Lake: Restore Command | 7:28 |  |
| 60 | Databricks | Pyspark | Delta Lake: Optimize Command - File Compaction | 13:16 |  |
| 61 | Databricks | Pyspark | Delta Lake: Vacuum Command | 15:32 |  |
| 62 | Databricks | Pyspark | Delta: Z-Order Command | 14:16 |  |
| 63 | Databricks | Pypark | Delta: Schema Evolution - MergeSchema | 7:53 |  |
| 64 | Databricks | Pyspark | Dataframe InsertInto Delta Table | 8:09 |  |
| 65 | Databricks | Spark | Pyspark | Data Skewness| Interview Question: SPARK\_PARTITION\_ID | 7:02 |  |
| 66 | Databricks| Pyspark| Input\_File\_Name: Identify Input File Name of Corrupt Record | 10:47 |  |
| 67 | Databricks | Pyspark | Window Functions: Lead and Lag | 15:04 |  |
| 68 | Databricks | Pyspark | Interview Question: Explain Plan | 27:27 |  |
| 69 | Databricks | Pyspark | UDF to Check if Folder Exists | 6:45 |  |
| 70 | Databricks | Pyspark | Interview Question: Sort-Merge Join (SMJ) | 16:46 |  |
| 71 | Databricks | Pyspark | Performance Optimization - Bucketing | 22:03 |  |
| 72 | Databricks|Pyspark:Interview Question|Scenario Based|Max Over () Get Max value of Duplicate Data | 8:27 |  |
| 73 | Databricks | Pyspark | Create\_map(): Convert Dataframe Columns to Dictionary (Map Type) | 8:34 |  |
| 74 | Databricks | Pyspark | Performance Optimization: Delta Cache | 7:47 |  |
| 75 | Databricks | Pyspark | Split Array Elements into Separate Columns | 9:56 |  |
| 76 | Databricks | Pyspark | Tips: Write Dataframe into Single File with Specific File Name | 12:09 |  |
| 77 | Databricks | Pyspark | Workspace Object Access Control | 8:52 |  |
| 78 | Databricks | Pyspark | Databricks Secret Scopes: Azure Key Vault Backed Secrets | 19:12 |  |
| 79 | Databricks | Pyspark | Databricks Workflows: Job Scheduling | 17:17 |  |
| 80 | Databricks | Pyspark | Azure Data Factory + Azure Databricks: Execute Notebook Via ADF | 13:23 |  |
| 81 | Databricks | Pyspark | Notebook Activity in Azure Data Factory with Input Parameter | 10:27 |  |
| 82 | Databricks | Pyspark | Notebook Activity in Azure Data Factory with Output Parameter | 8:35 |  |
| 83 | Databricks | Pyspark | Real Time Project: ETL Pipeline Integrating ADF, ASQL, ADLS, Key Vault | 22:01 |  |
| 84 | Databricks |Pyspark |Notebook Scheduling through Schedule Based Trigger using Azure Data Factory | 13:11 |  |
| 85 | Databricks | Pyspark | Notebook Scheduling through Event Based Trigger using Azure Data Factory | 15:35 |  |
| 86 | Databricks | Pyspark | Interview Question: Read Excel File with Multiple Sheets | 10:13 |  |
| 87 | Databricks | Pyspark | Interview Question |Handlining Duplicate Data: DropDuplicates vs Distinct | 11:41 |  |
| 88 | Databricks | Pyspark | Interview Question | Performance Optimization: Select vs WithColumn | 11:33 |  |
| 89 | Databricks | Pyspark | Interview Question | Schema Definition: Struct Type vs Struct Field | 15:40 |  |
| 90 | Databricks | Pyspark | Interview Question | Schema Definition: Struct Type vs Map Type | 13:22 |  |
| 91 | Databricks | Pyspark | Schema | Different Methods of Schema Definition | 15:32 |  |
| 92 | Databricks | Pyspark | Real Time Scenario | Schema Comparison | 12:34 |  |
| 93 | Databricks | Pyspark | Data Security: Enforcing Column Level Encryption | 11:48 |  |
| 94 | Databricks | Pyspark | Interview Question: Pyspark VS Pandas | 9:09 |  |
| 95 | Databricks | Pyspark | Real Time Use Case: Generate Test Data - Array\_Repeat() | 13:52 |  |
| 96 | Databricks | Pyspark | Spark Architecture: Internals of Partition Creation Demystified | 55:50 |  |
| 97 | Databricks|Pyspark|Automation|Real Time Project:DataType Issue When Writing to Azure Synapse/SQL | 14:06 |  |
| 98 | Databricks | Pyspark| Transformation: Subtract vs ExceptAll | 8:37 |  |
| 99 | Databricks | Pyspark| Window Function: First and Last | 12:27 |  |
| 100 | Databricks | Pyspark| Spark Reader: Skip First N Records While Reading CSV File | 6:31 |  |
| 101 | Databricks | PySpark| Spark Reader: Skip Specific Range of Records While Reading CSV File | 13:19 |  |
| 102 | Databricks | Pyspark| Query Dataframe Using Spark SQL | 10:46 |  |
| 103 | Databricks | Pyspark| Spark SQL: Except Columns in Select Clause | 8:54 |  |
| 104 | Databricks | Pyspark| AutoLoader: Incremental Data Load | 34:56 |  |
| 105 | Databricks | Pyspark| Delta Live Table: Introduction | 24:25 |  |
| 106 | Databricks | Pyspark| Delta Live Table: Declarative VS Procedural | 10:17 |  |
| 107 | Databricks | Pyspark| Delta Live Table: Datasets - Tables and Views | 15:09 |  |
| 108 | Databricks | Pyspark| Delta Live Table: Data Quality Check - Expect | 8:16 |  |
| 109 | Databricks | Pyspark | Downloading Files from Databricks DBFS Location | 8:00 |  |
| 110 | Databricks | Pyspark| Built-In Function: TRANSFORM | 15:36 |  |
| 111 | Databricks | Pyspark| Delta Lake: Deletion Vectors | 25:03 |  |
| 112 | Databricks | Pyspark| Delta Lake: Change Data Feed | 17:26 |  |
| 113 | Databricks | Pyspark| Built-in Function: ZIP\_WITH | 11:54 |  |
| 114 | DataBricks Learning: System Variable \_SQLDF | 8:46 |  |

**Steps to perform the learning:**

SSIS:

1. Introduction to SSIS | Create your first SSIS package

-> SSIS Control Flow Level ToolBox:

1. **Data Flow Task:** The Data Flow Task is a core SSIS component used to move and transform data from one or more sources to one or more destinations. It defines a workflow where data is extracted, transformed, and loaded (ETL).

Use Case: Use a Data Flow Task when you need to move data between different systems while applying data cleansing, transforming, or enrichment rules during the transfer.

2. **Execute SQL Task:** The Execute SQL Task is used to run SQL queries or ` statements against a relational database. This includes SELECT, INSERT, UPDATE, DELETE, DDL statements, or calling stored procedures.

Use Case: Use the Execute SQL Task when you need to interact with a database to perform data manipulation or administrative operations, such as creating tables, updating records, or retrieving configuration values before running ETL processes.

3. Analysis Services Processing Task: The Analysis Services Processing Tasks

4. Bulk Insert Task

5. Data Profiling Task

6. Execute Package Task

7. Execute Process Task

8. FTP Task

9. Hadoop File System Task

10. Hadoop hive Task

11. Hadoop Pig Task

12. Script Task

13. Send Mail Task

14. Web Service Task

15. XML Task

16. For Loop Container

17. Foreach Loop Container

18. Sequence Container

19. Analysis Services Execute DDL Task

20. Back Up Database Task

21. CDC Control Task

22. Check Database Integrity Task

23. Execute SQL Server Agent Job Task

24. Execute T-SQL Statement Task

25. History Cleanup Task

26. Maintanence Cleanup Task

27. Message Queue Task

28. Notify Operator Task

29. Rebuilt Index Task

30. Reorganize Index Task

31. Shrink Database Task

32. Transfer Database Task

33. Transfer Error Messages Task

34. Transfer Jobs Task

35. Transfer Logins Task

36. Transfer Master Stored Procedures Task

37. Transfer SQL Server Objects Task

38. Update Statistics Task

39. WMI Data Reader Task

40, WMI Event Watcher Task

-> SSIS Data Flow Level Components:

1. Destination Assistant

2. Source Assistant

3. Aggregate

4. Balanced Data Distributor

5. Conditional Split

6. Data Conversion

7. Data Streaming Destination

8. Derived Columns

9. HDFS File Destination

10. HDFS File Source

11. Lookup

12. Merge

13. Merge Join

14. Multicast

15. OData Source

16. ODBC Destination

17. ODBC Source

18. OLE DB Command

19. Row Count

20. Script Component

21. Slowly Changing Dimension

22. Sort

23. Union All

24. Audit

25. Cache Transform

26. CDC Splitter

27. Character Map

28. Copy Column

29. DQS Cleansing

30. Export Column

31. Fuzzy Grouping

32. Fuzzy Lookup

33. Import Column

34. Percentage Sampling

35. Pivot

36. Row Sampling

37. Term Extraction

38. Term Lookup

39. Unpivot

40. ADO NET Source

41. CDC Source

42. Excel Source

43. Flat File Source

44. OLE DB Source

45. Raw File Source

46. XML Source

47. ADO NET Destination

48. DataReader Destination

49. Dimension Processing

50. Excel Destination

51. Flat File Destination

52. OLE DB Destination

53. Partition Processing

54. Raw File Destination

55. Recordset Destination

56. SQL Server Destination

2 SSIS Tutorial for beginners

3 Install Visual Studio 2019 Community | SSIS 2019 Installation | Integration services 2019

4 Install SQL Server 2019 with custom option | Install SQL Server 2019

5 Install SQL Server 2019

6 Import csv file to sql server using SSIS | Load CSV File in SSIS

7 Load Fixed Width File in SQL Server - SSIS | Import Fixed width file in SSIS

8 Load Excel File in SQL Server - SSIS | Import excel file in SQL Server using SSIS

9 Overwrite data to Excel file in SSIS | Export data to excel file using SSIS

10 Load XML File in SQL Server - SSIS | Import XML file in sql server using SSIS

11 Copy data from One SQL Sever Instance to another - SSIS

12 Export data from SQL to CSV File - SSIS | Export data to csv file in SSIS

13 Export data from SQL to Fixed Width File - SSIS | export data to fixed file in SSIS

14 Export data from Sql Server to Excel using SSIS

15 Execute SQL Task in SSIS | Use variables in Execute SQL task in SSIS

16 File System Task in SSIS | How to use File System Task in SSIS

17 Zip and UnZip files in SSIS

18 Script Task in SSIS | Using script task in SSIS

19 Debugging Script Task in SSIS

20 Sequence container in SSIS example

21 For loop Container in SSIS

22 Foreach Loop Container in SSIS Example

23 Foreach loop ado enumerator in SSIS example

24 Foreach loop Item Enumerator in SSIS Example

25 Load multiple excel sheets using SSIS

26 Aggregate Transformation in SSIS

27 Row Count Transformation in SSIS | RowCount Transformation in SSIS

28 Data Conversion Transformation in SSIS

29 Character Map Transformation in SSIS

30 Copy Column Transformation in SSIS

31 Derived Column Transformation in SSIS

32 Multicast Transformation in SSIS

33 Conditional Split Transformation in SSIS

34 Union All Transformation in SSIS

35 Merge Transformation in SSIS

36 Merge Join Transformation in SSIS

37 Script Component Transformation in SSIS

38 OLE DB Command Transformation in SSIS | OLEDB Command Transformation

39 Lookup Transformation in SSIS

40 Fuzzy Lookup Transformation in SSIS

41 Fuzzy Grouping Transformation in SSIS

42 Slowly Changing Dimension in SSIS

43 Error handling in SSIS | How to handle errors in SSIS

44 Execute Package Task in SSIS

45 Implement Logging to SQL Server in SSIS

46 Implement Logging to Flat File in SSIS

47 Send Email in SSIS

48 Send HTML Email in SSIS

49 Pass value from Parent Package to Child Package in SSIS

50 Execute SSIS Package from command line | Execute SSIS Package from batch file

51 Import Export Wizard in SQL Server | Import Export Wizard

52 Execute package utility in SSIS | how to execute ssis package from dtexecui

53 SSIS Package deployment step by step | Deploy SSIS package | Deploy SSIS project to ssis catalog

54 Converting from Project Deployment to Package Deployment

55 XML Configuration in SSIS

56 SQL Server Configurations in SSIS | SSIS configuration sql server

57 Checkpoints in SSIS with example

58 Load Multiple Excel Sheets with different schema

59 Breakpoints in SSIS | how to use breakpoints in ssis

60 Data Viewer in SSIS | SSIS Data Viewer

61 Schedule SSIS package in SQL Server agent | create sql job to run ssis package

62 Schedule an SSIS package using Windows Task Scheduler

63 Insert update differential data in SSIS

64 Create SQL Job from SSIS Catalog Package

65 How do I stop Package Execution in SSIS catalog

66 Precedence Constraints in SSIS

67 SSIS Transactions | Sequence Container in SSIS

68 Project Deployment in SSIS and Create environment

69 How do I debug an SSIS package?

70 Foreach loop item enumerator in ssis example

71 What are expressions in SSIS ? How to write expressions ?

72 How to write expressions in SSIS ?

73 How to get FileName of all different types of Files in a Folder and write to a SQL table ?

74 How to unzip multiple zip files in SSIS ?

75 How to delete duplicate records in SSIS

76 What are variables and Parameters in SSIS

77 How to execute multiple child packages from parent package

78 Synchronous vs Asynchronous Transformations in SSIS

79 Delay Validation in SSIS

80 Load current date file into SQL server table using SSIS

81 Audit Transformation in SSIS

82 Send Mail Task in SSIS | Send Notifications using Send Mail Task in SSIS

83 Pivot in SSIS | Pivot Transformation in SSIS

84 How to use sql authentication connection in SSIS | Make the connection dynamic using variables

85 How to use Windows authentication and sql authentication in a single connection in SSIS

86 How to Execute SSIS packages in parallel | Execute Tasks in Parallel in SSIS

87 How to import dbf files into SQL Server using SSIS

88 How to read data from Stored Procedure and insert into SQL table using SSIS

89 How to read data from Stored Procedure and pass input parameter from a variable using SSIS

90 How to pass the SSIS variable value from SQL Agent Job

91 How to skip rows in flat file in SSIS | How to skip rows in CSV file in SSIS

92 How to Execute SQL Queries which are stored on disk using SSIS

93 How to convert Y and N to 1 and 0 using SSIS expressions

94 How to move files from one folder to another using SSIS | File System Task in SSIS

95 How to import a csv file into SQL Server using bulk insert SQL query

96 Deploy SSIS package using manifest file | Deploy SSIS package to File System

97 How to fix Connecting to Integration Services Access is denied error

98 how to deploy SSIS package to SQL server | how to deploy SSIS package to msdb

99 How to export data to Pipe delimited format using SSIS

100 Export data from multiple databases to CSV files using SSIS

101 export data to csv in sql server | sql server data export to csv using bcp

102 bulk insert task in ssis

103 Web Service task in SSIS | How to use Web service task in SSIS

104 consume web service from C#

105 how to use global temp table in ssis | how to use temp table in ssis

106 Execute SSIS package from Powershell

107 How to load json file in SSIS | Import json file to SQL Server using SSIS

108 How to load data from CSV file that has extra delimiter | delimiter for column "X" is not found

109 Event Handlers in SSIS | How to use Event Handlers in SSIS

110 How to check if File exists in SSIS

111 Invalid character value for cast specification in SSIS

112 The value violated the integrity constraints for the column

113 Removing special characters from data in SSIS

114 Convert data to proper case in SSIS | How do you convert a string to a proper case?

115 How to skip rows to load from excel SSIS

116 Execute Process Task in SSIS

117 Execute SQL Task in SSIS result set

118 Merge join transformation in SSIS with example

119 how can we load that special char rows into error table

120 How do you implement SCD Type 2 in SSIS Using Lookup

121 Replace null values with blank in SSIS

122 How do you write a case statement in a derived column in SSIS

123 Replace column value in SSIS | Replace a string with new string in SSIS

124 data conversion failed while converting column status value 4 in SSIS

125 Send email on SSIS package failure | Send Error description from SSIS package

126 What are third party components in SSIS

127 How to import a csv file into SQL server using C# or SSIS

128 How to load dimension table in sql using ssis

129 How do you load a fact table

130 FTP Task in SSIS

131 How do you upload multiple files to ftp server in SSIS

132 How to download files from sftp server using SSIS or C#

133 How to upload files to sftp server using SSIS or C#

134 How to import data from xml to sql server using C#

135 Percentage Sampling Transformation in SSIS

136 Row Sampling Transformation in SSIS

137 Export data to csv file with header using bcp command

138 How to create dynamic connection manager in ssis

139 What are different flat file formats in ssis

140 How to encrypt password in ssis package | How to save password in sql connection

141 Cache transformation in ssis

142 How to use lookup transformation with cache file in ssis

143 How to import multiple excel files into sql server using ssis

144 How to handle errors in script task in ssis

145 How do I redirect error rows to another table in ssis | Error handling at data flow in SSIS

146 How to import multiple xml files to sql server using ssis

147 How to store File Path into sql server table using ssis

148 How to do incremental load to mysql from sql server using ssis

149 data from excel file is getting inserted as float in sql in ssis

150 How to move files to month wise folders in ssis

151 How to merge data from multiple flat files using ssis

152 The connection string components cannot contain unquoted semicolons in SSIS

Azure Data Factory:

1 Introduction to Azure Data Factory

2 Top level Concepts in Azure Data Factory

3 Create your First Azure Data Factory

4 Different ways to work with Azure Data Factory

5 Pipelines and Activities in Azure Data Factory

6 Linked Services and Datasets in Azure Data Factory

7 Triggers in Azure Data Factory

8 Schedule Trigger in Azure Data Factory

9 Tumbling Window Trigger in Azure Data Factory

10 Tumbling Window Trigger Dependency in Azure Data Factory

11 Event based Triggers in Azure Data Factory

12 Integration runtime in Azure Data Factory

13 Azure Integration runtime in Azure Data Factory

14 Self Hosted Integration runtime in Azure Data Factory

15 Setting up Self Hosted Integration runtime in Azure Data Factory

16 Shared Self Hosted Integration runtime in Azure Data Factory

17 Parameterize Linked Services in Azure Data Factory

18 Parameterize Datasets in Azure Data Factory

19 Parameterize Pipelines in Azure Data Factory

20 System Variables in Azure Data Factory

21 Connectors Overview in Azure Data Factory

22 Supported File Formats in Azure Data Factory

23 Copy Data Activity in Azure Data Factory

24 Copy Data Activity in Azure data factory Continuation

25 Monitor Copy Data Activity in Azure Data Factory

26 Delete Activity in Azure Data Factory

27 Variables in Azure Data Factory

28 Set Variable Activity in Azure Data Factory

29 Append Variable Activity in Azure Data Factory

30 User Properties in Azure Data Factory

31 Execute Pipeline Activity in Azure Data Factory

32 Filter Activity in Azure Data Factory

33 ForEach Activity in Azure Data Factory

34 Get Metadata Activity in Azure Data Factory

35 If Condition Activity in Azure Data Factory

36 Wait Activity in Azure Data Factory

37 Until Activity in Azure Data Factory

38 Web Activity in Azure Data Factory

39 WebHook Activity in Azure Data Factory

40 Switch Activity in Azure Data Factory

41 Validation Activity in Azure Data Factory

42 Lookup Activity in Azure Data Factory

43 Transform Data Activities Overview in Azure Data Factory

44 Stored Procedure Activity in Azure Data Factory

45 Data flow in Azure data factory

46 Mapping Data Flow in Azure Data Factory

47 Data Flow Activity in Azure Data Factory

48 Mapping Data Flow Debug Mode in Azure Data Factory

49 Filter Transformation in Mapping Data Flow in Azure Data Factory

50 Aggregate Transformation in Mapping Data Flow in Azure Data Factory

51 JOIN Transformation in Mapping Data Flow in Azure Data Factory

52 Conditional Split Transformation in Mapping Data Flow in Azure data factory

53 Derived Column Transformation in Mapping Data Flow in Azure Data Factory

54 Exists Transformation in Mapping Data Flow in Azure Data Factory

55 Union Transformation in Mapping Data Flow in Azure Data Factory

56 Lookup Transformation in Mapping Data Flow in Azure Data Factory

57 Sort Transformation in Mapping Data Flow in Azure Data Factory

58 New Branch in Mapping Data Flow in Azure Data Factory

59 Select Transformation in Mapping Data Flow in Azure Data Factory

60 Pivot Transformation in Mapping Data Flow in Azure Data Factory

61 Unpivot Transformation in Mapping Data Flow in Azure Data Factory

62 Surrogate Key Transformation in Mapping Data Flow in Azure Data Factory

63 Window Transformation in Mapping Data Flow in Azure Data Factory

64 Alter Row Transformation in Mapping Data Flow in Azure Data Factory

65 . Flatten Transformation in Mapping Data Flow in Azure Data Factory

66 Parameterize Mapping Data Flow in Azure Data Factory

67 Validate Schema in Mapping Data Flow in Azure Data Factory

68 Schema Drift in Mapping Data Flow in Azure Data Factory

69 Wrangling Data Flow Overview in Azure Data Factory

70 Merge Queries in Wrangling Data Flow in Azure Data Factory

71 Group By in Wrangling Data Flow in Azure Data Factory

72 Different Author Modes in Azure Data Factory

73 Setup GitHub Code Repository for Azure Data Factory

74 Setup Azure DevOps Git Code Repository for Azure Data Factory

75 Use Azure Key Vault Secrets in Azure Data Factory

76 Continuous integration and deployment in Azure Data Factory

77 How to read JSON output of one Activity in to another Activity in Azure Data Factory

78 Annotations in Azure Data Factory

79 Templates Overview in Azure Data Factory

80 Global Parameters in Azure Data Factory

81 Rank Transformation in Mapping Data Flow in Azure Data Factory

82 Cache Sink and Cached lookup in Mapping Data Flow in Azure Data Factory

83 Session log in Copy Activity | Log Copied File names in Copy Activity in Azure Data Factory

84 Write Cache Sink to Activity Output in Azure Data factory

85 Parse Transformation in Mapping Data Flow in Azure Data Factory

86 Fail Activity in Azure Data Factory

87 Inline Dataset in Azure Data Factory

88 Stringify transformation in Mapping Data Flow in Azure Data Factory

89 Assert Transformation in Mapping Data Flows in Azure Data Factory

90 Flowlets in Mapping data flow in Azure Data Factory

91 Script Activity in Azure Data Factory or Azure Synapse

92 User defined Functions in Mapping data flows in Azure Data Factory

93 Fuzzy Joins Using mapping data flows in Azure Data Factory or Synapse Pipeline

94 Parameterize Linked Services using Advanced section in Azure Data Factory

95 Cast Transformation in Mapping data flows in Azure Data Factory or Azure Synapse Analytics

96 Extract Data from table of website page using Azure Data Factory

97 Per Pipeline Billing View for Azure Data factory

98 Time To Live(TTL) Setting in Azure IR to reduce cluster spin up time for dataflows

99 Create Alert rules in Azure Data factory for Pipeline or activity Failures

100 Pipeline return value in Set variable in Azure Data Factory & Azure Synapse Analytics

101 Copy activity - Pagination rules - When API response have URL for next page

102 Copy activity - Pagination rules - Variables in QueryParameters in Azure Data Factory

103 CDC(change data capture) for SQL Source in Mapping data flows in Azure Data Factory or Synapse

104 CDC (change data capture) Resource in Azure Data Factory

105 Deactivate an Activity in Azure Data Factory

106 Managed Virtual Integration Runtime in Azure Data Factory

107 TTL in Managed Vnet IR in Azure Data Factory or Azure Synapse Analytics

Azure Databricks:

1 Databricks: Spark Architecture & Internal Working Mechanism

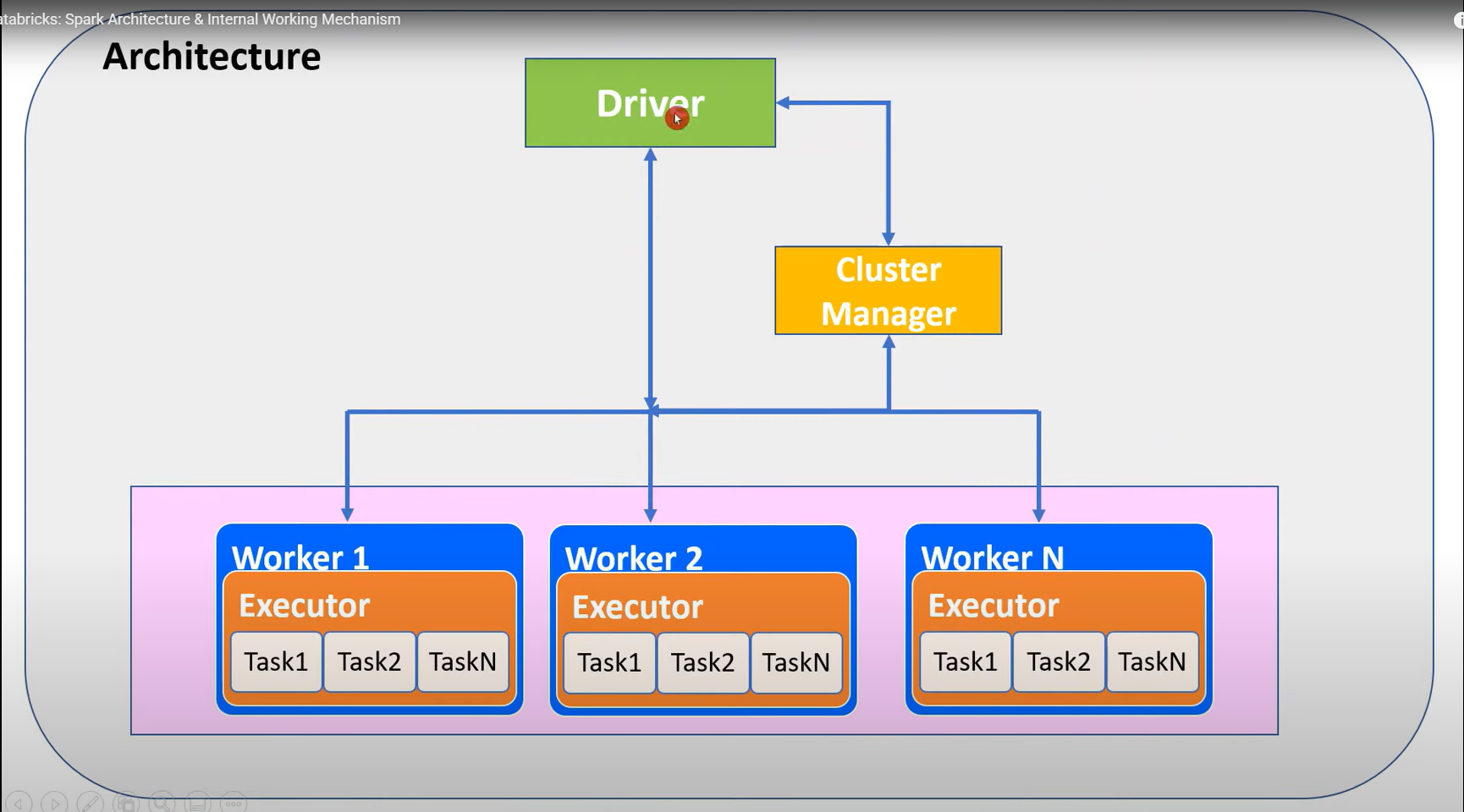
-> What is Spark:

-- Spark is an open source distributed computing engine. We use it for processing and analysing a large amount of data. Like Hadoop, Spark also works in distributed computing nature but differs in in-memory processing.

-- Compared to Hadoop or traditional systems, 100 times faster in memory and 10 times faster in disc.

-- How it handles the processes at lightning speed?- Its lightning speed is from in-memory and parallel processing

-- Spark has well-designed layered architecture. It follows master/slave concept. That is driver and worker concept. Between driver and worker layer comes Cluster manager layer. These layers driver, worker and cluster manager and designed well within its boundary and loosely coupled to each other.



Flow:

* Driver node demands cluster manager to give resources to driver node…the cluster manager will create worker nodes to provide to driver node.
* Driver node will directly make connection with worker node after worker nodes are provided to driver node by cluster manager.
* Cluster manager is responsible to manage the health of worker nodes.
* Driver node will directly assign the tasks to worker node and will be responsible to maintain the tasks.
* Worker nodes will execute the piece of logic using executor and in return it will give the result jar file to driver node.

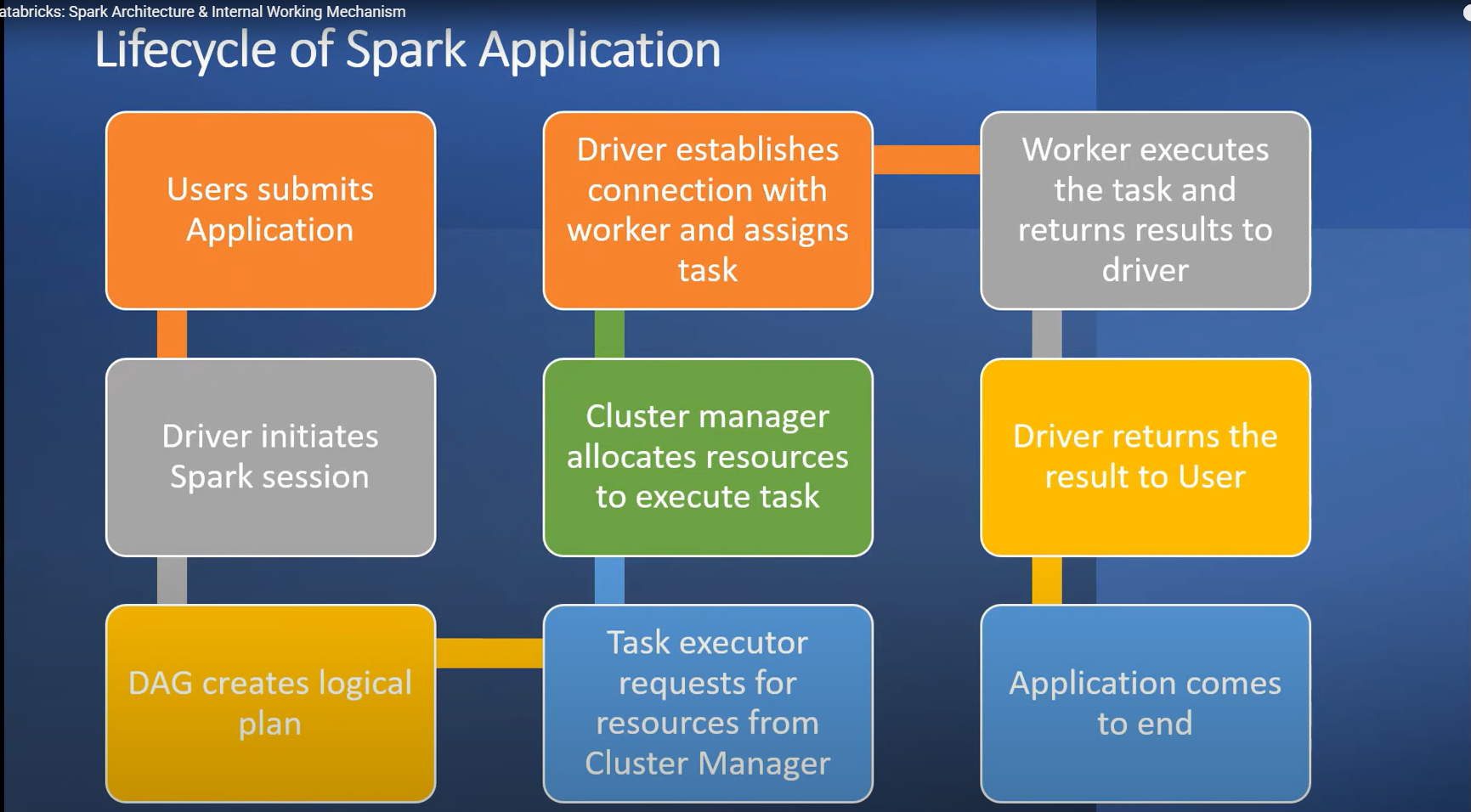
A screenshot of a computer

AI-generated content may be incorrect.

Worker Node Architecture :   
The whole memory is divided into 4 executor parts…the processor is divided equally amongst the executors equally – 4 – 32gb executors with 16-16 gb core processing capability to each. This is totally based on how we define or configure the spark settings to make the nodes.

Use case: If we have 2GB RDD size file to execute amongst a worker node, we will partition the whole file into 200 partitions and each partition will have 10 mb size, once the partitions are done, each executor will take 50-50 partitions each and each core will take 1-1 partition to perform the task and cycle runs until all 50 partitions allocated to it are completed to transform or load, whatever task is given to it.

Lifecycle of Spark Application:



* - Users submits Application: pyspark or spark command or multiple notebooks for execution
* - Driver initiates Spark session
* - DAG creates logical plan: Directed Acylcing Graph will create the logical plan for execution and keep the lineage plan…it will break the code into multiple chunks of task and will give it to task executor.
* - Task executor requests for resources from Cluster Manager: will request resources from the cluster manager
* - Cluster manager allocates resources to execute task: will allocate the worker nodes to driver nodes
* - Driver establishes connection with worker and assigns task: it will establish connection with worker nodes and start assigning tasks.
* - Worker executes the task and returns to driver: The driver will run the tasks using executors and pass the jar file back to the driver
* - Driver returns the result to user
* - Application comes to end

Spark follows Lazy Evolution Model: Spark will not immediately execute the plan, instead it will plan the whole thing out and then create a lineage graph and store in DAG. – Why this is used? So that if 1 worker node goes down – we can recover the work plan from the lineage graph given to DAG.

Spark Attributes:

- Scalable:

- Fault Tolerance: The Lineage Graph plan is used for fault tolerance

- Polyglot: Spark supports multiple languages in the spark also in notebook to support all types of developer to make use of the spark environment.

- Real time

- Speed: In-memory computational model , Distributed frameworks enables the parallel processing to higher degree.

- Rich Libraries  
  
Terminologies:

- Driver and worker process: These are nothing but JVM process. Within one worker node, there could be multiple executors. Each executor runs its own JVM process.

- Application: It could be single command or combination of multiple notebooks with complex logic. When code is submitted to spark for execution, Application starts.

- Jobs: When an application is submitted to Spark, driver process converts the code into job.

- Stage: Jobs are divided into stages. If the application code demands shuffling the data across nodes, new stage is created. Number of stages are determined by number of shuffling operations. Join is example of shuffling operation.

- Tasks: Stages are further divided into multiple tasks. In a stage, all the tasks would execute same logic. Each task will process 1 partition at a time. So number of partition in the distributed cluster determines the number of tasks in each stage.  
- Transformation: transforms the input RDD and creates new RDD. Until action is called, transformations are evaluated lazily.

- DAG: Directed Acyclic Graph keeps track of all transformations. For each transformation. logical plan is created and lineage graph is maintained by DAG.

- Action: When data output is needed for developer or for storage purpose, action is called. Action would be executed based on DAG and processes the actual data.

- RDD: Resilient Distributed Dataset is basic data structure of Spark. When spark reads or creates data, it created RDD which is distributed across nodes in the form of partition.

- Executor: Each worker node can consists of many executors. It can be configured by spark settings.

- Partition: RDD/Dataframe is stored in-memory of cluster in the form of partition

- Core: Each executor can consists of multiple cores. This is configurable by spark settings. Each core(if single thread) can process one task at a time.

- On-heap memory: The executor memory that lies within JVM process managed JVM

- Off-heap memory: The executor memory that lies outside JVM process managed by OS

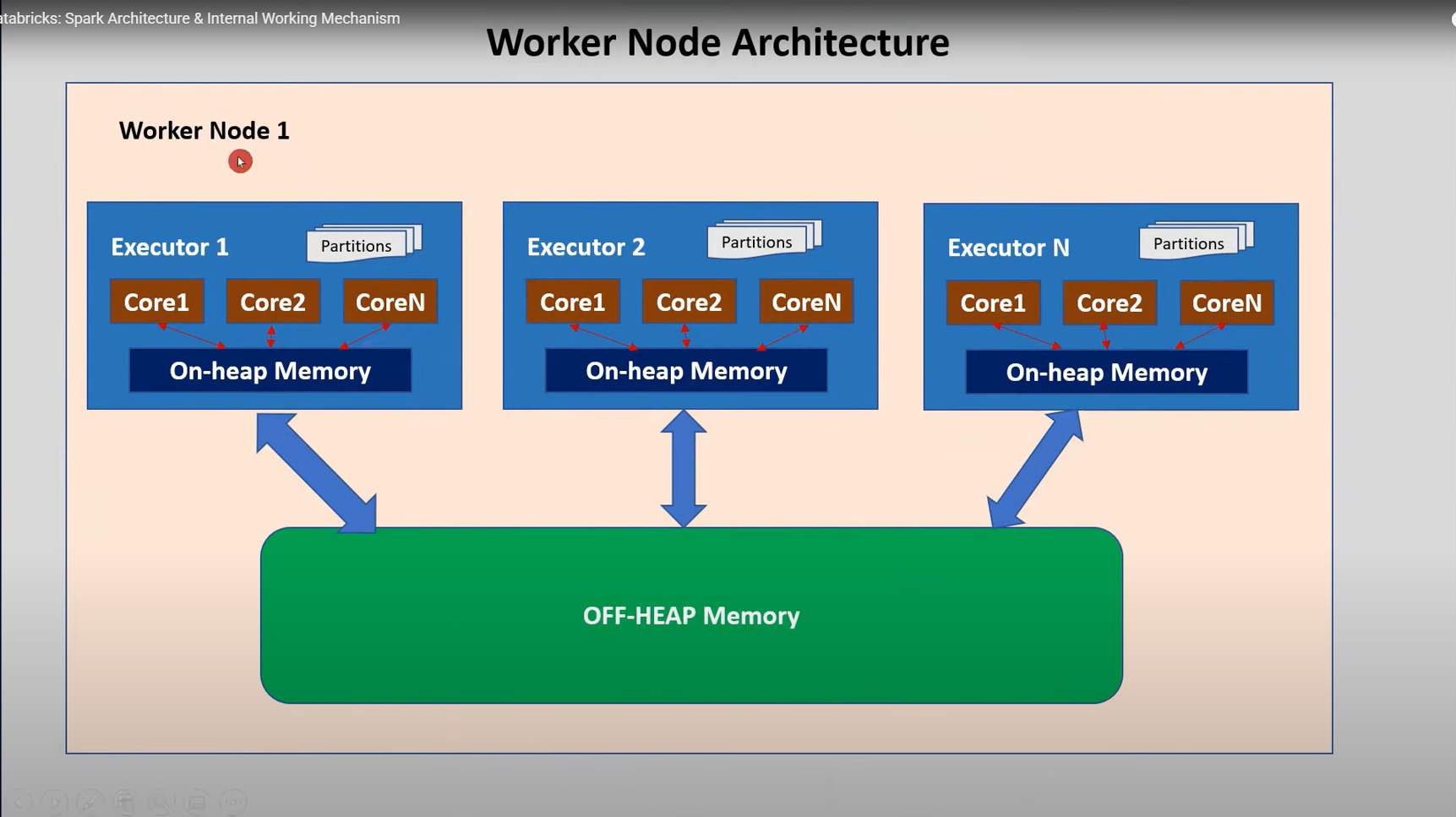
Libraries having massive support for databricks: SQL, Streaming, MLLib, GraphX

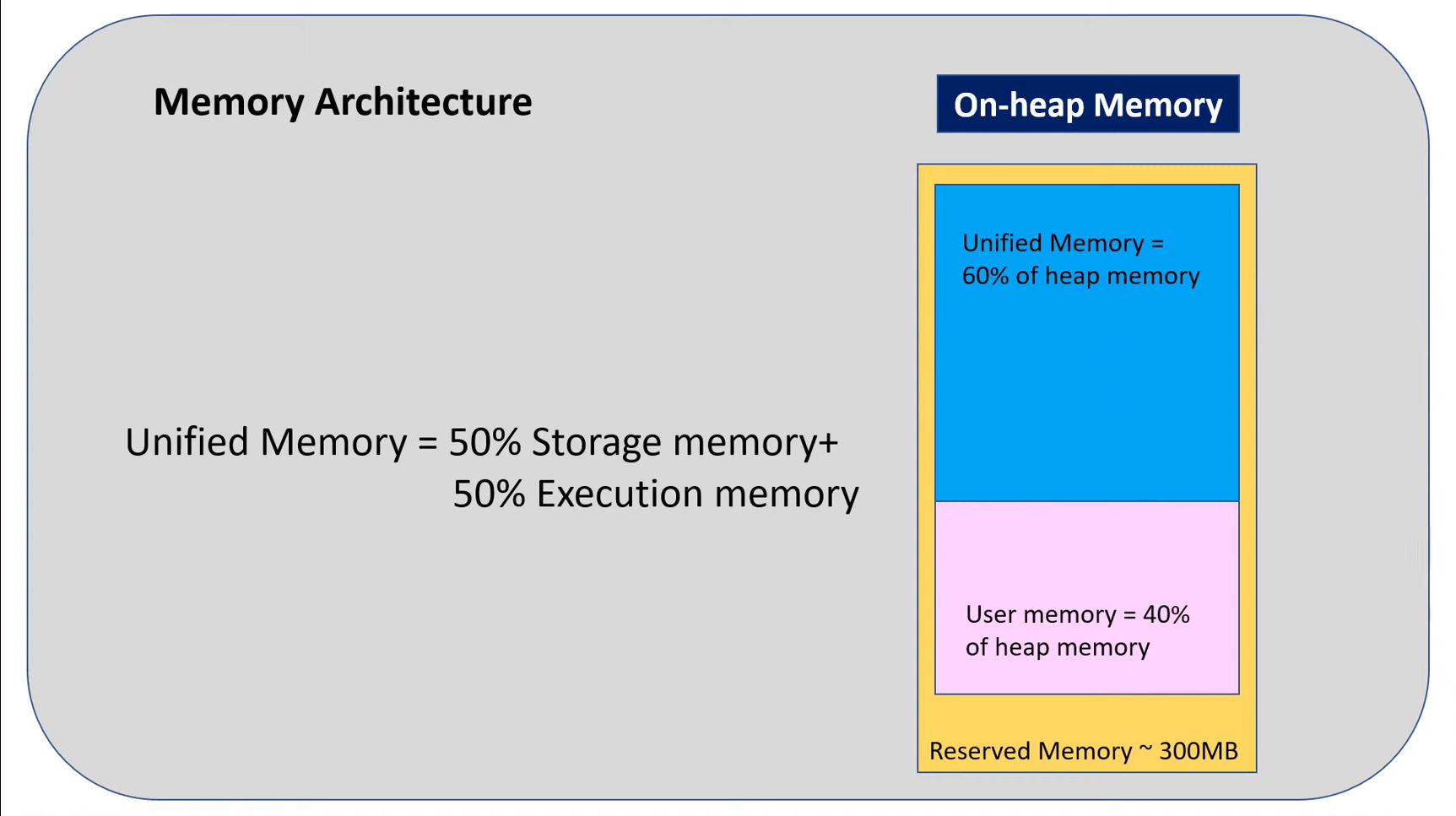
Languages usable in databricks: Scala, Java, Python, SQL, R

Driver Node Architecture:  
A computer screen shot of a computer

AI-generated content may be incorrect.

Worker Node Architecture:



Worker Node Memory Architecture:  


2 Databricks | PySpark: RDD, Dataframe and Dataset

-> They all are APIs provided by Spark for developers for data processing and analytics

-> In terms of functionality, all are same and returns same output for provided input data. But they differ in the way of handling and processing data. So, there is difference in terms of performance, user convenience and language support etc.

-> Users can choose any of the API while working with Spark.

-> Basic Histroy of RDD, Dataframe, Dataset:

RDD – Spark 1.0 -> 2011 (Features : Programming control OOPs, Type Safety)

Dataframe – Spark 1.3 -> 2013 (Features : Relational format, Optimization, Memory Management)

Dataset – Spark 1.6 -> 2015 (Features : All features of RDD + Dataframe)

Similarities:

- Fault tolerance

- Distributed in nature

- In-memory parallel processing

- Immutable

- Lazy evaluation

- Internally processing as RDDs for all APIs

Differences:

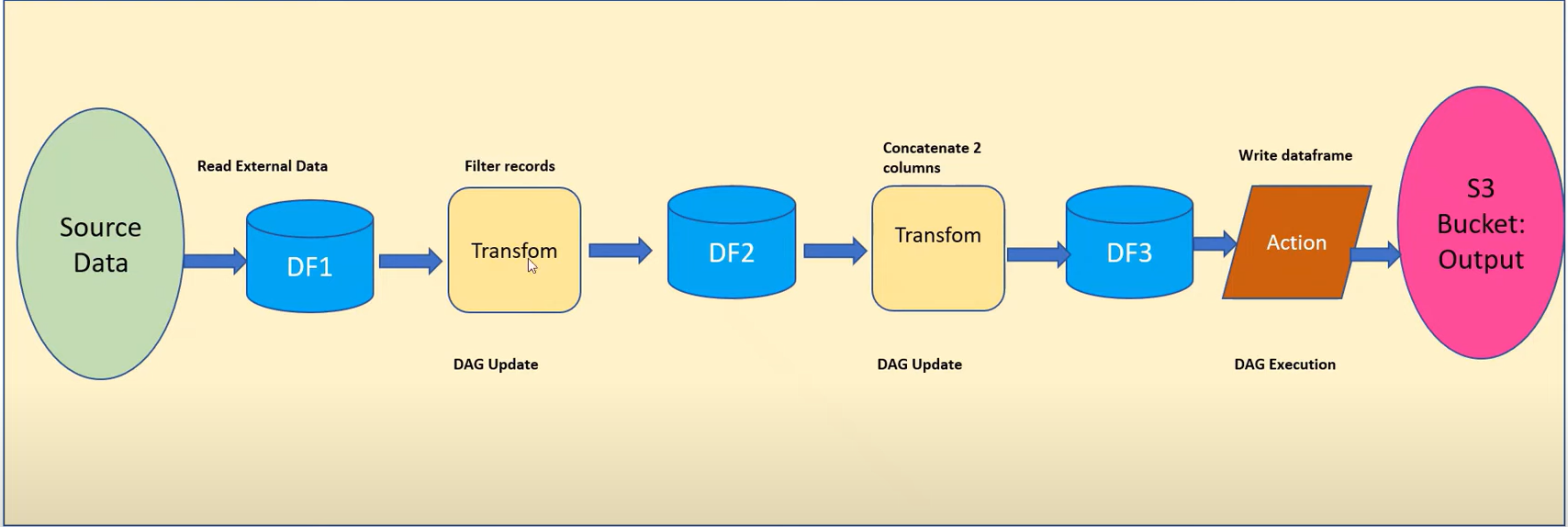
|  |  |  |  |
| --- | --- | --- | --- |
| Sr. No. | RDD | Dataframe | Dataset |
| 1 | Program have to do | Program what to do | Program what to do |
| 2 | OOPs Style API | SQL Style API | OOPs Style API |
| 3 | On-heap JVM objects | Off-heap also used | Off heap also used |
| 4 | Serialization unavoidable | Serialization can be avoided (off heap) | Serialization can be avoided(encoder) |
| 5 | GC impacts performance | GC impact mitigated | GC impact mitigated |
| 6 | Strong Type Safety | Less Type Safety | Strong type safety |
| 7 | No optimization | Catalyst Optimizer | Optimization |
| 8 | Compile time error | Run time error | Compile time error |
| 9 | Java, Scala, Python, and R | Java, Scala, Python, and R | Scala and Java |
| 10 | No schema | Schema Structured | Schema Structured |

3 Databricks | PySpark: Transformation and Action

-> Transformation: Transformation is kind of operation which will transform Dataframe from one form to another. We will get new Dataframe after execution of each transformation operation. Transformation is lazy evaluation based on DAG (Directed Acyclic Graph). Filter, Union etc.

-> Action: When we want to work with actual data, call the action. Action returns the data to driver for display or storage into storage layer. Count, Collect, Save. Etc.

Execution flow of Action and Transformation in Lazy Evaluation and DAG:



Flow:

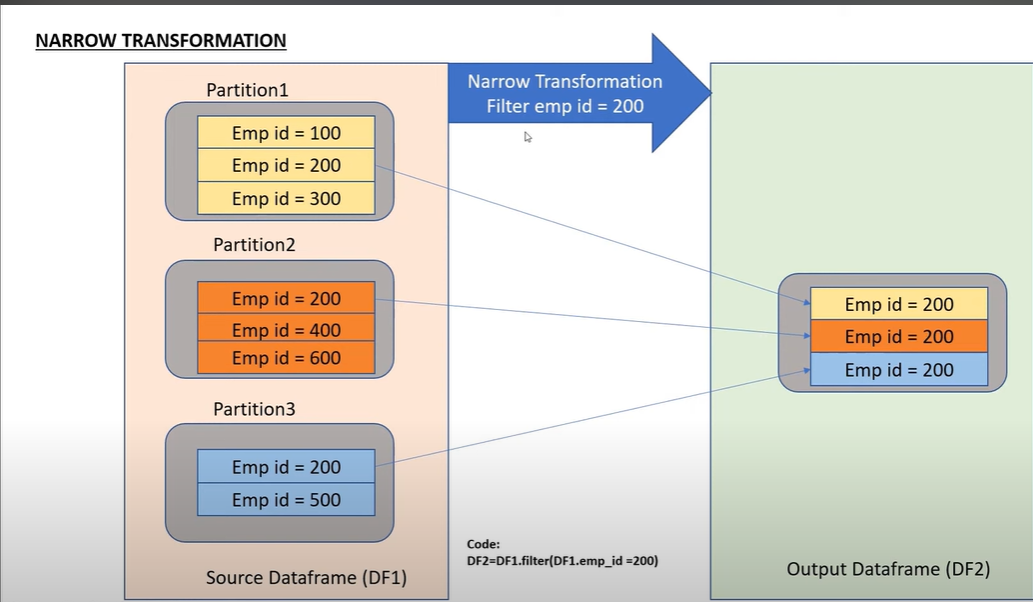
* First Source Data will get converted to DataFrame 1 (DF1)
* DF1 will provide instructions which needs to be performed and store the instructions in DAG i.e. will update DAG document
* If any other transformation needs to be performed, DAG will give an metadata of how data will look after transformation, and based on that it will create DF2
* Once DF2 is created, it will again store next set of instructions to be performed in DAG file
* The cycle will continue till Action is found to be performed in the flow.
* Once action is found, it will go to first set of DAG instructions and start execution, and from DF1 to DF2 the transformation will occur, same way all set of instructions written in DAG will be performed, until final Action step is found
* Once action step is found it will output the data as per the action asked.

Types of Transformation:

* Narrow Transformation – Cost efficient
* Wide Transformation – Too costly

Narrow Transformation:

Code Example: DF2 = DF1.filter(DF1.emp\_id = 200)



Wide Transformation:

Code Example: DF2 = DF1.groupBy(“Dept”).count()

A diagram of a group of data

AI-generated content may be incorrect.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Sr. No. | Transformation | | 1 | Map | | 2 | filter | | 3 | flatMap | | 4 | mapParitions | | 5 | mapParitionsWithIndex | | 6 | groupBy | | 7 | sortBy | | 8 | Union | | 9 | intersection | | 10 | subtract | | 11 | Distinct | | 12 | cartesian | | 13 | Zip | | 14 | Sample | | 15 | randomSplit | | 16 | keyBy | | 17 | zipWithIndex | | 18 | zipWithUniqueID | | 19 | zipPartitions | | 20 | Coalesce | | 21 | Repartition | | 22 | pipe | | |  |  | | --- | --- | | Sr. No. | Actions | | 1 | Reduce | | 2 | collect | | 3 | Aggregate | | 4 | fold | | 5 | first | | 6 | Take | | 7 | forEach | | 8 | Top | | 9 | treeAggregate | | 10 | treeReduce | | 11 | forEachPartition | | 12 | collectAsMap | | 13 | count | | 14 | takeSample | | 15 | Max | | 16 | Min | | 17 | sum | | 18 | Histogram | | 19 | mean | | 20 | variance | | 21 | stdev | | 22 | sampleVariance | | 23 | countApprox | | 24 | countApproxDistinct | | 25 | Save | |

4 On-Heap vs Off-Heap| Databricks | Spark | Interview Question | Performance Tuning

-> On-Heap Memory:

-

-> Off-Heap Memory:

5 Databricks | Pyspark: Cluster Deployment

6 Databricks | Pyspark| Spark Reader: Read CSV File

7 Databricks | Pyspark: Filter Condition

8 Databricks | Pyspark: Add, Rename and Drop Columns

9 Databricks | PySpark Join Types

10 Databricks | Pyspark: Utility Commands - DBUtils

11 Databricks | Pyspark: Explode Function

12 Databricks | Pyspark: Case Function (When.Otherwise )

13 Databricks | Pyspark: Union & UnionAll

14 Databricks | Pyspark: Pivot & Unpivot

15 Databricks| Spark | Pyspark | Read Json| Flatten Json

16 Databricks | Spark | Pyspark | Bad Records Handling | Permissive;DropMalformed;FailFast

17 Databricks & Pyspark: Azure Data Lake Storage Integration with Databricks

18 Databricks & Pyspark: Ingest Data from Azure SQL Database

19 Databricks & Pyspark: Real Time ETL Pipeline Azure SQL to ADLS

20 Databricks & Pyspark: Azure Key Vault Integration

21 Databricks| Spark Streaming

22 Databricks| Spark | Performance Optimization | Repartition vs Coalesce

23 Databricks | Spark | Cache vs Persist | Interview Question | Performance Tuning

24 Databricks| Spark | Interview Questions| Catalyst Optimizer

25 Databricks | Spark | Broadcast Variable| Interview Question | Performance Tuning

26 Databricks | Spark | Adaptive Query Execution| Interview Question | Performance Tuning

28 Databricks| Pyspark| Handling Null Part 2

29 Databricks | Spark | Pyspark | UDF

30 Databricks - Spark: Data Skew Optimization

31 Databricks & Spark: Interview Question - Shuffle Partition

32 Databricks: Autoscaling | Optimized Autoscaling

33 Databricks | Pyspark: Dataframe Checkpoint

34 Databricks | Pyspark | Interview Question | Compression Methods: Snappy vs Gzip

35 Databricks | Spark | Pyspark Functions| Split

36 Databricks | Spark | Pyspark Functions| Arrays\_zip

37 Databricks | Spark | Pyspark Functions| Part 2 : Array\_Intersect

38 Databricks | Spark | Pyspark Functions| Part 3 : Array\_Except

39 Databricks | Spark | Pyspark Functions| Part 4 : Array\_Sort

40 Databricks | Spark | Python Functions| Join

41 Databricks | Spark | Pyspark | PartitionBy

42 Databricks | Spark | Pyspark | Number of Records per Partition in Dataframe

43 Databricks | Spark | Pyspark | Null Count of Each Column in Dataframe

44 Databricks - Pyspark: Find Top or Bottom N Rows per Group

45 Databricks & Spark: Interview Question(Scenario Based) - How many spark jobs get created?

46 Databricks | Pyspark: Greatest vs Least vs Max vs Min

47 Databricks | Pyspark | Delta Lake: Introduction to Delta Lake

48 Databricks| Pyspark| Delta Lake Architecture: Internal Working Mechanism

49 Databricks| Pyspark| Delta Lake: Solution Architecture

50 Databricks | Delta Lake| Pyspark: Create Delta Table Using Various Methods

51 Databricks| Pyspark| Delta Lake: Delta Table Instance

52 Databricks| Pyspark | Delta Lake: Different Approaches to Insert Data Into Delta Table

53 Databricks| Pyspark| Delta Lake: Different Approaches to Delete Data from Delta Table

54 Databricks | Pyspark | Delta Lake : Update Delta Table

55 Databricks Pyspark:Slowly Changing Dimension|SCD Type1| Merge using Pyspark and Spark SQL

56 Databricks & Pyspark: Delta Lake Audit Log Table with Operation Metrics

57 Databricks | Pyspark | Delta Lake : Slowly Changing Dimension (SCD Type2)

58 Databricks | Pyspark | Delta Lake: Time Travel

59 Databricks | Pyspark| Delta Lake: Restore Command

60 Databricks | Pyspark | Delta Lake: Optimize Command - File Compaction

61 Databricks | Pyspark | Delta Lake: Vacuum Command

62 Databricks | Pyspark | Delta: Z-Order Command

63 Databricks | Pypark | Delta: Schema Evolution - MergeSchema

64 Databricks | Pyspark | Dataframe InsertInto Delta Table

65 Databricks | Spark | Pyspark | Data Skewness| Interview Question: SPARK\_PARTITION\_ID

66 Databricks| Pyspark| Input\_File\_Name: Identify Input File Name of Corrupt Record

67 Databricks | Pyspark | Window Functions: Lead and Lag

68 Databricks | Pyspark | Interview Question: Explain Plan

69 Databricks | Pyspark | UDF to Check if Folder Exists

70 Databricks | Pyspark | Interview Question: Sort-Merge Join (SMJ)

71 Databricks | Pyspark | Performance Optimization - Bucketing

72 Databricks|Pyspark:Interview Question|Scenario Based|Max Over () Get Max value of Duplicate Data

73 Databricks | Pyspark | Create\_map(): Convert Dataframe Columns to Dictionary (Map Type)

74 Databricks | Pyspark | Performance Optimization: Delta Cache

75 Databricks | Pyspark | Split Array Elements into Separate Columns

76 Databricks | Pyspark | Tips: Write Dataframe into Single File with Specific File Name

77 Databricks | Pyspark | Workspace Object Access Control

78 Databricks | Pyspark | Databricks Secret Scopes: Azure Key Vault Backed Secrets

79 Databricks | Pyspark | Databricks Workflows: Job Scheduling

80 Databricks | Pyspark | Azure Data Factory + Azure Databricks: Execute Notebook Via ADF

81 Databricks | Pyspark | Notebook Activity in Azure Data Factory with Input Parameter

82 Databricks | Pyspark | Notebook Activity in Azure Data Factory with Output Parameter

83 Databricks | Pyspark | Real Time Project: ETL Pipeline Integrating ADF, ASQL, ADLS, Key Vault

84 Databricks |Pyspark |Notebook Scheduling through Schedule Based Trigger using Azure Data Factory

85 Databricks | Pyspark | Notebook Scheduling through Event Based Trigger using Azure Data Factory

86 Databricks | Pyspark | Interview Question: Read Excel File with Multiple Sheets

87 Databricks | Pyspark | Interview Question |Handlining Duplicate Data: DropDuplicates vs Distinct

88 Databricks | Pyspark | Interview Question | Performance Optimization: Select vs WithColumn

89 Databricks | Pyspark | Interview Question | Schema Definition: Struct Type vs Struct Field

90 Databricks | Pyspark | Interview Question | Schema Definition: Struct Type vs Map Type

91 Databricks | Pyspark | Schema | Different Methods of Schema Definition

92 Databricks | Pyspark | Real Time Scenario | Schema Comparison

93 Databricks | Pyspark | Data Security: Enforcing Column Level Encryption

94 Databricks | Pyspark | Interview Question: Pyspark VS Pandas

95 Databricks | Pyspark | Real Time Use Case: Generate Test Data - Array\_Repeat()

96 Databricks | Pyspark | Spark Architecture: Internals of Partition Creation Demystified

97 Databricks|Pyspark|Automation|Real Time Project:DataType Issue When Writing to Azure Synapse/SQL

98 Databricks | Pyspark| Transformation: Subtract vs ExceptAll

99 Databricks | Pyspark| Window Function: First and Last

100 Databricks | Pyspark| Spark Reader: Skip First N Records While Reading CSV File

101 Databricks | PySpark| Spark Reader: Skip Specific Range of Records While Reading CSV File

102 Databricks | Pyspark| Query Dataframe Using Spark SQL

103 Databricks | Pyspark| Spark SQL: Except Columns in Select Clause

104 Databricks | Pyspark| AutoLoader: Incremental Data Load

105 Databricks | Pyspark| Delta Live Table: Introduction

106 Databricks | Pyspark| Delta Live Table: Declarative VS Procedural

107 Databricks | Pyspark| Delta Live Table: Datasets - Tables and Views

108 Databricks | Pyspark| Delta Live Table: Data Quality Check - Expect

109 Databricks | Pyspark | Downloading Files from Databricks DBFS Location

110 Databricks | Pyspark| Built-In Function: TRANSFORM

111 Databricks | Pyspark| Delta Lake: Deletion Vectors

112 Databricks | Pyspark| Delta Lake: Change Data Feed

113 Databricks | Pyspark| Built-in Function: ZIP\_WITH

114 DataBricks Learning: System Variable \_SQLDF