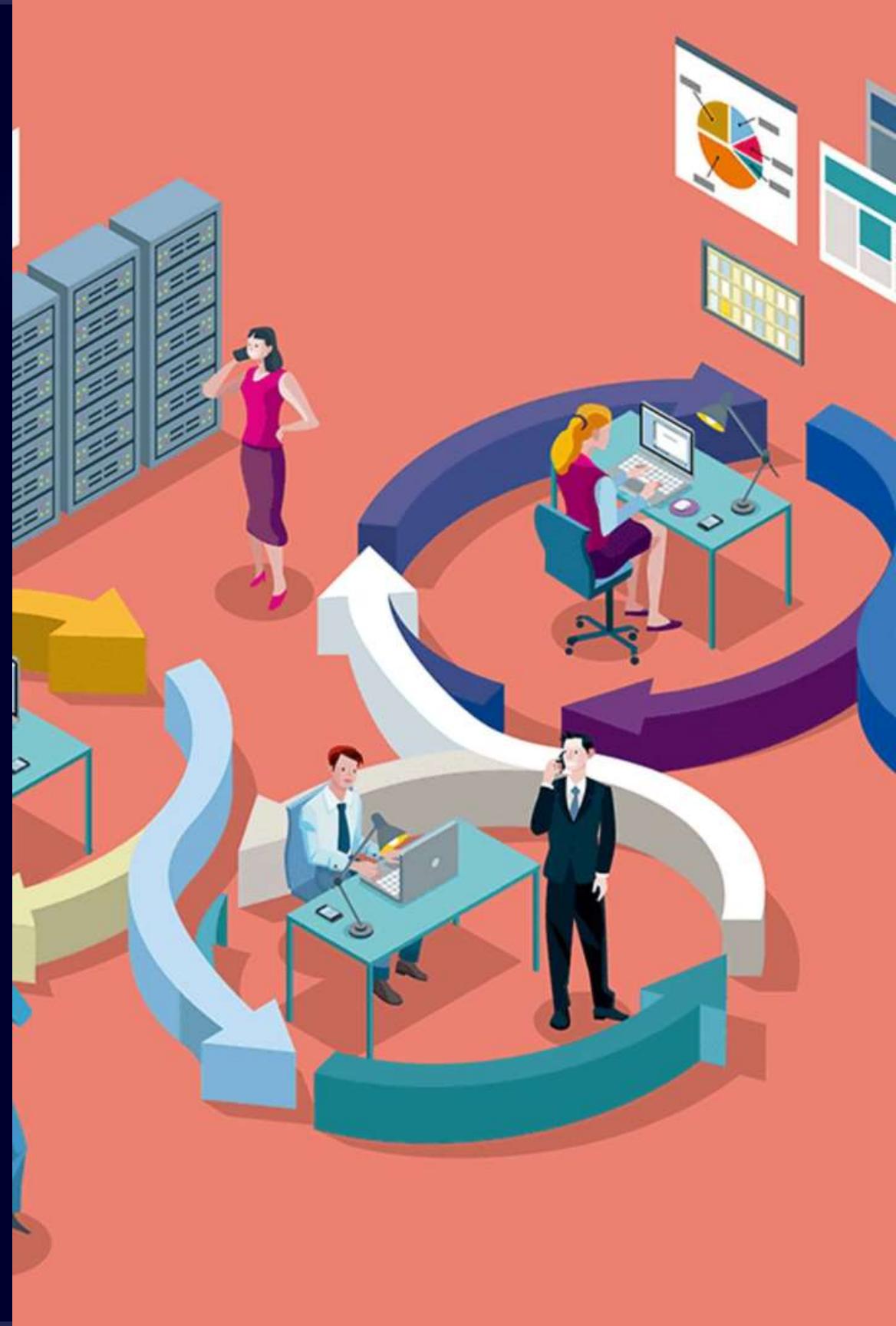


Foundation of Software Development

By-
Tanya Yadav



SDLC

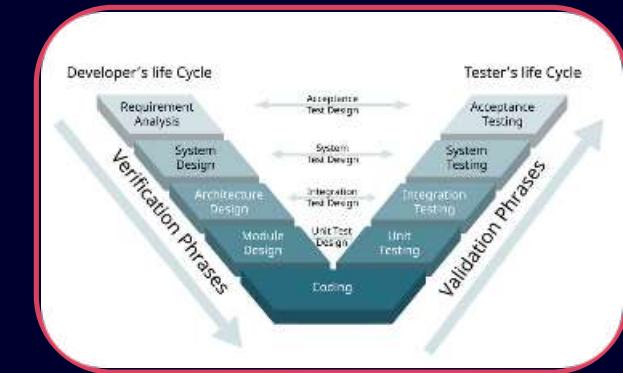


Software Models



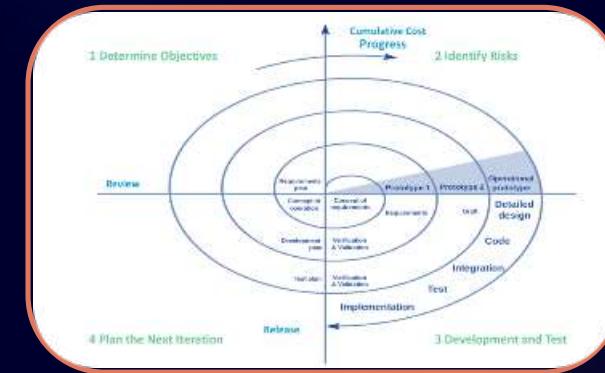
Waterfall Model

This model is a linear sequential approach to software development. It divides the software development process into phases where each phase is completed before moving onto the next one.



V Shape Model

It is based on the association of a testing phase for each corresponding development stage. Development of each step directly associated with the testing phase. The next phase starts only after completion of the previous phase i.e. for each development activity, there is a testing activity corresponding to it.



Spiral Model

This model is based on a Risk-Driven approach where the software development process is divided into smaller cycles. Each cycle includes risk assessment, building, and testing, followed by risk reassessment in the next cycle.

Agile methodology



The Agile methodology is a project management approach that involves breaking the project into phases and emphasizes continuous collaboration and improvement. Teams follow a cycle of planning, executing, and evaluating.

- Faster implementation of solutions
- Waste reduction thanks to the minimization of the resources
- Greater flexibility and adaptability to change
- More success thanks to more focused efforts
- Faster delivery times



Gantt Chart

Gantt charts are an essential tool in project management. They help in planning, coordinating, and tracking specific tasks. The chart consists of a bar graph that displays the project's schedule. Each bar represents a specific task, and the length of the bar represents the time needed to complete that task.

Scrum Overview

Product Owner

A person who defines the product vision and is responsible for the product's development and its delivery.

Scrum Master

A person responsible for ensuring that the Scrum team adheres to the Scrum framework and helps them to remove any impediments.

Dev Team (Dev + Tester)

A cross-functional team responsible for delivering potentially shippable products by the end of each sprint.

Events / Stages

- Product Backlogs
- Sprint Backlogs
- Release Planning
- Sprint Planning
- Daily Scrum
- Sprint Review
- Sprint Retrospective

Roles

- Product Owner
- Scrum Master
- Dev Team (Dev + Tester)

Artifacts

- Epic
- User Story
- Bug
- Task

Tools

- Product Backlog
- Sprint Backlog
- Burn Up Chart

Issues

- Backlogs

InPrivate

Home - Microsoft Az... | azure devops - Search | TF400813: The user... | My Information | Streamline Solutions Inc.

https://dev.azure.com/yadavtanya9/Streamline%20Solutions%20Inc.%20-%20Optimizing%20Software%20Deployment/_backlogs/

Azure DevOps yadavtanya9 / Streamline Solutions Inc. Boards Backlogs

Streamline Solutions Inc. +

Overview Boards Work items Boards Backlogs Sprints Queries Delivery Plans Analytics views Repos Pipelines Test Plans Project settings

Backlog Analytics + New Work Item View as Board Backlog Items

Order Work Item Type Title

Epic Develop and deploy spring boot application

Feature Feature1

+ Product Backlog User story1

Task Fork the main GitHub repository

Task configure a virtual machine and define a Maven-based build process

Task craft a Dockerfile tailored

Task Perform Pull Request Collaboration

Task outline a GitHub Actions workflow triggered by PRs

Task test the deployed Azure Web App's webpage

Task manage tasks on the Azure Boards Tasks Board

Task Project Closure and Reflection

Feature Feature 2

Planning

Drag and drop work items to include them in a sprint.

Streamline Solutions Inc. - Optimizing Software Deployment Team

Sprint 1

No work scheduled yet

InPrivate Home - Microsoft Az... | azure devops - Search | TF400813: The user... | My Information | Streamline Solutions Inc.

https://dev.azure.com/yadavtanya9/Streamline%20Solutions%20Inc.%20-%20Optimizing%20Software%20Deployment/_sprints/ta...

Azure DevOps yadavtanya9 / Streamline Solutions Inc. Boards Sprints

Streamline Solutions Inc. +

Overview Boards Work items Boards Backlogs Sprints Queries Delivery Plans Analytics views

Taskboard Backlog Capacity Analytics + New Work Item

Sprint 1 Person: All

Collapse all To Do In Progress Done

98: User story1 New Unassigned

103: test the deployed Azure Web App's webpage To Do Unassigned

102: outline a GitHub Actions workflow triggered by PRs In Progress Unassigned

94: Fork the main GitHub repository Done Unassigned

104: manage tasks on the Azure Boards Tasks Board To Do Unassigned

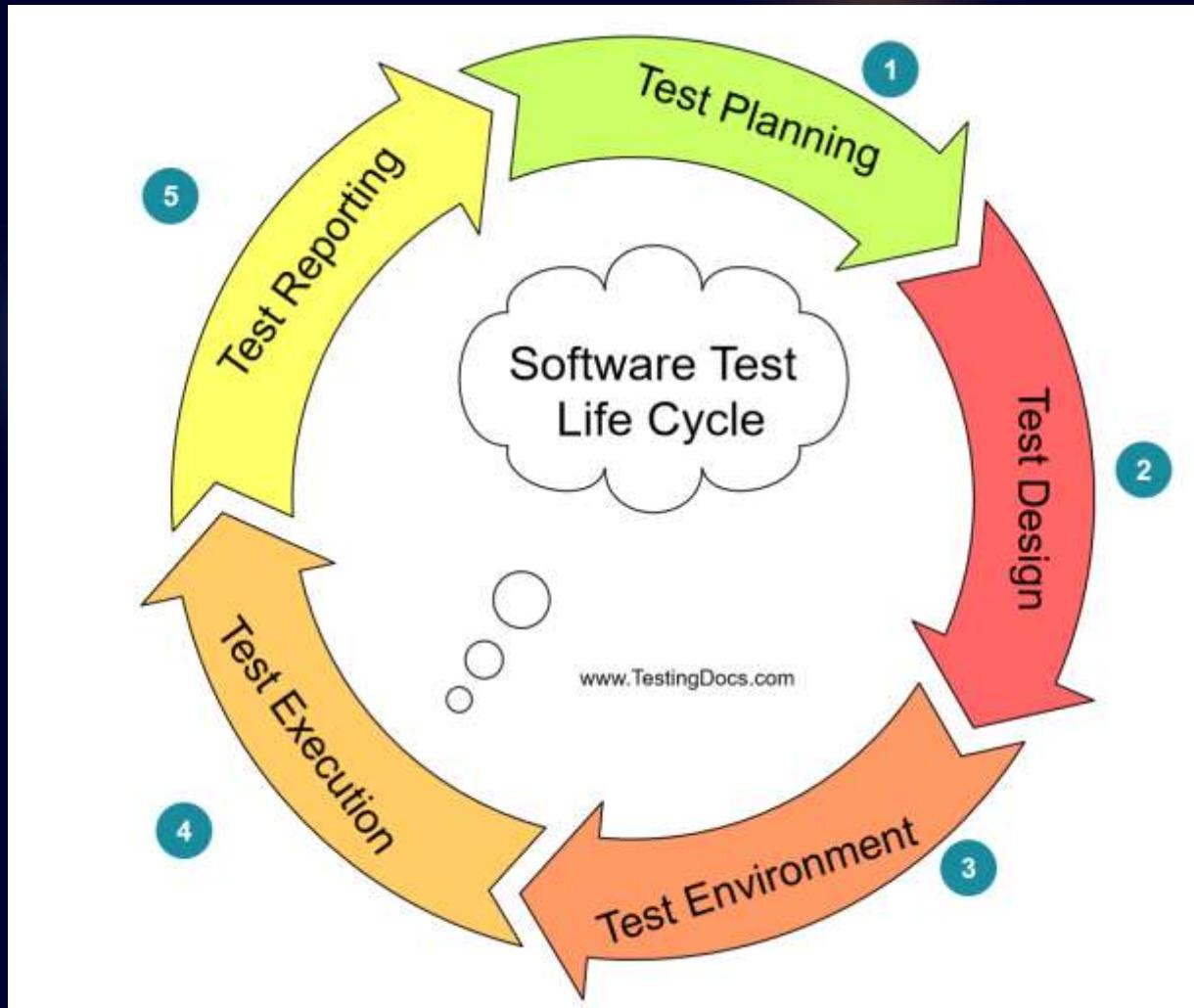
95: configure a virtual machine and define a Maven-based build process Done Unassigned

100: craft a Dockerfile tailored Done Unassigned

101: Perform Pull Request Collaboration To Do Unassigned

105: Project Closure and Reflection To Do Unassigned

Software Testing



Software Testing Life Cycle (STLC) is a sequence of specific activities conducted during the testing process to ensure software quality goals are met. STLC involves both verification and validation activities. Contrary to popular belief, Software Testing is not just a single/isolate activity, i.e. testing. It consists of a series of activities carried out methodologically to help certify your software product.

Software Testing Life Cycle

- 1 Requirement Analysis**

Understand the client's requirements and create a detailed plan for software creation and testing to ensure your software will meet and exceed their expectations.
- 2 Test Planning & Test Case Development**

Develop comprehensive test cases and create test plans accordingly. Testing planning is about attention to detail.
- 3 Environment Setup**

Ensure you have a set-up that makes testing efficient and effective.
- 4 Test Execution & Test Case Closure**

Thoroughly test the software in the designed environment to ensure it passes all test cases and meets high-quality standards. Close test cases when the listed requirements are satisfied.
- 5 Acceptance Testing**

Commit to ensuring the software is fully reliable and functional for the client in a live production environment before the final launch.

Software Testing Techniques

White Box Testing

Developers use this technique to ensure the maximum quality of individual application components.

Black Box Testing

Quality assurance engineers use this technique for integration testing to ensure the whole application meets high-quality standards.

Grey Box Testing

Automation engineers use this technique, relying on QA engineers to provide testing scenarios. Make sure to maintain consistent communication to ensure testing is efficient and effective.

Boundary Value Analysis

Check for edge cases to ensure the software works under less than ideal conditions and maintain high-quality standards.

C33

A	B	C	D	E	F	G	H	I	
Test Case ID	Test Case Description	Test Case Priority	Pre-Condition	Test Step	Test Data	Expected Result	Actual Result	Status	
TC_001	To validate facebook account creation functionality with Valid Input	1	Launch browser with the url "https://www.facebook.com/" and navigate to Account Creation page	1. Enter username in the "First Name" field	senthil	The entered data should be displayed in the "First name" field	The entered data is displayed in the "First name" field	Pass	
				2. Enter surname in the "Surname" field	nata	The entered data should be displayed in the "Surname" field	The entered data is displayed in the "Surname" field	Pass	
				3. Enter email in the "Mobile number or email address" field	sen@gmail.com	The entered data should be displayed in the "Mobile number or email address" field		Pass	
				4. Enter password in the "New Password" field	Qwertyu#21	The entered data should be displayed in the "Password" field		Pass	
				5. Select Date, Month and Year from "Date of birth" drop down filed	12, Feb, 2022	The selected date, month and year should be displayed in the "Date of birth" field		Pass	
				6. Select gender from "Gender" field	Male	The selected gender should be selected		Pass	
				7. Click on "Sign Up" button	NA	The message "Account Created Successfully" should be displayed		Pass	
TC_002	To validate facebook account creation functionality with invalid input	2	Launch browser with the url "https://www.facebook.com/" and navigate to Account Creation page	1. Enter username in the "First Name" field	senthil	The entered data should be displayed in the "First name" field		Pass	
				2. Enter surname in the "Surname" field	nata	The entered data should be displayed in the "Surname" field			
				3. Enter email in the "Mobile number or email address" field	sen@gmail.com	The entered data should be displayed in the "Mobile number or email address" field			
				4. Enter password in the "New Password" field	Qwertyu#21	The entered data should be displayed in the "Password" field			
				5. Select Date, Month and Year from "Date of birth" drop down filed	12, Feb, 2022	The selected date, month and year should be displayed in the "Date of birth" field			
				6. Select gender from "Gender" field	Male	The selected gender should be selected			
				7. Click on "Sign Up" button	NA	The error message "The email id is not valid" should be displayed			
TC_003	To validate facebook account creation functionality with field level validation of First Name field	3	Launch browser with the url "https://www.facebook.com/" and navigate to Account Creation page	1. Leave First name field bland and move to next field	NA	The error message "First Name" is mandatory	Project Name:	Project Start Date:	
				1. Enter username in the "First Name" field	senthil\$%^	The error message "First Name cannot accept special character" is displayed			
				1. Enter username in the "First Name" field	Senthil123	The entered data should display in the "Firstname" field			
TC_004	To validate facebook account creation functionality with field	4					Version Number:	Project End Date:	
Test Scenario Test Case Defect Log RTM +				Requirement ID	Test Scenario ID	Test Case ID	Status	Defect ID	Remarks
				REQ_001	TS_001	TC_001	Covered		
				REQ_002	TS_001	TC_002	Covered		
				REQ_003	TS_001	TC_003	Covered		
				REQ_004	TS_001	TC_004	Covered		
				REQ_005	TS_001	TC_005	Covered		
				REQ_006	TS_001	TC_006	Covered	DEF_001	

Test Case Design for Facebook Creation

Requirement Traceability Matrix

Project Name:

Project Start Date:

Version Number:

Project End Date:

Azure DevOps for Software Creation

Azure DevOps is a development platform that facilitates the entire process of software creation, from managing code to testing and deploying it on multiple platforms. It is essential for software development, allowing for the fast and efficient creation of high-quality software.

The screenshot shows the Azure DevOps Pipelines interface for the organization 'FabrikamFiber'. The top navigation bar includes 'FabrikamFiber / Pipelines' and a search bar. Below the header, there are tabs for 'Recent', 'All', and 'Runs', with 'Recent' currently selected. A section titled 'Recently run pipelines' lists two pipelines:

Pipeline	Last run
pipelines-dotnet-core	#20191209.2 • Set up environment
FabrikamFiber	#20191209.3 • Set up environment

Each pipeline entry includes a green checkmark icon, the pipeline name, the run number, the status message, and a note indicating it was manually triggered.

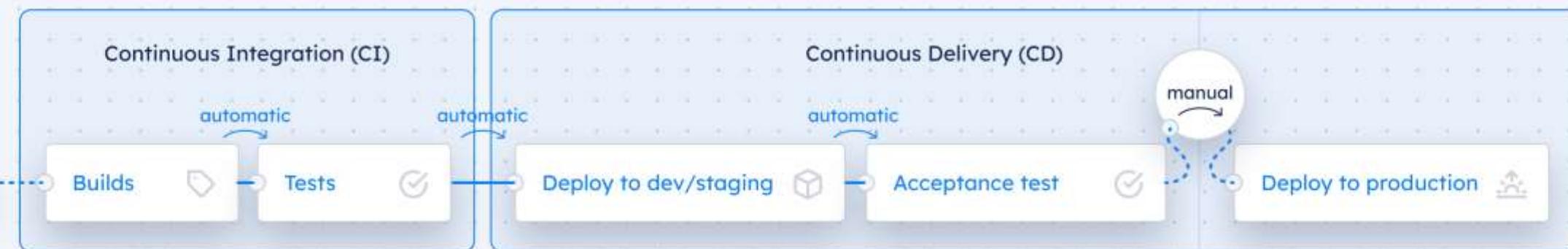
Working on Azure Devops

DevOps automates *software development, testing and IT operations* to shorten the development life cycle while delivering features, fixes, and updates in alignment with the business' objectives.

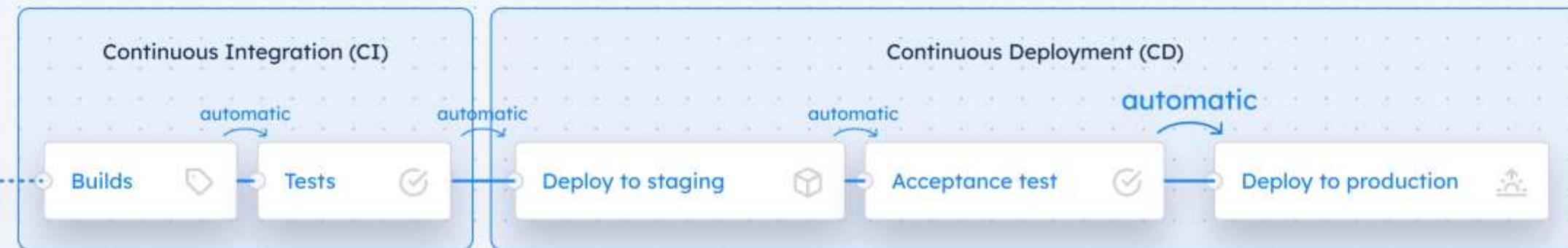
The screenshot shows the Azure DevOps interface for the 'Streamline Solutions Inc.' project. The left sidebar lists project services: Streamline Solutions Inc., Overview, Summary, Dashboards, Wiki, Boards, Repos, Pipelines, Test Plans, and Artifacts. The 'Summary' service is currently selected. The main content area displays a welcome message: 'Welcome to the project! What service would you like to start with?'. Below this, there are four buttons: Boards, Repos, Pipelines, and Test Plans. The background features a cartoon illustration of a person working at a desk with a dog nearby.

CI/CD

CONTINUOUS DELIVERY ...



...VS CONTINUOUS DEPLOYMENT



GitHub Actions

Automate, customize, and execute your software development workflows right in your repository with GitHub Actions. You can discover, create, and share actions to perform any job you'd like, including CI/CD, and combine actions in a completely customized workflow.

DOCKER IMAGE AND DOCKER HUB

Docker Hub is a collaboration tool and a marketplace for community developers, open source contributors, and independent software vendors (ISVs) to distribute their code publicly. Docker Hub provides a consistent, secure, and trusted experience, making it easy for developers to access software they need.

Docker Image is an executable package of software that includes everything needed to run an application. This image informs how a container should instantiate, determining which software components will run and how. **Docker Container** is a virtual environment that bundles application code with all the dependencies required to run the application. The application runs quickly and reliably from one computing environment to another.

https://hub.docker.com

DockerCon 2023: Our annual developer event is back – online & in person. [Learn more.](#)

Search Docker Hub

Explore Repositories Organizations Help

Upgrade tanya9941

tanya9941

Search by repository name

All Content

Create repository

tanya9941 / hello-world-java-demo

Contains: Image | Last pushed: 7 days ago

Inactive 0

1

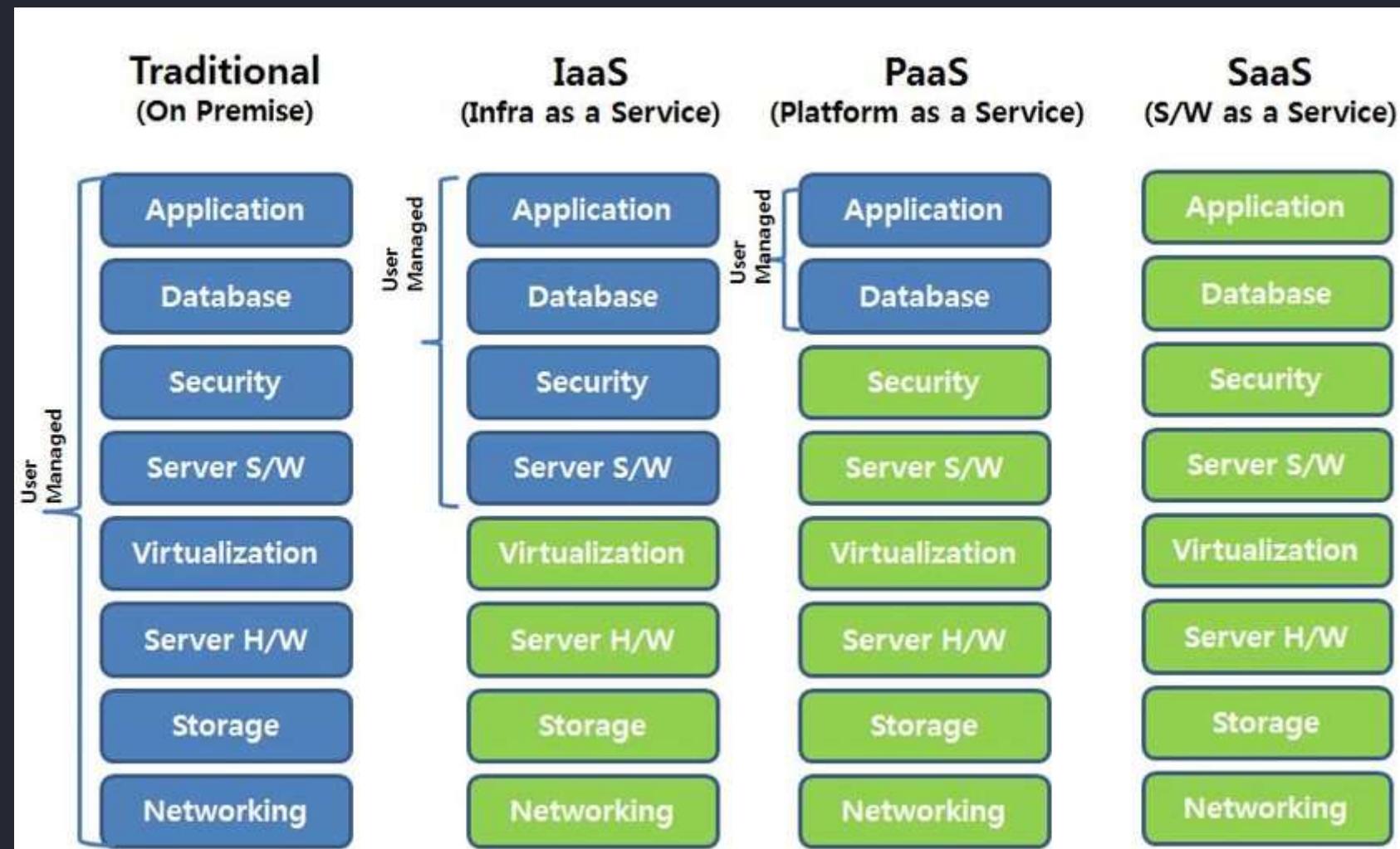
Public

Create an Organization
Manage Docker Hub repositories
with your team

The screenshot shows the Docker Hub homepage. At the top, there's a purple banner for DockerCon 2023. Below it is a blue header bar with navigation links like 'Explore', 'Repositories', 'Organizations', and 'Help'. A user profile for 'tanya9941' is visible on the right. The main content area shows a repository card for 'tanya9941 / hello-world-java-demo', which is inactive, has 0 stars, 1 download, and is public. To the right of the repository card is a sidebar with icons for creating an organization and managing repositories. The URL in the browser bar is https://hub.docker.com.

CLOUD

Cloud computing means storing and accessing the data and programs on remote servers that are hosted on the internet instead of the computer's hard drive or local server. Cloud computing is also referred to as Internet-based computing, it is a technology where the resource is provided as a service through the Internet to the user. The data which is stored can be files, images, documents, or any other storable document.

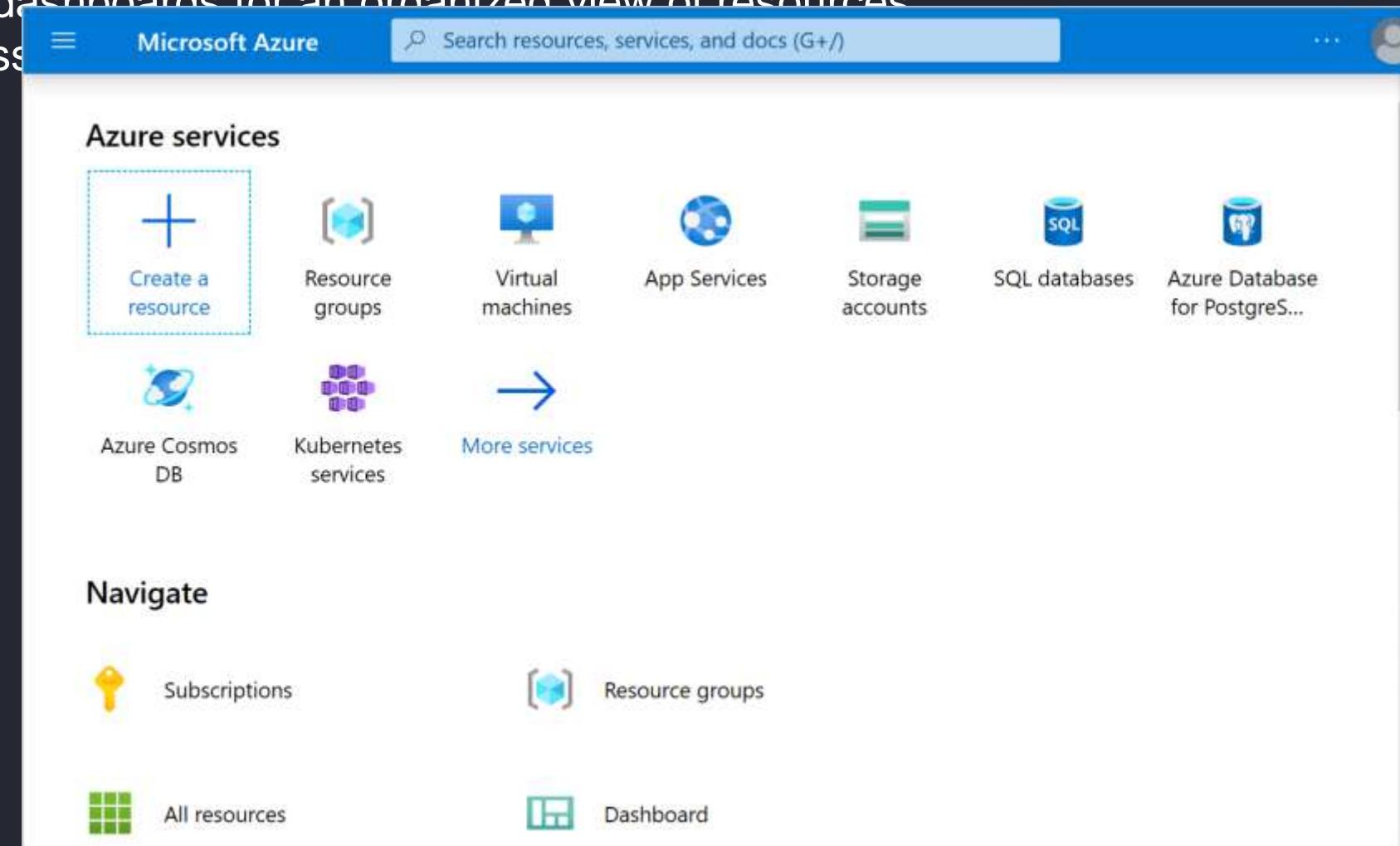


Azure Portal

Azure Portal is a web-based, integrated console that provides an alternative to command-line tools. With the Azure Portal, you can manage your Azure subscription using the graphical user interface.

Build, manage and monitor everything from simple web apps to complex cloud deployments.

- Create custom dashboards for an organized view of resources
- Configure access



Got the
chance to
visit Xenon
office



DBMS

Database management systems were developed to handle the following difficulties of typical File-processing systems supported by conventional operating systems.

1. Data redundancy and inconsistency
2. Difficulty in accessing data
3. Data isolation – multiple files and formats
4. Integrity problems
5. Atomicity of updates
6. Concurrent access by multiple users
7. Security problems

DDL stands for Data Definition Language. SQL queries like CREATE, ALTER, DROP, TRUNCATE and RENAME come under this.

DML stands for Data Manipulation Language. SQL queries like SELECT, INSERT, DELETE and UPDATE come under this.

DCL stands for Data Control Language. SQL queries like GRANT and REVOKE come under this.

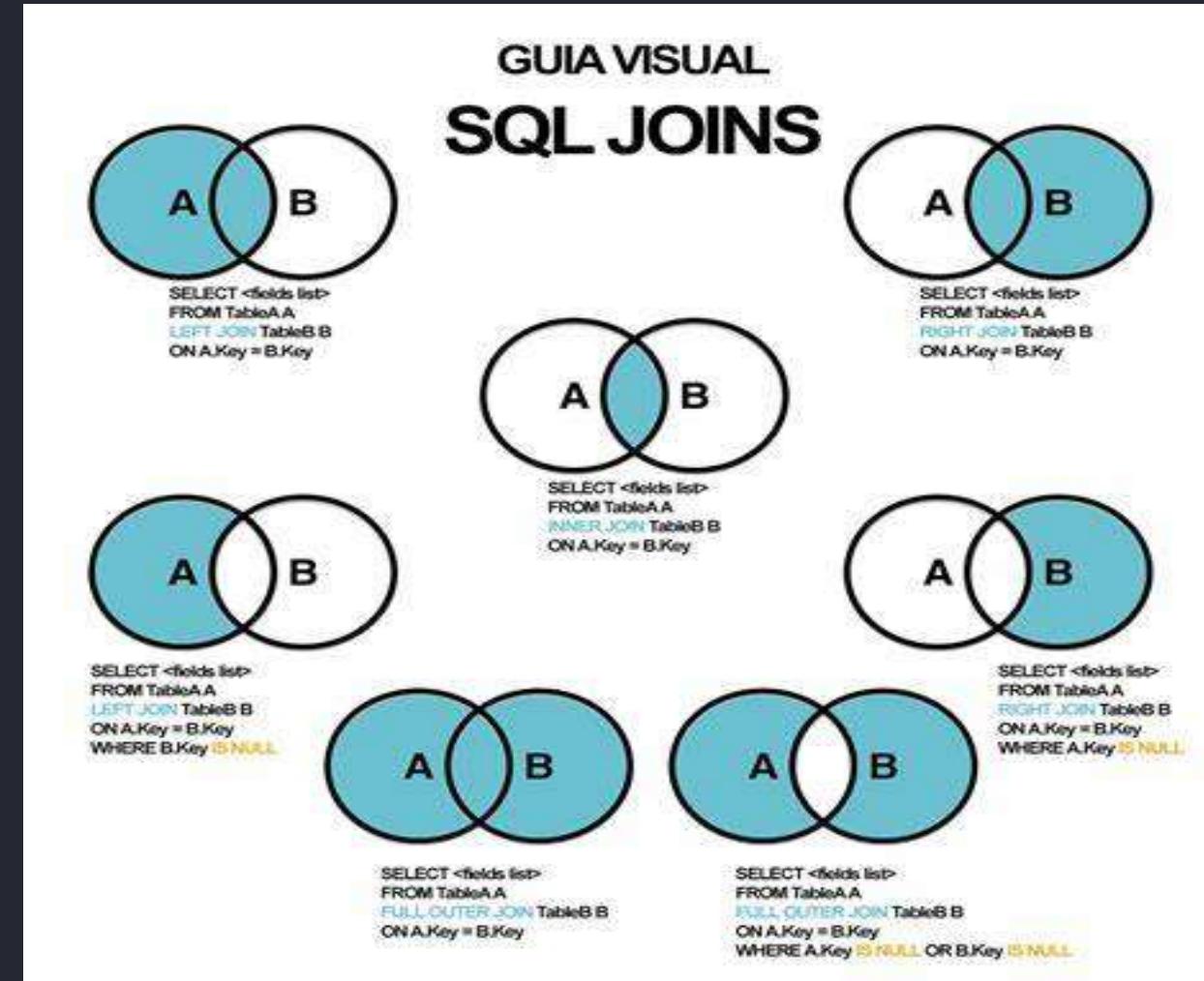
SQL JOINS

Inner Join: This type of join is used to fetch the data among the tables which are common in both tables.

Left Join: This returns all the rows from the table which is on the left side of the join but only the matching rows from the table which is on the right side of the join.

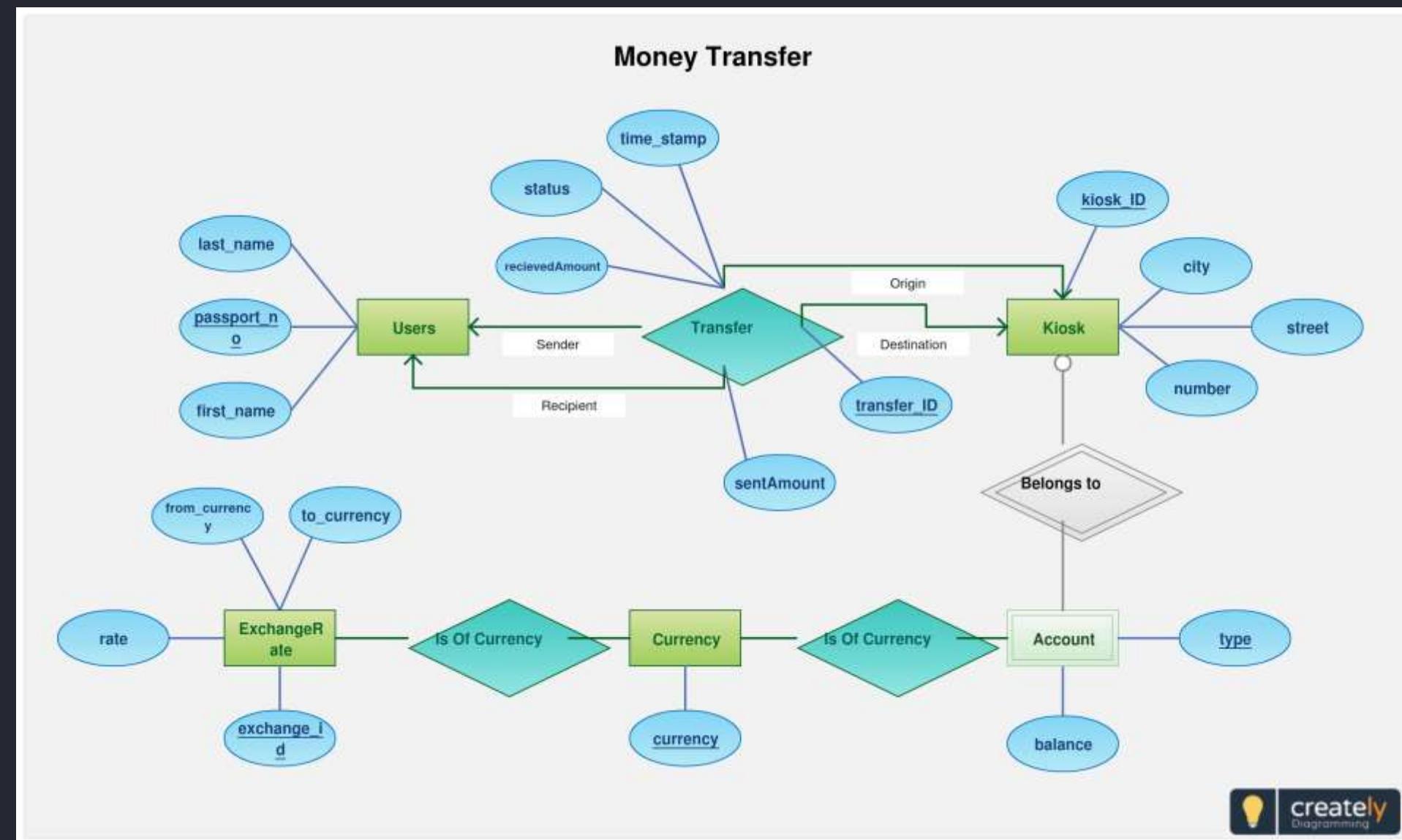
Right Join: This returns all the rows from the table which is on the right side of the join but only the matching rows from the table which is on the left side of the join.

Full Join: This returns the rows from all the tables on which the join condition has been put and the rows which do not match hold null values.



ERM MODEL

The Entity Relational Model is a model for identifying entities to be represented in the [database](#) and representation of how those entities are related. The ER data model specifies enterprise schema that represents the overall logical structure of a database graphically.



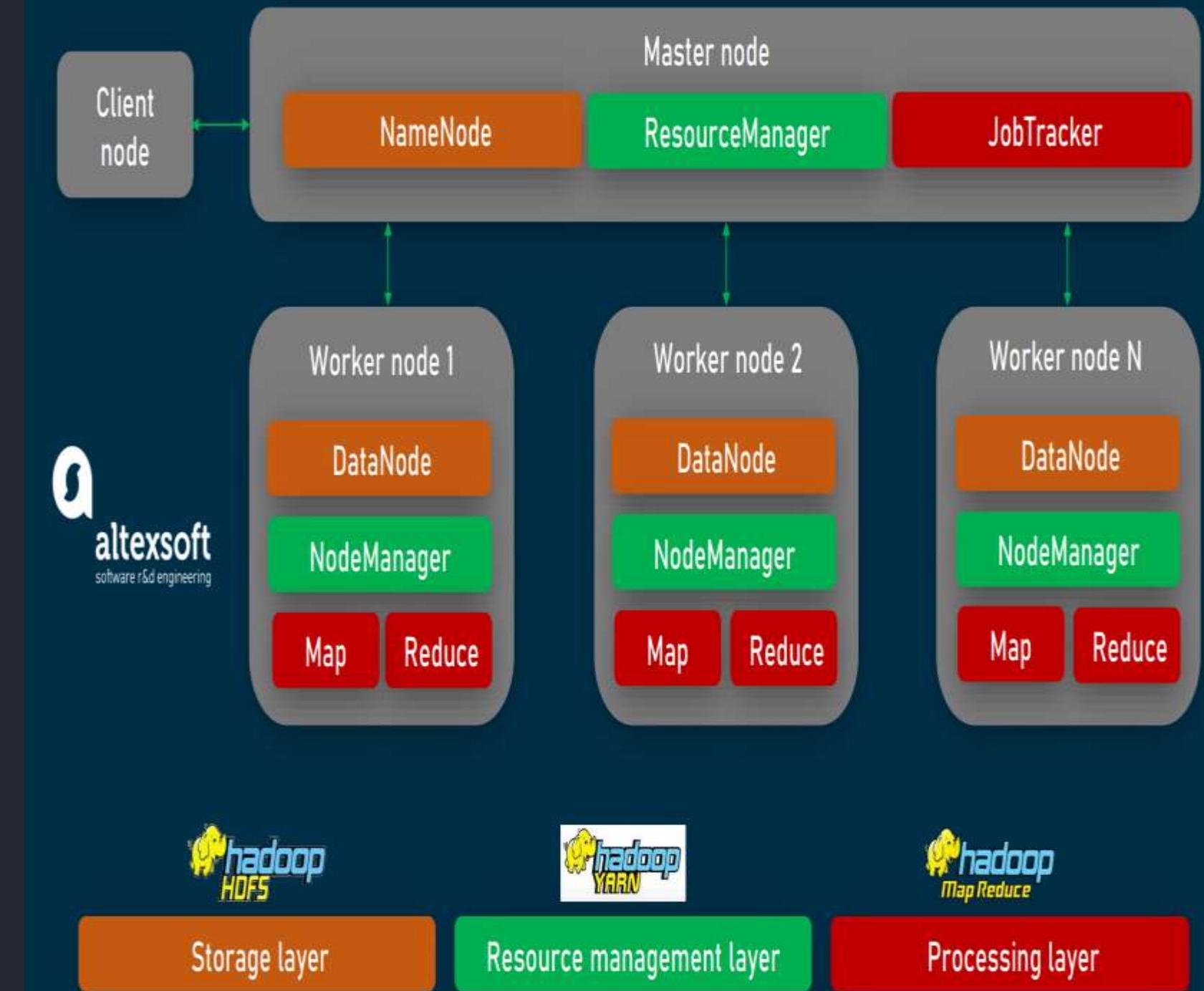
CLOUD DATA STORAGE PATTERNS

	Data Warehouse	Data Lake	Data Lakehouse
Storage Data Type	Works well with structured data	Works well with semi-structured and unstructured data	Can handle structured, semi-structured, and unstructured data
Purpose	Optimal for data analytics and business intelligence (BI) use-cases	Suitable for machine learning (ML) and artificial intelligence (AI) workloads	Suitable for both data analytics and machine learning workloads
Cost	Storage is costly and time-consuming	Storage is cost-effective, fast, and flexible	Storage is cost-effective, fast, and flexible
ACID Compliance	Records data in an ACID-compliant manner to ensure the highest levels of integrity	Non-ACID compliance: updates and deletes are complex operations	ACID-compliant to ensure consistency as multiple parties concurrently read or write data

Hadoop

The Apache Hadoop software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Rather than rely on hardware to deliver high-availability, the library itself is designed to detect and handle failures at the application layer, so delivering a highly-available service on top of a cluster of computers, each of which may be prone to failures.

HADOOP CLUSTER ARCHITECTURE



Creating SQL Database in Azure

The screenshot shows the 'Create SQL Database' page in the Microsoft Azure portal. In the 'Project details' section, the subscription is set to 'npunext-1673505441526' and the resource group is '(New) Tanya_training'. A cost summary table is visible on the right, showing 'General Purpose (GP_S_Gen5_1)' with a cost per GB of 41.6 and estimated storage and compute costs. In the 'Database details' section, the database name is 'tanya_database' and it is associated with the server '(new) tanyaserver1 (East US)'. The 'Workload environment' is set to 'Development'. Below the form are 'Review + create' and 'Next : Networking >' buttons.

The screenshot shows the deployment details for 'Microsoft.SQLDatabase.newDatabaseNewServer_ad4b7b6bca2d4c8ca0701'. The status bar indicates 'Deployment succeeded'. The deployment summary shows the deployment name, start time (8/31/2023, 10:16:43 AM), subscription, correlation ID, and resource group. The 'Overview' tab is selected, showing deployment details and next steps. The 'Deployment details' section includes a 'Go to resource' button. The right sidebar features links for 'Cost management', 'Microsoft Defender for Cloud', 'Free Microsoft tutorials', 'Work with an expert', and 'Find an Azure expert >'. A feedback section at the bottom encourages users to tell their experience with deployment.

The screenshot shows a Microsoft Edge browser window with several tabs open at the top, including Home, Lumen, Subscription, Tanya, portal.azure.com, and tanya_database. The main content area is the Microsoft Azure portal, specifically the Query editor (preview) for the 'tanya_database' on 'tanyaserver1'. The URL in the address bar is https://portal.azure.com/#@npunext.onmicrosoft.com/resource/subscriptions/324a722... . The search bar at the top says 'Search resources, services, and docs (G+ /)'. The user is signed in as 'Shellunext_1693422092... UNEXT'.

Home > tanya_database (tanyaserver1/tanya_database)

tanya_database (tanyaserver1/tanya_database) | Query editor (preview) ...



SQL database

Search

>Login

+ New Query

↑ Open query

Feedback

Getting started

Overview

Activity log

Tags

Diagnose and solve problems

Query editor (preview)

Settings

Compute + storage

Connection strings

Properties

Locks

Data management

Replicas

Sync to other databases

Integrations

Azure Synapse Link

tanya_database (tanya)

Query 1 ↻

1
2 sp_tables

Results Messages

TABLE_QUALIFIER	TABLE_OWNER	TABLE_NAME	TABLE_TYPE
tanya_database	dbo	students	TABLE
tanya_database	sys	trace_xe_action_map	TABLE
tanya_database	sys	trace_xe_event_map	TABLE
tanya_database	INFORMATION_SCHEMA	CHECK_CONSTRAINTS	VIEW
tanya_database	INFORMATION_SCHEMA	COLUMN_DOMAIN	VIEW

✓ Query succeeded | 2s

Azure Platform

account

An account that's used to access and manage an Azure subscription. It's often referred to as an Azure account although an account can be any of these: an existing work, school, or personal Microsoft account.

region

An area within a geo that does not cross national/regional borders and contains one or more datacenters. Pricing, regional services, and offer types are exposed at the region level.

resource group

A container in Resource Manager that holds related resources for an application. The resource group can include all of the resources for an application, or only those resources that are logically grouped together. You can decide how you want to allocate resources to resource groups based on what makes the most sense for your organization.

storage account

An account that gives you access to the Azure Blob, Queue, Table, and File services in Azure Storage. The storage account name defines the unique namespace for Azure Storage data objects.

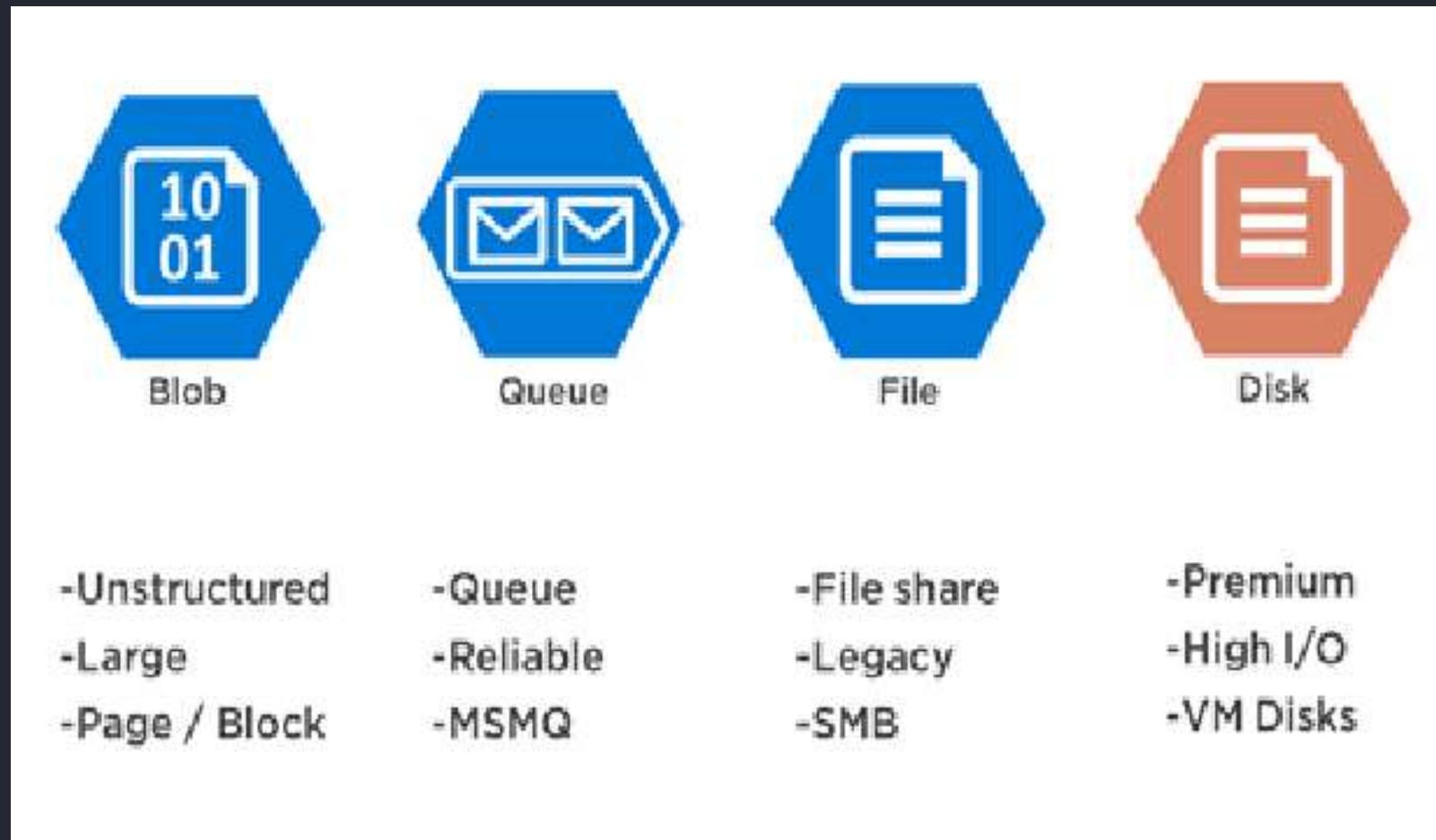
virtual machine

The software implementation of a physical computer that runs an operating system. Multiple virtual machines can run simultaneously on the same hardware. In Azure, virtual machines are available in a variety of sizes.

Azure Data Lake

Analytics is an on-demand analytics job service that simplifies big data. Instead of deploying, configuring, and tuning hardware, you write queries to transform your data and extract valuable insights. The analytics service can handle jobs of any scale instantly by setting the dial for how much power you need. You only pay for your job when it's running, making it cost-effective.

Types of Azure Storages



Azure Storage comparison matrix

	BLOCK Storage (Azure Managed disks)	BLOB/Object Storage (later Azure Data Lake Storage)	File Storage
Use Cases & Workload	<ul style="list-style-type: none"> OS applications that need service-side processing Structured Databases that need good disk read/write performance virtual machine file system (VMFS) volumes Transactional Data for Huge databases Container Storage Typical for SAN 	<ul style="list-style-type: none"> Easy and fast provisioning of storages in scale w/o locking Great for running Big Data workloads, decouple data storage architecture modern application architectures where stored data can be directly accessed by applications through REST API calls , SDK, URI's requests Static web content , images / Video / audio files, large unstructured data, logs , ingestion file storage , Archival , backup 	<ul style="list-style-type: none"> Replace / Lift and Shift On premise file server Better written as “traditional” on-premise applications. well-suited to working with applications that need file locking File storage has broad capabilities and can store just about anything. It’s great for storing an array of complex files and is fast for users to navigate Typical for NAS Storage
Data Organization	Keeps data in large groups, called blocks	<ul style="list-style-type: none"> considers all data equally each object independent of the others metadata to organize the information 	Data managed and organized on File level
Data Access	Data stored in block storage can only be accessed through the operating system to which it is attached to	Data in object storage can be accessed in a more flexible and secure manner	Data stored in files is organized and retrieved using a limited number of metadata as catalog that tells the system exactly where the file itself is kept.
Architecture	Tightly coupled with OS and accessed through the operating system to which it is attached	<ul style="list-style-type: none"> Decoupled the storage from compute or OS Data accessed in a more flexible, secure manner without depending on OS 	Data is organized in hierarchical storage by folders. Can be mounted across different deployments on different OS and cached on Windows server with Azure File sync
K8s Friendly Tiers	Yes , any type of Managed disk can be provisioned	Can be assigned to Azure Hot , Cool , Archive access tiers	Azure file storage can be mounted by AKS
limitation	Metadata handling capabilities including the security and management	There is literally no limit on object storage it doesn't provide the ability to incrementally edit one part of a file (as block storage does) , Object must be manipulated as a whole unit, requiring the entire object to be accessed, updated, then re-written in their entirety. Poor performance for Transactional DB	File-based storage systems must scale out by adding more systems, rather than scale up by adding more capacity. So not very scalable
Costs	block storage device data is always costly	Flexible Storage tiers can be cost effective for massive amount of data	Cheaper than premium and Hot tiers of Blob
Metadata Handling	It has limited capability to handle metadata which is layers of overhead for application developers	Object storage metadata can be extremely detailed and can store lot more information. To retrieve the data, the storage operating system uses the metadata and identifiers, which distributes the load better and lets administrators apply policies that perform more robust searches	Data stored in files is organized and retrieved using a limited number of metadata by using reference path where the files stored

Azure Data Factory



Pipeline

A data factory might have one or more pipelines. A pipeline is a logical grouping of activities that performs a unit of work. Together, the activities in a pipeline perform a task. For example, a pipeline can contain a group of activities that ingests data from an Azure blob, and then runs a Hive query on an HDInsight cluster to partition the data.

Mapping data flows

Create and manage graphs of data transformation logic that you can use to transform any-sized data. You can build-up a reusable library of data transformation routines and execute those processes in a scaled-out manner from your ADF pipelines. Data Factory will execute your logic on a Spark cluster that spins-up and spins-down when you need it. You won't ever have to manage or maintain clusters.

Activity

Activities represent a processing step in a pipeline. For example, you might use a copy activity to copy data from one data store to another data store. Similarly, you might use a Hive activity, which runs a Hive query on an Azure HDInsight cluster, to transform or analyze your data. Data Factory supports three types of activities: data movement activities, data transformation activities, and control activities.

Datasets

Datasets represent data structures within the data stores, which simply point to or reference the data you want to use in your activities as inputs or outputs.

Linked services

Linked services are much like connection strings, which define the connection information that's needed for Data Factory to connect to external resources. Think of it this way: a linked service defines the connection to the data source, and a dataset represents the structure of the data.

Azure Data Factory

Triggers

Triggers represent the unit of processing that determines when a pipeline execution needs to be kicked off. There are different types of triggers for different types of events.

Pipeline runs

A pipeline run is an instance of the pipeline execution. Pipeline runs are typically instantiated by passing the arguments to the parameters that are defined in pipelines. The arguments can be passed manually or within the trigger definition.

Parameters

Parameters are key-value pairs of read-only configuration. Parameters are defined in the pipeline. The arguments for the defined parameters are passed during execution from the run context that was created by a trigger or a pipeline that was executed manually. Activities within the pipeline consume the parameter values.

A dataset is a strongly typed parameter and a reusable/referenceable entity. An activity can reference datasets and can consume the properties that are defined in the dataset definition.

A linked service is also a strongly typed parameter that contains the connection information to either a data store or a compute environment. It is also a reusable/referenceable entity.

Control flow

Control flow is an orchestration of pipeline activities that includes chaining activities in a sequence, branching, defining parameters at the pipeline level, and passing arguments while invoking the pipeline on-demand or from a trigger. It also includes custom-state passing and looping containers, that is, For-each iterators.

Variables

Variables can be used inside of pipelines to store temporary values and can also be used in conjunction with parameters to enable passing values between pipelines, data flows, and other activities.

CREATING ALERTS

Subscription Details | Nuvepro tansssssss - Microsoft Azure WebContentNotFound

https://portal.azure.com/#@npunext.onmicrosoft.com/resource/subscriptions/324a722... ...

Microsoft Azure Search resources, services, and docs (G+) Shellunext_1693422092... UNEXT (NPUNEXT.ONMICROSOFT.COM)

Home > tanstorage1

tanstorage1 | Alerts Storage account

View as tile

al

Security + networking

Encryption

Settings

Resource sharing (CORS)

Monitoring

Alerts **1**

Logs

Help

Resource health

Name ↑

tansssssss

Showing 1 - 1

tanssssssss Alert details

Summary History

Severity: 3 - Informational Fired time: 9/5/2023, 11:52 AM Affected resource: tanstorage1 Hierarchy: npunext-16735054... > tanrg1

User response: New Alert condition: Fired Change user response

Why did this alert fire?

Egress (Sum) tanstorage1 41.21 kB

0B 4.88kB 9.77kB 14.65kB 19.53kB 24.41kB

11:30 AM 11:45 AM UTC+05:30

Evaluation window start time (for which alert fired): 9/5/2023, 11:45 AM

Evaluation window end time (for which alert fired): 9/5/2023, 11:50 AM

https://portal.azure.com/#

The screenshot shows the Microsoft Azure portal interface. On the left, there's a navigation sidebar with links for Home, tanstorage1, Security + networking, Encryption, Settings, Resource sharing (CORS), Monitoring, Alerts (which is selected and shows 1 alert), Logs, Help, and Resource health. The main content area is titled 'tanstorage1 | Alerts' and shows an alert for 'tanssssssss'. The alert summary indicates it's an informational alert that fired at 9/5/2023, 11:52 AM, affecting the resource 'tanstorage1'. The alert condition was 'Fired'. Below the summary, there's a section titled 'Why did this alert fire?' with a line chart showing the trend of egress traffic from 'tanstorage1'. The chart has a y-axis labeled 'Egress (Sum) tanstorage1' with values 0B, 4.88kB, 9.77kB, 14.65kB, 19.53kB, and 24.41kB. The x-axis shows times: 11:30 AM, 11:45 AM, and UTC+05:30. A blue line starts at 0B at 11:30 AM, remains flat until 11:45 AM, and then rises sharply to 41.21 kB at 11:50 AM. A red shaded area highlights the period from 11:45 AM to 11:50 AM. At the bottom of the alert details page, it says 'Evaluation window start time (for which alert fired): 9/5/2023, 11:45 AM' and 'Evaluation window end time (for which alert fired): 9/5/2023, 11:50 AM'.

AZURE DATA FACTORY

The screenshot shows the Microsoft Azure portal interface. The URL in the address bar is <https://portal.azure.com/#@npunext.onmicrosoft.com/resource/subscriptions/324a7224-f028-40f6-9a19-b8b16ed4f77a/resourceGroups/tanrg3/providers/Microsoft.DataFactory/factories/tanya11>. The page title is "tanya11 - Azure Data Factory".

The left sidebar shows the navigation path: Home > tanrg3 > tanya11 (Data factory (V2)). The main content area displays the "Essentials" section for the "tanya11" Data factory (V2). The essentials include:

- Resource group ([move](#)): tanrg3
- Type: Data factory (V2)
- Status: Succeeded
- Getting started: [Quick start](#)
- Location: East US
- Subscription ([move](#)): npunext-1673505441526
- Subscription ID: 324a7224-f028-40f6-9a19-b8b16ed4f77a

Below the essentials, there is a large blue icon of a factory building with smokestacks, followed by the text "Azure Data Factory Studio" and a "Launch studio" button.

The bottom of the screen shows the Windows taskbar with various pinned icons and the system tray indicating the date and time as 9/5/2023 at 3:00 PM.

Creating data pipeline

The screenshot shows the Microsoft Azure Data Factory pipeline editor interface. The left sidebar displays 'Factory Resources' with sections for Pipelines, Datasets, Data flows, and Power Query. The 'Pipelines' section shows one pipeline named 'pipeline1'. The main workspace shows a pipeline named 'pipeline1' with a single activity named 'DelimitedText1'. This activity is currently set to 'Copy data'. The pipeline editor includes tabs for Parameters, Variables, Settings, and Output, along with buttons for Validate, Debug, and Add trigger. The status bar at the bottom shows the date and time as 5:17 PM on 9/5/2023.

Creating Virtual Machine

The screenshot shows the Microsoft Azure portal interface. The top navigation bar has several tabs open, including 'Subscription', 'Home - Micr...', 'tanstorage1...', 'tanvm2 - Mi...', 'portal.azure...', 'tanya11 - Az...', 'Lumen...', and a '+' button. The main title bar says 'Microsoft Azure' and 'Search resources, services, and docs (G+)'. A user profile 'Shellunext_1693422092... UNEXT' is visible.

The main content area displays the details for a virtual machine named 'tanvm2'. The left sidebar shows navigation links like 'Overview', 'Activity log', 'Access control (IAM)', 'Tags', 'Diagnose and solve problems', 'Networking', 'Connect', 'Disks', 'Size', 'Microsoft Defender for Cloud', 'Advisor recommendations', 'Extensions + applications', 'Availability + scaling', 'Configuration', and 'Identity'. The 'Overview' section is currently selected.

Essentials (JSON View)

Resource group (move) tanrg2	Operating system Windows (Windows 10 Pro)
Status Running	Size Standard D2s v3 (2 vcpus, 8 GiB memory)
Location East US (Zone 1)	Public IP address 172.174.240.112
Subscription (move) npunext-1673505441526	Virtual network/subnet tanvm2-vnet/default
Subscription ID 324a7224-f028-40f6-9a19-b8b16ed4f77a	DNS name Not configured
Availability zone 1	Health state -
Tags (edit) Add tags	

Properties [Monitoring](#) [Capabilities \(7\)](#) [Recommendations](#) [Tutorials](#)

Virtual machine

Computer name	tanvm2
Operating system	Windows (Windows 10 Pro)

Networking

Public IP address	172.174.240.112 (Network tanvm2243_z1)
Interface	

Bottom status bar: 78°F Mostly cloudy, Search, File Explorer, Task View, 5:18 PM, 9/5/2023, battery icon.

Configuring to Github

The screenshot shows the Azure Data Factory interface with the 'Git configuration' section selected in the left sidebar. Two windows are open, both titled 'Configure a repository'.

Left Window (Main View):

- General:** Repository name: Tanya-shell-unext_single, Collaboration branch: main, Publish branch: adf_publish, Root folder: /
- Custom comment:** Use custom comment checked
- Import existing resources:** Import existing resources to repository checked
- Import resource into this branch:** (empty input field)
- Buttons:** Apply, Back, Cancel

Right Window (Details View):

- Repository type:** GitHub
- GitHub account:** Tanya-Yadav
- Repository name:** Tanya-shell-unext_single
- Collaboration branch:** main
- Publish branch:** adf_publish
- Root folder:** /
- Last published commit:** cc6c90902635b9ecd7bbd31a0d43b0ed8e46e70c
- Publish (from ADF Studio):** Enabled
- Custom comment:** Enabled

Configuring with Github

The screenshot shows a GitHub repository page for 'Tanya-shell-unext_single'. The repository is public and has 2 branches and 0 tags. A message indicates that 'adf_publish' had recent pushes about 1 hour ago. The main branch is not protected. The commit history shows several commits from 'Tanya-yadav' updating a pipeline named 'PL_Array'. The repository has 6 commits in total, with the most recent being 1 hour ago. The repository has 0 stars, 1 watching, and 0 forks. There are no releases or packages published.

InPrivate Tanya-yadav/Tanya-shell-unext_single + https://github.com/Tanya-yadav/Tanya-shell-unext_single/tree/main

Tanya-yadav / Tanya-shell-unext_single

Type / to search

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

Tanya-shell-unext_single Public

Pin Unwatch 1 Fork 0 Star 0

adf_publish had recent pushes about 1 hour ago

Compare & pull request

main 2 branches 0 tags

Go to file Add file Code

Your main branch isn't protected

Protect this branch

Tanya-yadav Updating pipeline: PL_Array

f140b46 1 hour ago 6 commits

File	Commit Message	Time
factory	Adding linkedService: AzureDataLakeStorage1	3 hours ago
linkedService	Adding linkedService: AzureDataLakeStorage1	3 hours ago
pipeline	Updating pipeline: PL_Array	1 hour ago
index	Create index	3 hours ago
publish_config.json	Update publish_config.json	3 hours ago

About

No description, website, or topics provided.

Activity

0 stars

1 watching

0 forks

Releases

No releases published

Create a new release

Packages

No packages published

Publish your first package

Creating Pipeline

The screenshot shows the Microsoft Azure Data Factory pipeline creation interface. Two windows are open, both titled "PL_Array".

Left Window (Main View):

- Properties:** General tab selected, Name: PL_Array.
- Output:** Pipeline run ID: 1d315a5b-f760-4cc0-a763-5a6472b424a8.
- Activities:** A single activity named "Filter1" is listed under "All status".
- Annotations:** Shows annotations for the Filter1 activity.

Right Window (Detailed View):

- Properties:** General tab selected, Name: PL_Array.
- Settings:**
 - Items:** @pipeline().parameters.Tanyaadav
 - Condition:** @not>equals(item(), 'ccc')

Working on filters

The screenshot shows two instances of the Microsoft Azure Data Factory Data Flow interface side-by-side.

Left Window: This window displays a data flow pipeline named "dataflowinput". It consists of a single step: "filterYear". The "Properties" panel indicates that the output stream name is "filterYear" and the column count is "8 total". The "Data preview" tab shows the following filter expression: `toInteger(year) >= 1910 && toInteger(year) <= 2000`. The "Filter settings" tab shows the incoming stream is "source1".

Film	Genre	Lead Studio	Audience	Profit
Zack and...	Romance	The Wein...	70	1.7475
Youth in ...	Comedy	The Wein...	52	1.09
You Will ...	Comedy	Independ...	35	1.2118
When in ...	Comedy	Disney	44	0
What Ha...	Comedy	Fox	72	6.2676
Water Fou...	Drama	20th Cen...	72	3.0814
WALL-E	Animation	Disney	89	2.8960
Waitress	Romance	Independ...	67	11.089
Waiting F...	Romance	Independ...	53	0.005

Right Window: This window also displays the same "dataflowinput" pipeline. The "Properties" panel shows the "Name" is "dataflowinput". The "Data preview" tab shows the same filtered data as the left window.

Film	Genre	Lead Studio	Audience	Profit
Zack and...	Romance	The Wein...	70	1.7475
Youth in ...	Comedy	The Wein...	52	1.09
You Will ...	Comedy	Independ...	35	1.2118
When in ...	Comedy	Disney	44	0
What Ha...	Comedy	Fox	72	6.2676
Water Fou...	Drama	20th Cen...	72	3.0814
WALL-E	Animation	Disney	89	2.8960
Waitress	Romance	Independ...	67	11.089
Waiting F...	Romance	Independ...	53	0.005

Working on filters

The screenshot shows the Microsoft Azure Data Factory authoring interface. The pipeline is named "dataflowinput". The data flow consists of the following stages:

- source1**: Imports data from DelimitedText1.
- filteryear**: Filters rows using expressions on column "Year".
- filter1**: Filters rows using expressions on column "Genre".
- filter2**: Filters rows using expressions on column "8 total".
- sink1**: Exports data.

The "Data preview" tab shows the following data:

Film	Genre	Lead Stu...	Audienc...	Profitabil...	Rotten T...	Worldwi...
Miss Pett...	Comedy	Independ...	70	0.2528949	78	\$15.17

Properties pane:

- Name**: dataflowinput
- Description**: (empty)

Bottom navigation bar:

- 79°F Partly sunny
- Search
- File Explorer
- Cloud Explorer
- Task List
- 12:14 PM 9/7/2023

Working on web

The screenshot shows the Microsoft Azure Data Factory pipeline editor interface. On the left, the 'Factory Resources' sidebar lists Pipelines, Datasets, Data flows, and Power Query. The main workspace displays a pipeline named 'pipeline1' with a single 'Web' activity named 'Web1'. The 'Properties' pane on the right shows the activity's name as 'PL_webapi'. The 'Settings' tab for the Web activity is open, showing the following configuration:

- Method:** GET
- Authentication:** None
- Headers:** (empty)
- Advanced:** (empty)

The pipeline editor also includes tabs for 'Validate all', 'Publish all', and 'Preview experience'.

New linked service configuration pane (right side):

- Name:** HttpServer1
- Description:** (empty)
- Connect via integration runtime:** AutoResolveIntegrationRuntime
- Base URL:** https://atlas.microsoft.com/weather/currentConditions/json?api-version=1.0&query=47.60
- Information:** Information will be sent to the URL specified. Please ensure you trust the URL entered.
- Server Certificate Validation:** Enable Disable
- Authentication type:** Basic
- User name:** tan
- Password:** (redacted)
- Azure Key Vault:** (disabled)

At the bottom right, there are 'Create' and 'Cancel' buttons, and a 'Test connection' link.

Connected to SQL Database

The image displays two side-by-side browser windows illustrating the connection and management of a SQL database.

Left Window (Microsoft Azure - Overview):

- Deployment:** Microsoft.SQLDatabase.newDatabaseNewServer_ade7efcb818c44d3bc614 | Overview
- Deployment Status:** Deployment is in progress
- Deployment Details:**
 - Deployment name: Microsoft.SQLDatabase.newDatabaseNewServer_ade7efcb818c44d3bc614
 - Start time: 9/8/2023; 10:15:09 AM
 - Subscription: npunext-1673505441526
 - Correlation ID: bdc609d7-eae5-4bde-b76...
 - Resource group: tan08septrg
- Deployment details table:**

Resource	Type	Status
tan08sept	SQL server	Accepted

Right Window (Microsoft Azure - Data Factory):

- Linked services:** New linked service
- Form Fields (Azure SQL Database):**
 - Name: AzureSqlDatabase1
 - Description: (empty)
 - Connect via integration runtime: AutoResolveIntegrationRuntime
 - Account selection method: From Azure subscription
 - Azure subscription: npunext-1673505441526 (324a7224-f028-40f6-9a19-b8b16ed4f77a)
 - Server name: tan08sept
 - Database name: tan08septsql
 - Authentication type: SQL authentication
 - User name: (empty)

Worked on Query Editor

The screenshot displays two separate instances of the Microsoft Azure Query Editor (preview) interface.

Left Window (Main View):

- Title Bar:** tan08septsql (tan08sept/tan08septsql) | Query editor (preview)
- Toolbar:** Search, Overview, Activity log, Tags, Diagnose and solve problems, Query editor (preview).
- Alert:** Your local network settings might be preventing the Query Editor from issuing queries. Please click here for instructions on how to configure your network settings.
- Query Editor:** Shows the creation of a table:

```
1  create table data_source_table
2  (
3      PersonID int,
4      Name varchar(255),
5      LastModifytime datetime
6 );
```
- Results Tab:** Shows the table name: data_source_table. A message indicates the query succeeded.

Right Window (Inset View):

- Title Bar:** tan08septsql (tan08sept/tan08septsql) | Query editor (preview)
- Toolbar:** Search, Overview, Activity log, Tags, Diagnose and solve problems, Query editor (preview).
- Alert:** Your local network settings might be preventing the Query Editor from issuing queries. Please click here for instructions on how to configure your network settings.
- Query Editor:** Shows a select query:

```
1
2
3  select * from data_source_table;
```
- Results Tab:** Shows the results of the select query:

PersonID	Name	LastModifytime
1	aaaa	2017-09-01T00:56:00.0000000
2	bbbb	2017-09-02T05:23:00.0000000
3	cccc	2017-09-03T02:36:00.0000000

A message indicates the query succeeded.

WEEK-5

Introduction to Azure Synapse Analytics

Large scale Data
warehousing
Advanced Analytics
Real time Analytics
Data Integration

The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes the URL `portal.azure.com/#@npunext.onmicrosoft.com/resource/subscriptions/324a7224-f028-40f6-9a19-b8b16ed4f77a/resourceproviders`. The main content area displays the 'Resource providers' page for the subscription 'npunext-1673505441526'. A search bar at the top right contains the prefix 'syn'. The page lists three resource providers:

Provider	Status
Microsoft.Synapse	Registered
Microsoft.StorageSync	Not Registered
Microsoft.Syntex	Not Registered

On the left, there is a sidebar with sections for 'Subscriptions' (listing 'npunext-1673505441526'), 'Cost analysis', 'Cost alerts', 'Budgets', 'Advisor recommendations', 'Billing' (with 'Billing profile invoices'), and 'Settings' (with 'Programmatic deployment', 'Resource groups', 'Resources', and 'Preview features').

Components of Azure Synapse Analytics

Azure Synapse SQL – give concept of analytical services.

Synapse Pipelines

Synapse link – Use to establish communication with Cosmos DB.

Synapse Studio

Synapse spark

Synapse Architecture

Highly scalable

Common libraries and multiple programming language

Integrated notebook experience

Explore data with data explorer

Introduction to Azure Synapse Analytics

65.0.107.73 - Remote Desktop Connection

Home > tan12septrg > Marketplace > Azure Synapse Analytics > Create Synapse workspace

Basics *Security Networking Tags Review + create

Create a Synapse workspace to develop an enterprise analytics solution in just a few clicks.

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all of your resources.

Subscription * Resource group * Managed resource group

Workspace details

Name your workspace, select a location, and choose a primary Data Lake Storage Gen2 file system to serve as the default location for logs and job output.

Workspace name * Region * Select Data Lake Storage Gen2 * From subscription Manually via URL

Account name * File system name *
 Assign myself the Storage Blob Data Contributor role on the Data Lake Storage Gen2 account to interactively manage it in this workspace.

Review + create < Previous Next >

https://npunext.azure.com/#/dashboards

RDP

65.0.107.73 - Remote Desktop Connection

Home > tan12septrg > Marketplace > Azure Synapse Analytics > Create Synapse workspace ...

Validation succeeded

Basics *Security Networking Tags **Review + create**

Product Details

Azure Synapse Analytics workspace by Microsoft Serverless SQL est. cost/TB ⓘ --

[Terms of use](#) | [Privacy policy](#)

Terms

By clicking Create, I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. For additional details see [Azure Marketplace Terms](#).

Basics

Subscription	npunext-1673505441526
Resource group	tan12septrg
Region	East US
Workspace name	(new) tan12sept
Data Lake Storage Gen2 account	(new) https://tan12septac.dfs.core.windows.net
Data Lake Storage Gen2 file system	(new) tan12septfs
Managed resource group	None
Role assignments	The Storage Blob Data Contributor role will be assigned on the specified Data Lake Storage Gen2 account to both the workspace managed identity and the current user.

Create < Previous Next > Download a template for automation

Working on SQL Script

65.0.107.73 - Remote Desktop Connection

Applications New Tab - Google Chrome yellow_tripdata_2023-01.parquet tan12sept - Azure Synapse TLC Trip Record Data - TLC

Tue 12 Sep. 09:21

Microsoft Azure | Synapse Analytics | tan12sept

We use optional cookies to provide a better experience. Learn more Accept Reject More options

Synapse live Validate all Publish all 1

Develop + SQL scripts SQL script 1

Other users in your workspace may have access to modify this item. Do not use this item unless you trust all users who may have access to the workspace.

Run Undo Publish Query plan Connect to Built-in Use database master

```
1 SELECT TOP 100 *
2 FROM OPENROWSET(
3     BULK 'https://tan12septac.dfs.core.windows.net/tan12septfs/yellow_tripdata_2023-01.parquet',
4     FORMAT = 'PARQUET'
5 ) AS[Result]
```

Properties General Related (0)

Name * SQL script 1

Description

Run Undo Publish Query plan Connect to Built-in Use database master

```
1 SELECT TOP 100 *
2 FROM OPENROWSET(
3     BULK 'https://tan12septac.dfs.core.windows.net/tan12septfs/yellow_tripdata_2023-01.parquet',
4     FORMAT = 'PARQUET'
5 ) AS[Result]
```

Results Messages

View Table Chart Export results

Search

VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance	RatecodeID	store_and_fwd_flag	PULocationID	DOLocationID
2	2023-01-01T00:00:00	2023-01-01T00:00:00	1	0.97	1	N	161	141
2	2023-01-01T00:00:00	2023-01-01T01:00:00	1	1.1	1	N	43	237
2	2023-01-01T00:00:00	2023-01-01T00:00:00	1	2.51	1	N	48	238
1	2023-01-01T00:00:00	2023-01-01T00:00:00	0	1.9	1	N	138	7
2	2023-01-01T00:00:00	2023-01-01T00:00:00	1	1.43	1	N	107	79
2	2023-01-01T00:00:00	2023-01-01T01:00:00	1	1.84	1	N	161	137
2	2023-01-01T00:00:00	2023-01-01T00:00:00	1	1.66	1	N	239	143
2	2023-01-01T00:00:00	2023-01-01T00:00:00	1	11.7	1	N	142	200
2	2023-01-01T00:00:00	2023-01-01T00:00:00	1	2.95	1	N	164	236
2	2023-01-01T00:00:00	2023-01-01T00:00:00	1	3.01	1	N	141	107
2	2023-01-01T00:00:00	2023-01-01T01:00:00	1	1.8	1	N	234	68
1	2023-01-01T00:00:00	2023-01-01T01:00:00	4	7.3	1	N	79	264

84°F Partly sunny

Search

Line

(none)

VendorID, tpep_pickup_datetime, tp...

bottom - center

Legend (series) label

Category label

0 3 6 9 12 15 18 21 24 27 30 33 36 39 42 45 48 51 54 57 60 63 66 69 72 75 78 81 84 87 90 93 96 99

VendorID RatecodeID payment_type mta_tax improvement_surcharge airport fee passenger_count PULocationID fare_amount extra tip_amount tolls_amount total_amount congestion surcharge

Working on SQL Script

65.0.107.73 - Remote Desktop Connection

Applications New Tab - Google Chrome

tan12sept - Microsoft Azure tan12sept - Azure Synapse TLC Trip Record Data - TLC +

web.azuresynapse.net/en/authoring/analyze/sqlscripts/SQL%20script%201?workspace=%2Fsubscriptions%2F324a7224-f028-40f6-9a19-b8b16ed4f77a%2FresourceGroups%2Ftan12septrg%2Fproviders%2FMicrosoft.SynapseAnalytics

Tue 12 Sep, 09:53

Microsoft Azure | Synapse Analytics > tan12sept

We use optional cookies to provide a better experience. Learn more ▾ Accept Reject More options

Synapse live Validate all Publish all 1

Develop + <>

SQL script 1

Other users in your workspace may have access to modify this item. Do not use this item unless you trust all users who may have access to the workspace. ×

Run Undo Publish Query plan Connect to Built-in Use database master

1 SELECT TOP 100 *
2 FROM OPENROWSET(
3 BULK 'https://tan12septac.dfs.core.windows.net/tan12septfs/yellow_tripdata_2023-01.parquet',
4 FORMAT = 'PARQUET'
5) AS[Result]
6
7 CREATE DATABASE DataexplorationDB
8 COLLATE Latin1_General_100_BIN2_UTF8
9
10 CREATE EXTERNAL DATA SOURCE tan12septfs
11 WITH (LOCATION = '<https://tan12septac.dfs.core.windows.net>')
12
13 CREATE LOGIN user_data_explorqer WITH PASSWORD='localhost@12345'
14
15 Create user usrdb_data_explorer FOR LOGIN user_data_explorer;
16 GO
17
18 GRANT ADMINISTER DATABASE BULK OPERATIONS TO usrdb_data_explorer;
19 GO
20
21 SELECT TOP 100 *
22 FROM OPENROWSET(
23 BULK '/tan12septfs/yellow_tripdata_2023-01.parquet',
24 DATA_SOURCE = 'tan12septfs',
25 FORMAT = 'PARQUET'
26) AS[Result]

Properties

General Related (0)

Name * SQL script 1

Description

Type .sql script

Size 0 bytes

Results settings per query ⓘ

First 5000 rows (default)

All rows

Working on Database diagrams

The screenshot shows the Microsoft SQL Server Management Studio interface. The title bar indicates a connection to '43.204.236.220 - Remote Desktop Connection' and the database 'EC2AMAZ-OIIIOAMA\SQLEXPRESS.sample db - Diagram_0*'.

The left pane displays the Object Explorer, showing the database structure including 'Databases', 'Tables' (with 'Customer' selected), and other system and security objects.

The central pane shows the 'Customer' table definition in the 'Table Designer'. The columns are:

Column Name	Data Type	Allow Nulls
CustomerID_PK	int	<input type="checkbox"/>
Fullname	varchar(500)	<input checked="" type="checkbox"/>
Custaddress	varchar(5000)	<input checked="" type="checkbox"/>
Phoneno	varchar(50)	<input checked="" type="checkbox"/>
	nchar(10)	<input type="checkbox"/>

The right pane shows the 'Properties' window for the 'Customer' table. Key properties include:

- (Name)**: Customer
- Database Name**: sample db
- Description**:
- Schema**: dbo
- Server Name**: ec2amaz-oiiioama\sq
- Identity Column**: CustomerID_PK
- Indexable**: Yes
- Lock Escalation**: Table
- Regular Data Spac**: PRIMARY
- Replicated**: No
- Row GUID Column**:
- Text/Image Filegr**: PRIMARY

The status bar at the bottom shows system information: 75°F Partly sunny, Search, 10:26 AM, 9/13/2023.

Running Queries

The screenshot shows the Microsoft SQL Server Management Studio (SSMS) interface. The title bar indicates the connection is to EC2AMAZ-OIIIOAMA\SQLEXPRESS.AdventureWorksLT2014 (EC2AMAZ-OIIIOAMA\Administrator (57))*. The main window displays a query results grid for the SELECT * FROM SalesLT.Customer statement. The Properties pane on the right shows connection details and aggregate status.

Object Explorer:

- EC2AMAZ-OIIIOAMA\SQLEXPRESS (SQL Server)
- Databases
 - System Databases
 - AdventureWorksLT2014
- destination
 - Database Diagrams
- Tables
 - System Tables
 - FileTables
- Views
- Synonyms
- Programmability
- Service Broker
- Storage
- Security
- ReportServer\$SQLEXPRESS
- ReportServer\$SQLEXPRESSSTempDB
- sample db (Offline)
- Security
- Server Objects
- Replication
- Management

SQLQuery1.sql - EC2AMAZ-OIIIOAMA\SQLEXPRESS.AdventureWorksLT2014 (EC2AMAZ-OIIIOAMA\Administrator (57))*

```
SELECT * FROM SalesLT.Customer
```

Properties

Current connection parameters	
Aggregate Status	
Connection failures	
Elapsed time	00:00:01.178
Finish time	9/13/2023 2:32:37 PM
Name	EC2AMAZ-OIIIOAMA'
Rows returned	847
Start time	9/13/2023 2:32:35 PM
State	Open
Connection	
Connection name	EC2AMAZ-OIIIOAMA'
Connection Details	
Connection elapsed	00:00:01.178
Connection finish	9/13/2023 2:32:37 PM
Connection rows	847
Connection start	9/13/2023 2:32:35 PM
Connection state	Open
Display name	EC2AMAZ-OIIIOAMA'
Login name	EC2AMAZ-OIIIOAMA'
Server name	EC2AMAZ-OIIIOAMA'
Server version	12.0.2000

Results

CustomerID	NameStyle	Title	FirstName	MiddleName	LastName	Suffix	CompanyName	SalesPerson	EmailAddress	Phone
1	0	Mr.	Orlando	N.	Gee	NULL	A Bike Store	adventure-works\pamela0	orlando0@adventure-works.com	245-555-0173
2	0	Mr.	Keith	NULL	Harris	NULL	Progressive Sports	adventure-works\david8	keith0@adventure-works.com	170-555-0127
3	0	Ms.	Donna	F.	Cameras	NULL	Advanced Bike Components	adventure-works\jillian0	donna0@adventure-works.com	279-555-0130
4	0	Ms.	Janet	M.	Gates	NULL	Modular Cycle Systems	adventure-works\jillian0	janet1@adventure-works.com	710-555-0173
5	0	Mr.	Lucy	NULL	Harrington	NULL	Metropolitan Sports Supply	adventure-works\shu0	lucy0@adventure-works.com	828-555-0186
6	0	Ms.	Rosmarie	J.	Carroll	NULL	Aerobic Exercise Company	adventure-works\linda3	rosmarie0@adventure-works.com	244-555-0112
7	0	Mr.	Dominic	P.	Gash	NULL	Associated Bikes	adventure-works\shu0	dominic0@adventure-works.com	192-555-0173
8	0	Ms.	Kathleen	M.	Garza	NULL	Rural Cycle Emporium	adventure-works\josé1	kathleen0@adventure-works.com	150-555-0127
9	0	Ms.	Katherine	NULL	Harding	NULL	Sharp Bikes	adventure-works\josé1	katherine0@adventure-works.com	926-555-0159
10	0	Mr.	Johnny	A.	Caprio	Jr.	Bikes and Motorbikes	adventure-works\garrett1	johnny0@adventure-works.com	112-555-0191
11	0	Mr.	Christopher	R.	Beck	Jr.	Bulk Discount Store	adventure-works\jae0	christopher1@adventure-works.com	1 (11) 500 555-0132
12	0	Mr.	David	J.	Liu	NULL	Catalog Store	adventure-works\michael9	david20@adventure-works.com	440-555-0132

Query executed... | EC2AMAZ-OIIIOAMA\SQLEXPRESS ... | EC2AMAZ-OIIIOAMA\Administr... | AdventureWorksLT2014 | 00:00:01 | 847 rows

Power BI

Power BI is a business analytics tool developed by Microsoft that helps you turn multiple unrelated data sources into valuable and interactive insights. These data may be in the form of an Excel spreadsheet or cloud-based/on-premises hybrid data warehouses. You can easily connect to all your data sources and share the insights with anyone.

- It helps build an interactable data visualization in data centers
- It allows users to transform data into visuals and share them with anyone
- It establishes a connection for Excel queries and dashboards for fast analysis
- It provides quick and accurate solutions
- It enables users to perform queries on reports using simple English words



Using SQL Server uploading data in BI Dashboard

The screenshot shows a dual-monitor setup. The left monitor displays Microsoft SQL Server Management Studio (SSMS) with a query window running against the AdventureWorksLT2014 database. The query is:

```
SELECT * FROM SalesLT.Customer
```

The results grid shows 847 rows of customer data. The right monitor displays a Microsoft Power BI dashboard titled "SalesLT.vGetAllCategories". This dashboard includes a table view of product categories and a sunburst chart.

Object Explorer (Left Monitor):

- Connected to EC2AMAZ-OIIIOAMA\SQLEXPRESS.AdventureWorksLT2014 (EC2AMAZ-OIIIOAMA\Administrator (57))
- Databases: AdventureWorksLT2014, destination, ReportServer\$SQLEXPRESS, ReportServer\$SQLEXPRESSTempDB, sample db (Offline)
- Tables: SalesLT.Customer, SalesLT.vGetAllCategories, SalesLT.vProductAndDescription, SalesLT.vProductModelCatalogDescription, SalesLT.Address, SalesLT.Customer, SalesLT.CustomerAddress, SalesLT.Product, SalesLT.ProductCategory, SalesLT.ProductDescription, SalesLT.ProductModel, SalesLT.ProductModelProductDescription, SalesLT.SalesOrderDetail, SalesLT.SalesOrderHeader, fx_ufnGetAllCategories, fx_ufnGetCustomerInformation, fx_ufnGetSalesOrderStatusText, destination, ReportServer\$SQLEXPRESS
- Views, Synonyms, Programmability, Service Broker, Storage, Security, Security, Server Objects, Replication, Management

Properties (Right Monitor):

- Current connection parameters
- Aggregate Status: Connection failures (0), Elapsed time (0), Finish time (9/1/2024 10:00:01), Name (El), Rows returned (847), Start time (9/1/2024 10:00:01), State (0)
- Connection: Connection name (El), Connection elapsed (0), Connection finish time (9/1/2024 10:00:01), Connection rows returned (847), Connection start time (9/1/2024 10:00:01), Connection state (0), Display name (El), Login name (El), Server name (El), Server version (1)
- Connection Details: BuildVersion, ErrorLog, SalesLT.Address, SalesLT.Customer, SalesLT.CustomerAddress, SalesLT.Product, SalesLT.ProductCategory, SalesLT.ProductDescription, SalesLT.ProductModel, SalesLT.ProductModelProductDescription, SalesLT.SalesOrderDetail, SalesLT.SalesOrderHeader, fx_ufnGetAllCategories, fx_ufnGetCustomerInformation, fx_ufnGetSalesOrderStatusText
- Name: The name of the connection (El)

Navigator (Right Monitor):

- EC2AMAZ-OIIIOAMA\SQLEXPRESS [4]
 - AdventureWorksLT2014 [18]
 - SalesLT.vGetAllCategories
 - SalesLT.vProductAndDescription
 - SalesLT.vProductModelCatalogDescription
 - BuildVersion
 - ErrorLog
 - SalesLT.Address
 - SalesLT.Customer
 - SalesLT.CustomerAddress
 - SalesLT.Product
 - SalesLT.ProductCategory
 - SalesLT.ProductDescription
 - SalesLT.ProductModel
 - SalesLT.ProductModelProductDescription
 - SalesLT.SalesOrderDetail
 - SalesLT.SalesOrderHeader
 - fx_ufnGetAllCategories
 - fx_ufnGetCustomerInformation
 - fx_ufnGetSalesOrderStatusText
 - destination
 - ReportServer\$SQLEXPRESS

SalesLT.vGetAllCategories (Right Monitor):

ParentProductCategoryName	ProductCategoryName	ProductCategoryID
Accessories	Bike Racks	30
Accessories	Bike Stands	31
Accessories	Bottles and Cages	32
Accessories	Cleaners	33
Accessories	Fenders	34
Accessories	Helmets	35
Accessories	Hydration Packs	36
Accessories	Lights	37
Accessories	Locks	38
Accessories	Panniers	39
Accessories	Pumps	40
Accessories	Tires and Tubes	41
Clothing	Bib-Shorts	22
Clothing	Caps	23
Clothing	Gloves	24
Clothing	Jerseys	25
Clothing	Shorts	26
Clothing	Socks	27
Clothing	Tights	28
Clothing	Vests	29
Components	Handlebars	8
Components	Bottom Brackets	9
Components	Brakes	10
Components	Chains	11
Components	Cranksets	12
Components	Derailleurs	13
Components	Forks	14

Working on the Database in Power BI

The screenshot displays two windows side-by-side. On the left is Microsoft SQL Server Management Studio (Administrator) showing a script named DB-Script.sql. The script creates a database (PetroCorpDB), uses it, creates a DrillingRigs table, inserts sample data, and creates a DrillingData table. The right window is Power BI Desktop (Untitled - Power BI Desktop) showing a report with a table visual. The table has columns RigName and Sum of DrilledDepth, with data for Rig A (1,500.00), Rig B (1,750.00), and a Total of 3,250.00. A tooltip 'Back to report' is visible over the table.

```
-- Create the database
CREATE DATABASE PetroCorpDB;
GO

-- Use the database
USE PetroCorpDB;
GO

-- Create the DrillingRigs table
CREATE TABLE DrillingRigs (
    RigID INT PRIMARY KEY,
    RigName NVARCHAR(50) NOT NULL
);
GO

-- Insert sample drilling rigs
INSERT INTO DrillingRigs (RigID, RigName)
VALUES
    (1, 'Rig A'),
    (2, 'Rig B');
GO

-- Create the DrillingData table
CREATE TABLE DrillingData (
    DataID INT PRIMARY KEY,
    RigID INT NOT NULL,
    DrilledDepth DECIMAL(10, 2)
);
```

RigName	Sum of DrilledDepth
Rig A	1,500.00
Rig B	1,750.00
Total	3,250.00

Loading Database in Power BI

The screenshot shows the Microsoft SQL Server Management Studio (SSMS) interface. On the left, a query window displays T-SQL code for creating a table and inserting sample data into the 'SalesTransactions' table. The 'Messages' pane shows successful execution with 10 rows affected. In the center, the Object Explorer shows the database structure, including the 'handson' database and its tables ('Customers', 'Products', 'SalesTransactions', 'Stores'). On the right, the Properties pane displays connection statistics: Elapsed time 00:00:00.145, Finish time 9/15/2023 9:26:48, Name EC2AMAZ-OIIIOOMA, Rows returned 0, Start time 9/15/2023 9:26:48, State Open. Below the Object Explorer, the 'Queries [4]' pane shows the results of a query against the 'Customers' table, displaying columns CustomerID, CustomerName, Email, and Phone for 7 rows of data.

```
CREATE TABLE SalesTransactions (
    TransactionID INT PRIMARY KEY,
    ProductID INT,
    StoreID INT,
    TransactionDate DATE,
    QuantitySold INT,
    Revenue DECIMAL(10, 2)
);

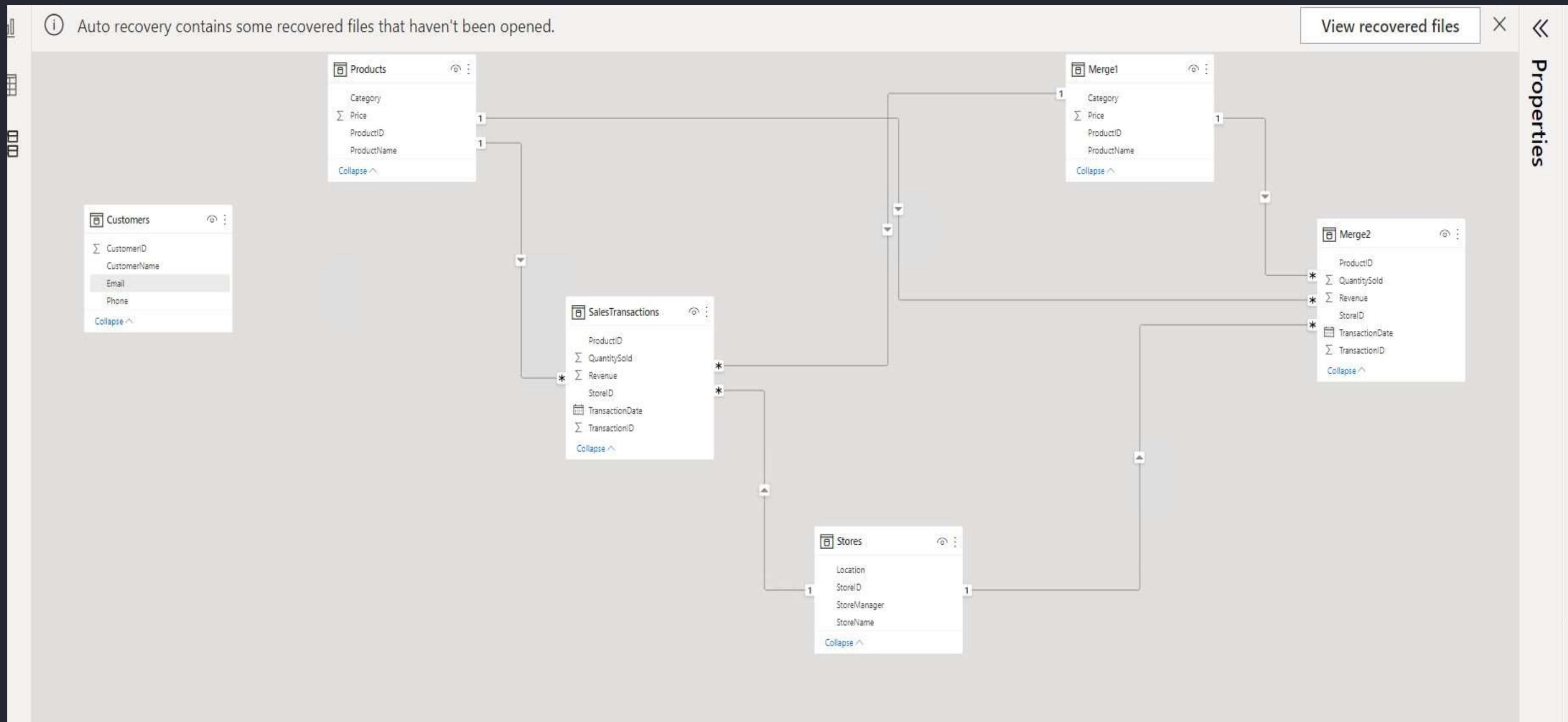
-- Insert sample data into Sales Transactions Data
INSERT INTO SalesTransactions (TransactionID, ProductID, StoreID, TransactionDate, Q
VALUES
    (1, 101, 1, '2023-01-05', 50, 500.00),
    (2, 102, 2, '2023-01-10', 40, 600.00),
    (3, 103, 1, '2023-02-15', 30, 300.00),
    (4, 104, 3, '2023-03-20', 60, 900.00),
    (5, 101, 2, '2023-01-25', 70, 700.00)
```

(10 row(s) affected)
(7 row(s) affected)
(3 row(s) affected)
(7 row(s) affected)

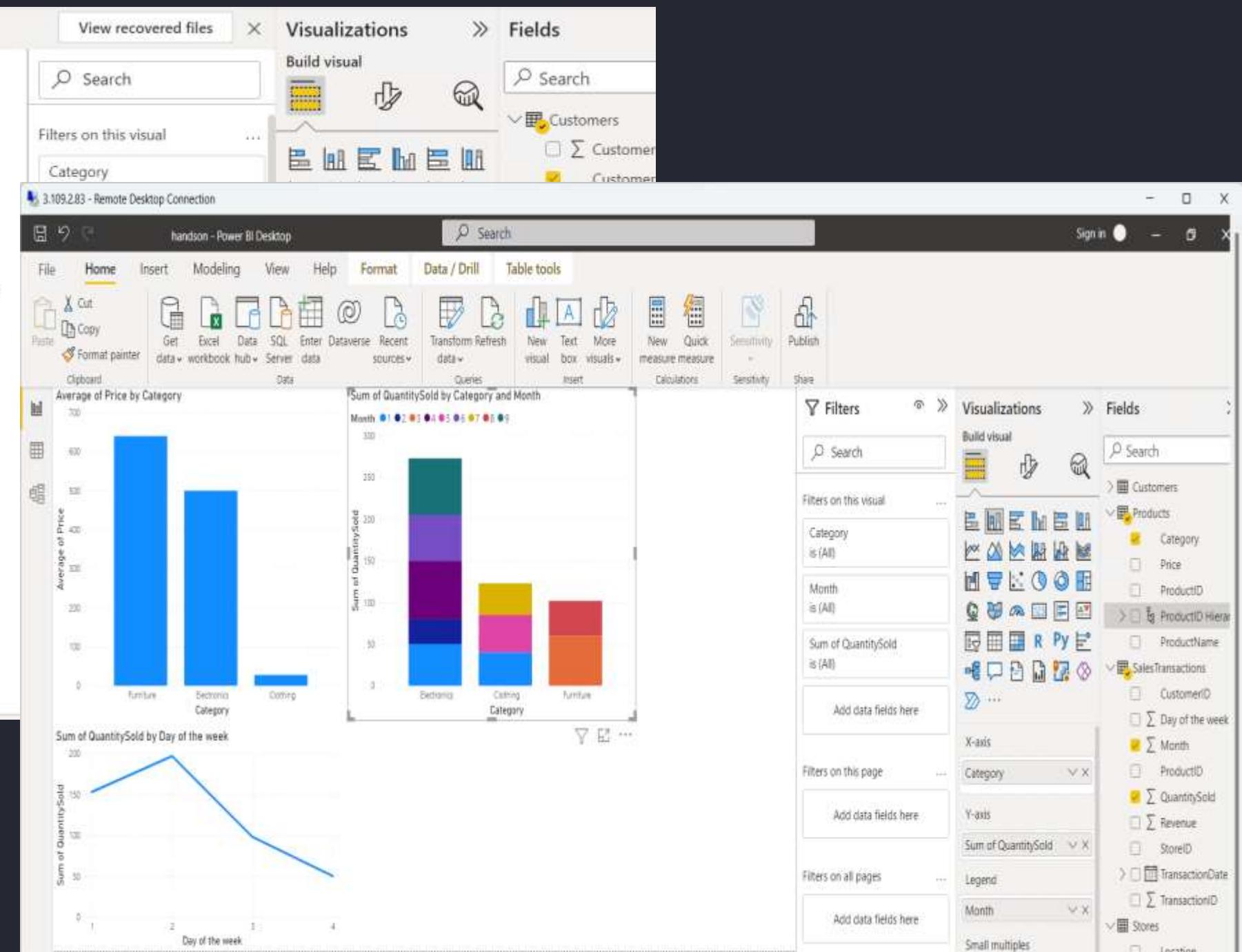
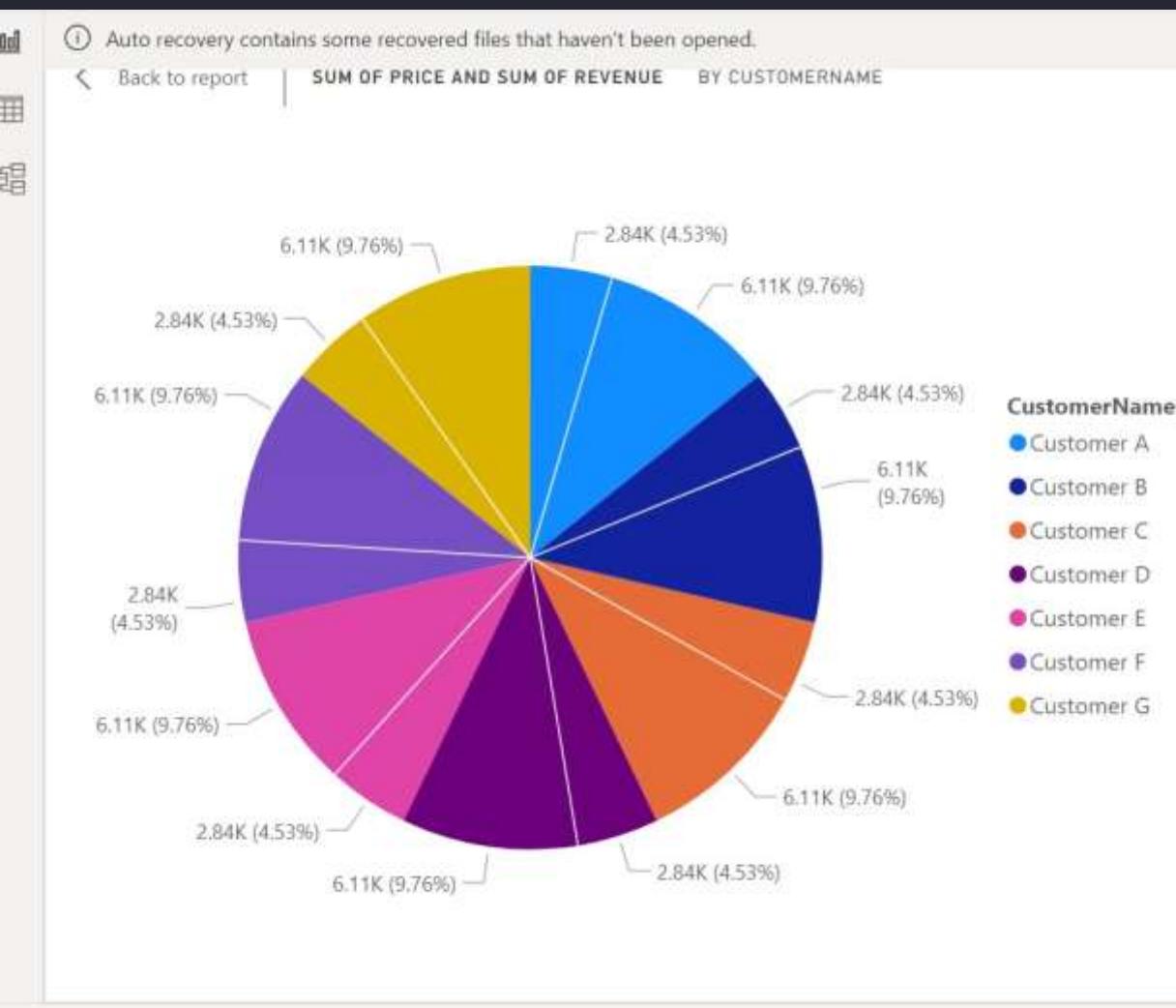
Queries [4]

	CustomerID	CustomerName	Email	Phone
1	101	Customer A	customerA@email.com	(123) 456-7890
2	102	Customer B	customerB@email.com	(234) 567-8901
3	103	Customer C	customerC@email.com	(345) 678-9012
4	104	Customer D	customerD@email.com	(456) 789-0123
5	105	Customer E	customerE@email.com	(567) 890-1234
6	106	Customer F	customerF@email.com	(678) 901-2345
7	107	Customer G	customerG@email.com	(789) 012-3456

Merging different tables



Visualizations



WEEK- 6

Python

Python is a widely-used general-purpose, high-level programming language. It was created by Guido van Rossum in 1991 and further developed by the Python Software Foundation. It was designed with an emphasis on code readability, and its syntax allows programmers to express their concepts in fewer lines of code.

The following are the benefits of using Python language:

- Object-Oriented Language
- High-Level Language
- Dynamically Typed language
- Extensive support Libraries
- Presence of third-party modules
- Open source and community development
- Portable and Interactive
- Portable across Operating systems



Basics of Python

The image shows two Jupyter Notebook windows side-by-side, illustrating basic Python operations.

Left Notebook (Untitled):

- In [1]:

```
languages=['java', 'python','c']
versions = [14,3,6]
```
- In [2]:

```
result=zip(languages,versions)
print(list(result))
[('java', 14), ('python', 3), ('c', 6)]
```
- In [3]:

```
data = {'Name':['Tanya','Surbhi','sia'], 'Age':[22,22,34]}
```
- In [6]:

```
import pandas as pd
dfemp1=pd.DataFrame(data)
print(dfemp1)
```

	Name	Age
0	Tanya	22
1	Surbhi	22
2	sia	34
- In []:

Right Notebook (Untitled):

- In [1]:

```
dfemp1=pd.DataFrame(data)
print(dfemp1)
```

	Name	Age
0	Tanya	22
1	Surbhi	22
2	sia	34
- In [7]:

```
seremplage = pd.Series(dfemp1['Age'])
```
- In [8]:

```
seremplage.head()
```

	Age
0	22
1	22
2	34

Name: Age, dtype: int64
- In [10]:

```
result = seremplage.apply(lambda x : 11 if x==19 else x)
print(result)
```

	Age
0	22
1	22
2	34

Name: Age, dtype: int64
- In [11]:

```
import numpy as np
array1= np.array ([1,3,5])
print("np.array():\n", array1)
```

np.array():
[1 3 5]
- In [12]:

```
array2= np.zeros((3,3))
print("\nnp.zeros():\n", array2)
```

np.zeros():
[[0. 0. 0.]
 [0. 0. 0.]
 [0. 0. 0.]]
- In []:

Some Python Snippets

```
[0. 0. 0.]
```

```
In [14]: array3= np.array([1,3,5,7,9,11])
array4= np.reshape(array3,(2,3))
array5= np.transpose(array4)
print("Original array:\n", array3)
print("Reshaped array:\n", array4)
print("Transposed array:\n", array5)
```

Original array:

[1 3 5 7 9 11]

Reshaped array:

[[1 3 5]

[7 9 11]]

Transposed array:

[[1 7]

[3 9]

[5 11]]

```
In [15]: array1 = np.array([[1,3,5], [2,4,6]])
np.savetxt('data.txt', array1)
loaded_data = np.loadtxt('data.txt')
print (loaded_data)
```

[1. 3. 5.]

[2. 4. 6.]

```
In [18]: a = np.array([[2,1],[4,-5]]) #2x+y=5,4x-5y=7
b= np.array([[5],[7]])
from scipy import linalg
x= linalg.solve(a,b)
print(a.dot(x)-b)
```

[[0.]

[0.]]

No products found.

Sales Management System

1. Add a new product
2. List all products
3. Add a new customer
4. List all customers
5. Create a new sales order
6. List all sales orders
7. Quit

```
In [ ]:
```

```
In [4]: import sklearn
```

```
from sklearn import tree
```

```
Bumpy = 1
```

```
Smooth =0
```

```
Orange = 0
```

```
Apple = 1
```

```
feature = [[140,Bumpy],[130,Bumpy],[150,Smooth],[170,Smooth]]
```

```
label = [Orange,Orange,Apple,Apple]
```

```
mlModel = tree.DecisionTreeClassifier()
```

```
mlModel = mlModel.fit(feature,label)
```

```
print(mlModel.predict([[160,Smooth]]))
```

PySpark

PySpark is a python-based API used for the Spark implementation and is written in Scala programming language. Basically, to support Python with Spark, the Apache Spark community released a tool, PySpark. With PySpark, one can work with RDDs in a python programming language also as it contains a library called Py4j for this. If one is familiar with Python and its libraries such as Pandas, then it is a good language to learn. It is used to create more scalable analyses and pipelines. One can opt for PySpark due to its fault-tolerant nature. Basically, it is a tool released to support Python with Spark.

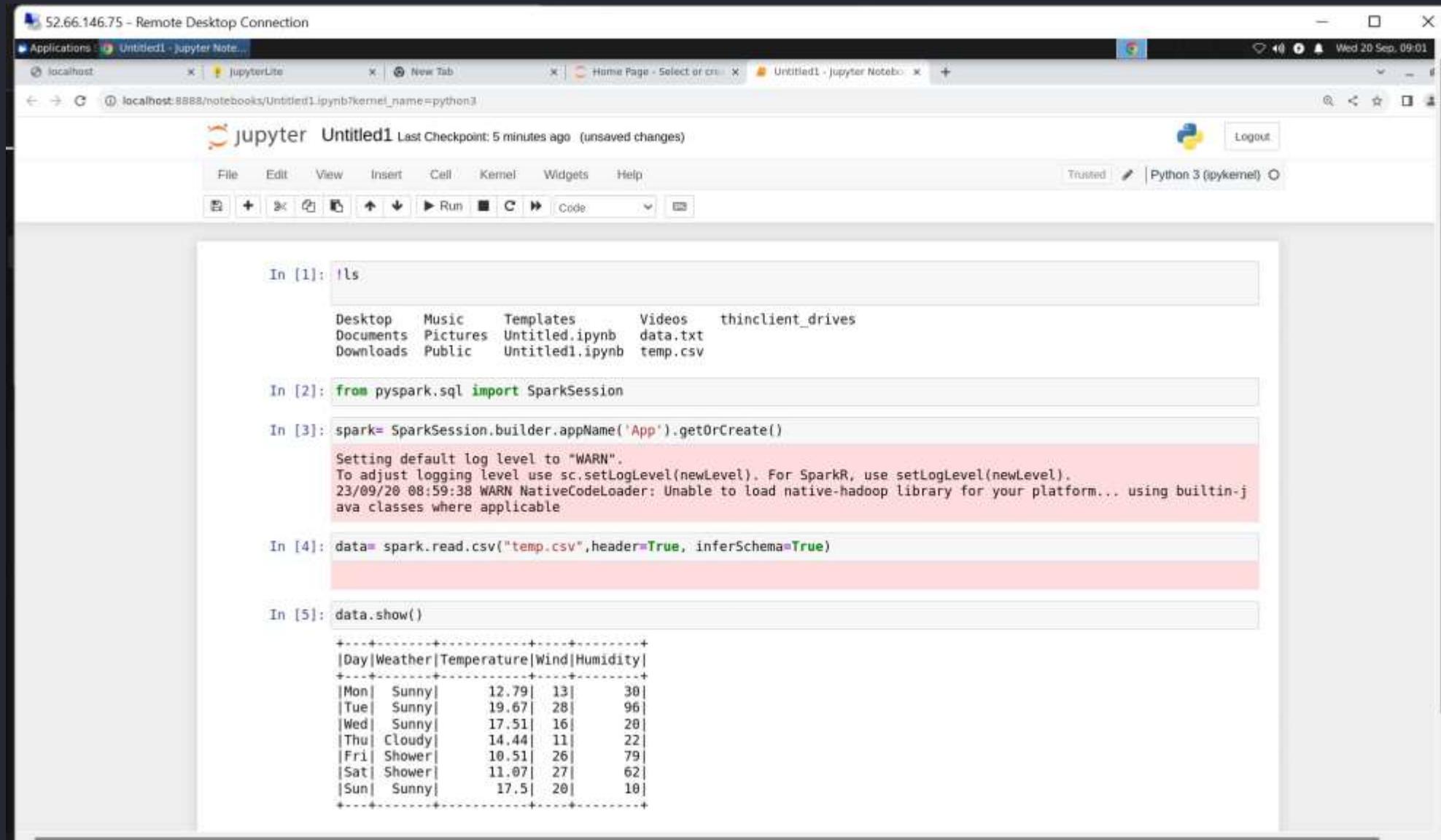
Features of PySpark

- It shows low latency.
- It is immutable.
- It is fault tolerant.
- It supports Spark, Yarn, and Mesos cluster managers.
- It has ANSI SQL support.
- It is dynamic in nature.

Limitations of PySpark

- It is hard to express.
- Less efficient
- If one requires streaming, then the user has to switch from Python to Scala.

PySpark



The screenshot shows a Jupyter Notebook interface running on a remote desktop connection. The notebook has five cells:

- In [1]: `!ls`
Output:

```
Desktop  Music  Templates  Videos  thinclient_drives
Documents  Pictures  Untitled.ipynb  data.txt
Downloads  Public  Untitled1.ipynb  temp.csv
```
- In [2]: `from pyspark.sql import SparkSession`
- In [3]: `spark= SparkSession.builder.appName('App').getOrCreate()`
Output:

```
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
23/09/20 08:59:38 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
```
- In [4]: `data= spark.read.csv("temp.csv",header=True, inferSchema=True)`
- In [5]: `data.show()`
Output:

```
+---+-----+-----+-----+
|Day|Weather|Temperature|Wind|Humidity|
+---+-----+-----+-----+
|Mon| Sunny| 12.79| 13| 30|
|Tue| Sunny| 19.67| 28| 96|
|Wed| Sunny| 17.51| 16| 20|
|Thu| Cloudy| 14.44| 11| 22|
|Fri| Shower| 10.51| 26| 79|
|Sat| Shower| 11.07| 27| 62|
|Sun| Sunny| 17.5| 20| 10|
+---+-----+-----+-----+
```

RDD (Resilient distributed dataset)

Perfect
optimization

Fault tolerant

In memory
processing

Data sharing

Compatibility

Integrate
with external
storage

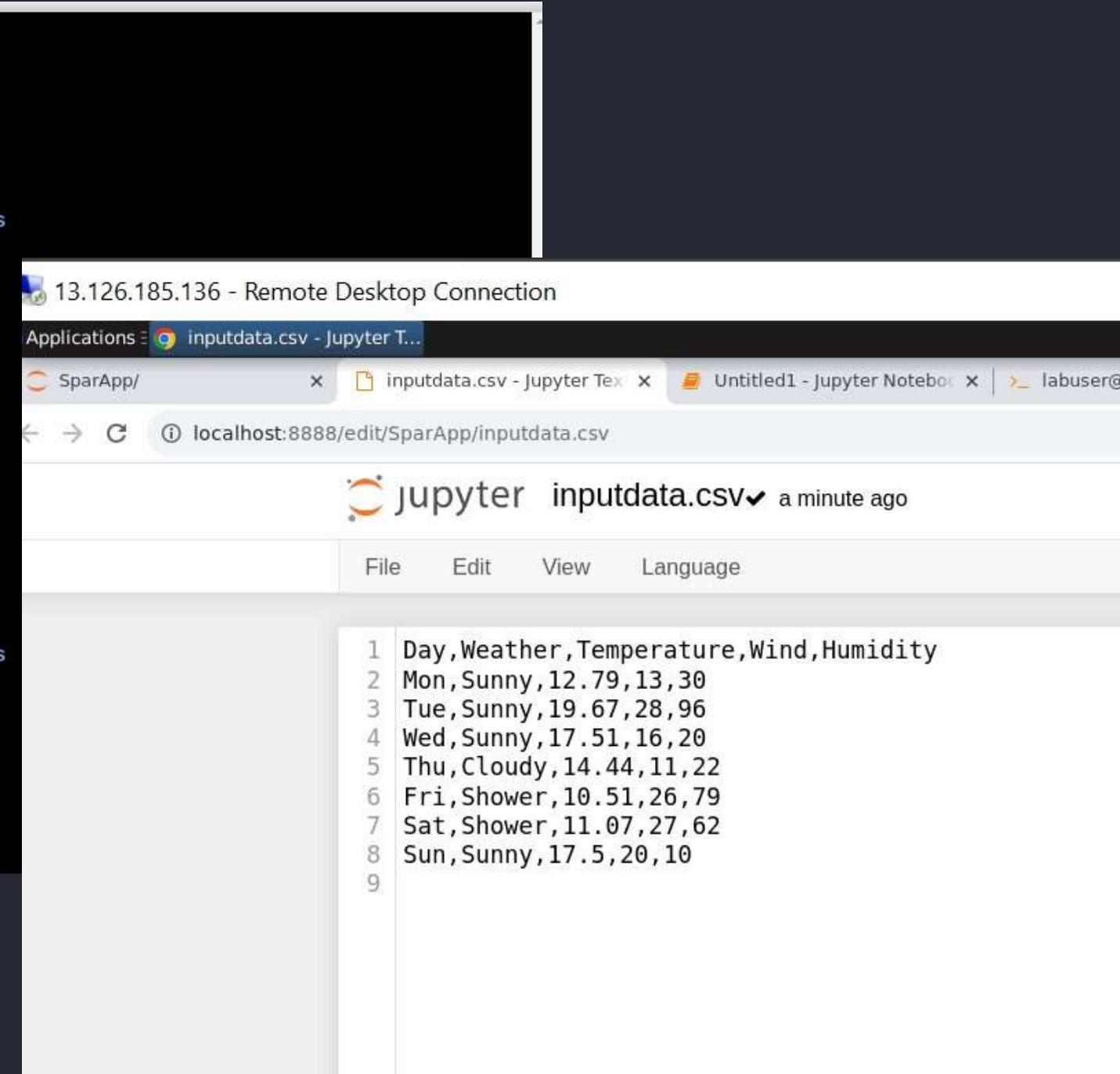
PySpark

```
1 from pyspark.sql import SparkSession  
2  
3 spark=SparkSession.builder.appName("MySparkApp").getOrCreate()  
4  
5 df=spark.read.csv("/home/labuser/Desktop/temp.csv",header=True,inferSchema=True)  
6  
7 df.write.csv("output_data.csv",header=True)  
8  
9 spark.stop()
```

```
1 from pyspark.sql import SparkSession  
2  
3 spark=SparkSession.builder.appName("RDDMapExample").getOrCreate()  
4 data=[1,2,3,4,5]  
5 rdd=spark.sparkContext.parallelize(data)  
6  
7 def double(x):  
8     return x*2  
9  
10 result_rdd=rdd.map(double)  
11 result=result_rdd.collect()  
12 print(result)  
13 spark.stop()
```

Working on Jupiter Lab

```
(base) labuser@ip-172-31-5-179:~$ cd Dekstop  
bash: cd: Dekstop: No such file or directory  
(base) labuser@ip-172-31-5-179:~$ cd Dekstop/  
bash: cd: Dekstop/: No such file or directory  
(base) labuser@ip-172-31-5-179:~$ mkdir SparApp  
(base) labuser@ip-172-31-5-179:~$ ls  
Desktop Downloads Pictures SparApp Untitled.ipynb Videos temp.csv  
Documents Music Public Templates Untitled1.ipynb data.txt thinclient_drives  
(base) labuser@ip-172-31-5-179:~$ cd SparApp  
(base) labuser@ip-172-31-5-179:~/SparApp$ cat > SparApp.py  
from pyspark.sql import SparkSession  
spark= SparkSession.builder.appName('App').getOrCreate()  
data= spark.read.csv("temp.csv",header=True, inferSchema=True)  
df.write.csv("output_data.csv", header=True)  
spark.stop()  
^Z  
[1]+ Stopped cat > SparApp.py  
(base) labuser@ip-172-31-5-179:~/SparApp$ ls ../../  
labuser ubuntu  
(base) labuser@ip-172-31-5-179:~/SparApp$ ls  
SparApp.py  
(base) labuser@ip-172-31-5-179:~/SparApp$ ls ~  
Desktop Downloads Pictures SparApp Untitled.ipynb Videos temp.csv  
Documents Music Public Templates Untitled1.ipynb data.txt thinclient_drives  
(base) labuser@ip-172-31-5-179:~/SparApp$ cp ~/temp.csv inputdata.csv  
(base) labuser@ip-172-31-5-179:~/SparApp$ ls  
SparApp.py inputdata.csv  
(base) labuser@ip-172-31-5-179:~/SparApp$
```



The screenshot shows a Jupyter Notebook interface with a single cell containing the following code:

```
jupyter inputdata.csv a minute ago  
File Edit View Language
```

The notebook displays the contents of the 'inputdata.csv' file, which contains the following data:

	Day	Weather	Temperature	Wind	Humidity
1	Mon	Sunny	12.79	13	30
2	Tue	Sunny	19.67	28	96
3	Wed	Sunny	17.51	16	20
4	Thu	Cloudy	14.44	11	22
5	Fri	Shower	10.51	26	79
6	Sat	Shower	11.07	27	62
7	Sun	Sunny	17.5	20	10
8					
9					

Hands On

```
In [1]: from pyspark import SparkContext
```

```
In [2]: sc= SparkContext('local','Sparkpractice')
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
23/09/22 04:17:28 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
```

```
In [3]: data=[1,2,3,4,5]
rdd = sc.parallelize(data)
result = rdd.collect()
print(result)
```

```
[Stage 0:> (0 + 1) / 1]
[1, 2, 3, 4, 5]
```

```
In [4]: def f(x): print(x)
```

```
In [6]: names =sc.parallelize(['Adam','tanya','surbhi','mark','john'])
```

```
In [7]: names.collect()
```

```
Out[7]: ['Adam', 'tanya', 'surbhi', 'mark', 'john']
```

```
In [8]: num = sc.parallelize([5,5,4,3,2,9,6],3)
num.getNumPartitions()
```

```
Out[8]: 3
```

```
In [ ]:
```