



Journey Presentation

Harsh Thakur

Harsh.Thakur@shell.com

Day-1 Data Fundamentals

- What Is Data?
- Types Of Data
- What Is Database?
- What Does Data Engineer Do?
- Normalization And Denormalization
- Normalization Forms (1NF,2NF,3NF,BCNF)
- Dependencies (Partial, Transitive, Functional)
- Attribute Keys (Candidate, Primary, Foreign, Super)
- Anomalies- Insert, Update, Delete
- Data Modelling And ER Diagram
 - Activity- To Make An ER Diagram For Ordering A Book From Online Store, Perform Normalization
- Dimensional Modelling
- Fact Table And Dimension Table
- Star Schema & Snowflake Schema
 - Activity- Implement The Concepts Of Star And Snowflake Schema By Drawing Respective Diagrams
- SCD (Slowly Changing Dimension)

Day-2 and 3 Big Data

- Big Data
- Extract, Load, Transform(ETL) Pipelines
- How To Process Big Data?
- Hadoop
- PySpark
- Parallel Processing And Distributive Computing
- Hadoop Used For Batch Processing & PySpark Used For Stream Processing

○ Types Of Data Structure

- Structured Data
- Semi-structured Data
- Unstructured Data

○ Data Warehouse -

- Save Data As Object
 - Tables
 - Views
- Structured Data (Fixed Schema-data Allocated Even If Value Is Null)
- Dimension Modelling
- ACID Property In DBMS -
 - Atomicity
 - Consistency
 - Isolation
 - Durability

- Data Lakehouse
 - Combination Of Data Warehouse And Datalake.
 - Took Advantages Of Both Data Warehouse And Datalake.
- Creating A VM Using Microsoft Azure.
- What Is Cloud Computing.
- Types Of Cloud -
 - Public Cloud
 - Private Cloud
 - Hybrid Cloud
- Cloud Benefits -
- High Availability
- Scalability
- Agility
- Fault Tolerance
- Elasticity
- Security
- Disaster Recovery
- Customer Latency Capabilities

- Compare Capex(capital Expenditure) Vs OpEx(operational Expenditure) -
 - Cloud Is OpEx.
 - Infrastructure As A Service(IAAS)
 - Platform As A Service(PAAS)
 - Software As A Service(SAAS)
 - Shared Responsibility Model
 - Serverless Computing
-
- Azure Architecture
 - Regions
 - Region Pairs
 - Availability Zones
 - Azure Resources
 - Azure Subscriptions -
 - Billing Boundary
 - Access Control Boundary
 - Azure Resource Manager
 - Resource Groups
 - Subscriptions -> Resource Groups

Day-3 Azure SQL

- Setting Up Azure SQL Services By Logging In On Azure Portal.
- SQL Deployment Options For Azure SQL:
 - SQL Database
 - SQL Managed Instances
 - SQL Virtual Machines
- Select SQL Database -> Single Database -> Create Database.
- Introduction To SQL
- Database Created On Microsoft Azure SQL.
- Running Query Editor On Our Server.

- DDL (Data Definition Language): create, Alter, Drop, Truncate
- DML(data Manipulation Language): Select, Insert, Update, Delete
- DCL(data Control Language):grant, REVOKE
- Constraints
- View: View Is A Virtual Table Based On The Result-set Of An SQL Statement.
- Schema In Sql
- Difference Between DELETE, DROP, TRUNCATE.
- Activity : Create Tables For Online Book Store
- Group By Clause
- Having Clause : For Filtration
- Where Clause
- GRANT: Grant Select Permission On <Table_name> To <User>
- Revoke

Note

- o Since we were getting connection issues on some of the laptops, UNext asked us to share the laptops and run the queries on a single laptop

<<

Delete Cancel Redeploy Download Refresh

Overview

Inputs

Outputs

Template

Your deployment is complete

Deployment name : Microsoft.SQLDatabase.newDatab... Start time : 8/31/2023, 10:28:52 AM
Subscription : [npunext-1680261348707](#) Correlation ID : d77af3b5-3d31-4ac7-bd1f-aa286...
Resource group : [RG_IDAShell_Anushka](#)

> Deployment details

∨ Next steps

[Go to resource](#)

Give feedback
 [Tell us about your experience with deployment](#)

Cost management
Get notified to stay within your budget and prevent unexpected charges on your bill.
[Set up cost alerts >](#)

Microsoft Defender for Cloud
Secure your apps and infrastructure
[Go to Microsoft Defender for Cloud >](#)

Free Microsoft tutorials

[Start learning today >](#)

Work with an expert

Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.
[Find an Azure expert >](#)

Search

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

Stream analytics (preview)

Add Azure Search

<< Login + New Query ↑ Open query Feedback Getting started

Anushka_IDA (Anushka0602)

Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

> Tables

> Views

> Stored Procedures

Query 1 ×

Run Cancel query Save query Export data as Show only Editor

1 select getdate()

Results Messages

Search to filter items...

2023-08-31T06:18:05.4100000

Query succeeded | 0s

[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

Stream analytics (preview)

Add Azure Search

Anushka_IDA (Anushka0602)

Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

Tables

dbo.emp_tbl

ven.emp_tbl

Views

Stored Procedures

Query 1 ×

Run

Cancel query

Save query

Export data as

Show only Editor

```
1 --create schema ven
2 --create table emp_tbl(emp_id int)
3 create table ven.emp_tbl(emp_id int)
4 
```

Results

Messages

Query succeeded: Affected rows: 0

Query succeeded | 0s

Search

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

Stream analytics (preview)

Add Azure Search

<< Login + New Query ↕ Open query Feedback Getting started

Anushka_IDA (Anushka0602)

Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

Tables

dbo.Author

dbo.Book

dbo.customer

dbo.Orders

dbo.Publisher

Views

Stored Procedures

Query 1 ×


Run Cancel query Save query Export data as Show only Editor




```
1 create table Author (Aid int primary key, A_name varchar(255),no_of_books int);
2 create table Publisher(Pid int primary key,p_name varchar(255),p_country varchar(255));
3 create table Book (bid int primary key,Author_id int foreign key references Author(Aid);
4 create table Orders (Oid int primary key,odate datetime,bid int foreign key references
5 create table customer(Cid int primary key,c_name varchar(255),oid int foreign key refer
6
```


Results Messages

Query succeeded: Affected rows: 0

Query succeeded | 0s

 Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

- >  Tables
- >  Views
- >  Stored Procedures


 Run ☐ Cancel query  Save query  Export data as  Show only Editor

```
1 Select TOP (1000) *
2 FROM [SalesLT].[Customer]
3 INNER JOIN [SalesLT].[CustomerAddress]
4 ON [SalesLT].[Customer].CustomerID = [SalesLT].[CustomerAddress].CustomerID;
```

Results Messages

 Search to filter items...

CustomerID	NameStyle	Title	FirstName
29485	False	Ms.	Catherine
29486	False	Ms.	Kim
29489	False	Ms.	Frances
29490	False	Ms.	Margaret

 Query succeeded | 5s

DAY-4

Azure SQL

- Create A Database And Elastic Pool.
- Joins:
 - Inner Join
 - Left Join
 - Right Join
 - Self Join
 - Outer Join
 - Cross Join
- Stored Procedure
- Views
- Functions
 - Scaler
 - Table Valued
- Complex Codes -> Stored Procedures
- Simple Code -> Views
- Temporary Table: # Before Table Name Creates A Temporary Table (i.e., No Physical Table Created). It Drops Automatically.
- Sub-queries
- UNION And UNION ALL
- Intersection
- Index
 - Clustered Index
 - Non-Clustered Index
 - Column Stored Index

Query 1 ×

▶ Run ☐ Cancel query ⬇ Save query ⬇ Export data as ▾  Show only Editor

```
1 SELECT MAX([SalesLT].[Product].ListPrice)
2 FROM [SalesLT].[Product]
3 WHERE [SalesLT].[Product].ListPrice < (SELECT MAX([SalesLT].[Product].ListPrice) FROM [SalesLT].[Product])
```

Results Messages

🔍 Search to filter items...

3399.9900

Query 1 × Query 2 × Query 3 × Query 4 ×

▶ Run ☐ Cancel query ⬇ Save query ⬇ Export data as ▾  Show only Editor

```
1 Select TOP (1000) *
2 FROM [SalesLT].[Customer]
3 RIGHT JOIN [SalesLT].[CustomerAddress]
4 ON [SalesLT].[Customer].CustomerID = [SalesLT].[CustomerAddress].CustomerID;
```

Results Messages

🔍 Search to filter items...

CustomerID	NameStyle	Title	FirstName	MiddleName
29485	False	Ms.	Catherine	R.
29486	False	Ms.	Kim	
29489	False	Ms.	Frances	B.

✓ Query succeeded | 1s

Query 1 × Query 2 ×

▶ Run ☐ Cancel query ⬇ Save query ⬇ Export data as ▾  Show only Editor

```
1 Select TOP (1000) *
2 FROM [SalesLT].[Customer]
3 LEFT JOIN [SalesLT].[CustomerAddress]
4 ON [SalesLT].[Customer].CustomerID = [SalesLT].[CustomerAddress].CustomerID;
```

Results Messages

🔍 Search to filter items...

CustomerID	NameStyle	Title	FirstName	Middle
1	False	Mr.	Orlando	N.
2	False	Mr.	Keith	
3	False	Ms.	Donna	F.

Query succeeded | 17s

Query 1 × Query 2 × Query 3 ×

▶ Run ☐ Cancel query ⬇ Save query ⬇ Export data as ▾  Show only Editor

```
1 Select TOP (1000) *  
2 FROM [SalesLT].[Customer]  
3 CROSS JOIN [SalesLT].[CustomerAddress]  
4
```

Results Messages

🔍 Search to filter items...

CustomerID	NameStyle	Title	FirstName	Middle
1	False	Mr.	Orlando	N.
1	False	Mr.	Orlando	N.
1	False	Mr.	Orlando	N.

✔ Query succeeded | 2s