

AZURE BIG DATA WITH AZURE DATA LAKE STORAGE AND AZURE DATA FACTORY

FIRST STEPS INTO BIG DATA

Navneet Sharma

Azure Data Engineer

About Speaker

Navneet Sharma

Azure Data Engineer

I have around 13 years of experience in the Microsoft technologies. I am a Microsoft certified Azure Data Engineer Associate. I am very passionate about Big Data and Microsoft Azure technologies.

<https://www.linkedin.com/in/nsharma7/>



1

2

3

4

5

AGENDA

INTRODUCTION

AZURE DATA STORAGE

AZURE DATA FACTORY

CERTIFICATION PATH

NEXT STEPS

Introduction

IMPORTANCE OF DATA



Data is the most critical asset



Storing data the right way is of utmost importance



Data needs to be moved from time to time



One-prem to cloud or between different cloud platforms



Data also needs transformations as it moves from one system to another





Data Storage

Data Storage Approach - Basics



Classify your data as structured, semi-structured, or unstructured



Determine how your data will be used



Determine whether your data requires transactions

Big Data Storage Technologies in Azure

File storage:

Azure Blob Storage (Hot, cool, and archive storage tiers)

Azure Data Lake Store (Hot, cool, and archive storage tiers)

NoSQL databases:

Azure Cosmos DB

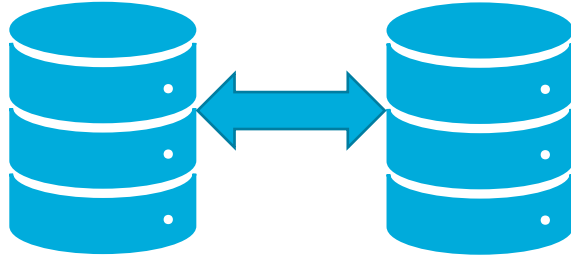
HBase on HDInsight

Analytical databases:

Azure Data Explorer

Azure Synapse Analytics – Dedicated SQL Pools

Storage Redundancy and Regional Pairs



Redundancy Options

Locally Redundant Storage (LRS)
Zone Redundant Storage (ZRS)
Geo Redundant Storage (GRS)
Geo-zone-redundant Storage (GZRS)
Read-accessible versions for GRS & GZRS



Region Pairs

Globally different region

Lifecycle Management

Hot

Cool

Archive

Blob
Deletion



Securing Data Storage

**Enable Soft
Deletes**

**Access via SAS
tokens**

**Setup Networking
to allow access
only from
required sources**

**Disable all Public
Blob access**

**Require secure
transfer**

**Lock your
resources**

Demo

Azure Data Lake – Gen 2

Azure Data Factory

Azure Data Factory – Code Free ETL

Ingest



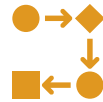
Multi-Cloud and on-prem hybrid copy data

90+ native connectors

Serverless and auto-scale

Use wizard for quick copy jobs

Control Flow



Design code-free data pipelines

Generate pipelines via SDK

Utilize workflow constructs: loops, branches, conditional execution, variables, parameters,...

Data Flow



Code-free data transformations that execute in Spark

Scale-out with Azure Integration Runtimes

Generate data flows via SDK

Designers for data engineers and data analysts

Schedule



Build and maintain operational schedules for your data pipelines

Wall clock, event-based, tumbling windows, chained

Monitor



View active executions and pipeline history

Detailed activity and data flow executions

Establish alerts and notifications

Azure Data Factory – Basic Flow

Connect and Collect Data

Transform Data

Data Pipelines (ETL) with Azure DevOps & GitHub

Publish data to various targets

Monitor & Alert

Key Technical Concepts



Pipelines



Activities



Datasets



Linked
Services



Data Flows



Integration
Runtimes



Types of Integration Runtime in Azure Data Factory

IR type	Public network	Private network
Azure	Data Flow Data movement Activity dispatch	Data Flow Data movement Activity dispatch

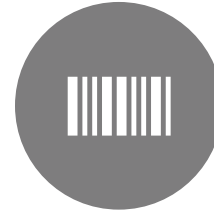
Azure Data Factory – Design Considerations



INITIAL SETUP
CONSIDERATIONS



PARALLELISM
MANAGEMENT



MANAGING
REUSABLE CODE



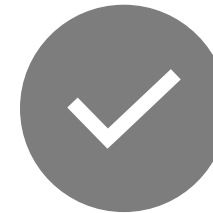
ENHANCING
SECURITY



MONITORING &
ALERTING



ENFORCING
STANDARDS



MISCELLANEOUS

Azure Data Factory – Security

Data Encryption in transit

- HTTPS
- TLS

Data Encryption at rest

- Microsoft managed keys
- Customer managed keys

Azure Key Vault

- For any connection strings
- Storing custom keys



Demo

Azure Data Factory

Certification Path

Certification Path

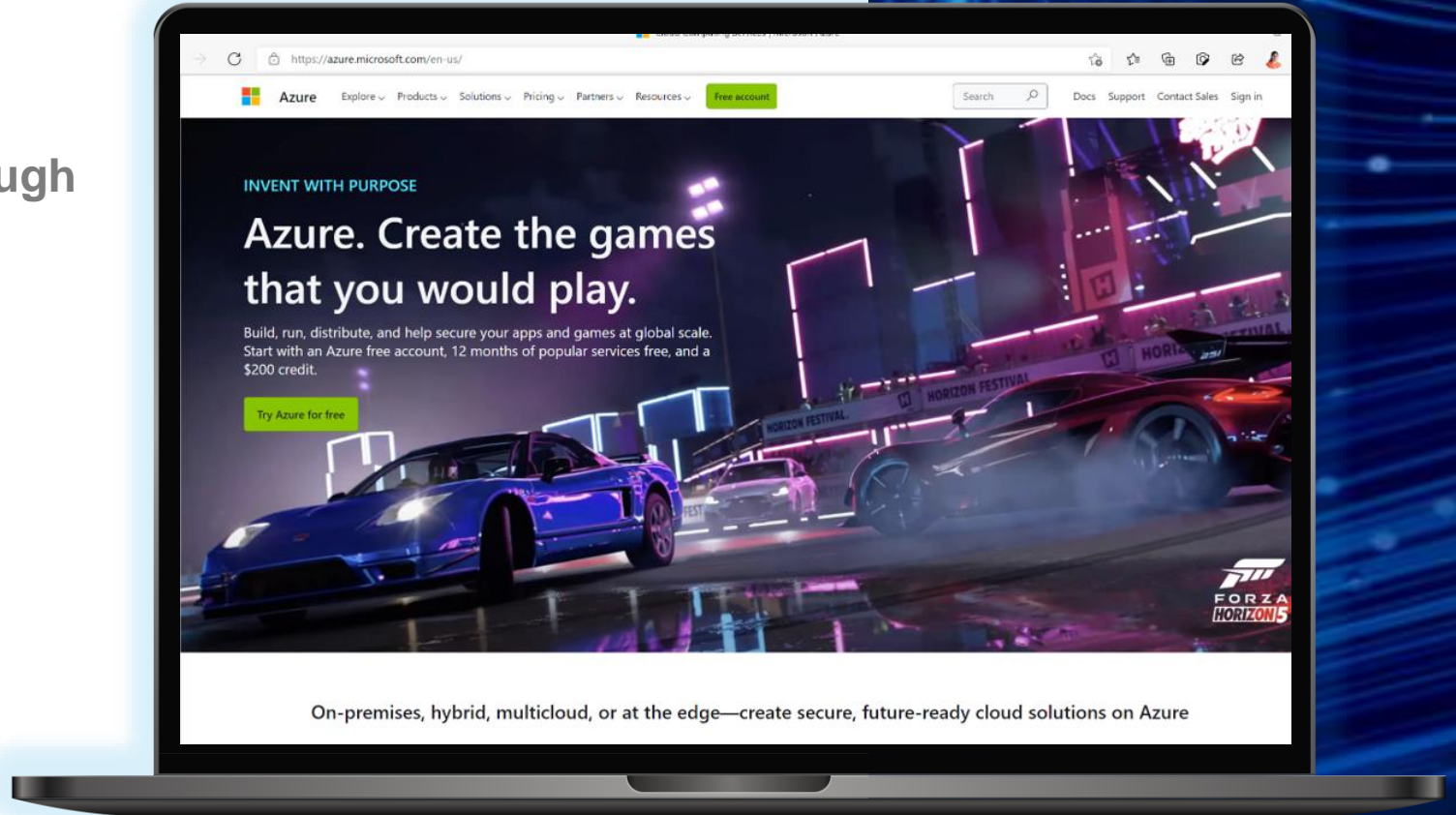
Optional certifications	Skills required for the certification	Exam Taken	Certification Earned
-------------------------	---------------------------------------	------------	----------------------



Get Azure Account Set Up

Sign up for free Azure Account through

<https://azure.com>



Key Takeaways



AZURE DATA
STORAGE



AZURE DATA
FACTORY





Next Steps



CREATE FREE
AZURE ACCOUNT



PREPARE FOR
DATA ENGINEER
CERTIFICATION



TAKE DP-203
AZURE EXAM



REACH OUT TO
SME

Reference Links

Topic	Link
Azure Blob Storage	https://docs.microsoft.com/en-us/azure/storage/blobs/storage-blobs-introduction
Azure Data Lake Store	https://docs.microsoft.com/en-us/azure/data-lake-store
Azure Data Explorer	https://docs.microsoft.com/en-us/azure/data-explorer/
Learning Paths - Azure Data Engineer Associate	https://docs.microsoft.com/en-us/learn/certifications/exams/dp-203
Azure Data Factory	https://docs.microsoft.com/en-us/azure/data-factory/introduction



Q&A

Thank you.