

1. Volume

The total amount of data generated, stored, and processed, ranging from gigabytes to petabytes and beyond.

Aspect	Definition		
Small-scale Data	Manageable datasets that can be processed on a single machine (e.g., local databases, small CSV files).		
Large-scale Data	Massive datasets requiring distributed computing (e.g., big data frameworks like Hadoop, Spark).		

4. Veracity

The accuracy, quality, and reliability of data, ensuring it is trustworthy for decision-making.

Type	Definition
Accurate Data	Reliable, precise, and error-free data that can be confidently used for analysis.
Biased Data	Data that may contain inaccuracies or distortions due to human or systemic biases.
Noisy Data	Data with irrelevant or misleading information that can obscure meaningful insights.

2. Variety

The different types and formats of data.

Туре	Definition
Structured Data	Data organized in fixed formats, such as tables (e.g., SQL databases, spreadsheets).
Unstructured Data	Data without a predefined structure, such as images, videos, or social media posts.
Semi-structured Data	Data with some structure, often stored in formats like JSON or XML files.

5. Value

The usefulness of data in generating insights, making informed decisions, and driving business impact.

Туре	Definition		
Business Insights	Actionable insights derived from data, helping organizations optimize operations and strategy.		
Predictive Models	Data-driven algorithms that forecast future trends or outcomes.		
Optimization	Using data to enhance efficiency improve processes, and maximize performance.		

3. Velocity

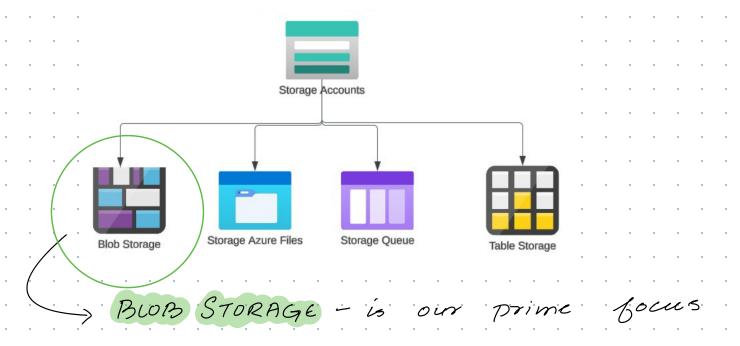
The speed at which data is generated, processed, and analyzed.

Type	Definition		
Batch Processing	Data processed in large sets or chunks over a specific time frame (e.g., ETL jobs).		
Real-time Data	Data processed immediately as it arrives (e.g., stock prices, live tracking).		
Streaming Data	Continuous, real-time data flow from sensors or IoT devices.		

Category	Data Type	Definition
Structured Data	Relational Databases	Data stored in tables with rows and columns, following a strict schema (e.g., SQL databases).
	Spreadsheets	Tabular data used for analysis and reporting (e.g., Excel, Google Sheets).
	ERP & CRM Systems	Business applications managing structured data for operations and customer relationships (e.g., SAP, Salesforce).
Semi-Structured Data	JSON & XML Files	Data with key-value pairs or tags , providing partial structure but no fixed schema (e.g., API responses, config files).
	NoSQL Databases	Schema-less databases storing data in key-value , document , column , or graph models (e.g., MongoDB, Cassandra).
	CSV Files	Text files storing tabular data without enforced schema.
	Parquet & ORC	Optimized columnar storage formats for big data, enabling efficient compression and fast querying (e.g., Apache Parquet, ORC).
Unstructured Data	Images & Videos	Media files lacking a predefined structure, requiring metadata for organization (e.g., PNG, MP4).
	Audio Files	Sound recordings used in voice assistants, music streaming, and call logs (e.g., MP3, WAV).
	Emails & Social Media	Text-heavy, unstructured data with mixed content like attachments, messages, and comments (e.g., Outlook, Twitter posts).







Storage Type	Description	Best Use Cases
Azure Blob Storage	A scalable object storage for unstructured data like text, images, videos, and backups. Supports Hot, Cool, and Archive tiers.	Media storage, backups, logs, big data processing.
Azure Data Lake Storage (ADLS)	A high-performance storage service optimized for big data analytics. Supports structured, semi-structured, and unstructured data with hierarchical namespaces.	Data lakes, machine learning, analytics workloads. Gen 2
Azure Files	Fully managed file shares in the cloud, accessible via SMB and NFS protocols .	File sharing, lift-and-shift applications, application storage.
Azure Queue Storage	A message queue service for asynchronous communication between applications, ensuring decoupled architecture.	Message queuing, event-driven architectures, background processing.
Azure Table Storage	A highly scalable NoSQL key- value store for structured, non- relational data, enabling fast lookups and massive data storage without complex relationships.	Logging events, telemetry data, sensor readings, metadata storage for web apps.

CREATE STORAGE ACCOUNT

Create a storage account

create a storage accou			۰	•	•	• •	•	•	•	• •	۰
storage accounts ರ			•	•	•		•	•	•		•
Project details			•		•			•	٠		
Select the subscription in which to create manage your storage account together w	the new storage account. Choose a new or existing reith other resources.	source group to organize and	•	•	•		•	•	•		•
Subscription *	CIC-Prod	~	•	•	•	• •	٠	•	•	• •	•
Resource group *	Tikku	~	•	۰	•	• •	٠	۰	۰	• •	۰
	Create new		•	•	•	• •	•	•	•		•
Instance details			•	۰	•		۰	۰	•		۰
Storage account name * ①	testingtiku1		•	۰	•	• •	٠	۰	٠		۰
Region * (i)	(US) West US	~	•	•	•		•	•	•		•
	Deploy to an Azure Extended Zone		•	۰	•		۰	٠	•		۰
Primary service ①	Select a primary service	~	•	•	•		•	•	•		•
Performance * ①	Standard: Recommended for most scenarios (general-purpose v2 account)	•	۰	•		•	•	•		•
	Premium: Recommended for scenarios that re	equire low latency.	•	•	•		•	•	•		
Redundancy * ①	Locally-redundant storage (LRS)	~	·>	7 0	ċa i		•	•	•		
neutrically ()							•	•	۰		
				91	1060	a /					
Previous Next Review	+ create		٠	•	•	• •	•	•	•		•
			۰	۰	•	• •	۰	۰	۰		۰
			٠		•		•	•	٠		۰
CREATE C	ONTAINER INSIDE	STORAG	ϵ	. 7	ACC.	00	~ T	:	•		•
				•			•	ر •	٠		
Container	- · · · · · · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · · · · · ·	. ,	100	ز ر د	l	- 1-	. 00	. 1	١.	
Container	s used to sto	72 13/055		U	r) > 7	ru	CTU	186	d)	
			. (0	g f	9	۰	•	"	۰
New container	×	• • • •	•	•	•	• •	•	•	•	• •	•
• •	•	• • • •	•	۰	٠	• •	۰	۰	•	• •	۰
Name *	•	· [We c	-0	ىۋە		つわ	oir	9	= '	• •	•
• •	•	· }· · · ·						•	•	• •	•
Anonymous access level	•	. (. exce	6	5	. /	رد ما ما	<u>e</u> .C	•	٠		۰
Private (no anonymous acces	· · · · · · · · · · · · · · · · · · ·			۰				۰	•		۰
	private because anonymous access is	. No+	P	re	50	. S.C.	اک		•		۰
disabled on this storage $^{\circ}$	account.		•	•	•		•	•	•		
· · · Advanced	•										
Advanced			•	•	•	• •	•	۰	•		
• •			•	•	•		•	•	•		•
			•		•		•	•	•		•
Once di	one we will	· · · · · · · · · · · · · · · · · · ·		•		• •		•	•		. م
Once di	one we will	uploac				a+		•	i	75) [']	de
Once do	one we will			•		at		•	i	75) [†]	de
Once do		uploace ew Edit		•		at			i	75)°	de
Once do	one we will ortained (vie					at			i	75)	de

ALLOW	ACCESS TO CONTAINER FICE	• •
		• •
<i>C</i> ,		• •
Storage	Account -> Configuration	7:
• • • • •		• •
	Allow Blob anonymous access ①	• •
	Disabled • Enabled	
		• •
		• •
Container	(-) File -) Change Access Le	vel
		• •
	Change access level	• •
	Change the access level of container 'data'.	• •
	Anonymous access level ①	• •
	Blob (anonymous read access for blobs only)	• •
	Private (no anonymous access)	• •
	Blob (anonymous read access for blobs only)	• •
	Container (anonymous read access for containers and blobs)	• •
		• •
	OK Cancel	• •
		• •
	. <u>.</u> <u></u>	• •
ACCESS	IN PYTHON;	
		• •
import pand	tas as nd	۰
pore pane	.as as pa	•
	os://testingtiku1.blob.core.windows.net/data/ActivityLog-01.csv	
df = pd.rea	ad_csv(urt)	•
<pre>print(df)</pre>		•
		• •
		• •
CREATE	GENZ (DATA LAKE)	• •
		• •
Same	Procedure (Hierarchical Namespo	ace)
° Hierarchical Name	,	
	pace, complemented by Data Lake Storage Gen2 endpoint, enables file and directory semantics, accelerates	
	orkloads, and enables access control lists (ACLs) Learn more ♂	
• Enable hierarchical r	namespace (i) Englie	• •
•		

. . .

. .

. .

۰

CREATE	CONTAINER	INSIDE	GENZ	STORAGE	ACCOUN
					• • • •
Gen 2	will	have m	ore op	tions	
• • • •			• • •	• • • •	• • •
Ubio	ad Data		• • •	• • • •	• • •
		0 0 0			
ONNEC	T AZUI	ZE BLOI	3 . 70.	POWER	B1.
					• • • •
• • • •			• • •		• • •
		get. Da-	t ca		• • •
		• • / • •	• • •		• • • •
			• • •		• • • •
		AZUXE	• • • •		• • • •
	• • • • •	• • •	• • • •		• • • •
					• • • •
• • • •					
	AZUre	Blob	570,090		
		(.)		Å (a	
E	nter Stor	age. Ac	count	Vane	
• • • •		• • / • •			
					• • • •
	Acce	, ie ke	:45 (:	security)	• • • •
		• • • •	0	security)	• • • •
key1 () R	lotate key				
Last rotated	l: 3/11/2025 (0 days a	igo)			
Key					
	•••••		• • • • • • • • • • • • • • • • • • • •	•••••	Show
			• • •		• • •
			• • •		• • •
7/10	WAYS	To	ACCESS	FILES	
	73				
			,		• • • •
	Allow Blob a	anonymous a	ccess (i)		• • • •
	O Disable	d 💿 Enak	oled		• • • •
	Disable	Lilar	,		• • • •
			,		
	Accour	nt Na	mc /	ACCESS	Key
			(σ