

### 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  Data organized in fixed formats, such as tables (e.g., SQL databases, spreadsheets).				
Structured Data					
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

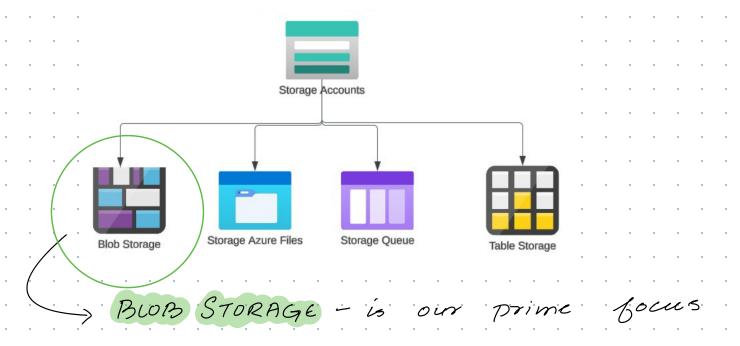
The speed at which data is generated, processed, and analyzed.

Туре	Definition  Data processed in large sets or chunks over a specific time frame (e.g., ETL jobs).			
Batch Processing				
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 <b>tables with rows</b> 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 <b>operations</b> and customer relationships (e.g., SAP, Salesforce).
Semi-Structured Data	JSON & XML Files	Data with <b>key-value pairs or tags</b> , providing partial structure but no fixed schema (e.g., API responses, config files).
	NoSQL Databases	Schema-less databases storing data in <b>key-value</b> , <b>document</b> , <b>column</b> , <b>or graph models</b> (e.g., MongoDB, Cassandra).
	CSV Files	Text files storing <b>tabular data</b> 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 <b>file shares</b> in the cloud, accessible via <b>SMB</b> and <b>NFS protocols</b> .	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 acc	count		•	٠	٠	٠	•	•	٠	•	٠	•	۰
storage accounts 🗗			٠	۰	٠	•	•	•	۰	•	۰	•	•
Project details			•	•	•	•	•	•	•	•	•	•	•
Select the subscription in which to cr manage your storage account togeth	reate the new storage account. Choose a new oner with other resources.	or existing resource group to organize and	•	•	•	•	•	•	•	•	•	•	•
Subscription *	CIC-Prod	V	•	0	٠	٠	۰	۰	•	•	۰	۰	۰
Resource group *	Tikku	V		۰	•	٠	•	٠	۰	۰	۰	٠	۰
	Create new		•	۰	•	٠	•	٠	۰	•	•	٠	۰
Instance details			۰	۰	۰	٠	۰	٠	۰	۰	۰	٠	۰
Storage account name * (i)	testingtiku1		] .	۰	•	٠	•	٠	۰	•	۰	٠	•
Region * ①	(US) West US	V		۰	•	٠	•	٠	٠	•	•	٠	۰
	Deploy to an Azure Extended Zone		۰			•	•			•			۰
Primary service ①	Select a primary service	~	] .										•
Performance * ①	Standard: Recommended for most	t scenarios (general-purpose v2 account)		۰	•	٠		٠	۰	۰	۰	٠	•
	Premium: Recommended for scen	arios that require low latency.	۰	٠		٠		٠	٠			٠	•
			1										
Redundancy * ①	Locally-redundant storage (LRS)	<u> </u>	) )		Ċa		•			٠	٠		
			-	G	106	a1	•	۰	۰	•	۰	۰	۰
Previous Next Re	view + create		•	•	•	•	•	•	•	•	•	•	۰
			۰	۰	•	۰	•	۰	•	•	•	۰	۰
CREATE						•	•	•	۰	•	۰	۰	۰
CREMIE	CONTAINER IN	ISIDE STORAG	$\epsilon$	. 7	ACC	00	<b>N</b>	7	;	•	•	۰	۰
			•	۰	•	٠	•	٠		۰	۰	٠	۰
Container	is used to	5th2 Blobs	. ,	(1)	ns.	tru	1° C =	h1	<b>プァ</b>	ار:	5		۰
• • • • •	is used to		. (		· · ·				•	•	,)		۰
						9+	-9			./			•
New containe	er ×		۰	٠			•	٠	٠			٠	•
Name *		Swe c		رود	. (	Cr	0	درو	9	5'	٠		
• • []			· _	٠	· /		,	۰	۰	۰	•	۰	۰
Private (no anonymous a		. Cucces	7	. /	20	ر (	_	•	•	•	٠	•	۰
		<b>N</b> 0+	·	· ~~	11	1 %	<u>e</u> c	1.	•	•	•	•	۰
	et to private because anonymous access is	7,404	P	0.0	50		•	٠	۰	•	•	٠	۰
disabled on this stor	age account.		•	٠	•	•	•	•	٠	•	٠	•	•
° ∨ Advanced			٠	۰	•	٠	•	٠	٠	•	•	٠	۰
•			۰	٠		•	•			•			۰
				۰	•	٠	•	٠	۰	۰		٠	•
Once e	done we upontained (	vill ublos	<i>ا</i> ل	٠	7	- بر ۱	<i>4</i> ~	<b>L</b> •		, ,	25	.·	0 .
		in in indigential	1		1	047	10				15	0	6
the CE	ontainer (	view Edit	)	۰	۰	۰	0	۰	۰	۰	۰	۰	۰
• • • • •			/	۰	۰	٠	•	۰	۰	•	۰	۰	۰
			۰	•	•	•	0	•	•	•	•	•	۰
maria	2.1 mB Blow	( 5:10)	۰	سد 'ر	•	•		٠		•	•	٠	۰

ALLOW	ACCESS TO CONTAINER FICE	• •
		• •
$\mathcal{C}_{i}$		• •
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	•
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	ACCOUNT
Gen 2 will to	ave m	ore op	tions	
Upload Date	i			
CONNECT AZUR	& Blog	3 . 70	DOWER	B1 · ·
		• • •		
	get Dan	La		
		• • •		
	Azure	• • • •	• • • •	
		• • •	• • • •	• • • •
Azure	Blob	Storage	• • • • •	
	/	• • •	• • • • •	
Enter Store		count	Name	
Enter Storo	nge.		• • • •	
· · · · · · · · · · · · · · · · · · ·	· Ke	:45 (	security)	
	55	<u></u>	· · · · · ·	
key1 ( ) Rotate key				
Last rotated: 3/11/2025 (0 days ag	10)			•
Key	,0)			,
•••••	•••••	•••••		Show
	• • • •	• • •		• • • •
lwo ways	To	ACCESS	FILES	5
Allow Blob ar	nonymous a	ccess (i)		
Disabled	d 💿 Enak	oled	• • • • •	
		)		
	• • • •	• • •		
- Accoun	+ Na	mc/	ACCESS	Keu
	• • •	/ .	, , , , , , , , , , , , , , , , , , , ,	7