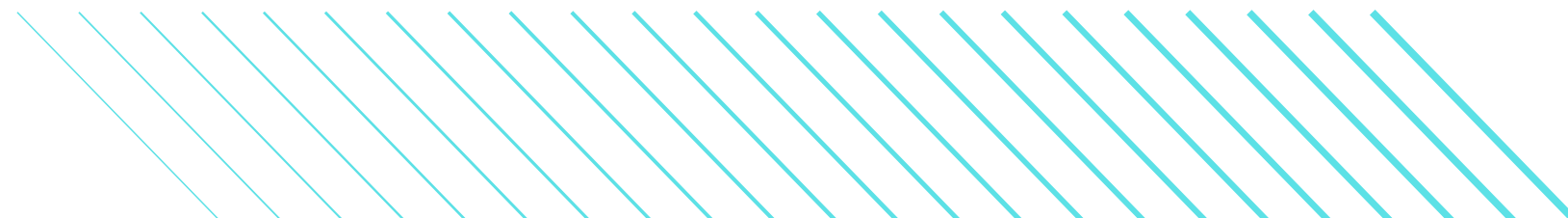
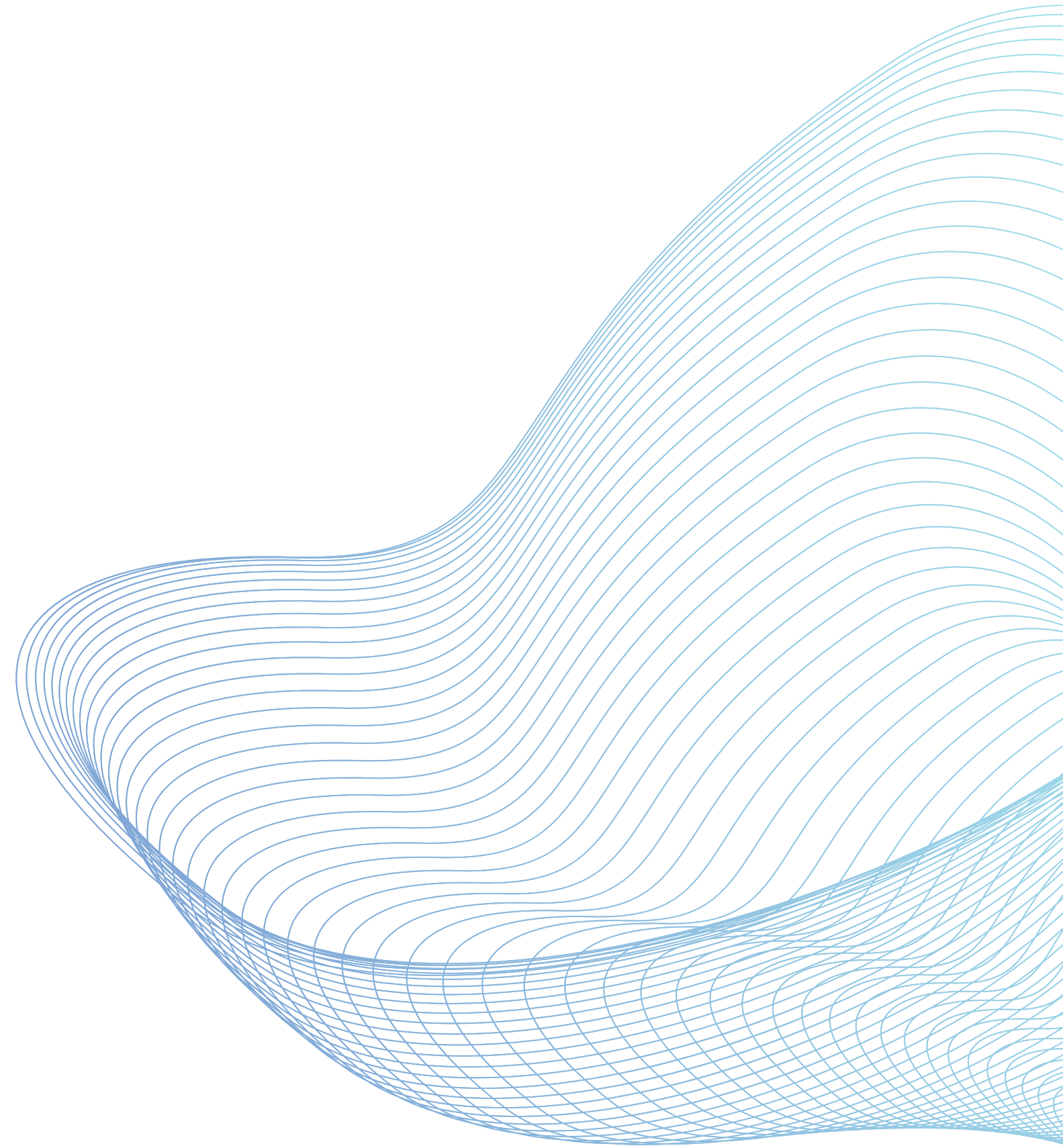


# **ARCHITECTURE**

# **BIG DATA**

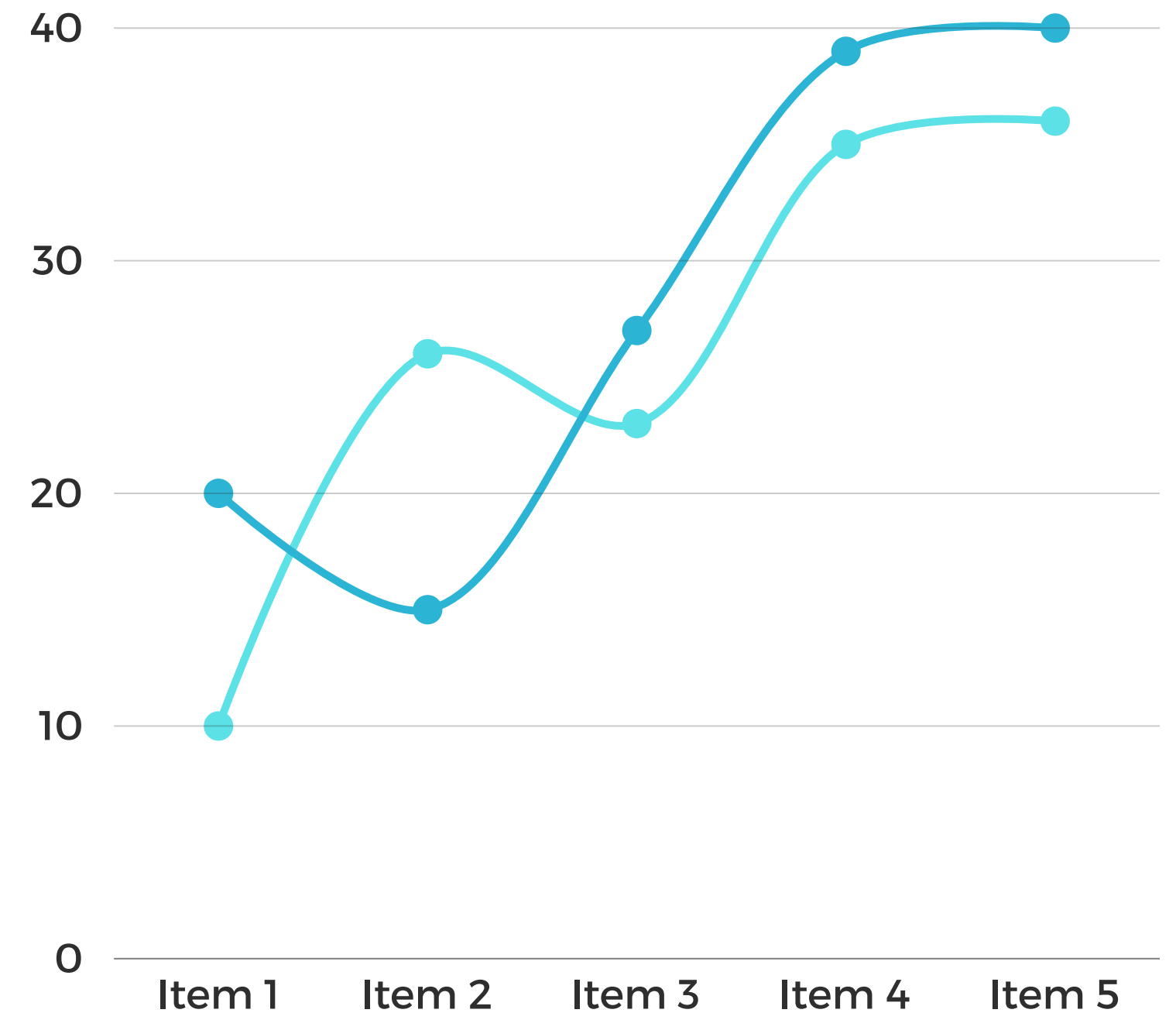
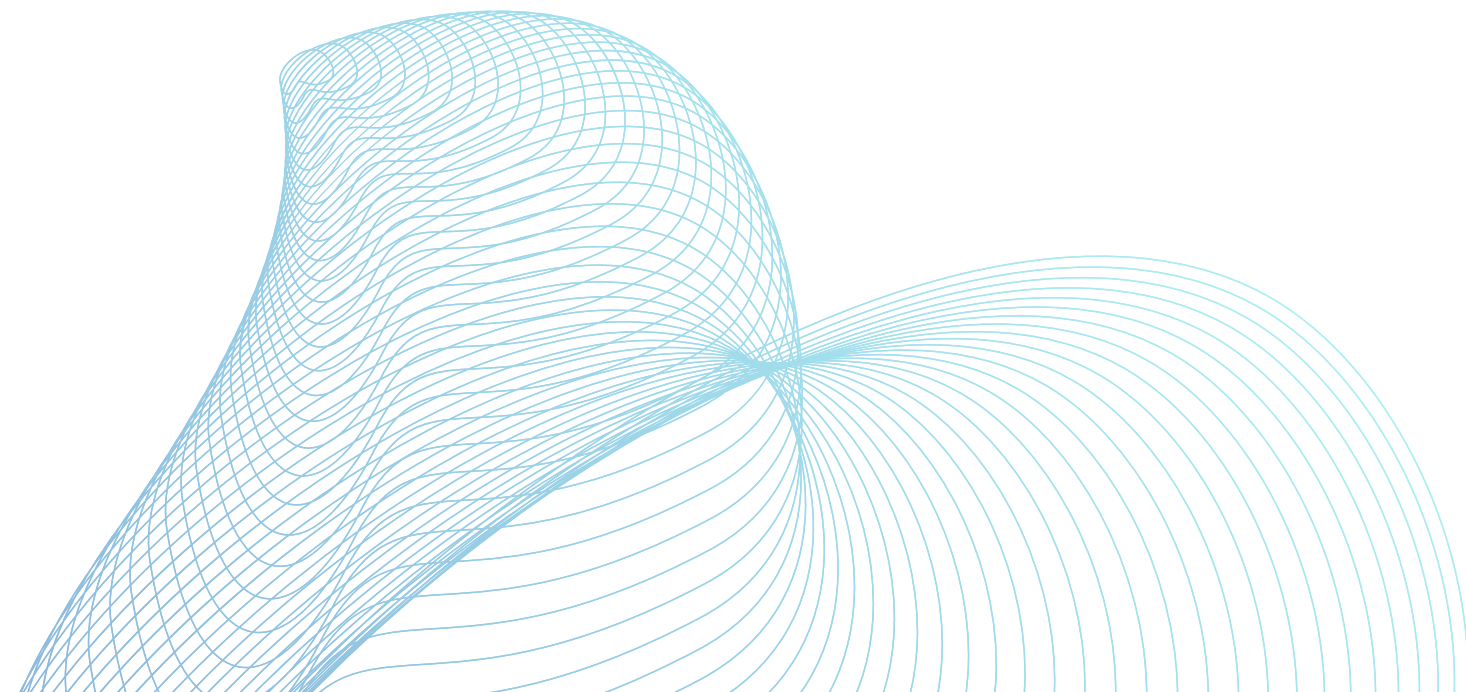
Par:

Brand TIKOUE TIKOUE  
Amine HAMZAOU



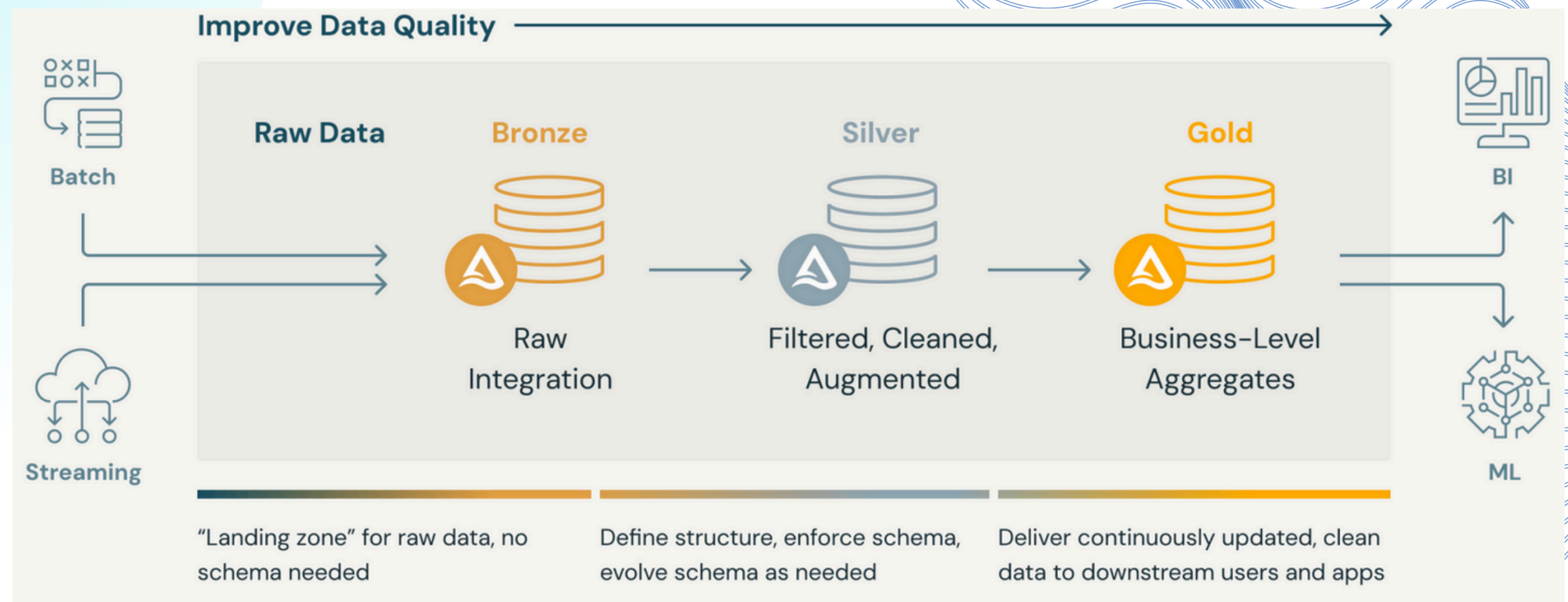
# CONTEXTE

Dans le but d'optimiser la prise de décision et de prédire les scores de crédit des clients bancaires à partir de divers attributs, tels que les données démographiques, financières et comportementales, nous disposons d'un jeu de données contenant des informations détaillées sur les utilisateurs.



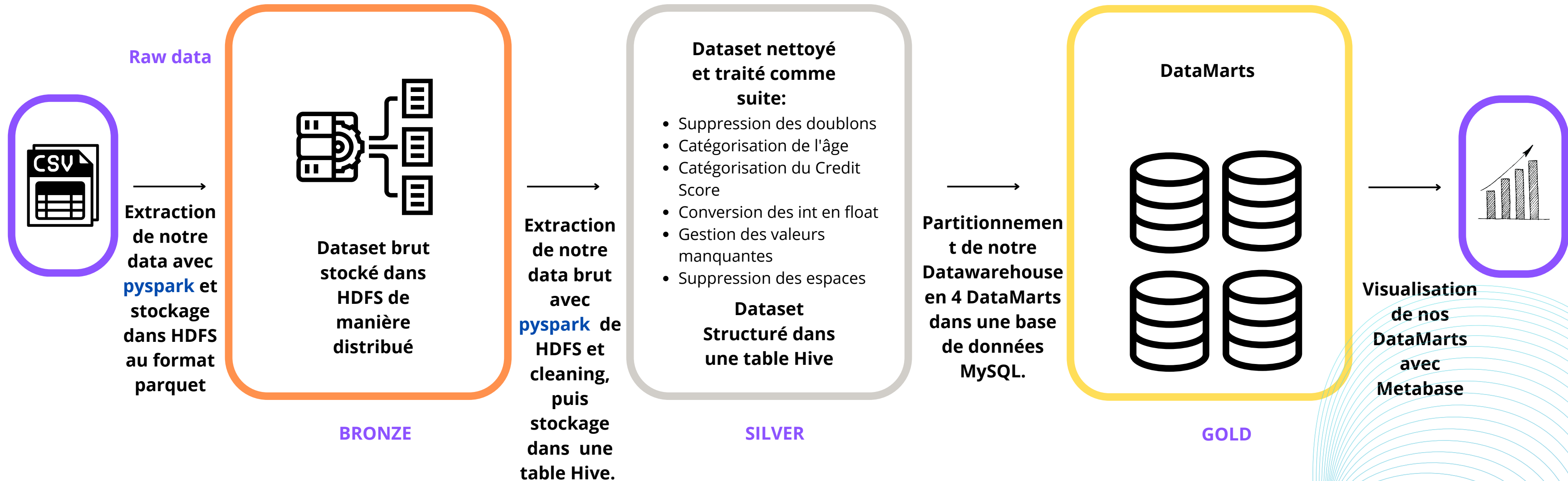
# OBJECTIF

L'objectif est de déployer une architecture Medallion pour rendre nos données accessibles, offrant ainsi aux équipes métiers et aux Data Analysts une visibilité claire et partitionnée de ces informations.





























































# WORKFLOW



# RUN DE NOTRE PIPELINE

<input type="checkbox"/>	 <a href="#">pipeline</a>		Running (11/11)		391.79%	11 seconds ago			
<input type="checkbox"/>	 <a href="#">metabase</a>	<a href="#">metabase/metabase:latest</a>	Running	<a href="#">3000:3000</a> 	233.35%	1 hour ago			
<input type="checkbox"/>	 <a href="#">mysql</a>	<a href="#">mysql:8.0</a>	Running	<a href="#">3306:3306</a> 	129.89%	1 hour ago			
<input type="checkbox"/>	 <a href="#">jupyter</a>	<a href="#">jupyter/pyspark-notebook:&lt;none&gt;</a>	Running	<a href="#">8888:8888</a> 	28.55%	1 hour ago			
<input type="checkbox"/>	 <a href="#">namenode</a>	<a href="#">bde2020/hadoop-namenode:2.0.0-hadoop3.2.1-java8</a>	Running	<a href="#">8020:8020</a> 	0%	18 seconds ago			
<input type="checkbox"/>	 <a href="#">datanode-2</a>	<a href="#">bde2020/hadoop-datanode:2.0.0-hadoop3.2.1-java8</a>	Running		0%	14 seconds ago			
<input type="checkbox"/>	 <a href="#">datanode-1</a>	<a href="#">bde2020/hadoop-datanode:2.0.0-hadoop3.2.1-java8</a>	Running	<a href="#">9870:9870</a> 	0%	14 seconds ago			
<input type="checkbox"/>	 <a href="#">spark-master</a>	<a href="#">bde2020/spark-master:3.1.1-hadoop3.2</a>	Running	<a href="#">7077:7077</a>  <a href="#">Show all ports (2)</a>	0%	13 seconds ago			
<input type="checkbox"/>	 <a href="#">hive-metastore</a>	<a href="#">bde2020/hive:2.3.2-postgresql-metastore</a>	Running	<a href="#">9083:9083</a> 	0%	13 seconds ago			
<input type="checkbox"/>	 <a href="#">hive-server</a>	<a href="#">bde2020/hive:2.3.2</a>	Running	<a href="#">10000:10000</a> 	0%	11 seconds ago			
<input type="checkbox"/>	 <a href="#">spark-worker-1</a>	<a href="#">bde2020/spark-worker:3.1.1-hadoop3.2</a>	Running		0%	11 seconds ago			
<input type="checkbox"/>	 <a href="#">spark-worker-2</a>	<a href="#">bde2020/spark-worker:3.1.1-hadoop3.2</a>	Running		0%	11 seconds ago			

# NAMENODE RUNNING

## Overview 'namenode:8020' (active)

Started:	Mon Oct 21 17:54:47 +0200 2024
Version:	2.8.0, r91f2b7a13d1e97be65db92ddabc627cc29ac0009
Compiled:	Fri Mar 17 05:12:00 +0100 2017 by jdu from branch-2.8.0
Cluster ID:	CID-d5b7ac2c-8958-4f07-ba7a-deeeb93c0011
Block Pool ID:	BP-393084982-172.18.0.2-1729523577545

## Summary

Security is off.

Safemode is off.

13 files and directories, 1 blocks = 14 total filesystem object(s).

Heap Memory used 63.39 MB of 206.5 MB Heap Memory. Max Heap Memory is 889 MB.

Non Heap Memory used 43.9 MB of 44.88 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	475.65 GB
DFS Used:	59.86 MB (0.01%)
Non DFS Used:	306.41 GB
DFS Remaining:	169.18 GB (35.57%)
Block Pool Used:	59.86 MB (0.01%)
DataNodes usages% (Min/Median/Max/stdDev):	0.01% / 0.01% / 0.01% / 0.00%
Live Nodes	2 (Decommissioned: 0)

# DATANODE RUNNING

## Datanode Information

✓ In service    ⚠ Down    🔧 Decommissioned    ⏻ Decommissioned & dead

### In operation

Show 

25

 entries

Search:

Node	Http Address	Last contact	Capacity	Blocks	Block pool used	Version
✓ 4400fabff82f:50010 (172.18.0.11:50010)	<a href="#">4400fabff82f:50075</a>	2s	237.83 GB <div></div>	1	29.93 MB (0.01%)	2.8.0
✓ 5c5640c1714c:50010 (172.18.0.10:50010)	<a href="#">5c5640c1714c:50075</a>	0s	237.83 GB <div></div>	1	29.93 MB (0.01%)	2.8.0

Showing 1 to 2 of 2 entries

Previous

1

Next

### Decommissioning

No nodes are decommissioning

# SPARK MASTER RUNNING



2.1.2-SNAPSHOT

Spark Master at spark://a84d5634265b:7077

URL: spark://a84d5634265b:7077  
REST URL: spark://a84d5634265b:6066 (cluster mode)  
Alive Workers: 2  
Cores in use: 8 Total, 8 Used  
Memory in use: 13.3 GB Total, 2.0 GB Used  
Applications: 1 Running, 2 Completed  
Drivers: 0 Running, 0 Completed  
Status: ALIVE

## Workers

Worker Id	Address	State	Cores	Memory
<a href="#">worker-20241021155441-172.18.0.12-37039</a>	172.18.0.12:37039	ALIVE	4 (4 Used)	6.6 GB (1024.0 MB Used)
<a href="#">worker-20241021155442-172.18.0.13-42037</a>	172.18.0.13:42037	ALIVE	4 (4 Used)	6.6 GB (1024.0 MB Used)

## Running Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
<a href="#">app-20241021161009-0000</a> (kill)	<a href="#">PySparkShell</a>	8	1024.0 MB	2024/10/21 16:10:09	root	RUNNING	10 min

## Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
<a href="#">app-20241021161912-0002</a>	PySparkShell	0	1024.0 MB	2024/10/21 16:19:12	root	FINISHED	14 s
<a href="#">app-20241021161643-0001</a>	PySparkShell	0	1024.0 MB	2024/10/21 16:16:43	root	FINISHED	1.1 min



**MERCI POUR  
VOTRE  
ATTENTION.**

