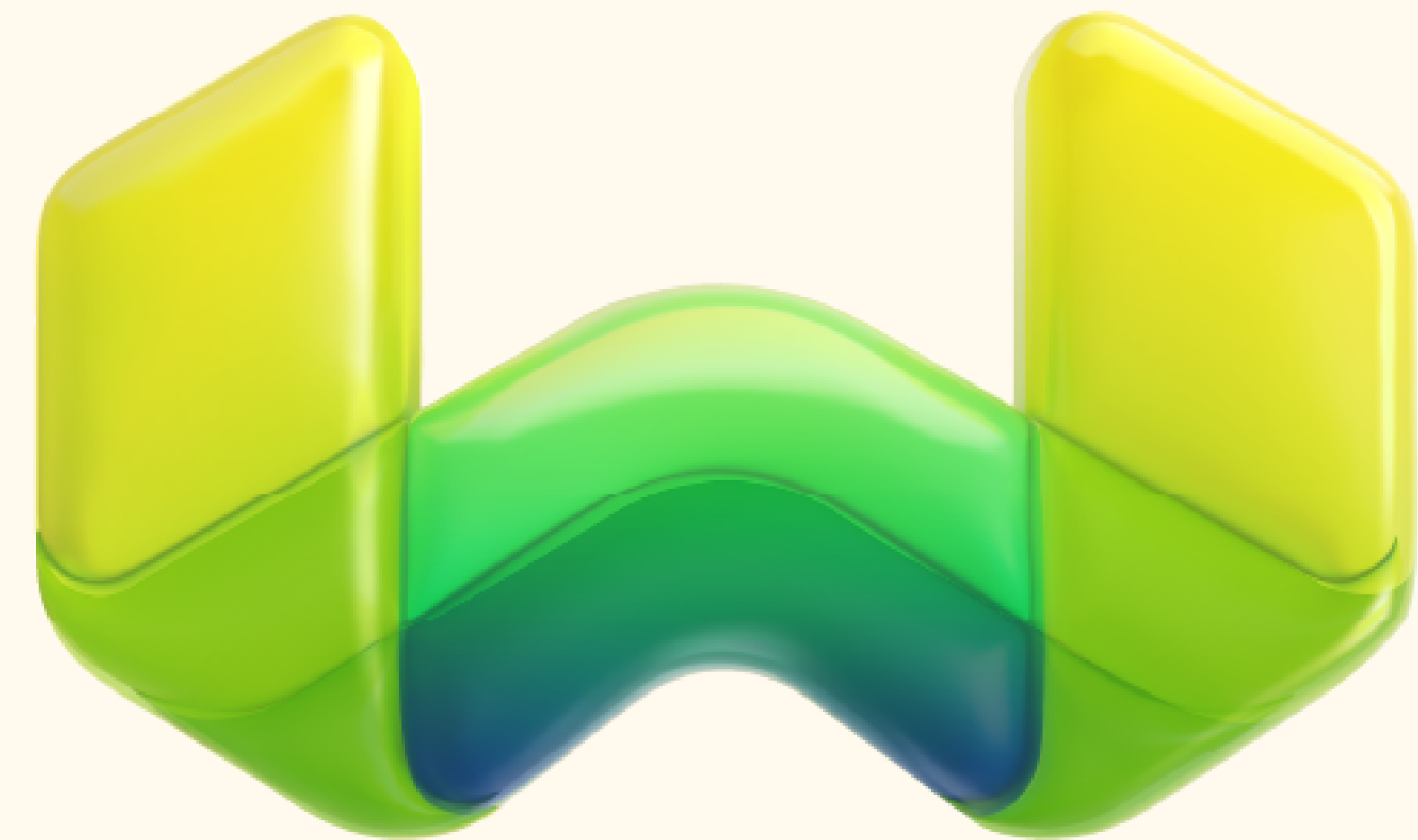


# **WEAVIATE EXPLAINED: CONNECTING SPARK PIPELINES TO WEAVIATE**



**CREATING SPARK ETL  
PIPELINE TO WEAVIATE  
IN JUPYTER NOTEBOOK**

# **CHALLENGE SOLVED : VECTOR STORAGE**

**1. BIGDATA IS INSIDE SPARK WAREHOUSE WHILE WEAVIATE NEEDS VECTORS TO ENABLE SEARCH**

**2. WEAVIATE SPARK CONNECTOR IS DEVELOPED FOR THIS PURPOSE. THE CONNECTOR MAKES THE ETL VERY SIMPLE.**

**3. JAVA AND SCALA **DEPENDENCIES TO BE RESOLVED****

**IN MY CASE THE CONNECTOR IS THROWING A METHOD NOT FOUND ERROR**

**1. READ THE DATA INTO SPARK SESSION AND DO THE TRANSFORMATION**

**2. CREATE WEAVIATE INSTANCE WITH LOCAL VECTORISER TO ACT AS STORAGE**

# LETS PRACTICE


- WEVIATE TEAM HAS WORKED ON SPHERE DATASET.
- THE CODE WILL SHOW HOW TO LOAD THE DATASET IF THERE IS ALREADY VECTOR STORE AVAILABLE.
- WE WILL LEARN A BIT ABOUT THE FACEBOOK MODEL USED FOR VECTORISING THE SPHERE DATASETS
- BE READY TO TAKE A NEW ROUTE ANY TIME

insightbuilder/  
**python\_de\_learners\_data**




Repo contains the code, data and supporting documents including presentations, playbooks and additional documents to support learning

 2  
Contributors

 0  
Issues

 34  
Stars


 16  
Forks



---

**python\_de\_learners\_data/weviateSpark\_integration.ipynb at main · insightbuilder/python\_de\_learners\_data**

Repo contains the code, data and supporting documents including presentations, playbooks and additional documents to support learning - python\_de\_learners\_data/weviateSpark\_integration.ipynb at mai...

 GitHub

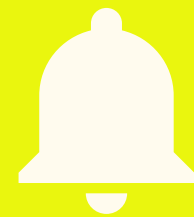
# THANKS FOR WATCHING



**LIKE**



**SHARE**



**SUBSCRIBE**