**מטלה 2 מחסני נתונים**

טבלאות משקיעים-

החלטנו להפריד את המידע על המשקיעים, הנתונים גולמיים גם מטבלת משקיעי עבר וגם ממשקיעים חדשים, לשתי טבלאות:

טבלת משקיעים, שמכילה פרטים על המשקיע, כולל קוד מדינה.

טבלת מדינה, שמכילה את הפרטים על המדינה- כולל קידומת, שם, קוד מדינה ומטבע.

בכך למעשה עשינו סטנדרטיזציה של שדה הטלפון, ששומר את מספר הטלפון ללא קידומת בטבלת משקיעים, ואת הקידומת שמתאימה לקוד המדינה בטבלת מדינה.

בנוסף, הוספנו טריגר שמוודא את פורמט המייל של המשקיע- ומאפשר הוספת כתובות מייל חוקיות נוספות, אך מעלה שגיאה כאשר כתובת מייל איננה חוקית.

לאחר מכן, יצרנו פייפ ליין שמהווה את שלבי הTLבתהליך ETL אפשרי. בהנתן טבלת לקוחות חדשים, אשר ניתנת בפורמט זהה לפורמט הנתון בטבלת הNEW INVESTORS הנתונה.

יצרנו פונקצייה שמנרמלת את המידע, ומעבירה אותו לפורמט של טבלת "משקיעים" שבנינו במחסן הנתונים שלנו- כולל הפרדה לתתי טבלאות שמכילות מידע על המדינה והמטבע וכו. פונקצייה זו מבצעת את שלב הטרנספורמציה.

בנוסף, הקמנו פרוצדורה שמורה שטוענת את הטבלה שעברה נירמול ע"י הפונקציה לתוך מסד הנתונים. ההגיון מאחורי ההפרדה היא שנוכל לבצע בדיקות של תקינות הקלט המנורמל לפני שנטען אותו לתוך המסד.

את בדיקת התקינות הנל מימשנו באמצעות שתי פונקציות נוספות. הראשונה בודקת שהטבלה לאחר הטרנספורמציה היא בפורמט הרצוי, על ידי יצירת עמודה שמציינת לכל רשומה האם היא ולידית או לא.

הפונקצייה השנייה בודקת האם כל הרשומות ולידיות, ומסירה את עמודת הולידציה אם כן.

כך למעשה, התהליך המלא מתנהל באופן הבא:

1. נכנסת טבלה של משקיעים חדשים.
2. היא עוברת סטנדרטיזציה.
3. נעשית בדיקה של הרשומות לאחר סטנדרטיזציה.
4. אם כל העמודות ולידיות, נסיר את עמודת הולידציה ונעבור לטעינת המידע.  
   אחרת, לא נוכל להסיר את העמודה, וטעינת המידע תיכשל- כך נוכל לבדוק את המידע ולתקן אותו.
5. נטען את המידע לדאטה בייס.

את מרכולי הנתונים בחרנו להציג גVIEWS. זה הגיוני כי למעשה מדובר על דלתות טמפורליות שבעלי התקפידים מעוניינים לבחון בזמנים משתנים.