**SQL Server**

**Business Intelligence Expert**

**פרויקט סיום מסלול**

**הנחיות**

## הפרויקט והיקפו

### מטרת הפרויקט

מטרתו של פרויקט ההתמחות במסלול ה-Business Intelligence Expert של מכללת נאיה, לספק לסטודנטים התנסות משמעותית, רחבת היקף ומעמיקה, בהקמת מערכת מידע שלמה המבוססת בעיקרה על מערך הטכנולוגיות של SQL Server.

בפרויקט יבוא לידי ביטוי כל הידע התיאורטי והמעשי שנרכש במהלך מסלול Business Intelligence Expert. לאחר ההתנסות בפרויקט התמחות זה, הסטודנטים רוכשים ניסיון משמעותי לקראת השתלבותם בעבודה.

בפרויקט זה הסטודנטים מתנסים באופן אישי, בכל שלבי הפיתוח של המערכת, החל משלב תחקור החלק התשתיתי של מסד הנתונים וכלה בבניית מערך ה-BI המבוסס עליו. הפרויקט מדמה פיתוח מערכת מידע בתרחיש ריאליסטי הכוללת את כל שכבות המידע הרלוונטיות לתפקוד עסקי מלא, וממחיש באופן מעשי את עבודת מפתחי ה-BI ברבדיה השונים, כפי שהיא מקובלת ונדרשת בשוק העבודה.

### היקף הפיתוח

הטכנולוגיות המשמשות לפיתוח המערכת הן:

* SQL Server Database Engine לתחקור מסד הנתונים התפעולי
* SQL Server Database Engine לפיתוח מחסן הנתונים הניתוחי
* SQL Server Integration Services לפיתוח תהליכי שינוע נתונים (ETL)
* SQL Server Analysis Services (SSAS) tabular model לפיתוח המודל האנליטי
* Power BI לפיתוח סביבת ניתוח, דוחות, Self Service ישירות מעל ה DWH או מעל מודל הטאבולר.
* SQL Server Agent לתזמון תהליכים

כמו כן בפרוייקט יש להשתמש בסביבות הפיתוח והניהול

* Visual Studio
* Management Studio
* Power BI

**שלבי הפיתוח העיקריים של המערכת הם:**

* מסד הנתונים התפעולי ומערכת OLTP
  + תחקור ולימוד המערכת התפעולית על פי הדרישות העסקיות.
* מערך BI
  + איפון ועיצוב של מחסן הנתונים (DWH) על פי הדרישות העסקיות.
  + פיתוח תהליכי ETL (MRR – STG – DWH) לטעינת מחסן הנתונים.
  + פיתוח מודל TABULAR כולל הגדרת מדדים מחושבים ו KPI.
  + הקמת מערך דו"חות מנהליים על פי דרישות הלקוח המבוססים על נתונים ב-DWH.

### שלבי ההתקדמות

כדי לאפשר דיונים מקצועיים רלוונטים ורב-שיח בנושאי פיתוח שונים, וגם כדי להמנע מפערי התקדמות בין המשתתפים בפרויקט, יחולק הפרויקט על ציר הזמן לכמה אבני-דרך. אבני-דרך אלו יהיו יעדי זמן לפיתוח שלבים שונים בפרויקט, ותבחן בהן מידת ההתקדמות של המשתתפים מתוך מטרה "ליישר קו" בפיתוח.

בכל שלב ושלב יערך סיכום של מצב הפיתוח ויבחנו מאפייני הפיתוח והפתרונות השונים שמימשו המשתתפים. בדרך זאת תתקבל הפריה הדדית של דרכי פיתוח שונות והתאמתן להמלצות המנטור המלווה אתכם.

שימו לב : את הפרויקט ניתן לעשות בזוגות או שלשות!

המתווה הבא מציין את אבני הדרך, אך הוא עשוי להשתנות במהלך הפרויקט עצמו במידת הצורך:

* **פגישות 1-2**
* Restore ל DB של המכללה ולקרוא לו בשם NAYA\_OLTP.
* לימוד מצב קיים של ה DB - NAYA \_OLTP.
* עיצוב קונספטואלי של מודל מסד נתונים DWH – בניית ERD מפורט (סכמת כוכב וכו')
* הצגת מסמך איפיון מפורט מקורות ליעדים (Source to Target)
  + בשלב זה יש להעזר בקובץ אקסל S2T, שימו לב ללשונית Demand כל יתר הלשוניות צריכות להיות מלאות.
* **פגישות 3-4**
* פיתוח מנגנוני ETL – SSIS בדגש על מודולריות ומתודת פיתוח DWH
* יש לעבוד על פי הדרישות הבאות:
  + ב SSIS שם הפרויקט יהיה CollegeProject\_[YourName]
  + ב SSIS יש לבנות 3 Packages לכל טבלה ב DWH בקונבנציה הבאה לדוג'
    - Mrr\_Customer\_N
    - Stg\_DimCustomer
    - Dwh\_DimCustomer
  + יש לייצר Connections דינמים עם פרמטרים.
* הקמת מחסן הנתונים (DWH) – שם מחסן הנתונים יהיה College\_DWH
* יש לקבוע שמות מפתחות זהים בין טבלה ראשית לטבלה המקושרת

לדוג' בטבלת DIM\_STUDENTS שם המפתח יהיהStudentID ואותו שם יופיע גם בטבלה המקושרת.

* כל טבלה לכל שכבה תהיה בקונבנציה הבאה לדוג'
  + - Mrr\_Customer
    - Stg\_DimCustomer
    - Dwh\_DimCustomer
* יש לבנות Job שיכיל את הפרויקט על פי הסדר, ולתזמן אותו ל 12:20 בלילה
* חובה טיפול ב NULL וטיוב נתונים (על שדות STRING בלבד ולא על שדות מפתח)
* יש להשתמש בשיטות העבודה שנלמדו בקורס
* בונוס – יש לייצר מנגנון טיפול בשגויים (עם טבלה ב DB ששם נרשמים השגויים במקרה ויש)
* **פגישה 5-6**
* פיתוח דוחות ופורטל הפרויקט –Power BI ו (SSAS) tabular model.

הדרישות לדוחות נמצאות מטה.

* **פגישה 7**
* הצגת הפרויקט

### ליווי מקצועי וחניכה

במשך כל תקופת ההתמחות וביצוע הפרויקט יזכו הסטודנטים לליווי מקצועי ולחניכה צמודה ממומחה BI בתחומי הפיתוח השונים. הליווי יהיה אישי וממוקד ובו כל משתתף יקבל משוב על עבודתו והנחיות מועילות לקראת השלבים הבאים בפרויקט. המומחה שילווה אתכם בתהליך יהיה זמין עבורכם למפגש פעם בשבוע למפגש פרונטלי במכללה שמטרתו לסייע לכם להתקדם ולהתייעץ על שלבים בפרויקט בהם אתם נתקלים בקשיים. במהלך השבוע המומחה יהיה גם זמין עבורכם במיילים לשאלות אם לא סובלות דיחוי עד ליום המפגש עימו.

### משימות הפרויקט

משימות הפרויקט לקוחות מצרכים ממשיים של לקוחות חברת נאיה טכנולוגיות.

### הנחיות כלליות

* מקור מסד הנתונים התפעולי הוא MSSQL + קבצי אקסל המנוהלים ידנית
* מסד הנתונים ה DWH יפותח בסביבת MSSQL
  + MSSQL – הכנת מימדים , טיוב נתונים ,טיפול בשגויים
* פיתוח מודל אנליטי יתבצע ב (SSAS) tabular model
* פיתוח דוחות יתבצע ב Power BI

## פרויקט - מכללה להכשרה מקצועית

**הדרישה העסקית**

הדרישה העסקית של הפרויקט היא להקים מערכת מידע אנליטי עבור מכללה להכשרה מקצועית טכנולוגית (כדוגמת נאיה קולג'). כאמור מערכת המידע מורכבת משתי שכבות עיקריות הקשורות ביניהן ועל כן הינה מערכת מידע שלמה:

* שכבת מסד הנתונים התשתיתי המכיל ומנהל את הנתונים האופרטיבים (OLTP).
* שכבת ה-BI המכילה נתונים מעובדים לטובת דוחות ואנליזות (DWH).

1. **מערכת OLTP**

כאמור, שכבה זאת הינה הבסיס האופרטיבי של המערכת ובה מתבצעת הפעילות האפליקטיבית של המשתמשים.

מסד הנתונים עליו נעבוד מכיל את המידע הבא :

* מודול תלמידים
  + רישום וניהול
  + ניהול תשלומים
* מודול מדריכים
  + רישום וניהול
  + ניהול תמחור
  + תחומי הדרכה
* מודול קורסים
  + פתיחת קורסים
  + תכנים (סילבוסים) ומועדים
  + לו"ז
  + ניהול מחזורי קורסים
  + שיבוץ מדריכים
  + שיבוץ תלמידים
  + הערכה וציונים
* ממד זמן (טבלה) משותפת לכל המודולים.(ראו קובץ Excel S2T בלשונית Demand)
* טבלאות עזר (Lookup tables) כנדרש, למשל:
  + רשימת סטטוסים רלוונטית
  + רשימות נושאי קורסים
  + ועוד...

בכל אחד מהמודולים נדרשת האפשרות לאחזר את הנתונים ולהציגם בחיתוכים שונים. לדוגמא:

1. הכנסות מתשלומים של סטודנטים.
2. תלמידים בקורס
3. ציוני תלמידים
4. לו"ז קורס
5. ועוד...
6. **מערכת BI**

מערכת זאת תכלול מנגנוני ETL ותשתית מחסן נתונים לבניית מודל אנליטי ולהפקת דו"חות.

דרישות ה-BI העסקיות הן כדלקמן:

1. המערכת תאפשר יצירה של דוחות דינמיים למשתמשי קצה בכלי BI כגון ,Excel, Power BI
2. המערכת תכיל נתונים כמותיים (סיכומים) ונתוני חיתוך עבור אותם סיכומים.
   * נתונים כמותיים על הרשמות
   * נתונים כמותיים על תשלומים
   * נתונים כמותיים על שיבוצים - כמה תלמידים בכל קורס
   * נתונים כמותיים עבור ציונים
   * חיתוכים ע"פ ישות (כל אחד מן המאפיינים של אותה ישות).
     + סטודנט
     + מדריך
     + קורס
   * חיתוכים ע"פ זמן והיררכיות זמן.

**דרישות נוספות**

* דוח שיראה פר קורס את התלמידים הכי נעדרים מאותו קורס ( מעל 10 אחוז העדרויות – שיהיה ניתן לשנות את האחוז הזה)
* באיזה יום בשבוע יש הכי הרבה העדרויות?
* באיזה קורס יש הכי הרבה העדרויות?
* ממוצע תשלומים פר קורס.
* דוח חייבים – סטודנטים שחייבים יותר מ X% מעלות הקורס (X = חותך לבחירה)
* התפלגות לפי צורות תשלום ברמת המכללה וברמת קורס.
* כמה מוקפאים פר קורס (יש להראות את מספר הקורס והקטגוריה)?
* באיזה קטגורית קורס הכי התעניינו ובאיזה קורס הכי פחות התעניינו?

3. **אופן הגשת הפרויקט:**

יש להגיש תיקיה שבתוכה יהיו:

* תיקיה אחת בשם SSIS בתוכה הפתרון.
* תיקיה אחת בשם Power BI בתוכה הפתרון.
* קובץ Bak של ה DB של ה DWH.
* קובץ S2T מלא.
* קובץ סקריפט SQL ובו את ההגדרה של כל האובייקטים שהוגדרו במהלך הקמת ה DWH.
* תיקיה אחת בשם (SSAS) tabular modelבתוכה הפתרון (בונוס).

**בהצלחה!**