

קורס מפתחי BI

מסמך מסכם לפרויקט

על הפרויקט

הפרויקט מדמה סביבה ארגונית של חברת סלולר גדולה בארץ. חברת הסלולר רוצה להקים פרויקט BI גדול על מנת לסייע למנהלי החברה בהחלטות הקריטיות והחשובות עמן הם נדרשים להתמודד מידי יום ביומו וביניהם:

* תמחור נכון של שירותים שונים
* העברת משאבים לשירות מסוים על חשבון אחר
* מציאת לקוחות הנוטים לנטוש את החברה ולתת מענה לפני הנטישה
* הקצאת משאבים נכונה כדי לחסוך בהוצאות

# שלב התכנון ויישום סביבת העבודה

* אפיון הסביבה העסקית
* הבנת הדרישות העסקיות ותהליכי העבודה בארגון. הבנת סביבת העבודה, הבנת תהליכים ארגונים,
* הבנת התחום בו אנו נדרשים לבצע פרויקט .BI
* יצירת קדימויות בשלבי הפרויקט
* אפיון משתמשי הקצה של המערכת
* ניתוח מערכות המידע של הארגון , מקורות ויעדים
* פיתוח הפרויקט בכלי מיקרוסופט
* הקמת מערכת מידע תפעולית לארגון
* הקמת מחסן נתונים עבור הארגון בטופולוגית סכמת כוכב

# BI Microsoft Solution

# שלב פיתוח פתרון ה BI לארגון בעזרת כלי מיקרוסופט -

פיתוח תוכניות להעברת נתונים מבסיס הנתונים התפעולי למחסן הנתונים על ידי שימוש באחת הטכניקות שנלמדו בקורס:

* כלי ה ETL של מייקרוסופט - SSIS
* פיתוח בשפת T-SQL
* פיתוח קובייה או מודל בעזרת כלי הSSAS של מייקרוסופט שיועלה אחר כך לסביבת model tabular server sql שתיתן מענה מהיר ואיכותי לשאלות העסקיות של הארגון במינימום מאמץ על ידי משתמשי הקצה
* דוחות מנהלים בעזרת מחולל דוחות (  services reporting ) ו - BI Power
* הקמת פורטל מקומי לאתר הדוחות .

# סביבת העבודה

על התלמיד לבנות פרויקט BI על פי הכללים שהוגדרו לעיל כדי לענות על שאלות עסקיות בחברת סלולר גדולה. חברת הסלולר (The Voice) מספקת שירותי מידע וקול ללקוחותיה בארץ ובעולם ולחברה מערכות תפעוליות רבות המשמשות את החברה במתן הפתרונות הרבים אותם מצפים לקוחותיה לקבל לפי רמת השירות הגבוהה מאוד המקובלת בשוק. המערכת התפעולית מורכבת מטבלת רשומות שיחה ענקית (CDR) ממנה ניתן לגזור את כל המידע הרצוי בתוספת של טבלאות מידע עסקי כגון:

* לקוחות
* שרותים
* אזורים
* מקורות

מקורות המידע של הפרויקט הינם בסיס הנתונים התפעולי של החברה + קבציי אקסל ו CSV  שונים.

החברה מבקשת ליצור פרויקט BI גדול מאוד אשר ישמש מודל שיעזור למשתמשים בקבוצות שונות וישמש כמערכת DSS (Data Support System) כדי לעזור להם בהחלטות הנדרשות מהם:

המודל יספק מענה לשאלות עסקיות הבאות:

* ניתוח של סה"כ תנועה של שימושי קול לפי מדינה
* ניתוח שימושים שונים לפי סוגי שיחות – call types
* ניתוח דקות לפי מספר הטלפון
* ניתוח שימוש duration לפי חבילה (לפני ואחרי הנחה)
* ניתוח של כמות שיחות וכמות לקוחות ביום ושעה
* ניתוח לקוחות רדומים – שאינם משתמשים בשרות
* ניתוח תנועה של דקות יוצאות ברמה יומית/ חודשית/ שנתית

# אופן הגשת הפרויקט

יש לנהל את הפרויקט בסביבת **GitHub**, כל הפרויקט יועלה ויימסר דרך חיבור ל – repo של Git

יש להקים  repo  אחד , לשתף אותו בין חברי הקבוצה,  ולהגיש קישור אליו בלבד.

ה- repo  יכיל את כל הקבצים הבאים:

1. מסמך מלא בפורמט וורד לתיאור כל תהליך העבודה התוצרים. יש לפרט במסמך:

* שמות מגישי הפרויקט
* דיאגרמה של המערכת התפעולית
* הנחות שנלקחו בפרויקט בראשי פרקים
* סכמת הכוכב של מחסן הנתונים
* שלבים בפיתוח הפרויקט
* צילומי מסך מתהליך בניית הפרויקט, חלקים מה SSIS , tabular , PBI ועוד.
* בעיות שונות שנתקלתם בהן בפרויקט ודרכים להתמודדות

1. קבצי SQL:

* מחסן הנתונים - סקריפט מלא ליצירת כלל האובייקטים – DB , טבלאות וכו'
* ג'ובים - סקריפט מלא ליצירת כלל הג'ובים הנדרשים לתהליכי ה ETL
* לולאה – סקריפט מלא ליצירת רשומות חדשות בטבלת USAGE\_MAIN – על הרשומות להיות תואמות

לשאר הרשומות בטבלאות האחרות – נדרש לייצר Data עבור שנה לפחות תחת שדות ה TIME וה-DATE

1. One Solution
2. פרויקט SSIS + קבצי CSV + Folders (עבור תהליך הטעינה הדינאמית )CSV
3. פרויקט SSAS
4. פרויקט SSRS
5. קבציי Power Bi
6. הנגשת הדוחות -  SSRS + PBI  -  צילומי מסך במסמך המקדים הכוללים כתובת URL
7. סקריפטים של Dax  עבור Queries Dax כפי שיפורטו בהמשך .

**הערה חשובה:** את הסעיפים הנ"ל יש לשלוח באמצעות קובץ ZIP של כל הפרויקט , כך ניתן לבדוק גם את הפיתוח וגם את התוצאות.

# מידול החברה הסלולרית

## מערכת המתגים הארגונית:

|  | מקור – ישויות במערכת התפעולית | תיאור | דוגמא |
| --- | --- | --- | --- |
| .1 | –OPFILEOPP קובץ מפעילים על פי קידומת | מתאר את המפעילים הקיימים בארץ על פי קידומת | 050 - פלאפון |
| .2 | קובץ –XXCOUNTRYPRE  קידומות מדינה | מתאר את המדינות  בעולם עלפי קידומת. | ישראל - 972 |
| .3 | CUSTOMER\_INVOICE | טבלה המתארת את חשבוניות הלקוח | חשבונית עבור לקוח 972520002375+ , עבור חודש דצמבר 2013 , ניתנה בינואר 2014 , חשבון לחיוב 400 שקלים |
| .4 | Package\_Catalog | טבלת קטלוג חבילות | חבילה מספר 1 – חבילת משפחה , חבילה 2 – מסלול לחיילים לקוח שמספרו 972525154837+ שייך לחבילת משפחה , סטטוס פעיל , מקבל הנחה של 35% בשיחות בתוך המשפחה |
| .5 | customer\_lines | טבלת קווי לקוח |  |
| .6 | countries | קודי מדינה בעולם | מדינה: אנגליה , יבשת: אירופה |
| .7 | call\_type | סוגי שיחה/שירות - | Call - Cellular Call ,TEXT - Text Messaging |
| .8 | Customer | לקוחות האירגון | לקוח מספר 1 , מספ טלפון  972541185146+ , אלברט אינשטיין , תאריך הצטרפות לחברה 5.2.2010 |
| .9 | USAGE\_MAIN | מכילה את כל התנועה האירגונית |  |

# הנחות מקובלות בפרויקט

הנחות מקובלות אותם נניח בפרויקט על מנת לפשט אותו:

* קידומת בת שלוש ספרות של מספר הטלפון מהוה אינדיקציה חד ערכית לשם החברה (לדוגמא 005 = פלאפון)  , התעלמו משיוך מספרים.
* קידומת מדינה היא עד כדי שלוש ספרות (לדוגמא =1 אמריקה , 972 = ישראל (
* מספר הטלפון של אדם מסוים מהוה יחוס חד ערכי למספרו כלקוח, כלומר לקוח מזוהה ע"י מס' טלפון אחד. במידה ובטבלאות יש יותר מאחד , יש לקחת את המעודכן ביותר בלבד.

# תהליך הETL

תהליך ה- ETL מעביר את הנתונים מהמערכת התפעולית למחסן הנתונים במנגנון מתוזמן. פרויקט אחד בלבד לנושא ה.ETL

# דרישות תהליך ה - ETL

1. יש לתכנן מפת גזירה כוללת של כל השלבים MRR,STG,DWH
   1. MRR – העתקה מלאה של הנתונים.
   2. STG – ניקוי והמרות של הנתונים, חישוב עמודות.
   3. חלוקה לפי DIM\FACT
2. בתוך DB DW , יהיו 3 סכימות של MRR, STG , DW. יש ליצור עוד סכימות של לחלוקה בין טבלאות integration  וטבלאות DIM , FACT
3. יש לבצע טעינה דינאמית של קבצי ה CSV
4. יש ליישם את התהליך באמצעות SSIS על פי כל כללי המתודולוגיה שלמדנו.
5. ניתן להניח הנחות שאינן סותרות את הלוגיקה העסקית המתוארת.
6. יש להריץ את התהליך באופן יומי.  ע"י   .SCHEDULER
7. יש לבצע תהליך Cleansing Data מלא לרבות המרת ערכי null
8. אחרי כל ,lookup במקרה של לא נמצא ערך יש להשלים עם 1- או unknown
9. יש לשמור על טבלת תיעוד טעינות בשם lineage ובה , מועד הריצה האחרון , שם הטבלה , שם המחשב  המריץ, שם משתמש,  שם ה package שהופעל.
10. יש לשלוף DELTA בלבד מטבלאות שהם temporal table

דרישות לבנית מחסן הנתונים מפורטות במסמך  EXCEL המצורף Target" To Source The Voice "

# טעינה דינאמית CSV

ישנם 2 סוגי קבצי ,CSV הראשון OPFILEOPP והשני .call\_type

נדרש לייצר תהליך ETL דינאמי עבור כל סוג קובץ ולטעון לטבלה מתאימה (בשלב המתאים(

על תהליך הטעינה להיות דינאמי -  יש לבחון את הרשומות שמתקבלות בכל קובץ ולייצר תהליך גנרי עבור כל סוג קובץ   call\_type /  OPFILEOPP    . בסיום טעינה יש להעתיק את הקבצים לספריית Process. ולשנות את שם הקובץ – לשם מקורי בתוספת חתימה של תאריך ושעה.

הנחות מקובלות:

1. כל קובץ נטען לתיקייה ספציפית על השרת בו מותקן server sql - יש לבצע משיכה יומית משם.
2. שמות הקבצים שונים מפעם לפעם.
3. קובץ OPFILEOPP מכיל את ה opfileopp כמחרוזת.
4. קובץ call\_type מכיל את ה-  ’call\_type‘ כמחרוזת.

 בתחילת התהליך יש לבחון את הצורך בהרצת סקריפט להכנת כל סביבת העבודה מוכנה. כולל יצירת ספריות, והעתקת קבצים.

ייתכן מידע כפול ויש לסנן רשומות אילו

# DW מעל Tabular Model פיתוח

מודל "הטבולר" הוא למעשה החלק המאפשר גישה מהירה למשתמשי הקצה של מערכת ה- BI על מנת שיקבלו תשובות לכל השאלות העסקיות של הארגון במהירות רבה תוך הצגה גרפית יפה על מנת לקבל החלטות נכונות לעתיד החברה. בניית המודל, במסמך Target" To Source "TheVoice ישנה הגדרה של:

## usage dimension cube

נדרש לממש את הסעיפים שלהלן:

1. Rename שמות עמודות וטבלאות
2. הגדרת טבלת Table Date
3. יצירת יחסים בין הממדים וטבלאות ה Fact
4. וודא Type Data ו – Format מתאים לעמודות – יש לתקן במידת הצורך.
5. בצע Column By Sort בעמודות הנדרשות
6. "החבא" עמודות במידת הנדרש
7. למודל יוגדרו Measures שיתבססו על טבלת ה :fact
   1. Count
   2. Count distinct on customer
   3. Call Duration
   4. Billable Call Duration
   5. Amount
   6. Billable Amount
   7. Call Duration Average
   8. :Calculated Columns יוגדרו למודל   .
   9. Billable Amount – Amount :Discount
   10. Discount / Amount :Discount Percentage
8. יוגדרו 3 KPI וערכי מטרה דינמיים
9. יש ליצור 3 פרספקטיבות ובהן כל המימדים והמדדים הרלוונטיים
10. בניית המימדים הבאים: ניתן לוותר על היררכיה ולבנות כל מימד בטבלה אחת.
11. מימד מדינות  - היררכיה ו- attributes:
    1. region->area->country
    2. Areas
    3. Countries
    4. Regions
12. מימד זמן - היררכיה ו- attributes:
    1. Year -> month -> Date
    2. Months
    3. Days In Week
13. מימד שעות - היררכיה ו- attributes:
    1. Hours -> Minutes
    2. Hours
    3. Minutes
14. מימד לקוחות  - היררכיה ו- attributes:
    1. Packages -> Customers
    2. Country -> Customers
    3. Operator -> Customers
    4. Customers
15. מימד סוג שיחה  - היררכיה ו- attributes:
    1. Call Type Category -> Call Types
    2. Call Types
16. מימד חבילה  - היררכיה ו- attributes:
    1. Package Status -> Packages
    2. Packages
17. מימד מפעיל  - היררכיה ו- attributes:
    1. Operators
18. מימד מקור שיחות - היררכיה ו- attributes:
    1. Call Origin Type

יש להקפיד להביא את התיאור (Description) ולא להציג למשתמש קצה קוד או מספר.

1. בנית הקובייה
2. במסמך " The Voice Source To Target " ישנה הגדרה של cube dimension usage
3. לקובייה יהיה measure group אחד, שיתבסס על טבלת ה fact. ה measures שיש להביא:
   1. Count
   2. Count distinct on customer
   3. Call Duration
   4. Billable Call Duration
   5. Amount
   6. Billable Amount

# דוחות

יש ליצור לפחות שישה דוחות שונים לבחירה מתוך בסיס השאלות העסקיות הבאות:

1. ניתוח של סה"כ תנועה של שימושי קול  לפי מדינה
2. ניתוח שימושים שונים לפי סוגי שיחות types( )call
3. ניתוח דקות לפי מספר הטלפון – SSRS
4. ניתוח שימוש duration לפני ואחרי ההנחה
5. ניתוח של כמות שיחות וכמות לקוחות ביום ושעה
6. ניתוח לקוחות רדומים – שאינם משתמשים בשרות
7. ניתוח תנועה של דקות יוצאות ברמה יומית/ חודשית/ שנתית

את כל הדוחות יש לפרוס (deploy) בשרת הדוחות – צילומי מסך ישלחו בהתאם במסמך המקדים

# דגשים לדוחות ה SSRS/PBI

יש ליישם את הפונקציונאליות שלהלן:

הקמת Server Report BI Power – קונפיגורציה מלאה

יישום האובייקטים הבאים בשרת הדוחות: KPI דינאמי  ,   Source ,Data Set ,Data SLICER

פרמטרים לדו"ח  מתוך רשימת בחירה

עבור הסעיפים הקשורים לשרת הדוחות server report BI Power - נדרשים צילומי מסך מלאים במסמך המקדים הכוללים כתובת URL

# Dax Calculation Formulas

יש להדגים שימוש של תשאול המודל בעזרת שימוש ב- dax queries

עבור הסעיפים הקשורים ל -  DaxStudio נדרשים צילומי מסך מלאים במסמך המקדים