



פרויקט מסכם לקורס BI

תוכן עניינים לפרויקט BI

2.....	מבוא:.....
3.....	על הפרויקט:.....
4.....	שלב התכנון ויישום סביבת העבודה.....
5.....	שלב פיתוח פתרון BI לארגון באמצעות כל מיקרוסופט.....
6.....	סביבת העבודה:.....
7.....	אופן הagation הפרויקט.....
8.....	אפיון מערכתפעולית ומבחן נתונים של החברה הסולולרית.....
9.....	מידול החברה הסולולרית.....
10.....	הנחות מקובלות בפרויקט:.....
11.....	תהליך ETL.....
12.....	פיתוח Tabular Model מעל DW:.....
14.....	דוחות:.....
15.....	dagשים לדוחות ה SSRS.....
16.....	dagשים לדוחות ה BI (Dax – I) Power BI.....



מבוא

קורס מפתח בינה עסקית הוא מסלול מקיף המכשר את המשתתפים לאחד המיקצועות המבוקשים ביותר בתחום זה והוא הטיפול במידע הנאסף במערכות המידע של הארגונים.

מטרת הפרויקט

מטרת הפרויקט המענייני היא לדמות סביבת ארגון גדולה במיוחד במינוח כך שהתלמיד יתנסה בכל מתודולוגיות BI תוך שימוש בכל האס派קטים הנלמדים בקורס וירכש ידע מעשי בכל שלבי החיים של פרויקט BI. התלמיד יתנסה מעשית בניהוועה מושתת בניתוח מידע, הבנת סביבה עסקית של הארגון ותהליכי העבודה, תכנון ארכיטקטורה ופיתוח הפרויקט בכל BI הפופולריים (Power BI , SSIS , SSRS , SSAS , T-SQL) תוך דגש על פיתוח בסביבת visual studio .

במהלך הכנות הפרויקט יתנסה ויחווה התלמיד בעיות שונות בעולם היפוי כגון מציאת מקורות מידע לאקלואס הנתונים, תכנון מערכות מידע, ניתוח נתונים, הבנת סביבה עסקית של חברה מורכבת, התמודדות עם הבעיות שגיאה, טיעיות בתיכנן וכו'.

הפרויקט מאפשר לתלמיד להבין את מורכבות עולם BI תוך כדי מציאת פתרונות אינטואיטיביים לקידום מערכת BI שתעצור למנהל האגפים ומנהלי המערכת בתכנון החלטות ומענה לבעיות שונות של רמת שירות, נטישת משתמשים, תמחור נכון וכל האתגרים העדכניים בפני חברות סלולריות בארץ ובעולם.

הפרויקט בו מתנסים התלמידים מדמה סביבה ארגונית של חברת סלולר גדולה בארץ ולמעשה אפילו הדרישות והగדרות הפרויקט נלקחו מסביבת ייצור אמיתית גדולה במיוחד.



על הפרויקט

חברת הסולולר רוצה להקים פרויקט BI גדול על מנת לאפשר למנהל' החברה בהחלטות הקרייטיות והחישבות בהם הם נדרשים להתמודד מיד' יומם ביוםו ובניהם:

- תמחור נכון של שירותי שונים
- העברת משאבי לשירות מסוים על חשבון אחר
- מציאת ליקוחות הנוטים לנוטש את החברה ולתת מענה לפני הנטיישה
- הקצאת משאבי נכונה כדי לחסוך בהוצאות

הדרישות מהתלמידים כוללות בין היתר את השלבים הבאים



שלב התכנון ויישום סביבת העבודה

- אפיון הסביבה העסקית
- הבנת הדרישות העסקיות ותהליכי העבודה בארגון. הבנת סביבת העבודה, הבנת תהליכי ארגונים, הבנת התפקיד בו אנו נדרשים לבצע פרויקט BI.
- יצירת קידימות בשלבי הפROYיקט
- אפיון משתמש הקצה של המערכת
- ניתוח מערכות המידע של הארגון, מקורות ויעדים
- פיתוח הפROYיקט בכל מיקרואופט
- הקמת מערכת מידע תפעולית לארגון
- הקמת מחסן נתונים עבור הארגון בטופולוגיה סכמת כוכב



שלב פיתוח פתרון ה BI לארגון בעזרת כלי מיקרוסופט

- פיתוח תוכניות להעברת נתונים בסיס הנתונים התפעולי למחסן הנתונים על ידי שימוש באחת הטכניקות שנלמדו בקורס
 - כלי ETL של מיקרוסופט - SSIS
 - פיתוח בשפת T-SQL
- פיתוח קוביה או מודל בעזרת כלי SSAS של מיקרוסופט שיועלה אחר כך לסביבת sql server tabular model
- שתיתן מענה מהיר ואיכותי לשאלות העסקיות של הארגון במינימום מאמץ על ידי משתמשי הקצה
 - דוחות מנהליים בעזרת מחולל דוחות (reporting services) ו - BI
 - הנגשת הדוחות ב Power BI report server



סביבה העבודה

על התלמיד לבנות פרויקט BI על פי הכללים שהוגדרו לעיל כדי לענות על שאלות עסקיות בחברת סלולרגדולה. חברת הסלולר (The Voice) מספקת שירותי מידע וקול ללקוחותיה בארץ ובעולם ולחברה מערכות תפעוליות רבות המשמשות את החברה במשךן הਪתרונות הרבים אותם מצפים ללקוחותיה לקבל לפי רמת השירות הגבוה מאוד המקובל בשוק.

המערכת התפעולית מורכבת מטבלת רשומות שיחה ענקית (CDR) ממנה ניתן לגזר את כל המידע הרצוי בתוספת של טבלאות מידע עסקי כגון :

- לקוחות
- שירותים
- אזוריים
- מקורות
- וכו'.

החברה מבקשת ליצור פרויקט BI גדול מאוד אשר ישמש מודל שיעזר 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 תנובה של שימושי קול לפי מדינה
- ניתוח שימושים שונים לפי סוג שיחות (call types)
- ניתוח דקות לפי מספר הטלפון
- ניתוח שימוש (duration) לפי חבילת (לפני ואחרי הנחה)
- ניתוח של כמות שיחות וכמות לקוחות ביום ושעה
- ניתוח לקוחות דומים – שאינם משתמשים בשירות
- ניתוח תנובה של דקות יוצאות ברמה יומית/חודשית/שנתית



אופן הגשת הפרויקט

הפרויקט כולל את הרכיבים הבאים

1. מסמך מקדים (WORD) בו יפרט התלמיד בדיאגרמות וסבירים את הגורמים הבאים:

- שמות מagiשי הפרויקט
- דיאגרמה של המערכת התפעולית
- הנחות שנלקחו בפרויקט בראשי פרקים
- סימת הכוכב של מחסן הנתונים
- שלבים בפיתוח הפרויקט
- הביטים שונים שנלקחו במהלך הפרויקט
- בעיות שונות שנטקלתם בפרויקט ודריכים להתמודדות

2. קבצי SQL:

- מחסן הנתונים - סקריפט מלא ליצירת כל האובייקטים – נתונים, טבלאות וכו'
- ג'ובים - סקריפט מלא ליצור כל הג'ובים הנדרשים לתהליכי ה ETL
- לולאה – סקריפט מלא ליצור רשומות חדשות בטבלה USAGE_MAIN – על הרשומות להיות תואמות לשאר הרשומות בטבלאות האחרות – נדרש ליצור Data עברו שונה (תחת שנות ה Time וה Date)

3. ה Solution שמכיל

- פרוייקט SSIS + קבצי Folders + CSV (עבור תהליכי הטעינה הדינמית CSV)
- פרוייקט SSAS
- פרוייקט SSRS

במקום להגיש solution אחד שיכיל 3 פרויקטים, אפשר גם 3 solutions, אחד ל SSIS, שני ל SSAS שלישי ל SSRS

4. קבצי BI + Power BI שיועלו לשרת

5. הנגשת הדוחות (PBI + SSRS) ב Power BI report server - צילומי מסך במסמך המקדים הכלולים כתובות URL

6. סקריפטים של Dax (עבור Dax Queries, פירוט בהמשך)

הערה חשובה:

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



אפיון מערכת תפעולית ומבחן נתונים של החברה הסלולרית

כללי

המערכת התפעולית מכילה רשומות שיחה (CDR) המציגות שיחות יוצאות באירגון ומאפשרת לנתח אותם בחחכים שונים .
מודול החברה הסלולרית ישמש עבור :

1. מענה לשאלות העסקיות
2. בסיס נתונים למודלים סטטיסטיים , DM ומודלים נוספים



מידול החברה הסלולרית

מערכת המתגים הארגונית

דוגמא	תיאור	מקור – יישיות במערכת התפעולית
050 - פלאפון	מתאר את המפעלים הקיימים בארץ על פי קיידומת	1. OPFILEOPP – קובץ מפעלים על פי קיידומת
ישראל - 972	מתאר את המדינות בעולם על פי קיידומת.	2. XXXCOUNTRYPRE – קובץ קיידומות מדינה
972520002375+ עבור חדש דצמבר 2013 , ניתנה בינואר 2014 , חשבון לחיבור 400 שקלים	חשבונית עבור לקוחות + עבור חדש דצמבר 2013 , ניתנה בינואר 2014 , חשבון לחיבור 400 שקלים	CUSTOMER_INVOICE .3
חכילה מס' 1 – חכילת משפחה , חכילה 2 – מסלול לחילימ	טבלת קטלוג חכילות	Package_Catalog .4
לקוח שמספרו + 972525154837 שיין לחכילת משפחה , סטוטוס פעיל , מקבל הנחה של 35% בשיחות בתוך המשפחה	טבלת קוי לקוחות	customer_lines .5
מדינה: אנגליה , ישות: אירופה	קווי מדינה בעולם	countries .6
Call - Cellular Call , TEXT - Text Messaging	סוגי שיחה/שירות -	call_type .7
לקוח מס' 1 , מס' טלפון 972541185146+ , אלברט אינשטיין , תאריך הצטרפות לחברת 5.2.2010	לקוחות הארגון	Customer .8
	מכילה את כל התנועה הארגונית	USAGE_MAIN .9



הנחות מקובלות בפרויקט

הנחות מקובלות אותן נניח בפרויקט (על מנת לפחות אותו):

- קידומת בת שלוש ספרות של מספר הטלפון מהו אינדיקציה חד ערכית לשם החברה (לדוגמא 054 = אורנג , התעלומו משירות מספרים)
- קידומת מדינה היא עד כדי שלוש ספרות (לדוגמא 1 = ארצות הברית , 972 = ישראל)
- מספר הטלפון של אדם מסוים מהו ייחוס חד ערכי למספרו קלוקוח,
- כלומר לקווי מסוימים בעלי מספר טלפון אחד בלבד
- לקווי מזוהה חד ערכית על ידי מספר טלפון כלומר לקווי מסוימים מספר טלפון אחד



תהליך ETL

תהליך ETL מעביר את הנתונים מהמערכת התפעולית למחסן הנתונים במנגנון מותזמן. חשוב לפתח פרויקט לנושא שיטוף בכל הנושא. פרויקט אחד בלבד לנושא ETL.

דרישות תהליכי ה- ETL

1. יש לתכנן מפת גזירה כולל כל השלבים MRR,STG,DWH
2. יש לבצע טעינה דינמית של קבצי CSV
3. יש לישם את התהליך באמצעות SSIS על פי כל כללי המתודולוגיה של מダンו.
4. ניתן להניח הנחות שאין סותרות את הלוגיקה העיסקית המתוארת

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

dagshim בתהליכי ה- ETL

1. יש להריץ את התהליך באופן יומי. (מנגןון SQL SERVER SCHEDULER)
2. יש לבצע תהליכי Data Cleansing מללא לרבות המרת ערכי null
3. אחרי כל lookup, במקרה של לא נמצא ערך יש להשלים עם -1 או unknown

טעינה דינמית CSV

ישנם 2 סוגי קבצי CSV, הראשון OPFILEOPP והשני call_type.

נדרש ליצור תהליכי ETL דינامي עבור כל סוג קובץ ולטעתו לטבלה מתאימה (בשלב המתאים).

על תהליכי הטעינה להיות דינامي (בדומה לתרגיל בכיתה) – שימושו לב בבחירה מתודת הטעינה, יש לבחון את הרשומות שמתאפשרות בכל קובץ וליצור תהליכי גנרי עבור כל סוג קובץ (OPFILEOPP / call_type).

בנוסף, בתהליכי הטעינה יש להעביר את הקובץ ("עליו רצים") לתיקייה Process בה תבוצע הטעינה לטבלה המיועדת, בסיום יש להעביר לתיקייה Done ולבצע Renaming לקובץ (להשאר את שמו המקורי של הקובץ ולהוסיף תאריך ושעה)

הנחות מקובלות:

1. כל קובץ נתון לתיקייה ספציפית על השרת בו מותקן sql server - יש לבצע משיכת יומית ממש.
2. שמות הקבצים שונים מפעם לפעם.
3. קובץ OPFILEOPP מכיל את ה string c opfileopp מכך call_type call_string
4. קובץ call_type מכיל את ה string c 'call_type'

שימוש ל-

עבור טעינת CSV יש לגשת ל Flat File Other Sources וatz להלן 2 קישורים לנושא:

<https://analytics4all.org/2018/03/13/ssis-lesson-2-import-data-from-csv-into-database/>

<https://www.mssqltips.com/sqlservertip/2923/configure-the-flat-file-source-in-sql-server-integration-services-2012-to-read-csv-files/>



פיתוח Tabular Model מעל DW

מודל "הטבולר" הוא למעשה החלק המאפשר גישה מהירה למשתמשי הקצה של מערכת BI על מנת שיקבלו תשובה לכל השאלות העסקיות של הארגון ב מהירות רבה תוך הצגה גרפית יפה על מנת לקבל החלטות נכונות לעתיד החברה.

בבנייה המודול, במסמך "TheVoice Source To Target" ישנה הגדרה של cube dimension usage אשר מגדיר את הסעיפים של להלן:

- .1 Rename שמות עמודות וטבלאות
 - .2 הגדרת טבלת Date Table
 - .3 יצירת יחסים בין הממדים וטבלאות ה Fact
 - .4 וידא Data Type ו – Format מתאים לעמודות (תקן במידת הצורך)
 - .5 בצע Sort By Column בעמודות הנדרשות
 - .6 "החבא" עמודות במידת הנדרש
 - .7 למודול יוגדרו Measures שיתבססו על טבלת ה fact: fact
- | | |
|----------------------------|---|
| Count | • |
| Count distinct on customer | • |
| Call Duration | • |
| Billable Call Duration | • |
| Amount | • |
| Billable Amount | • |
| Call Duration Average | • |
- .8 למודול יוגדרו Calculated Columns

Billable Amount – Amount : Discount	•
Discount / Amount : Discount Percentage	•
 - .9 יוגדרו 3 KPI וערכים מטרה דינמיים
 - .10 יש ליצור 3 פרוטוקטיבות ובהן כל הממדים והמדדים הרלוונטיים
 - .11 יש ליצור מחיצות (Partitions) בהתאם למתחם הנתונים על מנת ליצור עבודה אופטימלי למודול
 - .12 נדרש ליצור תהליכי עבודה למודול אוטומטי
 - .13 יש ליצור 3 Roles שונים ולחברים למשתמשים, לכל Role הגדרות Permission שונות
 - .14 – (RLS) Row Level Security
 - .15 יש להשתמש ב Excel browser על מנת לוודא נכונות של תוצאות Deploy Model

דgeships:

לסעיף 12 נדרש להויסף Package לפרויקט ה SSIS וכן Job בהתאם.
עבור סעיפים 14, 15 – נדרש צילומי מסך מלאים במסמך המקדים

16. בנייה המידים וההיררכיות הבאות:

- מימד מדינות - היררכיה
 - region->area->country
 - Areas
 - Countries
 - Regions
- מימד זמן - היררכיה
 - Year -> month -> Date
 - Months
 - Days In Week
- מימד שעות - היררכיה
 - Hours -> Minutes
 - Hours
 - Minutes
- מימד לקוחות - היררכיה
 - Packages -> Customers
 - Country -> Customers
 - Operator -> Customers
 - Customers
- מימד סוג שיחה - היררכיה
 - Call Type Category -> Call Types
 - Call Types
- מימד חבילה - היררכיה
 - Package Status -> Packages
 - Packages
- מימד מפעיל - היררכיה
 - Operators
 - Operators
- מימד מקור שיחות - היררכיה
 - Call Origin Type
 - Call Origin Type



דוחות

יש ליצור לפחות שישה דוחות שונים על בסיס השאלה העיסוקיות הבאות וdock נוסף לבחירה = סה"כ 7 דוחות

- ניתוח של סה"כ תנוצה של שימושי קול לפי מדינה
- ניתוח שימושים שונים לפי סוג שיחות (call types)
- ניתוח דקות לפי מספר הטלפון
- ניתוח שימוש (duration) לפי חביבה (לפניהם ואחרי הנחה)
- ניתוח של כמות שיחות וכמות לקוחות ביום ושעה
- ניתוח לקוחות רזרביים – שאינם משתמשים בשירות
- ניתוח תנואה של דקות יוצאות ברמה יומית/ חודשית/ שנתית

נדרש לבנות את הדוחות 3 ב SSRS ו 4 ב BI Power BI
את כל הדוחות יש להפנות (deploy) בשורת הדוחות – צילומי מסך ישלחו בהתאם במסמך המקיים



דגשים לדוחות ה SSRS

יש לישם את הפונקציונליות שליהן:

- הקמת Power BI Report Server – קונFIGורציה מלאה
- יישום האובייקטים הבאים בשרת הדוחות: KPI (динמי – כתוצאה משאלתה), Data Set ,Data Source ,Other Resources ,Power BI ,Paginated reports
- הרשאות אבטחה:
 - Report Server Access - user on server (Windows User) – התחברות לשרת הדוחות דרך כתובת URL, נדרש להשתמש על השרת (Windows User)
 - Report server security - role in folder + inherited & customize
 - Report running query - authentication in data source
- דוחות (RDL)
 - פרמטרים
 - Drop Down from Query ✓
 - לוח שנה ✓
 - Default Values ✓
 - Available Values from a Query ✓
 - עיצוב ◦
 - Execution time (build in fields)
 - Sort by Another column
 - Subtotal | Total ◦
 - זה"כ + Drill Down ◦
 - Parameters עם סימן חותם Drill Through ◦
 - תרשימים PIE ◦

עבור הסעיפים הקשורים לשרת הדוחות Power BI report server – נדרשים צילומי מסך מלאים במסמך המקדמים הכלולים כתובות URL



דגשים לדוחות ה Power BI (Dax – ו)

יש לישם את הפונקציונליות של להלן:

Power BI

- Dashboards •
- Visualizations •
- תרשיימי קומבינציה ○
- סלי'סרים ○
- מפות ○
- טבלאות ומטריצות ○
- Scatter Charts ○
- תרשיימי תפוזרת ○
- תרשיימי משפר ומפל ○
- תרשיימי גאוג' וקרטיסים ○
- תרשיימי KPI ○
- Filters: •
 - Visual level ○
 - Page level ○
 - Report level ○
 - Top 10 filter ○
- Interaction •
- Custom Tooltip •
- Drill through •
- Drill Down 3 ways - Different type of Drills using hierarchies •
 - What If •
 - Design •
- Image/Logo ○
- Title ○
- Paint Brush** ○
- Compare Year to Year •
- Deploy to Report Server (on-premise) •



Dax Calculation Formulas

- יש למשב במודול ה Dax FormulasCalculated Columns – ו Measures ,Tabular באמצעות (וממוסים בדוחות ה BI):

Calculate	<input type="radio"/>
Filter	<input type="radio"/>
SumX	<input type="radio"/>
Related	<input type="radio"/>
All	<input type="radio"/>
TotalYTD	<input type="radio"/>
TotalQTD	<input type="radio"/>
TotalMTD	<input type="radio"/>
SamePeriodLastYear	<input type="radio"/>
Variables	<input type="radio"/>

Dax Queries

- תשאול המודול ב Dax Studio (שימוש ב Subsets – יצרת 3 Dax Queries) עם אוריינטציה עסקית (להוציא סקריפטים של Dax)

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

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