



Project Assignment: Melodica Business Intelligence Pipeline

Ομάδα 1:

Ηλιάννα Ζαχαροπούλου
Δήμητρα Στρουμπάκου
Αναστασία Ντογραματζίδου
Γεωργία Γιαννάκη
Ασημίνα Τζάνα



REGENERATION®

 Code.Hub

Η Melodica Media Corp. είναι μία πολυεθνική εταιρία η οποία διαχειρίζεται το ηλεκτρονικό κατάστημα μουσικής Melodica. Στόχος της εργασίας αυτής είναι η δημιουργία ενός business intelligence pipeline για την ανάλυση των πωλήσεων, όπως επίσης και η υποστήριξη της διαδικασίας λήψης αποφάσεων της εταιρίας. Επομένως, για τη διευκόλυνση της υποβολής αναφορών και της ανάλυσης αρχικά σχεδιάσαμε ένα Data Warehouse σε star schema, το οποίο αποτελείται από 5 dimension πίνακες και τον fact ο οποίος περιλαμβάνει τις πωλήσεις. Έτσι λοιπόν έχουμε υλοποιήσει δύο πίνακες που κρατούν όλα τα στοιχεία των πελατών και των υπαλλήλων από την OLTP αντίστοιχα, έναν πίνακα που κρατάμε τις ημερομηνίες στον οποίο έχουν προστεθεί και οι Ελληνικές αργίες και γιορτές, έναν πίνακα με τις διευθύνσεις χρέωσης από τον οποίο έχει αφαιρεθεί το συνολικό ποσό παραγγελίας το οποίο υπολογίζεται στον Fact και τον πίνακα που περιέχει όλες τις πληροφορίες για τα τραγούδια (καλλιτέχνες, είδος μουσικής κτλ) στον οποίο ενσωματώσαμε σαν πρόσθετη πληροφορία τον αριθμό των playlist που υπάρχει κάθε τραγούδι.

Παράλληλα με την παραπάνω διαδικασία υλοποιήθηκαν τα παρακάτω ETL tasks. Το Slowly changing dimensions type 2 στον πίνακα του costumer και το Incremental loading στον πίνακα του Fact. Όταν αλλάζει η τιμή ενός attribute στον πίνακα του πελάτη, προσθέτουμε μια νέα εγγραφή στη βάση με την καινούργια πλέον τιμή της. Αυτή η νέα εγγραφή γίνεται τρέχουσα. Αντίστοιχα, για το incremental loading, λόγω μεγάλου όγκο εγγραφών, φορτώνουμε σταδιακά στον Fact, μόνο όσες εγγραφές είναι είτε καινούργιες είτε τροποποιημένες.

Όσον αφορά το SQL Server Integration Services (SSIS) αποφασίσαμε να φτιάξουμε ένα project στο Visual Studio, το οποίο να περιλαμβάνει 2 πακέτα. Το πρώτο πακέτο δημιουργεί το staging area αντλώντας δεδομένα από την OLTP βάση, δημιουργεί τον πίνακα με τις ημερομηνίες και το data warehouse κρατώντας το star schema του SQL server. Σε αυτό το πακέτο προσθέσαμε και το Slowly Changing Dimension για να συμπληρώνονται μελλοντικές εγγραφές στους dimension πίνακες. Το δεύτερο πακέτο φορτώνει τα δεδομένα στην εκάστοτε βάση, διαγράφοντας τα δεδομένα της προηγούμενης φάσης ώστε να μην υπάρχουν διπλές εγγραφές κάθε φορά που εκτελείται το project. Σε αυτό το πακέτο βάλαμε την επιλογή να δίνονται κάθε φορά indexes, να υπάρχει το incremental loading του fact πίνακα και να δημιουργείται backup της βάσης μας. Αυτό το project έγινε deploy στο SQL Server και μέσω του SQL Server Agent έχουμε δημιουργήσει μια εντολή όπου το πακέτο τρέχει συστηματικά, για την ακρίβεια κάθε πρώτη Κυριακή του μήνα, το πακέτο του loading data που δημιουργήσαμε.

Η Microsoft μέσω των Analysis Services για τη δημιουργία Business Intelligence (Data Warehouse) εκτός από τη δημιουργία πολυδιάστατων κύβων παρέχει και τη προσέγγιση Tabular μοντέλων. Στα πλαίσια της συγκεκριμένης ανάλυσης φτιάξαμε ένα Tabular model. Το Tabular μοντέλο δημιουργήθηκε στο Visual Studio με τη προέκταση Analysis Services. Στη συνέχεια έγινε deploy προκειμένου να μπορεί να γίνεται ορατό από τον SQL Server. Μέσω του SQL Server Analysis Services (SSAS) μπορούμε να δημιουργήσουμε calculated columns και DAX measures.

Το τελευταίο αλλά και πολύ σημαντικό στάδιο κάθε μελέτης που αφορά την ανάλυση δεδομένων, αφορά την εξαγωγή των τελικών συμπερασμάτων καθώς και την οπτικοποίηση τους. Αυτό βοηθά στο να παρουσιαστούν τα αποτελέσματα μιας έρευνας με τρόπο ο οποίος θα είναι κατανοητός στο ανθρώπινο μάτι. Το στάδιο αυτό ουσιαστικά αφορά την απάντηση των ερωτημάτων σχετικά με τα δεδομένα που έχουμε στη διάθεση μας με σκοπό τη λήψη αποφάσεων μέσω γραφημάτων. Αυτά τα γραφήματα τοποθετούνται σε ένα dashboard, το οποίο είναι στην ουσία μια σελίδα γραφημάτων τα οποία αλληλοεπιδρούν μεταξύ τους και αναπαρίστανται σε πραγματικό χρόνο.

Το τελευταίο μέρος της μελέτης μας λοιπόν, αποτελεί τη δημιουργία δυναμικών και διαδραστικών γραφημάτων καθώς και μετρήσεων με το Power BI. Αυτό πραγματοποιήθηκε, κάνοντας import τα

δεδομένα που προήλθαν από τις υπηρεσίες Analysis Services στο Power BI. Όπως αναφέραμε και προηγουμένως, η Melodica είναι μια εταιρία η οποία αποτελεί ένα eshop μουσικής. Η ανάλυση των δεδομένων της βάσης της Melodica μας βοήθησε να καταλήξουμε σε συμπεράσματα που αφορούν τις πωλήσεις της εταιρίας. Πιο συγκεκριμένα, η ανάλυσή μας μας οδήγησε σε συμπεράσματα που αφορούν τις πωλήσεις ανά χώρα, είδος μουσικής, καλλιτέχνη. Καταμετρήθηκαν οι πελάτες καθώς και οι παραγγελίες που έχουν πραγματοποιηθεί ανά έτος. Αναλύθηκαν επίσης τα πιο κερδοφόρα τραγούδια ανά χρονιά. Επίσης, αναλύθηκε η ποσότητα τραγουδιών που υπάρχουν στη βάση δεδομένων μας. Η ανάλυση των παραπάνω θα βοηθήσει την εταιρία στη λήψη μελλοντικών αποφάσεων οι οποίες θα βοηθήσουν στην αύξηση κερδών της.