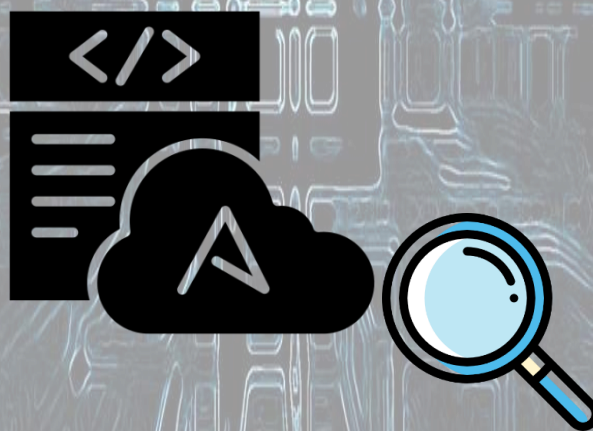




Εργασία στο μάθημα «Τεχνολογίες Ψηφιακών Υποδομών»



Καθηγητές: **Βασιλική Καλογεράκη**,
Βοηθοί Καθηγητών: **Αριστοτέλης Πέρη**

Ονόματα φοιτητών:

Καραχάλης Κωνσταντίνος, f3312307, kon.karachalis@aueb.gr

Μιχάσι Μπερνάλντο, f3312311, aldo.mihasi@aueb.gr

Τουμάσης Δημήτριος, f3312313, dim.toumasis@aueb.gr



Αρχιτεκτονική Συστήματος - Azure Taxi Application (UML Schematics)

Επιλογή εργαλείων:

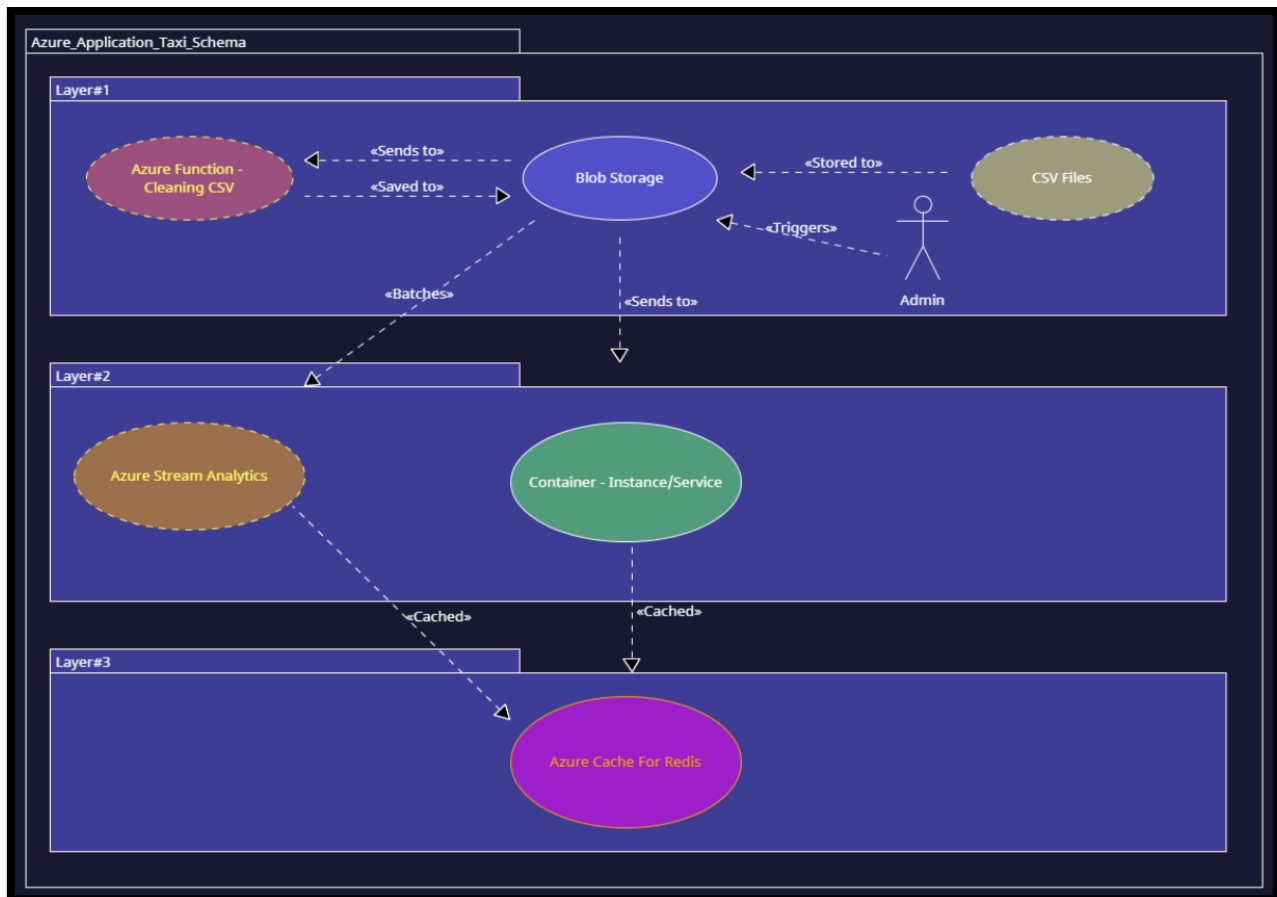
- Azure Blob Storage
- Azure Function
- Azure Container Instance
- Azure Stream Analytics
- Azure Cache for Redis

Αρχικά, ο διαχειριστής ανεβάζει τα δεδομένα σε μορφή .csv στο Azure Blob Storage. Τα δεδομένα στέλνονται σε ένα Azure Function έτσι ώστε να γίνει ο καθαρισμός τους από τυχόν μη έγκυρα και κενά δεδομένα και επιστρέφονται στο Azure Blob Storage.

Έπειτα, ο διαχειριστής μέσω του Azure Blob Storage κάνει trigger το Azure Container Instance με σκοπό την εκτέλεση των ερωτημάτων. Τα αποτελέσματα του container θα αποθηκεύονται στο Azure Cache for Redis έτσι ώστε να μπορούμε να πραγματοποιήσουμε ταχεία προσπέλαση χωρίς να χρειάζεται να ξανά υπολογίζονται τα αποτελέσματα. Πρόσβαση στη cache υπάρχει μέχρι ο διαχειριστής να κάνει trigger την όλη διαδικασία για διαφορετικό μήνα. Για παράδειγμα, έστω ότι ζητούνται τα αποτελέσματα του Ιανουαρίου και αυτά υπολογίζονται. Η αποθήκευσή τους γίνεται στην cache. Εάν ο διαχειριστής ζητήσει ξανά τα αποτελέσματα του Ιανουαρίου τότε τα λαμβάνει γρήγορα από τη cache. Σε περίπτωση που ζητήσει τα αποτελέσματα του Φεβρουαρίου τότε αυτά πρέπει να υπολογιστούν καθώς δεν υπάρχουν ήδη στη cache.

Όσο αφορά τη διαδικασία του streaming, θα χρησιμοποιηθεί το Azure Stream Analytics το οποίο θα τραβάει τα batches από το Azure Blob Storage κάθε φορά που ανεβαίνουν και θα υπολογίζει το αποτέλεσμα του ερωτήματος το οποίο θα αποθηκεύεται στο Azure Cache for Redis μέχρις ότου να ανέβει το νέο batch και το αποτέλεσμα να αλλάξει.

Επιπρόσθετα, να αναφερθεί πως ο κώδικας που θα χρησιμοποιηθεί για την εργασία θα είναι η γλώσσα προγραμματισμού Python.



Εικόνα 1: Διάγραμμα UMLet - Εφαρμογή Ταξί Azure



Reasons of Preference - Azure Taxi Application (Azure Tools)

Azure Container Instances (ACI): Το συγκεκριμένο εργαλείο είναι αρκετά βοηθητικό, διότι συμβάλλει στην γρήγορη εκκίνηση και διαχείριση των containers στο azure χωρίς την χειροκίνητη κατασκευή αντίστοιχης υποδομής. Παρέχει χαρακτηριστικά όπως κλιμακοσιμότητα, καθώς μας επιτρέπει να πραγματοποιήσουμε deploy για κάθε container ξεχωριστά σε σχέση με τον φόρτο εργασίας ανεξάρτητα το ένα από το άλλο. Μας επιτρέπει την κλιμάκωση με βάση την ζήτηση, με αποτέλεσμα να αυξάνονται τα instances που απαιτούνται για καλύψουν τον αυξανόμενο φόρτο. Ακόμη ένα χαρακτηριστικό που το διέπει αποτελεί η ανοχή στα σφάλματα αφού σε περίπτωση αποτυχίας κάποιου container, πραγματοποιεί επανεκκίνησή του σε έναν υγιή κόμβο.

Azure Blob Storage (ABS): Ο αποθηκευτικός χώρος blob (Binary Large Object) για το Azure μας παρέχει την δυνατότητα να αποθηκεύουμε δεδομένα χωρίς κάποια συγκεκριμένη δομή, σε μορφή αντικειμένων. Κατέχει την κλιμακωσιμότητα, διότι δεν περιορίζεται από τον παράγοντα χώρου για ανεξάρτητα blobs που μπορούν να αποθηκευτούν. Δηλαδή, υπάρχει η δυνατότητα αποθήκευσης έως και πολλαπλών terabytes και δημιουργίας όσων blob απαιτούνται. Από την άλλη πλευρά, το ABS προκαλεί αντίγραφα των δεδομένων μας 3 φορές προ-επιλεγμένα με αποτέλεσμα να διέπεται από μεγάλη ανθεκτικότητα και ανοχή σε σφάλματα υλικού, λογισμικού.



Azure Stream Analytics (ASA): Αποτελεί μία υπηρεσία ανάλυσης δεδομένων σε πραγματικό χρόνο που επεξεργάζεται streams από δεδομένα εφαρμογών. Είναι κλιμακώσιμο καθώς μπορεί να κλιμακώνεται δυναμικά ανάλογα με τον διαφορετικό φόρτο εργασίας ανά περίπτωση. Συνεπώς, μας επιτρέπει να χειριζόμαστε και μικρά και μεγάλα stream δεδομένα αποδοτικά. Το ASA έχει την δυνατότητα δημιουργίας αντίγραφων από δεδομένα εντός πολλαπλών κόμβων σε μία υποδομή Azure και συνεχίζει χωρίς απώλεια δεδομένων σε περίπτωση λάθους.

Azure Cache For Redis (ACR): Χρησιμοποιείται για την πραγματοποιήσει caching σε δεδομένα που επεξεργάζονται ανά τακτά χρονικά διαστήματα αποσκοπώντας στην βελτίωση της απόδοσης και της κλιμακωσιμότητας. Χάρη την δυνατότητα του εργαλείου να υποστηρίζει κάθετη και οριζόντια κλιμάκωση (δηλαδή, στο 1^ο θα έχουμε διαφορετικό μέγεθος cache και στο 2^ο χρησιμοποιούνται clusters για τον διαμοιρασμό μεταξύ πολλών instances στην cache) διατήρουμε την κλιμακωσιμότητα στις εφαρμογές μας. Επιπρόσθετα, συμπεριλαμβάνει πολλά χαρακτηριστικά μεταξύ αυτών και την αντιγραφή πολλαπλών κόμβων εντός ενός cluster στο Azure με αποτέλεσμα εάν υπάρξει σφάλμα σε έναν από τους κόμβους τα δεδομένα να αποκαθίστανται από τα αντίγραφα.



Βιβλιογραφία

- ❖ <https://azure.microsoft.com/en-us/products/storage/blobs>
- ❖ <https://azure.microsoft.com/en-us/products/container-instances>
- ❖ <https://azure.microsoft.com/en-us/products/cache>
- ❖ <https://learn.microsoft.com/en-us/azure/reliability/reliability-containers>
- ❖ <https://learn.microsoft.com/en-us/azure/stream-analytics/stream-analytics-time-handling>
- ❖ <https://learn.microsoft.com/en-us/azure/storage/common/storage-redundancy>
- ❖ <https://learn.microsoft.com/en-us/azure/azure-functions/functions-overview?pivots=programming-language-python>