

Κατανεμημένα Συστήματα

Distributed Systems

Εαρινό Εξάμηνο 2022-2023

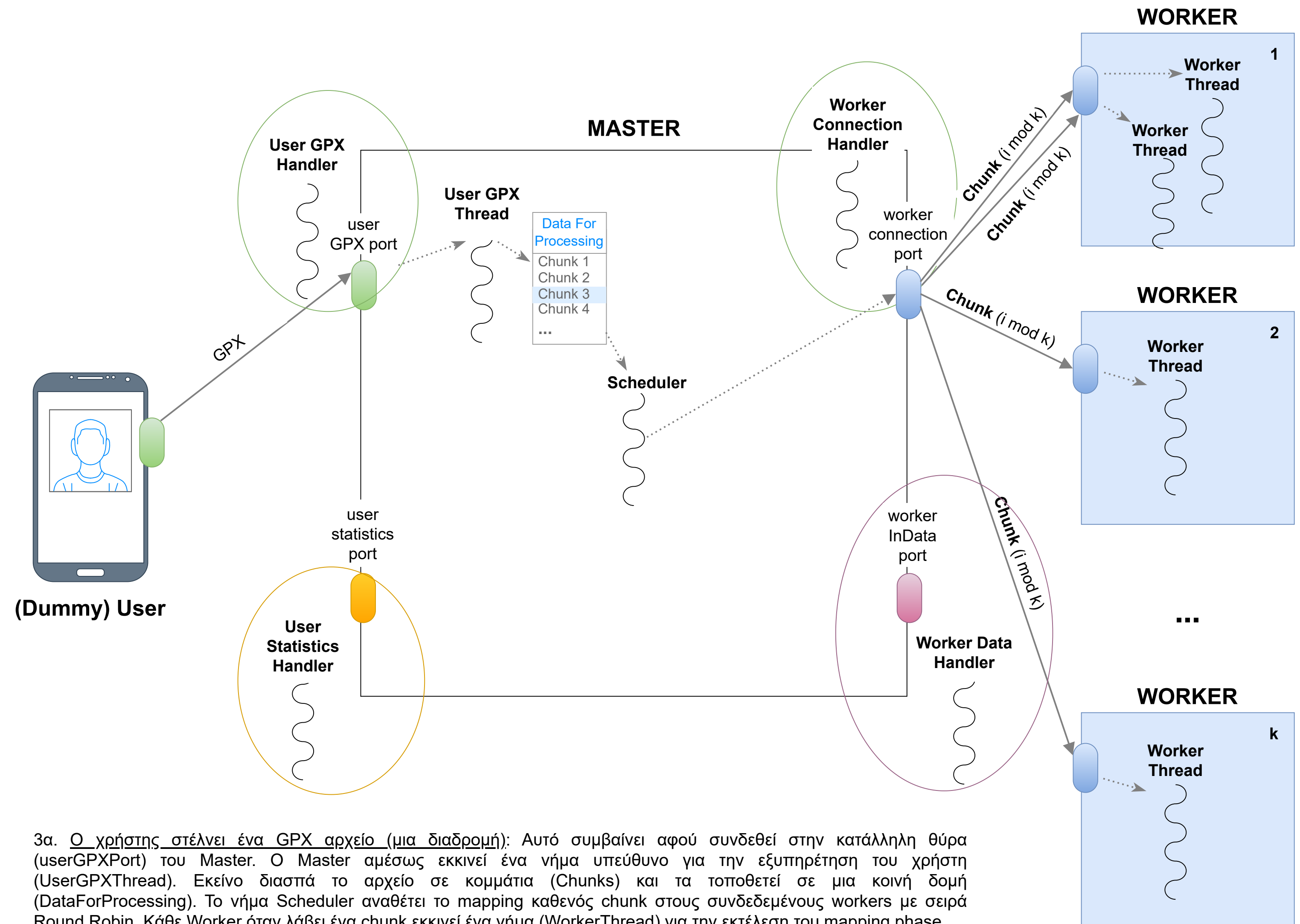
1ο Παραδοτέο

Αλβιόνα Μάντσο - 3200098

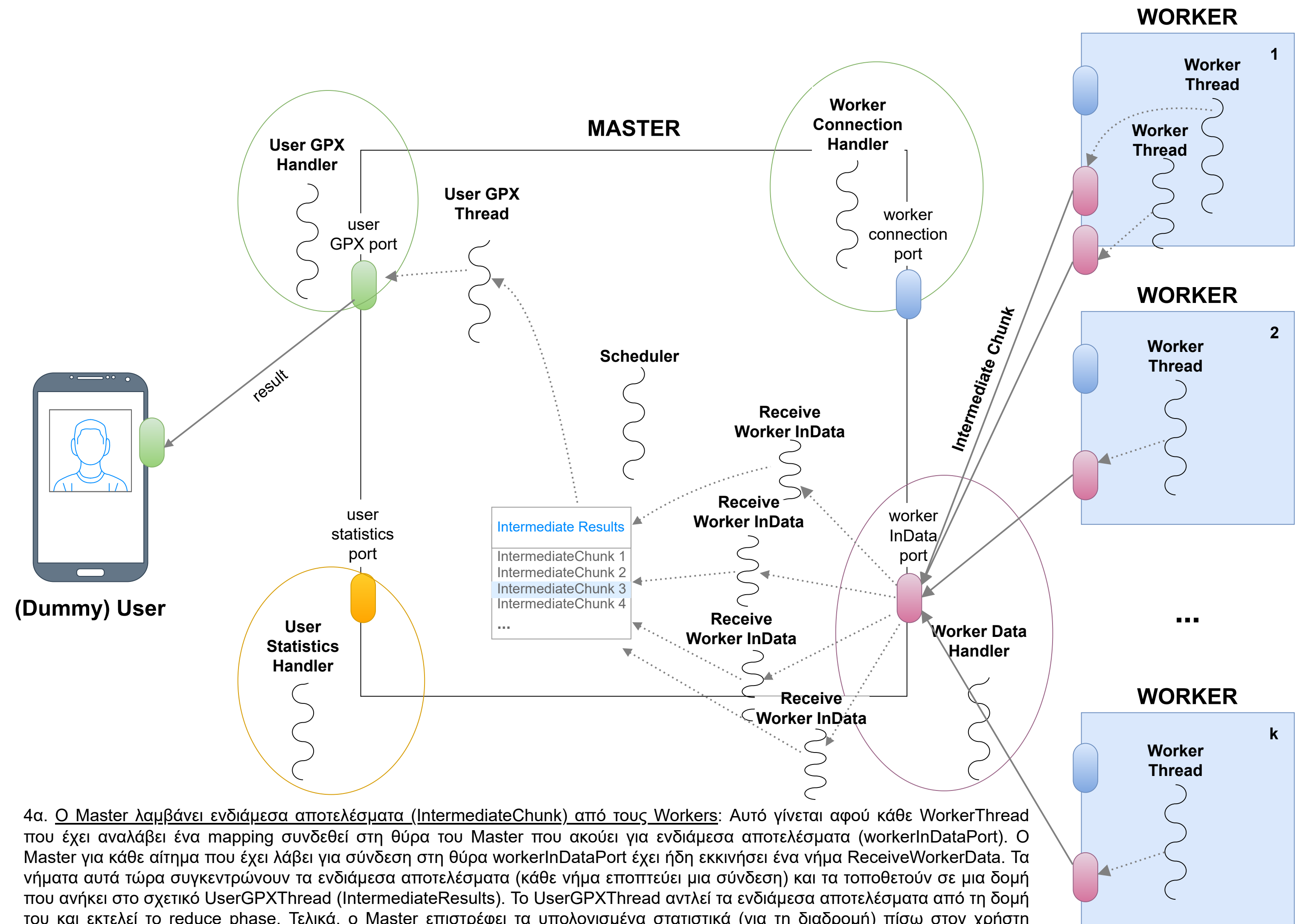
Χρήστος Πατρινόπουλος - 3200150

Διονύσης Ρηγάτος - 3200262

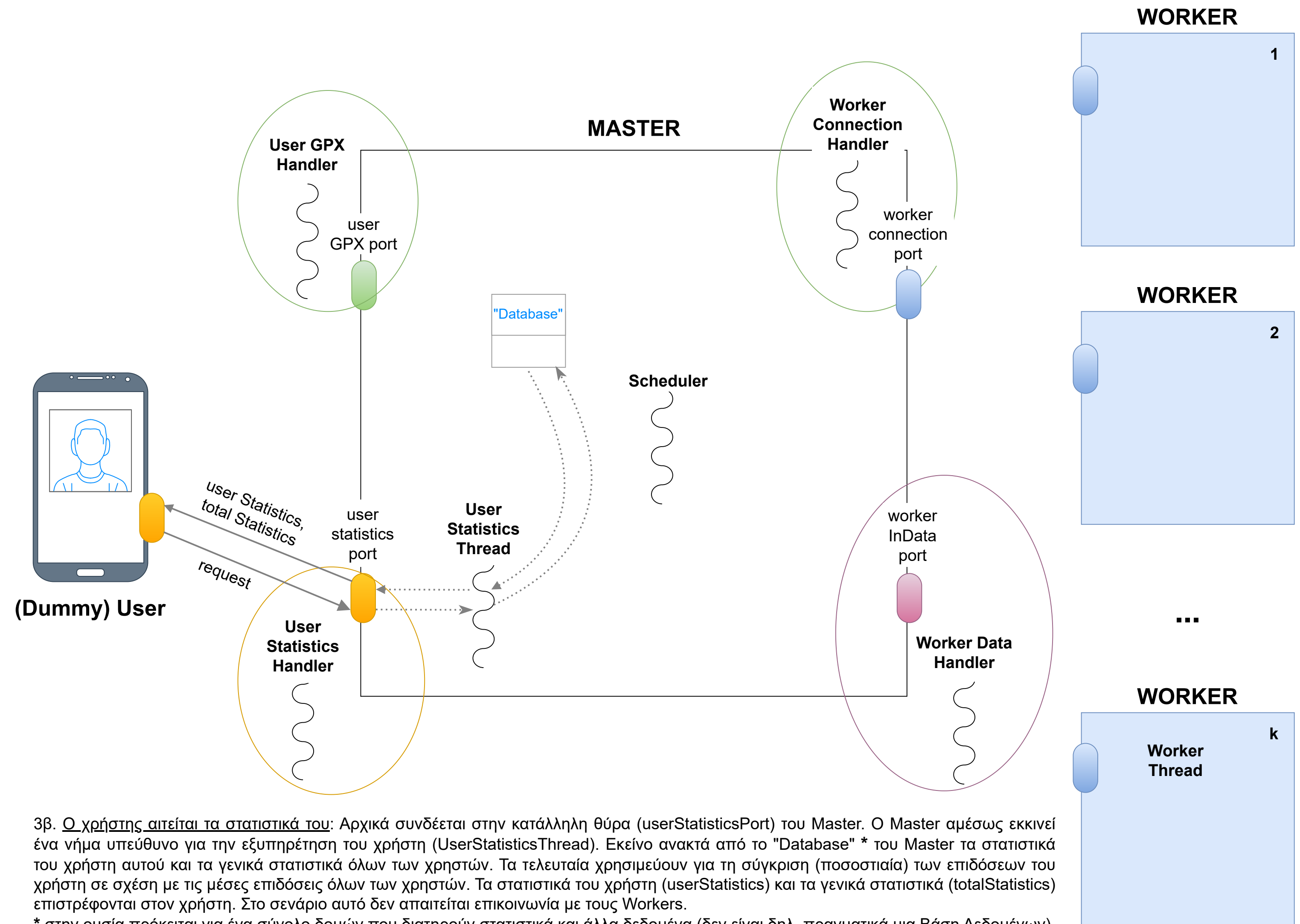
Αναφέρονται συνοπτικά τα βασικά σημεία της υλοποίησης του MapReduce με έμφαση στις δικτυακές συνδέσεις μεταξύ των συστημάτων του User, του Master και των Workers. Περιλαμβάνονται ενδεικτικά σχήματα και σύντομες περιγραφές της υλοποίησης των βασικών δυνατοτήτων που υποστηρίζονται για τους χρήστες της εφαρμογής.



3α. Ο χρήστης στέλνει ένα GPX αρχείο (μια διαδρομή): Αυτό συμβαίνει αφού συνδεθεί στην κατάλληλη θύρα (userGPXPort) του Master. Ο Master αμέσως εκκινεί ένα νήμα υπεύθυνο για την εξυπηρέτηση του χρήστη (UserGPXThread). Εκείνο διασπά το αρχείο σε κομμάτια (Chunks) και τα τοποθετεί σε μια κοινή δομή (DataForProcessing). Το νήμα Scheduler αναθέτει το mapping καθενός chunk στους συνδεδεμένους workers με σειρά Round Robin. Κάθε Worker όταν λάβει ένα chunk εκκινεί ένα νήμα (WorkerThread) για την εκτέλεση του mapping phase.



4α. Ο Master λαμβάνει ενδιάμεσα αποτελέσματα (IntermediateChunk) από τους Workers: Αυτό γίνεται αφού κάθε WorkerThread που έχει αναλάβει ένα mapping συνδεθεί στη θύρα του Master που ακούει για ενδιάμεσα αποτελέσματα (workerInDataPort). Ο Master για κάθε αίτημα που έχει λάβει για σύνδεση στη θύρα workerInDataPort έχει ήδη εκκινήσει ένα νήμα ReceiveWorkerData. Τα νήματα αυτά τώρα συγκεντρώνουν τα ενδιάμεσα αποτελέσματα (κάθε νήμα εποπτεύει μια σύνδεση) και τα τοποθετούν σε μια δομή που ανήκει στο σχετικό UserGPXThread (IntermediateResults). Το UserGPXThread αντλεί τα ενδιάμεσα αποτελέσματα από τη δομή του και εκτελεί το reduce phase. Τελικά, ο Master επιστρέφει τα υπολογισμένα στατιστικά (για τη διαδρομή) πίσω στον χρήστη (result).



3β. Ο χρήστης αιτείται τα στατιστικά του: Αρχικά συνδέεται στην κατάλληλη θύρα (userStatisticsPort) του Master. Ο Master αμέσως εκκινεί ένα νήμα υπεύθυνο για την εξυπηρέτηση του χρήστη (UserStatisticsThread). Εκείνο ανακτά από το "Database" * του Master τα στατιστικά του χρήστη αυτού και τα γενικά στατιστικά όλων των χρηστών. Τα τελευταία χρησιμεύουν για τη σύγκριση (ποσοστιαία) των επιδόσεων του χρήστη σε σχέση με τις μέσες επιδόσεις όλων των χρηστών. Τα στατιστικά του χρήστη (userStatistics) και τα γενικά στατιστικά (totalStatistics) επιστρέφονται στον χρήστη. Στο σενάριο αυτό δεν απαιτείται επικοινωνία με τους Workers.

* στην ουσία πρόκειται για ένα σύνολο δομών που διατηρούν στατιστικά και άλλα δεδομένα (δεν είναι δηλ. πραγματικά μια Βάση Δεδομένων)