Tujuan dari penggunaan Microservices adalah untuk scale up dari segi business dan developer. Pada umumnya refactor dari monolith ke microservices terjadi karena terdapat permasalah saat development dan deployment aplikasi. Misalkan perusahan sudah memiliki banyak engineer, tetapi arsitektur aplikasi masih menggunakan monolith. Ini tentu saja akan jadi masalah saat ada perubahan (penambahan fitur, dll). Masalahnya antara lain:

1. Proses development menjadi lambat karena setiap fitur baru atau perubahan baru akan di cek semuanya terlebih dahulu, apakah perubahan baru tersebut akan berpengaruh ke fitur yang tidak ada perubahan.
2. Pada proses release rentan terjadinya bug/error karena bisa saja ada indikasi error pada fitur baru yang dampaknya ke fitur yang lama.
3. Proses deployment ke server juga akan menjadi lambat, karena semakin besar aplikasi dalam suatu monolith maka akan semakin lama juga proses deploymentnya. Biasanya lama proses deploymentnya itu karena semakin banyak unit test/integration test, proses build aplikasi yang cukup lama, dan proses run aplikasi di server yang cukup lama juga.
4. Untuk scaling aplikasi monolith lebih cocok menggunakan scaling tipe vertical, yaitu dengan menambah jumlah resource pada server atau menambah server baru yang dimana cara ini lebih rumit.

Dari masalah diatas maka microservices bisa menjadi pilihan yang tepat, jika aplikasi menjadi microservices maka akan mendapatkan keuntungan antara lain:

1. Proses development akan menjadi cepat, karena setiap tim yang mengerjakan fitur tidak perlu khawatir akan mengganggu tim lain yang sedang mengerjakan fitur yang berbeda karena setiap tim mengerjakan service yang berbeda (per service per team).
2. Pada saat release service dan terjadi error tidak akan mengganggu service lain karena yang berubah hanya service tersebut.
3. Proses deployment akan menjadi lebih cepat, karena releasenya berdasarkan service.
4. Untuk scaling bisa menggunakan scaling tipe horizontal, yaitu dengan menambahkan jumlah aplikasi pada suatu server. Biasanya scaling horizontal dilakukan pada aplikasi yang sudah berjalan di container dan sudah di manage dengan Container Orchrestation Engine seperti Kubernetes.