

این پیام استاد بوده:

باید یکی از مقالات و به عنوان پایه انتخاب کنید با یه دیتا ست پیاده سازی کنید و بعد روش خودتون و هم با همون دیتاست پیاده سازی کنید و نتایج و براساس معیار های مختلف باهم مقایسه کنید
انتخاب مقاله باید باید بر اساس شباهت روش به عنوان انتخابی شما باشه انتخاب دیتاست هم می تونه با شما باشه لزوما دیتا ست مقاله پایه می تونه نباشه شما دیتاست مرتبط با claad database رو از سایت kaggle.com دانلود می کنید
بررسی و نرمال سازی می کنید و الگوریتم مقاله پایه و الگوریتم خودتون و روش پیاده سازی می کنید و بعد مقایسه می کنید باید توی روش پیشنهادی تون از lstm یا ترانسفورمر استفاده کنید ضمناً پیاده سازی رو روی یک دیتا ست از اول با هر دور روش مقاله پایه و روش خودتون انجام بدین
ضمناً موارد توزیع داده و پارتیشنینگ جداول پایگاه داده رو باید حتما بیارین چون در موضوع پایان نامه تون هست بنابراین دیتاست باید مناسب انتخاب بشه

اینم روش من که تایید کردن

روش مقاله اینه که با استفاده از داده های تاریخی workload و Polynomial Regression ، workload آینده رو پیش بینی می کنه و بعد بر اساس workload پیش بینی شده، میزان منابع مورد نیاز رو محاسبه کرده و منابع ابری رو به صورت پیش بینانه Scale-out/Scale-in می کنه. مقاله این روش رو هم در محیط CloudSim ارزیابی کرده است.
ارتباطش با موضوع من در بحث Elasticity و Predictive Auto-scaling در محیط ابری هست، مقاله منابع رو عمده تاً به صورت یکپارچه و در قالب VM در نظر گرفته و Disaggregated Cloud Database و Hybrid Workload رو بررسی نکرده.

و من روش پیشنهادیم رو به این صورت در نظر می گیرم:

در روش پایه، Polynomial Regression مقاله برای پیش بینی workload و سپس فرمول تخصیص منابع و Scale-out/Scale-in پیاده سازی می شود.
در روش پیشنهادی، Polynomial Regression با یک LSTM ساده جایگزین می شود تا workload ترکیبی OLTP و OLAP پیش بینی شود. بعد با حفظ منطق محاسبه منابع مقاله پایه، نیاز Compute و Storage به صورت جداگانه محاسبه و Scaling آنها مستقل انجام می شود. همچنین Data Distribution و Table Partitioning نیز متناسب با workload پیش بینی شده در تصمیم گیری مربوط به منابع و Storage در نظر گرفته می شوند.
هر دو روش از ابتدا و با یک دیتاست یکسان پیاده سازی و با معیارهای یکسان مقایسه خواهند شد و دیتاست نیز به گونه ای انتخاب می شه که اطلاعات لازم برای Hybrid Workload ، منابع، توزیع داده و Partitioning را داشته باشه.