**بهبود توازن بار در مدل MapReduce برای خوشه­های ناهمگن**

چكيده

در سال­های اخیر، مدل­های برنامه‌نویسی مختلفی برای پردازش داده­های بزرگ ارائه شده­اند. در این میان، MapReduce توانسته است محبوبیت زیادی، به ویژه در محیط­های رایانش ابری بدست آورد. MapReduce علاوه بر مزایا و ویژگی­های منحصر به فردی که دارد، طبیعتاً دارای چالش­های مختلفی نیز هست. مسئله توازن بار یکی از چالش­های مهم در مدل­های برنامه‌نویسی توزیع‌شده از جمله MapReduce است، و در صورتی که سازوکار مناسبی برای آن اتخاذ نشود می­تواند باعث افزایش زمان اجرای برنامه­ها و در نتیجه کاهش کارایی شود. از جمله معیارهایی که در توازن بار نقش کلیدی دارند، محلیت داده­­ها و انحراف داده­ها هستند. ضمن اینکه برای برقراری توازن بار، باید به همگن یا ناهمگن بودن خوشه­های محاسباتی نیز توجه داشت.

در این تحقیق، روشی بر اساس کنترل ظرفیت به نام LBC2 ارائه می­شود که توازن بار را در MapReduce برای خوشه­های ناهمگن بهبود می­بخشد. بر اساس ارزیابی­ها و آزمایش­های انجام شده در این تحقیق، LBC2 با افزایش محلیت داده­ها و در نتیجه بهبود توازن بار، زمان اجرا را به طور میانگین حدود 20% بهبود بخشیده است.

**واژگان كليدي:** MapReduce، Hadoop، توازن بار، داده­های بزرگ، خوشه­های ناهمگن.

Improvement of load balancing in MapReduce for heterogonous clusters

Abstract

Numbers of various programming models have been proposed to process Big Data in recent years. However, MapReduce is the most famous programming model amongst cloud computing environments and includes many advantages, yet there are several challenges to deal with. Load balancing is considered as one of the most significant downsides of MapReduce which causes the increase in applications’ runtime and accordingly results in less-efficiency, where there is no appropriate proposed mechanism. Although, Data Locality and Data Skew are known as two main key factors for determination of load balancing, yet it is remarkable that load balance highly depends on whether the computational clusters are homogeneous or heterogeneous. This research propose a load balancing scheme using capacity controlling (LBC2) that improve MapReduce performance in heterogeneous clusters. The experimental results indicate that LBC2 reduce the job execution time by 20% using improvement of data locality.

**Keywords:** MapReduce, Hadoop, Load Balancing, Big Data, Heterogeneous Clusters.