ملخص البحــث رقـــم (5)

عنوان البـــــحث باللغة العربية

استغلال الفرص المعاد استخدامها في تحسين لاستعلامات المتعددة في البيانات الضخمة

اسماء المؤلفين:

RadhyaSahal, Mohammed H. Khafagy and Fatma A. Omara

مكان النشر و تاريخه:

Journal of Computational Science, Elsevier,2017

ملخص البـــحث باللغة العربية :

أصبح التحسين متعدد الاستعلامات في البيانات الضخمة اتجاهًا بحثيًا رائدانظرًا في مجال أنظمة تحليل البيانات الضخمة (مثل MapReduce و Flink). حيث يتم ترجمة الاستعلام المتعدد إلى وظائف. وأيضا تقديم هذه الوظائف بشكل روتيني مع مهام مماثلة للأنظمة التحليلية للبيانات الضخمة. هناك بعض التقنيات الحالية التي تم اقتراحها لاستغلال مهام المشاركة في التخصيص متعدد استعلامات البيانات الضخمة (مثل MRShare و Relaxed MRShare)هذه التقنيات هي عوامل مُحسّنة ومُصممة بشكل كبير للفرص الدقيقة المعاد استخدامها. وفقًا لتحسين استعلامات البيانات الضخمة المتعددة ، فإن التقنيات الدقيقة الحالية معنية فقط بحجم مجموعات متساوية وتوزيع منتظم للبيانات. لا تنطبق هذه المشكلات على التطبيقات الموزعة في العالم الحقيقي ، مثل حجم المجموعات غير المتكافئ وتوزيع البيانات غير الموحد. تحظى هاتان المسألتان باهتمام أكبر في تحسين استعلامات البيانات الضخمة المتعددة ، لتقليل البيانات التي تمت قراءتها أو إعادة كتابتها إلى البنى التحتية للبيانات الضخمة (على سبيل المثال ، Hadoop). في هذا البحث ، تم اقتراح تحسين متعدد الاستعلامات باستخدام نظام Tuple Size و Histogram (MOTH) للنظر في دقة الفرص المعاد

استخدامها. يستغل نظام MOTH المقترح coarse grained للفرص المعاد استخدامها كليًا وجزئيًا بين الاستعلامات مع مراعاة حجم المجموعات غير المتكافئة وتوزيع البيانات غير المنتظم لتجنب الحسابات المتكررة. وفقًا لنظام MOTH المقترح ، تم تقديم تقنية مشتركة لتقدير الفرص المعاد استخدامهاباستخدام coarse grained أفقياً ورأسياً. تم إجراء التقدير الأفقي لحجم المجموعات غير المتكافئة عن طريق استخراج البيانات الوصفية في مستوى العمود ، بينما يتعلق التقدير الرأسي لتوزيع البيانات غير المنتظم باستخدام المدرج التكراري المحسوب مسبقًا في مستوى الصف. بالإضافة إلى ذلك ، يقدّر نظام MOTH الفرص القائمة على إعادة استخدامCoarse grained مع مراعاة التخزين البطيء (أي الموارد المادية المحدودة أو الموارد الافتراضية المخصصة أقل) لإنتاج تقدير دقيق لتكاليف النتائج المعاد استخدامها. أيضا تم تقديم خوارزمية إرشادية قائمة على التكلفة لتحديد أفضل فرصة قائمة على إعادة الاستخدام وإنشاء خطة تنفيذ فعالة لاستعلامات متعددة.نظرً ا لأنه تم النظر في الفرص الجزئية المعاد استخدامها ، هناك حاجة إلى حسابات إضافية لاسترداد النتائج غير المشتقة. حيث تم تصميمهُحسِّن جزئي قائم على إعادة الاستخدام وإضافته إلى نظام MOTH المقترح لإعادة صياغة خطة الاستعلامات المتعددة التي تم إنشاؤها لتحسين الاستعلامات الجزئية المشتركة. وفقًا للنتائج التجريبية لنظام MOTH المقترح باستخدام معيار TPC-H ، فقد وجد أنه تم تقليل وقت تنفيذ الاستعلامات المتعددة من خلال مراعاة دقة النتائج المعاد استخدامها.

البحث مشتق من رسالة علمية

يقع البحث ضمن مجالات البحث بالقسم العلمي