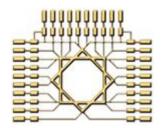
الجمهورية العربية السورية المعهد العالي للعلوم التطبيقية والتكنولوجيا



مقابلة بمشروع أساسيات النظم الموزعة البحث الموزع في الوثائق

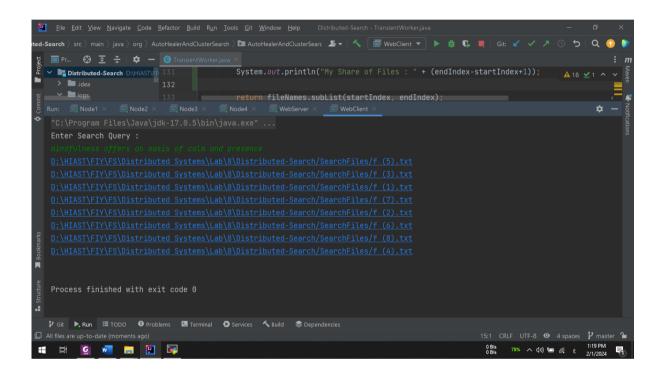
تقرير الطالب: محمد سلامه

ثم أجب عن الأسئلة التالية:

- 1 وضمّح الخرج للمستخدم النهائي الذي يطلب الخدمة مع
 - وصف مكونات النظام عدد العمال.
 - عدد الملفات الموزعة.
- 2. وضمّح من خلال الخرج على كل عامل عدد الملفات التي سيوكل بالبحث فيها عند استلام طلب البحث.
- 3 في حال فشل أحد العمال أجري البحث من جديد ووضح كيف سيتم إعادة تقسيم الملفات على العدد الجديد من العمال
 - 4 في حال فشل القائد وضبح كيف سيتعامل النظام مع هذا الفشل.
- 5. وضمّح رسائل JSON المرسلة والمستقبلة من الزبون مع مخدم الوب.
- 6. وضمّح كيف تم معالجة عدم وجود كلمات البحث في الوثائق الموزعة.
- 7. وضمّح كيف تم معالجة وجود وثائق لها نفس النتيجة Score بالنسبة لاستعلام البحث.
- 8 بين مزايا مشروعك عن طريق معالجتك لحالات خاصة أو إضافية تغنى تجربة المستخدم أو تحسن الأداء أو التوسعية في المشروع.

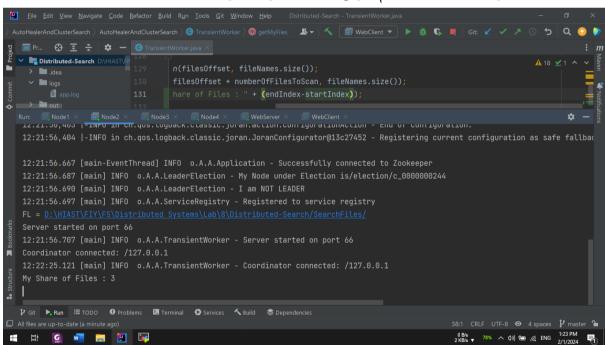
1- المستخدم يقوم بإدخال جملة تمثل query عبر Web Client وهو صف جافا، ويأتي الخرج على شكل strings كل منها يمثل رابط الملف الكامل (موقعه في الجهاز) وتأتي الأجوبة مرتبة تنازلياً وفقاً لأهميتها والتي يتم حسابها بالخلفية (tf, idf).

2- الدخل (باللون الأخضر) والخرج باللون الأزرق يكون كما يلي:

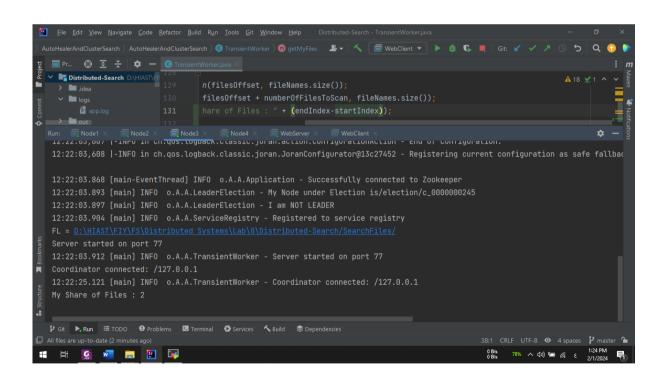


لدينا 4 عقد وواحدة هي الـ Coordinator فهي لا تعمل بالبحث، في حالتنا Node1 هي هذه العقدة:

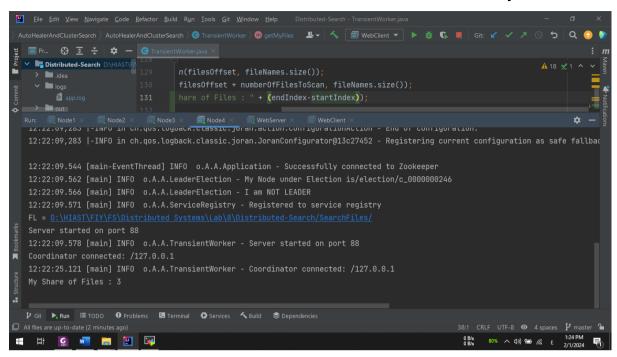
لدينا 3 عقد و 8 ملفات فيتم توزيعها 3 و 3 و 2:



العقدة الثانية:

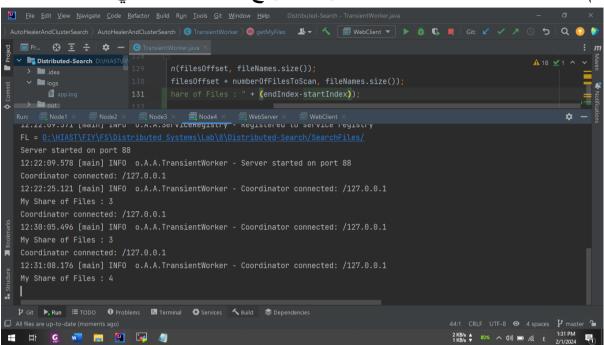


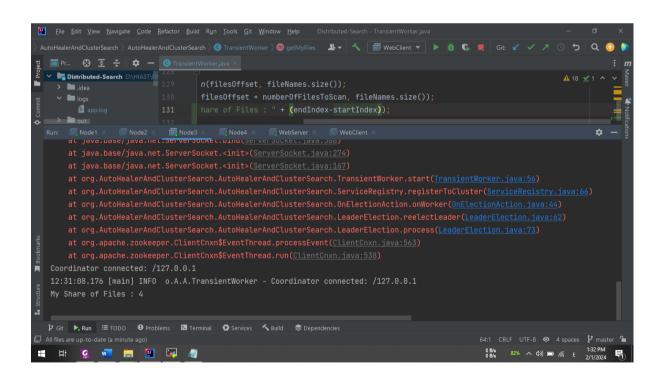
العقدة الثالثة:



3- حالة فشل أحد العمال:

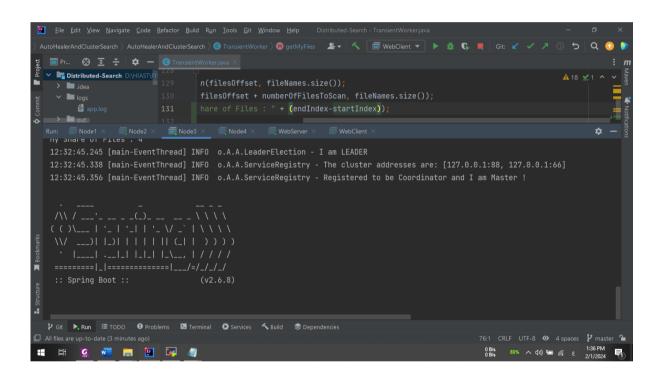
تم أيقاف Node2 عن العمل فنرى توزيع الملفات كما يلي:



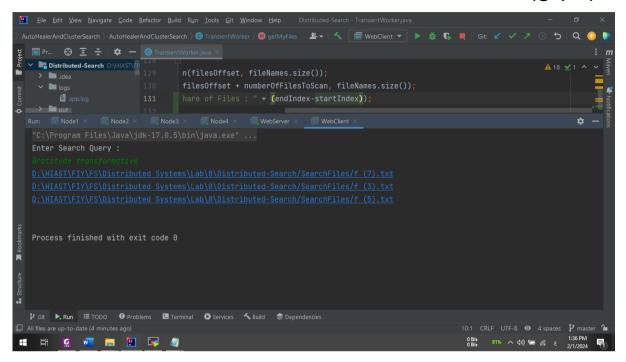


4- عند فشل القائد يتم تعيين قائد آخر من العمال ويقوم ذلك العامل بالانسحاب من خدمة البحث في الملفات والتسجيل ك Coordinator ويتسلم مهام استقبال اتصال grpc وتوزيع الملفات على العمال وحساب idf.

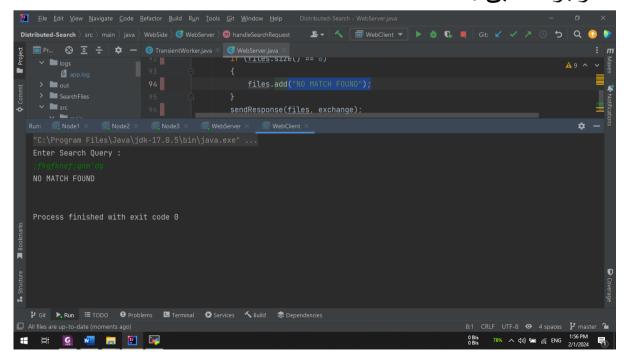
تم إيقاف القائد عن العمل وهي Node1 فتم انتخاب Node3 كقائد جديد ويطبع عناوين العقد الموجودة في الخدمة



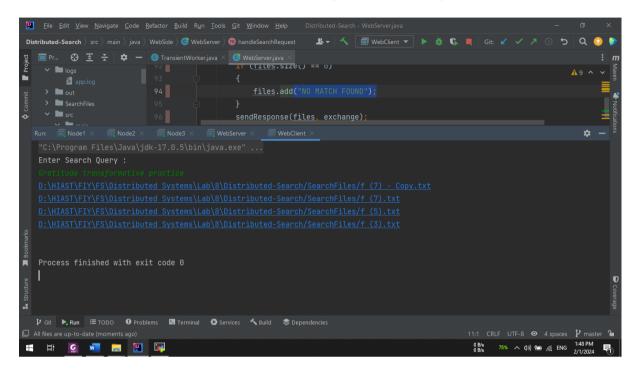
وتم تنفيذ بحث جديد بعد التغيير لاختبار تغييرات عناوين الاتصالات (grpc) وعنوان coordinator لتبادل الرسائل ضمن cluster:



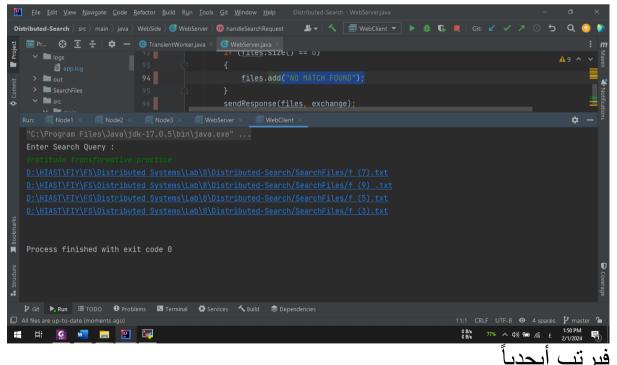
- 5- (لم يتم تحقيق واجهة web) بل التواصل عبر bytes ضمن http.
- 6- عند عدم وجود كلمات البحث في الوثائق يتم طباعة رسالة عدم وجود تطابق:



7- عند تساوي اثنان من الملفات بالـ score يتم عرضهم حسب الترتيب الأبجدي (حسب النظام):



فمثلاً للتوضيح المثال السابق تم تغيير اسم f(7) -copy إلى (9):



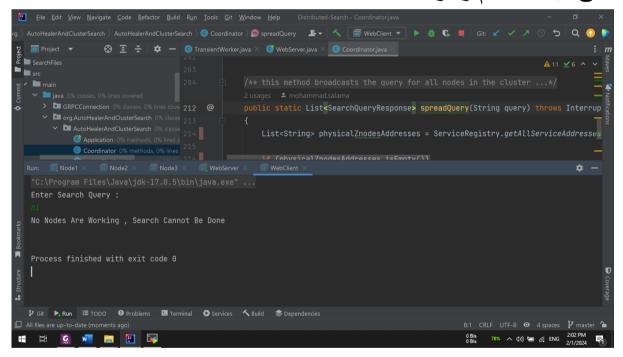
8- عند عدم وجود ملفات لا ينتج خطأ بل فقط كل عامل يأخذ 0 ملف.

حالة مهمة: بفرض أن الملف فارغ فعند حساب نسبة تكرار كلمة سوف نواجه حالة تقسيم 0/0 فتمت معالجة حالة أن عدد كلمات الملف يكون 1 إذا كان الملف فارغاً.

طبعاً نستطيع أساساً آلا نعالج الملف ولكن تم افتراض حالة تعديل ملف ما أثناء البحث (إضافة أو حذف كلمات).

تم فصل اعتماديات الصفوف عن بعضها قدر الإمكان لتحقيق قدر معين من التجرد، فبفرض أردنا تغيير عمل Coordinator نقوم بالتعديل عليه دون التأثر بالاتصال grpc أو عمل workers أو حتى طرق التواصل بين الجميع.

عند عدم وجود عقد عاملة وطلب بحث يتم طباعة رسالة عدم القدرة على البحث لعدم وجود عقد تعمل:



تم الاعتماد على فكرة threads و ExecutorService لتوزيع الأعمال وتحقيق التواصل بين Coordinator والعمال وبالتالي لا يؤثر تأخر أحدهم عن الآخر.