مقدمه‌ای بر الستیک‌سرچ

درس ارائه مطالب علمی و فنی

پارسا محمدیان

دانشگاه صنعتی شریف

۲۱ خرداد ۱۴۰۰

فهرست مطالب

- موتور جستجو متن‌باز بر پایه‌ی لوسین (Lucene)
  - لوسین موتورجستجو متن‌باز نوشته شده به زبان جاوا است

- توزیع شده (Distributed)
- مقیاس‌پذیر (Scalability)
- امکان جستجو Full-text
- رابط HTTP و RESTful
- اسناد غیروابسته به قالب (Scheme-free) و بر پایه‌ی JSON

Elasticsearch چیست؟

کاربردهای الستیک‌سرچ در صنعت

شکل: نمونه‌های کابربردهای مختلف و مشتری‌ها[۱]

[۱] تصویر از [۱]

- خوشه یا Cluster ⇐ تعدادی از سرورهای الستیک‌سرچ که به هم متصل هستند
- گره یا Node ⇐ هر یک از سرورهای الستیک‌سرچ در Cluster
- سند یا Document ⇐ هر یک از اسناد متنی که بارگذاری می‌شود
- مخزن یا Index ⇐ مخازنی که دارای تعدادی سند با قالب یکسان هستند
- نگاشت یا Mapping ⇐ قالب هر Index
- تکه داده یا Shard ⇐ هر Index تعدادی Shard دارد
- بخش یا Segment ⇐ هر Shard تعداد Segment دارد

الستیک‌سرچ	دیتابیس‌های رابطه‌ای
Index	Database
Type[a]	Table
Document	Row
Field	Column
Mapping	Schema

[a] Removed since version 6.00

کمی نکات منفی در مورد الستیک!

- مفهوم Transaction ندارد
- پشتیبانی محدودتر از Join

نکته مهم

این محدودیت‌ها برای دیتابیس‌های NoSQL و Distributed قابل انتظار است.

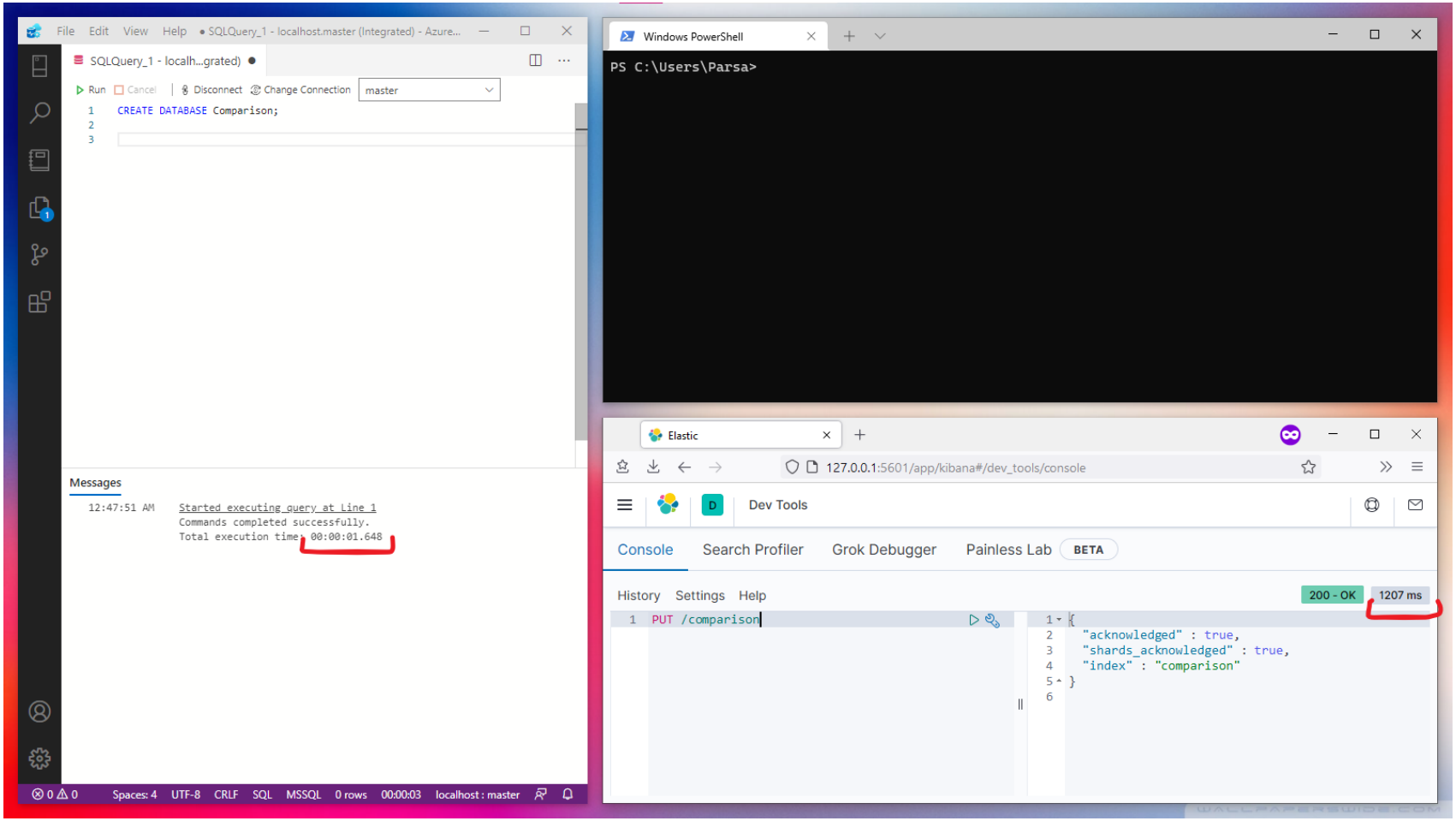
- داده بارگذاری شده اطلاعات املاک تهران موجود در دیوار است. [۴]
- این داده شامل ۳۷۲۳۳۲ رکورد و حجم خام 143 MB است.
- تست‌ها بر روی یک کامپیوتر شخصی با رم 8 GB و پردازنده Core i7-8550 انجام شده است.

ابزارهای مورد استفاده

برای کوئری زدن به الستیک از ابزار Kibana که یک وب اپلیکیشن است استفاده شده است. همچنین برای بارگذاری داده در الستیک با استفاده از ابزار Curl درخواست HTTP فرستاده شده است. برای ارتباط با SQLServer از Azure Data Studio استفاده شده است.

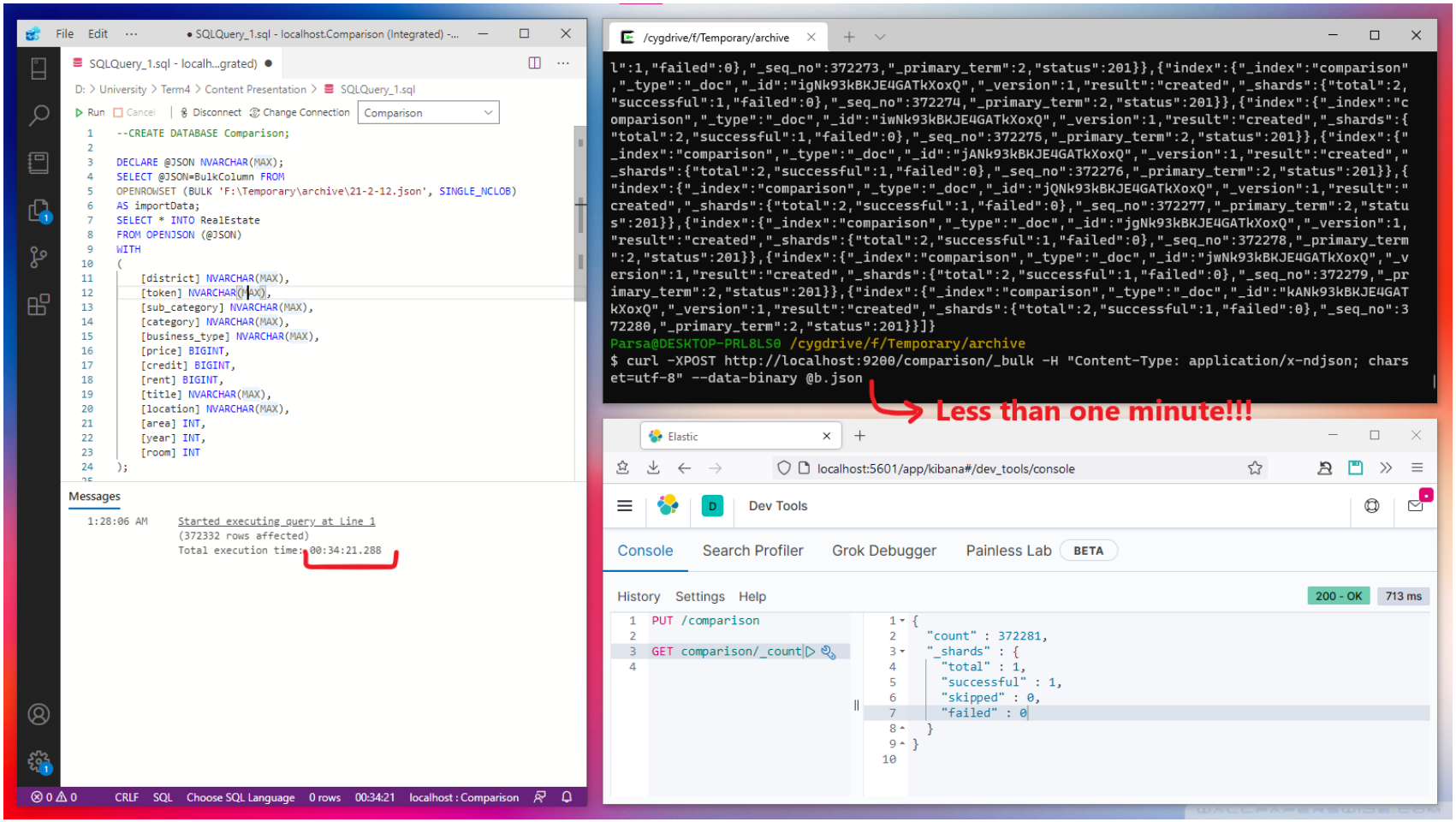
کارایی در عملیات مختلف

عملیات ساختن مخزن



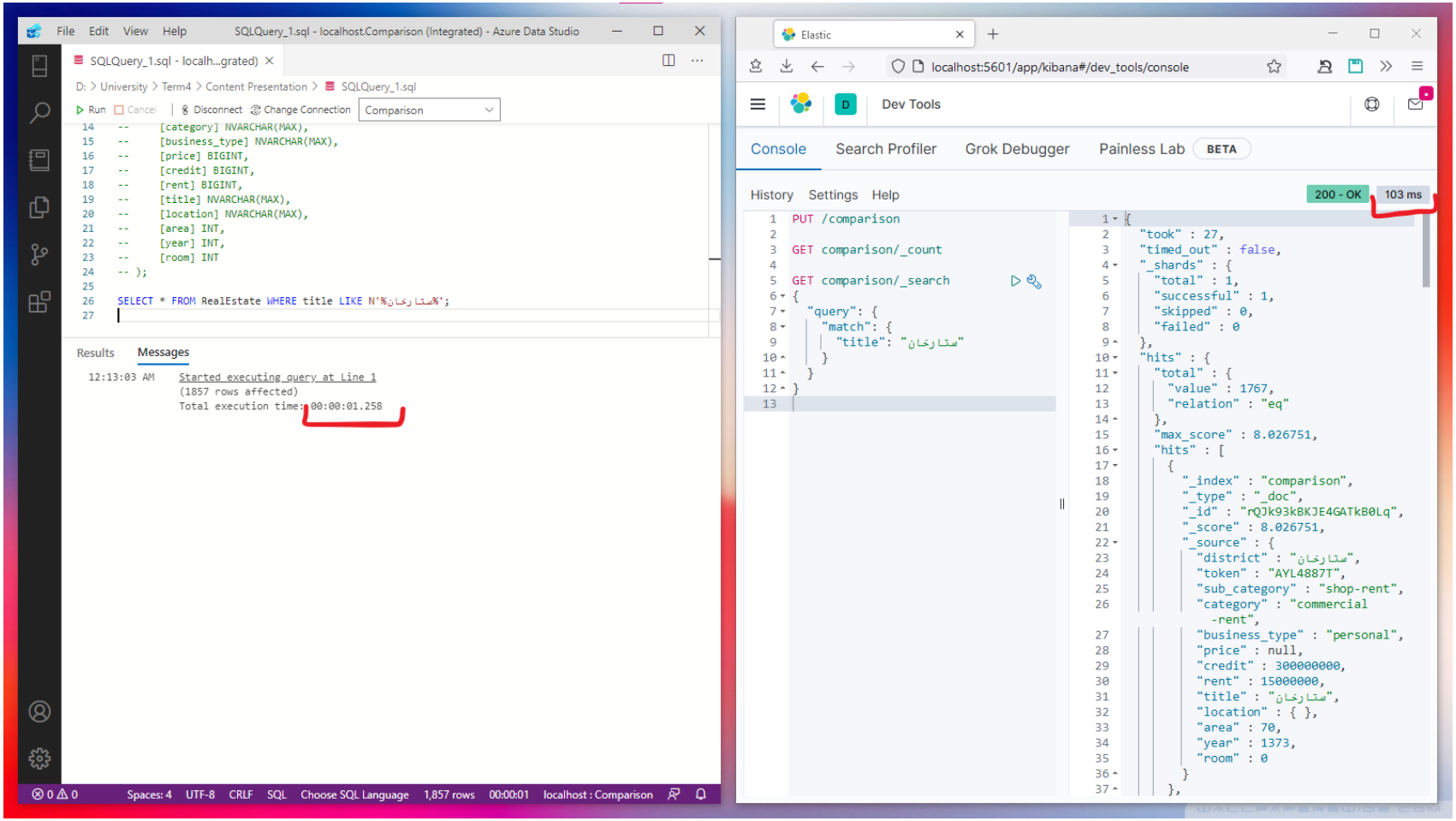
کارایی در عملیات مختلف

عملیات بارگذاری داده



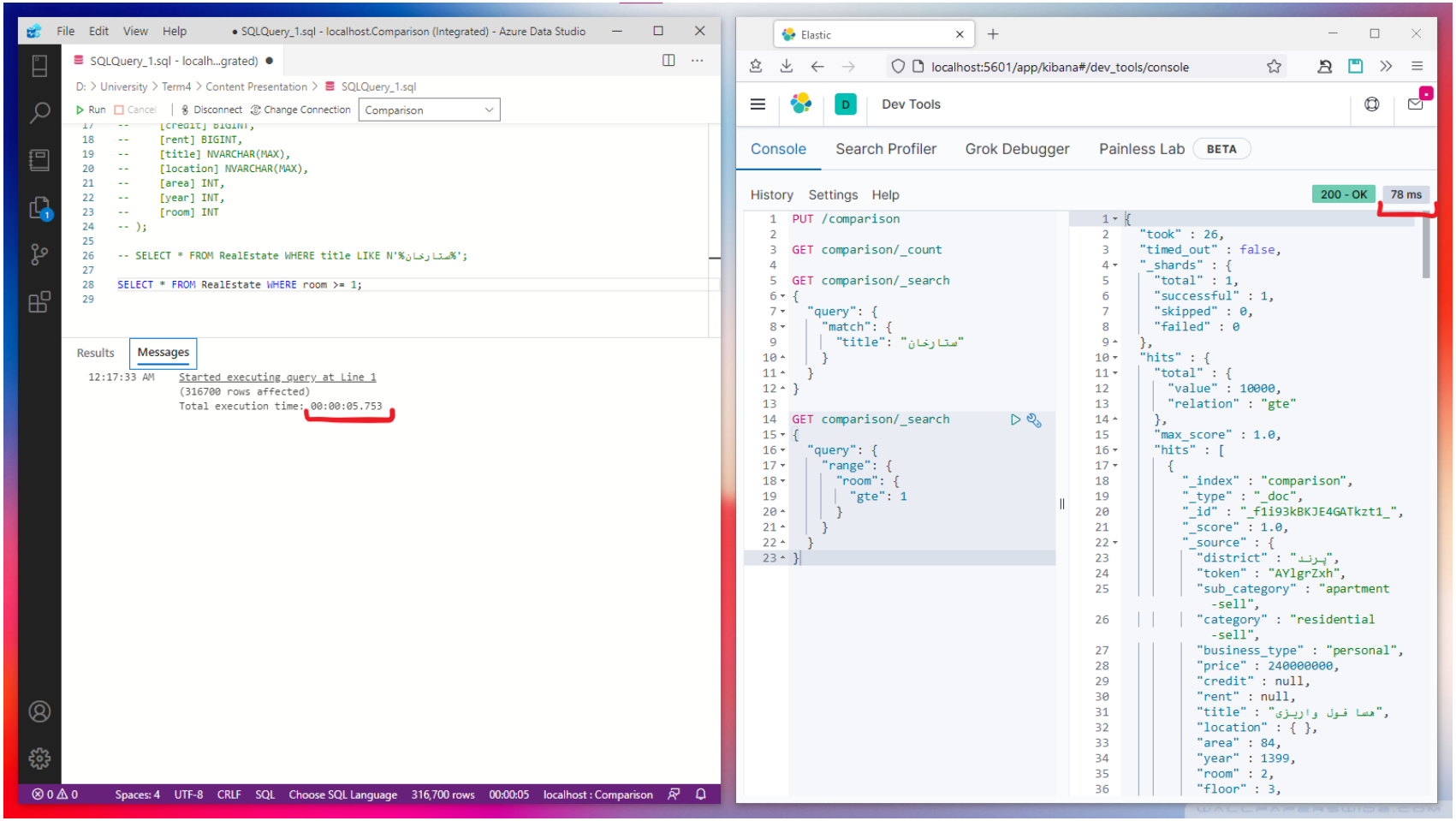
کارایی در عملیات مختلف

عملیات جستجو زیررشته



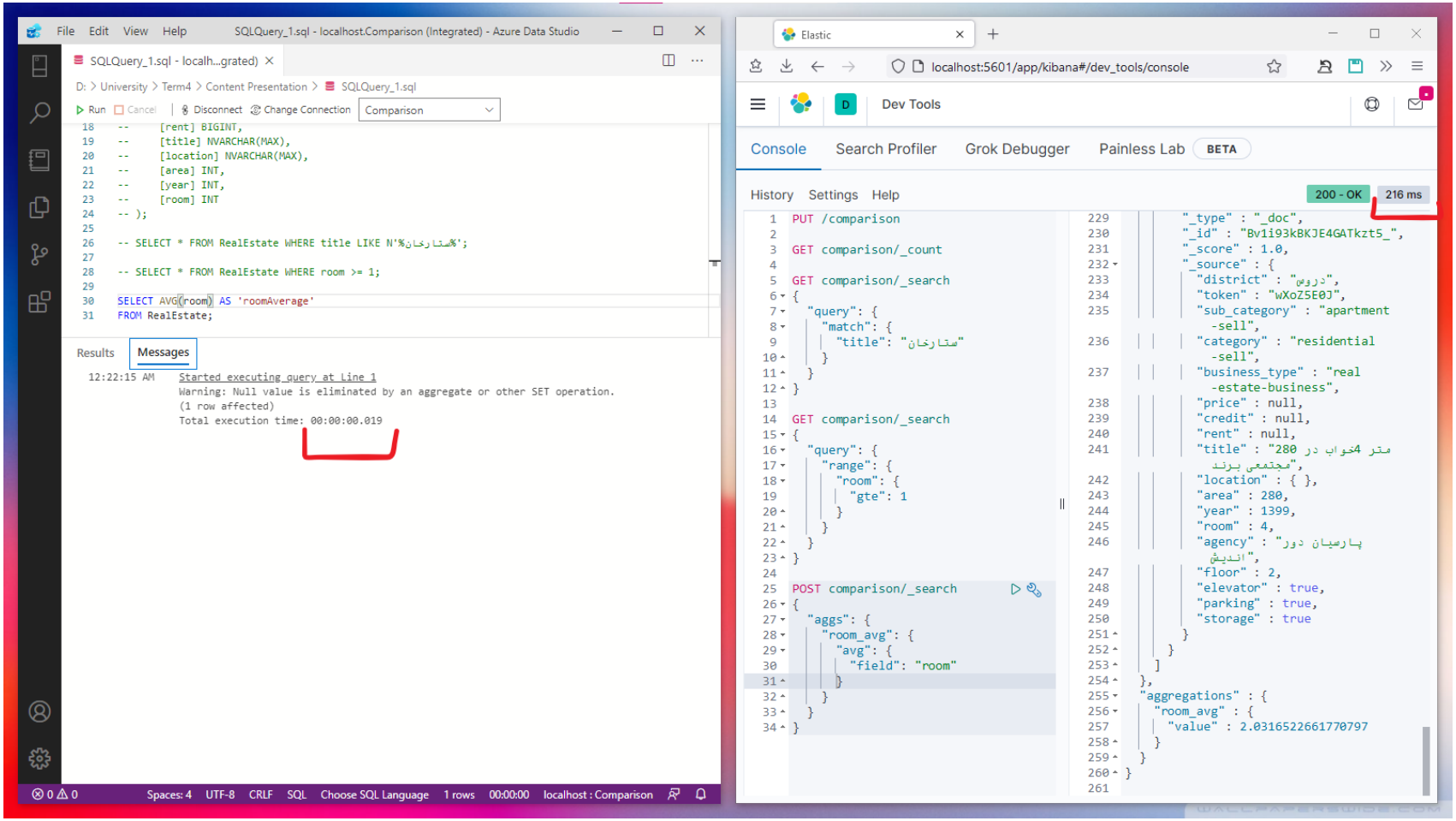
کارایی در عملیات مختلف

عملیات جستجو بازه



کارایی در عملیات مختلف

عملیات تجمعی (Aggregation)



سند۱ :

سند۲ :

منابع

[1] Howard Chen. *Elastic Q1 FY2021 Analysis.* URL: https://blog.publiccomps.com/elasticsearchq1fy2021/ (visited on 06/06/2021).

[2] Zachary Tong Clinton Gormley. *Elasticsearch: The Definitive Guide: A Distributed Real-Time Search and Analytics Engine.* O'Reilly Media, 2015. ISBN: 1449358543.

[3] *Elastic Stack and Product Documentation.* URL: https://www.elastic.co/guide/index.html (visited on 04/19/2021).

[4] Amirali Taheri. *tehran-real-estate-prices.* URL: https://www.kaggle.com/amiralitaheri/tehranrealestateprices (visited on 06/11/2021).

اگر جایی منبع ذکر نشده از [۳] و [۲] استفاده شده است.