

**ElasticSearch&Attivio  
替换验证**

**验证环境：上海市第六人民医院**

珀金埃尔默企业管理（上海）有限公司

2021年3月5日

摘要

附件：Attivio&ElasticSearch替换验证文档

目录

[一、验证背景 - 2 -](#_Toc65878956)

[二、验证环境 - 2 -](#_Toc65878957)

[三、部署服务器说明 - 2 -](#_Toc65878958)

[四、验证结果 - 3 -](#_Toc65878959)

[五、结论 - 7 -](#_Toc65878960)

### 一、验证背景

在过去的项目实施中，我们常用Attivio这款产品做为文本数据解析和搜索引擎。但在实际应用中发现，Attivio产品不稳定，服务不易启动；无法做分布式部署，当数据量庞大时，解析和搜索性能低；功能不完善，一些常用功能缺失，且存在丢失数据现象。所以我们需要寻求一款，更高效，更稳定的软件来替代Attivio。ElasticSearch是一款分布式多用户能力的全文搜索和数据分析引擎，是目前一种流行的企业级搜索引擎。具有实时搜索、稳定、可靠、快速、安装使用方便等特点。基于此，我们尝试用真实客户数据环境，验证Attivio被ElasticSearch替换的可能性，此文档为功能、性能及稳定性的替换验证结果文档。

### 二、验证环境

上海市第六人民医院住院EMR数据量2千万+，门诊EMR数据量1千万+，总计3千万+，以半结构化文本的形式存储在Sql Server数据库表里。具有数据量大，但数据存储较规范的特点。Attivio已经在上海六院上线使用。上海六院可提供多台服务器供使用。基于以上情况，上海六院较适合做为Elasticsearch&Attivio替换验证环境。

### 三、部署服务器说明

Master Node 主节点，部署在172.30.100.146服务器。

Slave Node1 附节点1，部署在172.30.100.158服务器。

Slave Node2 附节点2，部署在172.30.100.148服务器。

### 四、验证结果

| 序号 | 功能描述 | Attivio证据 | ElasticSearch证据 | | 验证结果 | 验证时间 |
| --- | --- | --- | --- | --- | --- | --- |
| 1. 数据源 | | | | | |  |
| F1.1 | 支持接入Sql Server数据源中的表和视图。 |  |  | Pass | | 03/05/2021 |
| F1.2 | 支持在软件中直接写入sql脚本。 |  |  | Pass | | 03/05/2021 |
| F1.3 | 支持对加载入的表做分类标签。 |  |  | Pass | | 03/05/2021 |
| F1.4 | 支持字段显示名称的更改。 |  |  | Pass | | 03/05/2021 |
| F1.5 | 支持接入多种字段类型及输出，例如字符型、数值型、日期型。 |  |  | Pass | | 03/05/2021 |
| F1.6 | 支持生成表的删除。 |  |  | Pass | | 03/05/2021 |
| F1.7 | 表或索引为真删除。 | 略 | 略 | Pass | | 03/05/2021 |
| F1.8 | 已生成表，支持新增字段添加。 |  |  | Pass | | 03/05/2021 |
| F1.9 | 解析数据量，支持住院2千万+门诊1千万=3千万+以上的数据解析。 |  |  |  | |  |
| F1.10 | 3千万+以上数据量解析性能，对于数据源为Sql Server数据库并且半结构化存储，应在72小时之内。 | 48小时 | 时间 |  | |  |
| 2. 语义 | | | | | | |
| F2.1 | 对指定某字段，可以解析生成两个新字段，为肯否定字段。例如：  “肯定\_病历内容”  “否定\_病历内容” |  |  | Pass | | 03/05/2021 |
| F2.2 | F2.1新生成字段同时，原字段仍可输出显示。 |  |  | Pass | | 03/05/2021 |
| 3. 搜索 | | | | | | |
| F3.1 | 支持简单搜索：and  （多个关键词，全文检索） |  |  | Pass | | 03/05/2021 |
| F3.2 | 支持简单搜索：or  （多个关键词，全文检索） |  |  | Pass | | 03/05/2021 |
| F3.3 | 支持简单搜索：not  （多个关键词，全文检索） |  |  | Pass | | 03/05/2021 |
| F3.4 | 支持and、or、not三者混合连用检索 |  |  | Pass | | 03/05/2021 |
| F3.5 | 支持指定字段，指定关键词检索。 |  |  | Pass | | 03/05/2021 |
| F3.6 | 支持指定表，指定字段，指定关键词检索。 |  |  | Pass | | 03/05/2021 |
| F3.7 | 检索结果支持字段全部内容显示。 |  |  | Pass | | 03/05/2021 |
| F3.8 | 检索结果中可高亮显示关键词。 |  |  | Pass | | 03/05/2021 |
| F3.9 | 检索性能，亿条数据达到秒级结果返回。 | 略 |  |  | |  |
| 4. SearchUI自定义功能 | | | | | | |
| F4.1 | 可以嵌入在Spotfire页面。  （支持客户端和Web端） |  |  | Pass | | 03/05/2021 |
| F4.2 | 在Spotfire页面免登录。 |  |  | Pass | | 03/05/2021 |
| F4.3 | 点击“Export”按钮，将符合关键词的ID号（例如住院号）导出到Spotfire的table中。字段名字为搜索关键词。导出哪个字段，是可以在文件中自定义配置的。（支持客户端和Web端） |  |  | Pass | | 03/05/2021 |
| F4.4 | 可设置导出限制，点击“Export”按钮，将符合关键词的ID号（例如住院号）导出到Spotfire的table中，高于某个条数，不能导出。 |  | C:\Users\linda\AppData\Local\Microsoft\Windows\INetCache\Content.Word\33.jpg | Pass | | 03/05/2021 |
| 5. 维护 | | | | | | |
| F5.1 | 支持增量处理 |  | 略 | Pass | | 03/05/2021 |
| F5.2 | 千万级数据量增量处理时间应在72小时内（紧针对Sql Server） | 略 | 时间 |  | |  |
| F5.3 | 支持自动化定时执行，且生成log |  |  | Pass | | 03/05/2021 |
| F5.4 | project可备份，且可导入其它环境 |  |  | Pass | | 03/05/2021 |

### 五、结论

以上各功能点、性能测试、稳定性测试结果均为Pass，由此证明，ElasticSearch可替换Attivio做为新的一款全文搜索和数据分析引擎。