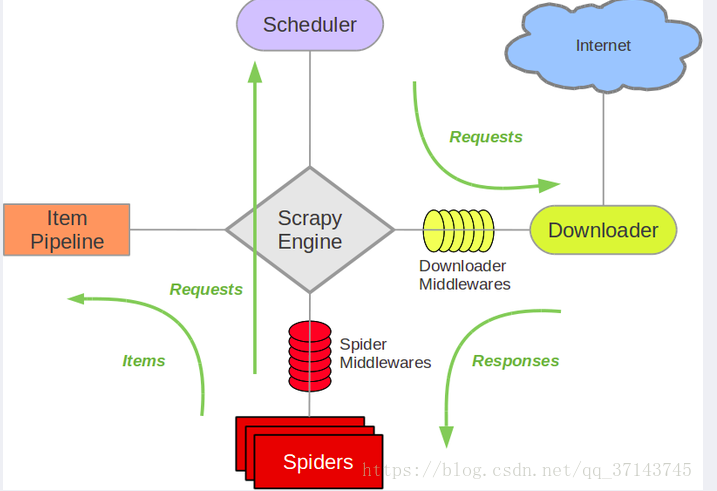
# 爬虫

### scrapy框架

* 
* **spider**
* **pipeline**
* **download**
* **engine**
* **scheduler**

### 数据清洗

* re**正则表达式库**

### 数据存储

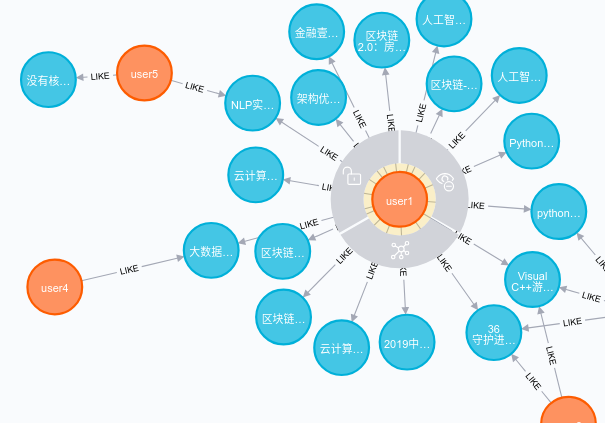
* **利用redis生成自增主键**
* **mongo**
* **mysql**

# 搜索引擎

### elasticsearch

* Elasticsearch是一个高度可扩展的开源全文搜索和分析引擎。它允许您快速，近实时地存储，搜索和分析大量数据。它通常用作底层引擎/技术，为具有复杂搜索功能和要求的应用程序提供支持。
* Elasticsearch是一个近实时搜索平台。这意味着从索引文档到可搜索文档的时间有一点延迟（通常是一秒）。
* **python整合elasticsearch框架**
* **用elasticsearch搜索文章**

# 知识图谱



### neo4j图数据库

* 我们周围的世界无处不在。Neo4j旨在有效地存储，处理和查询数据模型中高度连接的数据。借助功能强大且灵活的数据模型，您可以在不损失丰富性的情况下呈现真实世界，可变结构的信息。
* 可视化通常可以比文本格式支持更好地理解您的数据。正如我们的一位客户所说的图形可视化，“我可以查看交易和欺诈活动的图表。”
* Neo4j提供了可视化数据的方法，例如面向开发人员的Neo4j浏览器，面向分析人员的Neo4j Bloom以及寻求自然语言搜索的其他人，以及供开发人员将图表直接嵌入其应用程序的库。我们还有几个合作伙伴在这个领域提供额外的选择，以满足各种可视化需求。
* **python整合neo4**j
* **neo4j存储用户文章知识图谱**
* **neo4j查找关系**