**APACHE SOLR**

APACHE SOLR INTRO:

* It stands for**“Searching On Lucene w/ Replication”.** Solr is an open-source search platform which is used to build search applications.
* It is a search engine based on Apache Lucene library. Lucene is a free and open-source search engine software library written in Java by Doug cutting.
* Solr is also used as document-based NoSQL database with transactional support that can be used for storage purposes and even a key-value store.
* Solr runs as a standalone full-text search server.

SOLR

Structured or semi structured

The data is indexed and stored

Various sources

Or un-structured data

Here the **examples of various sources** is data in **csv, json, xml files** and also you can **directly upload the xlsx files** also

FEATURES:

* **Restful APIs** − To communicate with Solr, it is not mandatory to have Java programming skills. Instead you can use restful services to communicate with it. We enter documents in Solr in file formats like XML, JSON and .CSV and get results in the same file formats.
* **Full text search** − Solr provides all the capabilities needed for a full text search such as tokens, phrases, spell check, wildcard, and auto-complete.
* **Enterprise ready** − According to the need of the organization, Solr can be deployed in any kind of systems (big or small) such as standalone, distributed, cloud, etc.
* **Flexible and Extensible** − By extending the Java classes and configuring accordingly, we can customize the components of Solr easily.
* **NoSQL database** − Solr can also be used as big data scale NOSQL database where we can distribute the search tasks along a cluster.
* **Admin Interface** − Solr provides an easy-to-use, user friendly, feature powered, user interface, using which we can perform all the possible tasks such as manage logs, add, delete, update and search documents.
* **Text-Centric and Sorted by Relevance** − Solr is mostly used to search text documents and the results are delivered according to the relevance with the user’s query in order.

Apache Solr Terminology:

Few important terminology in solr for ease in understanding:

1. **Solr Instance**: a Solr Instance is an instance a Solr running in the JVM. In Standalone mode, Solr contains only one instance where as in Cloud mode, it contains one or more instances.
2. **Solr Core**: a Solr Core is also known as simply “Core”. A Core is an Index of texts and fields available in all documents. One Solr Instance can contain one or more Solr Cores. In other words, a Solr Core = an instance of Apache Lucene Index + Solr Configuration (solr.xml,solrconfig.xml etc.)

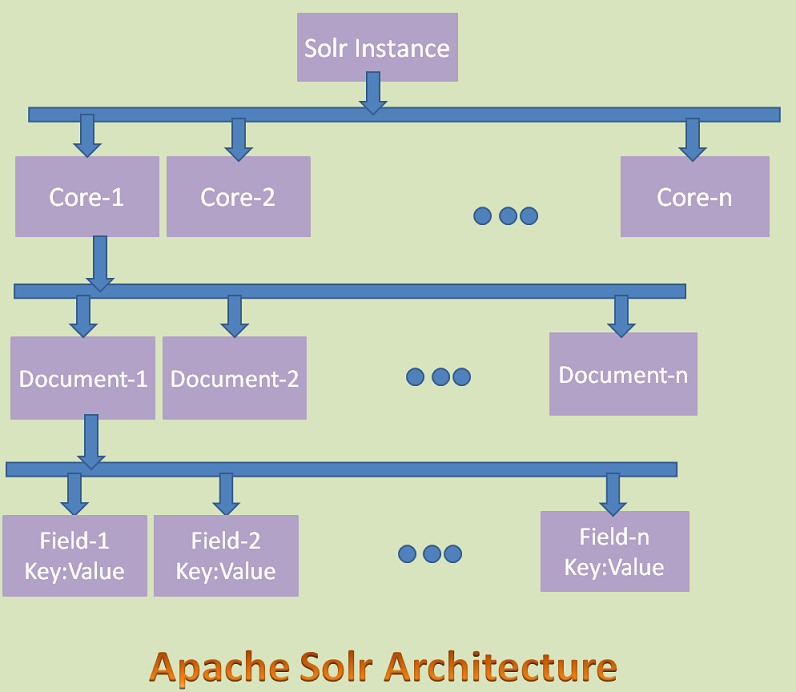
Apache Solr uses the following important configuration files:

a.solr.xml

b.solrconfig.xml

c.core.properties

1. **Indexing:** Indexing is a technique of adding Document’s content to Solr Index so that we can search them easily. Apache Solr uses Apache Lucene Inverted Index technique to Index it’s documents. That’s why Solr provides very fast searching feature.
2. **Document:** Document is a group of fields and their values. Documents are the basic unit of data we store in Apache Cores. One core can contain one or more Documents.
3. **Field:** a Field is actual data stored in a Document. It is a key & value pair. Key indicates the field name and value contains that Field data. One Document can contain one or more Fields. Apache Solr uses this Field data to index the Docuemnt Content.



Important points to remember:

* Apache Solr Standalone Architecture has only one Solr instance where as Solr Cloud Architecture have more instances.
* Apache Solr uses SolrCloud technology to support Cloud Architecture.
* Each Solr Instance can have zero or more Cores.
* Each core can contain zero or more Documents.
* Each Document can contain zero or more Fields.
* Each Filed contains a Key:Value pair. Key is the name of the Filed and Value is the data of the Field.

BASIC COMMANDS: (in cmd)

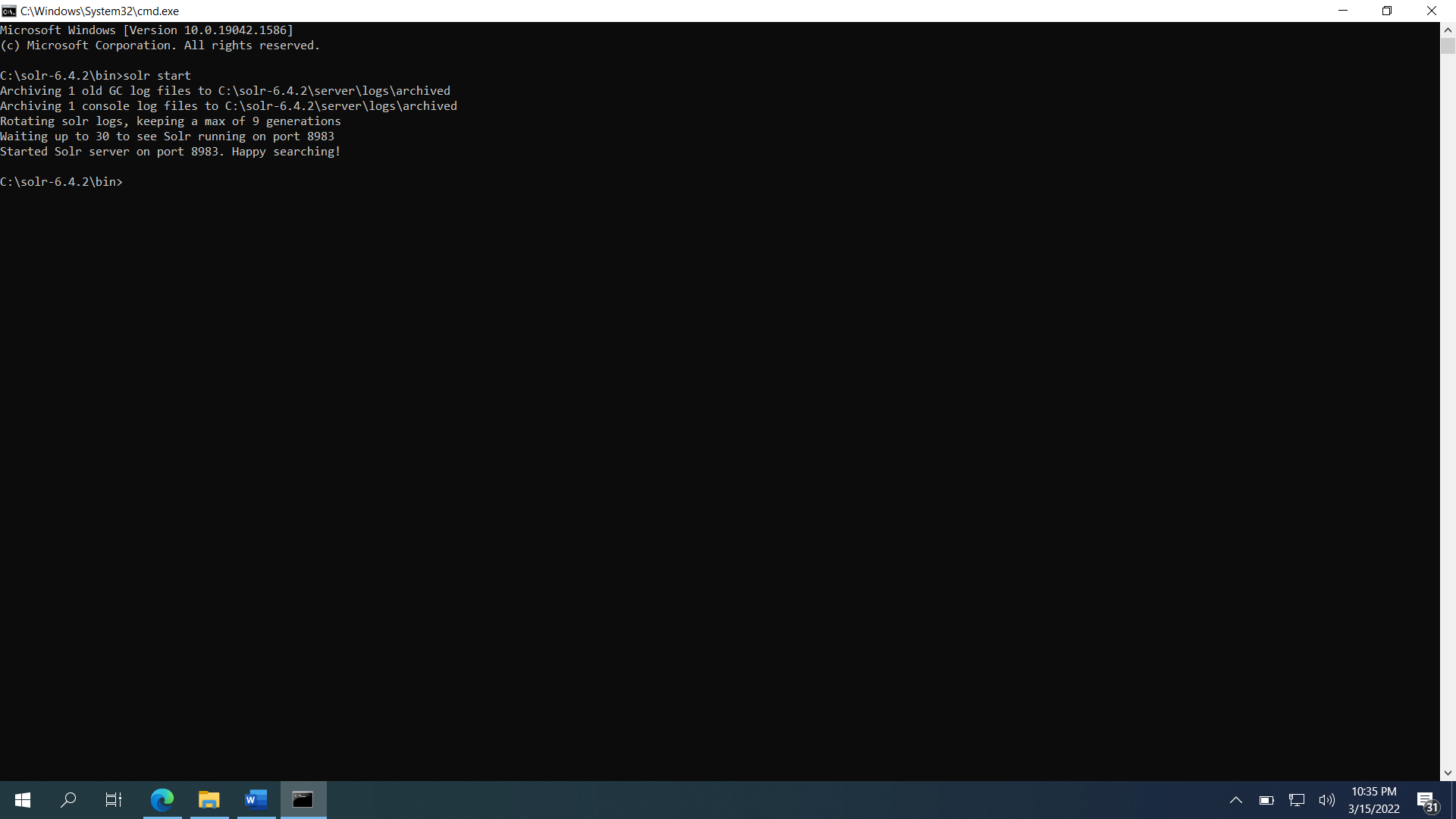
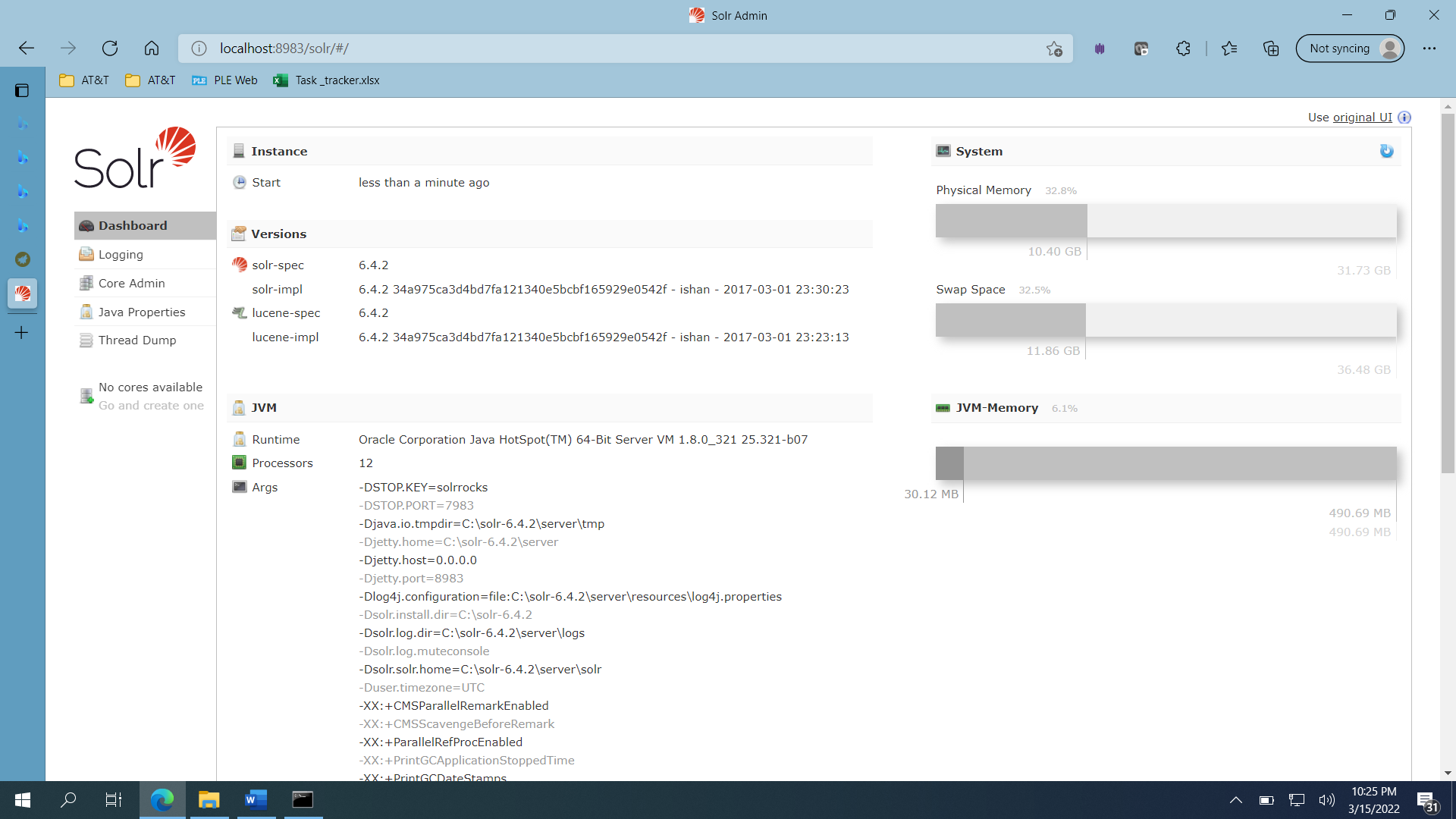
After installing Solr, browse to the **bin** folder in Solr home directory and start Solr using the following command.

* FOR START: $ solr start
* For stop: $ solr stop
* To create core: $ solr create -c <core name>
* For another port: $ solr start -p <port number>
* For help: $ solr -help

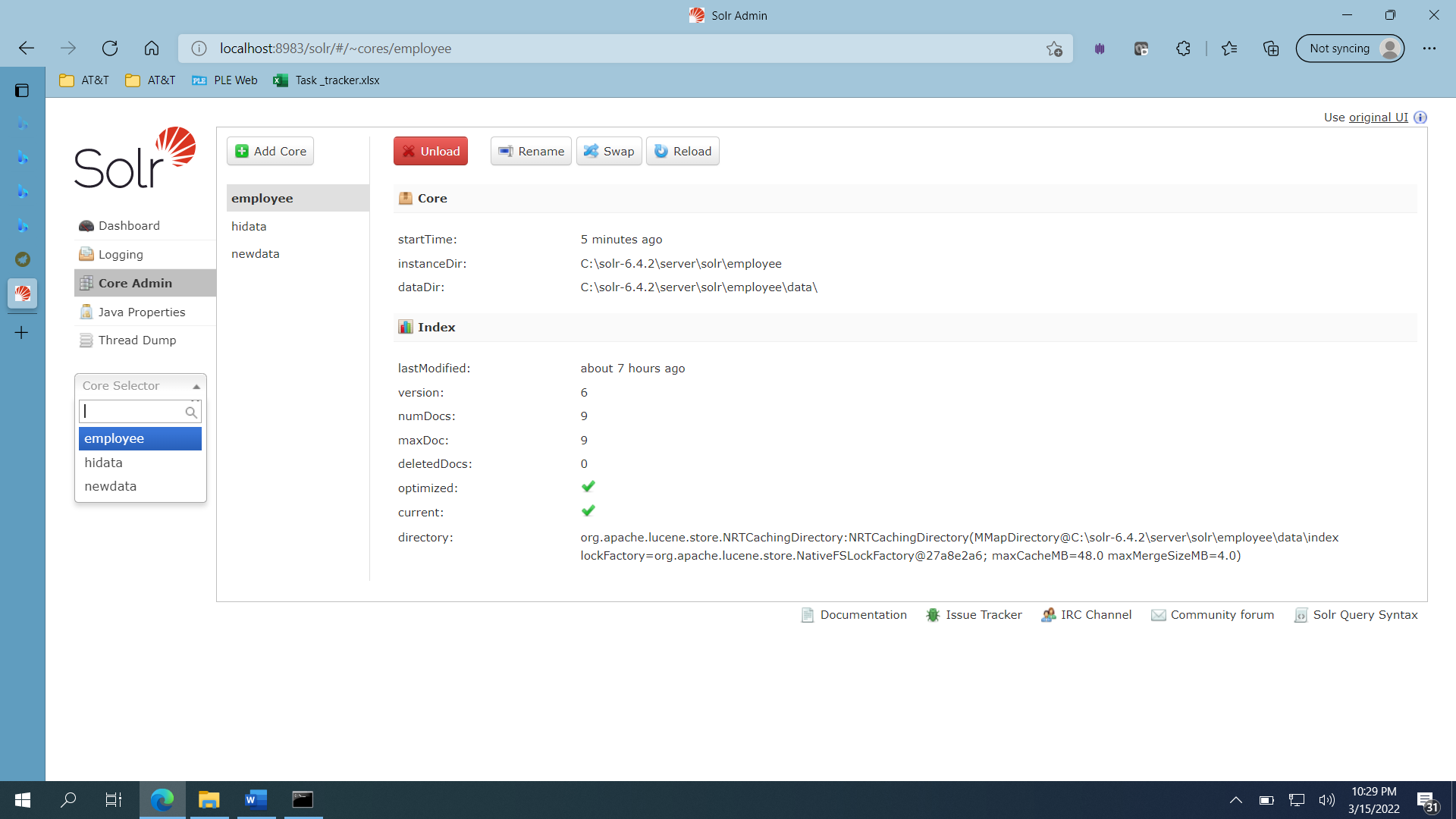
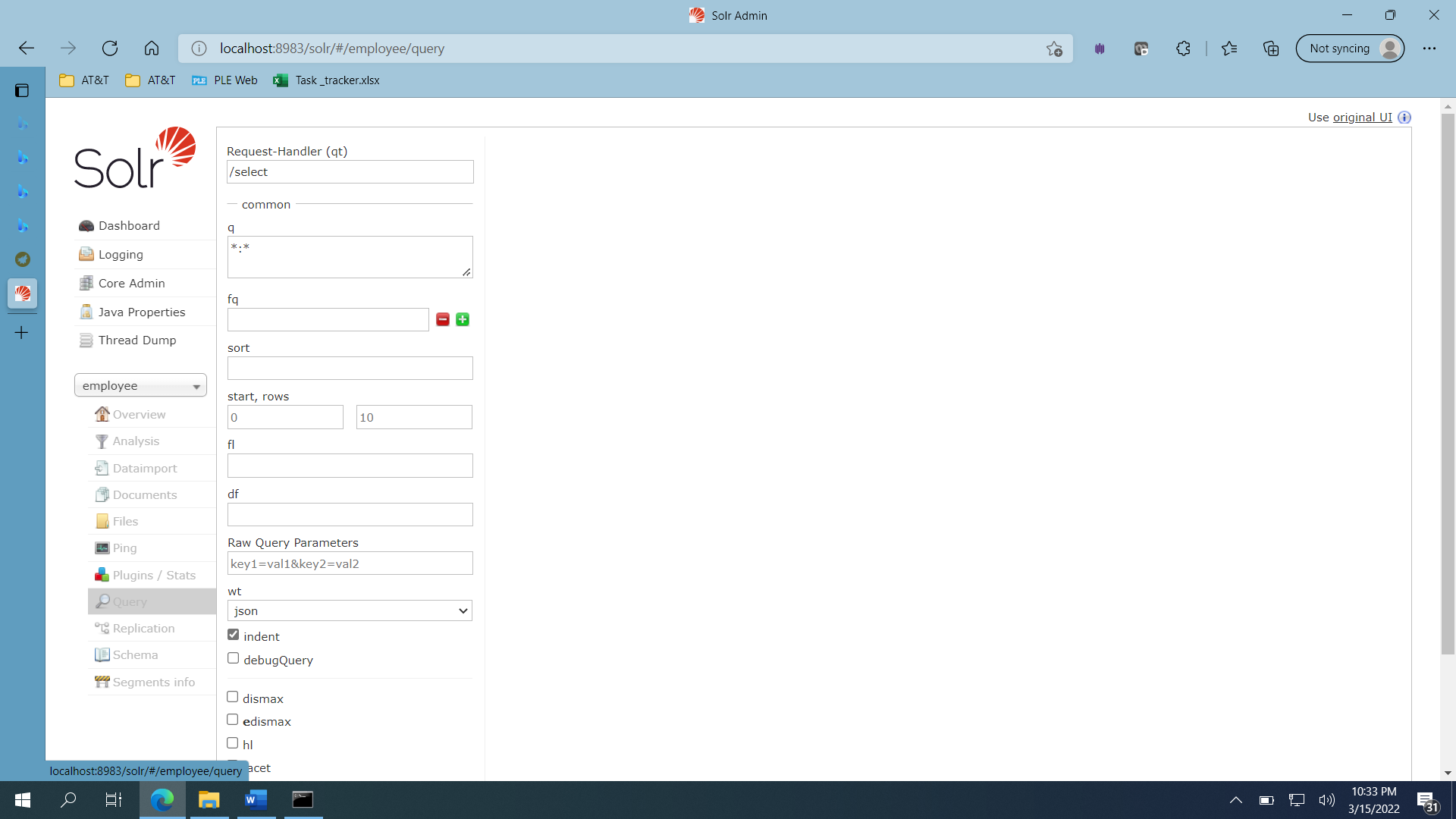
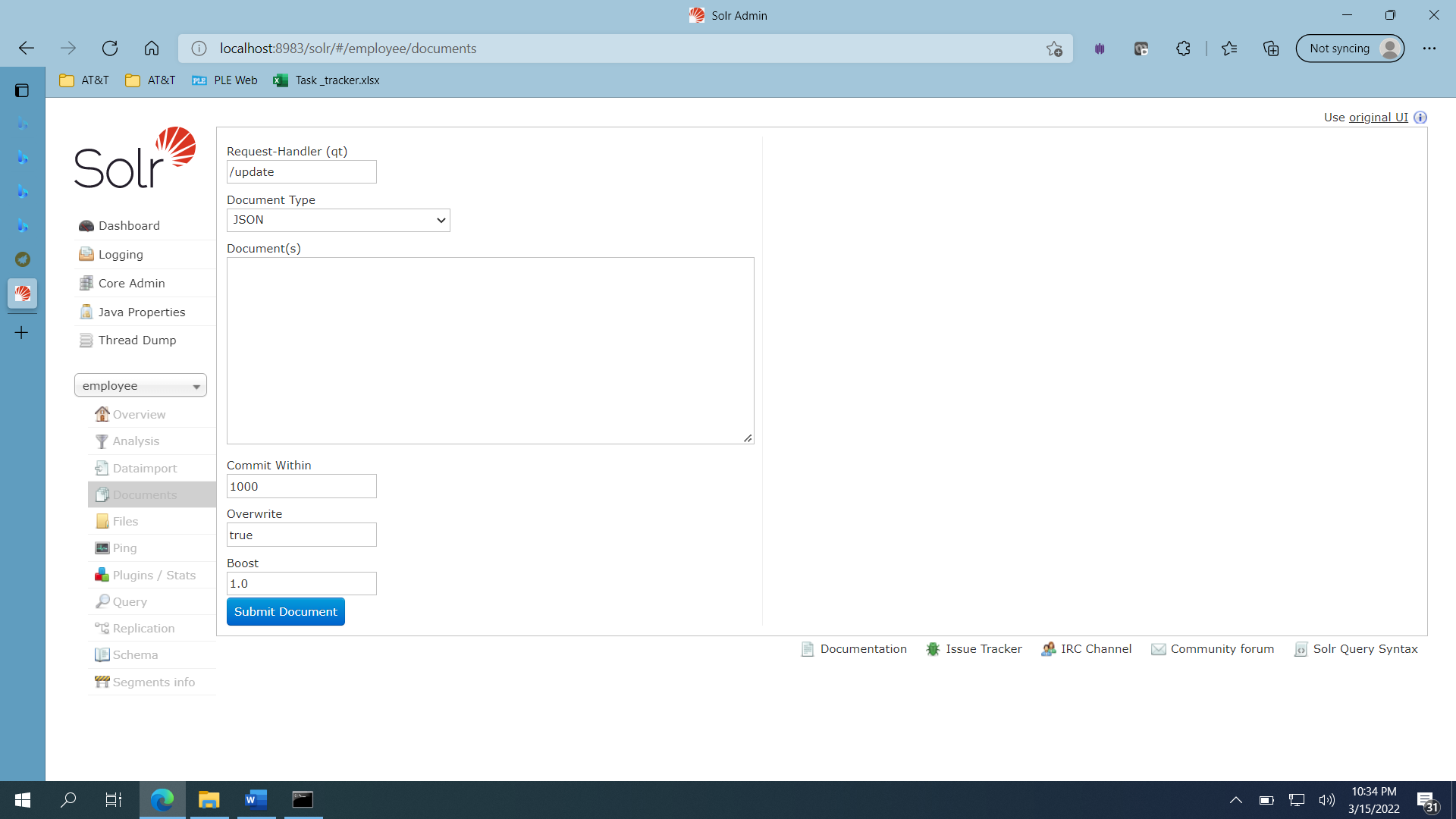
PROCESS:

* Install solr in pc, extract the files and navigate to bin folder and use cmd

[Index of /dist/lucene/solr/6.4.2 (apache.org)](https://archive.apache.org/dist/lucene/solr/6.4.2/) :link to install(solr 6.4.2 zip folder).

* 
* Type $ solr start , this starts solr admin page on port “localhost:8983”
* 
* Here to create a core inorder to insert the data goto cmd and type

“$ solr create -c <core name>”.

* 
* Click on the employee, now the data can retrieved by querying or data can be added in documents and so on..
* 
* 

PROS:

(+) Solar is very flexible and can be customized to meet your specific needs and requirements

(+) Solar is fast at returning search results

(+) Solar is pretty much an industry standard so there is a good chance other software programs have modules designed to work with it

(+) Solar allows you to build your own custom request handlers and allows you to import data from pretty much any data source.

(+) speed, scalability, flexibility.

CONS:

(-) It has a steep learning curve it's not intuitive how to configure it or customize it

(-) It doesn't include a web crawler indexing your website requires adding separate web crawler or using their API to add information to the index

(-) Debugging and troubleshooting query issues can be difficult task

(-) The admin UI is good but could be a bit more user friendly the field names are not very intuitive and require a learning curve

(-) It would be good to handle authentication natively in an enterprise scale. Currently it has wrapped it has to be wrapped by another service that does authentication

LATEST RELEASE:

* The most recent version is Solr 8.11.1
* The major upgrades are:

1.Support for multiple authentication schemes: two new authentication and authorization plugins provide support for configuring multiple authentication schemes

The two plugins are “MultiAuthPlugin” allows combining two or more authentication approaches such as JWT and basic authentication. And other plugin is “MultiAuthRuleBasedAuthorizationPlugin” used when MultiAuthPlugin is also in use and combines the various roles defined for all plugins to determine the proper role assignment for the user account.

2. ZooKeeper chroot: to create the ZooKeeper chroot at startup if it doesn’t exist already

ALTERNATIVES:

* Elasticsearch
* Google Cloud Search(new)
* SharePoint 2013