Hardware Datasheet

Hardware requirements for the SciBite ecosystem

TERMite

The TERMite engine underpins the SciBite Ecosystem, as it's responsible for the Named Entity Recognition (NER). It can be run in live 'Server' mode where it responds to markup requests in an HTTP environment, or it can be run in offline or 'Batch' mode where it is used to markup larger collections of documents on disk. Server mode powers SciNav, the user interface to TERMite and the RESTful webservices. Batch mode powers DOCstore. The amount of memory required by the TERMite engine is dependent on the number of dictionaries used. The standard set of 15 dictionaries is shown below, the full set contains about 50 dictionaries. It is possible to run TERMite in enhanced mode where the speed of indexing increases by 30%.

DOCstore

DOCstore is a searchable index of TERMite results. It's memory requirements also depend on the number of dictionaries used during the NER phase and on the text corpus size. DOCstore can be horizontally scaled, such that it's index and RAM needs can be split across different machines.

Additional Requirements

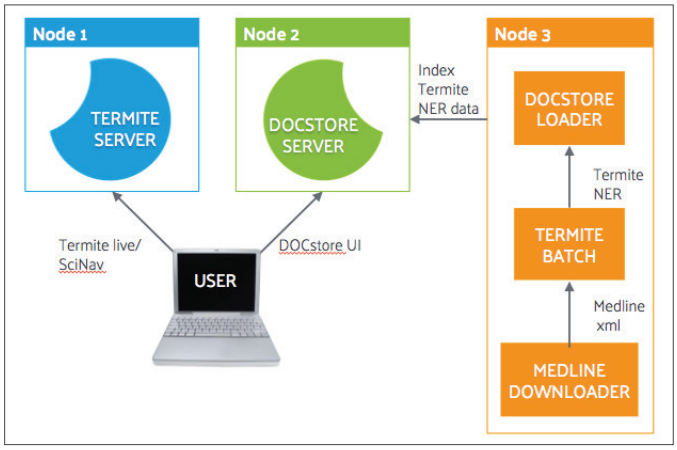
TERMite and DOCstore run as standalone java applications (Java 1.8+). No additional applications are required.

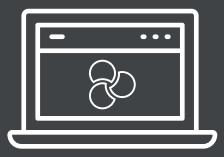
Standard Vocabulary Collection

Vocabularies included in the default library:

- **Gene/Protein**
- **Protein Type**
- **Drug**
- **Indication**
- **Orphan Disease**
- **Adverse Event**
- **Biomedical Concepts/Procedures**
- **Biomed Verb**
- **Biochemical**
- **Gene Ontology**
- **Human Endogenous Cells**
- **Anatomy**
- **Human Phenotype**
- **MicroRNA**
- **Single Nucleotide Polymorphism**

Example Configuration of Termite/DOCstore





SciBite
an ELSEVIER company

Minimum Hardware Specification – Medline Base Release

	DEFAULT LIBRARY		ALL VOCABULARIES	
	Memory (Gb)	CPUs	Memory (Gb)	CPUs
TERMite Batch	8	4	8	4
TERMite Batch (Enhanced)	8	4	16	4
TERMite Server	8	4	8	4
TERMite Server (Enhanced)	8	4	16	4
DOCstore Server	8	4	16	4
DOCstore Loader	8	4	8	4

Example Configurations for Standard Dictionary Set

Hardware Config	TERMite Batch	DOCstore Loader	DOCstore Server	TERMite Server
1	8Gb Node A	8Gb Node A	8Gb Node B	8Gb Node C
2	8Gb Node A	8Gb Node A	8Gb Node A (TERMite runs when DOCstore off)	8Gb Node B
3	32Gb Node A	32Gb Node A	32Gb Node A	32Gb Node A

Example Configurations for All Dictionary Set

Hardware Config	TERMite Batch	DOCstore Loader	DOCstore Server	TERMite Server
1	8GB Node A	8GB Node A	16Gb Node B	16GB Node C
2	16GB Node A	16GB Node A	16GB Node A (TERMite runs when DOCstore off)	16GB Node B
3	32GB Node A	32GB Node A	32GB Node A (TERMite runs when DOCstore off)	32GB Node A

About SciBite

SciBite's data-first, semantic analytics software is for those who want to innovate and get more from their data. At SciBite we believe data fuels discovery and we are leading the way with our pioneering infrastructure that combines the latest in machine learning with an ontology-led approach to unlock the value of scientific content. Supporting the world's leading scientific organisations with use-cases from discovery through to development, SciBite's suite of fast, flexible, deployable API technologies empower our customers, making it a critical component in scientific, data-led strategies. Contact us to find out how we can help you get more from your data.

To learn how SciBite can unlock the value of your data, speak to one of our experts today or email us at contact@scibite.com

Head Office:
SciBite Limited
BioData Innovation Centre
Wellcome Genome Campus
Hinxton, Cambridge CB10 1DR
United Kingdom

www.scibite.com
contact@scibite.com
LinkedIn: SciBite
Twitter: @SciBite
+44 (0)1223 786 129