



BOUDICA: THE AI WARRIOR QUEEN OF ENCRYPTED ANALYTICS.

An Introduction to OmniIndex's SLM AI Engine

PostgresBC enables the same analytics available on plaintext data to be performed on encrypted data.

Data is stored in the user's own blockchain and is protected from unwanted third-party eyes and Ransomware attacks.

This means customers can add their most sensitive and confidential data to collaboration, productivity, and analytics tools with no risk of exposure.

One of the core pieces of technology powering this Web3 data platform is OmniIndex's AI engine: Boudica.

WHAT DOES BOUDICA DO?	03
HOW DOES BOUDICA WORK?	04
BOUDICA IN ACTION	06
BENEFITS OF BOUDICA & SLMS	09

WHAT DOES BOUDICA DO?

Boudica is a Small Language Model (SLM) AI engine.

Its job is to determine what a user is requesting and to fulfill their wishes by working out the optimal response.

Some of Boudica's key tasks within PostgresBC are:

- Determine and derive the user's encryption key(s)
- Create the Queries to interrogate the blocks
- Process the encryption routines
- Manage the formatting of the output
- Process a Natural Language Understanding (NLU) input
- Determine context of a block
- Determine the sentiment of a block
- Collate the analysis to provide the required query to drive data analytics from the chain

HOW DOES BOUDICA WORK?

Boudica uses a range of dictionaries, lexicons, and thesauruses (referred to as 'ontologies') to work out an optimal response to any queries it's asked. These responses are then put through a separate probability matrix before being sent back to the user to determine the most appropriate response.

Boudica does this by replicating how the human mind processes questions and responses.

First, it needs to determine what is actually being requested. For example:

- Is it a question?
- Is it a statement?
- Is it a command?

Next, it needs to create an actionable request from that. To do this, it needs to:

1. Get answers.
2. Respond in a manner which matches the request.
3. Run the command.

To aid this, Boudica uses Natural Language Understanding (NLU) to break down the input into actionable commands.

Because each OmnilIndex customer has a different requirement when it comes to using their data, Boudica is supplied with a different ontology to match these specific needs. This means that the AI is able to provide specialist insights without the need for any expensive or lengthy training as it is able to immediately utilize the inputted information.

For example a medical organization would use a health ontology of medical definitions and terms while a financial institution would use financial ones.

What's more, the language of these dictionaries is determined by the customer's needs with Boudica able to work just as effectively in any language – so long as it is supplied with accurate dictionaries and ontologies.

The responses generated off the back of this NLU are then put through a separate probability matrix before being sent back to the user.

This probability engine uses a combination of language and vertical inputs to determine which response is top and which is bottom and scores them on a percentage of probability with '100%' being definite and '60%' being probable.

Anything less than 60% accuracy is dropped in order to enable a margin of error keeping the response on the correct side of probable.

BOUDICA IN ACTION

Health Innovation Hub (HIH) is a Life Science organization working with highly confidential medical and personal data.

Their goal is to use this customer data to give personalized and actionable health and lifestyle feedback via their App.

It is crucial for them to have a data platform that gives them complete protection from ransomware, while also enabling them to get insights from their regulated and sensitive data through AI and ML analytics.

As such, they use OmniIndex PostgresBS and Boudica.

The following is an example query Boudica might be asked by the team at HIH of their fully encrypted data:

“How many clients prone to gastric anomalies have signed up within the last 6 months?”

Boudica's process for answering this is:

1: NLU analyzes the content to determine if there is a question, statement, or command. It does this by looking for question words and phrases. In this case:

- 'How'
- 'How many'

2 : Next the NLU analyzes the content to look for any verbs and relevant words that give the context to those question words and phrases. In this case:

- 'Many'
- 'Have'
- 'Within'

3: Once done, the NLU looks for what item(s) the identified action needs to work on. In this case:

- 'Gastric'
- '6 months'

4: As items have been identified, the NLU has the basics to create a knowledge base and it can therefore look at the words that surround the item for more context and clarity.

To do this, it first identifies any such words, and then defines them to see if they are needed to clarify the command. In this case:

- 'Anomalies'
- '= irregular, not normal'

5: The NLU can now put together a statement based on the context of the processed question and the identified data the question is being asked of. In this case:

- 'Retrieve number of DNA_SEQUENCE_TYPE where signups have been created within 6 months.'

6: Boudica now checks its data store to see if any programmed commands align with the statement. If they do, then they will perform the given query.

7: The final step is to return the results to the user. This is only done once Boudica has checked that the required authorization and authentication is still in place.

BENEFITS OF BOUDICA

Efficient & Optimized:

SLMs do not require extensive training on huge pools of data in order to be used and can be adapted to offer specialist services in different languages and areas simply by changing their ontologies.

They also require less powerful hardware because they use less energy and are a substantially greener and cheaper option.

Accurate & Unbiased:

SLMs only work on small pools of controlled data that they are supplied with and are therefore far less likely to contain biases and inaccuracies than LLMs which learn from huge pools of varied data.

Boudica's use of multiple separate thesaurus models and its probability matrices also ensure that only the optimum response is given to a user.

Secure & Private:

Thanks to a combination of the Web3 data storage and fully homomorphic encryption that PostgresBC uses, Boudica is able to provide users insights on their fully encrypted data. This means at no stage is a user's private data at risk of exposure.

This is unique to OmniIndex thanks to our patented technology.

However speaking more broadly, SLMs are also a more secure and private option for AI analytics than LLMs because they never have to process data externally with no data leaving your infrastructure and protection and none of your data ever being used to train the AI model outside of your own uses.