Bloor

InDetail

InDetail Paper by Bloor
Authors **Philip Howard**
Publish date **August 2018**

SynerScope Ixiwa

In this sort of report, we normally like to provide a *"warts and all"* evaluation of the product under consideration. However, we cannot think of any relevant warts.

Executive summary

The idea behind SynerScope's products – Ixiwa and Iximeer – is that the industry has been more focused on the volume and velocity of big data and not enough on its variety. There is justification for this argument. There has been a lot of emphasis on scaling data lakes on the one hand and deploying stream processing on the other. However, these are essentially hardware issues. Of course, there is a software element but the advent of more powerful processors, the reduction in price of in-memory processing, and the use of graphical processing units all mean that the volume and velocity issues of big data can be readily resolved. However, variety is another matter. Certainly, there has been a focus on sensor data within Internet of Things environments, but sensor data is much more nearly structured, and can fairly easily be put into, say, a relational table format, than data that is in a Word document, or an image or an audio file.

So Ixiwa, the subject of this paper, and Iximeer, its companion product, are focused on and specialise in processing the variety of big data. This does not mean that they do not also process conventional, structured data, they do. And they have some unique capabilities for doing so. But it is especially when processing other forms of data that SynerScope has some significant competitive differentiators.

Fast Facts

SynerScope Ixiwa (which is from a Dutch dialect meaning *"I see something"*) might best be described as a data lake management product that covers everything from automated ingestion, through discovery and cataloguing to data preparation. Its companion product Iximeer (meaning *"I see a lot [more]"*) is an analysis and visualisation tool which may be used in conjunction with Ixiwa. While Ixiwa will often be deployed as a stand-alone solution, it can also be viewed as complementary to third party data cataloguing tools, which tend to focus on structured data only and/or have only limited unstructured capability.

Key Findings

Normally at this point we would provide some highlights of the product under examination and its key features. However, if you haven't seen Ixiwa or are not familiar with it, such a list would not make a lot of sense. We will therefore defer consideration of our key findings until after we have described the product. Since Ixiwa is highly visual we will do so as much by talking around screenshots as in general description.

"SynerScope Ixiwa (which is from a Dutch dialect meaning *"I see something"*) might best be described as a data lake management product that covers everything from automated ingestion, through discovery and cataloguing to data preparation."

The product

As we have noted, Ixiwa is a data lake (Hadoop and Spark-based) management product that ingests data automatically, collects metadata about the ingested data (automatically) and classifies that data for you. These are all good things, but it is how you explore the data that is most interesting, and we will focus on that to begin with, followed by an exploration of how Ixiwa supports this exploration. We will start by considering its approach to structured data and then move on to image and text processing.

Structured data

While you can view data sets that have been catalogued by Ixiwa in conventional tabular format if you want to, the more usual interface is illustrated in ***Figure 1***.

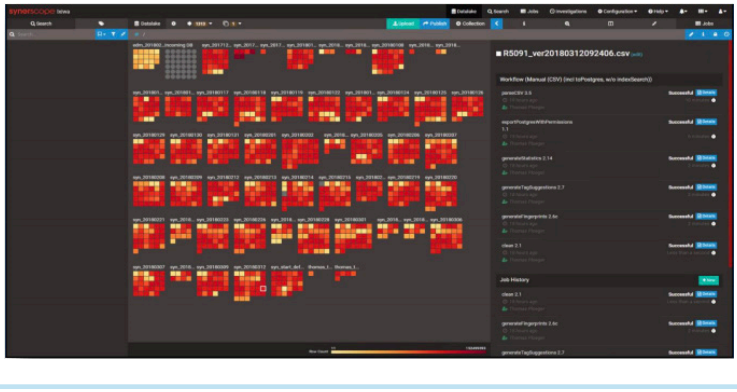

Figure 1 – Ixiwa's visual interface with heat mapping

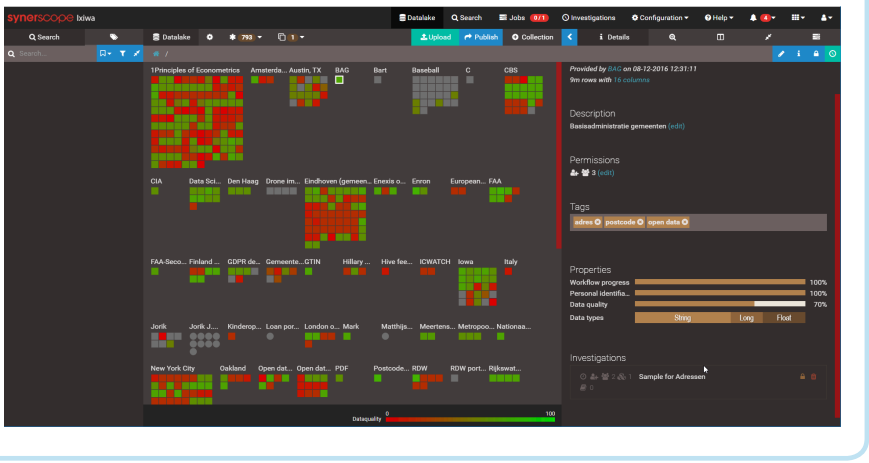

Figure 2 – An Ixiwa heat map showing data quality

This screenshot requires some explanation. Each of the shapes represents a data source and these source systems are heat mapped by table. In other words, each of the small squares represents a table and the redder the box the more it is of interest. Thus, if you are looking for customer data this representation would show you where the data that is of most interest resides. You can define filters that will further break down these boxes: you could look for a specific customer or look for customer by region, say. You can click on any particular box and details will be brought up to the right of the screen. Various options for these details are provided. For example, you can show profile data (graphically in a histogram), you can see where other users have tagged this data with a *"like"* flag or, perhaps more pertinently, where they have actually used this data in their own analyses.

Figure 2 shows a variant of the same type of screenshot, but this time showing data quality metrics.

Here green and red are used as classical traffic light colours. Note the scale bar at the bottom of the screenshot: in ***Figure 1*** it went from yellow (uninteresting) to red (very interesting) while it here goes from red (poor data quality) to green (good quality). On the right, the user has drilled down to the specific details with respect to a particular cell.

That's two options, but you can also explore the data in any way that it is classified. Thus, for example, you could bring up a similar heat map based on the sensitivity of the data (the product integrates with technologies such as Apache Atlas and Ranger). This might be based on PPI or PHI definitions or other standards that SynerScope supports, or you can create your own definitions. Other possible heat maps might be based around data size, the costs associated with particular data sets and so on. Other types of data can also be clustered and explored in this way. For example, this technique can be used for exploring sensor data.

Unstructured data

Figure 3 shows an Iximeer screenshot on top of Ixiwa related to traffic monitoring and vehicle identification. Arguably, the most interesting part of this is the image (or set of images) in the top right quadrant and we have blown this up in *Figure 4*.

What can be seen here are images that have been classified using an object detection algorithm, so that all the cars in the red boxes have a common characteristic (in this case: you can't read the number plate), as do the other coloured boxes.

In other words, Ixiwa can classify individual elements within a single image. Whereas auto-captioning, as offered by Google and others, could identify that these images show cars, it wouldn't be able to classify them in the way that Ixiwa can. For example, if you want to differentiate between a Mercedes and an Audi then Ixiwa can potentially do that, while auto-captioning cannot. However, you can use auto-captioning in conjunction with Ixiwa if you want to, but often pixel sorting will be more useful, especially if you project the Ixiwa results into Iximeer.

We won't show it, but Ixiwa does something very similar with text, for which it has a *"text sorter"*. For audio files it uses voice to text conversion first and you can use Ixiwa in conjunction with products that do feature extraction from text, if that is appropriate.

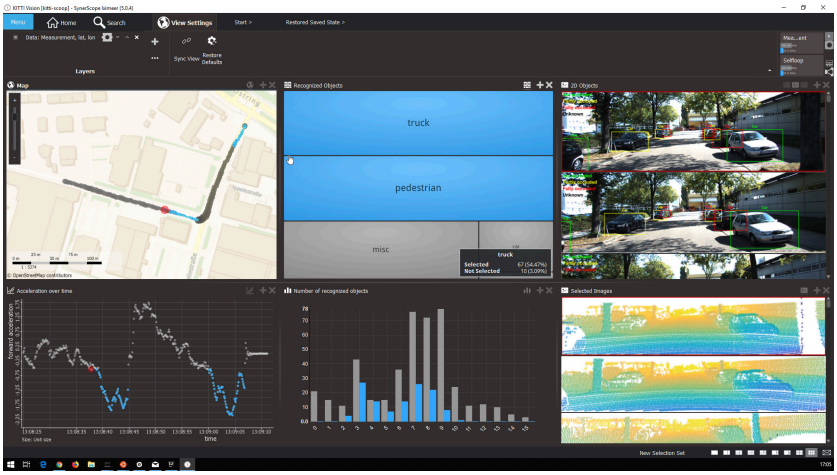

Figure 3 – A traffic control dashboard using Iximeer on top of Ixiwa

Figure 4 – Auto-correlation of images using Ixiwa

Under the hood

The preceding discussion gave an overview of what Ixiwa can do in terms of how you can explore the data in your data lake. However, the CEO of SynerScope recently blogged about potential users that they *"think it is too good to be true. For example, people do not believe that we can tag, analyze, categorize and match content of a data lake at a record level without having pre-created metadata. Other people struggle with believing that we can match structured and unstructured data without pre-defined search logic. They question if our patented many-to-many correlator truly can group content, elements or objects in logical clusters. Or they wonder if it is true that business users can easily add and use data-sets in the lake without IT support."*

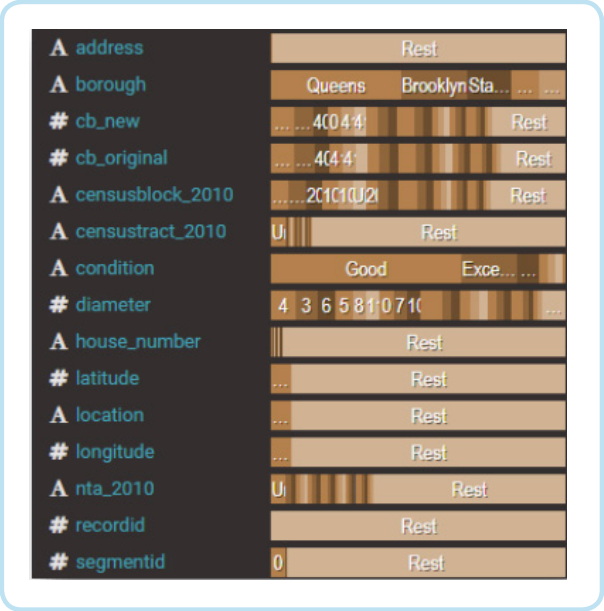

Figure 5 – Data profiling in Ixiwa

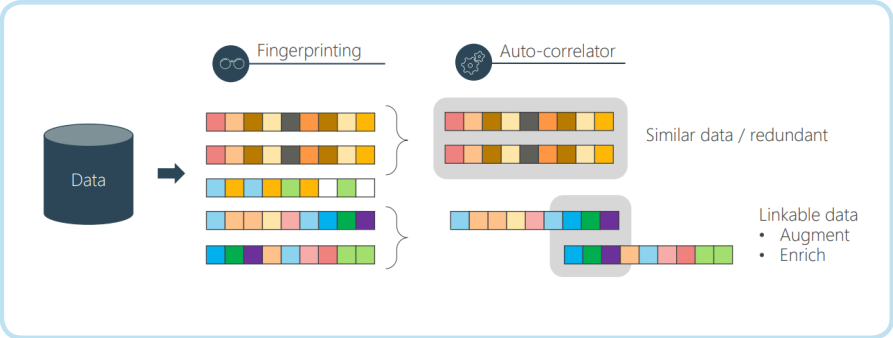

Figure 6 – Auto-correlation in Ixiwa

So, it is important to explain not just what Ixiwa does but how it does it. To begin with, it starts with the automated collection of metadata. There are three different types of such metadata that it collects, the first of which is *"provided metadata"*. That is, the metadata that accompanies the source system such as descriptions, field names and defined keys. This can be explored in the right-hand panel in ***Figures 1*** and ***2***, with more detail being shown in ***Figure 5*** where *"rest"* means that all the (other values) have an occurrence of 1. You can drill down further into these details if required, to see this information presented as a histogram.

The second type of metadata that Ixiwa supports is what the company refers to as *"inferred metadata"*. This would be metadata that is inferred from scanning the data itself. For this, Ixiwa has embedded machine/deep learning (based on TensorFlow) so that its inferences, and the tagging that it generates will improve over time. To illustrate how this works to support the sort of heat map clustering we discussed previously, consider ***Figure 6***. Here, Ixiwa has discovered metadata about five different sets of data and then, using its auto-correlator has discovered that the top two are either duplicates or at least have the same format, while the bottom two have some similarities that would suggest that it might be useful to join these two datasets for analysis purposes.

Finally, the third type of metadata that Ixiwa collects is what the company calls *"attributed data"*. This is where you can bring in external metadata about such things as data quality, cost and usage information, as shown in ***Figure 7***. Full data and user level audit trails are captured, to support these sorts of metrics.

Finally, we should comment on the Pixel and Text Sorters. These support exactly the same sort of clustering as the auto-correlation above: extracting similarities about, say, images, then inferring clusters (same colour, same shape, same maker of car, facial recognition and so forth) and then tagging those images appropriately. Again, these sorters use TensorFlow to enable this sort of processing. An example of the use of pixel sorting is shown in ***Figure 8***. Here, the image at top left shows meter locations with the blue area being the area of interest. In the top middle are selected photos of meters within the area of interest while top right is using the pixel sorter where each dot represents an individual photo, with different islands containing different photo contents. Along the bottom, from left to right, is the distribution of the year of construction of the different houses in which the meters are located, a data linkage diagram (so the right parts and the right engineer can be assigned to the right meter) and a differential analysis of what is typical within the area of interest which, in this case, is properties with porches built since 1970 (highlighted in blue).

Figure 7 – Discovery based on attributed data

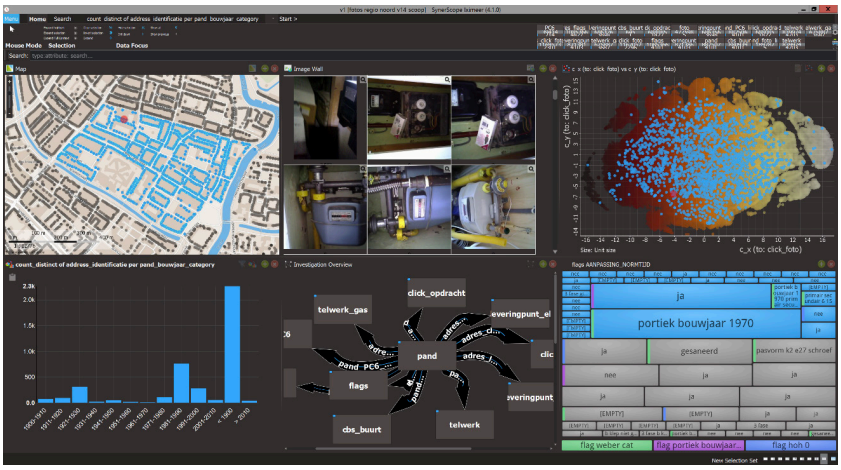

Figure 8 – A dashboard illustrating pixel sorting

The company

SynerScope was founded as a spin-off from research conducted at Eindhoven University. It is headquartered in the Netherlands but also has an office in California. It has a number of technical partnerships, most notably with Amazon, Microsoft (Azure) and Google for cloud services and with HortonWorks, IBM and Nvidia. The last two of these are particularly significant. In the case of IBM because Ixiwa can run on Power platforms as well as x86 (Dell is also a partner) and because it can leverage Nvidia GPUs (which are built into the IBM Power AI platform) to accelerate performance. The relevance of this is that SynerScope's view is that in large data lakes there is a potential scalability issue, especially (but not only) when you are doing complex things like image processing, so you need the extra scalability that a powerful hardware platform can provide.

The company also has a number of business and academic partners.

SynerScope's view is that in large data lakes there is a potential scalability issue, especially when you are doing complex things like image processing, so you need the extra performance and scalability that a platform like Power AI can provide.

Key findings revisited

Having now provided an overview of what Ixiwa does and how it works, we are in a position to reflect on the product's capabilities more generally. In the opinion of Bloor Research, the following represent the key features of SynerScope Ixiwa:

- Ixiwa is one the most visual products it has ever been our privilege to evaluate.
- Both the structured and unstructured capabilities built into the product are impressive.
- The use of machine/deep learning means that Ixiwa really can do what the company claims for it.
- In addition to auto-tagging there are also facilities for users to *"like"* a particular dataset. Further, because everything is logged you can distinguish between what is liked and what is actually used for analytic purposes.
- While the company has partnerships with platform providers we would like to see it partner with complementary software providers.
- With its support for IBM Power (as well as x86) and Nvidia GPUs Ixiwa should have no scalability or performance issues.

The bottom line

The difficulty SynerScope has is that there is nothing else quite like Ixiwa on the market: it could be used as a data cataloguing tool but the company, rightly, does not want to be seen as competing with the likes of Waterline and Alation, not least because in many ways it is complementary to those products. This is especially true if you are interested in text, audio or image processing, for which these vendors have very limited or no capability. This is why we have characterised Ixiwa as a data lake management tool though even though that suggests competition to some other vendors despite the fact that *"co-opetition"* would be more appropriate.

Leaving all that aside, we are truly impressed by Ixiwa. In this sort of report, we normally like to provide a *"warts and all"* evaluation of the product under consideration. However, we cannot think of any relevant warts. If the sort of features outlined here fall into your sphere of interest, then Ixiwa is a must-see product.

"Ixiwa is one the most visual products it has ever been our privilege to evaluate."

FURTHER INFORMATION

Further information about this subject is available from:
www.BloorResearch.com/update/2383

About the author

PHILIP HOWARD
Research Director / Information Management

Philip started in the computer industry way back in 1973 and has variously worked as a systems analyst, programmer and salesperson, as well as in marketing and product management, for a variety of companies including GEC Marconi, GPT, Philips Data Systems, Raytheon and NCR.

After a quarter of a century of not being his own boss Philip set up his own company in 1992 and his first client was Bloor Research (then ButlerBloor), with Philip working for the company as an associate analyst. His relationship with Bloor Research has continued since that time and he is now Research Director, focused on Information Management.

Information management includes anything that refers to the management, movement, governance and storage of data, as well as access to and analysis of that data. It involves diverse technologies that include (but are not limited to) databases and data warehousing, data integration, data quality, master data management, data governance, data migration, metadata management, and data preparation and analytics.

In addition to the numerous reports Philip has written on behalf of Bloor Research, Philip also contributes regularly to ***IT-Director.com*** and ***IT-Analysis.com*** and was previously editor of both ***Application Development News*** and ***Operating System News*** on behalf of Cambridge Market Intelligence (CMI). He has also contributed to various magazines and written a number of reports published by companies such as CMI and The Financial Times. Philip speaks regularly at conferences and other events throughout Europe and North America.

Away from work, Philip's primary leisure activities are canal boats, skiing, playing Bridge (at which he is a Life Master), and dining out.

Bloor overview

Technology is enabling rapid business evolution. The opportunities are immense but if you do not adapt then you will not survive. So in the age of Mutable business Evolution is Essential to your success.

We'll show you the future and help you deliver it.

Bloor brings fresh technological thinking to help you navigate complex business situations, converting challenges into new opportunities for real growth, profitability and impact.

We provide actionable strategic insight through our innovative independent technology research, advisory and consulting services. We assist companies throughout their transformation journeys to stay relevant, bringing fresh thinking to complex business situations and turning challenges into new opportunities for real growth and profitability.

For over 25 years, Bloor has assisted companies to intelligently evolve: by embracing technology to adjust their strategies and achieve the best possible outcomes. At Bloor, we will help you challenge assumptions to consistently improve and succeed.

Copyright and disclaimer

Bloor Research International Ltd
20–22 Wenlock Road
LONDON N1 7GU
United Kingdom

Tel: **+44 (0)20 7043 9750**
Web: **www.Bloorresearch.com**
email: **info@Bloor.eu**