POLY//ATICA

BUSINESS INTELLIGENCE

A JOURNEY TO UNLIMITED SCALE

13 SOHO SQUARE, LONDON, W1D 3QF
T +44 020 3468 1974 E INFO@POLYMATICA.COM

BUSINESS INTELLIGENCE - A JOURNEY TO UNLIMITED SCALE

Polymatica has an unusual product development history – seeing a gap in the market for a Business Intelligence tool able to handle large volumes of data, Polymatica set out to create one from scratch. From day one, all technological decisions were made with unlimited scaling in mind. After 7 years of success stories, we are confident our planning has paid off.

Linux as an OS

One of the first decisions with any new software is which operating system to support. Planning for high performance optimisation, Linux was an obvious answer. The way it handles memory allocation makes it many times faster than the other server OS options. Another advantage is the fact that it is open source, reducing the cost of implementing Polymatica.

Proprietary MultiSpheres

Creating a data structure fit for quick analytical processing was the next step. Polymatica's MultiSphere was born. The end result was to allow access to any node of the data containing more than 10 billion records within seconds.

The following pillars were always kept at the heart of development:

Multidimensional conceptual view

The user will view their data, which is multidimensional in nature, so every data point, such as transaction price, timestamp and a geographical region will all have a

number of dimensions associated with it. This multidimensional data views allow for intuitive data manipulation including slicing and dicing in Polymatica.

SHOP_DATE week			15			
SHOP_DATE week day			Monday		Tuesday	
SHOP_DATE hour						
CUST_CODE	PROD_CODE	BASKET_ID	Σ QUANTITY	Σ SPEND	Σ QUANTITY	Σ SPEND
CUST0000000040			6	8.43	0	0.00
CUST0000000068			4	8.26	0	0.00
CUST0000000091			2	2.70	5	20.20
CUST0000000094			0	0.00	0	0.00
CUST0000000103	901539		1	2.95	0	0.00
	900895		5	5.65	0	0.00
	901057		1	1.28	0	0.00
	901824		1	0.73	0	0.00
	903872		1	1.22	0	0.00
	903517		1	3.59	0	0.00
	Total		10	15.42	0	0.00
CUST0000000106			0	0.00	0	0.00

Figure 1: Multidimensional conceptual data view allows for intuitive data manipulation. Views and hierarchies can be created in real time by dragging and dropping dimensions into the left (CUST_CODE, PROD_CODE, BASKET_ID) or top (week, week day, hour) pane.

Transparency and accessibility

The end user should not be concerned about the details of data access or conversions. This means Polymatica is able to connect to any structured data source with an available ODBC connector, with support for full or incremental data updates. The end user can manipulate the data without needing to understand how it got there.

Consistent reporting performance

Performance should not degrade as the number of dimensions in the model increases (i.e. as the data goes from “skinny” to “fat”). Performance should also not degrade as more data rows are added. Polymatica can handle very different

datasets, ranging from "skinny" and short (few columns and few rows) to "fat" and long (many columns and many rows).

Client/server architecture

The product should have a client/server architecture, and the server component must allow for different clients to be attached with minimal effort. Polymatica's functionality is fully modular, making adding new features quick and easy. Access to the data structure is done through open REST API commands, so there are no restrictions on creating your own visualisations or connecting to existing workflows.

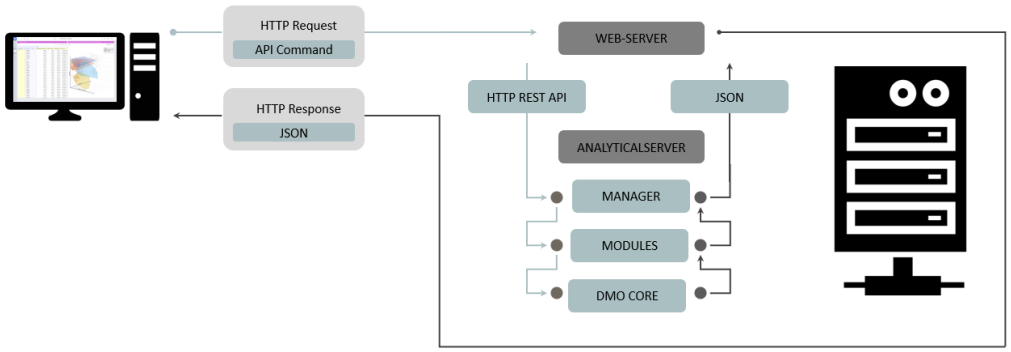

Figure 2: Polymatica's client/server architecture works through REST API commands. All the APIs are open, and all the functionality outside of the MultiSphere CORE is modular, so any part of the system can be integrated with

Generic dimensionality

The system should not be limited by a number of allowed dimensions and should not be biased toward any particular dimension. Polymatica has no pre-defined hierarchies, and all dimensions are treated as equal. There are no limits on the number of dimensions a data structure can have, nor on the depth of hierarchies the user can create and analyse. Any calculation can be performed across all dimensions.

Dynamic sparse-matrix handling

A sparse matrix will have cells that contain no data, and the system should be able to handle these efficiently. Polymatica's data structure compresses the original dataset up to 12 times, which is in part thanks to its excellent handling of sparse matrices.

Multiuser support

The system must support multiple concurrent users, including their individual views or slices of a common database. Polymatica allows its architecture to scale for support of unlimited concurrent users and handles user based access rights and restrictions.

Intuitive data manipulation

Users shouldn't have to use menus or perform complex multiple step operations when an intuitive drag and drop action will do. Polymatica's design is inspired by the study of the human brain. Having done research on what people find most intuitive when handling and manipulating data, Polymatica was designed to be intuitive. It has no need for scripting.

Flexible reporting

Users should be able to print and share the reports they need, and any changes to the underlying model should be automatically reflected in these reports. Polymatica allows you to save and share any part of a user's analysis.

Unlimited dimensional and aggregation levels

There are no limits on hierarchical levels that the user can create on the fly in Polymatica. No results are pre-aggregated.

Data can be loaded into Polymatica from a source database via full or incremental updates. The MultiSphere is compressed up to 12 times and is stored on an Solid State Drive. Analytical processing takes place in memory, where only the relevant

dimensions are loaded at one time. The MultiSphere is not finalised upon creation, meaning that new data can be added easily, without rebuilding the whole structure.

GPU acceleration

Seeing that the growth in data is heavily outpacing the growth of CPU processing power, Polymatica saw the immediate need to utilise GPU acceleration. While very good at handling sequential tasks, CPUs are not nearly as good at parallelism as GPUs. This is particularly important when considering typical BI tasks:

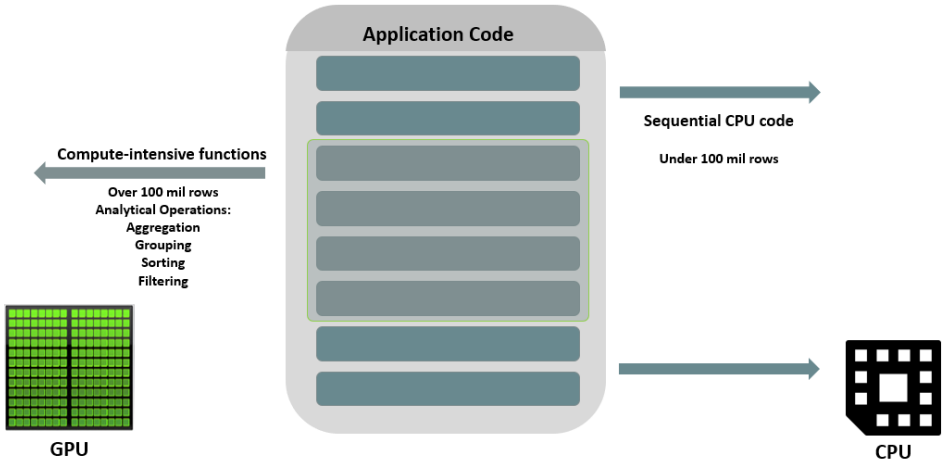

Figure 3: Due to the large number of cores, GPUs are much faster than CPUs at processing compute-intensive functions

Polymatica's server has an in-built algorithm to intelligently distribute tasks between available CPUs and GPUs, resulting in the fastest possible performance for a given

task. This becomes more relevant as the data scales, processing up to 40 times faster with GPU accelerated technology with datasets over 1 billion rows.

GPU acceleration advantages increase as the datasets grow larger. For example, a CPU architecture can comfortably handle datasets of under 100 million rows but processing a larger dataset will be done faster using GPUs. When we start talking about datasets in the order of billions of rows, GPUs are the obvious answer, as CPUs simply cannot handle the processing load in a reasonable time frame.

Polymatica has partnered with Nvidia, and runs extensive optimisation exercises each time a new Nvidia GPU is released.

Unlimited horizontal scaling

Having the most powerful server will still only allow one to process so much data. For example, an Intel Power8 machine running Polymatica with the following specifications, will handle up to 4 billion rows of data:

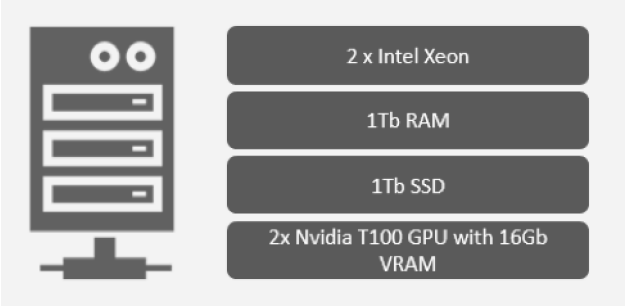

Figure 4: A Power 8 server can handle processing of up to 4 billion rows of data

Not wanting to be limited by even the best hardware, Polymatica can be comfortably distributed across multiple servers. The problem of scaling is two-fold: scaling for a large number of users (distributed server architecture) and scaling for large data sets (distributed data structures), both of which are supported.

Distributed servers

A single Polymatica server can handle approximately 100-150 active users (i.e. users who are actively manipulating data, not just viewing existing reports). If more users require access to the same data structure, the server can be replicated, and Polymatica will assess server loads and route a user accordingly. This configuration can also be used as a failover, in case one of the servers is down.

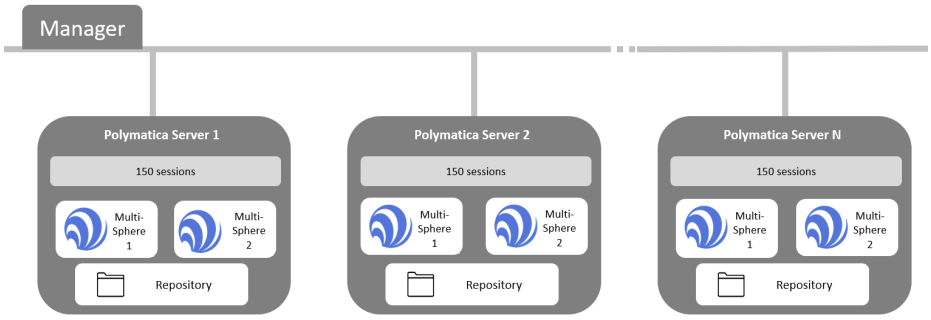

Figure 5: Distributed server architecture allows for scaling for many users

Distributed data structures

For very large datasets, the data structure can be distributed across multiple servers. The data can be split by a chosen dimension – usually by geography or time. If kept in the same place, the servers need to be connected by Infiniband or other fast networks for increased communication speeds. In the case of a distributed data structure architecture, if an analytical processing request is made requiring data from both servers, each will process the analysis on its respective dataset, and the results will be merged before being sent back to the user.

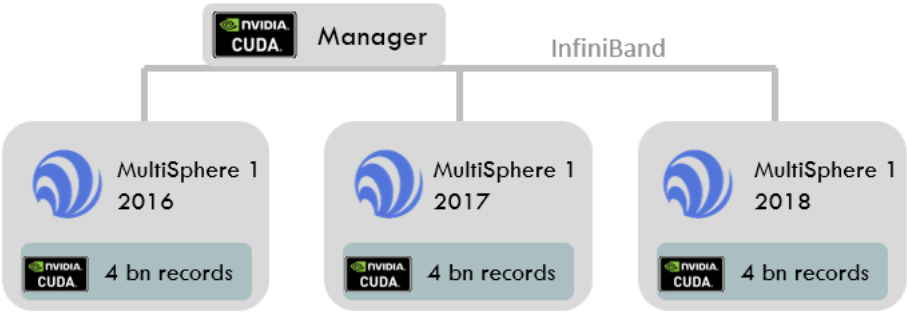

Figure 6 Distributed MultiSphere architecture allows for scaling for large datasets

Distributed structures can also be hosted in different locations. For example, a common data structure can be created across the UK, and split by geography, so that only the relevant region data is stored on the local server. This was, Scottish data can be stored on a server in Scotland, English data can be stored on a server in England, but it is still possible to have both comprise a single data structure.

SUMMARY

Polymatica is a versatile BI tool that allows unrestricted access to a full dataset, however large. Developed from scratch for performance, scale and ease of use, Polymatica offers a variety of ways to scale your analysis as your analytical needs increase. From data exploration to complex Data Science tasks like customer segmentation, association rules and forecasting, Polymatica covers all the needs of a Citizen Data Scientist.

About us

Polymatica is a data science company, pioneering new and accessible ways for global businesses to understand and use their data. The Polymatica platform uses embedded intelligence to deliver results and insights, at incredible scale and speed. Focused on business users throughout an organisation, Polymatica offers intuitive data analysis software, assisted decision-making solutions and visual storytelling tools, as well as consulting services and training designed to build the expertise of our clients.

For more information on Polymatica, please visit www.polymatica.com or call +44 (0) 20 3468 1974.