# Enterprise Information Architecture for Gen AI and Machine Learning

*Solix Common Data Platform –*
*Operating System for the Enterprise*

The SOLIXCloud Solution Architecture is an Enterprise Information Architecture. Architectures of this kind can be viewed as an *"Enterprise Operating System"* that orchestrates a data fabric, a multi-cloud IT environment and DevOps management software, serving a diverse user population with applications and services. SOLIXCloud's newly released AI products, Solix ML and Solix GPT, illustrate this as they are in effect applications that can be deployed within the EIA. Aside from their impressive functionality, among the benefits they bring is that they are fully integrated within the corporate computing environment.

# Executive summary

Generative AI or Gen AI caught the imagination of the web-using public at the beginning of 2023, racking up millions of users and attracting the attention of CIOs everywhere. In this paper, we describe and discuss the nature of Enterprise Information Architecture (EIA) and the data fabric requirements for Gen AI and machine learning (ML) using the Solix Common Data Platform (CDP) as a reference point.

In doing so, we conclude that a well-constructed Enterprise Information Architecture can be thought of as being an "Enterprise Operating System" organised as a data fabric, a multi-cloud IT environment providing the computer resources, DevOps to maintain and manage all the software and applications, and a large population of users running the applications and consuming the services.

Having discussed this idea, we then examine the recent additions to the SOLIXCloud Enterprise AI product set, Solix ML and Solix GPT, which have been built to support the machine learning and generative AI lifecycle within the Enterprise IT Environment.

We conclude that organisations will be well-served if they implement ML and Gen AI, as Solix has provided for in these two new software environments. On the one hand, Solix is doing a lot of the heavy lifting by integrating these two important AI capabilities. On the other hand, Solix provides an environment that is fully instrumented to accommodate the needs of enterprise application development in these areas.

In our view, organisations will be well advised to examine the capabilities Solix provides here when considering their data and AI strategy.

# The Foundation of Enterprise Information Architecture

The primary goal of this paper is to discuss the coherent adoption of both AI and Gen AI within the context of a functional Enterprise Information Architecture (EAI). Our discussion of EAI will be based on the capabilities provided by the Solix CDP Solution Architecture. This is illustrated in *Figure 1* below.

Among other things, Solix CDP establishes a fully-fledged Data Fabric, embodying an enterprise data lake and a comprehensive archiving capability. It also caters to Consumer Data Privacy, Content Services (for workgroup content data), and an Enterprise AI capability.

As illustrated, the Solix Common Data Platform can be integrated with a Multi-Cloud environment to provide a genuine EIA. There are few, if any, platforms that can compare.

(The broad capabilities of this whole ensemble are examined in depth in a Bloor Research paper entitled *Data Fabric and the Future of Data Management*. Readers who wish to ponder it in greater detail are advised to read that paper.)

Here, we will refer to the Solix product portfolio components as and when the need arises as we discuss the architectural implications of the corporate adoption of AI and Gen AI – which Solix has released as Early Access in their recent Solix CDP 3.0.11 product release.
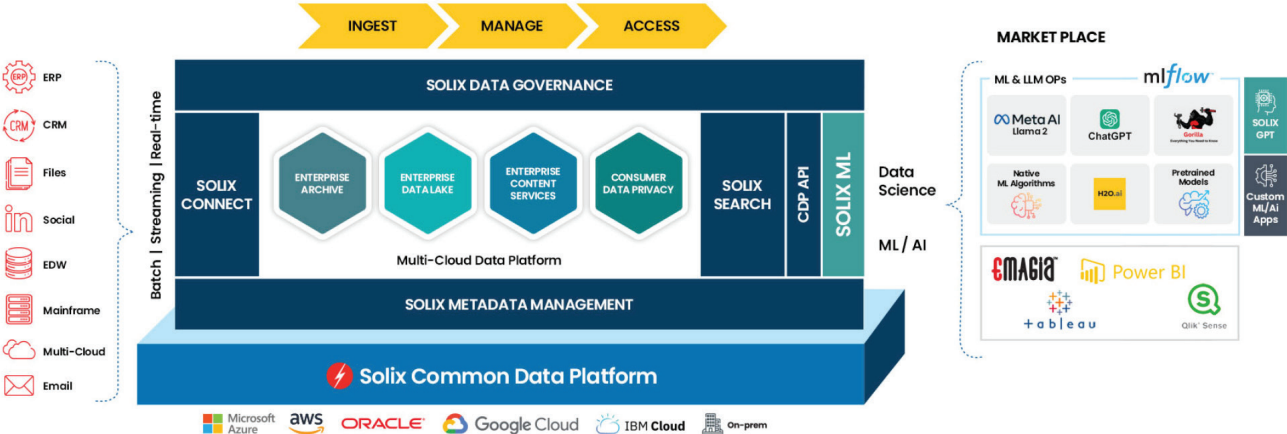


*Figure 1 – SOLIXCloud Solution Architecture*

# A Model of Enterprise Computing and EAI

We can discuss EAI using Bloor's Simple Model of IT Components and their relationships, shown below in *Figure 2*.

The diagram illustrates that for any application of IT, there are always four components: **User, Software, Data, and IT hardware.** An example might be sending an SMS message on a cell phone or accessing a website from a PC. So, the central part of *Figure 2* illustrates this ubiquitous situation: some user is running software to manipulate data using some device. And naturally, all the relationships indicated by the double-headed arrows exist.

It is important to note that, even in the case of a simple application, the situation is not static. The hardware can get upgraded somehow (one iPhone model replaces another, say), more data may gradually accumulate, the user may find new uses for the application, and the software itself may be upgraded. Put simply, the application needs to be maintained over time for it to continue providing the expected service.

The most important aspect of this model is that it scales. We can use a similar model to represent a system with many users, many different applications, and perhaps a database, all distributed between the cloud and local user devices. The diagram can be applied to whole business processes, multiple business processes, and enterprise IT in total.

When we consider the whole of Enterprise IT, the four terms are illustrated in the diagram, they are:

### Global User Population:
The user population will consist not only of staff but also customers, agents, and other third parties, and computer processes that are proxies for users.

### Data Fabric:
The data fabric is a data management architecture that provides a unified and consistent way for software to access and manage all data across the Enterprise.

### DevOps:
We can think of the software development aspects of DevOps as constituting a separate suite of applications in their own right that create, maintain, and serve operational software. Our interest here focuses mainly on operational and business applications that must be managed coherently over time.

### Distributed IT Resources:
Finally, the modern enterprise is usually widely distributed regarding hardware resources – possibly with multiple cloud instances, on-premise servers, desktop and mobile devices, and maybe even embedded processors.

Establishing a genuine EAI means having these IT categories properly defined and mapped. It then becomes possible to deploy software and data onto IT resources in ways that will deliver the service levels required and manage the whole environment over time, expanding it or changing it when necessary.
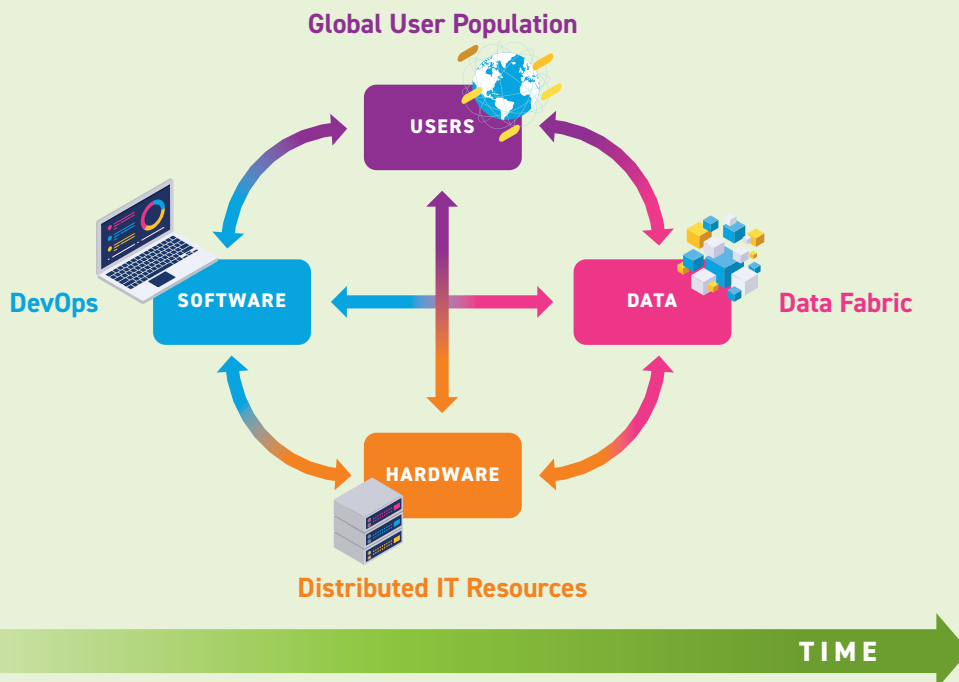


*Figure 2 – The Four Fundamental (IT) components*

# The Operating System of The Enterprise

Consider an Enterprise IT environment that delivers the capabilities we have described above. It establishes and maintains accurate catalogs of all four large groups of "IT assets." It knows the interrelationships between these groups and can take action based on this knowledge to run and manage all the applications that an organisation uses. In such an environment, the EAI that delivers these capabilities can legitimately claim to be an Operating System for the Enterprise.

To implement such an EAI ought to be the strategic IT goal of all organisations. Within such an EAI, all applications (with their users-software-data-hardware requirements) should be deployable like an application is deployed when you launch it on a PC. The PC knows the apps and the data at its disposal and provides them when requested by the user.

Solix's product portfolio can provide a substantial proportion of the "Operating System for the Enterprise." We discuss this below.

## The Solix Data Fabric

The Solix Data Fabric provides a rich and comprehensive capability. Solix categorises data into three groups:

1. The production data of the primary business applications.

2. The analytical data used by AI and BI applications.

3. Content data, such as documents, images, and video, are used by office and web systems.

The data fabric includes a layer of abstraction that hides the underlying complexity of the archipelago of data. It maintains and manages an information catalog with metadata, defining and providing access to all data. It also caters to many necessary data-related activities, including security, data distribution, and data governance. In respect of governance, it gives IT organisations a single point of control over the information lifecycle. This includes check-in and check-out management for data pipelines.

Cloud data management requires a data archiving capability. Archiving offloads data rarely used from production databases, and retires data from analytical activities and workgroup content stores. Nowadays, archiving data usually involves moving it to less expensive computer resources (usually in the cloud). It is a means of managing the economics and performance of Enterprise IT.

In respect of the enterprise IT Environment (and after the fashion of an operating system), Solix places data in the optimum location for its use and brings data, in a timely manner, to the place where it will be processed. Data is located according to demand: either in an archive store, in the Enterprise Data Lake, or with the application that uses it.

There are complexities within the enterprise IT environment. The Solix data catalog includes a data usage model and plays a vital role in data governance. Detailed metadata management is necessary here as regulatory requirements (such as GDPR) and most AI applications require data lineage tracking. Regulatory requirements also impose data security rules that require encryption and data masking. These, too, need to be managed by the data fabric.
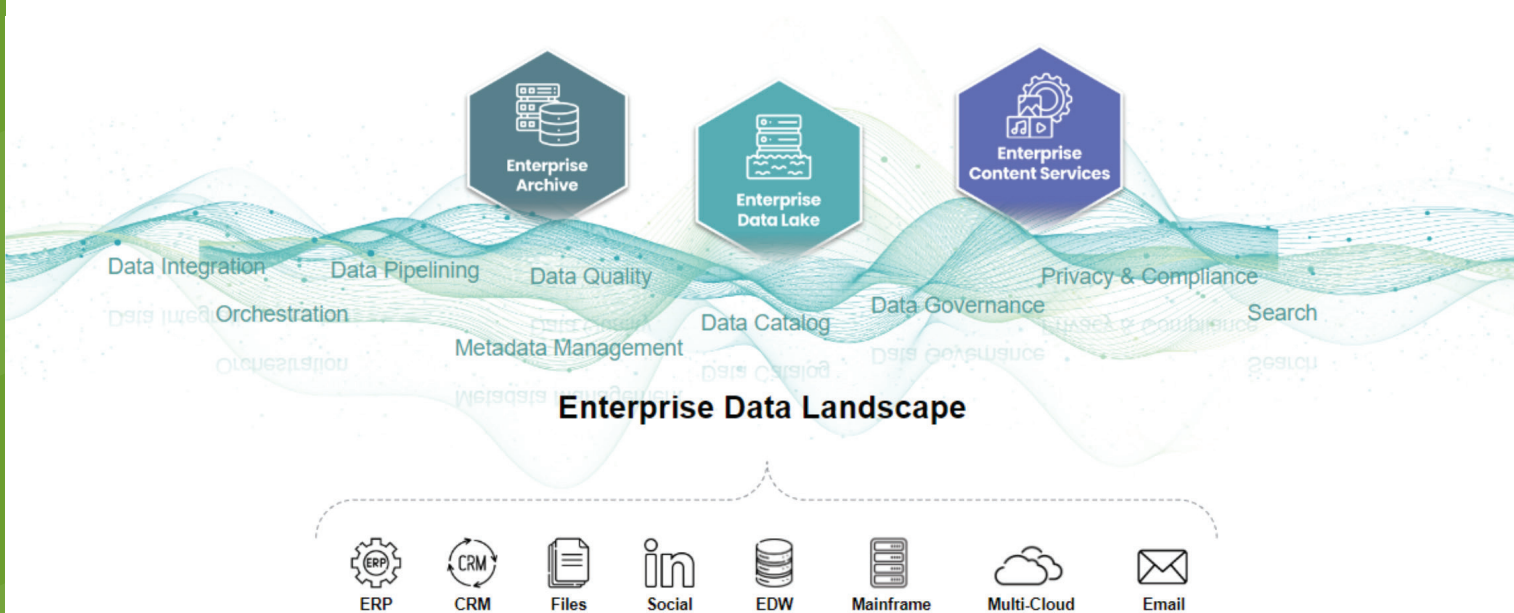


*Figure 3 – SOLIXCloud Enterprise Data Fabric*

Generally, the data fabric is tasked with meeting all application requirements within defined service levels. Its job is to get the required data to each user and application in a timely manner. This can be viewed as an Enterprise Operating System function.

## The Global User Population

Aside from the employees of an organisation, the enterprise user population must be global and may include a host of third parties, external agents, customers, and potential customers. The various groups of users need to have their details recorded and maintained and to have their application usage monitored. Essentially, there are two significant things to be managed.

1. **Security must be bulletproof.**
   Unauthorised use of Enterprise data, applications or resources must never occur.

2. **Role-based Access Control is needed.**
   Authorised users must be provided with the capabilities and services to which they are entitled within the limits of defined service levels.

The service level provided to any given user within the context of any given application depends on the efficiency of the data fabric and the versatility and management of the Distributed IT Resources.

## Distributed IT Resources

The distributed IT resources of the Enterprise include cloud resources of every kind, including the resources provided by any SaaS applications, PaaS or DaaS or anything else that supplements the on-premises capabilities. It also includes all mobile devices or embedded systems that may be part of the corporate computing network.

A detailed map of the whole resource space needs to be maintained, and resource usage needs to be monitored to avoid future shortfalls in capacity. This is the kind of capability that Solix provides with SOLIXCloud.

Thus, an orchestration must occur across the whole of Enterprise IT that links together the data fabric, the distributed IT resources, and the demands of the global user population.

## DevOps

At the Enterprise level, DevOps includes all software development activity and all upgrades to software (applications, databases, communications software, operating systems, etc.). Under normal circumstances, this environment is managed with a professional discipline that avoids unnecessary hiccups or failures. The most challenging activity in this area usually occurs when entirely new applications are introduced, as this will affect the data fabric and impact the demands on the distributed IT resources involving at least some of the user population. Almost everything is affected.

# AI and Gen AI

Regularly, about every 5 or 10 years, a sparkling new set of applications and capabilities emerges into the IT world. With hindsight, it is easy to identify such revolutions: The PC and office systems, The Internet, Social Networks, Cloud Computing, Big Data, and Machine-Learning (ML) based AI.

Such developments bring obvious business benefits but disrupt EIA, usually necessitating significant changes to whatever enterprise architecture has been assembled. Triggered by the release of the open-source Hadoop file system, Big Data caused extensive disruption. On the positive side, it gave birth to the idea of a data lake, which, in turn, made constructing a data fabric much more practical.

On the negative side, while Hadoop made large-scale parallel processing possible, it took many years for the Hadoop environment, with its many components, to mature. Big Data encouraged the development of AI environments built to exploit Machine Learning algorithms.

From a practical perspective, it became possible for data scientists to create data sandboxes within data lakes and run resource-hungry ML algorithms on the data they created, thus discovering useful knowledge. Nevertheless, there were problems in the evolution of ML. Effort was required to harvest data in a way that recorded its lineage, and it was also challenging to implement the ML models into production systems and existing business processes. The evolution of ML AI was slow, although the benefits were clear.

## Generative AI (Gen AI)

The sudden appearance of Gen AI in late 2022 was a shock to many observers. Its primary capabilities, which have gradually emerged, can be summed up as follows:

1. It has an unprecedented natural language ability, which makes it suitable for interactive conversations with people or (in some situations software). In this area, it can be thought of as a talented chatbot. In many situations, it can naturally augment the user interface of other applications.

2. It is highly skilled in creative activity, writing anything from email responses to blog posts or creating original artwork from simple verbal prompts.

3. It is also skilled in creating and documenting program code.

4. The large language models (LLMs) that drive its capability can be integrated with an organisation's data. This requires the implementation of a private LLM.

Beguilingly, its potential is still being determined as the technology is still in its infancy, and it will likely become more powerful. Nevertheless, its emergence prompted fast buy-ins from major IT providers, including Google, Microsoft, IBM, and also Solix, as we shall now discuss.

# Solix ML and Solix GPT

Solix currently offers an Enterprise AI capability that uses AI to provide the following functionality:

- **Data discovery and classification:**
  SOLIXCloud Enterprise AI surveys all enterprise systems to discover and classify structured and unstructured data. Thus, the data fabric remains current as the data catalog and metadata maps are kept up-to-date.

- **Data security and compliance:**
  It implements various security and compliance features, including encryption, data masking, and role-based access control, assisting organisations with their compliance obligations, including HIPAA, GDPR, and CCPA.

- **Information Lifecycle Management (ILM):**
  It assists in managing the data lifecycle. It allows for defining data retention and disposal policies, legal hold management, data archiving strategies, and reports when action is required.

- **Predictive analytics:**
  It employs predictive analytics to identify potential data risks and compliance issues so that proactive action can be taken.

In its recent announcement, Solix is adding two new components to SOLIXCloud Enterprise AI that directly support the implementation of AI applications. They are Solix ML and Solix GPT.

## Solix ML

Solix ML is a machine learning platform that helps organisations deploy, monitor, and maintain machine learning models. It is built to handle large datasets and supports the entire machine-learning lifecycle. It integrates naturally with Solix CDP, so all data, both structured and unstructured (including text, image, video, etc.), can be accessed and analysed.

It views the ML lifecycle as having the following stages: Data preparation, Exploration, and analysis, Feature engineering (for model training), Model training, Model validation, Pipeline/deployment and monitoring.

Solix ML is built for scalability and thus can be employed on the largest ML implementations. Because of its integration with the Solix CDP, organisations that already have active ML applications will likely be able to accommodate them within Solix ML.

## Solix GPT

Of the two new products, Solix GPT is the one most likely to attract strong interest as many organisations are still struggling with the problem of adopting Gen AI. They see obvious applications for it and know they will adopt it. Still, they have yet to determine how it will fit within the enterprise IT environment and where it will prove most effective within their business processes.



*Figure 4 – Solix GPT*

Solix GPT provides an excellent place to start. It implements Gen AI using a popular private LLM and connects it to the Solix Enterprise Data Lake via the Solix CDP. This is usually the first step in adopting Gen AI at a corporate level because the LLM is trained on corporate data and can be used immediately to investigate and answer questions that involve or are focused on corporate data.

Solix GPT can be used by any employee and requires little or no training, as it behaves like Chat GPT. The difference is that an employee could ask it to write a marketing document for a company product, write a blog post about industry news, or produce an HR report. Developers can use it to generate code or comment on and explain code in existing programs.

Because it is integrated with Solix CDP, Solix GPT has role-based access control, can implement end-to-end encryption, and applies dynamic and static data masking. This protects corporate data from unauthorised access.

### The impact of Gen AI on EAI

For organisations that do not have a well-organised enterprise IT environment, the adoption of Gen AI is likely to be as disruptive as any previous technology revolution. Without a data fabric providing a practical catalog of all corporate data, it will be difficult even to identify all the corporate data that should be made available to the Gen AI software. Security of data usage may prove problematic, as will any regulatory compliance that the organisation is subject to.

Data flow through the enterprise IT environment will inevitably increase, and the economics of enterprise computing will shift. Nevertheless, organisations will undoubtedly adopt Gen AI technology – the benefits are far too compelling.

Most organisations have already concluded that they will need a private Gen AI capability rather than risk exposing their valuable corporate data to ChatGPT or any other public Gen AI service. Solix provides it – a private Gen AI service to which is, by virtue of the data fabric, already able to access all corporate data.
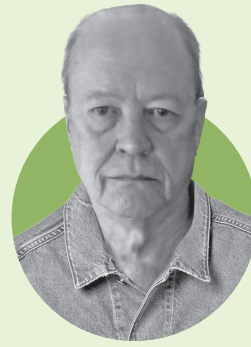
### Solix ML and GPT in summary

Ultimately the implementation of AI (ML and Gen AI) is all about data integration. Taken together, Solix CDP, Enterprise Archiving, the Enterprise Data Lake in combination with Consumer Data Privacy, provides a versatile environment for delivering safe, secure, and governed data ML and Gen AI.

# About the author

**ROBIN BLOOR**
Founder

Robin Bloor is a leading authority ansd influencer in the industry. In his role as an industry analyst, Robin has become an influential commentator on many corporate IT issues and is in great demand as a presenter at conferences, user groups and seminars addressing audiences across the world on a variety of technology topics from eCommerce through to IT Strategy and trends.

For a decade and a half he was the driving force behind the research effort at Bloor Research, and authored many of its industry reports and product comparisons. He has expertise across the whole field of IT with particular expertise in database, development tools, system management, IT security and hardware technology.

His best-selling business book *"The Electronic Bazaar: From the Silk Road to the eRoad"* was featured as book of the week by the Times, referred to as *"a classic"* by Publisher's Weekly (in the US), and described by the Library Journal (also in the US) as *"One of the Best Business Books of 2000".*

Now living in Austin, Texas, he is still a regular visitor to Europe. He is also a Partner in Hurwitz & Associates, a partner company to Bloor Research which provides IT analysis services to US companies.

He has been influential in shaping the direction and thinking of a generation of IT strategists and continues to provide insight on the direction of IT as it moves forward.

## 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**