Mobilizing data for an AI-powered world
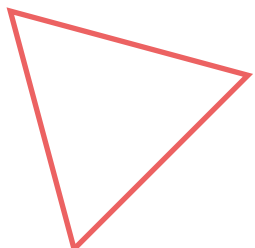
Why organizations must focus on modernizing legacy data platforms, ETL & data warehouses to enable AI analytics-led business transformation

Table of Contents

Introduction

For many decades, organizations have used the "traditional" extract-transform-load (ETL) process, spreadsheets, and BI tools to perform analytics to make recommendations for how to solve their business problems. Cut to the advent of generative AI, which is set to accelerate the pace of delivered data insights and predictions. In the age of AI technology, companies can no longer afford to generate data-driven insights with extensive human effort or time.

Artificial Intelligence (AI) is only expected to grow across industries in the coming years. According to 2024 statistics, 35% of businesses have adopted AI technology - while 77% of devices use AI in some form. Among the latest developments, gen AI can maximize business value across industries. The next wave of AI-powered innovations will depend on the generation of high-quality data.

The 2024 Google Cloud report on Data and AI Trends predicts that the roles of data engineering, analytics, and AI will blur in the years ahead. Here's a snapshot of some of the other trends highlighted by the Google Cloud report.

- 84% of business decision-makers believe generative AI and AI models will help their companies generate data-driven insights faster
- 2024 is the year of data engineering modernization
- 14% of companies are satisfied with the AI support provided by their legacy data systems

Are we witnessing the next wave of AI-powered innovation in analytics? Seems like it. Through its natural language capabilities, AI enables data analysts and other workers to interact directly with raw data and "ask" random questions. What's more, AI tools automatically provide direct insights that humans previously could not have.

Effectively, AI solutions are poised to transform data engineering by:

- Replacing the traditional legacy ETL-based data pipelines.

- Reducing the response time for answering business questions from weeks to a few minutes – or even seconds.

- Reducing human labor and costs in the ETL and Business Intelligence (BI) domain – by a whopping 90%

Why are legacy warehouses and data platforms falling short?

Image Source

Legacy warehouses like Oracle, Teradata, Db2, MS SQLServer, and Netezza have been the "backbone" of BI and data analytics for decades. On the flip side, data warehouses suffer from limitations, including:

- Traditional data ingestion and transformation processes are time-consuming and prone to human errors
- Limited scalability make them complex and expensive for increasing data volumes
- The lack of real-time data analysis capabilities hindered enterprises from making timely decisions
- The presence of data silos prevented a holistic view of the organizational data
- Raw data is not always retained or available for analytics purposes due to the scalability and cost restrictions of these legacy platforms
- Lack of support for advanced machine learning and AI models at scale

In an attempt to overcome these limitations, some enterprises have migrated to cloud-powered data platforms like Teradata Vantage and Snowflake. However, in an AI-powered world, these data platforms struggle to deliver real-time performance and scalability at an affordable cost or have other limitations. As the industry moves to AI-driven analytics, machine learning models need access to unfiltered operational data in near to real-time, unhindered by the restricted data models and layers of refinement that traditional ETL and data warehouses demand.

One of the primary benefits of AI-driven analytics is the ability of AI models to infer meaning and insights from raw operational data that were previously invisible, difficult, or time-consuming to discover. As the industry matures AI capabilities, the underlying data engineering, governance, and user interaction methods must be modernized, and traditional ETL, data warehousing, and BI are going to fade; they will be augmented and eventually replaced by natural language queries, AI/ML models, automated governance and rich knowledge graphs, based on unfiltered, real-time rich operational data, enriched by third-party data. This modernization to an AI-driven analytics world will deliver insights, KPI trends, alerts, recommendations, and automated actions based on data, at a speed and depth of knowledge far more capable than traditional approaches to BI and analytics.

Data and analytics modernization benefits

"I need to solve problems and do it on my own terms and timeline."

With the growing reliance on data, "traditional" data engineering practices, which include ETL and data warehouses, struggle to deliver on business needs. This led to the emergence of shadow IT – which empowers individual business teams (or citizen developers) to choose their analytics tools and processes to solve their business problems.

This "decentralized" approach primarily prioritizes speed over security. Here's how it compares to a "centralized" approach:

Decentralized	Centralized
• Self service	• Gate keepers
• Tool selection	• Tool direction
• Limited oversight	• Management oversight
• Business led and driven	• IT led and driven
• Speed over security	• Security over speed
• Ad hoc and on demand development	• Planned and prioritized development
• Flexible (unstable / undefined) operating model	• Rigid (stable / well defined) operating model
• Localized billing and contract governance	• Centralized billing and contract governance

Modernization of legacy ETL, data warehouses, and BI, and moving to an AI-driven analytics landscape will provide enterprises with a host of benefits, including:

- Agility with governance

- Platform scalability to leverage ML and AI models

- Flexible and immediate answers from data for business users without the latency of traditional IT tools or processes

- Real-time data processing for faster insights and decision-making

- Effective handling of various forms and volumes of data
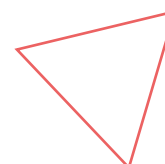
Here are some benefits of data pipeline and data warehouse modernization:

1. **Faster and broader analytics**
   Through the modernization process, enterprises can significantly improve their analytics capabilities in terms of scalability, security, and governance, reduce maintenance costs, and improve data quality. Besides BI, modern businesses thrive on effective analytics. A modernized data platform can integrate and extract real-time insights from incoming data (across touchpoints) stored not only in a traditional centralized repository.

2. **Business agility**
   As the number of data sources increases, the modernization of traditional data platforms, ETL, and data warehouses provides the foundation for organizations to achieve business agility. AI-driven replacement for ETL and data warehouses and BI is bringing great promise for business agility, efficiency, and reaction time without the massive investment in people costs and time to build traditional data pipelines.

3. **Centralized data or not**
   Through AI-based modernization, organizations can answer business questions without the traditional need to unify data from various sources. This helps in eliminating data silos with filtered or limited data sets and minimizing the adoption of shadow IT because these users have less need to "build their own" analytics solutions. With access to rich AI-enabled data and platforms, business teams can engage in effective analytics – with real-time insights into current business trends and patterns.

4. **Improved scalability**
   With modernization, enterprises can address a common limitation among "traditional" warehouses – lack of scalability. Cloud-powered data and analytics solutions can overcome this limitation with their ability to process diverse data formats at any scale at an affordable cost while unlocking ML and AI capabilities.

Next, let's discuss the possibilities of how AI technology can transform the data modernization process.

There are many ways to mobilize data for AI but only one way to make it soar.

We'll deliver it:

2X AT 1/2 FASTER THE COST

The urgent need to transform the "traditional" data engineering process

About Dr. Phil Shelley:

Currently the CSO at Onix, Dr. Phil Shelley was the co-founder and president of Datametica – a global leader in migrating and modernizing data warehouses to the cloud. With a PhD in biomedical engineering, Dr. Shelley is an industry expert on Data and advanced analytics.

Dr. Phil highlights the current state of how traditional data warehouses are functioning to deliver data-driven business insights, as illustrated below:
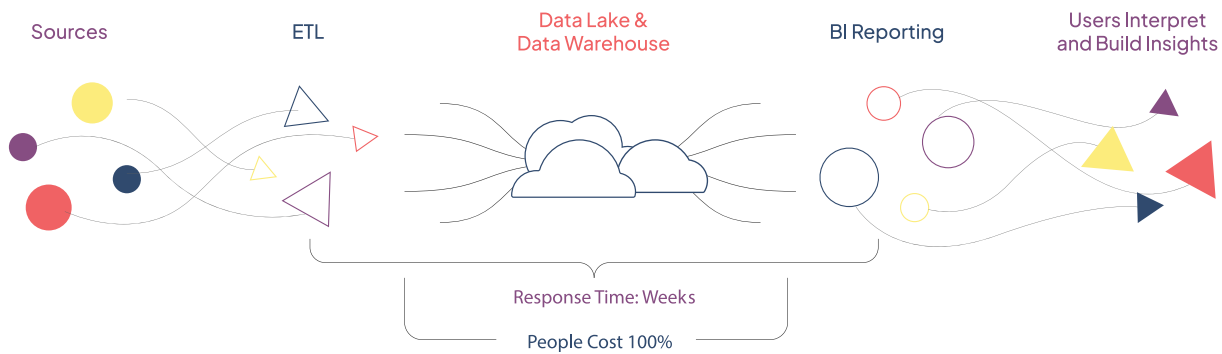
This "traditional" process has successfully delivered results for many decades. However, this method falls short when it comes to providing faster data-driven insights. Here are some of its obvious limitations:
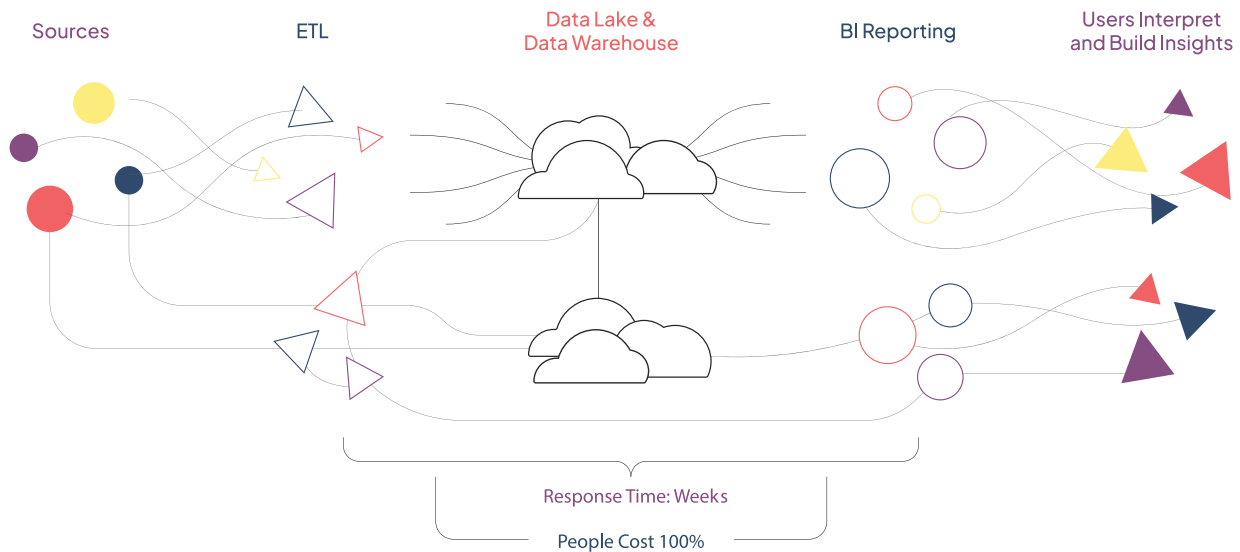
• For every new user request, enterprises take a few weeks to create or modify ETL pipelines, which creates delays and encourages shadow IT behaviors.

• Apart from the time consumed, the ETL stage requires expensive ETL professionals to connect data sources, thus adding to the employment costs and manual efforts.

• To generate or edit BI reports, enterprises need a high number of high-cost BI professionals and BI licenses – making the BI process slow and immensely expensive.

Dr. Phil emphasizes that while organizations have achieved some benefits by moving this ETL and data warehousing workload to the cloud, they are still essentially following the same time-consuming process. Despite their cloud efforts and spending, organizations are still not getting faster responses (or solutions) to their business problems. This is not the only challenge.

In reality, shadow IT is making this process even more complex and layered. Here's the current state of how this process "actually" works in the real world (with the shadow IT being the second layer):

Sources   ETL   Data Lake & Data Warehouse   BI Reporting   Users Interpret and Build Insights

Response Time: Weeks

People Cost 100%

According to Dr. Phil, the presence of shadow IT is a "waste of time and money." It creates security and data governance exposures, as individual teams are finding answers to their queries with their own data pipeline and copies of data – commonly comprising of an ETL process, data warehouse (or lake), and BI tools/spreadsheets.
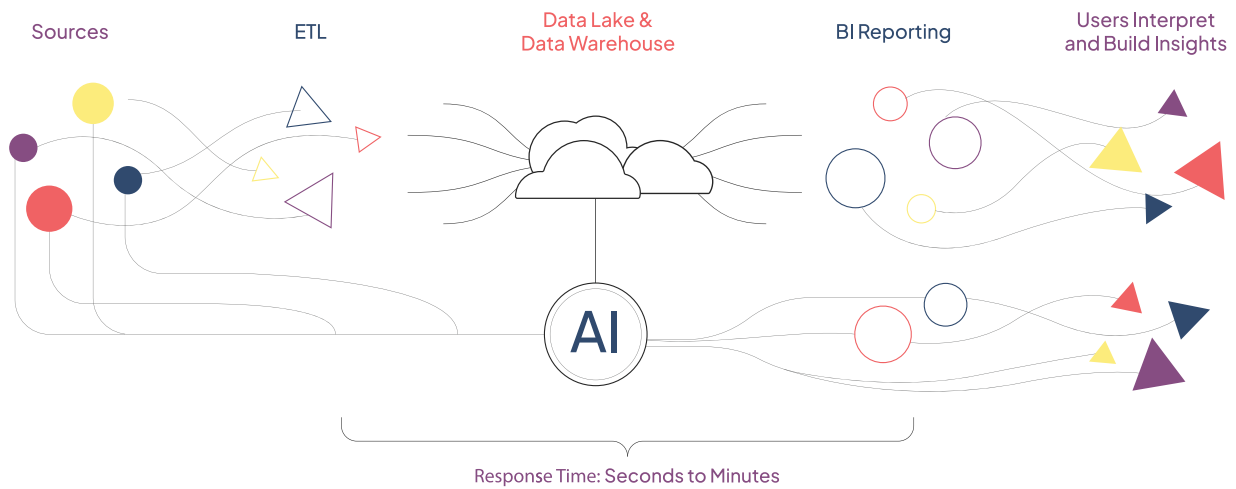
Artificial Intelligence: The game-changer

How can AI technology address the challenges of these traditional methods? Here are a few interesting prospects:

- AI models can be trained to replace the traditional ETL tools and work directly against the data sources.
- AI can deliver answers in seconds – with labor costs at a fraction of the cost – compared to traditional data engineering and data warehouses.
- AI can work with BI tools and deliver "intelligent" responses with its ability to understand and interpret existing BI reports and add the ability to ask questions in natural language.
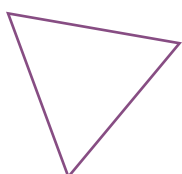
Dr. Phil projects how AI models can deliver BI insights by directly working on the data sources as illustrated below:

Sources ETL Data Lake & Data Warehouse BI Reporting Users Interpret and Build Insights
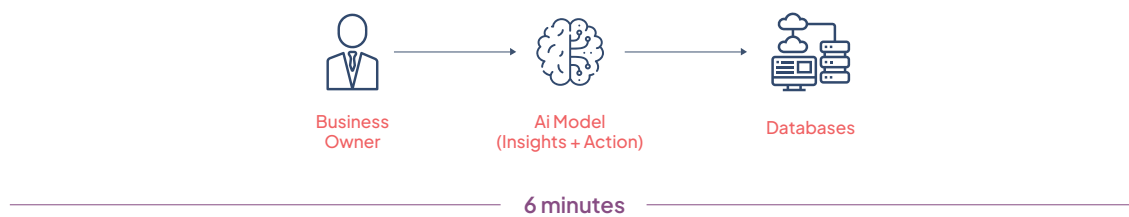
AI

Response Time: Seconds to Minutes

Here's an illustration of how the AI approach can deliver insights in just minutes – as compared to the conventional ETL/warehouse/BI approach, which takes weeks.

Conventional Workflow Approach of Insights

Business Owner → Business Analyst → Reports Dashboards → Data Warehouse → Data Sources

6 Weeks

AI Approach of Insights

Business Owner → Ai Model (Insights + Action) → Databases

6 minutes

In the era of AI dominance, legacy data warehouses cannot prepare data for a variety of AI applications or opportunities – including the following:

- AI-augmented analytics and AI-over-BI
- Real-time proactive decision-making
- AI-powered KPI and inference generation
- AI-first data engineering strategy

With the capabilities of Onix's Bird suite, companies can now mobilize data for AI faster and more cost-effectively. Let's see how in the next section.
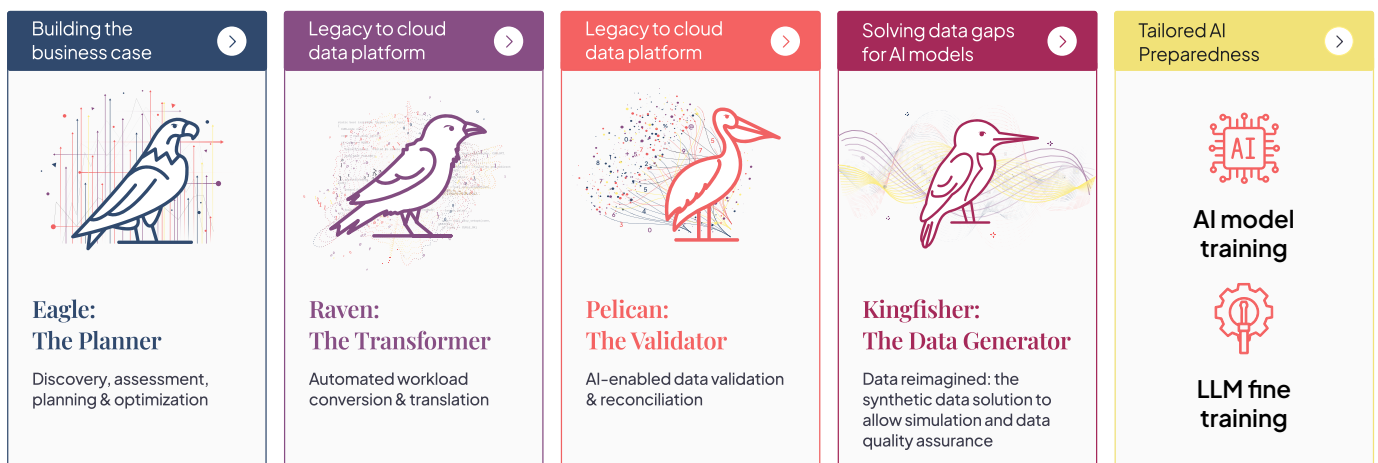
Accelerating data modernization with Onix's Birds

How can the Onix Birds product suite empower AI capabilities for data-driven companies?
Here's a short video.

According to the Google Cloud report, "2024 will be the year of rapid data platform modernization." Dr. Phil stresses that this process can be accelerated with the capabilities of Onix's Birds.

Data to AI journey: our approach & IPs in action

| Building the business case	Legacy to cloud data platform	Legacy to cloud data platform	Solving data gaps for AI models	Tailored AI Preparedness
Eagle: The Planner	**Raven: The Transformer**	**Pelican: The Validator**	**Kingfisher: The Data Generator**	**AI model training**
Discovery, assessment, planning & optimization	Automated workload conversion & translation	AI-enabled data validation & reconciliation	Data reimagined: the synthetic data solution to allow simulation and data quality assurance	**LLM fine training**

Here are some reasons that open up exciting possibilities:

- With its knowledge graphs capability, the Eagle tool can map out any company's existing ETL/ warehouse process and eliminate AI hallucinations.
- Onix's Birds can help organizations create a strong data foundation by migrating data to Google Cloud – thus opening up more AI-related possibilities.
- Training innovative AI models for improved outcomes using a synthetic data generator tool.
- A front-end AI-powered tool that can respond to user queries in natural language.
- Finding relevant business insights from generated BI reports without any human intervention.

How can your enterprise benefit from our Birds suite? Here are some key benefits:

- 2× faster migration to the cloud
- 50% lower migration costs
- Near 100% validation of data, including AI
- 100% code conversion guaranteed
- Fewer business risks and costs
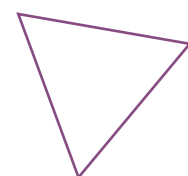
Takeaway

To leverage the emerging power of AI and explore new opportunities, organizations need to modernize and mobilize data for the AI future. At the same time, an innovative modernization approach is required to dispel the market's "disillusionment" with their AI investments.

In partnership with Google Cloud, Onix is best positioned to pioneer this AI-driven transformation. We are exploring how to replace traditional ETL tools with AI-powered tools that can save you both time and money. Our Birds suite provides you with the necessary foundation for mobilizing high-quality data quickly and cost-effectively. We can bring AI inference and drive KPIs and meaning from existing BI reporting. We can work together to reduce shadow IT and clean up the governance and security exposure that shadow IT brings.

If you are on a transformation journey to modernize your business data, talk to our experts today!

Get in touch

🌐 onixnet.com

✉ connect@onixnet.com