Unlocking Business Value through Data Engineering

A BUYER'S GUIDE

MAY 2023

SG Analytics
Life's Possible

Table of Contents

Introduction

When do I need Data Engineering? Why do I need Data Engineering? What can I do with Data Engineering? What? When? Why?

Feeling Confused about Data Engineering?

Let us help you!

You may need data engineering WHEN

1 **You have large volumes of data**

If you are dealing with large volumes of data, you may need data engineering to help you process and analyze the data efficiently.

2 **You have multiple data sources**

If you have data coming in from multiple sources in different formats, you may need data engineering to help you integrate and clean the data.

3 **You need to scale your data processing**

If you need to process your data quickly and efficiently at scale, you may need data engineering to help you build a robust and scalable data processing infrastructure.

4 **You need to automate your data workflows**

If you need to automate your data workflows, you may need data engineering to help you build pipelines that automate the flow of data from source to destination.

5 **You need to build machine learning (ML) models**

If you need to build ML models, you may need data engineering to help you prepare the data for analysis and build pipelines that feed data into your models.

WHY do you need data engineering?

You need data engineering because it can **HELP** you:

1 **Manage your data**

Data engineering can help you manage data more efficiently, ensuring that your data is accurate, consistent, and available when you need it.

2 **Optimize your data processing**

Data engineering can help you optimize data processing workflows, ensuring that your data is processed quickly and accurately.

3 **Build scalable data infrastructure**

Data engineering can help you build a robust and scalable data infrastructure that can handle large volumes of data and support your growing data processing needs.

4 **Automate your data workflows**

Data engineering can help you automate data workflows, freeing up your time and resources to focus on other tasks.

5 **Build ML models**

Data engineering can help you prepare data for ML analysis and build pipelines that feed data into your ML models.

With data engineering, **WHY** can you do:

You need data engineering because it can **HELP** you:

1 **Collect and store data**

Data engineering can help you collect and store data from multiple sources, ensuring that your data is clean, consistent, and reliable.

2 **Process and analyze data**

Data engineering can help you process and analyze data, providing insights that can help you make informed decisions.

3 **Visualize and report data**

Data engineering can help you visualize and report data, making it easier to communicate your findings to others.

4 **Build ML models**

Data engineering can help you prepare data for ML analysis and build pipelines that feed data into ML models, allowing you to build predictive models that can help you make better decisions.

Business Context of Data Engineering

Let us proceed step-by-step to understand more about Data Engineering and its NEED today

Data engineering is the discipline of designing, building, and maintaining data infrastructure and systems, which enable organizations to collect, store, process, and analyze data effectively. The field of data engineering has become increasingly important in recent years as organizations generate and collect vast amounts of data from various sources, including social media, IoT devices, and other digital platforms.

Today's data engineering ecosystem is a rapidly evolving field driven by the increasing demand for data-driven insights in organizations of all sizes and across all industries. Some of the key components of the data engineering ecosystem include:

First Wave – Traditional On-premise Application

In this traditional approach, organizations would typically purchase and install their own hardware and software, hire a team of IT professionals to manage and maintain the infrastructure, and build custom data processing pipelines using tools like ETL (extract, transform, load) and data warehousing solutions. This approach provided organizations with a high degree of control and customization over their data infrastructure and processing, but it also required significant capital expenditures and ongoing maintenance costs. For example, Oracle MS SQL Server, DB2, Informatica, Cognos, etc.

Second Wave – Big Data Platforms

These platforms provide a distributed computing environment for processing and storing large volumes of structured and unstructured data. Popular big data platforms include Apache Hadoop, Apache Spark, and Apache Flink.

Third Wave – Cloud Computing

Cloud computing platforms such as Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) provide scalable and cost-effective infrastructure for storing, processing, and analyzing large volumes of data.

Some Key Considerations Are

Selecting the right data engineering services for your organization can be a daunting task, as there are many different providers and offerings to choose from. Here are some key factors to consider when selecting data engineering services:

Business Needs

The first step in selecting data engineering services is to understand your organization's specific data engineering needs. What types of data do you need to collect, process, and analyze? What are your goals for data processing and analysis? What are your data security and compliance requirements? Understanding these factors will help you identify the right data engineering services for your organization.

Technology Stack

The data engineering services you select should be compatible with your organization's existing technology stack. This includes the programming languages, databases, data storage and processing tools, and other technologies you may currently be using. You must evaluate your chosen data engineering services to ensure they support your technology stack and can integrate with your existing systems.

Expertise and Experience

Look for data engineering service providers that have a track record of success and expertise in your industry. You must also consider the qualifications and experience of the individual data engineers who will be working on your project. Look for teams that have experience working with similar data sets, tools, and technologies to ensure they can deliver high-quality results.

Scalability and Flexibility

Your data engineering needs may change over time, so it is important to select services that can scale and adapt as your organization evolves. Look for providers that offer flexible pricing models and can quickly ramp up or down their services as needed. You must also consider providers that offer a range of services, from simple data processing to more complex data analytics and ML.

Security and Compliance

Data security and compliance are critical factors to consider when selecting data engineering services. Look for providers that have strong security protocols in place, including data encryption, access controls, and regular security audits. You must also ensure that the data engineering services you select comply with relevant data privacy regulations and industry standards.

Support and Maintenance

Once you have selected data engineering services, you will need ongoing support and maintenance to ensure the continued success of your data projects. Look for providers that offer comprehensive support services, including technical support, training, and ongoing maintenance and updates, for their software and tools.

By carefully evaluating these factors and selecting the right data engineering services for your organization, you can ensure that you have the right tools, expertise, and support to meet your data processing and analysis needs, and to drive business success.

How do I Ensure the Quality of Data Engineering Services?

Here are some best practices for data engineering:

Data Modeling

Data modeling is the process of defining the data structure, relationships, and constraints of a database system. Best practices include designing normalized data models, defining primary and foreign keys, and ensuring data consistency.

Data Processing

Data processing is the process of transforming and manipulating data to create useful insights. Best practices include using scalable data processing frameworks, performing data validation and verification, and optimizing data processing workflows.

Data Integration

Data integration is the process of combining data from multiple sources into a single, unified view. Best practices include identifying and resolving data conflicts, creating a data integration roadmap, and using data integration tools.

Data Storage

Data storage is the process of storing data in a secure and scalable way. Best practices include selecting the right data storage technology, ensuring data security and privacy, and performing regular data backups and disaster recovery tests.

Data Quality

Data quality is the process of ensuring that data is accurate, complete, and consistent. Best practices include setting data quality goals, implementing data quality controls, and performing regular data quality checks.

Data Governance

Data governance is the process of managing the availability, usability, integrity, and security of the data used in an organization. Best practices include defining data governance policies and procedures, assigning data ownership and stewardship, and implementing data governance tools and technologies.

Agile Methodology

Agile methodology is an iterative and collaborative approach to software development that prioritizes customer satisfaction and responds to changing requirements. Best practices include using agile data engineering methodologies, conducting regular stand-up meetings, and continuously delivering value to stakeholders.

Documentation

Documentation is the process of creating and maintaining records of the data engineering process, including data models, data flow, data processing workflows, and data governance policies. Best practices include creating comprehensive documentation, using standardized documentation templates, and regularly reviewing and updating documentation.

How do I gain SUCCESS in my Data Engineering Services?

The success criteria of data engineering can vary depending on the specific project or organization, but here are some general metrics that can be used to measure success:

1

Data Quality

Data engineering should ensure that data is accurate, complete, and consistent. Metrics for data quality can include data completeness, data accuracy, and data consistency.

2

Data Availability

Data engineering should ensure that data is available when needed. Metrics for data availability can include data uptime, system uptime, and response time.

3

Data Processing Time

Data engineering should ensure that data processing is efficient and meets performance requirements. Metrics for data processing time can include data processing speed, data latency, and data throughput.

4

Data Security

Data engineering should ensure that data is secure and protected from unauthorized access. Metrics for data security can include data security incidents, data breaches, and data access controls.

5

Data Governance

Data engineering should ensure that data is governed appropriately and in compliance with regulations and policies. Metrics for data governance can include data governance violations, compliance audits, and data stewardship.

6

Cost-effectiveness

Data engineering should deliver value at a reasonable cost. Metrics for cost-effectiveness can include total cost of ownership, return on investment, and cost per unit of data processed.

7

User Satisfaction

Data engineering should meet the needs of the users and stakeholders who depend on the data. Metrics for user satisfaction can include user feedback, user adoption rates, and user retention rates.

Where can I find my Data Engineering Partner for Consulting, Solutioning, Delivery, Winning Relationships, and Thought Leadership?

SGA – A TRUSTED and WINNING PARTNER in the data engineering journey

A vast experience in Investment Banking, Capital Market, Healthcare, and Media & Entertainment has proven SGA's mettle to provide scalable service and solutions for their valued data engineering customers.

SGA has brought value to data engineering through its expertise, capabilities, and flexibility for its customers. SGA has a talented and customer-focused team of data engineers, which helps in making customer processes efficient, effective, and focused.

Our team has developed proprietary software and applications to reduce the time to market (TTM) and increase the transparency of data lineage. Our data migration framework and accelerators have helped reduce 50% of tasks for moving data from one database to another.

SGA has all the latest data technologies to help customers in their Data Cloud journey. SGA is a full-fledged company that enables customers to execute end-to-end data analytics and data engineering projects with ease and confidence.

SGA has expertise in AWS, Azure, GCP, and Snowflake Data Cloud. SGA also has expertise in ELT tools such as IICS, Databrick, DBT, Matillion, and FiveTran.

We have gained competitive benefits through our business domain expertise:

Investment Banking

Completed 50+ projects of data engineering using a modern data platform.

Capital Markets

Completed 25+ projects of data engineering and migrated 4 projects from the old traditional platform to the cloud data platform.

Healthcare

Partnered with 20+ providers for their data journey and end-to-end data analytics.

Media & Entertainment

All major media houses in the world are working with SGA for their business data processing using data engineering.

SGA Technology Expertise

SAAS Applications
Salesforce, SAP HANA, Oracle Fusion, Workday

Big Data Cloud Analytics
AWS, Azure, GCP, Hadoop, Spark

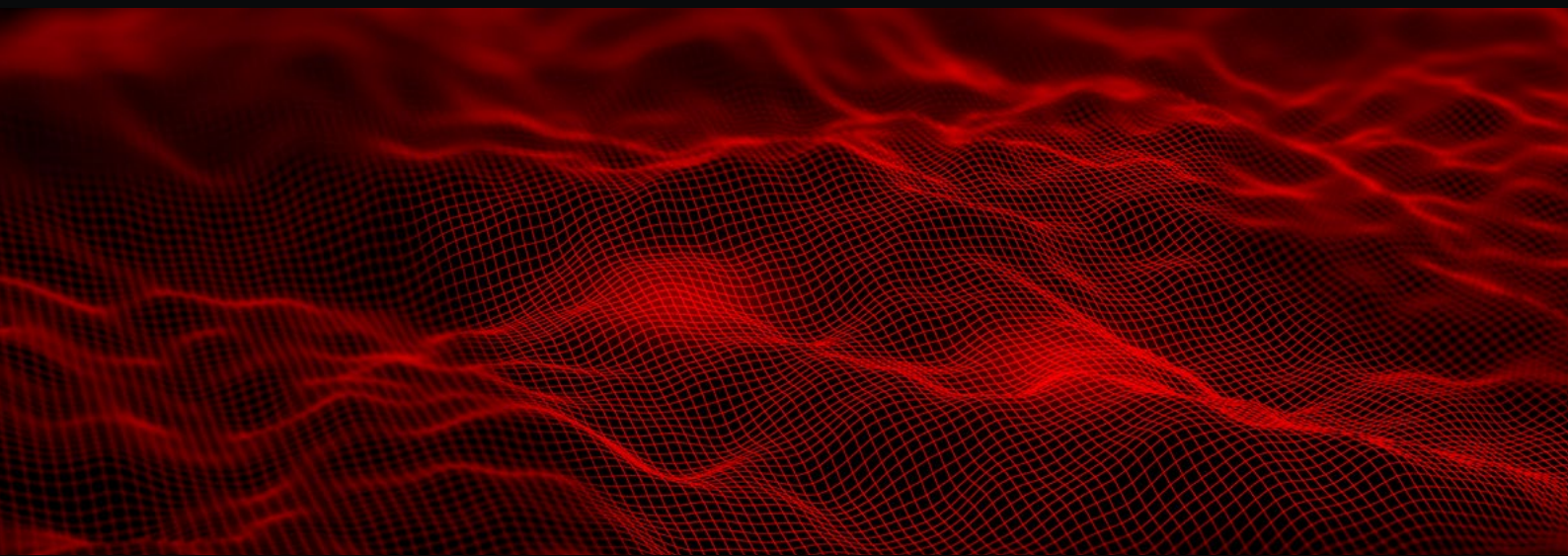
Cloud Databases
Snowflake, RedShift, Bigquery, Azure Data Warehouse, Synapse

Data Engineering
AWS Glue, EMR, Azure Data Factory, Databrick, Fivetran, Matillion

Data Science
AWS Sage Maker, Azure ML Studio, GCP ML Ops

Reporting/Visualization
Tableau, Power BI, AWS Quicksight, Looker

Databases
Oracle, Teradata, Netezza



What is Your Takeaway?

In conclusion, data engineering plays a critical role in harnessing the power of data and turning it into valuable insights and actions. By designing, building, and maintaining the infrastructure and systems required to store, process, and analyze data, data engineers enable organizations to make informed decisions and drive business growth.

Modern data engineering ecosystems offer a wide range of tools and technologies that can be tailored to each stage of the data journey, from data ingestion and processing to storage, analytics, and visualization. In addition, cloud platforms have opened up new opportunities for organizations to leverage data and gain a competitive edge, with benefits such as scalability, flexibility, and cost-effectiveness.

To succeed in data engineering, organizations must follow best practices such as data governance, security, and quality, and partner with trusted providers that can offer expertise and support throughout the data journey. By doing so, organizations can unlock the full potential of data and drive innovation and growth in today's data-driven world.

SGA can be your Trusted Partner with domain expertise, capabilities, and flexibility to understand your business needs.

Life's Possible

SG Analytics
Life's Possible

www.sganalytics.com

SG Analytics, a Great Place To Work® (GPTW) certified company, is one of the leading research and analytics firms to offer data-centric research and contextual analytics services. With presence in the US, the UK, Switzerland, and India, SG Analytics has been consistently meeting and exceeding customer expectations by its knowledge-based ecosystem and impact-oriented solutions.

Pune | Hyderabad | Bengaluru | London | Zurich | New York | San Francisco | Amsterdam | Toronto