aws

DESIGNER CLOUD
Powered by TRIFACTA

alteryx

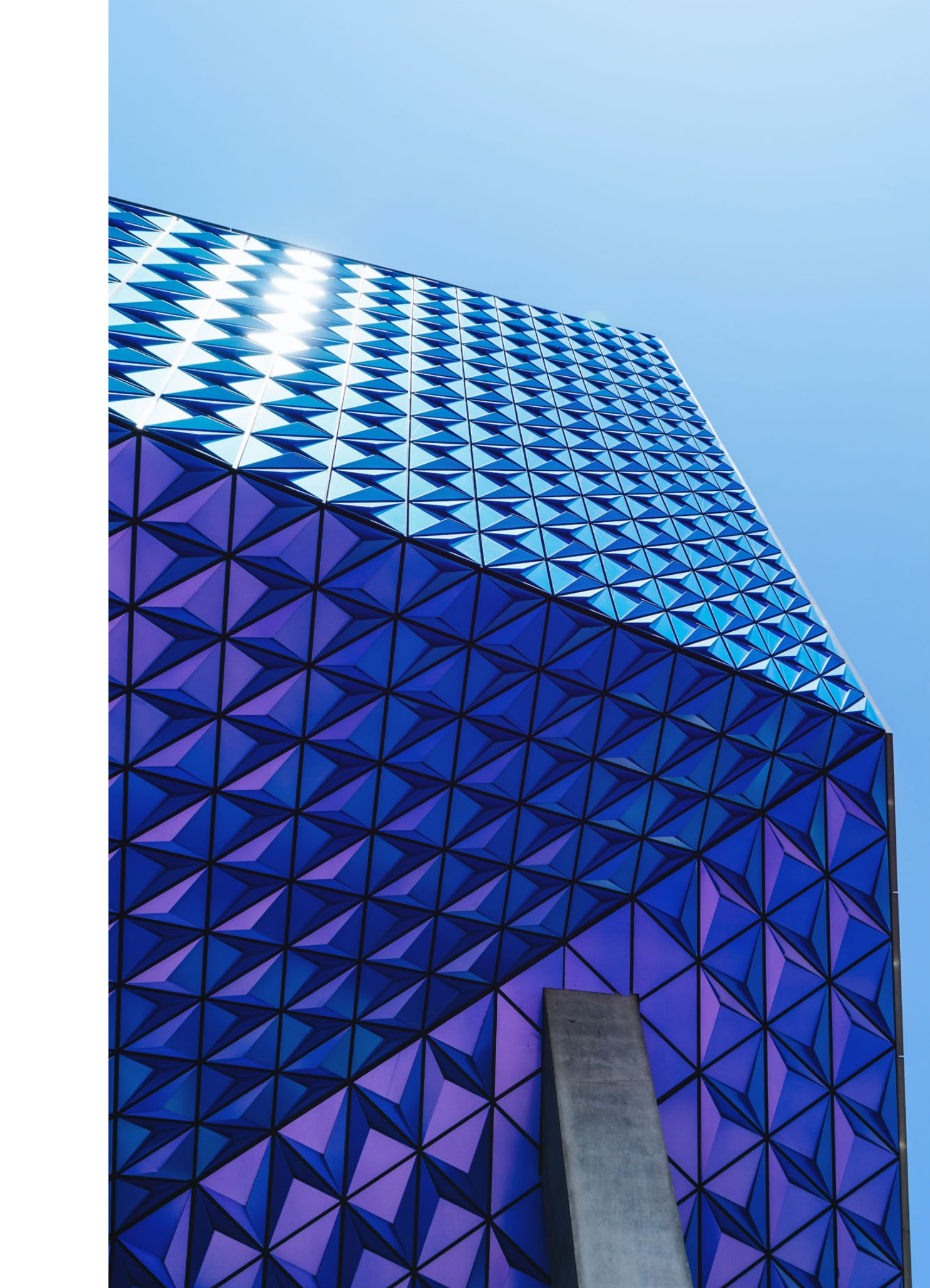# HOW TO SELECT A MODERN DATA PREPARATION SOLUTION FOR AMAZON WEB SERVICES

Your organization is looking to explore varying ranges and volumes of data on the Amazon Web Services (AWS) cloud platform to improve your BI reporting, ML predictions, and Data Services.

But great analytics starts with great data. And great data takes preparation. Data preparation can end up being the most time-consuming step in the analytics workflow if you don't have the right solution.

Analysts spend up to **80%** of their time preparing data instead of analyzing and interpreting it

alteryx

This eBook outlines the six capabilities a modern data preparation solution should have to maximize your organization's investment in AWS:

1. Seamlessly integrate with the AWS ecosystem
2. Improve stakeholder efficiency and collaboration
3. Address diverse data types and new use cases
4. Enable self-service while maintaining data governance
5. Support data quality at scale
6. Centralize the orchestration of data preparation

alteryx

Seamless Integration with AWS Ecosystem

A cloud-native data prep solution should be tightly integrated with AWS services to capitalize on the scalability, agility and speed offered on the cloud platform. Your organization needs a modern data preparation solution that supports:

- **Storage** by leveraging Amazon S3 to reduce the time it takes to collect and move data to an intermediate system for data preparation.

- **Data management and processing** through tight integration with AWS Glue, Amazon Elastic MapReduce (EMR), and Amazon Redshift to make it easier and faster to access data, schedule, and execute data preparation jobs, while maintaining centralized data governance to eliminate data silos and gain transparency throughout the data prep workflow.

- **Security** by leveraging AWS Identity and Access Management (IAM) and AWS Key Managment Service (KMS) to ensure data is available to only authorized users during preparation.

Improve Stakeholder Efficiency and Collaboration

Your data preparation solution should enable all stakeholders to participate in the data preparation process.

BUSINESS / DATA ANALYSTS should be able to access raw data in an Amazon S3 data lake or Amazon Redshift data warehouse to discover and explore data set(s) to answer business questions themselves, without having to rely on IT's technical expertise.

alteryx

A data preparation solution that offers interactive, visual guidance and ML-powered recommendations for data transformation can help all stakeholders make the best use of AWS.

DATA SCIENTISTS, who use Amazon Sagemaker and other Amazon AI and ML services to build and train machine learning models and algorithms, should be able to collaborate with lines of business users to interpret the results and apply the insights to drive better business decisions.

DATA ENGINEERS, who are responsible for managing the data management infrastructure should be able to ensure connected and trusted data is easily available to the business users at all times.

Address Diverse Data Types and New Use Cases

Amazon S3 data lakes and Amazon Redshift data warehouses allow your organization to store an enormous volume of diverse data—structured, semi-structured, and unstructured—making it easy to access all available data in a centralized location.

While traditional ETL tools are good at handling structured data in operational systems or databases, they're not equipped to clean and prepare large-scale, raw data stored on AWS. Such limitations hinder organizations from tackling exploratory or experimental analytics projects, impacting the organization's competitiveness.

Your organization needs a modern data preparation solution that's designed and architected to handle any scale of diverse, complex data in AWS to test new use cases and algorithms.

alteryx

Enable Self-Service While Maintaining Data Governance

Self-service data preparation is critical. Business users who know data best should be able to explore, profile, structure, clean, and enrich data themselves without having to rely on limited IT resources.

But how does your organization find the right balance between empowering all users to derive value from data while maintaining a holistic data governance process? Here are two ways:

1. **Centralize data preparation** rather than letting data silos proliferate across business units where users run their own preparation routines. A cloud-native solution runs data preparation processes centrally on AWS as users explore and interact with different types of data.

2. **Use data catalogs or glossaries** to manage data definitions, metadata, and data lineage to help users find data faster and empower your organization to govern and monitor the entire data life cycle. A data preparation solution that's natively integrated with AWS can leverage services such as AWS Glue to centrally manage the data for IT.

Support Data Quality at Scale

SAs growing volumes and varieties of data are ingested and integrated into your AWS platform, your organization needs to continuously validate that data quality is being maintained. Continuous validation means that users shouldn't have to wait until the end of a data prep process to view and test results—a process that's integral to agile development methods.

alteryx

Modern data preparation solutions like Designer Cloud can automatically detect potential data quality issues in large volumes of data and are capable of monitoring data flows and alerting users when new data is ready for validation.

By applying AI and machine learning to automate data cleansing and transformation, a modern data preparation solution that's natively integrated with AWS supports ongoing data quality at scale in dynamic AWS environments.

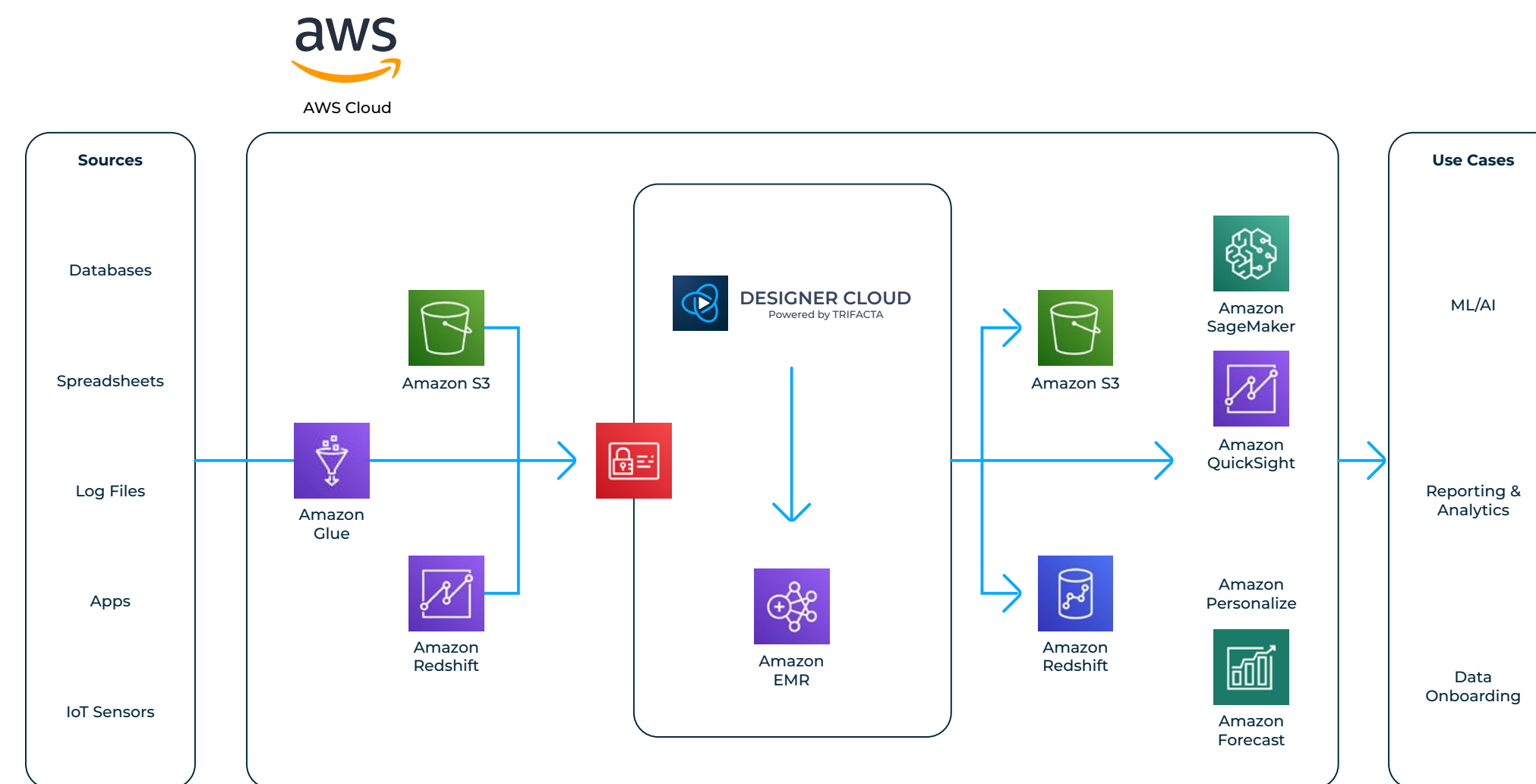## Centralize the Orchestration of Data Preparation

Since multiple data preparation workloads can be executed in parallel on AWS, data preparation routines must be scheduled, published, and operationalized in a well-governed manner so that all stakeholders have usable data immediately, whenever they need it. Centralized data preparation orchestration helps to ensure consistency and performance, reduce maintenance and operational costs, and drive faster time to results.

Centralization, plus integration with data catalogs such as AWS Glue, increases the potential for reuse among stakeholders for various analytics initiatives, including BI reporting, machine learning, and data services. Centralization also improves portability, allowing your organization to change front- or back-end technologies and services without compromising investments in data preparation.

alteryx

Designer Cloud powered by Trifacta is a Modern Data Preparation Platform Architected for AWS

Designer Cloud powered by Trifacta is a serverless data preparation service on Amazon Web Services (AWS) leveraging Amazon ecosystem services.

It enables data analysts, data scientists, data engineers, and business users to prepare data of any form and size, and quickly transform data from its raw format into a refined state for analytics and/or machine learning initiatives.

Want to try it for yourself? Start a Free Trial of Designer Cloud today!

alteryx

alteryx

Accelerate Towards Insights with Alteryx

Try Designer Cloud
for free

> Start a free trial

See Designer Cloud
in action

> Watch the video

Talk to an expert

> Book a consultation

See what our customers
are saying

> Read their success stories