

Features of Ab Initio and AWS Glue

The following tables provide details for Ab Initio and AWS Glue related to the key features for Architecture, Pricing, ETL Development, Orchestration, Data Quality and Data Governance.

Architecture	Ab Initio	AWS Glue
Deployment Options	Ab Initio has been traditionally used as an on-premise ETL tool. However, Ab Initio applications can now be deployed on all standard cloud environments, including Amazon Web Services (AWS), Google Cloud, and Microsoft Azure.	Cloud, Serverless.
Ecosystem Integration	Integrates with Ab Initio specific tools like Enterprise Meta Environment and Continuous Flows. Connector licenses needed for interacting with additional systems.	Integrates with AWS ecosystem that has technology and tools for enterprises all the way from IoT, Streaming, Analytics and Machine Learning.
Scalability	Since most Ab Initio implementations in the market are on-premise, run time scalability is not easy to achieve.	Since Glue is serverless and scaling is automatically handled. You provide the number of DPUs (Data Processing Unit) needed for running the jobs and Glue provisions the necessary compute.

Pricing	Ab Initio	AWS Glue
Pricing Structure	Strictly confidential pricing structure. Costs exclude any cloud or infrastructure costs.	DPU based pricing plus additional pricing for Connectors in Connector Marketplace. Costs include any cloud or infrastructure costs.
Total Cost of Ownership	Costs exclude any cloud or infrastructure costs.	Costs include the infrastructure costs. No additional charges for running jobs.

ETL Development	Ab Initio	AWS Glue
Connect to Different Sources	Free connectors. Licensed connectors.	Free connectors. Licensed connectors. Connector Marketplace. Build your own connectors.
Build ETL Pipelines	Visual Development, code can be embedded.	Dev Endpoints - Notebook styled interactive Development. Glue Studio - Visual ETL and code can be embedded. Glue DataBrew - No code ETL for Data Analysts.

ETL Development	Ab Initio	AWS Glue
Reusability	Reusable graphs/sub-graphs and custom components supported.	Parameterized jobs, programmable blueprints.
Code Lock-in	Yes, objects are stored as graphs which are specific to Ab Initio software.	No, code is written in open source PySpark or Scala. Code generated by Glue Studio is extremely easy to understand and modify without visual interface.
Debugging	Yes, can debug jobs on graphs using various features like watchers, checkpoints, breakpoints, step by step execution, etc.	Yes, can debug using Interactive notebook.
Automated Code Generation	Code generation can be automated with the help of meta programming functions supported in the transformation functions in Ab Initio. Dynamic script generation creates a shell script which can be executed via Ab Initio co-op.	Jobs can be parameterized. Jobs can also be automatically created using preview feature called programmable blueprint.
Complex Transformations	Yes, offers transformations as "components" and are very rich.	Yes, offers all transformations in Spark and Dynamic frames that allows customers to handle semi-structured data effectively. Not all transformations are available in the visual editor yet, but can be easily coded using custom transformations.
Data Preparation (Allow non-data engineering team to build ETL pipeline)	This can be done using Express IT tool. This tool enables business analysts to write business rules and perform testing using easy to use excel-like user interface. This tool needs to be licensed separately.	Glue DataBrew can do this with a pay as you go model.
Built-in Metadata Catalog	Ab Initio supports reusable datasets which can be stored as datasets in the project. Users can drag and drop the datasets in the graphs.	Yes, Glue Data Catalog is integrated with Glue.

Orchestration	Ab Initio	AWS Glue
Lake House Architecture Support (Ability to easily synchronize data across purpose-built databases such as Document databases, NoSQL databases, etc)	Manual.	Glue Elastic Views support creation of materialized views from sources and replicates them in real time. This is in preview.
Event Driven Workflows (Trigger a workflow by listening to events)	Ab Initio supports this kind of feature using Ab Initio Plans and continuous flows, etc. However, it needs to be licensed separately.	In private preview. Glue jobs can easily trigger jobs based on Eventbridge notification without any licensing. You will incur cost for using Eventbridge.

Orchestration	Ab Initio	AWS Glue
Monitor Results	Test results can be monitored using Ab Initio control center.	Supported via open source testing and Data Quality solution DeeQu. Learn more about DeeQu and AWS Glue.
CI/CD Pipeline	Ab Initio has exposed a Jenkins plugin to facilitate effective CI/CD. Version change is provisioned by Path monitoring and tag monitoring on the Ab Initio SCM, i.e. EME.	Yes, CI/CD pipelines and approval processes can be setup.

Data Quality & Data Governance	Ab Initio	AWS Glue
Define Data Quality Rules	Ab Initio Data Quality Environment and the testing framework enables users to build automated test framework to test Ab Initio data pipelines. Newer versions of Ab Initio also support writing unit tests at graph level.	Supported via open source testing and Data Quality solution DeeQu. Learn more about DeeQu and AWS Glue.
Data Lineage	Yes, licensed separately under Ab Initio Metadata Hub.	Private preview available.
Compliance Certificates	Not available for public reference.	Third-party auditors assess the security and compliance of AWS Glue as part of multiple AWS compliance programs. These include SOC, PCI, FedRAMP, HIPAA, and others.