

4.0

Ab INITIO

Contents

3 Introduction

4 Cloud Native

- Built for the cloud
- Containers
- Build once, deploy anywhere

6 Data Agility

- Understanding your data
- Adding meaning to the catalog
- Simplifying data access
- Focusing on business logic
- Enabling microservices, real-time services, and streaming services
- Governing data
- Accelerating testing
- Empowering data science

16 Innovative Applications

- Customer experience
- Entity resolution

18 Taking the Lead in the Digital Economy

19 About Ab Initio

Introduction

The companies that derive the maximum benefit from the expansion of the digital economy are the companies that innovate rapidly and use data aggressively to reinvent products, services, business processes, and customer interactions.

Ab Initio accelerates digital innovation and data analysis for many of the world's most successful and forward-thinking organizations in banking, retail, healthcare, pharmaceuticals, insurance, logistics, and government. With our 4.0 release, Ab Initio enables your business to take significant steps toward becoming an agile and data-driven organization by introducing advanced new capabilities in cloud technologies, data agility, and business applications that:

- Make it possible to build cloud native applications that are portable across private and public cloud environments
- Provide elastic scalability for both batch and real-time applications, enabling you to respond to rapidly changing markets
- Accelerate extraction of value from data
- Use automation to increase agility, productivity, and speed of data orchestration
- Access, understand, and leverage data with ever-increasing speed and ease
- Provide full data virtualization across your corporate ecosystem
- Empower users to do their jobs faster, more efficiently, with higher quality results
- Dynamically interact with millions of customers in real time, with unprecedented insight into their needs, wants, and behaviors
- Address the entity resolution problem with comprehensive and tunable entity matching and data disambiguation features

The advanced new features of Ab Initio's 4.0 release enable you to drive digital innovation in your organization.



Cloud Native

Built for the cloud

Ab Initio is a native cloud processing environment. Ab Initio 4.0 supports RESTful APIs, microservices, cloud messaging (Amazon Kinesis, Google Pub/Sub, Azure Message Bus, Kafka), and cloud databases (Snowflake, Redshift, Cassandra, RDS Aurora, BigQuery, and many others).

When your company uses Ab Initio in the cloud, you enjoy significant advantages:

- You can easily implement hybrid and multicloud ecosystems. Transitioning between cloud environments is simple. You are never locked in to a single cloud provider.
- With Ab Initio's advanced Just-In-Time execution model, applications can run unchanged on the environments of all major cloud providers.
- Designed for high throughput and massively parallel applications, Ab Initio's unrivaled compute efficiency and high performance mean your applications require far less cloud computing power than applications built with other technologies.
- With elastic processing, Ab Initio scales on demand. You use only the resources you need, as you need them — allowing you to rapidly respond to changing business conditions.
- Ab Initio supports DevOps in the cloud, including the ability to automatically deploy an entire environment — complete with all libraries and other supporting software — to production.

Ab Initio seamlessly integrates many cloud technologies

Storage

-  Amazon S3
-  Azure Blob Storage
-  Google Cloud
-  Amazon EFS, EBS
-  HDFS

Messaging

-  Amazon Kinesis
-  Amazon MQ
-  Kafka (MKS)
-  Google Pub Sub
-  JMS

Databases

-  Google BigQuery
-  Redshift
-  Amazon RDS Aurora
-  Snowflake
-  Cassandra

Runtime

-  Containers
-  Kubernetes
-  Hadoop
-  Google Dataproc
-  Amazon EMR
-  Virtual Machines

Containers

Ab Initio is fully integrated with standard container technologies, including Kubernetes (k8s).

Because Ab Initio software was designed from the beginning to scale horizontally and vertically across an arbitrary number of servers, it naturally scales across clusters of pods and containers as well, enabling one processing job to run in parallel across multiple containers. The number of containers used is elastic: at various points, a running job can request additional containers or release containers that are no longer needed. Ab Initio manages job execution for you, facilitating exactly-once processing and handling job recovery automatically within the cluster.

Developers do not need to be experts in Kubernetes to take full advantage of containers — a single command within Ab Initio is sufficient to go from an Ab Initio application to a container image.

Ab Initio supports buildpacks — pluggable, modular tools that translate application sources into Open Container Initiative (OCI) images. Because the people who are responsible for your company's infrastructure configure buildpacks, you have complete control over the contents of each layer of your container images.

Building portable cloud-native applications is simple with Ab Initio 4.0. You can build and run fully scalable applications in public, private, and hybrid clouds, and you have the flexibility to change platforms. On premises or in any cloud, it's all the same with Ab Initio.

Build once, deploy anywhere

With Ab Initio software, you can write applications once and deploy them anywhere: in the cloud, in containers, on Hadoop, Linux, Unix, Windows, and z/OS (mainframe). Ab Initio 4.0 expands our native support for cloud object storage, databases, cloud-based messaging, and operating within containers. You can migrate applications designed for your on-premises environment to the cloud, or from Hadoop to AWS or Google Cloud. If your cloud provider's pricing model changes, you can readily switch to a new cloud provider or run in a hybrid mode.

Because Ab Initio software was designed from the beginning to scale horizontally and vertically across an arbitrary number of servers, it naturally scales across clusters of pods and containers as well, enabling one processing job to run in parallel across multiple containers.

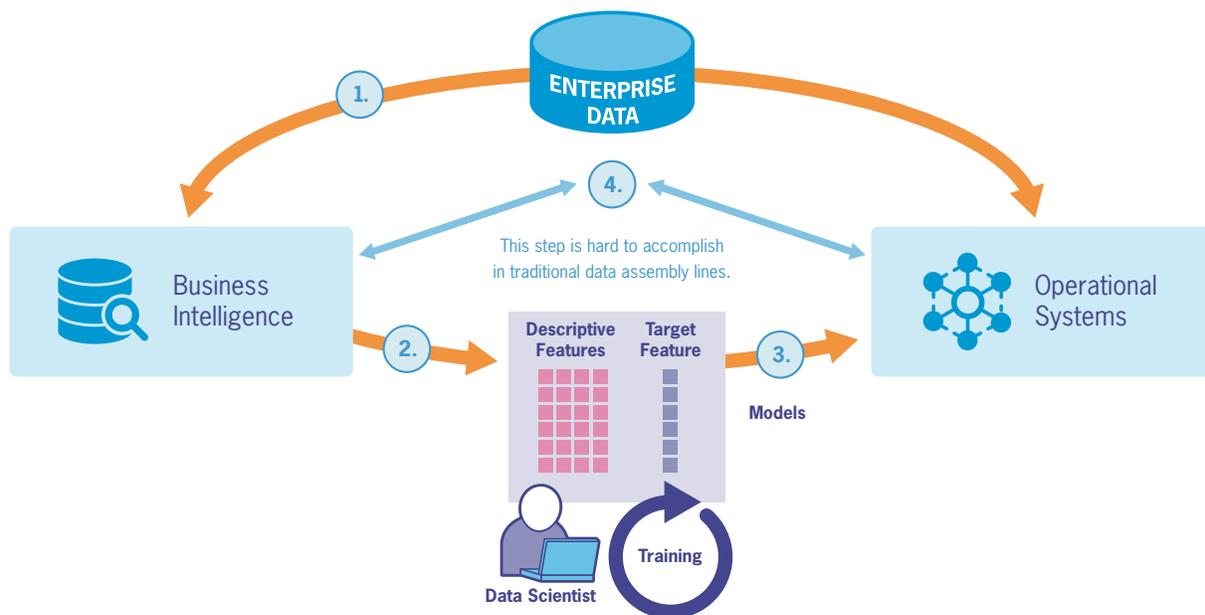
Data Agility

Data is only valuable if you can use it quickly enough to serve your customers, lead your markets, and accomplish your goals.

The diagram below illustrates a common “data assembly line” employed by many large enterprises.

1. Enterprise data is fed into Business Intelligence data repositories (data warehouses, data lakes, and so on). In these repositories, the data must be cleansed, integrated, governed, and enriched with meaning.
2. Data scientists and analysts access these Business Intelligence repositories for reporting and for creating and training their machine learning models.
3. These work products must be promoted to production and operationalized through the deployment of robust and highly performant real-time systems.
4. The results captured from operational systems are captured and fed back into the data assembly line to improve model performance.

Data Assembly Line



But the problem is that this “data assembly line” is extremely time-consuming and costly. Going from raw data to applications in production can take months or longer. With Ab Initio 4.0, instead of taking months, the process takes only days or even hours. Here are some of the ways we make this happen:

- Ab Initio’s new Data>Catalog technology automates the discovery, enrichment, and understanding of data sources across your organization. Application developers and data scientists can use Data>Catalog to directly access the right data from wherever it exists across the enterprise.
- Using the metadata held in the Data>Catalog, Ab Initio 4.0 accelerates application development — for example, by automatically generating data masking and data quality rules, or determining how best to join data and run applications in parallel.
- Ab Initio’s new Easy>Data™ technology makes it easy to quickly access and assemble data to build data repositories or wrangle data into feature tables needed for machine learning applications.
- Ab Initio 4.0 enhances our existing real-time capabilities with support for stateless and stateful services, using low latency, highly resilient in-memory technology.
- Ab Initio 4.0 enhances our already powerful data governance capabilities with extensive new automation and enhanced lineage capabilities.
- Ab Initio 4.0 introduces important testing automation capabilities for driving Continuous Integration/Continuous Deployment (CI/CD), further reducing time to production.
- Ab Initio 4.0 accelerates training machine learning models and putting them into production.

With Ab Initio 4.0, you can rapidly turn your data into actionable intelligence.

Discover
data sources

Access
the right data
at the right time

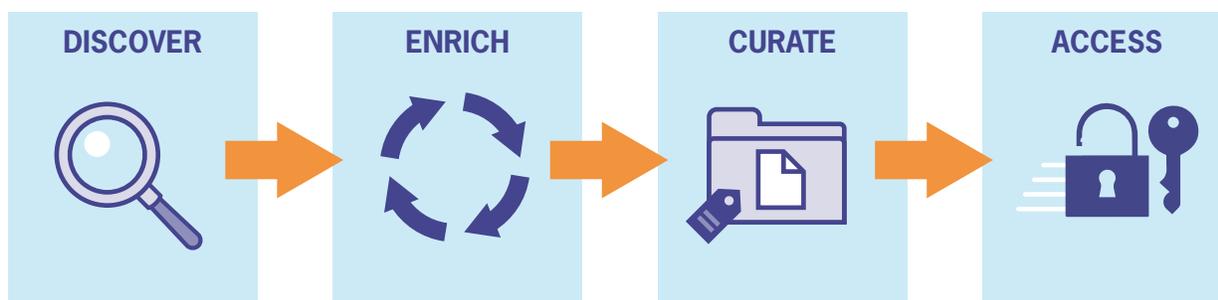
Resilient
real-time
services

Decreased time
to production

Understanding your data

Data>Catalog is key to the data assembly line. As the central clearinghouse for all data in your ecosystem, it provides a single portal to all your enterprise data. Data>Catalog accelerates the discovery, cleansing, integration, governance, and enrichment of datasets across your organization. As new data and metadata are generated, Data>Catalog makes that information available throughout the enterprise. With Data>Catalog, you can quickly identify data that is trusted and appropriate for use, significantly decreasing the time required for data-centric application development. You can easily perform semantic searches over metadata to find the most relevant datasets and then browse and filter the results.

4.0 Data>Catalog



- | | | | |
|--|--|--|---|
| <ul style="list-style-type: none"> • Discover enterprise data <ul style="list-style-type: none"> • Cloud • On Premises • Relational databases • Non-relational databases • Legacy databases | <ul style="list-style-type: none"> • Metadata management • Governance • Business glossaries • Lineage • Data quality • Data profiles | <ul style="list-style-type: none"> • Tagging / Hierarchies • Semantic searches • Browse / Filter • Navigate relationships • Collaborate | <ul style="list-style-type: none"> • Data Virtualization <ul style="list-style-type: none"> • ODBC • JDBC • Web Services • Ab Initio applications • Tools of your choice |
|--|--|--|---|



Adding meaning to the catalog

Ab Initio 4.0's patented Semantic Discovery technology, a component of Data>Catalog, accelerates the enrichment of the data assembly line. Semantic Discovery uses algorithmic/machine learning techniques to automatically discover and add business meaning to physical datasets.

With Semantic Discovery, organizations can analyze sets of data, identifying their contents through a highly automated process. It attributes meaning to new data and discovers the linkages and relationships among datasets.

Because Semantic Discovery processes datasets in bulk, you can use it to:

- Identify key relationships and map fields to business terms and domains.
- Identify Personally Identifiable Information (PII) within datasets.
- Discover fields that are not currently in your data dictionary, but that should be added.
- Discover the contents of data sources that are received with limited or no documentation (for example, old mainframe data). This includes understanding key relationships between two data sources.

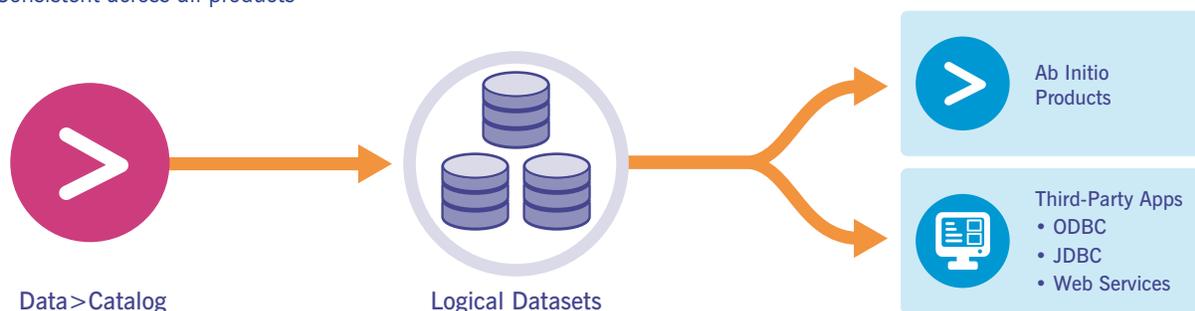
After Semantic Discovery analyzes data, the results are made available through Data>Catalog. The enriched data makes building Business Intelligence repositories easier and faster.

Simplifying data access

Data>Catalog enables data virtualization by abstracting out the physical location and underlying storage of your enterprise data assets and instead providing applications with access to logical datasets. Data scientists and application developers can immediately gain access to the right data wherever it is stored, accelerating the pace of the data assembly line.

Uniform definition of a dataset

Consistent across all products



Focusing on business logic

Ab Initio's Easy>Data gives business analysts, data scientists, and other domain experts the power to wrangle, process, and integrate data and prepare reports — without developer input. It automates and simplifies many of the technical details of application development by pulling in metadata held in Data>Catalog. You don't need to know how or where the data is stored; Data>Catalog automatically provides logical datasets and all the required metadata.

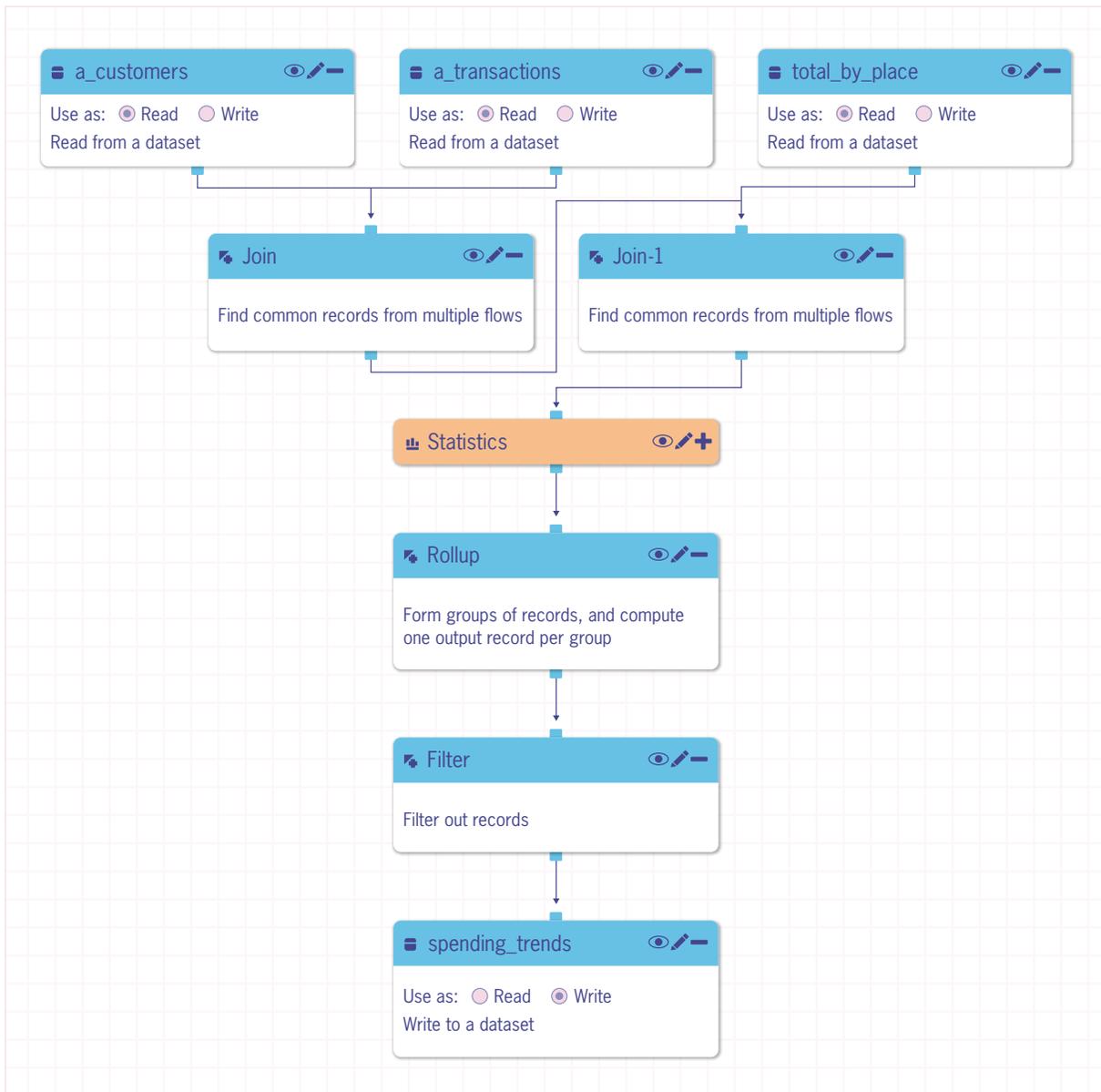
Working with datasets is as easy as dragging and dropping them onto the Easy>Data canvas. Data>Catalog gives you the information to join two datasets and auto-populates the key value. Easy>Data produces highly efficient applications — for example, it identifies and drops unused fields as early as possible during processing, thereby improving performance. It automatically decides if certain processing, such as a filter, could be performed faster using an underlying database. Efficiently generated applications allow you to promote analyst-developed code directly to production, without the need for the tuning or complete rewrites by developers or coding experts that characterize traditional code hand-offs.

Easy>Data includes powerful components that simplify common tasks. For example, with the Statistics component, you select which field you are interested in, and the component automatically computes a variety of useful summary statistics.

In Easy>Data, business logic is created using Ab Initio's Business Rules Environment®, enabling logic to be quickly specified and tested — again without developer involvement. Easy>Data and Business Rules Environment make it possible for nonprogrammers to build powerful applications in a fraction of the time — and cost — required for traditional development methodologies or open source.

With Easy>Data, you can focus on the business logic and let Ab Initio fill in the technical details.

Easy>Data graph



Enabling microservices, real-time services, and streaming services

Real-time systems are critical, whether you are protecting customers against fraud, maintaining up-to-the-second customer information, or conducting interactive marketing efforts.

Ab Initio processes data in real time with low latency and high throughput. Ab Initio can ingest data in any format and from any source — including change data capture and all common messaging systems — while providing horizontally scalable, reliable processing with exactly-once guarantees. With Ab Initio, data is never lost or duplicated.

Microservices written in Ab Initio scale horizontally, integrate with service mesh technologies, and support containerization, DevOps, OpenAPI, and other industry standards.

With Ab Initio's new Active>Data® technology, Ab Initio 4.0 adds high-performance state management to microservices. Active>Data lets you scale out your microservice architecture along with an equally scalable, resilient and performant stateful layer. Active>Data is built around a highly available, distributed, in-memory data repository. The repository can be accessed from any client or service to process data on demand. All data is replicated to multiple servers for high availability. Active>Data can be scaled out dynamically as the data volume or service load increases.

Active>Data combines the best features of an in-memory data grid and an in-memory compute grid to provide an exceptionally powerful real-time backing service.

Active>Data combines the best features of an in-memory data grid and an in-memory compute grid to provide an exceptionally powerful real-time backing service.

Governing data

Data that can't be trusted can't be used. If you don't know which data can be trusted and which cannot, then none of it can be trusted. It's not enough just to have data available: you must know where that data came from and what operations have been performed on it. If multiple copies of data exist, it's critical to know which version is the correct one. Trust in data is critical for adopting large data initiatives.

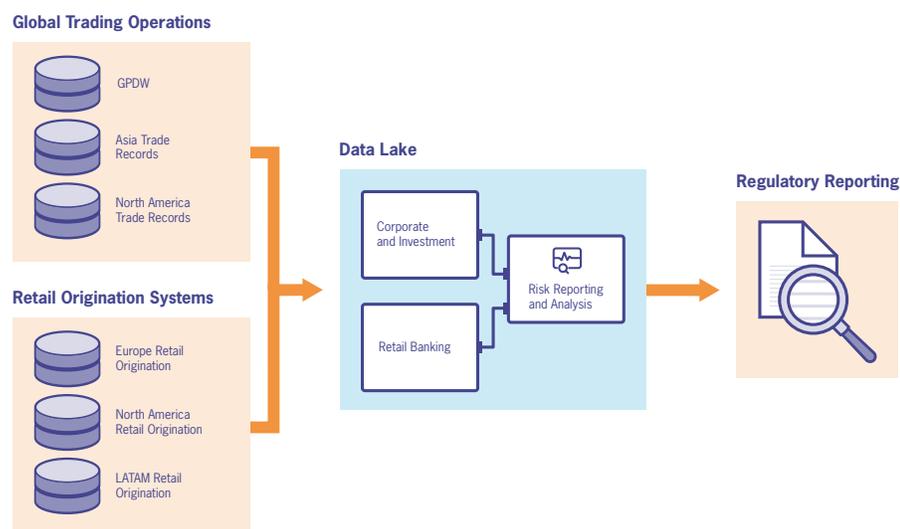
Data governance together with data quality is necessary to build that trust. Ab Initio 4.0 significantly extends our data governance capabilities.

Ab Initio technology can discover datasets and generate data lineage views across cloud, on-premises, and legacy sources, enabling compliance and impact analysis on source data assets. Data>Catalog datasets are integrated into the larger metadata management and governance system, simplifying governance tasks like assigning data stewards and identifying PII. Data quality rules can be automatically generated from the enriched metadata held in Data>Catalog.

Users with appropriate permissions can view data quality results, profiles, and virtualized access to physical data. Data specialists can find, vet, and curate data to be used in specific projects, campaigns, and applications. Curated data is tagged, shared, and made available by Data>Catalog for use in applications.

Ab Initio 4.0 uses a combination of algorithms and machine learning to detect data anomalies and provides extensive automation capabilities to generate data quality controls at appropriate control points. We have augmented our technical data lineage capabilities to support business-friendly views of data lineage, giving the business greater transparency and control of data processing. Data and metadata are up to date and at your fingertips. You always know what data you have. With Ab Initio, you can trust your data.

Data Lineage



Accelerating testing

Long development and test cycles kill progress. Even the best ideas get bogged down with missing test data or failing tests. Ab Initio 4.0 accelerates every aspect of testing, from generating test data to test automation.

Ab Initio 4.0 makes it easy to generate test data. Semantic Discovery can automatically discover PII and, using information stored in Data>Catalog, Ab Initio 4.0 can automatically generate masking rules to obscure sensitive data (for example, credit card numbers). Further, it can create minimal subsets of test data that enable efficient yet complete testing of a piece of business logic.

Ab Initio makes it easy to capture, reuse, and update test data; to author tests; and to run tests in an automated way. You can generate unit tests from a known good run of an application. Ab Initio 4.0 integrates with Jenkins to provide extensive test automation capabilities. You can orchestrate the running of a full test suite, engage in continuous regression testing, and implement Continuous Integration/Continuous Deployment pipelines.

With Ab Initio 4.0's testing features, problems are identified sooner, enabling them to be fixed earlier in the cycle. Development time is reduced, and revisions are easier and significantly faster.

Projects come in on time and on budget — and organizations no longer suffer the opportunity costs of nonfunctional software.

Ab Initio makes it easy to capture, reuse, and update test data; to author tests; and to run tests in an automated way. You can generate unit tests from a known good run of an application. Ab Initio 4.0 integrates with Jenkins to provide extensive test automation capabilities. You can orchestrate the running of a full test suite, engage in continuous regression testing, and implement Continuous Integration/Continuous Deployment pipelines.

Empowering data science

Ab Initio 4.0 enables you to rapidly move machine learning models from concept to deployment. Ab Initio makes it easy to catalog your data, cleanse it, identify appropriate data subsets, train models, and put the trained models into production: the data assembly line becomes an artificial intelligence assembly line.

Much of the time spent in training a model is actually spent preparing data. Within the artificial intelligence assembly line, Data>Catalog makes it easy to find and access appropriate data and metadata from across the enterprise. Easy>Data simplifies turning that data into feature sets for use in training and validating AI models.

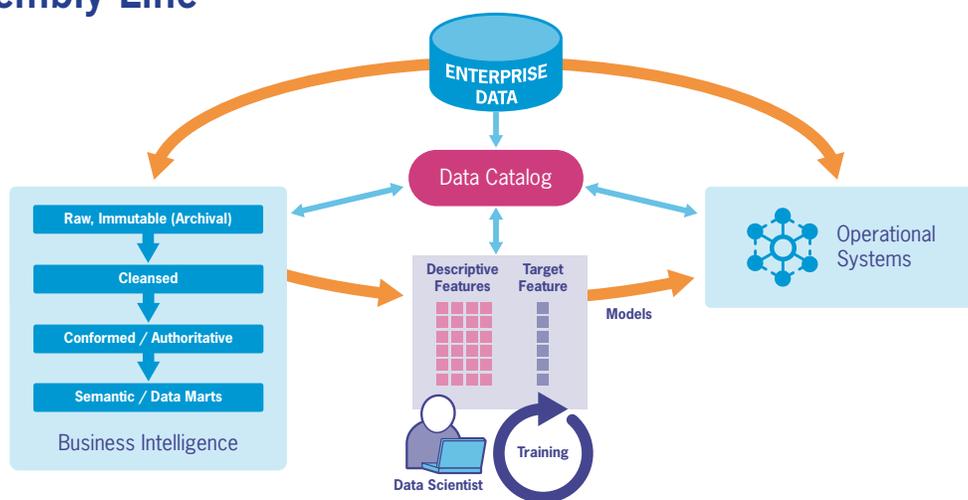
Through its support for PMML, Ab Initio software directly supports data science models such as regression, clustering, trees and random forests, and neural networks. You can export data science models in PMML, import them into Ab Initio, and take advantage of Ab Initio's performance, scalability, and platform independence in creating your operational system.

In an Ab Initio application, you can call pre-existing R and Python function libraries through external packages. This feature allows you to use non-Ab Initio language functions as if they were written in Ab Initio. You can do all pre- and post-processing of data in Ab Initio.

Ab Initio's extensive support for the development of robust and high performance real-time systems enables models to be rapidly deployed and operationalized.

With Ab Initio 4.0, you can maximize the value of your investments in AI and machine learning.

AI Assembly Line



Innovative Applications

Ab Initio 4.0 includes two new business-oriented applications that address common problems: customer experience management and entity resolution. Each application is built on core Ab Initio technology.

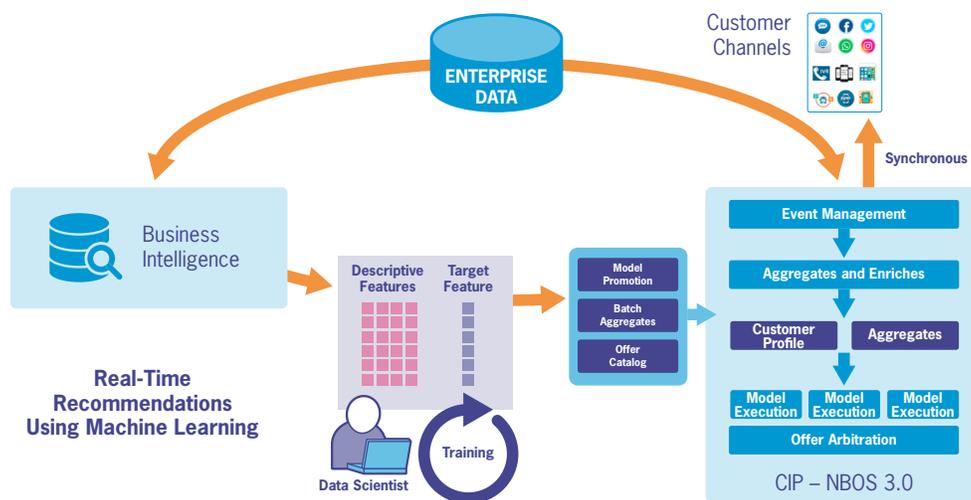
Customer experience

Successful businesses meet their customers *where the customers are* — in context, in the customer’s moment of need — connecting at every point in the customer’s journey. The goal is not just to respond but to anticipate, not just to serve customers but to surprise them.

Digitally connected brands flourish when they distinguish themselves with a compelling customer experience flawlessly delivered at speed and at scale. Ab Initio’s new Customer Interaction Platform (CIP) makes that responsiveness possible. CIP orchestrates interactions with millions of customers, generating billions of events, every millisecond of every day. CIP focuses on:

- **Data** — CIP turns insight into action, processing customer data from every possible source, at speed and at scale, providing the foundation for meaningful engagement.
- **Decisions** — CIP uses intelligent, end-to-end automation to enrich the customer experience by processing complex inputs and brokering complex choices, making smart decisions at moments that matter.
- **Dialog** — CIP maintains a personal dialog throughout the customer journey, rewarding the customer’s investment of time, trust, and money — in real time across every point of contact.

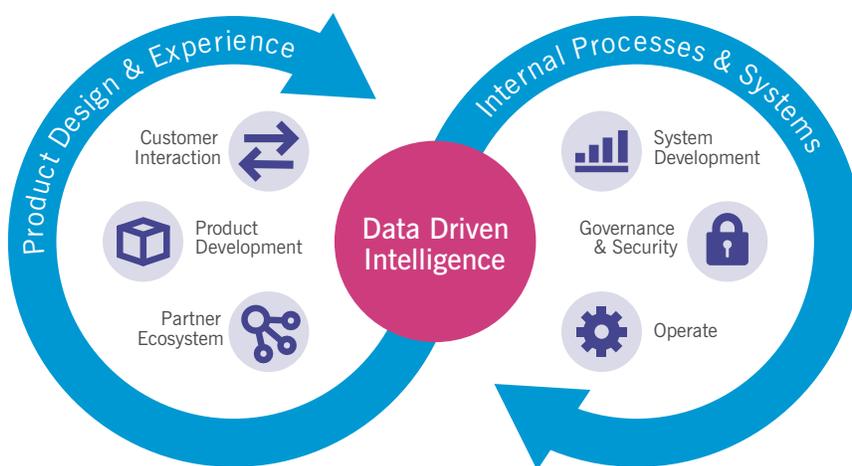
CIP’s flexible and intuitive user interface allows marketing and customer service teams to define and configure responsive and relevant communications, from location-based offers to event-triggered alerts; from personally priced incentives to tailored notifications. Because the platform maximizes flexibility, it empowers the business to rapidly adjust to changing market conditions. As the graphic below shows, CIP is an example of an operational system that you can build in Ab Initio 4.0.



Taking the Lead in the Digital Economy

Ab Initio 4.0 brings both power and agility to businesses seeking to be leaders in the digital economy. It accelerates software development with advanced cloud capabilities, data automation and agility improvements, data virtualization, machine learning, real-time and streaming applications, and extensive data governance capabilities. Ab Initio's Customer Interaction Platform gives businesses unprecedented control over customer interactions, including the ability to interact in real-time with millions of customers generating billions of events daily. And Co>Relate provides comprehensive, tunable entity matching and data disambiguation.

With our 4.0 release, businesses can take full advantage of cloud technologies, enjoy improvements in business agility and productivity, and rapidly transform data into actionable intelligence.



Because of Ab Initio, we no longer wonder *if* something is possible, but *what* is possible — we are no longer limited by technology.

— VP Architecture, Ab Initio Customer

About Ab Initio

Ab Initio works with the largest and most sophisticated organizations in financial services, telecommunications, healthcare, retail, high-tech, transportation, manufacturing, and government, among others, to assure their business success.

Ab Initio products are designed from the beginning to provide a single, cohesive technology platform for scalable, high performance data processing, integration, and governance.

Ab Initio has offices in Lexington, London, Paris, Munich, Warsaw, Rome, Istanbul, Johannesburg, Singapore, Tokyo, and Sydney. Ab Initio software is transforming the way large institutions manage and process their data.

Contact us to learn more.



To find out how Ab Initio can help you lead in the digital economy, contact us.

Ab Initio

201 Spring Street
Lexington, MA 02421
USA

T +1 781-301-2000

F +1 781-301-2001

Ab Initio is growing around the world. Find your regional office at www.abinitio.com
or send inquiries to solutions@abinitio.com

Copyright ©2020 Ab Initio. All rights reserved. Reproduction, adaptation or translation without prior written permission is prohibited, except as allowed under copyright law or license from Ab Initio. Ab Initio®, >®, Acquire>It™, Active>Data®, BRE®, Business Rules Environment®, Co>Operating System®, Co>Relate®, Conduct>It®, Control>Center®, Continuous>Flows®, Data Quality Assessment™, Easy>Data™, EME®, Enterprise Meta>Environment®, Express>It®, GDE®, Graphical Development Environment™, Metadata Hub®, Metadata Portal®, Query>It® are trademarks or registered trademarks of Ab Initio.