Ab Initio in Enterprise Application Integration

The recent pace of technological and environment changes have led to massive structural as well as operational changes as to how the organization is managed and how it interacts with customers in a highly competitive market economy. The main factors contributing to change are Mergers & Acquisitions, business expansion, globalization, and a growing and diverse customer base. To meet these changes companies need to implement increasingly complex software architecture, with much enhanced capability than the legacy systems, with the help of Enterprise Application Integration (EAI) solutions.
About the Authors

Bikram Das
Bikram is a senior consultant with TCS. He has been associated with BFSI, Telecom and Healthcare industry verticals. He has rich experience in database administration, performance tuning and in providing solutions for Data Warehouse ETL environments. He currently heads TCS' Ab Initio Center of Excellence.

Arnab Das
Arnab is a consultant with TCS BI practice. He has been involved in the execution of various BI projects and has earned kudos from customers for his value adding suggestions and creative solutions. He is currently a consultant for Ab Initio based solutions.
Table of Contents

1. Introduction 3
2. Application Integration Challenges 3
3. Ab Initio Capabilities 3
4. Ab Initio Product Components 4
5. Application Integration Implementation Using Ab Initio 5
Introduction
Information Technology systems have grown exponentially over the past few decades. The revolutions in technology and variation in environment have led to immense structural and operational changes in the way organizations are managed and the way they interact with the customers in the highly competitive market economy. The primary factors contributing to the change are mergers, acquisitions, business expansion, globalization, and a diversified and expanding customer base. To meet these changes, companies need to implement increasingly complex software architecture with much enhanced capability than the legacy systems.

The shift of focus toward tangible result-oriented customer relationship management projects is further highlighting the need for enterprise information integration capabilities as a key factor in reducing the time to market and the cost of ownership for customer-centric applications.

The best-of-breed software architecture for integrating disparate and distributed systems is Enterprise Application Integration (EAI). It essentially performs the following functions:
- Reduces the complexity of an organization’s IT framework
- Provides a step improvement in its efficacy, flexibility and scalability
- Supports new organizational initiatives.

Application Integration Challenges
The major challenges in implementing a suitable IT solution in this varying world of business operations are that the system must be able to do the following:
- Provide cost-effective integrated systems across the enterprise along with reduced time cycle for implementation.
- Adapt to the complexity of changing business models and customer demand.
- Achieve faster return on investment in integration of business-to-business and business-to-consumer systems.
- Establish an architectural framework that rapidly develops integration and reusability of applications.
- Remove the redundancy of applications resulting from mergers and acquisitions.

Ab Initio Capabilities
Ab Initio, classically an ETL vendor, has gradually emerged as a strong player in the application integration spectrum with its wide range of enterprise scale, mission-critical applications that include:
- data warehousing (ETL)
- real-time analytics
- customer relationship management (CRM)
- enterprise application integration (EAI)

Ab Initio provides a robust architecture that would allow simple, fast, and highly secure integration of systems and applications. It can run heterogeneously with parallel execution over distributed networks. It can integrate diverse, complex, and continuous data streams ranging in size from several gigabytes to tens of terabyte, providing both ETL (Extraction, Transformation and Loading) and EAI (Enterprise Application Integration) tasks within a single, consistent framework.
Organizations of all sizes – in industries as diverse as telecommunications, finance, insurance, health care, e-commerce, retail, and shipping – use Ab Initio software to manage their most complex data. The power and flexibility that make Ab Initio so effective at the enterprise level also empower smaller, departmental organizations.

## Ab Initio Product Components

The Ab Initio product portfolio consists of three core building blocks:

- **The CO>Operating System** – This system delivers distributed and parallel execution and platform-independent data transportation. It can run across a huge variety of operating systems and hardware platforms, from OS/390 on mainframes, to the different flavors of UNIX, to Windows NT or 2000.

- **The Component Library** – This is a set of reusable software modules that can perform the basic ETL functions as load/unload from different sources, data sorting, data transformation etc, including complicated tasks like checksum computation, XML reformation, multi stage aggregation, etc.

- **The Graphical Development Environment** – This is a development environment to create Ab Initio applications with drag & drop facility, connecting the components into executable flowcharts.

A major capability of Ab Initio is its ability of parallelism. It can distribute data load into heterogeneous systems and can execute a transformation on the entire volume of data simultaneously.

Ab Initio also provides a user friendly and effective mechanism for real time or near real time data processing. The various components in Ab Initio that cater for the real time data mechanics are:

- **Ab Initio Queues**, an adaptation of the FIFO queue concept, are a reliable method for storing continuous flow data.

- **Subscriber** is a component that writes data from various sources into a continuous flow process, and originates computepoints (the marking of blocks of records that are to be processed as groups) and checkpoints (intermediate states –wherein data processing can be restored from this point).

- **Publisher** is a component that writes data to various destinations, and consumes computepoints and checkpoints.

Ab Initio queues provide a method for storing records in an ordered sequence of files that can be fed to the ETL process, which runs for an indefinite period of times, continually taking in new input and producing new, usable output.
Application Integration Implementation Using Ab Initio

The following section discusses an application integration process facilitated by Ab Initio. The diagram below describes the general design architecture followed to integrate data from disparate sources into an Enterprise Data Warehouse and loaded into target CRM application.

Figure 1: Implementation Architecture

The major challenges in such an implementation are:

- **Multiple Sources** - Data from disparate sources like Mainframe /Oracle table using different technologies, data formats, with different data load frequencies.

- **Complex Business Logic** – Achieving common data format aligned with the target systems, generic entities, and data cleansing requirements.

- **Redundancy** – Multiple truth of source data due to data duplication.

A cost-effective solution can be provided using the Ab Initio batch or real time (continuous flow) execution mechanism. A scalable solution that extracts data from distributed and disparate systems, transforms multiple format data into common format, creates Data Warehouse, operational data stores, aggregations/derivations for Business intelligence, and loads data into target systems can also be provided.
The following schematic diagram explains the solution to implement the above system architecture.

**Figure 2**: A Design Option for Ab Initio in Application Integration

In this design, data from different sources are loaded into the data reception area (DRA) in periodic batch execution as well as near real time data flow using MQ series/AI queue. The DRA can handle multiple source data in different data format. The data is then transported to the data staging area where the data is converted to common format data. The generic loader or extractor process is Ab Initio based applications that can perform miscellaneous functions:

- Load data into operational data stores
- Use metadata driven rules engine to generate code
- Provide PAI facility to perform Query
- Interface with database for data extraction and data load
- Provide delta and before-after image of data
- Feeds target system/reporting tools or message queue
About Business Intelligence Practice
The Business Intelligence (BI) Practice delivers solutions spanning across industry verticals and technologies to a majority of Global Fortune 500 companies under the umbrella of its proprietary BI Solutions BIDS™ (Business Intelligence for Decision Support). It has delivered 200+ projects in the last 3 years and has a consultant base of 2200 (as of March 2005).

The BI Practice has developed solutions for Banking, Insurance, BFSI, Telecom, Manufacturing and Retail domains. It has mature Technology Centers of Excellence and Alliances with all major BI tool vendors.

About Tata Consultancy Services
Tata Consultancy Services (TCS) is among the leading global information technology consulting, services and business process outsourcing organizations. Pioneer of the flexible global delivery model for IT services that enables organizations to operate more efficiently and produce more value, TCS focuses on delivering technology led business solutions to its international customers across varied industries.

For more information contact
Ranjita Saran
Tata Consultancy Services Ltd
Akruti Business Port
Road No 13, MIDC
Andheri (East)
Mumbai 400096
India

Phone: +91-22-5550 6868
Fax: +91-22-5660 6855

Email: ranjita.saran@tcs.com
Website : www.tcs.com/datawarehouse