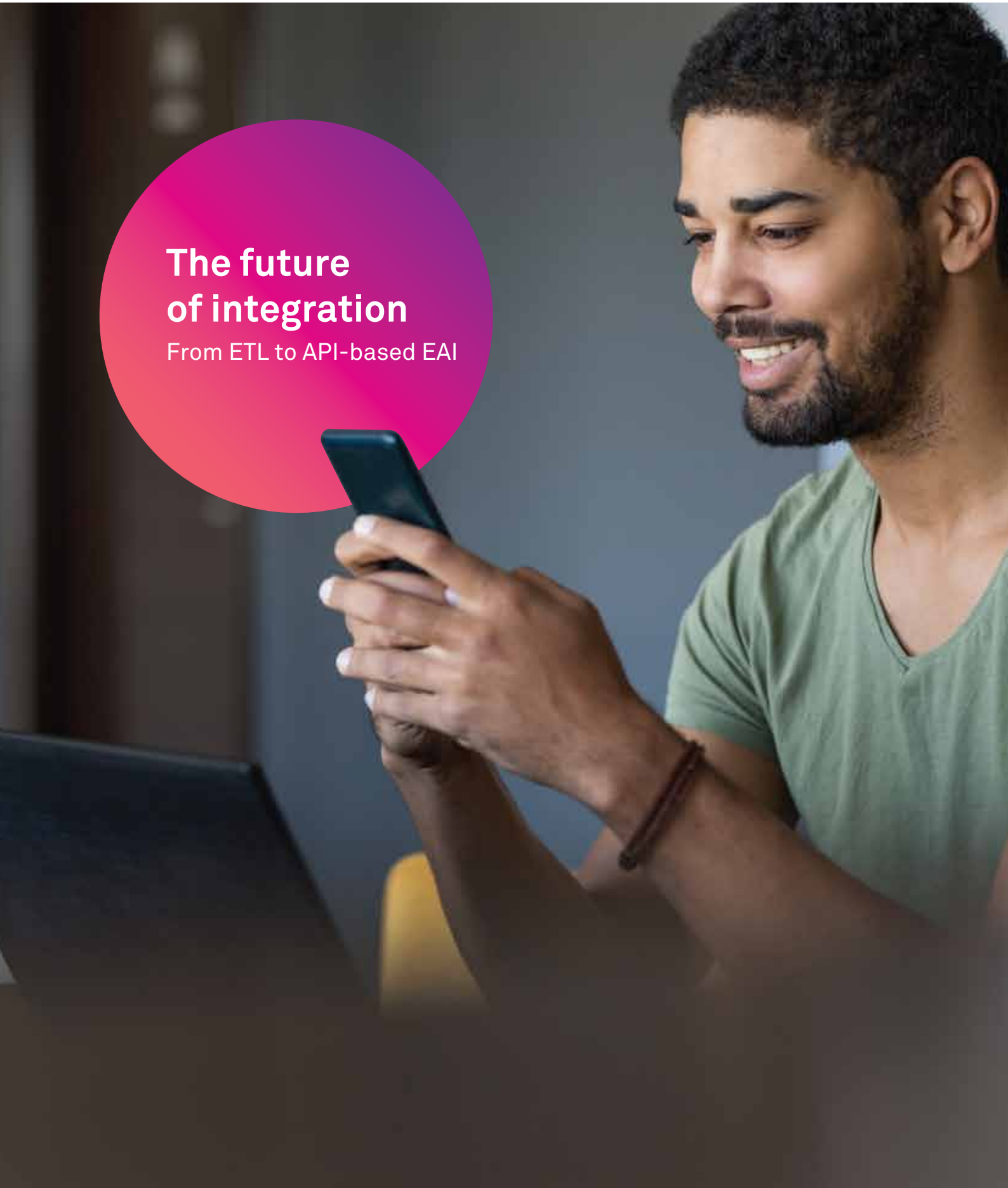

The future of integration

From ETL to API-based EAI

The Web API or API in short, is not a new technology. It has been there for more than a decade and today Programmable-Web directory lists more than 20,000 public APIs available for consumption. Transactions over billions of dollars are executed successfully in this technology every day and the volume is growing every passing quarter.

Though initially developed with system integration over HTTP protocol in mind, the recent developments have spilled over to various other peripheral areas. The growth of API and related technologies are influencing the established market players in these areas to modify their strategy to incorporate API in their product offerings, if not completely re-doing their product strategy around API driven architecture.

In this article, we will explore how the traditional on premise data integration approach is changing towards API based integration of both on-premise and cloud based systems.

The integration battleground

System integration is not a new concept. Humankind have discovered the power of collaboration in prehistoric ages and utilized the concept in every domain since, only constrained by the available technology of that point in time.

Divide and conquer strategy ensured that the systems continued to be developed independent of each other, while business survival compulsions forced them to get connected more and more. In the process two broad categories of high-level integration emerged—application integration and data integration.

Applications started talking to each, mostly driven by the usage requirement—either asynchronously via messaging or synchronously via HTTP and other online protocols. The availability of bandwidth and processing power constrained them to use minimum amount of data for processing. The predecessors of current

Taking advantage of increased security over public network, cheaper cost of connectivity and larger on-demand capacity of cloud, Integration platform as a service is steadily gaining ground over traditional on premise ETL tools

day web APIs trace their root in this type of Enterprise Application Integration strategy.

Figure 1 gives the relative advantages of each strategy - ETL vs EAI.

Operations	ETL Tools	EAI Tools
Data migration: high volumes of data, diverse schemas, formats, validation needs	Strongly supported	Not a key strength. APIs have bulk data handling capabilities so this is app dependent. File transfer is possible but lacks high volume data processing capability
Diverse applications: Mainframes, Cloud apps, on premise data warehouses etc.	Strongly supported. Provide the ability to connect to apps via latter's SDK/APIs as well as can leverage native DBC/ ODBC drivers to directly connect to the sources. Can support legacy systems like VSAM files and mainframes	Can support as long as the system in question has an API to integrate with. Ill-suited for legacy systems.
Data warehousing	Strong fit	Good fit for real time data updates for key operations/ transactions
Data quality and validation	Strong fit	Suitable for real time low data but synchronous operations
Real time semantics	Traditionally, had poor support, however today ETL platforms do have substantial real time data sync and integration capabilities	They are very good at orchestrating real time APIs as well as complex workflows
Security	Can connect to schemas and actually move data through network so the security risks are much higher depending on how ETL pipelines are implemented	EAI tools, when using app provisioned APIs, are authorized by the respective app, so data access is controlled by the app

Figure 1: Relative advantages of ETL and EAI

Development drivers for API based integration

There are six major drivers pushing towards the convergence of application and data integration via API.

Cheap bandwidth: Advances in fiber optics are enabling wires to carry terabytes of data at the speed of light, whereas 4G (and imminent arrival of 5G) are sending gigabytes over ether to mobile devices. The abundant supply is driving down the effective cost of bandwidth to the point of insignificance.

Sensors, systems and data explosion: Increasingly, we are living in a connected world, filled with various sensors and systems, generating enormous amounts of machine data. The maturity of a global IT manufacturing supply chain has lowered the cost of IT Hardware significantly, allowing collecting, collating, processing and making sense out of this humongous data load via artificial intelligence, completely removing the human bottlenecks in the process.

Privacy and security: Deployment of cheap biometric devices for identification by various government organizations worldwide have necessitated foolproof data protection legal framework. The advent of social network has sensitized the masses about the danger in lack and/or loss of privacy.

Cloud—private, public and hybrid: While on-premise systems are not going away anytime soon, every organization worth its name are spending significant amounts of effort on their hybrid cloud strategy to cut down their operational costs. Other than optimizing idle capacity, the 'pay as you go' model of cloud is inherently more suitable for the business, driving towards its adoption.

Handheld devices everywhere: Cheap bandwidth combined with significant processing power available in handheld devices have increased the appetite for real time data of the data consumers. The information latency of traditional batch processes is no longer acceptable. Enriched data must appear for user consumption as soon as the event occurs and tracked till the logical closure in all four space-time dimensions.

Everything has an app for it: The data consumer is ready to pay for the on-demand service, giving rise to the 'subscription' model of instant revenue generation for the data. Taking advantage of this, an ecosystem of connected apps has come up, providing every kind of service at every corner of the world. The data they use are not always their own. The interaction patterns and the data generated by them changes rapidly. This necessitates a different kind of agile tool that is lightweight, fault tolerant and horizontally scalable.

IPaaS: The wholesome integration

In the last few years, Integration Platform as a Service has gained significant traction. Taking advantage of increased security over public networks, cheaper cost of connectivity and larger on-demand capacity of cloud, they are steadily gaining ground over traditional on-premise ETL tools.

The current global revenue has already crossed a billion[1] and can only go up since all the traditional ETL vendors are increasingly focusing on this space. Using their existing customer base, existing vendors have either retained strong presence or regained the initial lost ground. Simultaneously new players have come up with their own niche. There is still entropy in the market and it will go through a consolidation phase in future but today's business can ill afford to wait it out.

A. The first advantage the IPaaS tools have is their dual capability of API connectivity and data processing. Today's SaaS offerings are not simple. Take for example ServiceNow, SalesForce, Ariba or Office365; each provides significant capability at the cost of a very complex and complicated API. They also need to transfer huge amount of data back and forth. With the deep inroads by them in every domain, an organization has a serious business disadvantage if it is not ready to consume the relevant ones in its own business process.

However, the design and development of an API, its gateway or even the development of a public or B2B API client in bespoke manner is not simple and can quickly turn into an expensive and time-consuming proposition completely outweighing its benefit. Another consideration is the adaptation of own API by the outside world. In case of no or low adaptation, the entire expense goes out of the window without any return.

Given the diverse nature of the public and B2B APIs, the tools already have significant capability of transforming data. Traditionally there has been three patterns of data transform which every ETL tool excel in:

a. Movement with enrichment from one system to another

b. Synchronization of data in multiple systems - either uni or bidirectionally and

c. Aggregation from many systems into one

B. The second advantage of IPaaS tools are their tight integration with high performing messaging platforms. While the traditional ETL tools have developed connectors for them, the cloud native IPaaS tools have distinct advantage in streaming mode for better near real time asynchronous performance. The combination of high throughput and low latency allows creation and deployment of new architectures like hybrid transactional analytical platform.

C. Typically, the on premise ETL tools have a perpetual license model with upfront payments connected to number of CPUs. The IPaaS, on the other hand, typically uses a subscription model tied to actual usage. This gives optimized cost of capability and business prefers this model more.

Conclusion

The days of ETL are far from over. The current customer base is still an order of magnitude more than the nearest IPaaS offering. Nevertheless, there is no denying that IPaaS is steadily gaining ground and is the way to the future. Given such a scenario, the prudent approach would be to retain the existing ETL solutions but simultaneously prepare a 'to-be' state involving an IPaaS tool. For any new development, a due diligence should be done based on the following key considerations:

- Data volume to move: Both one time as well as ongoing
- Complexity of data formats and transformations needed
- Number of real time operations between two or more systems
- Nature of systems/data sources to integrate: Are all participating systems API based? Or are some legacy and only allow data access via files and SQL queries?

Reference

[1]https://www.gartner.com/doc/3885665/market-share-analysis-integration-platform

About the authors

Swagata De Khan
Senior Architect - Data, Analytics & AI,
Wipro Limited.

Swagata has over 19 years of data warehouse experience, and has successfully executed large engagements for global MNCs. He is an AWS certified Solution Architect and currently focuses on solutions involving Cloud, Integration Technologies, Artificial Intelligence and Machine Learning.

Prasanna Raghavendran
Head of Data Integration - Data, Analytics & AI,
Wipro Limited.

Prasanna has over 18 years of experience in the IT industry and is currently leading the enterprise data integration practice in Wipro globally. He has performed several leadership roles in the area of data & analytics including sales, pre-sales, consulting and portfolio management

Wipro Limited
Doddakannelli, Sarjapur Road,
Bangalore-560 035,
India

Tel: +91 (80) 2844 0011
Fax: +91 (80) 2844 0256
wipro.com

Wipro Limited (NYSE: WIT, BSE: 507685, NSE: WIPRO) is a leading global information technology, consulting and business process services company. We harness the power of cognitive computing, hyper-automation, robotics, cloud, analytics and emerging technologies to help our clients adapt to the digital world and make them successful. A company recognized globally for its comprehensive portfolio of services, strong commitment to sustainability and good corporate citizenship, we have over 175,000 dedicated employees serving clients across six continents. Together, we discover ideas and connect the dots to build a better and a bold new future.

For more information, please write to us at **info@wipro.com**

IND/B&T/MAR-AUG2019