# datawave
**An HCL Technologies Company**

Orchestrate. Automate. Transform.

The State of iPaaS and

# where Datawave

sits within it

There's a widespread view that although existing on-premises, large-scale data integration developed by integration specialists or configured for application-to-application integration is not going away any time soon; adoption of Cloud technology is the direction of travel.

This change is fuelled by an increasing appetite among companies to take advantage of low-cost cloud storage and processing power when building new applications and, when the opportunity arises, overhauling existing ones.

## Key facts

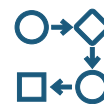Pattern-based code generation and separate job control framework

Data Integration for both operational and analytical purposes

Support for on-premises & cloud-based data store connectivity

Integrated metadata support to govern Data stores and processes

Orchestration of workflows to support data and application integration

www.datawave.io

**Orchestrate. Automate. Transform.**

# The Three Triggers for Change

## 1. Volume of Data

The 'State of Data' in terms of the platforms that it is stored on has therefore become significantly more complex. At the same time, the volume and rate of growth of data has increased significantly. New technologies have enabled support of new Data Analytics technologies that exploit distributed processing modes and larger different modes of data storage.

## 2. New Technology Platforms

In turn, the technology platforms used to source, transform and locate data have also changed the ways in which data integration processes are built. The impact of this is to increase the proportion of manual code development that has always existed in the ETL space. This in turn has increased the importance of API's. The ease of connectivity provided by ETL tools is not available to hand-coded processes and complex data-type specific calls within the hand-coded applications are costly to build. These frequently need to be refactored because new storage technologies can adapt and change frequently.
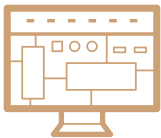
## 3. Primary Source of Data

The third shift in the data space is a re-shaping of what a traditional DWH or EDW is for. Instead of being the primary foundation for analysis and reporting, they have gradually moved to become the primary source of data for regulatory, structured and company-wide reporting. Ad hoc analysis (Data discovery, data exploration etc.) and some line-of-business reporting has shifted to use relatively unstructured, high-volume data stores, (referred to Data Lakes, Reservoirs and so on depending on the amount of data curation involved) in their assembly and ongoing support.

Alongside and sometimes over-lapping with ad hoc analysis, there is now a new class of 'citizen integrator'. Their role is to create new analytical insights as part of a business or analytics function using correlation-driven methods within the organisation. This is far removed from the structured data models on which DWH and EDWs are largely built.

These three drivers are transforming the data landscape. Data is now stored and maintained on new technology platforms that previously included only on-premises RDMS and file systems.

These new and existing data stores are growing faster than ever before and demand for data has increased.

It is being driven by a democratisation of business analytics to both shape and respond to market conditions and customer demand.

The bank's data-driven re-thinking of how they communicate with their customer base is an example of how data volumes, modes of access and availability support analytics applications that drive business change.

Orchestrate. Automate. Transform.

Traditional ETL companies and megavendors such as Oracle or IBM recognise that having a software suite made up of products that are either Cloud-only or on-Premises-only does not meet these emerging hybrid data needs. They are all trying to simplify their software stack to address the customer perception that they are too complex and offer poor integration between products.

Integrated Platform as a Service (iPaaS) is the market's attempt to harmonise product support for this new data landscape. The number of players in this area is growing daily. There are up to ten ETL specialists and megavendors offering enterprise-level capabilities, at least twenty smaller players offering enterprise support to some degree, and at least another fifty who provide niche support.

## Expectations of an iPaaS

In order to compare the Datawave platform to the trends and competition in the marketplace then it is helpful to look at what an Integration Platform as a Service (iPaaS) typically offers. Functional support for some or all of the topics below is expected from an enterprise-level iPaaS:

Those types of iPaaS software that offer the highest levels of functionality often also support:

Process and Application integration – for example, Informatica's mature support for Salesforce to ServiceNow for bi-directional data interchange of Salesforce cases, users, accounts, opportunities, etc. with ServiceNow

Data Integration for both operational and analytical purposes:
- Support for cloud-based data store connectivity e.g. cloud-based Hadoop
- Support for on-premises data store connectivity e.g. operational applications and transactional systems
- Data integration for Cloud and On-premises DWH & EDWs including:
  - Support for data sources in multiple clouds
  - Data integration to populate Cloud and On-premises MDM systems
  - Data ingestion & integration to prepare data for analysis on Cloud and On-premises HDFS and RDBMS systems
  - Scalable streaming data ingestion into cloud data stores from on-premises and other cloud systems

Integrated metadata support to govern Data stores and processes

An information catalogue that publishes data, processes and patterns as shared, reusable assets that can be used in any context

Productivity and simplification via patterns that deploy when connected

Orchestration of workflows to support data and application integration

Push-down execution in both legacy and current modes – e.g. PowerCenter & SPARK

Range of pre-built connectors (APIs)

Developer APIs to use services that are provisioned on the iPaaS

Migration of existing ETL processes on-premises to the same vendor's Cloud platform

File transfer between cloud and on-premises systems

Metadata cataloguing capability across technologies to locate & identify data assets

Data and API security, privacy, auditing and monitoring ideally using a policy-based approach

Dynamic auto-scaling based upon user / admin-defined policies

Hub-based approach that decouples the data sources from how they are integrated with each other and the targets that these sources and integrations populate in favour of a publish and subscribe model that captures sources once and integrates and consumes in one or more distinct use cases

Orchestrate. Automate. Transform.

The ETL Specialist and Megavendor providers include Informatica, IBM, Oracle, Microsoft, Dell Boomi, Mulesoft, Snaplogic, Fujitsu, SAP and Talend; the niche players include Jitterbit, Adaptris and Actian.

The niche players are predominantly cloud-focused; where they do provide support for on-premises applications then they tend to do so in an industry vertical context such as healthcare or food, see Adaptris for example. For the purposes of this paper we can discount these niche players as competitors because Datawave is not primarily focused on the Cloud at the present time largely because our track record and user-base are mainly in financial services where cloud adoption is proceeding at a slower pace. This could change in the future, of course.

## Does Datawave's platform qualify as an iPaaS?

Datawave's new platform provides an abstraction and orchestration layer over the underlying technologies that it supports; currently these are PowerCenter (Classic only), Datastage (limited support) and SPARK (batch mode). For this reason, the platform qualifies as an iPaaS. The concept of 'service' is taken further by companies like Informatica to incorporate licensing on a pay-per-use and capacity models; platform hosting by Informatica; dedicated technical and account support, etc.

At the present time Datawave do not envisage the platform primarily as a stand-alone service so those aspects of service provision have not been considered here. Instead the Datawave iPaaS platform is likely to be marketed as part of a services package with the option for the customer to maintain the use of the IPaaS software after the services engagement comes to an end on a subscription license basis.

## The Competition

For this first draft of our paper we will compare Datawave only to Informatica's Intelligent Cloud Services (IICS) which is the winter release of their iPaaS platform. Future drafts of this paper can widen the comparison net but since Informatica is the vendor that is best known to Datawave and provides the widest set of integrated capabilities (at a significant cost) then it provides a useful point of comparison.

However, before considering a feature-by-feature comparison, let's look at how Informatica stacks up against its competition. Informatica has stolen a march on the competition by going all out to integrate the Data Integration, Master Data Management, Data Quality & Governance and Data Security products into IICS. Their competitors suffer by either being slower to integrate products onto one platform – for example IBM has a disjointed offering: IBM App Connect Professional for ad hoc and citizen integration and IBM Integration Bus on Cloud (based on Datastage) as its high-control iPaaS component for the specialist integrator. Other parts of the competition are seen as too focused on their own ecosystem: see SAP and Microsoft as examples. The current competition for Informatica comes from Dell Boomi, Snaplogic, Mulesoft and Oracle; each has their strengths and they have all made inroads into the Informatica customer base by either offering better value for money or a simpler product stack.

Both of these factors are worth considering in terms of how the DataWave product is positioned with a prospective customer.

## Informatica's 'Cloud Journey'

Informatica's IICS platform provides the capability to migrate existing PowerCenter Classic applications to IICS theoretically enabling companies to move their PowerCenter on-premises installation base and computing capacity to the Cloud and shift traditional support and maintenance revenue from perpetual software licenses to Cloud subscription revenue.

This is crucial for Informatica who aim to achieve 80% of their revenue through repeat revenue, but it also means that code produced today on PowerCenter Classic is not seen as instant-legacy code thereby potentially eroding their existing customer base by opening them up for competition.

Prospective customers of Datawave may look for assurances that we support Cloud deployment even if initially only to the extent that PowerCenter code generated on our iPaaS can be deployed on IICS in the future, the same question will apply to how we support IBM products.

Orchestrate. Automate. Transform.

# Summary

Datawave has grown its pattern-based code generation and separate job control framework products into a single iPaaS to support the most common use cases that the company sees in its market segment at the moment: Informatica PowerCenter users, IBM Datastage users and companies that are looking for a governed method of ingesting data into their Big Data platform.

All of these use cases are for new projects not, for the purposes of clarity, migration from one technology platform to another. It is important to stress compatibility with other technologies as neutrality is very important in technology evaluation studies at the moment due to the risks associated with rapid technological change.

As the first iPaaS release takes shape then we need to consider:

- Support, connectivity and compatibility with other platforms and technologies
- Pricing
- Usage modes: as a differentiator when delivering a services engagement and/or stand-alone product used post service delivery. Others?
- Product support & training (both internally within HCL and to customer base)
- Ease of use and deployment
- Positioning the iPaaS to Customers e.g. offers simplicity and cost advantages over rival platforms such as IICS and backed by services? N.B. this is the Fujitsu approach which has been very successful for them.

**datawave**

An HCL Technologies Company

**Orchestrate. Automate. Transform.**