**Data- as- a- Service: Unlock the value of Data**

Prepared by: Mohd Hafeez, BDA Data Economy, MdeC.

Data-as-a-Service (DaaS) is the output from the emergence of cloud computing technology, whereby the scalability in the platform is the main factor, similar to the other X-as-a Service concept. DaaS offering the data or datasets that require by organisation/clients to fuel their analytic or insight on current trend, marketing analysis, identify churn rate and so on. Data is supplied “on-demand” via cloud platforms, as opposed to the traditional models in which the data remains in the customer’s hands, and the vendor provides the tools that make it easier to access, explore and analyse.

DaaS is based on the concept that the product, data in this case, can be provided on demand to the user regardless of geographic or organizational separation of provider and consumer. Data as a Service (DaaS) is a cloud strategy used to facilitate the accessibility of business-critical data in a well timed, protected and affordable manner.

Most companies, even though might necessarily have in-house Business Intelligent (BI) department–don’t have the necessary infrastructure to deal with truly big data, both in terms of procuring and storing the data and in terms of effectively analyzing it. Furthermore, there is no real reason for a manufacturing company (for example) to invest heavily in building in-depth analyses of social media outliers. This is where the DaaS can fill the gap, in order for company or organisation to truly unlock the value of the data and/or open data.

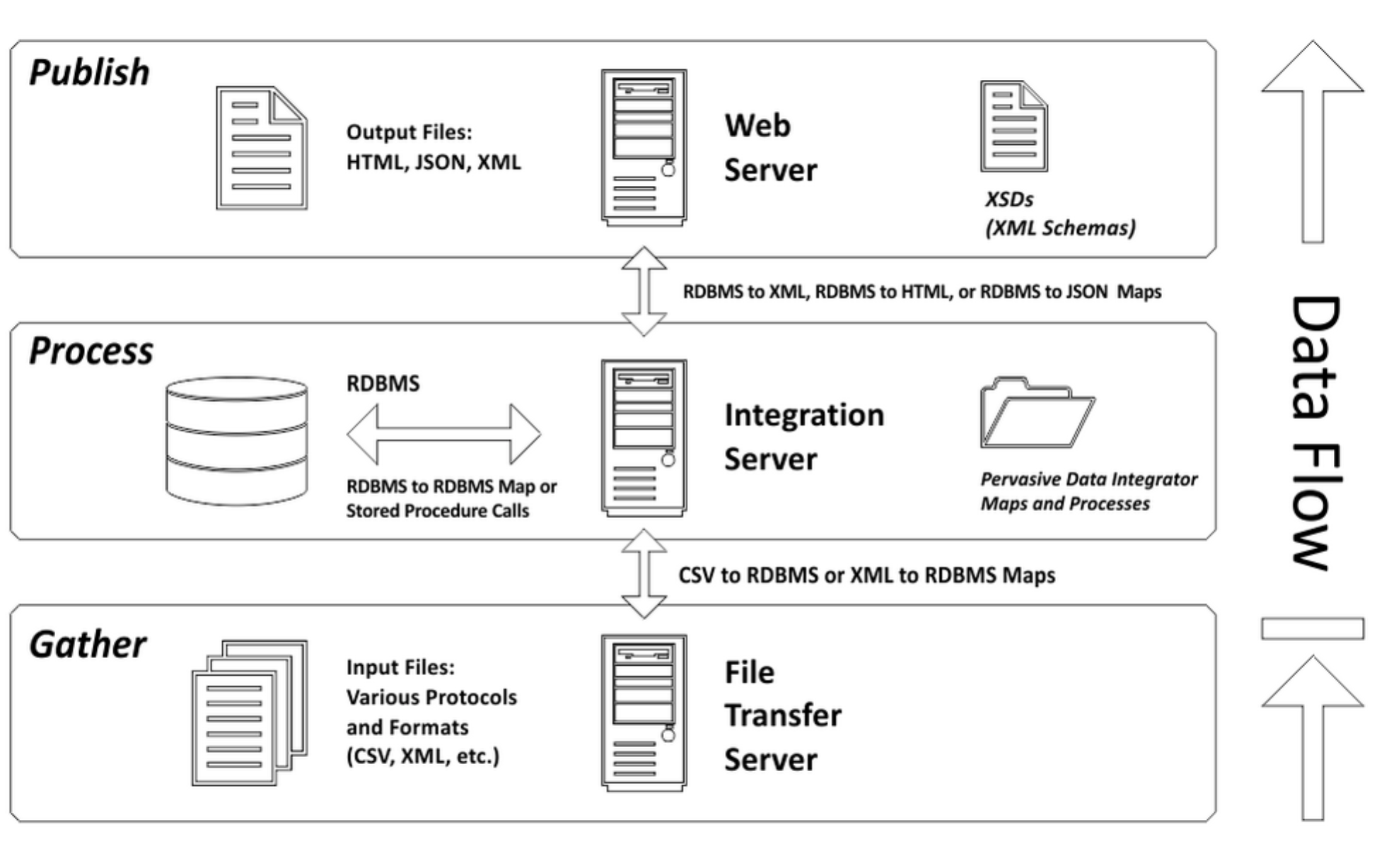
**Why DaaS is needed**

As mentioned above, one of the reason why DaaS is the new way moving forward is most of the companies did not have the facilities or/and neccessities to store large (real-time) data that procure from the emergence market or in the net. In additional, some organisation did not have the expertise, and technology know-how to converge and store the granular unstructured or structured data that use for BI, which then would require a large cost. For instance, SMEs that run the core business in trading would not know the approach on how to obtain and store the trending market data, or even competitors data that are useful in marketing analytics and BI. Their need for this data is limited in time and scope, and so it makes sense for them to purchase it from an external provider as their business demands, rather than devoting extensive resources to reach the same results internally.

Next it is all about cost and time. When we are dealing with data that is produced outside the organization, it could often be simpler and more cost-effective to outsource the related services, rather than store and process it locally, unless you want to mesh it with your own proprietary data. Of course, there are cases when you would also want to handle external data sources in-house (for example, to cross-reference them more efficiently against your own records), but I believe this to be the exception and done by more advanced organizations, and not the rule.

DaaS could increase agility and easily accessible as the customers can take immediate action and do not require in-depth understanding if actual data. Also, the data that was acquired from the service is high quality data, as it was controlled through data services (cleansing, filtering, meshing process), which improves data quality. Furthermore, customer do not require to build or develop a complex tools to either scrap the data from the net, also it allows the customer to easily design their own front end and not practically tied up with data protocols.

**DaaS Model Architecture**



DaaS cannot perform without the collection of data that will be gather and store in the transfer server. The input files can be of various formats such as XLS, CSV, XML, HTML and so on. This then will be store in the Integration Server than handle the storage of the source data and the request from the customer. On top layer, application can be build to publish the output files requested by the customer or consumer. Please take note that the data flows and storage happens in the clouds. The output files can be on the HTML, JSON, XML or even APIs that consist of metadata. The pricing model can be build depending on the complexity of the architecture and so on.

**Business Model**

The pricing model of DaaS can be categorize as such:

1. Volume based model.

This is the simplest model of pricing for DaaS, as it will depend on the volume of the data query by the customers. Also the DaaS provider will charge the customer by each call of the API.

1. Subscription based model.

Like what have been implemented by TFL for TransportAPI in UK, the pricing model is based on the subcription by developers. The premium subscriber will enjoy low-latency data query, full capacity and limitless data from the call of the API.

1. Data type model.

The type of data queries by the customer will determine pricing. For instance, geographic, financial and historical data necessary for customer business are examples of types of data upon which pricing may be based. Also, high demand datasets can be categorized into one data type.

1. Crowdsourcing based model.

Any entities that would want their data to be acquired by customers will need to pay depending on what type of data and volume of the data stored in the cloud.

**The Challenges**

There are certain obstacles that still pose a challenge for these services, and should be considered before one turns to a cloud solution.

1. Scalability

DaaS requires high scalable platform to store the data that was acquired either by manual collection, public data or private acquired datasets. DaaS solutions typically rely on per-usage subscription models, which seem great when a business is working with small amounts of data. But since there is in the near future, the typical organization will be dealing with much more data than it does now, every company has to ask itself not how much data it currently has, but how much it might have in a few years’ time.  These will proportionally affecting the cost of the cloud storage rental, as more scalable storage will cost you more. According to the current pricing of cloud storage, it is a less than optimal solution for Big Data, particularly if this data is frequently updated.

1. Data Ownership

Private and corporate data are always bound to be sensitive, as it tied to the legal and stringent requirement regarding to access and modify for other purposes. Transferring this data to third parties could raise legal and contractual issues, which must be worked out with the vendor beforehand to prevent complications down the road. There are also concerns on the public data, wherein the modification of the datasets will always be an issue.

1. Data Management

The way data is modeled and structured can significantly affect the types of analyses that can be performed on it, and the way different data sources can be joined, compared and referenced. Thus, different customer might want different type of schema data for their quicks soultioning/visualization. This might be the issues on how to handle the data management and publish as per say.

1. Data Sources

Public data can easily scrap from the public sector websites, however it’s difficult to really acquire more details and updated critical data such as satellite images, geospatial images and so on. On the other hand, private data is bound to the mechanism on how to mine the data itself. Most of the DaaS Company acquired data through crowsourcing and survey, which include the participations of end user (customer) itself. They get paid for their services.

1. Data Privacy and Integrity

If you talk about data mining and storage, how do you want to preserve the privacy, security and integrity of the data? There is a need to identify the best possible mechanism to ensure the security to avoid data breaching and risk the integrity of the data available.

**Conclusion**

There significant difference between Data-as-a-Service and Analytics-as-a-Service. Basically, DaaS is similar like data brokerage, which gather, store and publish data that query by the customer. Unlike Analytics-as-a-Services companies that provide BI from the data that gather and store in the systems. Both infrastructures used cloud and/or big data platform.