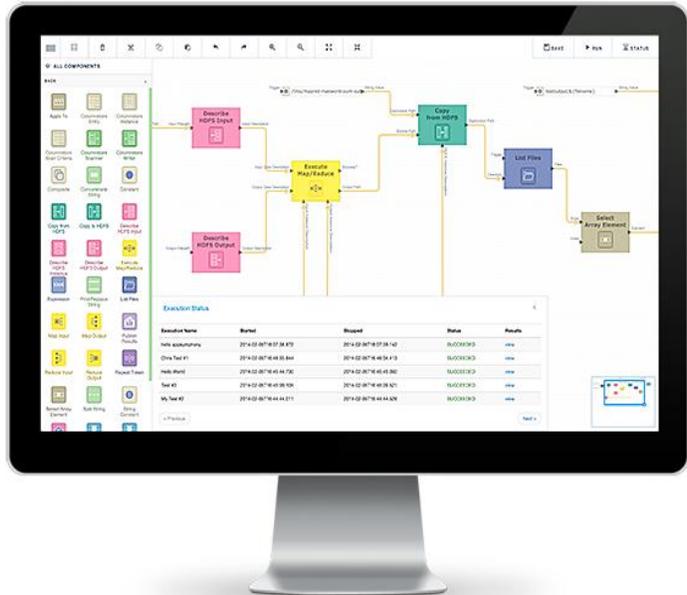


### *Secure Self-Service Analytics for Curated Digital Collections*

#### Introduction

Optensity, Inc. offers a self-service analytic app composition platform, AppSymphony™, which enables data scientists and analysts to rapidly build customizable mission critical analytics that execute and scale across multiple compute infrastructures. AppSymphony's highly intuitive browser-based graphical composition environment dramatically simplifies the technical diversity of big data analysis, enabling users to rapidly compose applications that combine structured and unstructured, real-time and archived, web, social media and multimedia data. Composed apps are seamlessly distributed and executed on one or more AppSymphony Execution Engines hosted on corporate infrastructure and/or at scale on multiple public, private and hybrid clouds to analyze big data at or near its source.



#### Use Case: Curated Digital Collections

This use case addresses the needs of curated digital collection providers – organizations who collect, curate and archive data into digital collections, and then provide their digital collections for customers to use. Curated data of all types is provided by a wide variety of government, academic, commercial and other organizations, with highly diverse content: national security defense and intelligence data, proprietary business data, creative journalism and entertainment products, academic content, personal, social and health data, scientific data sets, ISR data etc. The data may be an organization's primary product, it may be collateral or residual byproducts of other business, government, social, scientific, medical, or other real-world processes and sensors. It may also be aggregated from a number of sources and classification levels.

#### Example Case

A US Federal Government agency has a particularly challenging digital collection that accumulates on order of 1 petabyte of new data per day. They have established a data acquisition process that successfully captures, organizes and stores the data. But providing that data to researchers and policy makers in a comprehensive, sustainable and cost-effective way presents some challenging pain points.

### *Pain Point - Infrastructure Costs*

The most obvious pain point for this curator is the infrastructure costs of preserving the collection and providing customers access. The agency assessed existing software and hardware solutions that divide and simultaneously search large data sets to reduce search time. To achieve a significant reduction of search time, however, would require an extensive infrastructure of hundreds if not thousands of servers.

However, there is another more insidious pain point - one that hard limits the enduring value of the dataset, and therefore the provider's core mission effectiveness.

### *Pain Point - Analytic Lock-In*

The unavoidable fact is: curators must build a mission information system to make their curated data available for use. And common practice is that information systems are engineered and built to perform functions that are specified very early on in their life-cycle. This often means they are locked-in to specific data access functions, e.g. search, browse, select, transform, combine, filter, process, analyze and present the results. Such analytic lock-in is anathema to data collections that must support ongoing discovery.

### **Impact to Mission**

Public and private research in information and related sciences are continually advancing the state of the art in what can be learned from digital collections. New analysis methods, algorithms and technologies are being invented every day that can only be tested, refined, and ultimately proven against significant data sets. It is impossible for anyone to anticipate those breakthroughs and engineer a traditional information system that can support them, which ultimately limits the pace of innovation, scope of discovery and the effectiveness of the mission.

### **AppSymphony Solution**

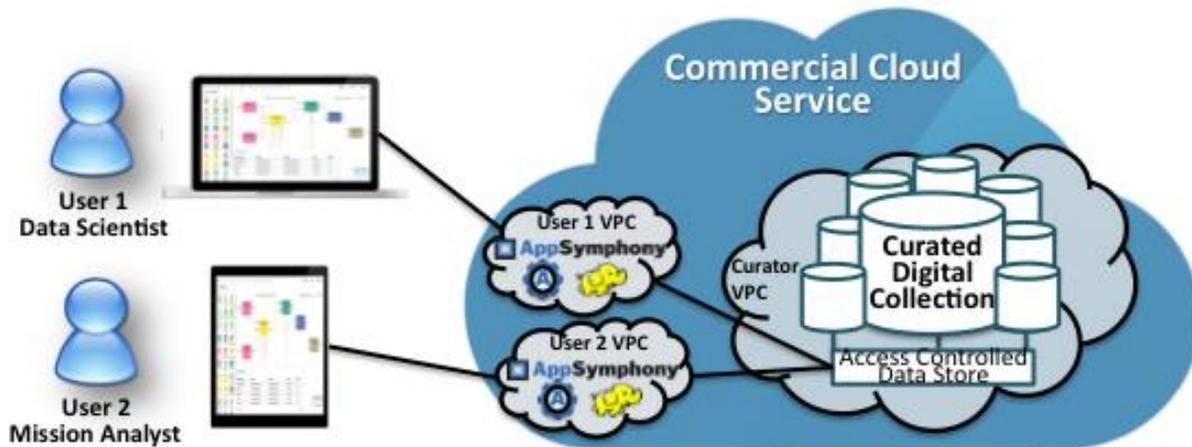
There are three distinct capabilities that digital collection curators need to implement:

- Acquire - capture, curate and conserve the value of ever expanding data collections
- Provide - make the curated data readily and affordably accessible to its customers for research and analysis
- Protect – ensure data integrity, security, privacy, anonymity, etc. of the data *and its consumers*

AppSymphony can be used to support all three of these needs. But AppSymphony's key technological innovations best support 'provide' and 'protect'.

**Optensity**

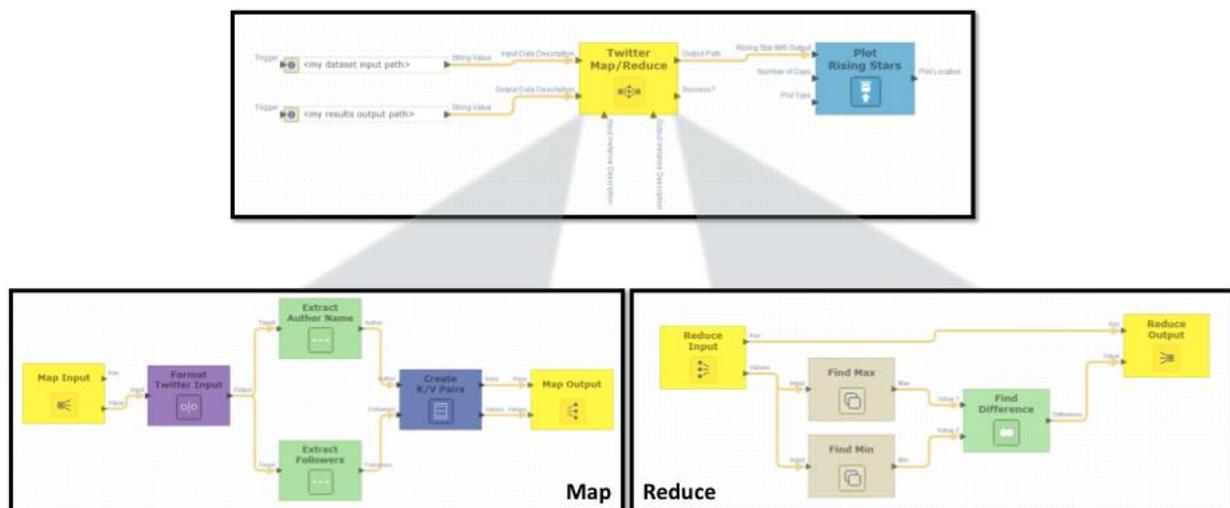
<http://www.optensity.com>



**Provide:**

*Self-Service Composable Analytics*

Data Scientists and analysts can compose their own analysis applications using AppSymphony’s intuitive graphical dataflow language shown in the figure above. Users drag-and-drop components onto an information workspace and ‘wire’ them together into executable dataflows. No software engineering or information technology (IT) skills are necessary to compose and execute analysis apps – just information domain knowledge (e.g. collection content) and analysis experience. AppSymphony effectively eliminates the underlying IT complexity, providing users with hundreds of easy to use data processing and analysis components, including ones that run composed map-reduce jobs natively on Hadoop clusters (see the figure), NoSQL data stores, etc.



AppSymphony’s composable analytics puts the full power of today’s state-of-the-art high-performance cloud computing directly into the hands of data scientists and analysts. *No longer will users suffer the analytics lock-in pain point* – they’ll be able to rapidly compose and apply their most innovative analysis methods, algorithms and technologies to the curator’s digital collection to be tested, refined and ultimately proven.

**Optensity**

<http://www.optensity.com>

### *Bring Your Analytics to the Data*

The real value to customers lies not in the curated data itself, but in the actionable insights that can be discovered and extracted from it. AppSymphony allows users to send their graphically composed analytic apps to execute on our patent-pending Execution Engines, which are physically collocated at or near the digital collections. This approach not only eliminates the analytic lock-in pain point, but also the time and expense to transport data from multiple physically distributed curators back to your own analytics - where that's even possible or practical at all! In other words, AppSymphony automatically brings your composed analytics to the data, while making its physical location transparent to the user.

### *Run In the Cloud*

The traditional approach has been for curators to manage their digital collections and make analysis capabilities available through corporate or agency data centers, which can be particularly expensive and time-consuming to build and operate, especially where content is continually acquired and added to the collection. Commercial cloud providers, on the other hand, offer on-demand massively scalable infrastructures as a service (IaaS), which enables curators to store and manage their digital collections at a more reasonable cost. The commercial cloud approach has been embraced by the Intelligence Community through the Commercial Cloud Service (C2S) and other comparable programs. Originally built on Amazon AWS from the ground up, AppSymphony runs composed analysis apps transparently across Amazon AWS and several other cloud providers – without having to move executable code.

### *Just in Time vs. Just-in-Case Infrastructure*

Traditional corporate/agency data centers are sized in advanced based on projected compute, storage and network capacity demands from the mission applications they host – usually with some capacity margin to support short term demand surge and long term growth. This approach might be called “just-in-case” provisioning. Since cloud service providers offer the ability to dynamically provision infrastructure resources on-demand, digital collection customers can access just the right data they need from the curator and just the right compute infrastructure they need from the cloud provider for just long enough to perform their own analysis on the data. AppSymphony interacts directly with multiple cloud providers’ APIs to dynamically provision resources “just-in-time” just where they are needed to perform user analyses, and then return the resources to the cloud for others to use until they are needed again. This capability *dramatically reduces infrastructure cost pain point* for both the curator and the consumer, all while making the entire process transparent to both.

### **Protect:**

#### *Dynamic Enclaves*

Cloud services provide a variety of mechanisms to simultaneously share and protect curators' digital collection. But there is more to protection than just the curated data. Many customers also need to protect the fact of, circumstances and results of their data use. In particular, customers who combine data from a number of diverse providers must often protect the aggregated data and resulting analyses at least as carefully as the original individual collections. AppSymphony makes full use of cloud provider protection mechanisms such as Virtual Private Clouds (VPCs) in conjunction with Just-in-Time dynamic resource provisioning - an approach we call Dynamic Enclaves. In this approach,

**Optensity**

<http://www.optensity.com>

AppSymphony interacts programmatically with the cloud provider API to dynamically stand-up limited access VPCs and provision compute and storage resources into the VPC. AppSymphony then executes users' composed analytics in the VPC, where they access, combine and analyze curated data from all relevant collections. When the analyses are complete, the compute resources, any contained data and the VPC may be stopped and saved until they are needed again, or terminated altogether. AppSymphony makes the entire dynamic provisioning process transparent to the end user, leaving her to focus on the analysis tasks at hand.

#### *Fine-Grained Access Controlled Data Stores*

AppSymphony provides composable components that connect to a number of data stores that provide varying degrees of fine-grained data access control. These include the Accumulo column-family store, Amazon DynamoDB noSQL database, and the Oracle relational database with Label Security.

#### **Separation of Concerns**

The AppSymphony solution puts curated data services responsibilities exactly where they belong:

- Digital collection curators use just enough resources to store, manage and protect their digital collections in the cloud – they no longer have to field analysis applications nor worry about analytic lock-in
- Intelligence analysts and data scientists graphically compose their own analytic applications based on their domain knowledge and tradecraft, and then execute them directly through AppSymphony. The full power of curated data and the cloud is placed directly in their hands
- AppSymphony dynamically provisions just enough cloud VPCs and compute resources to execute analysts' composed analysis apps, and then shuts them down and returns them to the cloud for others to use.

This separation of concerns may be easily extended to cloud provider billing as well.

#### **Summary**

AppSymphony brings unprecedented self-service analysis capabilities to curated digital collections of virtually any size or location. It enables analysts and data scientists to quickly compose analysis applications that extract actionable insight from diverse collections hosted throughout the community. AppSymphony puts analytic power and agility directly in analysts/data scientists hands – no software skills required. It brings your analytics to the data, eliminates IT and cloud complexity, makes highly efficient scalable secure use of dynamically provisioned cloud resources, and effectively eliminates analytic lock-in.

Contact: Frank White, CTO  
Optensity, Inc.  
[fwhite@optensity.com](mailto:fwhite@optensity.com)  
781-771-2798

**Optensity**

<http://www.optensity.com>