Orbis Technologies, Inc.

DataARCH®

Product Highlights:

The DataARCH® platform establishes a cloud-based semantic enterprise integration platform that is a turn-key solution for rapid integration of enterprise data.

- Reduces time required for integration of multiple enterprise data sources from months to minutes.

- Semantic data mediation allows an organization to receive all of the benefits of a cloud architecture without costly legacy database migration efforts.

- Automates the generation of database data models in order to eliminate the need for time consuming technical meetings to define common or standard schemas or ontologies.

- Provides multiple algorithms and approaches to resolve duplicate instances of enterprise data from multiple data sources to create a streamlined unified Big Data "Data Lake."

- Industry standard interfaces, effective incorporation of project ontologies as a data source, and the integration of legacy databases and data services allows previously created enterprise knowledge to evolve and grow over time.

180 Admiral Cochrane Drive
Suite 305
Annapolis, MD 21401
866-606-7247
info@orbistechnologies.com
www.orbistechnologies.com

DataARCH® Overview

DataARCH® is an innovative next generation enterprise integration product designed to significantly reduce the time and resources required to integrate a large number of heterogeneous data sources. The product provides key integration capabilities to mediate between different data architectures and data types (Hadoop, RDBMS, etc.) as well as resolving conflicts in instance data.

Clients typically use DataARCH® to physically and virtually integrate disparate legacy databases and data services to create a unified Big Data view of the enterprise. An easy to use User Interface (UI) provides Enterprise Architects, Data Scientists and integration team members a tool to perform a variety of critical Big Data enterprise integration tasks.

Innovative Features

DataARCH® delivers numerous product features that serve as the fundamental elements for any enterprise integration effort.

- **Data Model Governance:** Enterprise data models are efficiently managed to support regulatory and policy compliance, model mappings to databases, alignments between models and model versioning. Dashboard plug-ins provide metrics to support the monitoring and display of governance metadata.

- **Unified View of Enterprise Data:** Enterprise data can be virtually integrated (i.e. data stays at rest in legacy databases) or physically integrated (i.e. selectively transformed and published to a new database).

- **Enhanced Data Security:** Organization application level security and access policies can be applied as well as Big Data implementations that provide instance level data protection in support of compliance with numerous regulatory requirements associated with Personally Identifiable Information (PII) and Health Insurance Portability and Accountability Act (HIPAA).

- **Standards Based Implementation:** Data models and queries are implemented using World Wide Web Consortium (W3C) standards. Data transformed to reference models is available to the enterprise as semantic web-based linked data to assist in data interoperability.

- **Discoverable Data Registry:** Ontologies, vocabulary mappings, federation information, and associated pedigree metadata is stored and retrieved via a data registry.

- **Automated Model Generation for New Data Sources:** Data models, in industry standard format (i.e. Web Ontology Language – OWL) are automatically generated for legacy databases when they are integrated.

- **Advanced Query Brokering:** Multiple complex data schemas possessing hierarchical and logical relationships are mapped to a reference model to support query brokering and federated enterprise queries. Mappings can include the creation of mathematical rules and logic to support mediation between data schemas from different data sources.

- **Data Exporting:** Data mapped to a reference model can be exported using either the data source schema or the reference schema, allowing for basic ETL functionality to move the data from one repository to another.

- **Promotes Data Clarity:** Sophisticated algorithms link data from separate sources by specifying a relationship algorithm using easy to create user interfaces to find duplicates and disambiguate instance data. Additional tools enable clean-up, normalization, and formatting of the data before it is exported.

- **Expose flat files as a data source:** Structured local/remote files can be loaded into the platform and exposed as a data source, making the data available for querying and linking with other sources.

Integration of Customer Databases

A large international corporation wanted to integrate multiple databases with customer contact and account information in only a few days. By integrating these records they could create a unified integrated view of their client records. The effort had several key technical challenges that made DataARCH® the best product for the project. Select examples include:

### Unknown Data Architecture for Legacy Databases
The legacy relational databases were not provided with corresponding data models. As a result, the DataARCH® automatic generation of ontologies for the legacy databases allowed the integration team to rapidly identify the components of the legacy databases that were needed in order to build the queries supporting both a physical and virtual integration.

Data Duplication
All of the databases had overlapping client records. Unfortunately, the records were not identical which meant key information, such as a street address, was not entered in the same way across the databases. DataARCH® was used to selectively identify key database fields and rules were written in the DataARCH® rule builder to mediate between the records. Once implemented, duplicate records could be identified.

Integration of Business Intelligence Applications and Analytics
Most data integration approaches rely on transforming the disparate sources to a common format. This approach leads to redundant data storage, unnecessarily complex processing pipelines, and data stovepipes that result in higher-than-required data center operational costs. DataARCH® keeps data in its native format to avoid these issues.
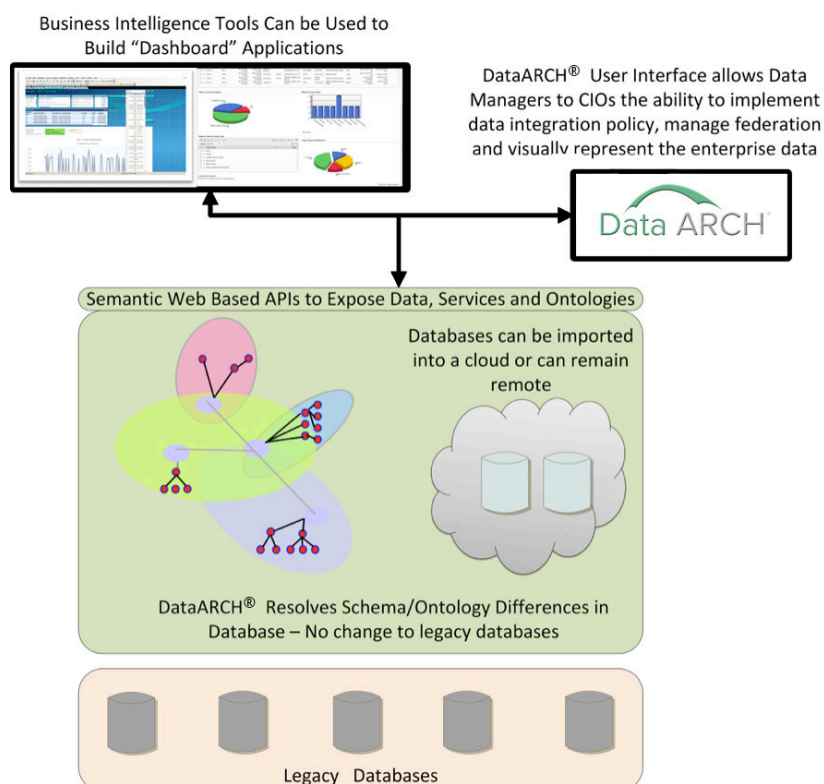
Q: What makes DataARCH® better than other enterprise database integration technologies and techniques?

DataARCH® represents a paradigm shift in enterprise integration – a turn-key solution specifically designed to accelerate the integration and availability of geographically disparate and heterogeneous data sources.

DataARCH® creates a virtual data lake, making any number of your data sources available in the cloud, without having to modify your existing data infrastructure. Data sources are registered and exposed incrementally with no impact to existing client applications.

By using open industry standards in a cloud environment, DataARCH® promotes interoperability among your current and future applications while ensuring platform independence. The open cloud integration approach offers significant advantages over proprietary and transitory implementations. This has proven to reduce the overall cost of IT infrastructures and data management systems.

Q: What differentiates the DataARCH® platform from traditional enterprise integration platforms?

In stark contrast to traditional SQL databases and web service architectures, the semantic web technologies leveraged by DataARCH® are specifically designed to simplify data access and open up traditionally closed data silos to enable information to be fused and new knowledge derived.

Q: How do I obtain a license?

DataARCH® licenses are available on term, project or enterprise basis. This simple approach to product licensing allows clients to discover the value of the product before entering into an enterprise agreement.

Learn more at www.orbistechnologies.com.

About Orbis Technologies, Inc.

Orbis Technologies, Inc. is a recognized leader in providing services and technologies for designing and developing next-generation information technology. Our award-winning company provides a distinguished team of technologists and developers to assist commercial and government clients with near-Internet sized data integration challenges. Orbis offers a range of discriminating cloud-based products and enterprise services, such as:

- Cloud Text Analytics (CTA®) designed to process and semantically correlate documents in a distributed storage and processing environment;

- DataARCH® designed to rapidly integrate disparate data sources and mediate data models and ontologies;

- Analysis and Data Efficacy Profiling Tool (ADEPT®) designed to leverage the value of an enterprise's data, and more effectively manage their data centers;

- CloudSTREAMS® designed to integrate third-party analytics for real-time Hadoop and visualization applications.