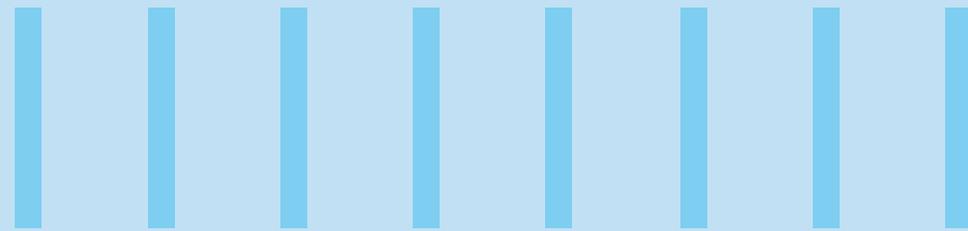




ANALYTICS IN HEALTH AND HUMAN SERVICES: OWNING YOUR DATA

Three approaches to build a strong analytics and data management system for generating actionable insights



Abstract

Health and Human Services (HHS) is approaching a renaissance in analytics. Many States have begun to warehouse their data, but many times the warehouse becomes a means to share data between agencies that lack interoperability, turning expensive data warehousing hardware and software into a very expensive data exchange gateway. Failing to meet the compliance needs through their data initiatives, States end up outsourcing critical analytics to external consultancies at a premium price. There are alternatives and ways to bring back control of data and generate actionable insights.

Building a single, web-based master data store

Health and Human Services (HHS) agencies manage varied data stored across a variety of devices. It is crucial to integrate all this data from multiple operational source systems into a single, web-based information data store. This exercise is multifaceted and involves:

- Integration and transformation of data interactions between different entities from myriad sources – exchanges, patients, citizens, etc. – into a set of meaningful metrics that can be tracked continuously besides being used for evaluation purposes

- Ingestion and processing of semi-structured or unstructured data (call center notes, voice transcripts, etc.) along with structured data
- Data ingestion in real time or batch mode

Given the aforesaid requirements (or constraints), there are three approaches that HHS agencies can use to build their master data store:

- Open source
- Traditional
- Hybrid

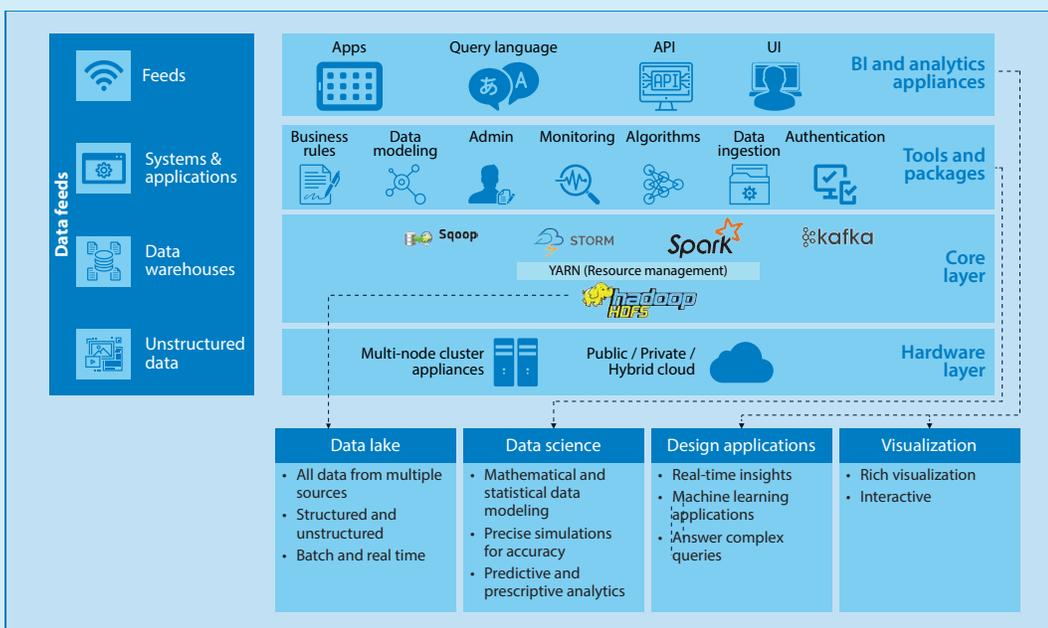
Choosing the right option is based on a combination of factors such as budgetary considerations, speed with which business needs the requirements to be implemented, ease of implementation of the technology solution, and compatibility with existing technology. The sections below introduce each approach and discuss the pros and cons of using them to build a national data warehouse.

Open source approach

This approach leverages open-source tools and technologies for data storage and computation (like Hadoop), transferring bulk data between Hadoop and relational

databases (like Sqoop), and setting up a distributed messaging system (like Kafka). The approach allows agencies to extract transactional data (structured /

unstructured / batch / real-time, etc.) into a Hadoop-based data lake that can serve as the information data store.



An open-source approach can be a cost-effective method to build analytics capability for those States that can manage on their own and are willing to invest in building in-house capabilities from the ground up.

Figure 1: Open Source Approach

Traditional approach

Traditional data warehousing approach requires handling data from multiple operational source systems in three layers:

- **Data integration layer –**
Extracts data from disparate source systems and provides a single view of source data
- **Data transformation layer –**
Transforms the data as per business needs by implementing data quality checks to remove data anomalies, duplications, and enrichment
- **Data delivery layer –**
Delivers data to end users through a reporting access layer as per users' needs

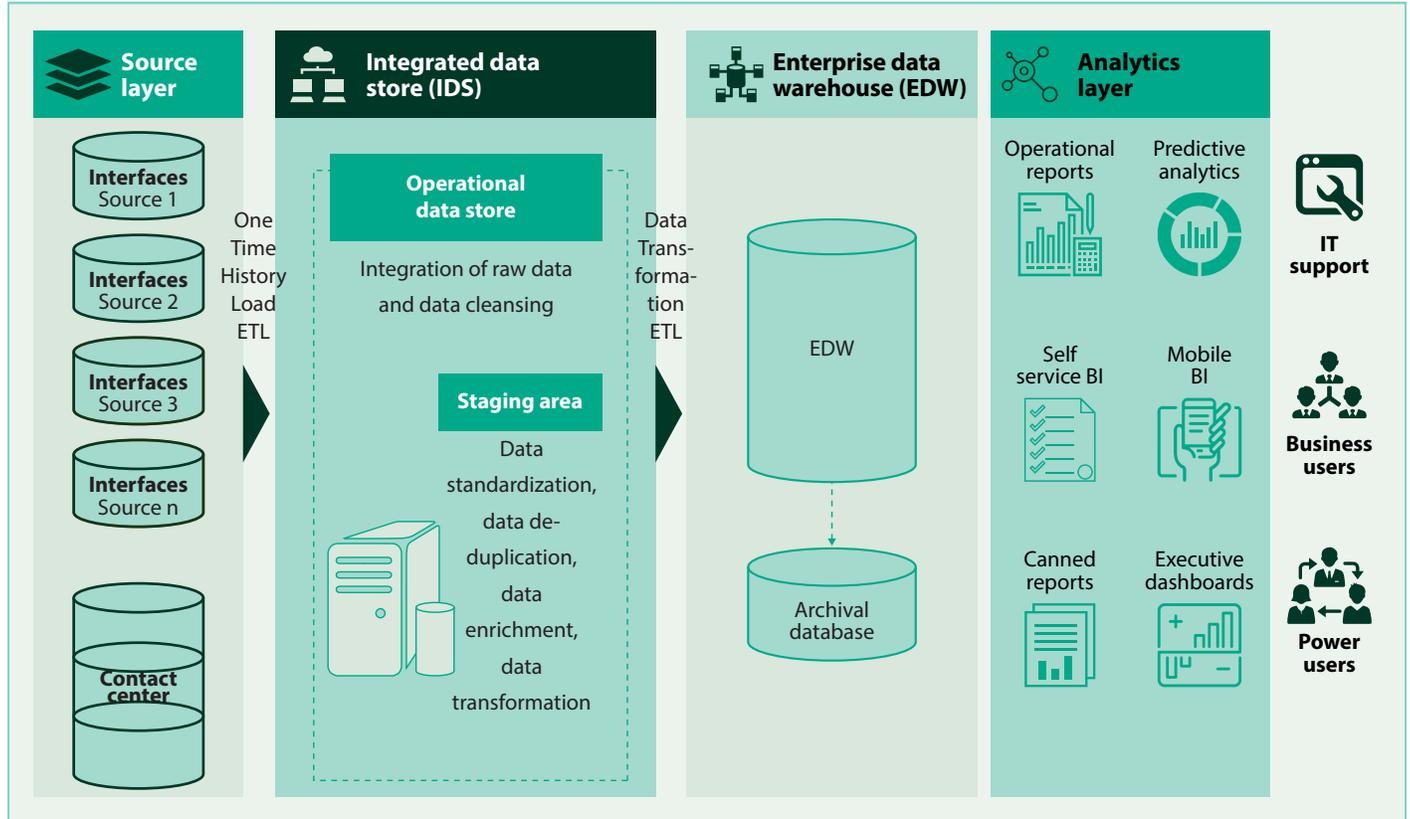


Figure 2: Traditional approach ▲

The integration layer consists of two major components – information data store (IDS) and a staging area. IDS will be populated to take care of operational and analytical reporting needs and will provide the single source of transactional data processed. Data quality checks like data de-duplication, data standardization, data enrichment, and the required transformations would be implemented in the staging area to provide the desired data for further business rules implementation in the IDS layer. Data quality is a continuous improvement process. The following illustration depicts the foundation blocks of data quality:

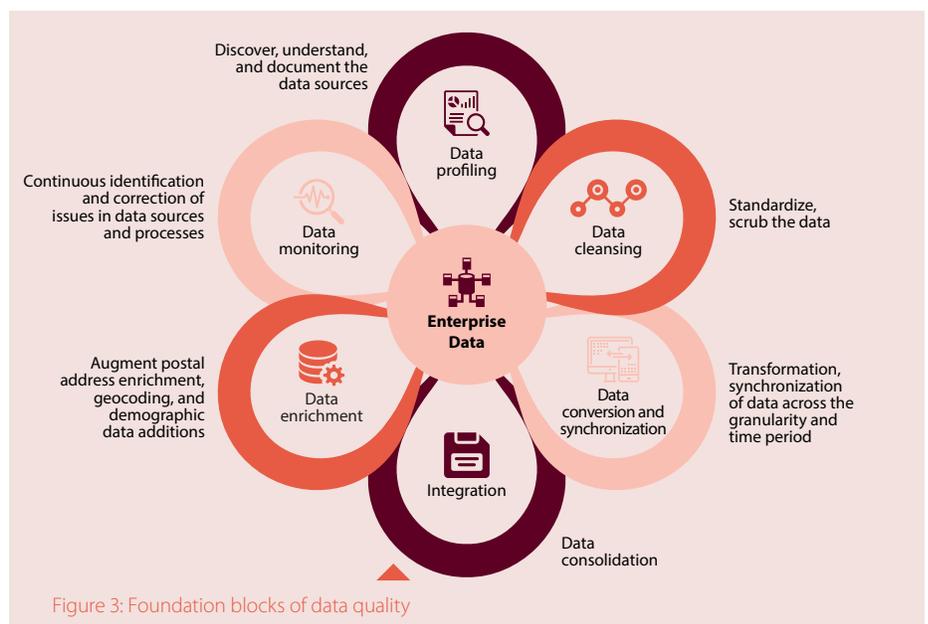


Figure 3: Foundation blocks of data quality ▲

Since the source of data being used to create any form of report is the data warehouse, there is a higher level of control over any type of report (standard, dashboard, ad-hoc query, etc.) being

created, and metrics that are tracked via that report. Data quality practices ensure that the errors are identified and caught before they reach the reports. A traditional approach may be more

efficient for those States with an existing staff and technology capability, building on current processes and resources. While the traditional approach may take more time, it provides a proven path.

Hybrid approach or the Augmented Data Warehouse Solution

The hybrid approach is in between the 'big bang, rip and replace' and 'expand on the traditional' approaches. This approach has the following components:

- Open-source technology that acts as the foundation layer for agile, iterative analytics
- Open-source platform to process data from subject areas / external sources
- Standardization / Normalization of data (both structured and unstructured) into a common format for easy ingestion
- Use of open source platforms like Hadoop to augment traditional EDW landscape by sharing ETL and for resource-intensive historical data processing

Big data augmentation can be taken up for key areas. The processed data will be made available to existing EDW / data marts and other systems for enterprise-wide integration, analytics, etc.

A hybrid approach enables a more rapid development of capability than a traditional approach, but maintains key elements of a traditional approach to lower technology risk and resource gap. It may be the best fit for a State's technology strategy.

In the event of challenges in integrating different data sources, like real-time data delivery requirements, complexity of data integration tools to handle various sources and formats in real-time, organizations can leverage a data virtualization approach. The virtualization process would involve abstracting the metadata of disparate data sources (databases, applications, file repositories, websites, data service vendors, etc.) through a single data access layer, which may be any of several data access mechanisms, while leaving the actual data in place.

Conclusion

HHS agencies have access to a lot of data, both structured and unstructured, and are in a unique position to generate actionable insights using this data by building a master, web-based data store. This can be done in three ways – open, traditional, and hybrid. HHS agencies need to carefully consider a number of factors including but

not limited to – budgetary considerations, technology risk appetite, ease of implementation, and compatibility with existing technology, before deciding on the right approach. The right approach will enable HHS agencies to own their data and generate actionable insights, quickly and cost-effectively. While selection of the right

approach to build a master web-based data store and analytics system is crucial, HHS agencies should not lose sight of their end goal – that of using their analytics system to generate insights that help deliver personalized services to citizens, improve care and outcomes, reduce costs, and eliminate fraud.

For more information, contact askus@infosyspublicservices.com

Infosys[®]
Public Services

© 2016 Infosys Public Services, Inc., Rockville, Maryland, USA. All Rights Reserved. Infosys Public Services believes the information in this document is accurate as of its publication date; such information is subject to change without notice. Infosys Public Services acknowledges the proprietary rights of other companies to the trademarks, product names and such other intellectual property rights mentioned in this document. Except as expressly permitted, neither this documentation nor any part of it may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, printing, photocopying, recording or otherwise, without the prior permission of Infosys Public Services and/ or any named intellectual property rights holders under this document.