

Database Trends and Apps Report – Building a Data Lake for the Enterprise



Data lakes are forming as a response to today's big data challenges, offering a cost-effective way to maintain and manage immense data resources that hold both current and future potential to the enterprise. However, enterprises need to build these environments with great care and consideration, as these potentially critical business resources could quickly lose their way with loose governance, insecure protocols, and redundant data. The following are best practices for making the most of data lakes in the enterprise:

- THINK ABOUT THE BUSINESS
- THINK EXPERTISE
- THINK ABOUT WHAT DATA REALLY NEEDS TO BE CAPTURED AND STORED
- THINK LONG-TERM, THINK ARCHITECTURALLY
- THINK SECURE
- THINK SELF-SERVICE

A key area in which data lakes are proving their potential is the healthcare sector. A semantic data lake for healthcare is underway at Montefiore Medical Center, which involves a sophisticated machine learning project that is slated to go live for patient care in the summer of 2017. The data lake supports predictive analytics to clinicians, starting with the ability to flag patients entering the health system at high risk of experiencing a serious crisis event within 48 hours. The system also generates customized checklists of intervention tasks sent to clinicians that may help to avert or lessen the impact of the crisis.

The Semantic Data Lake

A Semantic Data Lake is incredibly agile. At the core of a Semantic Data Lake model we find two W3C standards: the URI and RDF. The architecture quickly adapts to changing business needs, as well as to the frequent addition of new and continually changing data sets. No schemas, lengthy data preparation, or curating is required before analytics work can begin. Data is ingested once and is then usable by any and all analytic applications. Best of all, analysis isn't impeded by the limitations of pre-selected data sets or pre-formulated questions, which frees users to follow the data trail wherever it may lead them.

AllegroGraph – Semantic Graph Database

The Montefiore project uses AllegroGraph to store the data lake. Unlike traditional relational databases, AllegroGraph provides the unique ability to link data, without manual user intervention, coding, or the database being explicitly pre-structured. AllegroGraph processes data with contextual and conceptual intelligence to resolve queries and help the

clients to build predictive analytics, which help them to make better, real-time decisions.

[Download a pdf of the full report here](#)