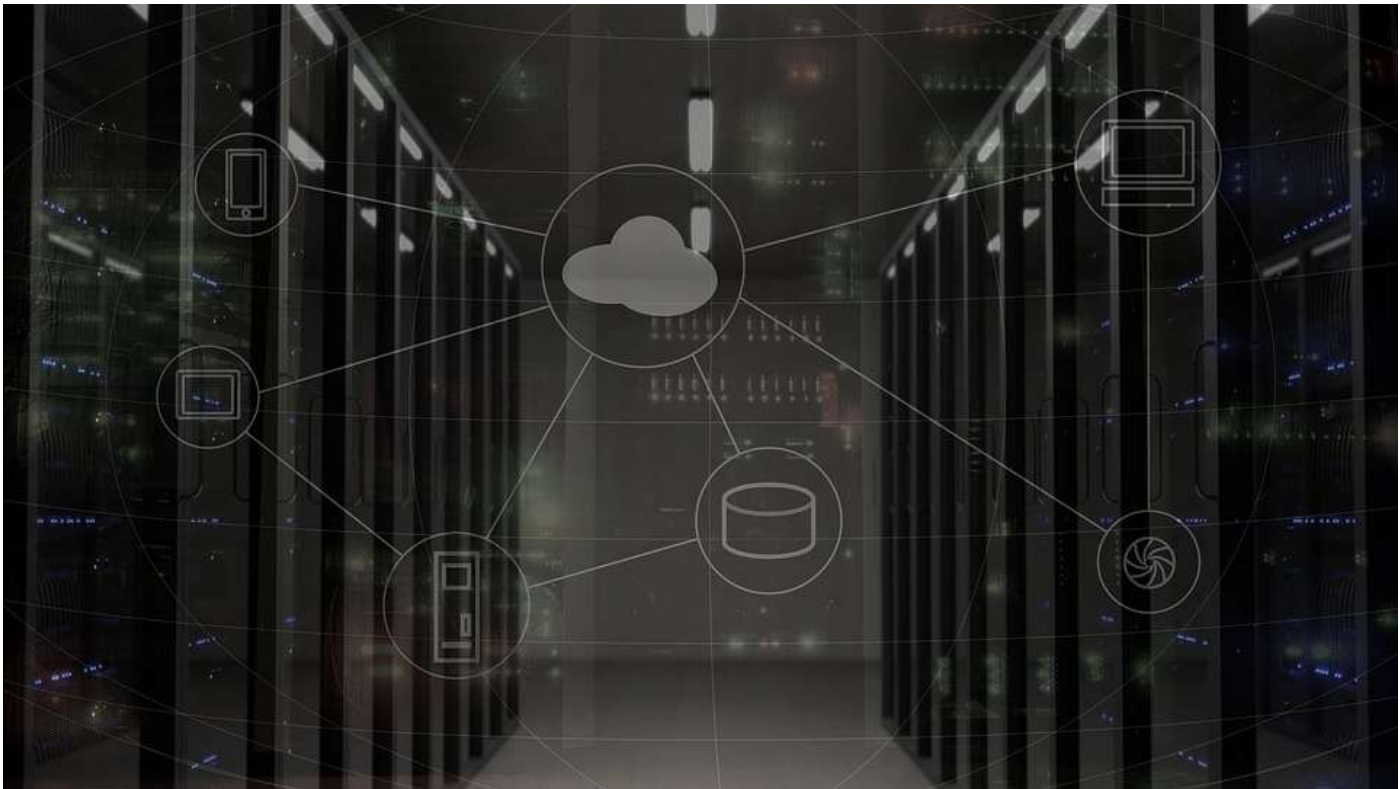


Data Lake on Cloud for Healthcare



About the Client

The client is a global leader in advancing personalized oncology treatment and supporting more effective and efficient cancer drug development. Founded in 2010, client is headquartered in Greater Boston area.



The Challenge

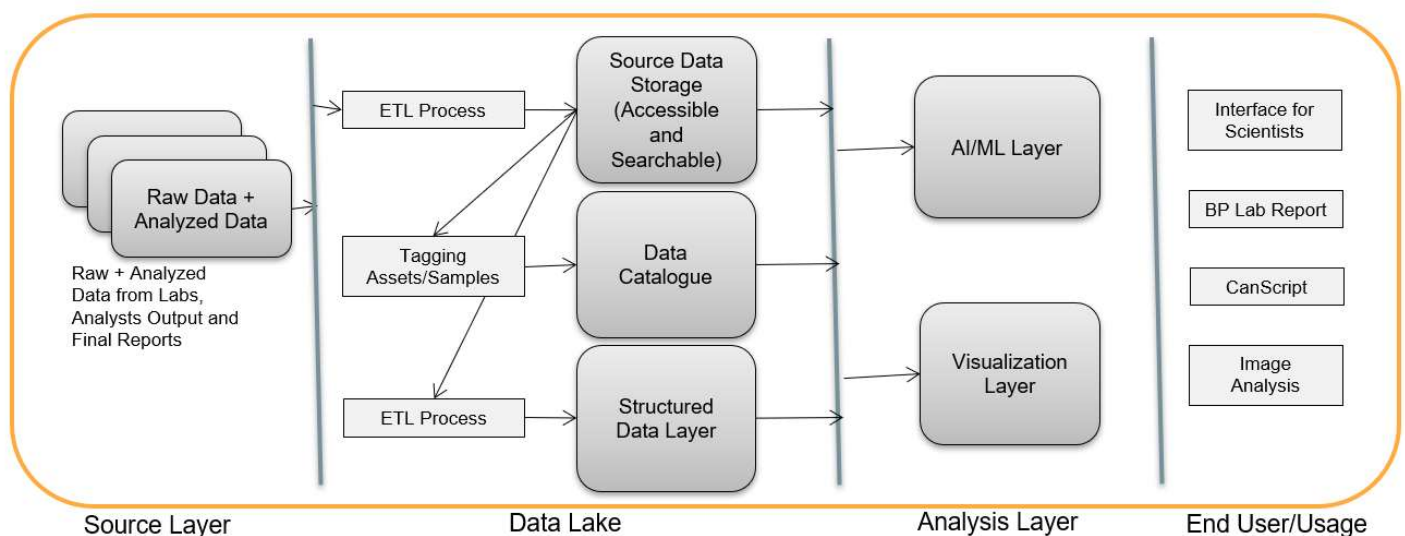
- ✓ How to handle large volume of unstructured and structured data in different machine formats coming from various clinical apparatus.
- ✓ Cataloging all this Raw data, reports from Lab scientist and making this data available for business applications as well as to scientist for future research.

Approach

- Business Scoping and Source Data Exploration
- Cloud based data lake was selected as per client requirement of multi geography data access.
- A solution architecture was defined which included data lake architecture and an analysis layer
- A detailed scoring mechanism was applied to find best cloud platform for client requirement. Platforms considered for this were Microsoft Azure, AWS, Hortonworks on AWS and Google.
- The factors which were considered for the scoring were Modularity, Identity/ Access Management, Usage in Industry, Local to Cloud Integration and Development and Maintenance cost.

Solution

- AWS S3 storage system was used as a base for the data lake
- ETL Framework based on AWS Glue to discover data and store the associated metadata in the AWS Glue Data Catalog.
- Cataloged data is immediately searchable using Elastic Search and available for various business application through various SQL queries.
- Semantic Layer (DWH): Redshift was chosen for the data warehouse services. Data stored in data warehouse was easy to search and retrieve for business intelligence reports and analytics





Copyright @ 2019 www.bizacuity.com

Do you want to empower your decision makers?

3379 Peachtree Road NE(Buckhead), Suite 555,
Atlanta, GA

Contact us: sales@bizacuity.com
US: +1 704 302 1836
Tel: +91 40 – 4311 1808