



Background

One of our Clients, a Fortune 100 health insurer, is migrating financial and operational data warehouse to Hadoop. Data integration from core operational systems will be managed by the IBM InfoSphere Information Server suite of products.

This is a very large migration, which requires redesigning thousands data extracts to integrate with the Hadoop HDFS / Hive file systems. The Client contracted with Technology Advance Partners (TAP) to provide architectural design, best practice recommendations and staff training for the migration.

Our responsibilities are both technical and process leadership.

Technical Goals

The Client has many developers but as this is a major migration spanning multiple years and thousands of hours of work effort, the Client believes appropriate planning, process and methodology is critical to a successful and efficient migration.

- Work with Enterprise architects to assess and remediate the designed framework for corporate analytics and data warehouse on the HortonWorks (Hadoop) platform. This environment analysis a wide scope including:
 - a. Analyzing and assessing installed applications and recommending technical design remediation steps for other required applications
 - b. Reviewing installation and configuration options and recommending changes eliminating functional gaps
 - c. Testing interoperability between all products / environments
- Recommend Big Integrate ETL technical framework best practices to be used to ingest millions of claim, plan and member records daily.
- Diagnose environment and design issues which impact stability or performance in the Big Integrate / Yarn cluster.
- Recommend application redesign eliminating eliminate stability issues and improve throughput.
- Design and build templates demonstrating best practice ETL job design processing XML data in an HDFS environment.
- Install and configure full Information Server / Big Integrate suite, including Information Governance Catalog, Information Analyzer, Metadata Asset Manager, Business Glossary, Subscription Manager, and QualityStage.

Process Goals

- Establish and document Best Practice Development Framework
 - a. Coding practices
 - b. Tools
 - c. Application usage recommendations (when and where to use)





- d. Metadata Management and Information Governance
- e. Provide training as required to implement framework
- Update existing code review practices (peer review) to incorporate new framework / best practices
 - a. Analyze and redesign pre-existing jobs to resolve performance or design issues
- Insure issues / recommendations / solutions are captured in a knowledge base for long term availability

