



Advanced Logic Analytics

Hadoop Datalake Proof of Concept

Outline

A Hadoop datalake is a method of consolidating multiple silos of structured and unstructured data in a way that they can be analysed and cross-referenced within a central repository, generating new insights.

This approach has been recognised to be superior to a SQL database for analytical purposes, especially for organisations with large volumes of data from many sources. A datalake is also a powerful way to break down internal silos and encourage effective co-operation between departments while having strong controls over who can access what particular data.

Objectives

The purpose of this proof of concept is to demonstrate our capability to create a datalake within Hadoop, demonstrate an effective security matrix on this data and provide some visualisation on the central repository for analytical purposes.

A typical proof of concept would involve consolidating 3 data silos into an initial datalake. The selection of these initial data silos would be driven by business needs and data availability for the project. A few examples could be:

- **360 customer analysis** – consolidate transactional, customer and online browsing data to enable better targeting of customers with marketing and to identify new sales opportunities.
- **Fraud and risk analytics** – consolidate transactional, customer and identity data to highlight areas of potential fraud.
- **Knowledge management** – consolidate documents, email, and employee information to allow contextual searching within the organisation for operational or regulatory purposes.

Project outline

- Week 1: Scoping and statement of work
- Week 2: Analysis of internal systems and data availability
- Week 3: Hadoop cluster set up and test imports
- Weeks 4-7: Integration and consolidation of data in Hadoop
- Weeks 8- 9: Implement security matrix
- Weeks 10-11: Presentation layer – dashboards or other visualisations
- Week 12: User acceptance testing

Working team: CTO and 2 data scientists

Indicative timescale: 3 months

Contact us on: + 44 (0)20 3893 7571 or contact@advancedlogicanalytics.com