

Thorogood client story
Big Data Analytics

Big Data Analytics with Microsoft and Hadoop at Whistl

Whistl (formerly TNT Post) is the UK's second-largest post provider handling 1.2 billion items of mail a year. Part of the large Whistl group, the organization has grown organically for the last 10 years and anticipates much further growth going forward.

The Vision

To maximize the analytic benefits of this increased volume of data, Whistl were keen to identify a cost effective solution that would deliver:

- easy, timely access to both current and historic business data
- a single corporate "active" archiving approach
- reduced cost of storing business data for reporting
- governed access to business data
- scaling regardless of data volumes
- a future-proof way to store and retrieve business data

A number of considerations needed addressing including technology choices, handling data access and governance, data maintenance and costs. Whistl already used Microsoft Business Intelligence tools to search for insights to help to improve service and grow market share.

Fulfilling the Vision

With its open-source framework, Hadoop provided a powerful option which allows data to be distributed and processed over a wide network of low-cost servers.

Whistl have invested heavily in Microsoft's Business Intelligence technology and have a wide range of Microsoft skills available within their organization. The key for this proof of concept was therefore to implement an architecture which allowed Hadoop to work in tandem with the Microsoft technology.

With the help of Thorogood, Whistl set up a proof of concept to link Hadoop to their existing Microsoft environment. This retains "hot" data (frequently accessed, high degree of change) in Microsoft SQL Server and moves "cold" data (infrequently accessed and not modified) into an active Hadoop archive that is always online. The exercise has shown how these two technologies can be combined in an analytic view with key business benefits:

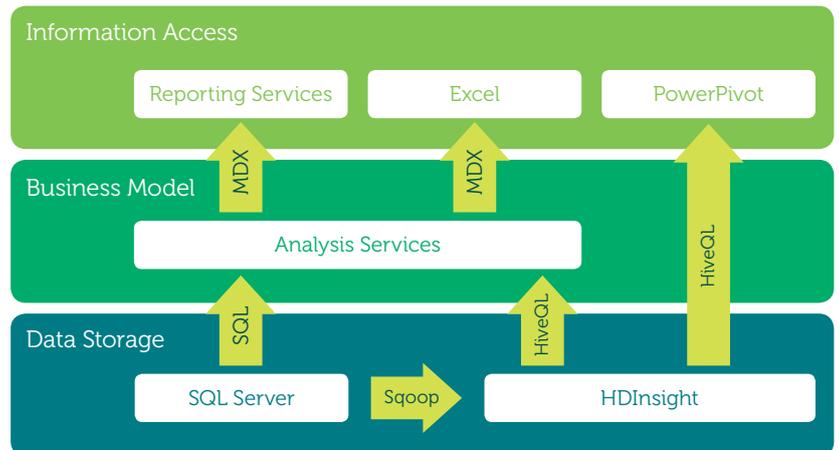
- Hadoop can be seamlessly integrated into the existing Microsoft Business Intelligence structures - business users do not have to learn any new tool sets to interact with the active archive
- All historical data throughout the organization can be available 24/7 for direct reporting

The use of Hadoop need not be limited to archiving alone. It can be further used to store and process weather data, image files, etc., thus providing other important benefits.

Technologies Used:

- Hadoop
- Hive
- Sqoop
- Microsoft HDInsight
- Microsoft SQL Server Database Engine
- Microsoft SQL Server Analysis Services
- Microsoft SQL Server Reporting Services
- Microsoft Excel
- Microsoft PowerPivot

The architecture combines a number of Microsoft technologies to deliver capability and value.



Next Steps

So, how does Whistl now see its choices? Doing nothing isn't an option if the organization wants to mature the way it handles data as a business. Both centralized and decentralized solutions based on their current Microsoft SQL Server solution have both scale and cost concerns. Hadoop looks to provide them with what they need.

The next step will be a proposal for the development of a fully integrated Hadoop solution as a basis for Whistl to fully exploit its extensive data assets. The team will be educating business stakeholders on the challenges and benefits of Hadoop through clearly explained data growth and coping strategies, showcasing data access options and highlighting likely costs going forward.

If you would like to know more about what can be achieved with Business Intelligence and Analytics and would like to discuss your options with an independent specialist, please get in touch with Evelyn Heyes in the UK (evelyn.heyes@thorogood.com) or Trevor Jones in the US (trevor.jones@thorogood.com).