Offloading Mainframe Applications to Hadoop

Abstract

Enterprises run critical applications on mainframe systems that, along with hosting critical corporate data, generate and process huge volumes of data. While legacy mainframe systems are an essential component of enterprise IT, they involve high maintenance costs and often lack the capability to support new business requirements through rapid development cycles.

Hadoop is a viable alternative to mainframe batch processing and storage because of its scalability, fault tolerance, quicker processing time and cost effectiveness. This paper discusses the main challenges facing mainframe systems and proposes alternatives based on Big Data ecosystems.

Introduction

Legacy mainframe systems used by organizations are scalable, secure, available, and reliable machines capable of processing large workloads. However, they incur huge hardware, software, and processing costs. Moreover, with the dearth of mainframe technicians in the current workforce, maintenance itself has become a challenge. Hence, many organizations feel the need to partly or wholly migrate business applications involving batch processing running on mainframe systems to contemporary platforms. This exercise forms an important part of their strategy to help reduce costs, meet evolving business requirements, mitigate risk, consolidate IT systems, and enable them to support the business in an agile manner.

With the emergence of Big Data technologies, which are cost effective, scalable, and fault tolerant, mainframes' maintenance and processing expenses can be reduced by integrating a Hadoop layer or completely off-loading batch processing to Hadoop. This paper highlights the role of Big Data technologies in addressing the limitations of mainframe systems and meeting organizations' IT needs.

Industry landscape

Legacy mainframe systems face several limitations such as high maintenance costs, inability to support new business requirements quickly and ineffective utilization of systems. In addition, performing analytics on mainframes systems is cumbersome because, compared with the latest visualization tools, Graphical User Interfaces (GUIs) are not adequately supported by mainframes systems.

These limitations can be effectively addressed by Hadoop, a programming framework that is reliable, scalable and enables distributed processing of large data sets across commodity server clusters with a high degree of fault tolerance.

Hadoop helps lower the cost of batch processing while enabling faster development cycles and easier maintenance of code to facilitate reductions in project delivery time.

Mainframe batch processes can be scaled up and efficiently run on Hadoop at a fraction of the cost and time. Migrating mainframe applications to Hadoop is now a viable proposition because of its flexibility in upgrading the applications, improved short term return on investment (ROI), cost effective data archival and the availability of historical data for querying.

Traditional data archival methods used by mainframe systems lack the flexibility to query historical data. Hadoop, however, allows optimal query performance so that historical data can be retrieved in time for further analysis. Hadoop offers several

advantages over mainframe batch processing. To limit the effort, time and cost that a total transition would entail, enterprises can implement an optimal solution where the mainframe systems co-exist alongside the Hadoop ecosystem.

Mainframe-Hadoop: The Co-existence Approach

To implement a mainframe Hadoop system, the following guidelines need to be kept in mind:

- Use of mainframe systems should be restricted to core activities to reduce costs by reducing Million Instructions Per Second (MIPS), and Hadoop should be used to supplement back end processing.
- Batch processing should be offloaded to Hadoop by migrating Central Processing Unit (CPU) intensive batch processes such as transformations and aggregations to reduce the batch processing cycle.
- Storage intensive data should be moved to Hadoop for warm archival to reduce operational and maintenance costs, and run ad-hoc queries as required for business and compliance purposes.

Data mining and analytics needs can be addressed using Hadoop and the distributed analytical models of Mahout Libraries. As with traditional BI analytics, analytics on Hadoop need not be restricted to limited volumes of data and also it can be extended to unstructured and semi structured data as well.

The following approach can be utilized to migrate some of the applications to a Hadoop ecosystem:

- Identify CPU intensive jobs with voluminous data and batch processes with 'write once read many' characteristics, and non-sequential processing.
- Remodel the existing systems to exploit the distributed processing capabilities of Hadoop, eliminate the limitations of legacy platforms, and re-engineering the business process to remove redundancy
- Develop the modules in parallel and deploy the solution. The code developed in one Hadoop distribution can also be deployed across other Hadoop distributions with minimal changes.

- Test performance and scalability.
- Integrate with existing enterprise applications and standards.

Benefits of adopting the Mainframe-Hadoop coexistence approach

Adopting a Mainframe-Hadoop coexistence approach allows enterprises to address data mining and analytics needs using Hadoop and the distributed analytical model of Mahout Libraries while leveraging accumulated legacy data for information discovery.

Batch processing can be performed effectively with Hadoop ecosystems using PIG, Hive and MapReduce. Batch jobs can be taken off from mainframe systems and processed on Hadoop, and the output can then be moved back to mainframe or other systems, thereby reducing MIPS costs.

Huge volumes of structured and unstructured data as well as historical data can be leveraged for analytics instead of restricting it to limited volumes of data in a bid to contain costs. This helps improve the quality of analytics and offers better insights on a variety of parameters to create value.

Conclusion

Hadoop can be an alternative to mainframe batch processing and storage, as it fits well among COBOL, VSAM, and other legacy technologies. By migrating to Hadoop, mainframe batch processing can be done efficiently, quickly and at a lower cost. Hadoop code is easily maintainable as well as flexible, which aids in building new functionality, and facilitates swift development with the added benefit of faster project delivery times. Migrating mainframe applications to Hadoop is now a viable proposition due to reduced infrastructure and batch processing costs. In addition, Hadoop programs offer flexible options for upgrading applications and help increase the return on investment in the short term.

About The Author

Rajasekhar Reddy

Rajasekhar Reddy Pentareddy has over 11 years of IT experience spanning multiple technologies and domains. His areas of expertise include Big Data, Data Warehousing and Business Intelligence and he currently works as a Technical Architect with the Analytics Big Data and Information Management (ABIM) practice of TCS.

Contact

Visit TCS' Analytics & Insights unit page for more information

Email: analytics.insights@tcs.com

Blog: Digital Reimagination

Subscribe to TCS White Papers

TCS.com RSS: http://www.tcs.com/rss_feeds/Pages/feed.aspx?f=w
Feedburner: http://feeds2.feedburner.com/tcswhitepapers

About Tata Consultancy Services Ltd (TCS)

Tata Consultancy Services is an IT services, consulting and business solutions organization that delivers real results to global business, ensuring a level of certainty no other firm can match. TCS offers a consulting-led, integrated portfolio of IT and IT-enabled, infrastructure, engineering and assurance services. This is delivered through its unique Global Network Delivery Model™, recognized as the benchmark of excellence in software development. A part of the Tata Group, India's largest industrial conglomerate, TCS has a global footprint and is listed on the National Stock Exchange and Bombay Stock Exchange in India.

For more information, visit us at www.tcs.com

TCS Design Services I M I 12 I 16