Solving the Biggest Challenges of Big Data

ADOPTION • SCALE • SPEED • ROI

EDW MODERNIZATION • BI ON BIG DATA • FAST DATA ANALYTICS
# Table of Contents

Introduction ............................................................................................................................... 3

Case study: Multinational Retail Chain  
Saves Millions by Migrating Workloads from  
Traditional Data Warehouse to Hadoop ................................................................................ 4

Case Study: Leading Global Investment  
Bank Minimized Risk Exposure ............................................................................................... 6

Case Study: Global Telecom Service Provider Offers  
Better Programming with Viewer Analytics ............................................................................. 8

Case study: Leading Wireless and Telecom Service  
Provider Reduced Annual Call Center Cost .............................................................................. 10

Case Study: Real-time Call Center Monitoring ....................................................................... 12
Introduction

Welcome to the real world of enterprise class big data analytics!

Our team at Impetus Technologies has been helping enterprises deliver on the full promise of big data since the very early days of Apache Hadoop. Along the way, we have developed service and software product solutions that help solve the biggest challenges that enterprises face in their big data journey:

- Driving for increased real world adoption of production scale technologies that deliver improved decision-making based on deeper insights at all levels and functions across the enterprise.
- Taking advantage of unprecedented scale in the number, size and variety of available data sources.
- Achieving greater speed at scale.
- Realizing true ROI from their big data investments.

Enclosed you will find five case studies representing a range of industries—all showcasing how these companies are overcoming the challenges inherent in big data projects and are delivering big business impact from big data.

These case studies revolve around the following Impetus capabilities:

- Impetus Data Warehouse Modernization services and solutions (www.impetus.com).
- Fast data solutions like Impetus StreamAnalytix (www.streamanalytix.com) that deliver near real-time decision velocity.
- Delivering immediate value from enterprise data lake investments using Kyvos (www.kyvosinsights.com)—a BI/OLAP platform that provides very fast, interactive reporting and analytics at massive scale by creating a powerful BI consumption layer on top of your enterprise big data lake.

Best wishes with on own journey into the world of the real-time, data-driven enterprise!
Multinational Retail Chain Saves Millions by Migrating Workloads from Traditional Data Warehouse to Hadoop

Business Challenge
Optimize and reduce the current cost of data warehouse and build a data lake for faster ELTL processing, advanced analytics and reduced time to analytics.

Benefits
- Cost Savings
- Migration to modern technology
- Improved user experience
- Risk mitigation

Client Overview
A U.S.-based multinational retail chain wanted to optimize the current Teradata Data warehouse costing millions of dollars every year and ensuring that its analytical applications are using the most appropriate data platforms. To achieve this, the client wanted to migrate batch applications and ad-hoc uses that consume expensive Teradata cycles to Hadoop/ Hive environment.

Requirements
The client wanted to identify and migrate the batch and ad-hoc workload from Teradata to Hadoop systems.
- Batch migration includes ingestion of dependent datasets from true source, building batch jobs, rollout, upload manufactured data to target systems, data validation, warranty support and production hand-off.
- Ad-hoc migration includes establishing access patterns, ingestion of dependent datasets from true source, query conversion including conversion of Teradata specific operators to the target systems and user-training assistance.
**Solution**

The Impetus solution involved two key migration:

**One time migration**
- Historical data (from Teradata or true source). Scripts and queries used in batch and ad-hoc fashion to access Teradata.
- Movement of analytical post-processed data from Hadoop to Teradata (if required).

**Incremental migration**
- Daily, weekly, and monthly increments of data from true source to Hadoop.
- Movement of analytical post-processed incremental data results from Hadoop to Teradata.

Workload Migration solution was implemented for historical data migration wherein it identified and offloaded data and workloads from the enterprise data warehouse to Hadoop.

The solution features automated utility that converts BTEQ and SQL transformation scripts into equivalent Spark QL/Hive QL and executes them on Hadoop/Hive environment. It also allows users to run a set of data validation checks.

Finally, the post-processed analytical data can be loaded back to the source enterprise data warehouse for reporting and usual access.

Further, a combination of Workload migration product and IBM Data Stage were used for incremental data migration from the true source of data to Hadoop and movement of processed results to Teradata or other database.
Benefits

- Improved risk compliance with a daily consolidated view and analysis across all asset classes, incorporating data from 56 billion trades
- Kyvos BI consumption layer allowed risk analysts to have interactive access to individual trade data via four familiar UIs - Tableau, Excel, Spotfire, and an in-house custom application
- Power to perform 90-day trend analysis, updated daily, to understand risk behavior and correlate the impact of risks across assets

Background

The investment bank arm of this global financial institution faced a requirement to limit its daily risk exposure across its entire business. While it was an easy decision to load the large data volume and variety into Hadoop, the Risk Analytics team couldn’t meet the mandate to analyze their trading positions daily across all asset classes: foreign exchange, equities, fixed income, and other special products. They wanted the ability to view risk trends to see how one asset class risk may affect another. They set out to deliver a daily single consolidated view of the investment bank’s risk position with the ability to drill down into the individual trades.

Challenges

- With over 700 million daily transactions, the bank was struggling to create a consolidated view of risk across their assets
- It was impossible to correlate risk across asset classes to understand trends
- Each asset class stored data in a different platform, silos: Teradata, Oracle SQL Server, and HDFS
- Different teams relied on different BI tools
  - Analysts were unable to drill down their data to understand the cause of problem transactions
  - Risk analysis was always late and deficient
“We tried all the open source tools—Hive, Impala, Spark; and found that with all their inefficiencies, nothing worked. Kyvos outperformed all these tools and other vendor tools we tried!”

—VP Risk Derivatives

**Solution**

Using Kyvos, the risk analytics team was able to prepare the data, create a common data schema for all asset classes, and build a BI Consumption Layer with transaction-level data directly from Hadoop on a daily basis. Kyvos used its built-in integration with Tableau, Excel and Spotfire, and integrated to the bank’s own custom application (using Kyvos’ Java APIs), allowing analysts to continue to work in a familiar environment while enabling access to data across assets. To ensure that access to the data lake had the same enterprise security that the bank required, Kyvos integrated with the company’s LDAP and Kerberos protocols.

The risk analytics team very quickly built an OLAP cube with 20 days of trading data – something the investment bank had not been able to achieve before with open source tools and other third party solutions. Data was added to the cube incrementally each day for analysis until the OLAP cube reached the team’s goal of 90 days of data—approximately 56 billion trades.

**Results**

Bringing data from all the asset classes together and analyzing them directly in Hadoop has transformed how the investment bank manages its risk. The bank was finally able to deliver risk assessments daily across all asset classes without delay. With Kyvos, the investment bank achieved:

- A daily consolidated view of its risk with 90-day trend analysis across asset classes
- Self-service analytics for its team to ensure adoption success
- Drill down access to the trade transaction level for troubleshooting

The risk analytics team now has a firm grasp of their risk positions on a daily basis and over a longer-term period. After nearly two years of trial and error with Hadoop, the global investment bank was finally able to tap into the power of big data to stay ahead of its risk management requirements.
Global Telecom Service Provider Offers Better Programming with Viewer Analytics

Benefits

- To better prepare for programming negotiations with media companies, the marketing team ran ad hoc queries on viewing data from nearly 30 million set-top boxes
- A 13-month OLAP cube of more than 14 TBs of data was deployed in production within three weeks
- Kyvos—Tableau integration means Tableau extracts were no longer needed, achieving time and resource savings

Background

Serving their customers is an important mission to one of the largest telecommunications companies in the world. With access to so much rich data from live TV, DVR, pay-per-view, video-on-demand, STB (set-top boxes), and other devices; the marketing department knew that this data could help them provide superior customer experience and a competitive advantage.

Starting with data from the STBs, the marketing department wanted to better understand viewer behavior to prepare for programming negotiations with media companies. Machine data from nearly 30 million STBs could provide insights into who was watching what television show at any point in time.

Challenges

- Massive volume of customer behavior data generated from STB’s with over 6 million accounts
- Ad-hoc requests from analysts’ queries took hours or even days to turn around
- Impossible to do year-over-year and month-over-month comparisons as it would take weeks to answer questions on such huge data
We were impressed with the Kyvos integration with Tableau. No more Tableau extracts is a huge savings for us in terms of time and resources. It’s like Kyvos opened up the world of big data for Tableau.

—Senior Director, Big Data

Solution

The telecommunications company believed that to achieve a successful adoption of the analytics tool, it was important to keep their analysts in the familiar environment of Tableau. IT successfully implemented Kyvos and built a BI Consumption Layer, which integrated with Tableau to provide a familiar analytics environment for their marketing team. To provide the interactive response time that marketing needed to analyze the data, IT built a 14 TB OLAP cube with multiple dimensions of data – all while achieving sub-second response times.

The Kyvos integration with Tableau enabled seamless access for analysts and that meant setting up and processing Tableau extracts was no longer necessary.

Results

Building a BI Consumption Layer in Hadoop with Kyvos was a significant advancement for the telecom company. The marketing team now has interactive access to 13 months and 14 TB of granular data from over 30 million STBs. Insights from the BI Consumption Layer can help the company build competitive advantage and save on operating costs. With the OLAP cube, they are able to understand media metrics and trends, programming successes and failures, viewer behaviors by demographics, etc.—crucial information to negotiate better programming venues from their media channel partners.

With the Tableau integration, the analysts’ environment is the same familiar tool but with interactive query response times to their big data. Marketing analysts can now answer all of their questions and drill down to the customer interaction details.

Ultimately, the Kyvos OLAP on Hadoop implementation enabled this multinational telecom company to have deeper customer behavior insights, which in turn helped to create better programming for their customers and operational cost savings.
Leading Wireless and Telecom Services Provider Reduced Annual Call Center Cost

Business Challenge
To optimize call center cost by tracking the desktop activities of the call center representatives in real-time

Client Overview
A leading U.S.-based wireless and telecommunications service provider wanted to improve call center performance, increase customer satisfaction, and have greater insight into the activities of its call center representatives. To achieve this, the Fortune 50 Company wanted to analyze the desktop activities of the call center representatives around the clock.

Requirements
In an effort to improve performance metrics, the client wanted to monitor desktop activities in real-time while the representatives are on duty. From an operational perspective, this meant creating a centralized system where operations personnel would be able to:

- Track idle time
- Track what websites are being used for how much time
- Track outlook usage
- Track various applications being used on the desktop

The client also wanted to track desktop activities when the agent are:

- On call
- Not on call
- On call and kept customer on hold

Benefits
- Annual overall cost reduction of $5 million
- Improved agent productivity with ability to handle more than 30 calls per day
- Improved customer experience
- Reduced Agent idle time to 15 minutes per day
- Reduced overall after call work activities of agents to 30 minutes per day
- Handling of CPNI information compliance
- Identification of anti-company
Solution

StreamAnalytix delivered a three part solution:

1. The team developed a Data Collector component to ingest data from multiple sources and send it to the respective Kafka topics.

2. Built-in Kafka Channels were used to ingest the data further in a Storm pipeline and process them. The following StreamAnalytix bolts were used to process the data:
   - Enricher Processor: For providing support to look up and enrich the raw data by adding more metadata required for further correlation.
   - Timer Processor: To collect the events within time-based window and sort them to maintain the sequence of events.

3. Further, StreamAnalytix persister components were used to persist processed data in HDFS, ElasticSearch and Apache Phoenix.

The solution enabled the client to improve agent productivity dramatically by reducing idle time. It also increased customer satisfaction and handled CPNI information compliance.
Real-time Call Center Monitoring

**Benefits**

Call centers process millions of minutes of calls per day across vast distributed networks around the globe. StreamAnalytix provides an infrastructure monitoring platform that allows a unified view and analysis of events in real-time.

**Client Overview**

A leading cloud-based communications technology company that offers hosted contact center services needed a way to improve performance metrics, eliminate the guessing game of problem resolution and dramatically increase customer satisfaction. To attain this, they wanted a unified view into their infrastructure that would allow them to monitor calls in real-time.

**Challenges**

In the battle for consumer loyalty, the contact center is at the heart of customer care strategies. It is the central hub of communications and customer service for enterprises and is responsible for the vast majority of consumer interactions and service-related transactions in today’s market. The customer service touch points—such as resolving a complaint, taking an order, renewing a warranty or up-selling a product—are pivotal in accomplishing strategic business objectives.

Enterprises are increasingly realizing that running a call center is not their core business and are therefore choosing cloud-based contact center services. The sole focus of these call centers is to create positive customer service experiences, and they must do so across multiple channels of interaction as quickly and as economically as possible.
As a result, contact centers are under pressure. They must adapt rapidly to the changing technology landscape and the demands of ever increasing customer expectations. As you can see in the diagram below of a hosted call center solution, events are scattered in different media servers and networks across vast geographies. Correlating the entire puzzle is a complex and incredibly complicated proposition.

In an effort to improve performance metrics such as call abandonment rate, average speed of answer, and average call length, the client wanted to monitor the activities of every call in real-time. From an operational perspective, this meant creating a centralized system where operations personnel could:

- View the behavior of the call center infrastructure
- Trace the complete call flow
- View current calls happening in real-time
- Generate reports that could predict future hardware capacity needs
- Search calls based on various filters such as phone number, caller name and time in queue

**Solution**

StreamAnalytix delivered a five-part solution:

1. **IVR Call Flow**
   Call “stitching” in real-time that includes the ability to view, sort, filter and zoom into a call.

2. **Dominant Path Flow**
   Insight into the top 10 most dominant paths a customer follows, including the ability to report the IVR abandonmen path.

3. **SLA Alerts**
   Service level alerts in real-time allow managers to escalate issues and resolve them as they are happening.

4. **Sentiment Analysis**
   The system performs real-time, multi-lingual classification and sentiment analysis of text data, including the ability to generate alerts on email and conversations happening in real-time.

5. **Predictive Analytics**
   A reporting tool provides the ability to generate historical reports for future pricing models and requirement identification. The reports can be viewed on the UI for analysis and enabling business decisions.

**Conclusion**

- Numerous person-months of productivity gain
- Improved customer complaint resolution speed
- Better customer satisfaction index
- Higher customer retention rates
About Us

Impetus Technologies, Inc. is a 20 year old U.S. corporation headquartered in the heart of Silicon Valley, California. We are in the business of helping Fortune 1000 enterprises deliver on the full promise of big data. We have more than 1600 employees in the U.S and in our three labs and engineering centers in India--where we have been recognized for four years in a row as a top 100 Great Places to Work company. We provide full life cycle software engineering and data science services, as well as software solutions for envisioning, architecting, implementing and supporting end to end big data analytics.

Kyvos is a disruptive Big Data solution that delivers the fastest BI on the planet at massive scale. Our patent-pending OLAP technology enables Fortune 500 companies to query billions of rows of data within seconds and helps business leaders make informed decisions. We harness the true potential of data lakes in partnership with industry leaders in BI, Cloud, and Hadoop technologies. Our breakthrough technology enables analysts to interact with data in real time, using their favorite BI tools. Kyvos has been shown to be 100x faster than Hive and Impala on standard benchmarks.

Impetus StreamAnalytix platform is enabling the real-time enterprise. The platform provides an elegant user-interface over Apache Spark Streaming and other Big Data technologies, allowing users to build, deploy and monitor applications easily, in hours or even minutes. Enterprise IT teams can now provide a centralized development platform for their developer teams building different streaming analytics applications.