HADOOP FOR ENTERPRISES: OVERCOMING THE MAJOR CHALLENGES
INTRODUCTION TO BIG DATA

BIG DATA ARE INFORMATION ASSETS THAT ARE HIGH VOLUME, VELOCITY, AND VARIETY. BIG DATA DEMANDS COST-EFFECTIVE, INNOVATIVE FORMS OF INFORMATION STORAGE AND PROCESSING FOR ENHANCED INSIGHT AND DECISION-MAKING.

VOLUME
Machine-generated data is produced in considerably larger amounts than traditional, human-generated data. For example, high-resolution video used in surgeries generates 25 times the data volume (per minute) of even the highest resolution still images such as CT scans require thousands of times more space than a page of text or numerical data. More than 95 percent of the clinical data generated in health care is now video.

VELOCITY
Social media produces a large, constant stream of data containing opinions and relationships that are highly valuable to customer relationship management. Facebook’s 600 million active users spend more than 9.3 billion hours a month on the site. Every month, the average Facebook user creates 90 pieces of content and the network itself shares more than 30 billion items of content including photos, notes, blog posts, Web links, and news stories.

VARIETY
Traditional data formats are strictly configured and change slowly. Conversely, non-traditional data formats change day by day. As new services are added, new protocols are deployed, or new marketing campaigns are executed that require new data types to be processed.
Big data typically refers to the following types of data:

**Traditional Enterprise Data**
Examples are customer information, transaction logs, and general ledger data.

**Machine- or Sensor-Generated Data**
Examples are Call Detail Records ("CDR"), smart meters, weblogs, equipment logs (often referred to as digital exhaust), and trading systems data. Forecasts by IBM, Intel, and others indicate that more than 60 billion intelligent devices exist in the world today and that this figure will rise to more than 200 billion by 2015 as the “Internet of Things” becomes a tangible reality.

**Social Data**
Examples are customer feedback streams, micro-blogging sites such as Twitter, and social media platforms such as Facebook.

Why Big Data for Enterprises?

For decades, companies have been making business decisions based on transactional data stored in relational databases. Beyond that critical data, however, is a potential treasure trove of non-traditional, less structured data such as weblogs, social media, email, sensor data, and photographs that can be mined for useful information.

When Big Data is distilled and analyzed in combination with traditional enterprise data, enterprises can develop a more thorough and insightful understanding of their business, which can lead to enhanced productivity, a stronger competitive position, and greater innovation — all of which have a significant impact on the bottom line.
BIG DATA USE CASE EXAMPLES BY INDUSTRY:

**FINANCIAL SERVICES**
Big Data technologies allow risk managers to improve statistics and forecast accuracy with greater granularity, resulting in better risk assessment and the creation of optimized product offerings targeted toward individuals.

**ONLINE BUSINESSES**
Using a Big Data framework, online businesses, media agencies and content publishers can get detailed data about customer preferences and social interactions, cross-selling, up-selling, and recommendation models leading to better understanding of customer behavior and priorities, as well as significantly increased revenue.

**ENERGY AND UTILITIES**
With Big Data technology adopted, Energy and Utility companies can receive detailed data about individualized use patterns. The broader incorporation of external demand influences leads to significantly improved demand management and groundbreaking energy savings.

**TELECOMMUNICATIONS**
Big Data provides telecom operators with proper storage and processing of Call Detail Records and allows proactive user behavior analysis, clustering of subscribers, and numerous new dimensions for prediction that offer the possibility of building more progressive business models and maximizing revenue.

**RETAIL**
In the Retail sector, Big Data allows users to get enhanced insights and understanding of customer likes, dislikes, influences, and behaviors, resulting in increased sales and profitability.

**TRANSPORTATION AND LOGISTICS**
By using Big Data approaches in Transport and Logistics, analysts are able to optimize routes, loads and resource usage, that leads to improved services and new service options.

**GENERAL BUSINESS**
Big Data adds value to every business by tapping into data about consumer behaviors, attitudes, social relationships, and influences, thereby optimizing customer engagement and ultimately, profitability.
WHAT ARE THE BIG DATA ADOPTION ISSUES AT THE ENTERPRISE LEVEL THROUGH 2015?

According to latest analytical researches, business analytics requirements within the Enterprise sector will drive 70% of investments in the expansion and modernization of the information infrastructure. However, recent analytical research shows that 85% of Fortune 500 organizations will be unable to exploit Big Data for competitive advantage. The top business constraints against Big Data technology adoption are:

- Analytics Capabilities and Skills
- Infrastructure and Architecture
- Investment, Budget, ROI
- Privacy and Security Risks

70% of IT budgets within the Enterprise sector will be spent on expansion and modernization of the information infrastructure.

30% of other Enterprise IT spendings.

85% of Fortune 500 organizations will be unable to exploit Big Data.

15% of Fortune 500 organizations will successfully exploit Big Data.
_without a doubt, big data is a major trend in the worldwide IT space. Various whitepapers, analytical research papers, and independent articles draw an attractive, but abstract picture of big data’s advantages. However, enterprises frequently don’t see the concrete ways in which big data can add value to their business and don’t have the experts inside the company who can profitably deploy and integrate a big data solution.

luxoft’s solution:
that’s why we created the big data solution accelerator: a unique set of big data application development services that combine innovative approaches to comprehensive information storage, processing, and analysis with fundamental multi-industry experience and efficient, low-risk system adoption for private enterprises.

the solution accelerator is based on an iterative approach that allows users to see results quickly, moving from small proof of concept to long range goal.

our customers benefit from our unique preliminary assessment, which contains a business evaluation and consulting with our big data scientist. this detailed review and comprehensive analysis allows us to find the right big data solution that will ideally balance cost-effectiveness, functionality, and needed capacity.

our experts have significant experience in creating enterprise-wide, business critical solutions based on big data technologies and are experts at delivering tangible benefits to our clients. our architects, consultants, and developers each have up to seven years of experience in big data technologies (both open source and proprietary) and can propose effective solutions for a variety of business cases.

luxoft - hadoop for enterprises: overcoming the major challenges
BIG DATA SOLUTION ACCELERATOR BY LUXOFT

Stage 1: Business Assessment
- Deep dive into the client's business with a detailed assessment of needs
- Resources identification (client resources, technologies, planned budget)
- Identification of potential benefits for Big Data technology adoption

Stage 2: Big Data Scientist Consulting
- Analysis of client's functional business needs
- Big Data technologies proposal
- Industry-specific implementation approach proposal
- Detailed assessment of client's technology environment
- Client's corporate standards, policies and security requirements assessment
- Identification and assessment of systems to be integrated with the Big Data solution

Stage 3: Architecture Design
- Architecture proposal that considers industry-specific business and technical environments and proposed technologies
- Architecture design, including vision, goals, priorities, and features
- Architecture optimization and governance, including rationalization, blueprints, guidelines, and refactoring

Stage 4: Application Engineering
- Software development
- Performance Engineering
- Software Quality Assurance and Testing

Stage 5: Integration Services
- Transparent incorporation to corporate infrastructure
- Seamless integration with corporate-level tools and platforms
- Security and processes compliance
- Systems orchestration
- Disaster recovery

Stage 6: Support
- L1: Front-end support
- L2: Administrative level support
- L3: High-end support
- L4: Escalation point beyond the organization
CONTRAINT:
The Big Data paradigm as well as the Hadoop framework was initially created for internet companies, such as Google and Facebook. Making the framework usable for business analytics on an enterprise level requires adapting it for private infrastructure, security, and process requirements, which are strict. Also, any enterprise has relational databases within its infrastructure that can’t be fully replaced by Hadoop at the moment. An effective merge of Hadoop with Relational principles can become a nightmare for an internal IT department.

LUXOFT’S SOLUTION:

Luxoft offers a smooth, low-risk Hadoop implementation. We know how companies can optimize the use of Hadoop with other enterprise systems to improve overall analytical throughput and to build new data-driven products. Our expertise includes:

• Ways to achieve high-performance integration between Hadoop and relational-based systems
• Hadoop+NoSQL versus Hadoop+SQL architectures
• High-speed, massively parallel data transfer to analytical platforms that can aggregate web log data with granular fact data
• Strategies for freeing up capacity for more explorative, iterative analytics and ad hoc queries

“WITH THE AIM OF REUSING EXISTING ARCHITECTURE, COMPANIES MUST PLAN TO ADOPT BIG DATA TECHNOLOGIES THAT COMPLEMENT CURRENT INVESTMENTS.”


Without question, Apache Hadoop is an open-source framework that provides a cost-effective and innovative approach to advanced Big Data analytics and potentially can be used as a core Big Data technology for enterprises. Hadoop provides scalable, reliable, distributed computing with significant cost savings because of its license-free distribution.

Luxoft’s architects, consultants, and developers understand the Hadoop platform and its unique risks to private enterprises. Luxoft has significant experience in the development of large, business-critical applications combined with a mature expertise in Hadoop-based systems development adapted to unique corporate requirements and rules.

We have “hands on” expertise in the development of large, business-critical applications combined with a mature expertise in Hadoop-based systems development adapted to unique corporate requirements and rules.

Luxoft is a partner of Cloudera, the market leader in Big Data platforms for enterprises, and implements their innovative technologies to maximize Hadoop adoption value for its clients.
The Enterprise IT community is still distrustful of solutions using Big Data technologies because the ROI versus the needed investment are unclear. Today’s CFOs want to see empirical evidence of business value before the company invests a significant amount of money.

**LUXOFT’S SOLUTION:**
To realize the substantial benefits of Big Data, Luxoft begins with a clear strategic understanding of how the data will be used (i.e., market clarity or product distinction) and its business value. Then, we apply a rigorous approach to spending that balances the cost of deployment with the benefits accrued to insure a speedy return on investment. Luxoft recommends an iterative approach, beginning with a Proof of Concept (PoC) and small, short-term projects that add value while working toward the proposed global solution. This way, the client will see benefits quickly and in terms of their own business goals and environment. Luxoft’s cost-effective Big Data deployment approaches include:

- Innovative processing that leverages low-cost servers/CPUs
- Use of the increasingly robust array of open-source and freeware technology
- “Start Small and Scale at Need” information processing model
PRIVACY AND SECURITY RISKS

CONSTRAINT:
Hadoop isn’t designed to be a secure processing environment, making security a risk and a requirement for enterprises using the technology for their Big Data initiatives. To compound the risk, many companies centralize traditional and non-traditional data - and centralized data is the most difficult to secure. Securing large volumes of highly dynamic, unstructured data is a known challenge.

LUXOFT’S SOLUTION:
At Luxoft, we provide Big Data solutions for many global enterprises. Multinational investment and retail banks, which have the strictest security policies, trust our company to secure their data. Based on our experience, we understand that there is no single solution to security challenges for all enterprises and our experts work diligently to identify the security strategy that will ideally fit the rules and processes within your company.

CONCLUSION

With Luxoft as your technology partner, you can begin to adopt the Big Data technologies that you need to stay competitive with low-risk system adoption that maximizes your current technology investments.

For more information about Luxoft’s Big Data development services, visit (www.luxoft.com/big-data) or contact us at bigdata@luxoft.com