Big Data – The next big thing
India to capitalize on the growing Big Data market opportunity

1. Global Big Data market opportunity is estimated to grow at 45% annually to reach ~USD 25 billion by 2015, from the current size of about ~USD 8.0 billion.

2. Big Data industry characterized by technological advancements, niche startups, and large number of M&A’s.

3. Limited talent availability and low awareness about the benefits of Big Data analytics may inhibit market growth.

4. Big Data ecosystem continues to evolve over 2012-2013 through sustained investments to strengthen existing technologies & building strong talent pool.

Indian Big Data industry expected to grow from ~USD 200 million in 2012 to ~USD 1.0 billion in 2015 at a CAGR of over 83%.

Big Data industry gaining traction in India, with IT services/Analytics providers offering business centric solutions.

India has early mover advantage vis-à-vis other geographies in creating a strong base of Big Data workforce.

Indian service providers mirroring their global counterparts with key strategies primarily focusing on talent development, portfolio enhancement, and go-to-market partnerships.

Source: CRISIL GR&A analysis
Big Data is defined by Volume, Variety and Velocity

**What is Big Data?**

Big Data relates to rapidly growing, *Structured and Unstructured datasets* with sizes beyond the ability of conventional database tools to store, manage, and analyze them. In addition to its size and complexity, it refers to its ability to help in “Evidence-Based” Decision-making, having a high impact on business operations.

---

Source: CRISIL GR&A analysis
The data being generated globally is undergoing exponential growth

Growth of global data, 2009-2020

<table>
<thead>
<tr>
<th>Year</th>
<th>Zetabytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>0.8</td>
</tr>
<tr>
<td>2011</td>
<td>1.9</td>
</tr>
<tr>
<td>2015</td>
<td>7.9</td>
</tr>
<tr>
<td>2020</td>
<td>35.0</td>
</tr>
</tbody>
</table>

Key drivers for deployment of larger datasets

- Creation of data from multiple sources and touch-points
- Distributed storage techniques and cloud computing enabling organizations to store large amount of data at lower costs
- Emergence of innovative open source software and architecture such as Hadoop Distributed File System and MapReduce
- Roll-out of 100G Ethernet cables for fast information retrieval

Implication on an organization

- Need for large storage capacity
- Need for quick retrieval of data
- Enable informed decision-making effectively, leveraging large data sets. For e.g.:
  - Turn 12 TB of Tweets created each day into improved product sentiment analysis
  - Convert 350 billion annual meter readings to better predict power consumption

Total storage costs, 2005-2015

<table>
<thead>
<tr>
<th>Year</th>
<th>USD / gigabyte</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>18.9</td>
</tr>
<tr>
<td>2011</td>
<td>1.6</td>
</tr>
<tr>
<td>2015</td>
<td>0.7</td>
</tr>
</tbody>
</table>

CAGR (2009-2020) 41.0%
Today 80% of data existing in any enterprise is unstructured data

### Introduction
- Variety of sources from where data is being generated has also undergone a shift
- The types of data being created has changed from structured to semi-structured to unstructured data

### Implication for organization
- Need to manage broad range of data types
- Process analytic queries across numerous data types

### Solutions required
- Need to extract meaningful analysis from this data has led to several technologies to gain traction
- Examples include NoSQL databases to store unstructured data as well as innovative processing methods like Hadoop and massive parallel processing (MPP)

### Structured Data
- Resides in formal data stores – RDBMS and Data Warehouse; grouped in the form of rows or columns
- Accounts for ~10% of the total data existing currently

### Semi-Structured Data
- A form of structured data that does not conform with the formal structure of data models
- Accounts for ~10% of the total data existing currently

### Unstructured Data
- Comprises data formats which cannot be stored in row/column format like audio files, video, clickstream data,
- Accounts for ~80% of the total data existing currently

Source: Industry reporting; CRISIL GR&A analysis
Big Data Analytics helps derive insights from about 1.0 million updates/minute from Facebook & Twitter alone

Big Data velocity enabling real time use of data

- Big Data is also characterized by velocity or speed i.e. frequency of data generation or the frequency of data delivery
- New age communication channels such as mobile phones, emails, social networking has increased the rate of information flows

Examples:
- Location based marketing based on user location sensed by mobile towers
- Satellite images can help monitor and analyze troop movements, a flood plane, cloud patterns, or forest fires
- Video analysis systems could monitor a sensitive or valuable facility, watching for possible intruders and alert authorities in real time

Source: Industry reporting; CRISIL GR&A analysis
Big Data analytics is application of advanced techniques on Big Datasets; answers questions previously considered beyond reach.

**Evolution of analytics**

- **Basic analytics**
  - What happened?
  - When did it happen?
  - What was its impact?

- **Descriptive analytics**
  - What happened?
  - When did it happen?
  - What was its impact?

- **Predictive analytics**
  - Why did it happen?
  - When will it happen again?
  - What caused it to happen?
  - What can be done to avoid it?

- **Prescriptive analytics**
  - Why did it happen?
  - When will it happen again?
  - What caused it to happen?
  - What can be done to avoid it?

**Big Data analytics**

- Complex event processing
- Optimization
- Extreme SQL
- Social network analytics
- Stochastic optimization
- Visualization
- Behavioral analytics
- Constraint based BI
- Analytic database functions
- Natural Language Processing
- Online analytical processing (OLAP)
- Constraint based BI

**Level of Complexity**

- **Standard reports**
  - Query drill down

- **Adhoc reports**
  - Data mining

- **Alerts**

**Time**

- **Late 1990s**
- **2000 onwards**

Source: CRISIL GR&A analysis
Technology and Services players are investing to build platforms, service models and Big Data skills

Key Players

<table>
<thead>
<tr>
<th>Hardware/Software Providers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SPSS</strong></td>
</tr>
<tr>
<td><strong>Datameer</strong></td>
</tr>
<tr>
<td><strong>cloudera</strong></td>
</tr>
<tr>
<td><strong>EMC</strong></td>
</tr>
<tr>
<td><strong>Sas</strong></td>
</tr>
<tr>
<td><strong>Oracle</strong></td>
</tr>
<tr>
<td><strong>SAP</strong></td>
</tr>
<tr>
<td><strong>Vertica</strong></td>
</tr>
<tr>
<td><strong>EMC</strong></td>
</tr>
<tr>
<td><strong>Teradata Aster</strong></td>
</tr>
<tr>
<td><strong>IBM</strong></td>
</tr>
<tr>
<td><strong>Zettaset</strong></td>
</tr>
<tr>
<td><strong>Amazon web services</strong></td>
</tr>
<tr>
<td><strong>HP</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IT and IT Enabled Service Providers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genpact</strong></td>
</tr>
<tr>
<td><strong>Accenture</strong></td>
</tr>
<tr>
<td><strong>Fractal</strong></td>
</tr>
<tr>
<td><strong>MetaScale</strong></td>
</tr>
<tr>
<td><strong>WNS</strong></td>
</tr>
<tr>
<td><strong>absolutdata</strong></td>
</tr>
<tr>
<td><strong>Infosys</strong></td>
</tr>
<tr>
<td><strong>CRISIL</strong></td>
</tr>
<tr>
<td><strong>Capgemini</strong></td>
</tr>
<tr>
<td><strong>EXL</strong></td>
</tr>
<tr>
<td><strong>Cognilytics</strong></td>
</tr>
</tbody>
</table>

- Most of the technology and software companies in the Big Data space are focusing on Storage technology, parallel computing and visualization techniques
- Most of the Indian IT players have set up internal COEs, dedicated teams to execute pilot projects and proof of concepts around Big Data
- Pure play Analytics companies are investing in technology/programming skills to be able to help clients in their Big Data projects

Source: Industry reporting; CRISIL GR&A analysis
Financial Services, Retail and Telecom are likely to be the early adopters of the Big Data owing to enormous digital data flow

- Indian service providers like Infosys, Fractal are enabling Big Data analytics in the area of fraud detection, CRM and customer loyalty program, trading pattern analysis, risk calculation on large portfolio of loans
- Key Adopters: JPMorgan Chase, Merrill Lynch, HSBC, American Express, Goldman Sachs, Barclays, Bank of America, Citigroup, and Wells Fargo

- Both brick and mortar as well as online retailers are increasing their adoption of Big Data analytics for real time analysis of purchase behavior and buying patterns, enhanced customer segmentation and customer loyalty
- Key Adopters: Walmart & Sears

- Telecom players are increasingly focusing on Big Data to limit churn rates, build loyalty and provide multi-channel and multi-service applications by proactively analyzing the subscriber and network data
- Key Adopters: Airtel (India)

- Indian service providers are enabling manufacturing companies through Big Data analytics in the areas of accurate demand forecasting, optimization of operations, inventory management, open innovation and better analysis of post sales feedback in real time

- Key benefits of big data in public sector include: Intelligence to counter national threats, Forecast economic events, Traffic management, Environment monitoring, energy/waste management, etc.

- Healthcare players use Big Data Next-generation sequencing and mapping for genomics, analysis of correlation between treatments & outcomes and real time data from medical devices for better patient care

Source: Industry reporting; CRISIL GR&A analysis
Large IT players leveraging M&As to add Big Data capabilities to their service portfolios

<table>
<thead>
<tr>
<th>Area</th>
<th>Acquirer</th>
<th>Target Company</th>
<th>Date</th>
<th>Deal value</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Management</td>
<td>ORACLE</td>
<td>ENDECA</td>
<td>Oct. '11</td>
<td>USD 1.1 billion</td>
<td>• Develop a comprehensive platform to analyze Big Data</td>
</tr>
<tr>
<td></td>
<td>ORACLE</td>
<td>Autonomy</td>
<td>Oct. '11</td>
<td>USD 10.2 billion</td>
<td>• Leverage Autonomy’s pattern-matching technology that recognizes unstructured data, understands it and processes it</td>
</tr>
<tr>
<td></td>
<td>TERADATA</td>
<td>aster data</td>
<td>Mar. '11</td>
<td>USD 263 million</td>
<td>• Strengthen position in data warehousing market through expertise in SQL and MapReduce-based analysis</td>
</tr>
<tr>
<td>Advanced analytics</td>
<td>IBM</td>
<td>tealeaf.</td>
<td>Jun. '12</td>
<td>N.A.</td>
<td>• Extend Smarter Commerce suite with qualitative analytics software</td>
</tr>
<tr>
<td></td>
<td>IBM</td>
<td>Vivísimo</td>
<td>May. '12</td>
<td>N.A.</td>
<td>• Leverage data navigation technologies for Big Data by automating discovery of through innovative index and search capabilities</td>
</tr>
<tr>
<td></td>
<td>IBM</td>
<td>VARICENT</td>
<td>May. '12</td>
<td>N.A.</td>
<td>• Addition of sales performance analytics</td>
</tr>
<tr>
<td></td>
<td>TERADATA</td>
<td>eCircle</td>
<td>May. '12</td>
<td>N.A.</td>
<td>• Enhance Big Data marketing analytics</td>
</tr>
<tr>
<td></td>
<td>OPERA</td>
<td>biq</td>
<td>Apr. '12</td>
<td>N.A.</td>
<td>• Acquisition of spend and procurement analytics</td>
</tr>
<tr>
<td></td>
<td>EMC</td>
<td>Vertica</td>
<td>Mar. '12</td>
<td>N.A.</td>
<td>• Accelerate development of Big Data analytic applications</td>
</tr>
</tbody>
</table>

**Key highlights**

- M&As in the Big Data space have tripled in the first half of 2012
- Acquisition targets are mainly innovative Big Data start-ups
- M&As with bigger deal value are happening in data management

N.A. is not available. Source: Industry reporting; CRISIL GR&A analysis
Global Big Data market to reach ~USD 25 billion by 2015, with a 45% share of IT & IT-enabled services

USD billion

Drivers
- Rising need to focus on leveraging multiple data sources, to maintain competitive advantage
- Continued investments from service providers in Big Data technologies and tools, particularly in Big Data analytics
- Companies to define Big Data implementation strategy for taking more informed decisions
- Big Data deployments to become mainstream in verticals like utilities, transportation, energy, etc.

Restraints
- Shortage of analytical and managerial talent
- Regulations about data privacy and security
- Roadblocks in managing change during transition to data driven organizations
- Restrained adoption in new geographies such as the Middle East

Source: Industry reporting; CRISIL GR&A analysis
India’s opportunity is expected to be ~8.9% of Global Big Data IT and IT enabled services market by 2015E

India Big Data outsourcing opportunity, 2011 – 2015E
USD billion

<table>
<thead>
<tr>
<th>Year</th>
<th>Opportunity (USD billion)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011E</td>
<td>~0.1</td>
</tr>
<tr>
<td>2012E</td>
<td>~0.2</td>
</tr>
<tr>
<td>2015F</td>
<td>1.1-1.2</td>
</tr>
</tbody>
</table>

CAGR 2011 – 2015E: 83%

Source: CRISIL GR&A analysis

India Big Data outsourcing opportunity, by category, 2015F, Percent

- Pure-play Analytics firms: 24%-27%
- Integrated IT/ BPO players: 73%-76%

100% = ~USD 1.1 billion

Key growth drivers include:

1. Aggressive efforts for Big Data specific talent development
2. Increasing domain expertise and breadth of services offered by Indian service providers
3. Effective strategies of Indian service providers through synergies of international partnerships
4. Rising domestic demand for Big Data implementations across industry verticals

Source: Industry reporting; CRISIL GR&A analysis
Emergence of niche startups and technological developments fostering growth in the Big Data industry

<table>
<thead>
<tr>
<th>Market Trends and Developments</th>
<th>Impact on Industry</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Converging technology trends in data storage, processing, and analytics are driving adoption</td>
<td></td>
</tr>
<tr>
<td>2 Emergence of niche Big Data startups driving technological innovation</td>
<td></td>
</tr>
<tr>
<td>3 Large IT players leveraging M&amp;A’s to add Big Data capabilities to their service portfolios</td>
<td></td>
</tr>
<tr>
<td>4 Talent shortage is one of the biggest challenges of the Big Data space</td>
<td></td>
</tr>
<tr>
<td>5 Lack of awareness about benefits of Big Data may limit enterprise adoption</td>
<td>Negative impact</td>
</tr>
<tr>
<td>6 Regulations driving the adoption in various industry verticals</td>
<td>Positive impact</td>
</tr>
</tbody>
</table>

Source: CRISIL GR&A analysis
Potential shortfall of 1.5 million Data-Savvy Managers and ~150,000 Data Scientists in the US in 2018

Demand-supply gap for data scientists* in US, 2018

- **Data Scientists**
  - 2018E Supply: 300K
  - 2018E Demand: 440K-490K
  - 50%-60% gap relative to supply

Demand-supply gap for data-savvy managers* in US, 2018

- **Data-savvy Managers**
  - 2018E Supply: 1.5 million
  - 2018E Demand: 4.0 million
  - 60% gap relative to supply

Demand-supply gap for Technical Engineers in US, 2018

- **Technical Engineers**
  - 2018E Supply: 2.5 million
  - 2018E Demand: 4.0 million
  - 60% gap relative to supply

<table>
<thead>
<tr>
<th>Role in Ecosystem</th>
<th>Requisite educational qualifications</th>
<th>Other expertise</th>
</tr>
</thead>
</table>
| **Data Scientists** | - Big Data analytics  
- Business intelligence  
- Visualization | - Advanced degree like M.S. or Ph.D., in mathematics, statistics, economics, computer science or any decision sciences  
- Expertise in data analytics skills to extract data, use of modeling & simulations | - Multi-disciplinary knowledge of business to find insights |
| **Data-savvy Managers** | - Project management across the Big Data ecosystem  
- Consulting services  
- Implementation  
- Infrastructure management  
- Analytics | - Advanced business degree such as MBA, M.S. or managerial diplomas  
- Knowledge of statistics and/or machine learning to frame key questions and analyze answers | - Conceptual knowledge of business to interpret and challenge the insights  
- Ability to make decisions using Big Data insights |
| **Technical Engineers** | - Technical support in hardware & software across the Big Data ecosystem for:  
- Data architecture  
- Data administration  
- Developer environment  
- Applications | - Having a degree in computer science, information technology, systems engineering, or related disciplines  
- Possessing data management knowledge | - IT skills to develop, implement, and maintain hardware and software |

*Analysts with deep analytical training; **Managers to analyze Big Data and make decisions based on their findings; Source: McKinsey Global Institute; CRISIL GR&A analysis
India has an early mover advantage vis-à-vis other geographies in creating a strong base of Big Data workforce

India's Big Data talent requirement*, 2011-2015

- 2011E: ~1,000 - 1,500
- 2013F: ~4,000 - 6,000
- 2015F: ~15,000 - 20,000

Expected CAGR: ~80%

Annual potential talent pool available for Big Data in India

| IT Professionals | 2,800,000 |
| Graduates in Math/science | 700,000 |
| Engineers | 500,000 |
| Graduates in Economics | 350,000 |
| Post graduates in Math/science | 300,000 |
| MBAs | 250,000 |
| PhDs | 14,000 |
| Graduates in statistics | 5,000 |

While total talent availability is high, ready to hire talent for Big Data would be just ~3%-5% of the total

State of Big Data talent across key geographies

<table>
<thead>
<tr>
<th>Source: CRISIL GR&amp;A analysis</th>
</tr>
</thead>
</table>

| United States | High | High | High | I |
| India | High | Medium | High | II |
| Poland | Medium | High | Medium | III |
| China | High | Low | Low | IV |
| Philippines | Low | Low | Low | V |

*Includes data scientists and technical engineers
Source: Industry reporting; CRISIL GR&A analysis

- India is expected to be in the forefront of Big Data analytics and related IT services, fueling demand for data scientists and IT engineers estimated at about 15,000 – 20,000 by 2015
- India follows closely behind the US in terms of Big Data talent availability and service provider’s initiatives to build such talent
- India is ahead of most outsourcing destinations like China, Poland and Philippines, in terms of talent availability.
  - IT companies like EMC, Oracle, IBM, Infosys, etc., are leveraging their academic alliance programs, with universities in India and overseas to introduce courses in various areas of Big Data