Leveraging Big Data and Analytics Techniques in Managing Large Projects

**Theme:** Emerging Trends in Project Management

**Author:** Mr. Venkata Jogarao Vallabhajosyula
1. **KEYWORDS**

   Big Data, Analytics, Project Management, Emerging Trends, HADOOP, HDFS

2. **ABSTRACT**

   **Project Management Practitioners** find it a great advantage if there are tools and techniques, which enable them to sift through huge volume of project data to provide accurate and non-intuitive dependent analytical output, preferably in dashboard format. The ability to slice and dice these dashboards is another definite advantage.

   **The conventional methods in managing project** issues such as cost deviation, schedule slippages, resource loading, what-if analysis of various permutations & combinations of project influencing factors are not an easy task, especially in large, globally diversified projects.

   **Recent advances in data processing techniques** such as Big Data and Predictive Analytics offer much faster and accurate information processing capabilities along with multi-layer intelligent dashboards. These latest techniques enable project managers to take informed decisions.

   **Big Data** is the ability to process large sets of different varieties of data in fuse at much faster pace. Traditional techniques cannot analyze data without compromising accuracy. Big Data is supported by HADOOP framework to processing multiple parameters of large data. Programming languages such as JAVA can be used to develop HADOOP frameworks.

   **Analytics** - Statistical programming language "R" is very popular in developing predictive analytics engines. These engines, when they are fed project data, provide intelligent drill down dashboards.

   **Dashboards** SAP BI/Oracle and OBIEE are popular front-end tools to develop dashboards, and SAP HANA / Oracle DB are back end data base engines.

   I would like to share how Big Data and Analytics techniques can be leveraged in managing large projects.
### 3. Table of Contents

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction to Big Data &amp; Analytics</td>
<td>4</td>
</tr>
<tr>
<td>Technology Landscape</td>
<td>7</td>
</tr>
<tr>
<td>Project Management Focus Areas</td>
<td>10</td>
</tr>
<tr>
<td>Methodology &amp; Approach</td>
<td>11</td>
</tr>
<tr>
<td>Critical Success Factors</td>
<td>13</td>
</tr>
<tr>
<td>Quantified Benefits to Business</td>
<td>13</td>
</tr>
<tr>
<td>Conclusion</td>
<td>13</td>
</tr>
<tr>
<td>References</td>
<td>14</td>
</tr>
</tbody>
</table>
4. **INTRODUCTION TO BIG DATA AND ANALYTICS**

Big Data can be defined as a large and complex data set, which includes collection of transactions, interactions and observations from various data sources and in different data formats as mentioned in Fig 1.

- **Features**
  - Data at Rest – Existing Data
  - Data in Motion – Batch to Streaming
  - Data in many Forms – Structured to Unstructured
  - Data in Doubt – Incomplete Data

- **Components**
  - Various Data Sources
  - Data Storage
  - Data Processing
  - Fast Loading Analytical Database
  - Analytical Applications – SAS, R, Mahout
  - Value Creation

- **Data Sources**
  - Enterprise Data (Emails, Word Documents)
  - PM Plan (MPP, Excel)
  - Social Media
  - Sensor Data
  - Public Data

- **Requirements**
  - New Tools
  - New Platforms
  - New Analytical Methods
  - Data Scientist
  - Analytical Sandbox

---

**Fig # 1**
### Big Data vs. Traditional Data

<table>
<thead>
<tr>
<th>Data Class</th>
<th>Size of Data</th>
<th>Platform</th>
<th>Space Consumption</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small Data</td>
<td>&lt;10 GB</td>
<td>Excel</td>
<td>Fits in one machine’s memory</td>
<td>Thousands of sales figures</td>
</tr>
<tr>
<td>Medium Data</td>
<td>10 GB to 1 TB</td>
<td>Indexed files, Monolithic DB</td>
<td>Fits in one machines’ disk</td>
<td>Millions of web pages</td>
</tr>
<tr>
<td>Big Data</td>
<td>&gt; 1 TB</td>
<td>Hadoop, Distributed DBs</td>
<td>Stored across many machines</td>
<td>Billions of web clicks</td>
</tr>
</tbody>
</table>

Fig # 2
Big Data Analytics

![Diagram of Big Data Repository and Analytics Process]

**Fig # 3**
Introduction to Analytics

The Big Data Analytics life cycle passes through 6 major phases, which are dependent on and associated with each other and revolve until each phase meets specific criteria.

Fig # 4
5. **Technology Landscape**

Technology landscape required

In today's world, umpteen numbers of tools are available to process large volumes of unstructured data. HADOOP technology stack is very popular among the tool set. A glimpse of software stack supported by HADOOP platform is given below.
Though a project manager is not expected to possess any deep technical know-how, it is highly desirable to be familiar with the merits and de-merits of various technology stacks available in the market. This would help the project manager arrive at an informed decision on tool selection and evaluation.
6. **PROJECT MANAGEMENT FOCUS AREAS**

Big data and Analytics techniques are especially useful in managing large projects, which are multi-site and multi-discipline, involving sizable resource strength such as construction of metro rail projects, O&M of power plants, steel plants etc, where the duration is large and any delay in schedule would result in huge losses. Managing these projects in conventional methods always carried the challenges of information coordination, data assimilation, manual intervention and interpretation, among other things. Big Data and Analytics can analyze the sources of information and provide a near real-time view of project status.

**Capital Projects**  
SPI – Schedule Adherence  
CPI – Cost Adherence  

**Project Management**  
1. Project progress status  
2. Risk management  
3. Budget status  
4. Safety compliance  
5. Scope control  

Mock screens are given below

![Mock screens](image-url)

Fig # 7
7. **Methodology & Approach**

The Big Data Analytics architecture includes 6 major steps, which consist of data gathering, real-time processing, analytics & model development, and value creation.

**Data Needs and Sources**: Typical project data required for monitoring and tracking project health such as baseline plan – work breakdown structure, current situation of the plan, dependencies, future impact etc.

**Data Segregation & Assimilation**: The data source would consist of structured and unstructured format. This data has to be segregated according to the project insights to be developed.

**Missing Data Analysis**: Data which are missing for whatever reason will have to be substituted by derived data for data stream completeness.

**Critical variable analysis**: Identify all the critical variables to be measured as part of the analysis.
**Event based analysis:** All project events, which are milestone linked, should be grouped. This would help detect any localized issues.

**Business Intelligence Reporting:** This is a set of dashboards, which gives summary level information. These dashboards should be drilldownable in nature.
8. **CRITICAL SUCCESS FACTORS**

1. Data sources – ability to feed quality data
2. Ability to tap data streams and ensuring that there is no data loss
3. Project Insights – converted to programmable requirements
4. Leadership team with dedicated & committed vision towards excellence
5. Strong integration & collaboration among discrete teams
6. Out-of-the-box thinking and approach
7. Smart people

9. **QUANTIFIED BENEFITS TO BUSINESS**

According to Gartner reports and other reports of similar nature, Big Data and Analytics techniques are the future of data analysis. Minimizing data manipulation and avoiding wrong interpretation are most critical. Big Data and Analytics techniques offer a systematic platform for executing project management better.

1. For capital projects worth Rs. 5000 crore, if 1% benefit can be offered, it is equivalent to Rs. 50 crore, whereas investment will be less than Rs. 1 crore.
2. There is huge potential for re-use as tools developed can be used for similar projects again and again.

10. **CONCLUSION**

This paper talks about leveraging Big Data and Analytics techniques in managing large projects. At the core, the tasks are data sourcing, data assimilation, intelligent interpretation and suggested decisions, and managing them all in a programmed environment and dashboard world.
11. References

4. https://analytics.google.com/