Abstract

Today, Big Data is drawing a lot of attention and popularity and is aspiring enterprises to utilise more efficiently their data to help them understand how to better function, grow and manage as a business. Enterprises are aware that to derive real business value from Big Data and seek competitive advantages, they need to have available the right tool to extract, capture and organize a wide variety of useful data and insights from different sources, and to be able to easily use and analyze it within the context of their data.

In this paper we present BigGraph, a solution for enterprises to manage Big Data and facilitate data integration, making the enterprise data more accessible and easier to use externally. BigGraph provides the technology and expertise to build a data-centric enterprise architecture by enabling intuitive navigation, knowledge discovery, extensibility and social analysis. It offers “social” capabilities that enable unified communications and collaborations to help enterprises to reduce the time and cost needed to analyze critical information and efficiently improve enterprise productivity. It is believed that BigGraph is a suitable tool capable of facilitating enterprises to create and link their data as a graph allowing information flow around their world and outside.

1. Introduction

Today, Big Data [14] has drawn a lot of attention and popularity. Enterprises are aware that to derive real business value from Big Data and seek competitive advantages, they need to have available the right tool to extract, capture and organize a wide variety of useful data and insights from different sources, and to be able to easily use and analyze that data[1]. The emerging research questions around Big Data mainly focus on how to elicit meaningful and actionable insights and value from an immense volume, variety and velocity of data and how this can be achieved in a flexible and dynamic way. This paper presents our solution to address these research questions - BigGraph, a project initiated at the Fujitsu Laboratories of Europe Limited1, that provides software and expertise to their customers to help them link together the data in their own Enterprise and provide support and services around that data. BigGraph is a data management platform for graph, backed by large-scale storage that enables to import, mine, aggregate, and analyze data stored as RDF triples in the graph to unlock hidden insights within an enterprise. BigGraph utilises the RESTful APIs as the mechanism to interact with the data, and follows the Linked Data principles to link data over distributed data sources. Linked Data [2] has growing enthusiastic support from industry and academia: it is a method of publishing and linking structured data on the web using Resource Description Framework, RDF [4], which is a data modelling language providing semantic mark-up of data. The use of linked data provides a more effective way to access, organise, and integrate enterprise data, aiming to improve the way to extract value from information. Linked data provides a highly scalable solution by distributing information into multiple, small, compute nodes [12].

Some of the features and functionalities that BigGraph offers are:

- Handling large number of triples with high throughput using a distributed Key-Value-Store, KVS [8].
- Allowing easier integration of systems or use data across silos and technologies with Linked Data [12].
- Enabling flexibility in data management and use by providing a unified view on previously isolated data silos [20]. Enabling flexibility to modify and extensibility to future changes through RDF, which provides a common and flexible way of describing information.
- The flexibility of RDF data model enable to gradually build the structure capable of describing the information resources in an

1http://www.fujitsu.com/emea/about/fle/
enterprise, and at any time, by creating new kinds of relations and concepts thus enabling dynamic extensibility of schema and nodes.

- Provides user-friendly interface for Big Data representation.
- Adaptable to any situations that need a platform to process, integrate, interlink and analyze Big Data in an effective and efficient way.

The core idea of BigGraph is to provide the technology and expertise to build a data-centric enterprise architecture, which comprehensively describes all the key elements and relationships that constitute an organization [19]. In the context of BigGraph, the main focus is on the data itself - and sharing of this data across the enterprise, thus unlocking hidden information that might be of interest to the enterprise.

BigGraph is based on the assumption that machine readable information structured as a graph is a natural way of visualizing, understanding and connecting existing and evolving data. By using graph for storage, processing and handling data, BigGraph enables intuitive navigation, knowledge discovery, extensibility and federation of data sources. Knowledge discovery provides dynamic and on-the-fly discovery of available and hidden data through inference rules, extensibility is achieved by modelling schema as data while federation is provided by linking named data items across systems.

The remainder of this paper is organised as follows: Section 2 provides the background and motivational factors that BigGraph is grounded. Section 3 details the architecture of BigGraph. This is followed by two use cases demonstrating the ‘social value’ of BigGraph, in Section 4. We then conclude the paper in Section 5.

2. Background and Motivation

Today’s relational, hierarchical, and object database technologies are not able to support the new and emerging data management challenges having requirements such as extensibility, greater flexibility, greater dynamism and adaptability that enterprises need today; this calls for flexibility in data schemata and data sources. Frequent and dynamic changes to both data models and their underlying database schemata are being driven by the need to analyze vast amounts of both structured and unstructured data. BigGraph is specifically designed to address these new challenges and offers necessary capabilities such as emphasize model flexibility to support the next-generation of data management applications. It provides a platform for developing applications that are dynamic in nature and incorporates either structured or unstructured data. The ‘social’ aspect to enterprise data has the potential to change the way enterprises do their business. Enterprises that are embracing social tools are finding their employees happier, more productive and coming up with greater ideas, and transforming their business from the ground up (see Table 1) [5]. Few of the key elements of the social aspect of an enterprise are to empower communications and collaborations across the enterprise (Figure 1).

![Figure 1. Social Interlinking](image)

According to [6], enterprises that adopt social networking tools respond 36% faster to customer feedback. This has been further investigated in [5] that shows an analysis of the various case studies in the real world that benefit from enabling the social aspect of their enterprise, as illustrated in Table 1.

However, in order for an enterprise to gain more competitive advantage from more effective social tools, they need to acquire keen awareness of what the Big Data currently situation involves and how to successfully face it. One of biggest challenge of social technology is how to manage or curate the volume, velocity and variety of data that enterprises create on a daily basis. According to [15], the volume of data is increasing in the just last four years by a factor of six to an estimated 988 exabytes and it is expected to reach 8 zettabytes by 2015 [16][17]. Today, enterprises struggle with how to establish and manage such data in a way that enables effective search, filtering, and future reference. Moreover, enterprises have to face with the vast amount of unstructured nature of their data. Although, recently, numerous communication tools have become prevalent within the enterprise, most of this data is unstructured, such as text for E-mail, Web site (Apache, Wiki, MS SharePoint, etc.), Enterprise networks (Yammer, etc.), shared schedule tools (MS Exchange Server) and booking meeting rooms.
Another issue of social networking is that most social tools are vendor-specific and tend to build their own social network in silos rather than building upon the rich data available about existing social relationships. Their applications are often restricted to execute within the confines of specific social network platform. One of the keys in developing an effective staff in a successful enterprise is to understand and improve the way different professionals extract value from information [7][13]. By its very nature, in most organisations, enterprise knowledge resides in multiple islands and application data silos [12], often in different languages and formats. However, many applications need to access data across multiple domains and systems for different purposes, e.g., support decision support, cross-domain process management, target cost planning, etc.

Linked Data can help to integrate and unify this information in the context of ‘closed’ environments such as private enterprises. Linked Data can enable groups of experts to create an enterprise in which large-scale information is fully shared across the organisation and where the act of creating information is intimately coupled with the act of sharing information [3]. Sharing data is now as important as producing it. Professionals will continue to produce and consume information that is specific to their own business needs, but it is now generated in a way that can be connected and shared to other aspects of an enterprise [11]. With Linked Data, knowledge from experts can be immediately transferred and shared throughout the enterprise and reused. Therefore, this paper brings together the concepts of 'Enterprise' with 'social aspects'. Enterprise data does not require being only 'open' but can be 'closed' and live inside the firewall of an enterprise.

Graph representation is the ideal and natural way to represent social networking aspects of a given enterprise, encompassing relationships among its employees, vendors, partners, customers, and the public. Recently, there has been a phenomenal interest in Big Data in the use of on-line social networking services, such as Facebook, Twitter and LinkedIn, using social graphs [18]. Social graphs run analyzes over terabytes of graph data while maintaining the relationships between the data, even as the data and the relationships constantly evolve. The growing interest in such social networking will create new drivers for the development of BigGraph and the use and support of graph technology in enterprises.

Table 1 Analysis of real case studies for ‘Enterprise Social’ Networks (Source: [5])

<table>
<thead>
<tr>
<th>Enterprise</th>
<th>Problem</th>
<th>Solution</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apache</td>
<td>Globally distributed teams</td>
<td>Cross-regional collaboration: internal social network integrated with SharePoint</td>
<td>Cut trouble costs</td>
</tr>
<tr>
<td>Macy’s</td>
<td>Collaboration hindered by ineffective technologies</td>
<td>Focused collaboration: subject-based social networking platform</td>
<td>Get ahead of competitors</td>
</tr>
<tr>
<td>SES</td>
<td>The need for mobile access to real-time data</td>
<td>Real-time collaboration: internal social network with native mobile apps</td>
<td>Help save lives</td>
</tr>
<tr>
<td>NCH</td>
<td>Misalignment between geographically dispersed sales teams</td>
<td>Close new (innovative) deals: community based enterprise social network</td>
<td>Close new (innovative) deals</td>
</tr>
</tbody>
</table>

Figure 2 provides a flavor of what could be considered as ‘social’ in enterprise networks. Enterprises could benefit from enabling interconnectivity among not only human-type resources but also other types of enterprise resources such as equipments, networks, etc.
BigGraph has the potential to realize this vision and in order to do so it is imperative to have a flexible architecture that enables storage, querying, processing, and analyzing Big Data in a scalable manner.

3. BigGraph Architecture

BigGraph is based on RDF (Resource Description Framework) [21], and provides the opportunity to connect together different and isolated data silos and transform into a connected data graph. Due to the fact that RDF could be used as a schema-less data representation format, it enables BigGraph to satisfy the important requirements of flexibility and dynamic extensibility such as adding new data easily and coping with constantly changing data. A schema-less data representation format enables any data to be added and linked together without the need to constantly make changes to any schema [22].

From the architectural point of view, BigGraph is a graph-based semantic platform over a distributed, ordered Key-Value-Store, KVS [8], where data is distributed across multiple machines. The distributed architecture enables to achieve a high level of scalability in terms of the data stored and an efficient approach of querying the data. The architecture of a single server is illustrated in Figure 3. Each server is composed by a Semantic Layer, an RDF Layer, a Triple Adaptor Layer, a KVS Layer and a Communication Layer.

The Semantic Layer interacts with end-users and relays user queries to the RDF processor. The semantic layer explicitly captures meaningful relationships among data, which are represented at conceptual and model level (as groups of related RDF triples). In this layer, semantic operations such as data reconciliation and social analysis are performed.

Figure 3. BigGraph Server Architecture

At the model level, BigGraph uses lightweight ontologies (in RDFS and/or OWL) [9] to define domain concepts and relationships to model social enterprise domain. Figure 4 shows an example of a shared enterprise ontology that BigGraph employs to model the essence of an enterprise in a coherent, comprehensive, consistent, and concise way. It corresponds to an enterprise model presented as a simple set of concepts and relationship taxonomies. In Figure 4 several interrelated concepts that are crucial for enterprise ontology are distinguished. Some of these concepts are Activity, Person, Organisation, Skill and Project. These concepts are linked by relationships such as worksOn between Person and Project; and hasActivity linking Person with Activity, and many others. The ontology was designed to offer a common understanding about the essence of an organization with a specific focus on collaboration and social interaction among people and things. The enterprise data currently used in BigGraph are "owned" by Fujitsu Laboratories of Europe Limited, and extracted from different sources within the company. By linking enterprise data to a shared ontology, BigGraph allows not just to "understand" the content, but also to derive new knowledge by "reasoning" about that content. Sophisticated ontological analysis, such as rule-based reasoning is used to enable such social analysis. The current BigGraph uses Jena2 as rule engine. It supports rule-based inference over RDF graphs and provides forward chaining, backward chaining and a hybrid execution model. Rules in BigGraph provide more expressivity on top of ontologies and the ability to discover dynamically and on-the-fly new knowledge about the social aspects of an enterprise. For example, rules in BigGraph have been used to

---

1 jena.sourceforge.net/inference/
analyze activity feed in terms of the email exchanges & meetings and therefore calculate, on a timeline, the intensity of communications among people within an enterprise. Moreover, the Semantic Layer is not bound to specific implementations and allows other rule inference engines to be easily plugged in.

The architecture of BigGraph can be applied to a wide number of application scenarios in various domains including health, agriculture, or enterprise. In this paper, we introduce the ‘social layer’ of BigGraph by presenting two use cases in the enterprise domain in the following section.

4. Enterprise Social: Use Cases

BigGraph offers the whole infrastructure in-house to manage, organise, store, link and analyzes enterprise data and so to enable better enterprise decision making. Moving from static, broadcast websites to interactive, social web applications have changed the way people live through personalization, collaborative networks, reduced reliance on hierarchies and increased organization ability. Today, many enterprises are adopting social networking principles for external, customer relationship or brand management purposes, and so on [7].

BigGraph also provides a web-graph visualisation tools for RDF data that aims for a fully interactive, scalable and responsive user interface. Graphs are presented to the user, where the user is able to expand the graph by adding new nodes that are connected to existing nodes or explore the graph dynamically by moving from node to node (Figure 5). The user is also able to customise the style of each individual node set and the edges that connect them, includes changing the colour and size of node and the strength of edges. The interface uses and extends the D3 Visualisation library (v 2.6.1)³.

The RDF Layer is the interface with client applications. It uses the Jena Graph API to add/remove triples, and parse SPARQL queries [10]. This layer includes also a SPARQL engine, responsible for converting the text-based query into a set of operations. These operations include low level search operations, filtering, and joining among others. Some of these features are implemented within the SPARQL engine. Low level data searching is implemented by the underlying Triple Adaptor Layer.

The Triple Adaptor Layer is responsible for converting triples into a key that can be stored in the KVS Layer. The Triple Adaptor Layer provides operation to add, remove and search triples according to a specific criteria. This layer uses the underlying KVS interface to store, remove and query data.

The KVS Layer implements a distributed ordered key-value store. It offers a regular ordered hash map interface to put a key-value pair, to get or remove a key-value and to execute range queries over the ordered KVS. This layer is responsible for distribution and retrieval of the data from other servers. This layer uses the underlying Communication Layer to communicate with other servers.

Another layer, currently on development, is the Data Aggregation Layer. The Data Aggregation Layer enables to access, extract and aggregate together different data from different source (streams, RDBs, XLS, HTML, etc.) and to convert those data into RDF triples, that can then stored. The current implementation of BigGraph is currently capable to extract and convert data from RDBMS databases.

³ d3js.org/
This paper identifies two use cases in the enterprise domain that demonstrate the ‘social aspect’ and leverage the business value of BigGraph and its concrete benefits to enterprises. BigGraph offers ‘social’ capabilities that enable unified communications and collaborations to help enterprises reduce the time and cost needed to analyze critical information and efficiently improve enterprise productivity. These use cases were implemented in the BigGraph development, which constitutes of the following ‘social’ features:

- Sharing: collaboration & communication.
- Linked enterprise.
- Dynamic interaction with data (activeLinkeData).
- Timeliness for customised view (e.g. comment; like; share; recommend; preferences, etc.).
- Flexibility/dynamic extensibility.
- Data-centric enterprise architecture.
- Ease of use.

The data for both use cases is imported from the enterprise database system, where the data is extracted and transformed into RDF triples then loaded into BigGraph platform.

4.1. UC1-Dynamic Extensibility

BigGraph can handle situations where prior knowledge is needed or not known upfront. One of the benefits of BigGraph is that knowledge can be dynamically built in contrast with traditional relational databases. This use case (UC1) leverages the dynamic extensibility of the BigGraph at schema level and instance level. BigGraph includes the functionality of dynamically adding new instances to the system and also extending the schema by adding new classes at runtime. This dynamic extensibility leverages BigGraph technology over the existing technologies in terms enabling identification and isolation of the key information that make or breaks a decision.

The aim of UC1 is to demonstrate social recommendations in enterprises upon which decisions can be made. This particular use case finds people in terms of potential collaborations according to their profiles that match the prerequisite skill-set for specific tasks. New nodes can be added at anytime at the instance or class levels demonstrating the dynamic extensibility of BigGraph. For UC1, when a new person- or project- type node is added to the BigGraph (as shown in Figure 6), a social rule is triggered to find suitable collaboration. The Timeline...
gets updated with every change made to the graph resources.

For UC1, the rules are dynamically triggered in order to discover knowledge whenever new information is added to the BigGraph. An example of rule finding candidate people to work for a project with specific skills is as follows:

```
[peopleWithMatchingSkillsProject (?person enterprise_social#hasMatchSkillsFor ?project) <-
 (?person enterprise_social#has_skills ?skillSet1)
 (?skillSet1 rdf-schema#label ?labelSkill1)
 (?project enterprise_social#requires_skills ?skillSet2)
 (?skillSet2 rdf-schema#label ?labelSkill2)
.equal(?labelSkill1,?labelSkill2)]
```

As shown in Figure 7, recommendations about potential collaboration for this new person (e.g., John Smith) to an existing project (e.g., BigGraph project) are made. This is also captured via Timeline in the sliding panel as shown in Figure 8.

Out of a pool of vast number of employees, the most appropriate people are suggested which informs the process of decision making for the higher level management to accept or deny that collaboration.

This supports the establishment of an informed enterprise architecture in which users engage socially where they have identity in the system and can 'follow' particular data items, comment on a data item, can assert approval of a link, etc. This also highlights the 'social' aspect of BigGraph.

### 4.2. UC2-Knowledge Discover via Rules and Active Data

BigGraph gives the ability to define many to many relationships and so gives the ability to extensively cross-referenced data, i.e. link data from one data-source with other data-sources. Having the ability to relate things to each other offers huge opportunities for practical purposes, such as structured browsing and data integration. This use case (UC2) leverages the social analysis capability of the BigGraph through knowledge discovery using rules and active data. One of the applicability of the rules could be, in social enterprise scenarios, to discover useful information about the social aspect of an enterprise. For example, rules in BigGraph have been used to analyze activity feed in terms of the email exchanges & meetings and therefore calculate, on a timeline, the intensity of communications among people within an enterprise. This knowledge discovery via rules and active data leverages BigGraph technology over the existing technologies in terms of turning straw into gold by sifting through and analysing the data in order to extract the essential information that provides a competitive advantage in the domain.

The aim of UC2 is to demonstrate social interactions in enterprises which could be analyzed to extract hidden information from corpuses of Big Data. This includes the functionality of analysing results by the system and discovering new knowledge whenever a request is made at runtime.

This particular use case finds the strength of active communications among employees by investigating available information (such as emails, meetings, etc.) in an enterprise. It is run on-demand when such information is requested. This action triggers another social rule, which will look at the level of communication one employee have with other team members to determine the communication...
activeness. Three levels of communications - low, medium, high - are considered based on frequency of occurrences of emails and meetings. These levels are represented in different colours and through different thickness of lines (the finest being the lowest), as illustrated in Figure 6.

For UC2, the rules are triggered on-demand in order to reveal hidden knowledge in the BigGraph. This also highlights the ‘social’ aspect of BigGraph where, along with collaborations, various social activities such as communications among various resources can be revealed.

Example of a rule to identify low communication between two peoples is shown below:

```
[LowCommunication:
 (?person1 enterprise_social#has_low_communication ?person2)
 <- (?person1 enterprise_social#hasLowMeeting ?person2)
 (?person1 enterprise_social#hasLowEmail ?person2)]

[LowEmail: (?person1 enterprise_social#hasLowEmail ?person2) <-
 (?person1 rdf#type enterprise_social#Person)
 (?person2 rdf#type enterprise_social#Person)
 lessThan(?numberOfEmails, 2) ]

[LowMeeting: (?person1 enterprise_social#hasLowMeeting ?person2) <-
 (?person1 rdfs#type enterprise_social#Person)
 (?person2 rdfs#type enterprise_social#Person)
 lessThan(?numberOfMeetings, 2) ]
```

Figure 9: Low Communication between Two People

As shown in Figure 9 and Figure 10, based on the number of meetings and email exchanges, the frequency of communication is highlighted between two people.

The benefits of BigGraph can be summarized as follows:

- Greater economic value of data: enabling to identify what data is valuable and then transforming and extracting that data for analysis – turning straw into gold.
- Improved data integration across highly distributed & disconnected different sources using linked data.
- Enable extensibility by modelling schema as data.
Enable serendipitous discovery of new information and data – finding needles in the haystack.

- User-friendly, scalable data representation - simple, flexible and intuitive way to represent knowledge.

There are a variety of graph visual analytics tools for the discovery of new relationships in enterprises [23]. However, BigGraph is not only a graph visualization tool. BigGraph differentiates from others by offering the whole platform, technology and expertise to handle, link, manage, integrate and analyse data within and between enterprises, in a scalable, standardised and flexible way. BigGraph thus can be seen as an end-to-end system that can import, mine, aggregate, and assemble data as a social graph to improve communication and cooperation within an enterprise. These activities are usually poorly supported by traditional social tools that often work in isolation. Also by providing a graph visual interface, users can have the opportunity to quickly view all data together and their relationships to assist them in their daily routine.

5. Conclusion

In this paper, we report on a BigGraph development carried out in the Fujitsu Laboratories of Europe Limited. BigGraph offers customers capabilities that go beyond other alternatives or similar products by employing semantic techniques developed and standardized by the W3C4, to help enterprises to enhance integration capabilities to support connections and interactions among employees and communities, between employees and information assets, and to facilitate enterprise activities in all of their possible combinations. There are several graphical tools that claim to analyze Big Data. BigGraph provides a step forward by enabling opening up to other social network platforms allowing cross-platform data information exchange and usage with the Linked Data approach, thus highlighting relationships among employees, places, assets, and other things.

There are three important aspects that contribute to the success of BigGraph: the use of Linked Data, the graph technologies, and the social analysis capability. By using Linked Data, BigGraph provides

---

4 www.w3.org/
a unique infrastructure that allows storing and managing data, facilitates data integration, and makes the enterprise data more accessible and easier to use externally. BigGraph has the ability to create dynamically the structure of enterprise data, without having prior knowledge, and accommodates rapid changes. BigGraph offers more than a graph database by providing the whole infrastructure for managing, interlinking information, analyzing, and visualizing very large complex datasets. BigGraph provides social analysis capability through the discovery of useful information about the social aspect of an enterprise using rules and active data, as presented in Section 4. Moreover, BigGraph fits ideally in the current situation, where enterprises are looking for new solutions and technologies for managing their peta-bytes of data that have sprung from social computing and data analysis applications. With its agile method, BigGraph has a great potential to accommodate rapid changes in its social environments and to address users’ needs.

BigGraph is an on-going project and future works include the development of data reconciliation components to reconcile the heterogeneity among various RDF resources by identifying and defining equivalent links, and evaluating in detail the scalability and performance of platform.

10. References


