The State of RE Research in the Context of Big Data Applications

Darlan Arruda (Ph.D. Candidate)
Nazim H. Madhavji

University of Western Ontario
London, Canada

Introduction

- Big Data is characterised by: Volume, Velocity, Variety and Veracity … of data.

- Currently, the body of knowledge on developing Big Data software applications is thin:
  - Gaps and challenges in the development of Big Data software applications (Madhavji et al., 2015; Anderson, 2015)
  - Need for processes for Big Data software projects (Kumar and Alencar, 2016).

- This motivated us to assess the situation in the field of Reqts Eng. (RE).

Research Questions

Q1. What RE activities, requirements types, and application domains are addressed in the literature on Big Data applications development ?

Q2. What solutions have been proposed in the domain of RE and Big Data applications?

Search Strategy

- **Sources:** ACM Digital Library, IEEE Xplore, ScienceDirect and Scopus.
- **Search String:**

*"**Big Data" AND** ("Requirements Engineering" **OR** "Requirements Elicitation" **OR** "Requirements Specification" **OR** "Requirements Analysis" **OR** "Requirements Validation" **OR** "Requirements Negotiation" **OR** "Requirements Prioritization" **OR** "Requirements Management")*

- **Selection Criteria:** Studies must …

(i) be in paper/article/chapter formats.

(ii) be written in English.

(iii) address any aspect of RE in the context of Big Data applications.

(iv) be published between 2010 and 2017.

Selection Process

- Three step selection process:
  - **First step:** results were filtered by their title and abstract.
  - **Second step:** results were assessed by reading their introduction and conclusion sections.
  - **Third step:** results were assessed by reading the entire paper.

Descriptive Data

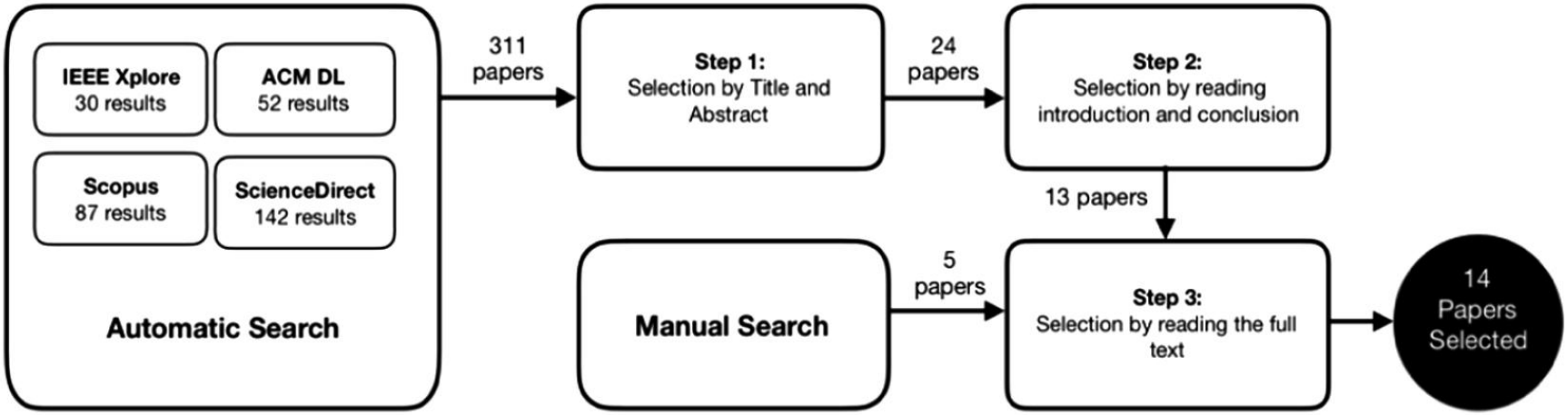


Fig. 1. Distribution of papers identified and selected organised by the phases in the selection process.

Descriptive Data

An observation:

- Conferences such as REFSQ and RE and
- Journals such as RE and IEEE Trans. on Big Data

… haven't yet published RE papers focused on the development of Big Data applications.

Descriptive Data

Table 2. Publication venue and number of papers from each venue

Publication venue	Issue, volume or year	Paper count
Conferences		
IEEE International Congress on Big Data	2013	1
International Conference on Data and Software Engineering	2014	1
International Conference on Cloud Computing, Data Science & Engineering	2017	1
IEEE International Conference on Big Data	2017	1
Workshops		
IEEE/ACM International Workshop on Big Data Software Engineering	2015 2016	2 1
International Workshop on Quality-Aware DevOps	2016	1
Journals		
International Journal of Ambient Systems and Applications	Vol. 2, No. 2/2014	1
IEEE Intelligent Systems	Vol. 30/2015	1
Books and Magazine		
Studies in Big Data – Springer	Vol. 05/2014	1
New Trends in Databases and Information Systems - Springer	Vol. 539/2015	1
Requirements Engineering Magazine	Issue 2016-01	1
Online Publication		
NIST Special Publication	Vol. 3/2015	1
Total		**14**

Descriptive Data

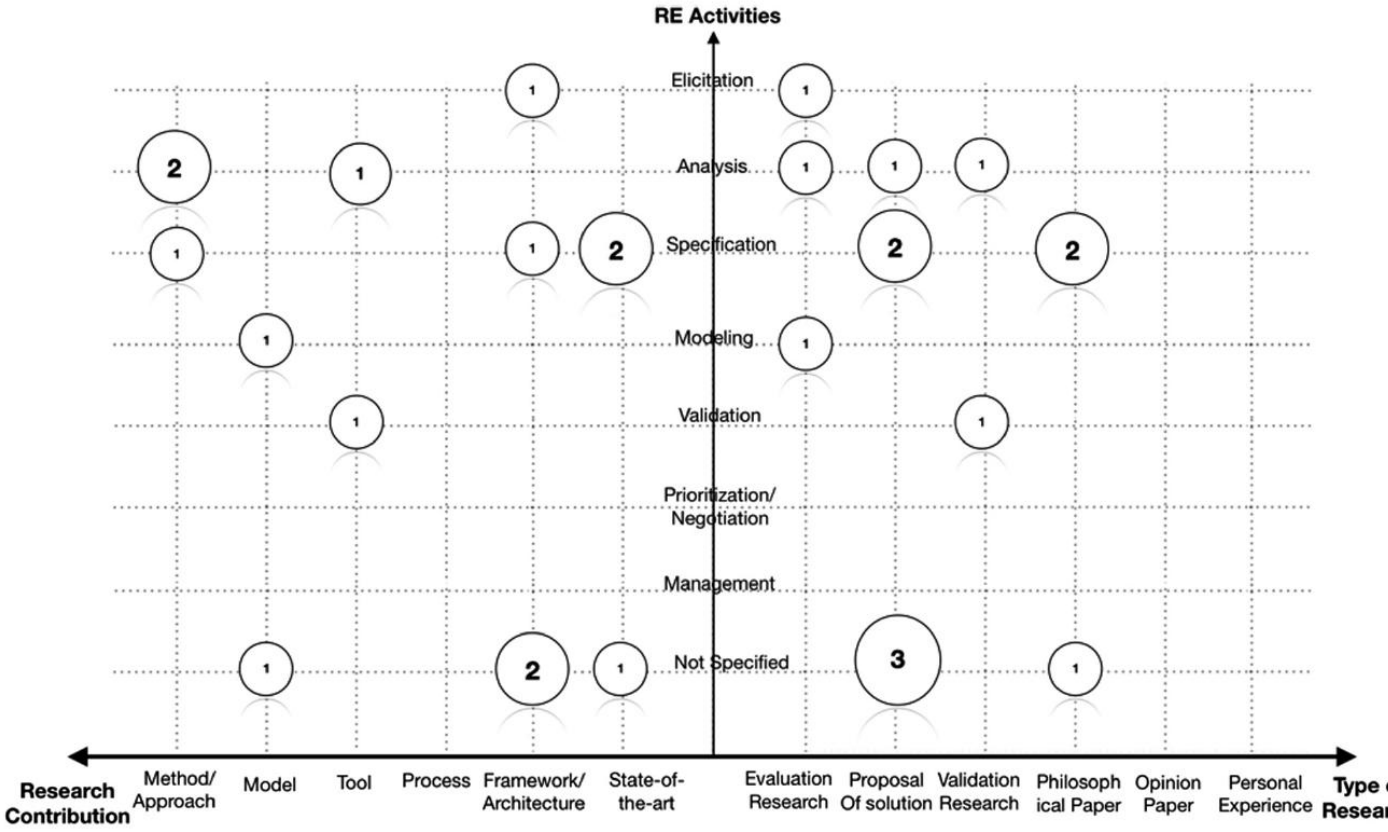


Fig. 3. Papers by contribution and type of research organised according to the RE activities they address.

Elicitation: 1
Analysis: 3
Spec.: 4
Modelling: 1
Validation: 1
Prioritisation: 0
Negotiation: 0
Management: 0
Unknown: 4

S-of-A: 3
Fram./Arch: 4
Process: 0
Tool: 2
Model: 2
Method/Appr,: 3

Evaluation: 3
Sol'n Proposal: 6
Validation: 2
Philosophical: 3

Results and Discussion

Results

Q1. *What RE activities, requirements types, and application domains are addressed in the literature on Big Data applications development?*

RE Activities	Paper Count	Type of Requirements	Paper Count	Application Domains
Elicitation	1	Functional Requirements	5	Healthcare Biomedical Research Government Marketing IT/Telecom Astronomy and Physics Earth Environmental and Polar Science Defense Commercial Social Media
Analysis	3	Quality Requirements	10	
Specification	4	Data Requirements	2	
Modelling	1	Architecturally Significant Requirements	1	
Validation	1	Not Specified	1	
Not Specified	4	**Note:** One paper could have discussed one or more types of requirements. Therefore, the sum of the papers in this table can be greater than the total number of papers selected.		

Results

RQ2. *What solutions have been proposed in the domain of RE and Big Data applications?*

Solutions Proposed	Author (s)
Approaches, Methods and Models	
Approach for handling non-functional requirements for Big Data projects in scrum	Sachdeva and Chung (2017)
Approach for analysing and specifying Quality Requirements	Noorwali *et. al.*, (2016)
Big Data System Design method	Chen *et al.*, (2015)
RE Generic model based on I* and KAOS	Eridaputra *et al.*, (2014)
RE Artefact Model in the Context of Big Data Software Projects	Arruda and Madhavji (2017)
Architectures and Frameworks	
Descriptive Architecture for Big Data Requirements Elicitation	Lau *et. al.*, (2014)
Requirements Specification framework for Big Data Collection	Al-Najran and Dahanayake (2015)
NIST Interoperability Framework*	NIST (2015)
Framework with security constraints	Youssef (2014)
Tools	
Verification Tool	Bersanini *et al.*, (2016)
UML extension for privacy requirements analysis	Jultla *et. al.*, (2013)

RE Artefact Model

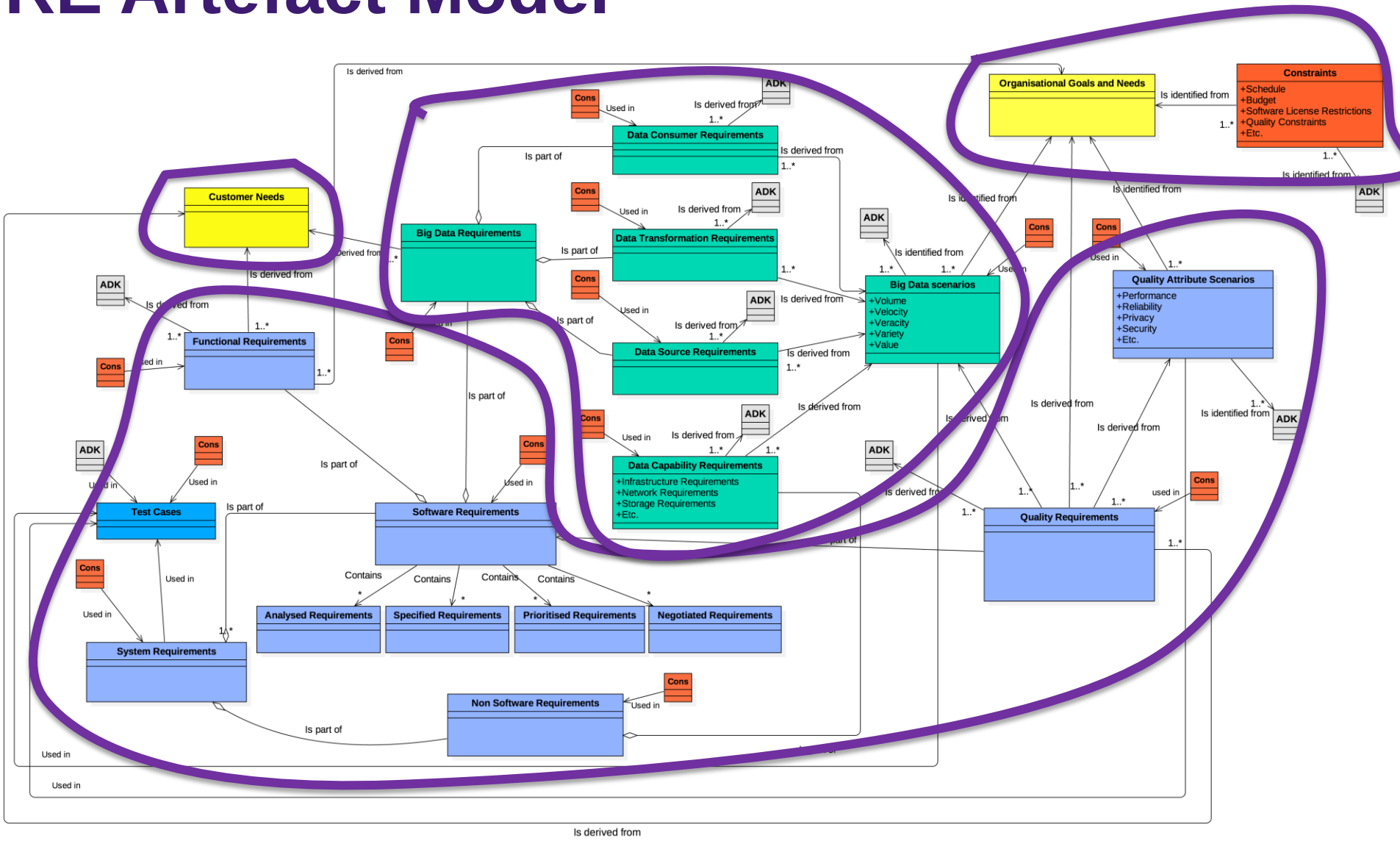
A model composed of three basic elements**: (a) Artefact , (b) Association** and **(c) Cardinality.** The following relationships are represented in the model:

- Is-derived-from
- Is-identified-from
- Is-part-of
- Contains
- Used in

Source: (Arruda and Madhavji, 2017)

Source: (Arruda and Madhavji, 2017)

RE Artefact Model



Results

RQ2. What solutions have been proposed in the domain of RE and Big Data applications?

Solutions Proposed	Author (s)
Approaches, Methods and Models	
Approach for handling non-functional requirements for Big Data projects in scrum	Sachdeva and Chung (2017)
Approach for analysing and specifying Quality Requirements	Noorwali *et. al.*, (2016)
Big Data System Design method	Chen *et al.*, (2015)
RE Generic model based on I* and KAOS	Eridaputra *et al.*, (2014)
RE Artefact Model in the Context of Big Data Software Projects	Arruda and Madhavji (2017)
Architectures and Frameworks	
Descriptive Architecture for Big Data Requirements Elicitation	Lau *et. al.*, (2014)
Requirements Specification framework for Big Data Collection	Al-Najran and Dahanayake (2015)
NIST Interoperability Framework*	NIST (2015)
Framework with security constraints	Youssef (2014)
Tools	
Verification Tool	Bersanini *et al.*, (2016)
UML extension for privacy requirements analysis	Jultla *et. al.*, (2013)

Approach for Specifying Big Data Quality Requirements

- In the non-RE but within the software engineering community, Big Data "V" characteristics are rarely explicitly discussed in conjunction with quality requirements.
    - Software analytics is the focus

- This raises the question as to the extent and the quality to which Big Data challenges are being addressed in the solution design.

Approach for Specifying Big Data Quality Requirements

[<Big Data characteristic> X <quality attribute>] + <requirement description>

If we intersect the first two parts, we get a number of permutations (e.g., variety × security; velocity × performance; veracity × privacy; volume × scalability; etc.).

This ensures that both quality and Big Data characteristics are addressed in requirements specifications.

Source: (Noorwali et al., 2016)

Approach for Specifying Big Data Quality Requirements

For example, in the requirement:

"*The system shall use a stream-processing engine with a latency of 0.5 – 2.0 seconds to respond to data in real-time between global earthquake sensors and the data centre*",

stream-processing engines (such as, Storm, S4, Spark or Samza) would handle velocity of Big Data; whereas, latency of 0.5-2.0 seconds specifies performance constraints.

This requirement is thus of the permutation: **velocity X performance.**

Source: (Noorwali et al., 2016)

Threats to Validity

- The paper describes threats and how we have addressed them.

Conclusions

- This study shows that, currently, research lacks:

  RE methods, tools and processes for: elicitation, negotiation, analysis, validation, prioritisation and management of requirements for Big Data applications.

- Research opportunities, but difficult access to real production environments.