

Astronomical big data testing of astronomical data analysis software and systems

Lixiao Zhou¹ and Maohai Huang²

Key Laboratory of Space Astronomy and Technology, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China

Email: lxzhou@nao.cas.cn, mhuang@nao.cas.cn



Abstract

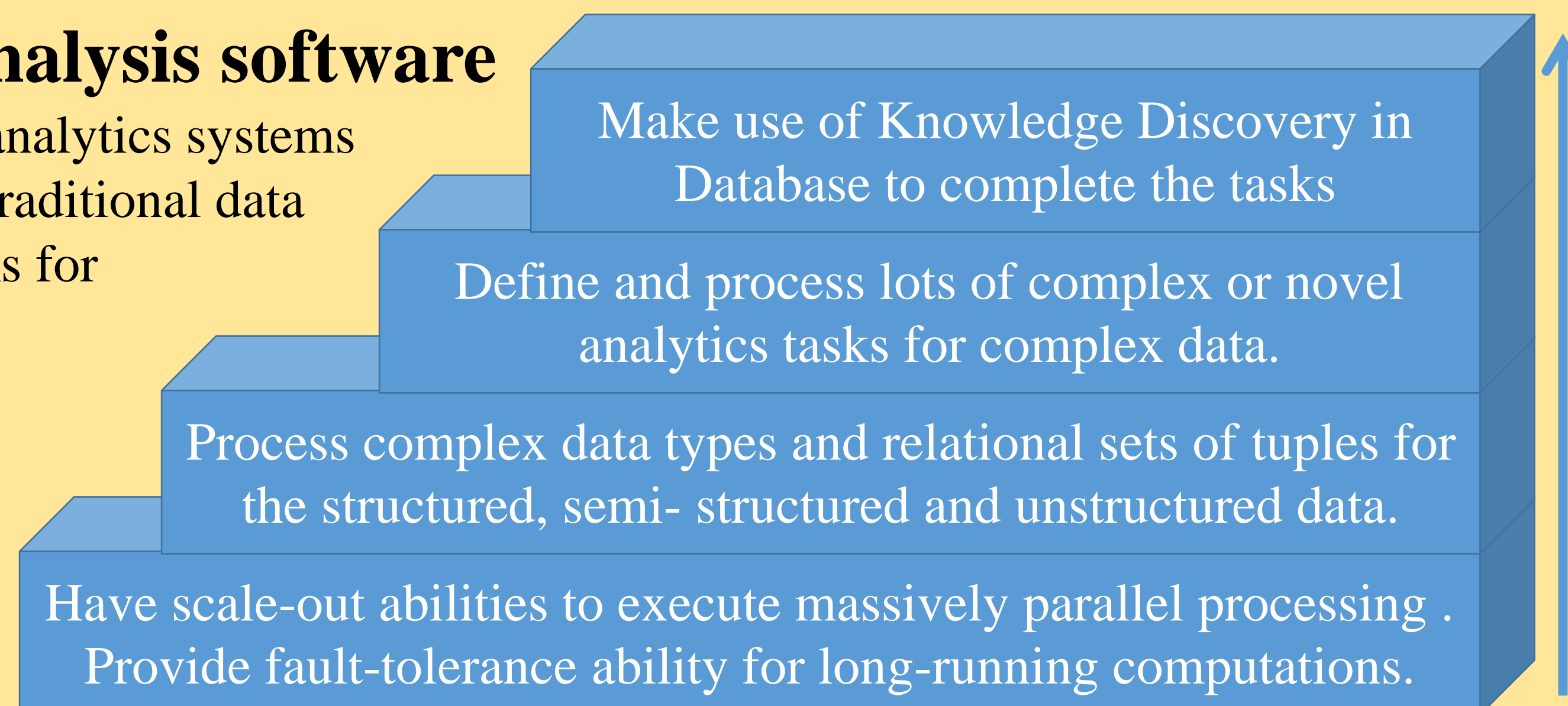
Big data is growing at a rapid pace, and one of the present hot trends in technology is the topic of big data and products targeting the big data problem.

“99% of organizations have a data quality strategy in place and 75% of businesses are wasting 14% of revenue due to poor data quality” -Experian Data Quality

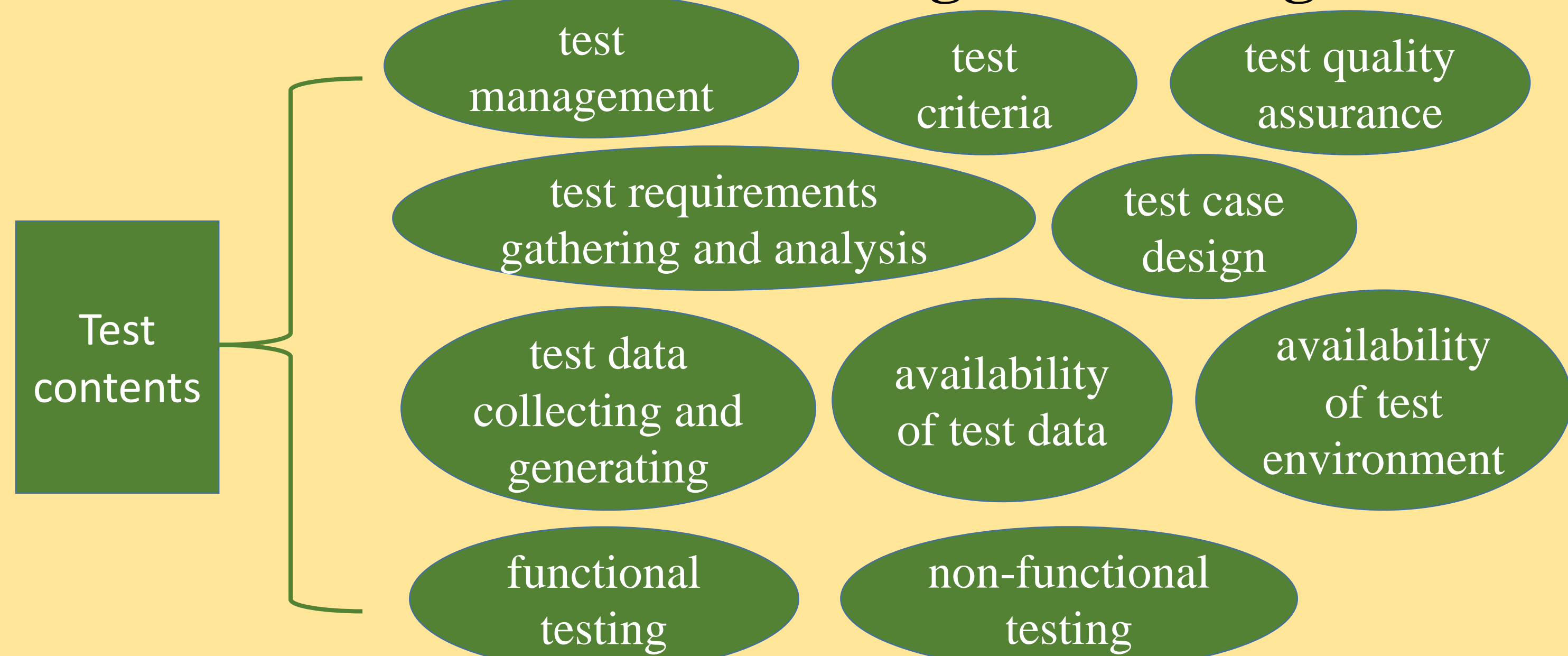
This paper will analyze and summarize characteristics, strategy, challenges for both general big data testing and astronomical big data testing.

Big data analysis software

Novel big data analytics systems differ from the traditional data analysis systems for their distinctive characteristics.



Contents and characteristics of big data testing



Test data

- Being unclear, complex, unstructured and non-semantic.
- Having low density but high value.
- Being from distributed and non-rational flat file storage.
- Having different storage formats, and is fit for No-SQL query.
- Being dynamic.
- Having no specific business rules to be applied.
- Having some unexpected changes with high velocity.
- Being implemented on the continuous data flow, and being done in the same platform of development with real time data at each stage.

test strategies and management methods

- Creating requirement traceability matrix between test requirement and software requirement specification or user requirement.
- Designing both positive and negative test cases.
- Setting up test environment and test platform, which is independent but just same as of development one.
- Using bug management tools, which cover bug reporting, monitoring and tracing.
- Analyzing and recording the root cause of bugs and the relevant solutions.
- Managing all test assets like texts and scripts in a unified platform.

Challenges in big data testing

Challenges for test requirement analysis and test cases design:

- There is huge difference in the thinking.
- Test scope cannot be defined easily.
- Many of them are “non-testable”, and the oracle problem obviously exists. The nature of prediction of big data leads to the difficulties for locating testing.
- How to correctly validate so complicated data types like unstructured and semi structured data should be considered.
- The data processing algorithms are very complex. How to decide the expected results and how to judge the actual results are a great challenge.
- The test cases are exploding due to all types of application scenarios.
- Adequately verifying and validating big data correctness is a grand challenge.

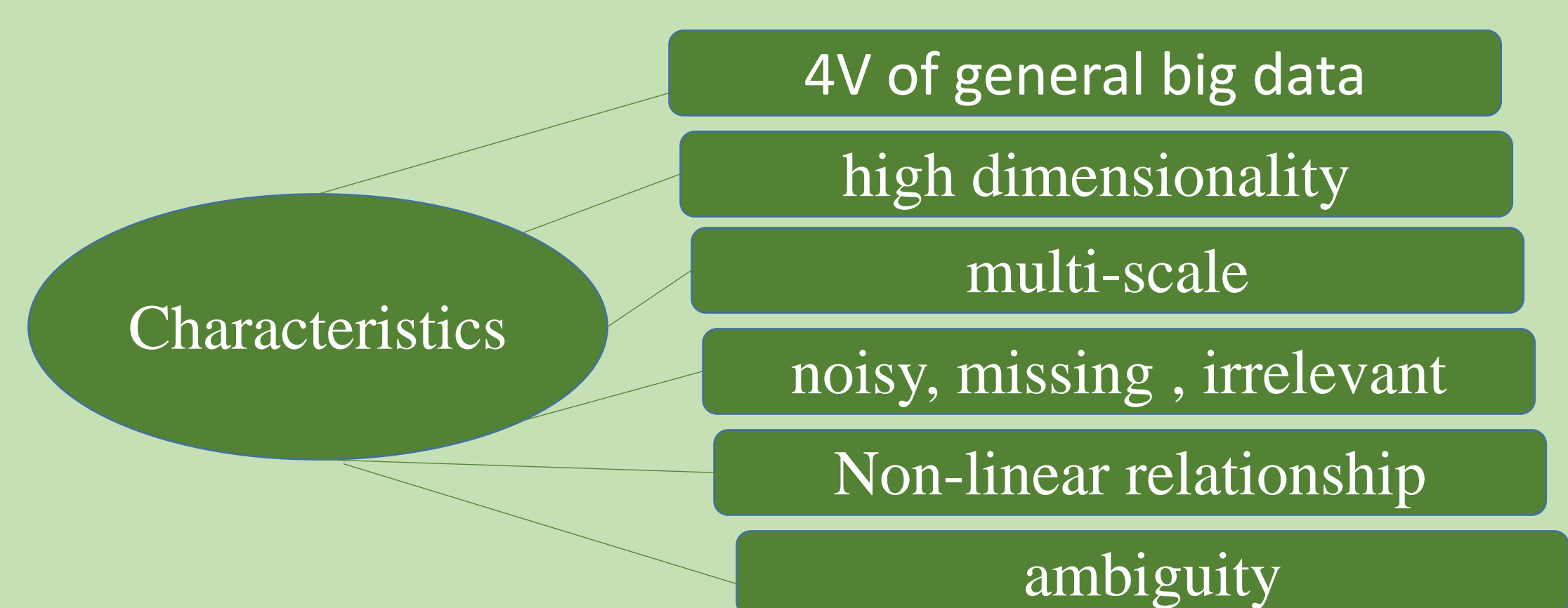
Challenges for test data preparation, test environment building and test case implementation:

- Making good preparation for test data when the data volume is so large
- Needing to be quickly generated huge test data sets
- Lacking realistic data sets in many industries
- Lacking a standardized data analysis language currently
- Looking for some small-scale test clusters which can detect some problems under the large-scale test clusters
- Testing big data requires new processes and a higher degree of automated tools.
- The visualization of big data testing results is worth studying.

Challenges for test management and the remainder:

- Shifting from being a quantitative to a qualitative challenge
- Increasing the test cost
- Needing many software testing engineers with several domains knowledge
- Having a contradiction between the complexity of software and the rhythm of new version release
- Having pesticide problem

Characteristics of astronomical data



Astronomical data analysis software and systems

In addition to functions and characteristics of general big data analytics systems, there are some special functions and characteristics in astronomical field.

- Having more focuses on timeliness, real-time, automaticity of data processing.
- Needing to provide more using modes, while general software can support one mode of GUI and console-batch.
- Needing to process more complex and diverse data.

Big data testing in astronomy

Some special contents and characteristics

- It is not easy to begin for a new test engineer without astronomical background.
- Most of test work is done through user trying the system.
- The test time has a longer duration because astronomical projects usually have long term and test data has strong complexity.
- It is difficult to develop or execute the entry and exit criteria of each test stage.
- The test effort of algorithm is more than that of general projects.
- The testing has more international cooperation and user collaboration.
- The expected test results are very difficult to make certain.
- Test work has to face the changing requirement due to so long-term project.
- There are some unexpected test problems.
- Proper and fully functional testing tools need to be introduced or developed.

Some special challenges

- Adequately testing scientific software
- Increasing new test contents
- Test management challenges because of distributed development teams, part-time efforts, highly constrained schedules and budgets

Conclusion: In order to better support astronomers, the pertinent methods, models and strategies for the better cost effective testing processes should be studied.

Future work: We are planning to study the required knowledge frame to prompt the astronomical big data testing advance in near future.

References: [1] Lu Luo, Software Testing Techniques Technology Maturation and Research Strategy Class Report for 17-939A, 2014. ...

[18] Evans, Janet D.; Plante, Raymond L.; Bonaventura, Nina, Managing distributed software development in the Virtual Astronomical Observatory, 2012