Distributed and Fault-Tolerant System for Tuple Streaming

Jorge Santos, Miguel Vera, José Semedo
Instituto Superior Tecnico
Computer Science Department
??IN-F?? Av. Rovisco Pais, 1, 1049-001 Lisboa
{jorge.pessoa, miguel.vera, jose.semedo}@tecnico.ulisboa.pt

Abstract

The ABSTRACT is to be in fully-justified italicized text, at the top of the left-hand column, below the author and affiliation information. Use the word "Abstract" as the title, in 12-point Times, boldface type, centered relative to the column, initially capitalized. The abstract is to be in 10-point, single-spaced type. The abstract may be up to 3 inches (7.62 cm) long. Leave two blank lines after the Abstract, then begin the main text.

1. Introduction

Please follow the steps outlined below when submitting your manuscript to the IEEE Computer Society Press. Note there have been some changes to the measurements from previous instructions.

2. Solutions

Describe the solutions in an overall way

2.1. Solution 1

Solution 1 advantages/disadvantages

2.2. Solution 2

Solution 1 advantages/disadvantages

3. Used solution

Describe the solutions in an overall way

3.1. Fault-tolerance algorithm

Fault tolerance algorithm: how it works, etc

3.2. Semantics algorithm

One of the biggest concern in the distributed tuple processing is being able to a guarantee a certain semantics associated with the processing of a tuple in the presence of system failures, and inevitably the way in needs to reconfigure in order to continue functioning properly. This issue is inevitably linked to the way that the system does fault tolerance. The algorithm used for at least once and exactly once is very simmilar, as well as using the same data structures and most of the same procedures, but the differences will be explained below. In the follow subsections we explain how the system assures that a tuple is processed at most once, at least once and exactly once.

3.2.1 At-Most-Once

To assure that a tuple is processed at most once, a system must only send any tuple once, indepent of failure. Since a configuration for this system is dependent on being an ?acyclic? graph, the only guarantee we need to provide is that there isn't any kind of mechanism to resend tuples in case of a takeover of a node that crashed by another node. As such the implementation of this strategy relies on not sharing any kind of information about tuples already processed, and not doing anything upon the takeover. This way, a certain tuple from an input is only sent once in the forward direction of the distributed network, and in case of a node failing all the tuples that it had processed but not sent yet won't be processed at all.

3.2.2 At-Least-Once

(???????????????????Maybe should be in a upper section) If the system must process a certain tuple at least once then there must a guarantee that in any kind of a node failure the tuples that weren't successfuly sent to a node in failure are sent either to another node or to the the original node if it recovers. There are however other scenarios in which a node can fail, for example after receiving a tuple but before sending it to the next operator if there is any. Keeping in mind that the system must obviously be ashynchronous in this confirmation, then the previous node can't easily know it has to resend the tuple, and when it doesn't, or how long it needs to keep the tuple. In this approach the tuples would need to be kept in every operator at least until the tuple was processed by every node, presenting scalability issues. (DATA STRUCTURES) To understand how the algorithm works we must first explain the used data structures: A tuple is identified by it's Tuple Id structure which represents a stream of a tuple along the processing chain. In each specific tuple there is an unique id that is usualy kept along all operators, unless there must an output of several tuples from the same one, efectively divering the tuple stream. The remaining information kept in the Tuple Id refers to the operator and replica it came from. In each node there is a delivery table (a simple Hash-based Map that stores each tuple by its unique id) and that keeps every tuple received in a replica until it can be disposed. There is aswell a shared tuple table in each node that stores Tuple Records associated to a tuple id (in a simillar Hash Map). A tuple record is a small representation of a tuple containing its Tuple Id, as well as the replica emiting the tuple record. This way the tuple record contains only the necessary information for a node to re-process a tuple that wasn't properly processed due to failure. For information storing purposes the tuple records can have a state (pending or purged) that is relevant since a tuple with the purged state is guaranteed to have been processed on that node. (ALGORITHM): When a node receives a tuple

3.2.3 Exactly-Once

4. Evaluation

Evaluation of the solution

4.1. Quantitative

Quantitive evaluation

4.2. Qualitative

Qualitative evaluation

5. Discussion

Evaluation of the solution

6. Conclusion

Evaluation of the solution

7. Style Guidelines

Please follow the steps outlined below when submitting your manuscript to the IEEE Computer Society Press. Note there have been some changes to the measurements from previous instructions.

7.1. Margins and page numbering

All printed material, including text, illustrations, and charts, must be kept within a print area 6-7/8 inches (17.5 cm) wide by 8-7/8 inches (22.54 cm) high. Do not write or print anything outside the print area. Number your pages lightly, in pencil, on the upper right-hand corners of the BACKS of the pages (for example, 1/10, 2/10, or 1 of 10, 2 of 10, and so forth). Please do not write on the fronts of the pages, nor on the lower halves of the backs of the pages.

7.2. Formatting your paper

All text must be in a two-column format. The total allowable width of the text area is 6-7/8 inches (17.5 cm) wide by 8-7/8 inches (22.54 cm) high. Columns are to be 3-1/4 inches (8.25 cm) wide, with a 5/16 inch (0.8 cm) space between them. The main title (on the first page) should begin 1.0 inch (2.54 cm) from the top edge of the page. The second and following pages should begin 1.0 inch (2.54 cm) from the top edge. On all pages, the bottom margin should be 1-1/8 inches (2.86 cm) from the bottom edge of the page for 8.5×11-inch paper; for A4 paper, approximately 1-5/8 inches (4.13 cm) from the bottom edge of the page.

7.3. Type-style and fonts

Wherever Times is specified, Times Roman may also be used. If neither is available on your word processor, please use the font closest in appearance to Times that you have access to.

MAIN TITLE. Center the title 1-3/8 inches (3.49 cm) from the top edge of the first page. The title should be in Times 14-point, boldface type. Capitalize the first letter of nouns, pronouns, verbs, adjectives, and adverbs; do not capitalize articles, coordinate conjunctions, or prepositions (unless the title begins with such a word). Leave two blank lines after the title.

AUTHOR NAME(s) and AFFILIATION(s) are to be centered beneath the title and printed in Times 12-point, non-boldface type. This information is to be followed by two blank lines.

The ABSTRACT and MAIN TEXT are to be in a two-column format.

MAIN TEXT. Type main text in 10-point Times, single-spaced. Do NOT use double-spacing. All paragraphs should be indented 1 pica (approx. 1/6 inch or 0.422 cm). Make sure your text is fully justified—that is, flush left and flush right. Please do not place any additional blank lines between paragraphs. Figure and table captions should be 10-point Helvetica boldface type as in

Figure 1. Example of caption.

Long captions should be set as in

Figure 2. Example of long caption requiring more than one line. It is not typed centered but aligned on both sides and indented with an additional margin on both sides of 1 pica.

Callouts should be 9-point Helvetica, non-boldface type. Initially capitalize only the first word of section titles and first-, second-, and third-order headings.

FIRST-ORDER HEADINGS. (For example, **1. Introduction**) should be Times 12-point boldface, initially capitalized, flush left, with one blank line before, and one blank line after.

SECOND-ORDER HEADINGS. (For example, **1.1. Database elements**) should be Times 11-point boldface, initially capitalized, flush left, with one blank line before, and one after. If you require a third-order heading (we discourage it), use 10-point Times, boldface, initially capitalized, flush left, preceded by one blank line, followed by a period and your text on the same line.

7.4. Footnotes

Please use footnotes sparinglyavoid footnotes altogether. To help your readers, avoid using footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence). and place them at the bottom of the column on the page on which they are referenced. Use Times 8-point type, single-spaced.

7.5. References

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example [**?**]. Where appropriate, include the name(s) of editors of referenced books.

7.6. Illustrations, graphs, and photographs

All graphics should be centered. Your artwork must be in place in the article (preferably printed as part of the text rather than pasted up). If you are using photographs and are able to have halftones made at a print shop, use a 100- or 110-line screen. If you must use plain photos, they must be pasted onto your manuscript. Use rubber cement to affix the images in place. Black and white, clear, glossy-finish photos are preferable to color. Supply the best quality photographs and illustrations possible. Penciled lines and very fine lines do not reproduce well. Remember, the quality of the book cannot be better than the originals provided. Do NOT use tape on your pages!

7.7. Color

The use of color on interior pages (that is, pages other than the cover) is prohibitively expensive. We publish interior pages in color only when it is specifically requested and budgeted for by the conference organizers. DO NOT SUBMIT COLOR IMAGES IN YOUR PAPERS UNLESS SPECIFICALLY INSTRUCTED TO DO SO.

7.8. Symbols

If your word processor or typewriter cannot produce Greek letters, mathematical symbols, or other graphical elements, please use pressure-sensitive (self-adhesive) rub-on symbols or letters (available in most stationery stores, art stores, or graphics shops).

7.9. Copyright forms

You must include your signed IEEE copyright release form when you submit your finished paper. We MUST have this form before your paper can be published in the proceedings.

7.10. Conclusions

Please direct any questions to the production editor in charge of these proceedings at the IEEE Computer Society Press: Phone (714) 821-8380, or Fax (714) 761-1784.