Key-Value stores Riak vs ArangoDB

Αγγελου Στελιος
αριθμος μητρώου

Γκέργκες Μάρκος
03117870

Παπαβασιλόπουλος Παναγιώτης-Φίλιππος
03117444

Abstract—**This document is a model and instructions for LATEX. This and the IEEEtran.cls file define the components of your paper [title, text, heads, etc.]. *CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.**

Index Terms—**component, formatting, style, styling, insert**

I. INTRODUCTION

II. Σκοπός της εργασίας

Σε αυτή την εργασία θα γίνει σύγκριση όσο αφορά την φόρτωση δεδομένων, select queries, range queries, update queries, delete queries, insert queries, των δύο key-value stores Riak και ArangoDB. Θα γίνει σύγκριση των δύο key-value stores σε διάφορες περιπτώσεις όπως αναφέρεται παρακάτω. Στην συγκεκριμένη εργασία θα χρησιμοποιηθεί το γνωστό tool YCSB benchmark.

III. Περιγραφή βημάτων για το στήσιμο του συστήματος και των πειραμάτων

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections VI-A–VI-E below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not number text heads—LATEX will do that for you.

IV. Περιγραφή της υποδομής και του SOFTWARE που χρησιμοποιήθηκε

A. Okeanos

B. Yahoo Cloud Service Benchmark

C. ArangoDB

Η ArangoDB είναι μια ανοιχτού κώδικά βάση δεδομένων NoSQL. Χαρακτηριστικό της είναι ο ευέλικτος τρόπος μοντελοποίησης των δεδομένων σε documents ή γράφους. Έχει σχεδιαστεί ως «βάση δεδομένων γενικού σκοπού», προσφέροντας όλες τις λειτουργίες που απαιτούνται συνήθως για τις σύγχρονες εφαρμογές Ιστού. Υποστηρίζει μοντέλα δεδομένων γραφημάτων, εγγράφων και κλειδιών-τιμών που επιτρέπουν στους χρήστες να συνδυάζουν ελεύθερα όλα τα μοντέλα δεδομένων σε ένα μόνο query. Καθώς οι εφαρμογές γίνονται όλο και πιο περίπλοκες, συχνά χρειαζόμαστε περισσότερες από μία βάσεις δεδομένων NoSQL. Η χρήση μιας βάσης δεδομένων πολλαπλών μοντέλων μπορεί να απλοποιήσει την αρχιτεκτονική μας συνδυάζοντας πολλούς τύπους NoSQL σε μια ενιαία υποδομή.

Σύμφωνα με αυτό, θα μπορούσε κανείς, για παράδειγμα, να χρησιμοποιήσει μια σχεσιακή βάση δεδομένων για να διατηρήσει δομημένα, πινακοποιημένα δεδομένα. ένα χώρο αποθήκευσης εγγράφων για μη δομημένα δεδομένα που μοιάζουν με αντικείμενα. ένα κατάστημα κλειδιού/τιμής για έναν πίνακα κατακερματισμού. και μια βάση δεδομένων γραφημάτων για ιδιαίτερα συνδεδεμένα αναφορικά δεδομένα.

Ωστόσο, η παραδοσιακή εφαρμογή αυτής της προσέγγισης θα οδηγήσει κάποιον στη χρήση πολλαπλών βάσεων δεδομένων στο ίδιο project. Μπορεί να οδηγήσει σε κάποιες λειτουργικές τριβές (πιο περίπλοκη ανάπτυξη, πιο συχνές αναβαθμίσεις) καθώς και σε ζητήμα τα συνέπειας και συντήρησης των ίδων δεδομένων πολλαπλές φορές.

Ποια ειναι τα κύρια χαρακτηριστικά της ArangoDB;

- Δυνατότητα για χρησιμοποίηση διαφορετικών τεχνολογιών αποθήκευσης, που ταιριάζουν καλύτερα με τον τρόπο που χρησιμοποιούνται τα δεδομένα από διαφορετικά στοιχεία της εφαρμογής.
- Προσφέρεται πρόσβαση σε οποιαδήποτε δεδομένα (ανεξάρτητα από το μοντέλο τους) χρησιμοποιώντας μια ενιαία γλώσσα ερωτημάτων, η οποία είναι αρκετά παρόμοια με μια τυπική γλώσσα ερωτήματος SQL με κάποιες μικρές διαφορές.
- Χρησιμοποιεί δυαδικό JSON για την αποθήκευση δεδομένων. Είναι λιτό, αυτόνομο, βελτιώνει τις απαιτήσεις από μνήμη και καλύπτει όλους τους JSON τύπους δεδομένων συν ημερομηνίες, δυαδικά δεδομένα, ακέραιους αριθμούς, καθώς και αυθαίρετους αριθμούς ακριβείας.

ArangoDB: A distributed open-source database with a flexible data model for documents, graphs, and key-values. A distributed free and open-source database with a flexible data model for documents, graphs, and key-values. Build high performance applications using a convenient SQL-like query language or JavaScript extensions;

D. Riak KV

A distributed, decentralized data storage system. Riak is a distributed database designed to deliver maximum data availability by distributing data across multiple servers. As long as your client can reach one Riak server, it should be able to write data. In most failure scenarios, the data you want to read should be available, although it may not be the most up-to-date version of that data.

V. περιγραφή των αποτελεσμάτων και των συμπερασμάτων
της εργασίας

A. Workload a

This workload has a mix of 50/50 reads and writes. An application example is a session store recording recent actions. Updates in this workload do not presume you read the original record first. The assumption is all update writes contain fields for a record that already exists; oftentimes writing only a subset of the total fields for that record. Some data stores need to read the underlying record in order to reconcile what the final record should look like, but not all do.
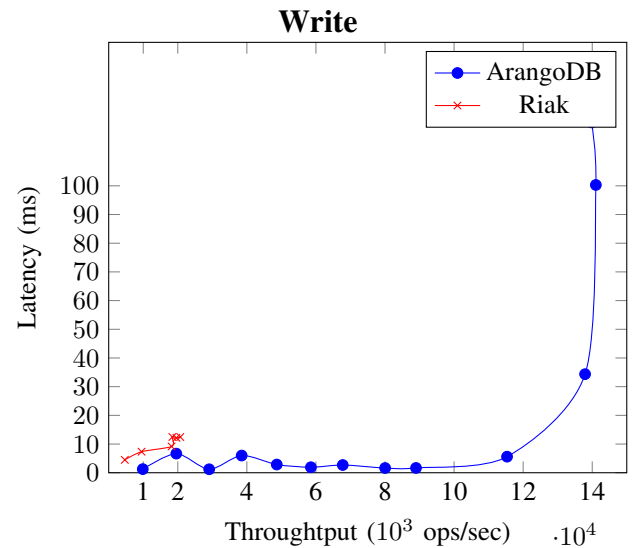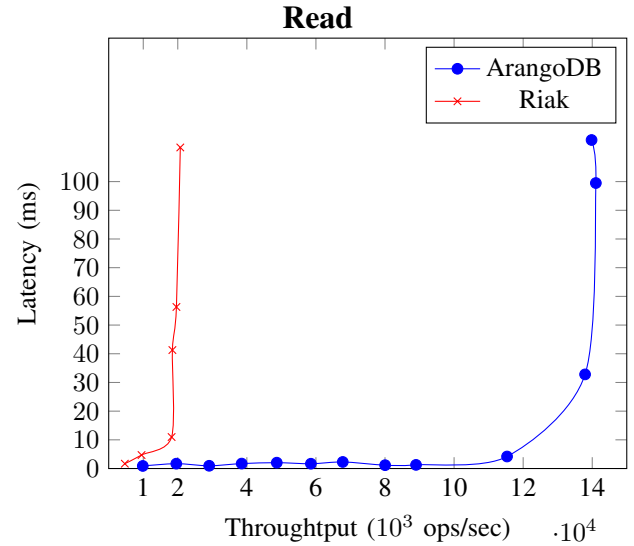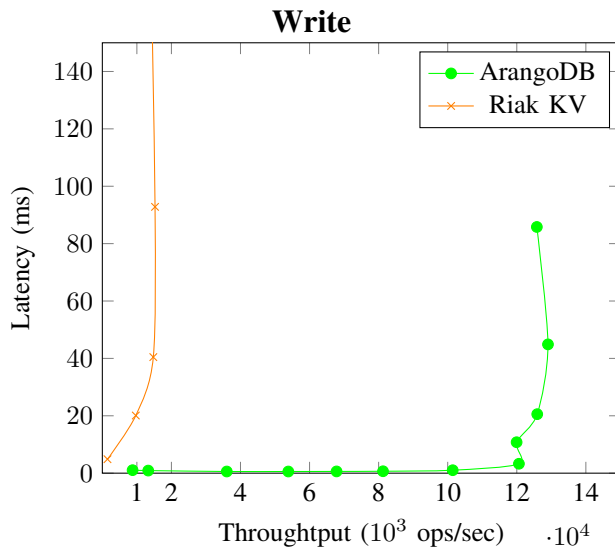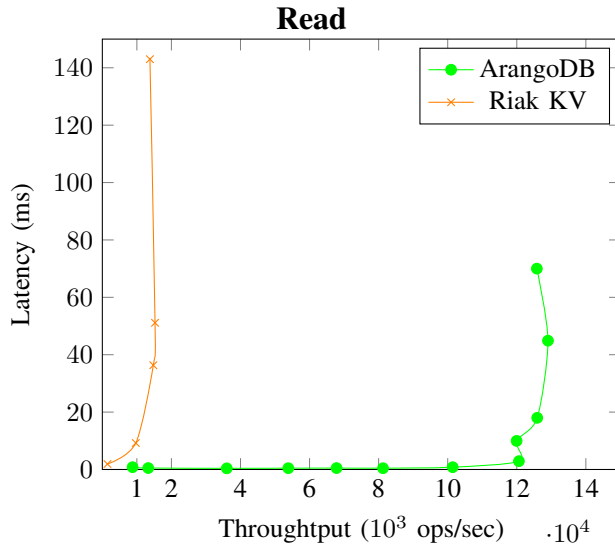
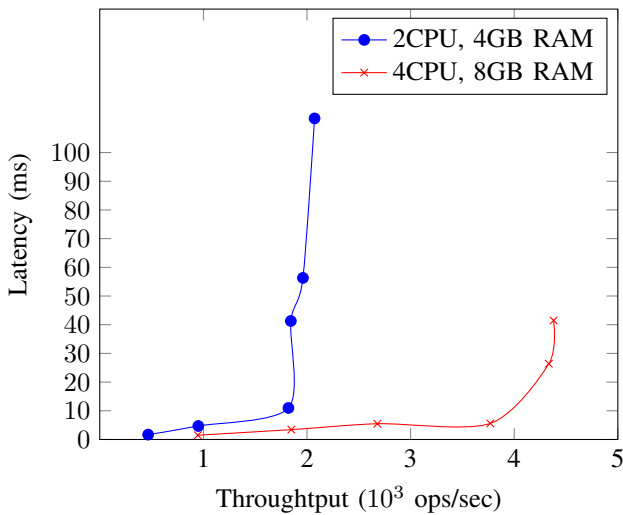*B. Workload b*

C. Διπλασιάζοντας του πόρους της Riak

1) Workload a - Updates: Από το worklfow a θα συγκρίνουμε τις επιδόσεις στα reads, μιας και τα reads θα εξετασθούν από το heavy-read workflow b.

Writes of workflow a

Latency (ms) vs **Throughput (10^3 ops/sec)**

Legend: 2CPU, 4GB RAM — 4CPU, 8GB RAM
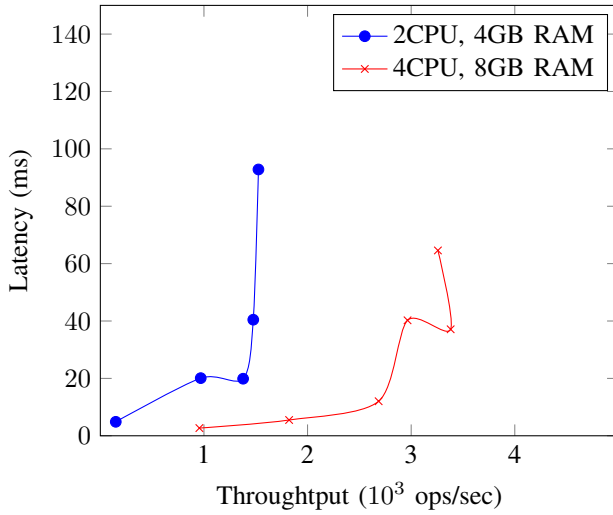
2) Workload b - Reads:

Heavy Read - Workload B

Latency (ms) vs **Throughput (10^3 ops/sec)**

Legend: 2CPU, 4GB RAM — 4CPU, 8GB RAM

VI. βιβλιογραφία και παραπομπές που χρησιμοποιήθηκαν στο κείμενο

A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, ac, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

B. abcd

- Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as "3.5-inch disk drive".
- Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: "Wb/m^2" or "webers per square meter", not "webers/m^2". Spell out units when they appear in text: ". . . a few henries", not ". . . a few H".
- Use a zero before decimal points: "0.25", not ".25". Use "cm^3", not "cc".)

C. Equations

Number equations consecutively. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

$$a + b = \gamma \tag{1}$$

Be sure that the symbols in your equation have been defined before or immediately following the equation. Use "(1)", not "Eq. (1)" or "equation (1)", except at the beginning of a sentence: "Equation (1) is . . ."

D. LaTeX-Specific Advice

Please use "soft" (e.g., `\eqref{Eq}`) cross references instead of "hard" references (e.g., `(1)`). That will make it possible to combine sections, add equations, or change the order of figures or citations without having to go through the file line by line.

Please don't use the `{eqnarray}` equation environment. Use `{align}` or `{IEEEeqnarray}` instead. The `{eqnarray}` environment leaves unsightly spaces around relation symbols.

Please note that the `{subequations}` environment in LaTeX will increment the main equation counter even when there are no equation numbers displayed. If you forget that, you might write an article in which the equation numbers skip from (17) to (20), causing the copy editors to wonder if you've discovered a new method of counting.

BIBTeX does not work by magic. It doesn't get the bibliographic data from thin air but from .bib files. If you use BIBTeX to produce a bibliography you must send the .bib files.

LaTeX can't read your mind. If you assign the same label to a subsubsection and a table, you might find that Table I has been cross referenced as Table IV-B3.

LaTeX does not have precognitive abilities. If you put a `\label` command before the command that updates the counter it's supposed to be using, the label will pick up the last counter to be cross referenced instead. In particular, a `\label` command should not go before the caption of a figure or a table.

Do not use `\nonumber` inside the `{array}` environment. It will not stop equation numbers inside `{array}` (there won't be any anyway) and it might stop a wanted equation number in the surrounding equation.

E. Some Common Mistakes

- The word "data" is plural, not singular.
- The subscript for the permeability of vacuum μ_0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter "o".
- In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
- A graph within a graph is an "inset", not an "insert". The word alternatively is preferred to the word "alternately" (unless you really mean something that alternates).
- Do not use the word "essentially" to mean "approximately" or "effectively".
- In your paper title, if the words "that uses" can accurately replace the word "using", capitalize the "u"; if not, keep using lower-cased.
- Be aware of the different meanings of the homophones "affect" and "effect", "complement" and "compliment", "discreet" and "discrete", "principal" and "principle".
- Do not confuse "imply" and "infer".
- The prefix "non" is not a word; it should be joined to the word it modifies, usually without a hyphen.
- There is no period after the "et" in the Latin abbreviation "et al.".
- The abbreviation "i.e." means "that is", and the abbreviation "e.g." means "for example".

An excellent style manual for science writers is [7].

F. Authors and Affiliations

The class file is designed for, but not limited to, six authors. A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

G. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is "Heading 5". Use "figure caption" for your Figure captions, and "table head" for your table title. Run-in heads, such as "Abstract", will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced.

H. Figures and Tables

a) Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation "Fig. 1", even at the beginning of a sentence.

Table I
TABLE TYPE STYLES

Table Head	Table Column Head		
	Table column subhead	*Subhead*	*Subhead*
copy	More table copy[a]		

[a]Sample of a Table footnote.

Figure 1. Example of a figure caption.

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity "Magnetization", or "Magnetization, M", not just "M". If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write "Magnetization (A/m)" or "Magnetization {A[m(1)]}", not just "A/m". Do not label axes with a ratio of quantities and units. For example, write "Temperature (K)", not "Temperature/K".

ACKNOWLEDGMENT

REFERENCES

Please number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use "Ref. [3]" or "reference [3]" except at the beginning of a sentence: "Reference [3] was the first . . ."

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove the template text from your paper may result in your paper not being published.