Performance and Cost Implications of Language Selection for Serverless Data Processing Pipelines (use style: *paper title*)

Subtitle as needed (style: *paper subtitle*)

Authors Name(s) first and last seperated by commas

School of Engineering and Technology

University of Washington
Tacoma, Washington USA

email1@uw.edu, email2@uw.edu, ermail3@uw.edu

*Abstract*—Please provide an abstract for your paper. The abstract should introduce the serverless application your team has implemented. Describe the key features provided by the application as well as the design tradeoffs you studied. (e.g. service composition, application flow control, alternate back-end databases, etc.) The abstract can include one or two key performance, throughput, or cost results your group measured after implementing the application. *\*CRITICAL: Do Not Use Symbols, Special Characters, or Math in Paper Title or Abstract*. (*Abstract*)

Keywords—This keywords for the paper: component; formatting; style; styling; insert (key words)

# Introduction

For the report title, please adjust the title above to describe your project. In the introduction, please introduce the serverless application you’ve implemented. If your group did not implement the standard “TLQ” application, please include background to fully describe the application from the perspective of someone who is unfamiliar with it. Be sure to define acronyms and include references where appropriate. If your group implemented the “TLQ” application, first describe your group’s interpretation of the TLQ application, and describe anything that may be different or unique regarding your group’s implementation. (e.g. languages used, tools, alternate cloud services, databases, etc.)

## Research questions

For your paper, write two or more “research” questions which you will answer through the performance experiments with respect to the design tradeoffs studied by your group. These research questions ask about how the tradeoffs of alternate implementations impact key indicators such as: performance, throughput, and/or cloud hosting costs ($). (e.g. switchboard vs. full service isolation, etc.) Your research questions may also consider the performance and cost implications of the memory reservation size on FaaS platforms. Since CPU power is coupled to memory size on AWS Lambda, the memory setting will impact performance, and also cost due to the FaaS pricing policy. Your research questions may also want to consider the freeze-thaw lifecycle of serverless computing. The switchboard architecture pattern, for example, is designed specifically to minimize infrastructure initialization overhead.

We investigate the following research questions: (Replace with your specific research questions)

1. How does service composition of application microservices impact the application performance when hosted using Function-as-a-Service platforms?
2. How does Function-as-a-Service memory reservation size impact application performance, throughput, and hosting costs for the “TLQ” pipeline application?

Please write or adapt the sample research questions and tailor them to your project. The paper should then present data that tries to answer the questions. Through the written narrative refer to the research questions by number (RQ-1), (RQ-2), etc. This helps tie together the presentation of your research questions and research approach with your experimentation and analysis. **Do not state research questions that are entirely unexplored by your group project.** Given time constraints in TCSS 562, it is entirely possible that research questions may not be entirely answered by the study. However, there should be at least some data, and an attempt to explore the question(s). Remember the experimental results simply provide data. They do not provide “right” or “wrong” answers or results. It is the role of the researcher to correctly design and execute experiments and report on the results regardless if the data is what we expected or not!

# Comparison Study

Section II of your paper should describe the serverless application implemented by your group. The first paragraph can provide additional detail not already in the introduction regarding the application.

One paragraph could describe why it is interesting to implement your application as a FaaS application case study. What is in particular challenging or interesting, (e.g. machine learning, data processing pipeline) where having access to the features of FaaS platforms (e.g. access to many CPUs in parallel for on-demand scaling, high availability, 24/7, pay only for resources actually used, no idle servers) is paramount?

## Design Tradeoffs

Sub-section A should describe the design tradeoffs studied by your group. (e.g. alternate service compositions, application flow control, alternate back-end relational databases, etc.) Include diagrams, tables, or textual descriptions to depict your alternate implementations. Consider labeling the alternate implications for easy reference throughout the text for abbreviation in other tables, graphs, etc. A brief description on the expectations of your alternate designs could be included. For example, what do you expect to see when testing alternate flow control implementations? In the end, it will be interested to see how expectations stand up to real world tests.

## Application Implementation

Sub-section B should describe the details for your application implementation. Describe any special languages, development libraries, tools, databases, external cloud services, and/or other technologies used in the implementation of your application. The description of your application implementation should help the reader understand how your application has been designed. When you report on the performance, throughput, and hosting costs of your tests, a thorough description of your application can help the reader understand what system you actually tested. This helps the reader understand if your testing results may be relevant to applications in their problem domain. Diagrams and figures can be included if helpful.

Please describe how ***state*** information is tracked and maintained in your implementation. For the “TLQ” application, how was data passed among the services? Are there separate SQL-Lite databases for each user session? If so, this would allow your TLQ pipeline to process many independent “flows” of data in parallel. If there is just one-database, do you differentiate between users? Has the application been designed to be multi-user?

If a user calls the “load” service to populate a database, is the application flow control among services asynchronous, synchronous, or does your application support both schemes? How does your application support asynchronous flow control? How are messages persisted by the server, and retrieved by the client?

## Experimental Approach

Sub-section C should describe your experimental approach to test the implications of, for example, service composition, application flow-control, memory reservation size, etc. Describe the client-side infrastructure used for testing. Described the server-side infrastructure configurations for testing. For example, How much memory was allocated to Lambda functions? Was the timeout adjusted? Were functions assigned to run in a VPC? A single availability zone? Which cloud region(s) were used? How did the client(s) communicate with your serverless application?

Section II.C should describe the test configurations, and tests that were performed so in the future someone could repeat them. Repeating tests may help to verify results and conclusions. Repeating tests can also help determine how public cloud service quality and implementations are changing over time. Are cloud providers addressing issues in class that we’ve talked about regarding performance variability? Is it easier from a user’s perspective to “know what you’re paying for”, let’s say in the future?

Groups may wish to write custom BASH/ Python scripts to implement all tests to make it easier to repeat tests later on. If test scripts were developed, these could be included in a git repository and the project report could refer to them.

## Subsections

Extra-subsection if helpful.

## Another Subsection

* Here is an example of a bulleted list-
* In case you need it.
* This is a third bullet.

## Equations

Equations, if used, are usually offset in the text. The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop. 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* 

Note that the equation is centered using a center tab stop. 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 . . .”

## Common Mistakes and Best Practices

* The word “data” is plural, not singular.
* The phrase “in order” usually has no useful meaning and can be deleted.
* The subscript for the permeability of vacuum **0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.
* In American English, commas, semi-/colons, 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].

# Experimental Results

Section III should present the results of your study. This should include data tables and graphs with captions where appropriate. Here are some simple examples of captions:

1. Table Type Styles

| Table Head | Table Column Head |
| --- | --- |
| Table column subhead | Subhead | Subhead |
| copy | More table copya |  |  |

1. Sample of a Table footnote. (*Table footnote*)

We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi TIFF or EPS file, with all fonts embedded) because, in an MSW document, this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on your frame, use the MSWord “Format” pull-down menu, select Text Box > Colors and Lines to choose No Fill and No Line.

1. Example of a figure caption. (*figure caption*)

Any graphs should be included as figures. Data should be presented in table format. In addition to presenting the experimental results, Section III should walk the reader through the presentation of the results with a discussion and analysis. The ***discussion and analysis*** will describe what the team believes results mean and why they were produced. After completing the tests, and collecting the data, what appears to be happening? Please state any conclusions from your data and your observations that you can now make. Conclusions not backed by data, observation, or experience (use cases, development experiences, etc.) are not scientific and should be avoided.

# Conclusions

Section IV should summarize the study. It should reiterate what the research questions were, and restate the key conclusions from section III based on the experimental results. Ideally, there are one or two conclusions that can summarize your observations for each research question. You can even identify the research question by number such as (RQ-1) when summarizing the key conclusion from your data. Someone reviewing your paper will look to see if in your conclusion section, you’ve answered the research questions proposed in the introduction. Good research papers **do not hide key results** from the reader. **They also do not make the reader do the analysis.** The authors of papers should not simply present a ton of data, and then leave it up to the reader to figure out what it all means! The authors should not present data to the reader and only “discuss” passively what it ***might*** mean. Good research papers make assertive conclusions based on facts and data. The core components of the research paper should be accessible in the introduction (Section I) and the conclusion (section IV). A reader wanting more detailed information can then read the intermediate sections if they find your paper of interest.

##### Acknowledgment *(Heading 5)*

Acknowledgements may be included. We can cite the use of AWS Educate cloud credits.

##### References

This section describes the format for the references. You should have at least 5!!! The template will 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 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].

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. *(references)*
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.