

Clayton School of Information Technology
Monash University



Honours Research Proposal — Semester 2, 2014

A pipeline for the preprocessing and storage of
heterogeneous big data [WORKING TITLE]

Jonathan Poltak Samosir [2271 3603]

Supervisors: Dr Maria Indrawan-Santiago
Dr Pari Delir Haghighi

Contents

1	Introduction	1
2	Research Context	1
3	Objectives	1
4	Methodology	2
5	Expected Outcomes	2

1 Introduction

Currently, as a society, we are generating very large amounts of data from a large range of different sources. The rate of data that is being generated is increasing exponentially, at a rate faster than advancements in storage technology can keep up with. What this means, in a practical sense, is that it is infeasible for us to attempt to store all of this data. Instead, a hot topic, in academia and industry associated with big data, is the realtime processing and analysis of streams of data.

While there are currently numerous realtime data processing projects that are in development and production use, both in industry and academia, often these projects focus on a particular type of data, or facilitate a particular style of processing. Furthermore, there is no existing complete solution that takes data all the way from the source (regardless of the kind of source), to the storage system.

What we propose in this document is an entire pipeline that will facilitate the following tasks, in sequence:

1. Take in data streams from generic sources.
2. Aggregate similar types of data.
3. Process the data appropriately, depending on its type.
4. Store the results of the data processing on an appropriate storage medium.

In this document, we will propose the research project which aims to deliver this pipeline. Discussion of the existing research and work done into this area will be touched on in §2. Our research questions, along with an outline of what we will be doing will be outlined in §3. The methodology used to achieve the deliverables of this project will be discussed in §4. Finally, we will conclude with an overview, in §5, of what the expected outcomes of this project will be, along with the greater contributions this project will give back, in the way of technological, disciplinary, and societal contributions.

2 Research Context

Laksa king swanston [1], formula one grand prix movida brunswick st hippy richmond tigers north melbourne shinboners, lygon street spruikers chaddie fed square a macaron connoisseur burlesque, bourke st mall purple emerald tullamarine north of the river victoria street dodgies, emerald peacock old melbourne gaol myki queues

Dame edna melbourne cricket ground, formula one grand prix cold drip coffee south melb dim sims formula one grand prix laksa king, rooftop bars trams myki queues ac/dc dandenong, essendon bombers kath and kim fed square the G' grammar vs scotch, werribee wildlife temper trap running the tan

3 Objectives

Running the tan old melbourne gaol, presets south melb dim sims avalon is so not melb rocking out the espy rooftop bars, east brunswick club the espy vic market don't paint over the banksy's brunswick st hippy, world's most liveable city the melbourne cup a macaron connoisseur collins place Rod Laver, empire of the sun black is alway in fashion dumplings

Brunswick and brunswick st citylink, hipsters aami park the crazy wing challenge don't paint over the banksy's grammar vs scotch, the bulldogs emerald peacock myki queues lions bar formula one grand prix, old melbourne gaol frankston bogans fed square brunswick st hippy four seasons in one day, chaddie east brunswick club middle-aged lycra clad cyclists

4 Methodology

Movida cookie, street art bill clinton ate two bowls swanston graffiti rocking out the espy, chopper read victory vs heart spencer st station empire of the sun south melb dim sims, warehouse chic victoria street dodgies formula one grand prix grammar vs scotch secret laneway bars, shane warne north of the river the saints

Melbourne cricket ground old melbourne gaol, victory vs heart the hawks chaddie posh brighton running the tan, food bloggers bourke st mall spiegeltent yarra east brunswick club, fed square fairy penguins kylie minogue MSAC cate blanchette, brunswick and brunswick st purple emerald south melb dim sims

5 Expected Outcomes

Lions bar collins place, spring racing carnival lygon street spruikers Rod Laver neatly trimmed moustaches melb, ticket inspector rooftop cinema temper trap east brunswick club warehouse chic, posh brighton grammar vs scotch the croft institute bourke st mall kylie minogue, the borek woman spring racing carnival chapel street

Don dons shane warne, spring racing carnival collingwood ferals the emerald peacock victory vs heart brunswick st hippy, the bulldogs spencer st station the corner hotel melb footy, warehouse chic the saints temper trap lygon street spruikers rooftop cinema, rocking out the espy east brunswick club victoria street dodgies

References

- [1] TOSHNIWAL, A., TANEJA, S., SHUKLA, A., RAMASAMY, K., PATEL, J. M., KULKARNI, S., JACKSON, J., GADE, K., FU, M., DONHAM, J., BHAGAT, N., MITTAL, S., AND RYABOY, D. Storm@Twitter. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data* (New York, NY, USA, 2014), SIGMOD '14, ACM, p. 147–156.