

Spotting Black Swans With Ease: The Case for a Practical Reproducibility Platform

Ivo Jimenez and Carlos Maltzahn (UC Santa Cruz)

Abstract

We make the case for *Black Swan*, a platform for implementing, maintaining and curating experimentation pipelines.

1 Introduction

Reproducibility is the cornerstone of the scientific method. Yet, in computational and data science domains, a gap exists between current practices and the ideal of having every new scientific discovery be *easily* reproducible [1]. Advances in computer science (CS) and software engineering slowly and painfully make their way into these domains, even in (paradoxically) CS research [2,3].

Popper [4] is an experimentation protocol and CLI tool for implementing experimentation pipelines following a DevOps approach. The goal of Popper is to bring the same methods and tools used for the agile delivery of software, known as DevOps [5], to scientists and industry researchers.

Our experience in the past four years developing and evangelizing the use of Popper in multiple domains has allowed us to identify opportunities where open-source software (OSS) can be used to close the existing gap in current research practices. While the Popper protocol is relatively easy to follow (wrap experimentation pipelines in the form of a sequence of Bash scripts), for many practitioners it still represents a big leap between current practices and where OSS development communities are (automated, portable and

versioned pipelines). To this end, in this position paper we make the case for *Black Swan*¹, a platform for *practical* reproducible research. In a nutshell, Black Swan enables the agile delivery of software created in universities and other research institutions, significantly accelerating technology transfers between research and operation.

This position paper is organized as follows. Next section (Section 2) expands on the need for *Black Swan*, while section Section 3 presents the components of the proposed platform and how they address the current gaps. Section section 4 illustrates the utility of the platform by describing use cases where Black Swan would be used. We close with related work and how Black Swan is different to other existing software (section 5).

2 Motivation

To illustrate the “gap”

- popper pipelines.
- they’re easy but represent a big paradigm shift (and cognitive burden).
- we illustrate with an example.

We have anecdotally shown (in our lab and from testimonies from “converts”) how Popper is agile but for

¹Black swans are typically used to illustrate the concept of falsifiability: the statement “all swans are white” is proven false whenever one can spot one or more black swans. We envision *Black Swan* (the platform) to be a tool for researchers to *easily* find black swans in their computational or data science theories (i.e. identify when their claims are false) and, more importantly, allow them to investigate **why**.

Listing 1 Contents of the GassyFS experiment.

```
paper-repo/experiments/gassyfs
  README.md
  baseliner
    config.yml
  docker
    Dockerfile
    entrypoint.sh
  geni
    cloudlab_request.py
  results
    output.csv
    visualize.ipynb
  run.sh
  setup.sh
  validate.sh
```

personal use and teams. However, from the point of view of “outsiders”, it is still “too much work”. Thus, we need to make it even easier, where the baseline is “I will just give it 2 minutes and then I’ll give up”, which in practice means just having time to create accounts on services and having budget for a few clicks.

3 Black Swan

Key Components:

- Pipeline catalog
- Pipeline builder
- Environment capture

4 Use cases

We describe two use cases for which we envision Black Swan to be applicable. We envision many more but we see these as key use cases.

4.1 Technology Transfer

Organizations with *Research and Development* units such as tech companies and government-funded institutions (DOE labs, NASA, universities) spend significant amounts of resources transferring technology from R&D to production environments. Deploying an instance *Black Swan* internally, using a public one (or connecting a public with a private one), would allow organizations to streamline tech (and knowledge) transfers.

Users: Industry and university researchers; software engineers.

4.2 Research Curation

In the past decade, institutional libraries have invested a significant amount of resources to the creation of *Data Repositories*. These efforts are aimed at systematically managing the output of research. For example, EU’s OpenAIRE initiative is a catalog of software and data containing more than 80 million entries (each being a code repository or dataset). Similar efforts are underway in the US such as the Center for Open Science’s Open Science Framework. We see *Black Swan* complementing these efforts, since Popper enables the curation of research by enabling all these existing repositories to be easily executed and validated over time.

Users: Librarians; university and industry researchers.

5 Related Work

We present how it is different to existing software.

References

- [1] D.L. Donoho, A. Maleki, I.U. Rahman, M. Shahram, and V. Stodden, “Reproducible Research in Computational Harmonic Analysis,” *Computing in Science & Engineering*, vol. 11, Jan. 2009.

- [2] G. Fursin, “Collective Mind: Cleaning up the research and experimentation mess in computer engineering using crowdsourcing, big data and machine learning,” *arXiv:1308.2410 [cs, stat]*, Aug. 2013. Available at: <http://arxiv.org/abs/1308.2410>.
- [3] C. Collberg and T.A. Proebsting, “Repeatability in computer systems research,” *Communications of the ACM*, vol. 59, 2016, pp. 62–69.
- [4] I. Jimenez, M. Sevilla, N. Watkins, C. Maltzahn, J. Lofstead, K. Mohror, A. Arpac-Dusseau, and R. Arpac-Dusseau, “The popper convention: Making reproducible systems evaluation practical,” *Parallel and distributed processing symposium workshops (ipdpsw), 2017 ieee international*, IEEE, 2017, pp. 1561–1570.
- [5] G. Kim, J. Humble, P. Debois, and J. Willis, *The DevOps Handbook*, O’Reilly Media, 2016.