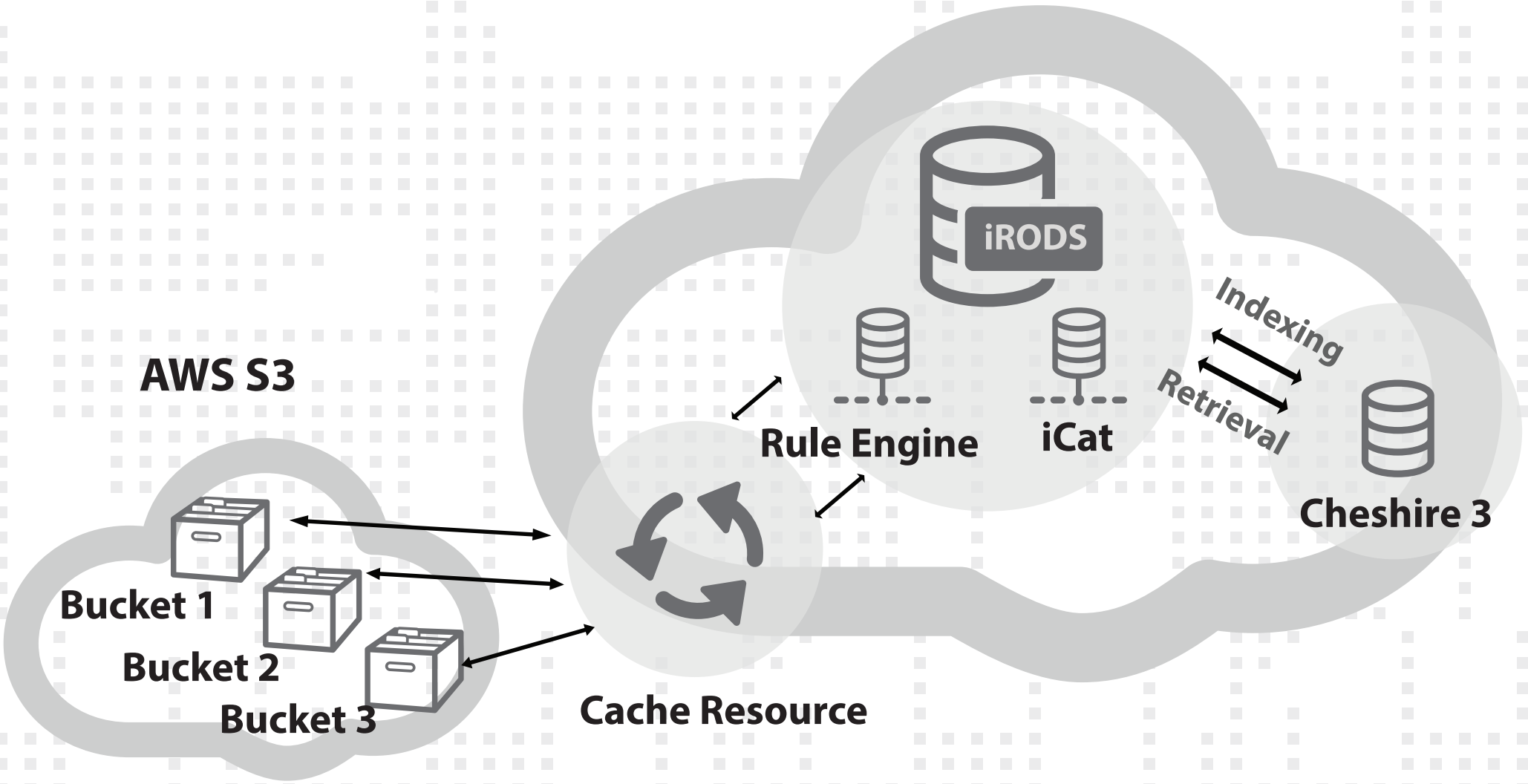
LUIS AGUILAR | SHREYAS
under Prof. Ray Larson

PROJECT POSTER

Digging Into Data

Integrating Data Mining and Data Management Technologies for Scholarly Inquiry



GOALS

Text mining and **NLP techniques** to extract content (named **Persons, Places, Time Periods/Events**) and **associate context**

DATA

- **Internet Archive Books** ~7.2T
- **Jstore** ~1T
- **Context sources**: SNAC
- **Archival and Library Authority records**.

TOOLS

- **Cheshire 3**
  (DL Search and Retrieval Framework)
- **iRODS**
  Policy-driven distributed data storage
- **Python NLTK NER**
- **Amazon S3 storage**
- **EC2 computing**

CONTACTS

Prof. Ray Larson <ray@ischool.berkeley.edu>
Luis Aguilar <luis@ischool.berkeley.edu>
Shreyas <shreyas@ischool.berkeley.edu>