Pipes vs Dams:

Collecting, Processing and Analyzing Networks

Bashar Ahmad

RISC Software GmbH

Softwarepark 35, 4232 Hagenberg, Austria

Content

- Introduction
- System architecture
- Current system
  - Data traffic
  - Sink Daemon
  - Collectors
  - LiveData
- Example
- Future Work

Introduction

- PIPES VS DAMS
  - Privacy Preserving Visual Dynamic Network Analysis for Advance Monitoring on Multiple Scale.
- Partners:
  - JKU – Institute of Computer Graphics
  - RISC Software GmbH
  - MOWIS GmBH
  - KT-Net Communications GmbH
- Project (840232) is funded by The Austrian Research Promotion Agency (FFG)
  - https://www.ffg.at

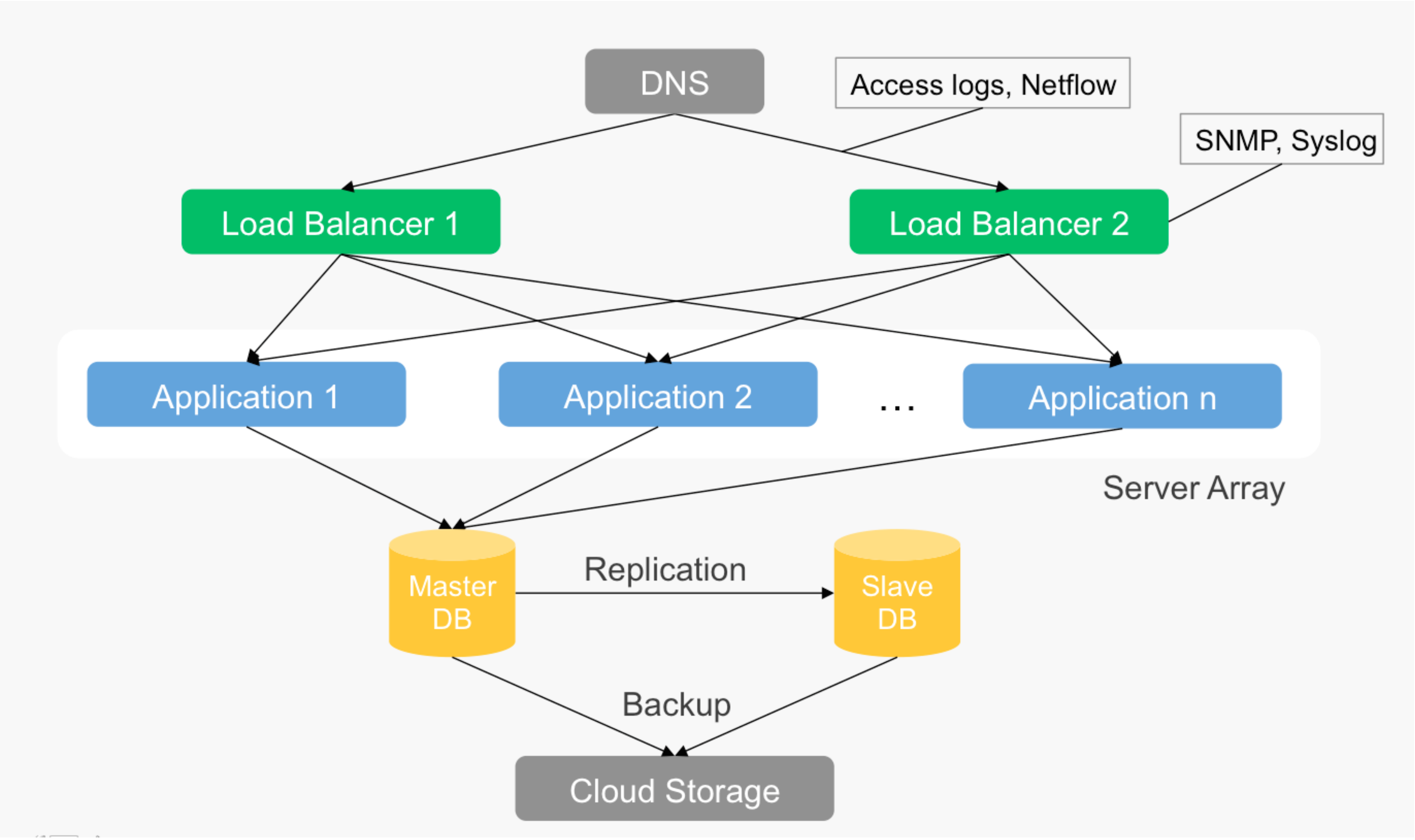
Introduction

- The project aims at automating the generation of a coherent and privacy-preserving overview of network infrastructures, allowing users to visually analyze the continuously increasing amount of data generated by the huge number of system components in complex infrastructures.

Introduction

- Current monitoring tools:
  - Provides an overview of current status of single system components (e.g. CPU load, memory usage, bandwidth).
  - Some provide historical overview and development of trends.
  - And less provide insights to possible future events.

Introduction

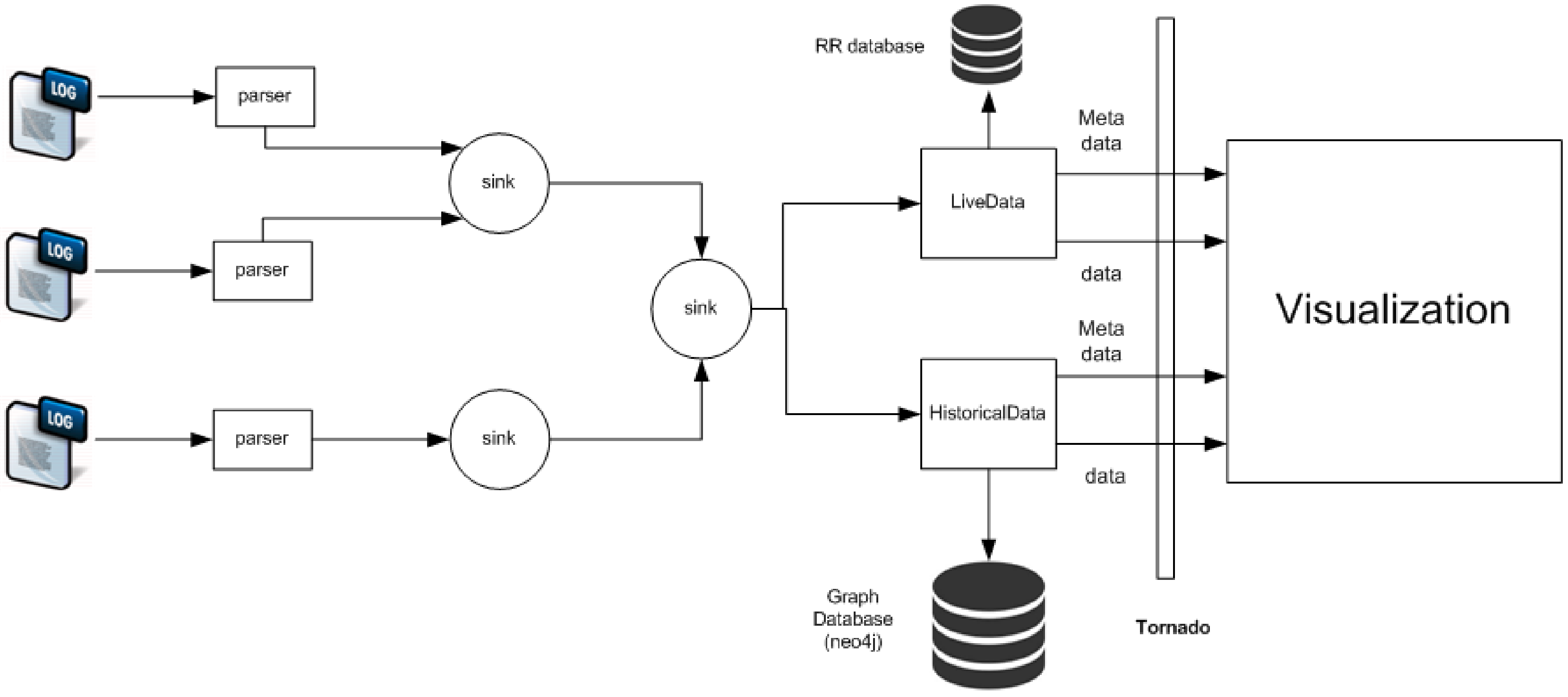


Introduction

- The project aims at building new approaches to analyze, monitor and simulate complex distributed systems that can be found on multiple scales (web platform providers, via cloud infrastructure providers, Internet service providers, etc.).

- This should be achieved by applying techniques from visual analytics, **data mining** and **dynamic network analysis**.

System Architecture



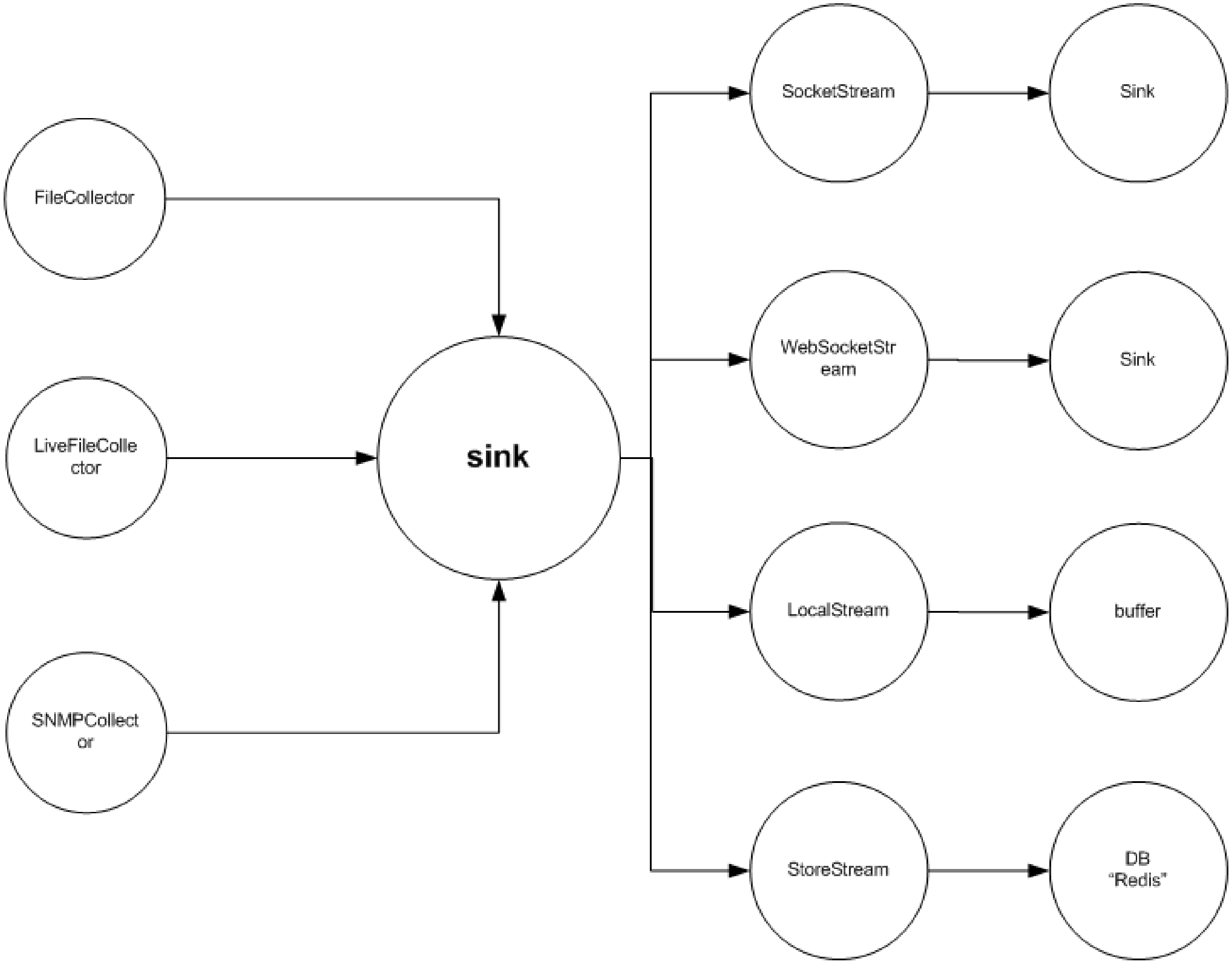
System Architecture

- Collector:
  - Collect and parse logs
  - Keep track of all parsed logs.
  - Discover new logs.
- Sink:
  - Stream parsed logs.
    - Parser to Sink
    - Sink to Sink
    - Sink to LiveData or HistoricalData
  - Applying early processing and filtering algorithms.

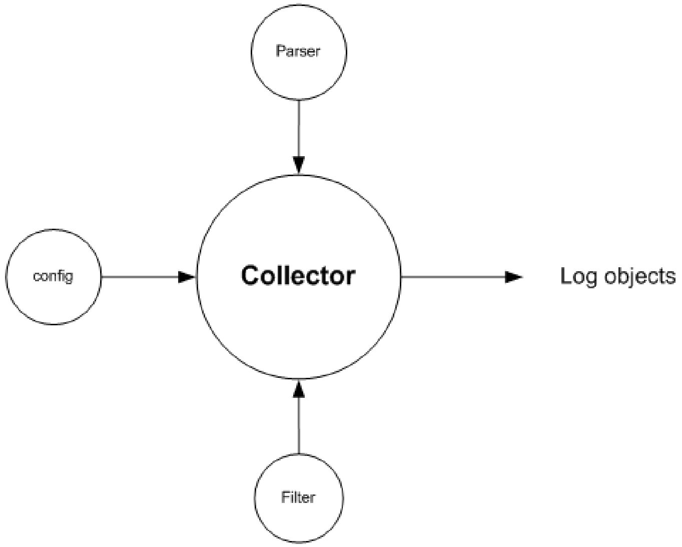
System Architecture

- Live Data interface
  - A central point where all live-data stream gathered.
  - Use NOSQL Database to buffer live stream.
  - Respond to live-data inquiries.
  - Offers a meta data of the live data.
- Historical-Data interface
  - Receives all live-data streams
  - Analyze and aggregates live-data and build historical data.
  - Store the historical data in graph database.
  - Respond to historical data inquiries.
  - Offers meta data of the historical data.
- Tornado interface
  - Links HD and LD with the visualization software.

Current System - Data Traffic



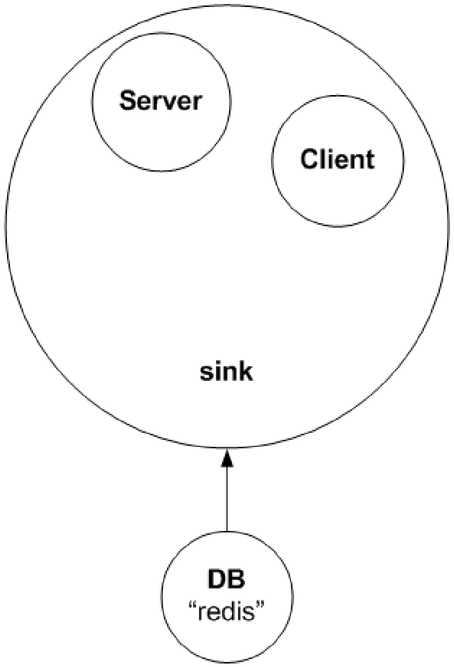
Current System - Collector



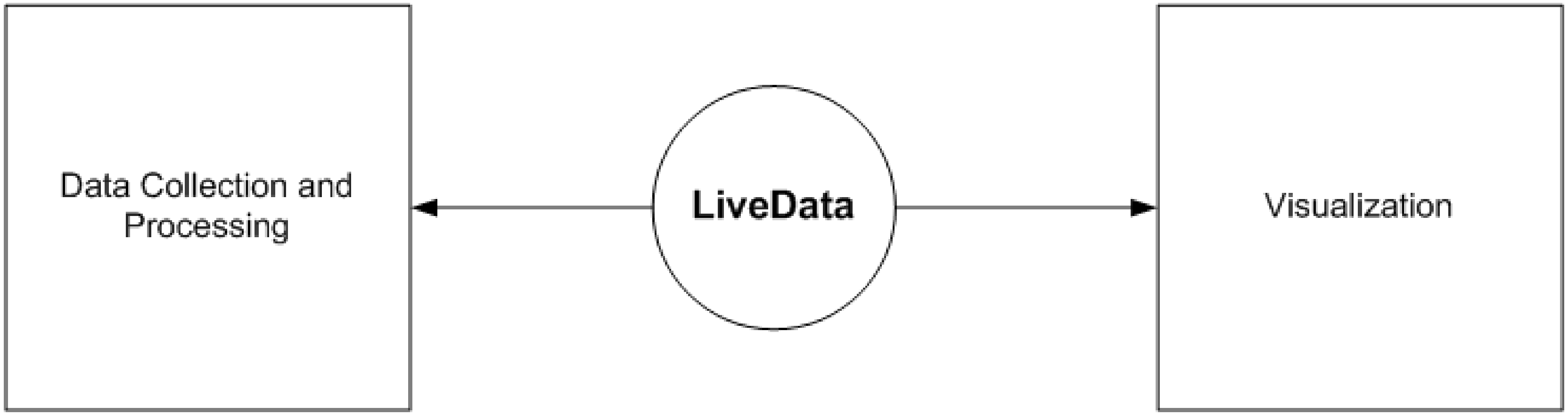
- **Collector**
  - **FileCollector**: is used to collect logs from Archived files.
  - **LiveFileCollector** : is used to collect logs from current log file.
  - **SNMPCollector**: is used to collect logs from a SNMP server (up to V3).

Current System - Sink

- Sink:
  - SinkServer: is used to received logs from other sink instances.
    - Socket
    - WebSocket
  - SinkClient: is used to forward logs to other sinks instances.
  - StoreStream: is used to store logs to a DB.

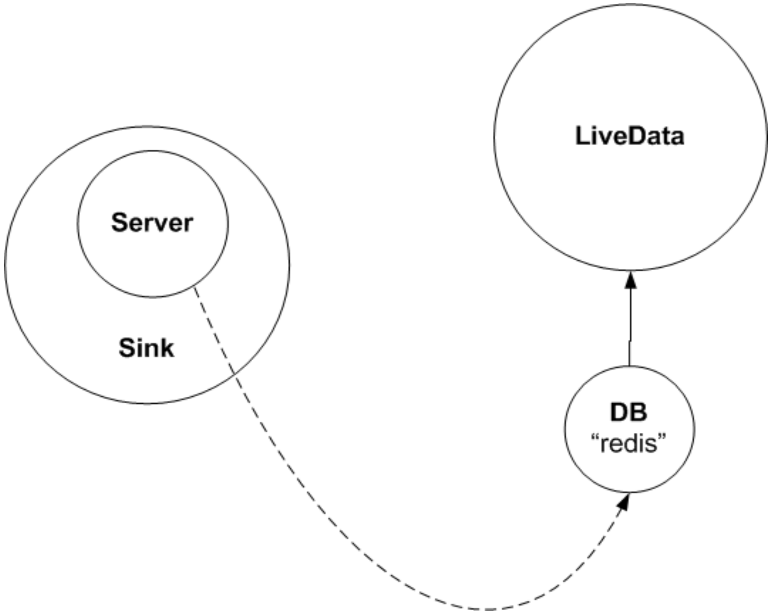


Current System - LiveData



Current System - LiveData

- **LiveData:**
  - Query live collected logs.
  - Query infrastructure meta-data.

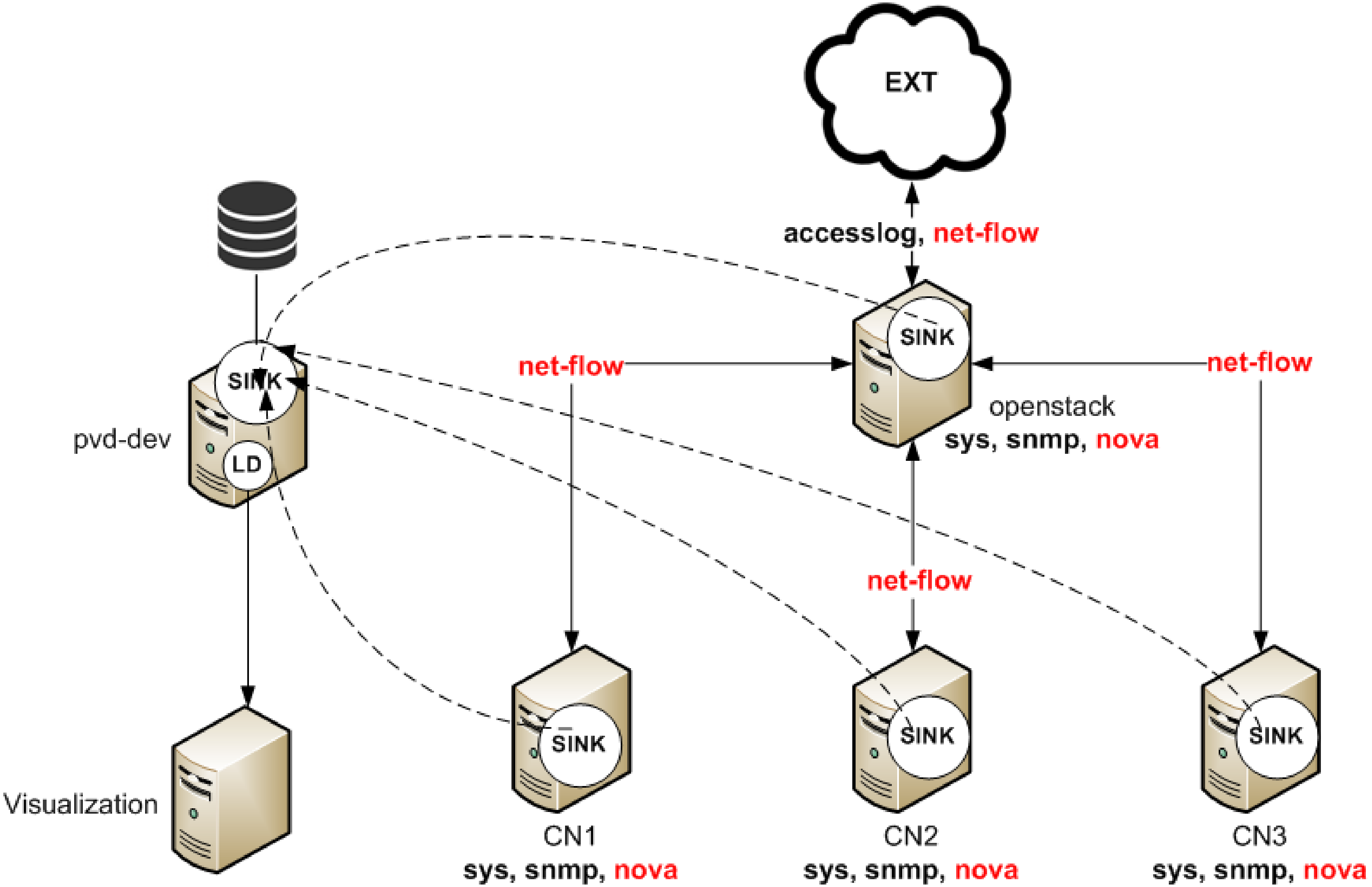


- Meta Data
  - Infrastructure (source)
  - Stream Objects (Wiki)

Current System

- Components:
  - Collector
  - Sink
  - DB Controller
  - Live Data interface
  - Log parser
  - Objects builder (JSON).
  - Stream (Websocket, Socket, File, DBStream)
- Python 2.7
- Component based (modular)
- Configurable

Example


EXT
accesslog, net-flow
SINK
openstack
sys, snmp, nova
net-flow
net-flow
SINK
LD
pvd-dev
net-flow
Visualization
SINK
CN1
sys, snmp, nova
SINK
CN2
sys, snmp, nova
SINK
CN3
sys, snmp, nova

Future Work

- New Collector (Net flow, Tomcat, ..etc)
- Historical Data Aggregation
- Historical Data Interface
- Prediction
- Privacy preservation (traffic anonymization) & access control – Master thesis
- Automatic dependency detection – Master thesis

Castor, 4228m

Pollux, 4092m

zwischen
Monte-Rosa-Massiv
und Matterhorn

Wallis, Schweiz

RISC
Software GmbH

www.risc-software.at