I. OVERVIEW OF KEY ELEMENTS IN BIG DATA SYSTEMS

The primary objective of this section is to delineate the crucial constituents of big data systems. The components under consideration are derived from the systematic literature review (SLR) as referenced by [1] and [2], complemented by a brief literary review of the existing knowledge.

Addressing the second research question (RQ2), the architectural references (ARs) presented in Table 1 were assessed and juxtaposed to pinpoint recurrent elements in big data architectural references. While some ARs, being brief, offered minimal insights, others, like NIST, were notably exhaustive. A prevailing observation was the tendency of ARs to be rooted in existing knowledge, substantiating the idea that crafting ARs from a foundation of pre-existing knowledge can be more potent than starting afresh.

In a systematic pursuit of this inquiry and subsequent to our data collation, we enumerated all the elements from every big data AR discussed prior. These elements find their description in Table 2. One challenge was the variation in terminologies used across different studies to delineate architectural components. While architectural definition languages, like Archimate, find limited adoption, most researchers resort to bespoke ontologies using simple diagrams. This renders comparison challenging, often necessitating translation from one conceptual framework to another.

An automated textual assessment was initially conducted on the naming of these components, aiming to discern trends and word preferences. The graphical representation of these terminologies is illustrated in Figure 3. Notably, terms such as 'big data application provider' and 'big data framework provider' emerged as common nomenclatures, possibly attributed to the influence of NIST BD AR (s17). Universally accepted terms included 'data consumer' and 'data provider', with 'layer' being the chosen descriptor for logically segmenting various AR components.

Distilling the above, to aptly address RQ2, we meticulously analyzed these components and sorted them based on functionality into: 1) Big Data Management and Storage, 2) Data Processing and Interface Applications, 3) Big Data Infrastructure.

REFERENCES

[1] P. Ataei and A. Litchfield, "The State of Big Data Reference Architectures: A Systematic Literature Review," in IEEE Access, vol. 10, pp. 113789-113807, 2022, doi: 10.1109/ACCESS.2022.3217557.

[2] Ataei, Pouya and Litchfield, Alan T., "Big Data Reference Architectures, a systematic literature review" (2020). ACIS 2020 Proceedings. 30.