Architecture for Decision Support in Precision Livestock Farming

Jonas S. Gomes[1], José Maria N. David[1], Regina Braga[1], Victor Ströele[1], Wagner Arbex[2], Bryan Barbosa[1], Wneiton Luiz Gomes[2], Leonardo M. Gravina Fonseca[2]

[1] Departamento de Ciência da Computação, Universidade Federal de Juiz de Fora (UFJF), Juiz de Fora, MG – Brazil

[2] Empresa Brasileira de Pesquisa Agropecuária – Embrapa, Juiz de Fora, MG – Brazil

jonas.gomes@estudante.ufjf.br, {jose.david, regina.braga}@ufjf.edu.br, {victor.stroele, bryan.barbosa}@ice.ufjf.br, {wagner.arbex, wneiton.gomes, leonardo.gravina}@embrapa.br

Abstract. *The use of sensors in the agricultural domain generates a massive volume of heterogeneous data that must be treated, stored, and processed for decision-making. These decisions must be taken considering the diversity of devices and contextual information, which is often not considered but is important to the decision-making process. This paper presents an architecture to integrate data from sensors related to precision livestock farms. The integration and processing of these data can support decision-making, lead to more accurate results and enhance agribusiness sustainability.*

1. Introduction

Agricultural farmers are increasingly using sensors to monitor food production, animal health, and welfare (Zhai et al. 2020). These sensors generate a large volume of heterogeneous data that needs to be treated, stored, and analyzed by intelligent applications. Based on this data, farmers can extract information and make strategic decisions.

In Agriculture 4.0, more specifically in livestock, sensors (Internet of Things devices, i.e., IoT) and data processing are called precision digital livestock (Zhai et al. 2020). Precision digital livestock currently has challenges due to the complexity of managing data heterogeneity and the complexity of its relationships. Moreover, it includes issues associated with the environment, sustainability, economy, coexistence with other land uses, and climate change (Bahlo et al., 2019). These challenges create a problem for decision-making since the number of IoT devices has increased considerably (Zhai et al. 2020), making it difficult to process and combine data. In livestock, decisions usually need to be made considering the great diversity of these data and devices. As a result, information from the data context[1], such as external temperature and humidity, rainfall data, and meteorological data for weather forecasting, to name a few, are often not used in the decision process.

[1] Context can be defined as a complex description of shared knowledge about physical, social, historical, or other circumstances in which an action or event occurs (Rittenbruch, 2002).

Combining technologies for precision livestock and precision agriculture can help farmers and managers monitor crops and animals in real-time. Besides, they can maximize productivity and profitability in agricultural operations, even improving the worker's quality of life. In the literature, some works propose solutions to this problem. Roukh et al. (2020) solve the problem of collecting, processing, storing, and visualizing Big Data in agriculture. However, the authors do not explore external data, such as contextual data. The decisions are made based only on data that are obtained by sensors. Xuan and Nhat (2019) present an architecture that uses external data for decision-making, but the architecture was designed to solve specific greenhouse problems and cannot be easily used in other domains.

This paper presents the e-Livestock architecture for collecting and processing heterogeneous data from sensors in real-time, also considering context information. The proposal is to specify an architecture that can be used in different agribusiness subdomains. The architecture was also designed to process the data traceability, considering the capture of provenance information[2]. The use of provenance data is due to external information can be added to the sensor data, and, thus, the architecture must be able to trace the origin of the information used in decision-making, reinforcing the data reliability.

This work is organized according to the following structure. Section 2 presents the background and related works. Section 3 presents the proposed architecture, as well as the evaluation. In Section 4, final remarks and future work are discussed.

2. Background and Related Work

The concept of Internet of Things (IoT) is related to the connection of a network of "objects" through the Internet without direct human intervention. The application of IoT in agriculture and livestock has advantages due to the possibility of monitoring and controlling many different parameters in an interoperable, scalable, and open context with the increasing use of automated sensors (Villa-Henriksen et al., 2020). In agribusiness, one of the innovations is its combination with data processing, allowing greater support in decision-making. One of the objectives of using IoT in livestock is to bring greater accuracy in the use of information for decision-making. According to Sprague (1980), a Decision Support System (DSS) combines models with data access and retrieval functions. A DSS is an application that supports different decision-making activities in a given domain (Belciug and Gorunescu, 2020).

Provenance provides a critical basis for assessing authenticity, trust, and reproducibility of decisions (Buneman et al., 2001). Provenance tracking allows data to be shared, discovered, and reused, simplifying collaborative activities (Ram and Liu, 2007). Currently, the standard model for specifying data provenance is the PROV (Belhajjame et al., 2013). In the era of Big Data, information integration often needs to go through the extraction and loading of large volumes of data and diverse sources. This distributed data needs to be collected by appropriate equipment or software, and data storage management models must be provided for this data to be processed (Wang et al., 2020).

[2] provenance data (Buneman et al., 2001) refer to "a type of contextual element that describes information about the origin of the data and its derivations".

Xuan and Nhat (2019) discuss how context data, such as temperature and humidity outside the sensor environment, can impact production and farming and contribute to an accurate prediction of future temperatures. However, the authors focused on the solution for forecasting temperatures in the greenhouse. Our research aims to add, in addition to real-time meteorological data, other sources, such as external databases, social networks, and weather forecasting services. Fote et al. (2020) present an architecture using Apache Kafka[3] to support the collection and Apache Storm for processing body temperature, vital data, movement patterns of livestock for animal welfare. However, the proposed architecture does not allow the integration of external data to enrich the context for decision-making. Roukh et al. (2020) propose an architecture for offline processing of multiple data sources, such as text files, spreadsheets, web services, and in real-time considering sensor data. However, the provenance of this data is not explored to enrich decision-making.The works discussed above present architectures related to specific contexts and cannot be generalized to other environments. Thus, to meet specific needs not discussed in the literature, the e-Livestock architecture was proposed to deal with agricultural data diversity and integrate context and provenance data to support decision-making.

3. e-Livestock Architecture

The architecture is composed of five main layers, as shown in Figure 1. ***Sensor Tier:*** This layer is responsible for monitoring all sensors used in the physical space, for example, devices connected to the animal body or in the environment (context data). This layer was designed to deal with different types of sensors responsible for monitoring different environments. The data collected includes information about the environment, the animal, and the sensor identification. ***Platform Tier:*** This layer is responsible for data collection and processing, and integration with other data sources. Heterogeneous data can come from different sources. Considering the collection of streaming data, i.e., a large volume of data produced by the sensors, we use Apache Kafka. The Kafka Connector makes an interface for these data integration and Apache Flink handles the processing (Akil et al. 2017). ***Integration Tier:*** This layer is responsible for receiving and integrating the processed data. The integration layer can also aggregate information from other databases, services, and external APIs, such as weather data. The main advantage of this layer is to support the integration and storage of context metadata. Figure 2 presents the data model of our solution. ***External Sources Tier.*** This layer represents external services, databases, historical bases, social networks, or other external data sources. As needed, new fonts can be easily coupled to the architecture through the integration layer. When integrating social network data, it is possible to provide a new perspective for decision-making. This layer also collects and store provenance data. ***Visualization Tier:*** The visualization layer allows the farmer to view the data in real-time, according to a time interval, using a dashboard. It is possible to analyze and interpret the data at different granularities.

---

[3] https://kafka.apache.org/

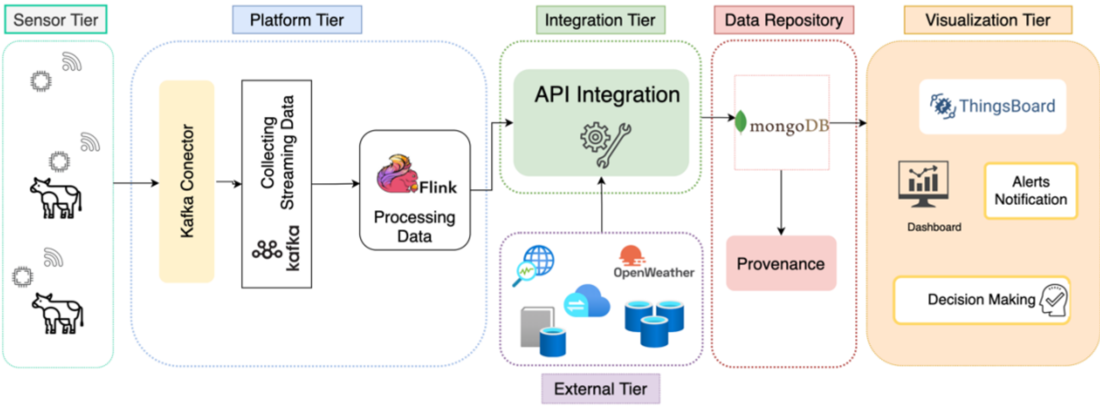

Figure 1 – e-Livestock Architecture Overview[4]

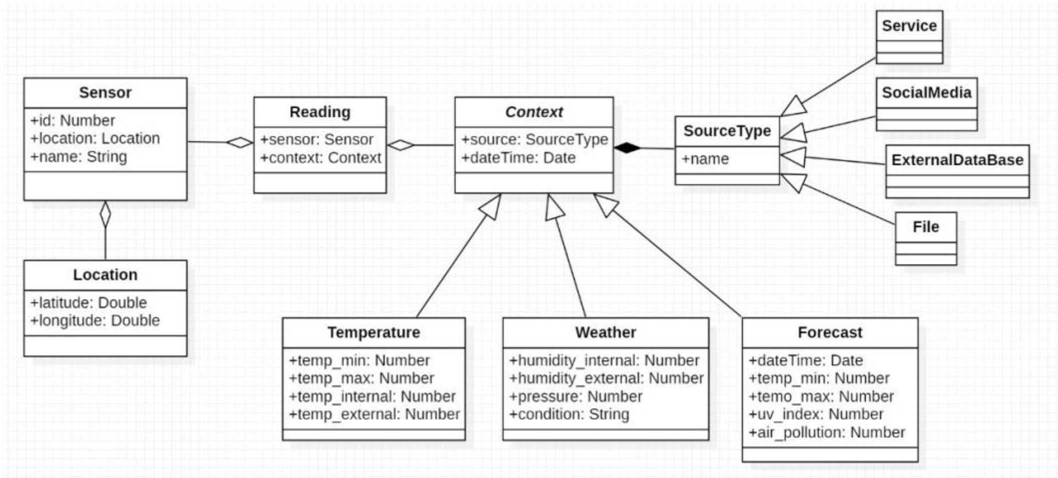

Figure 2. Sensors Data Model

3.1. Data Provenance

To capture the data provenance, we used the PROV model as illustrated in Figure 3. This model can be extended based on specific domain needs. In our work, entities represent mainly the animals, agents are sensors that monitor it, and activities represent smart farm activities such as animal weighing and the processing of the collected data. Considering provenance data, it is possible to identify the data source, the interactions that the farmers and users carry out, and, consequently, we can trace the decisions.

---

[4] Apache Kafka was used to collect heterogeneous data (Jafarpour et al. 2019). Apache Flink is an open-source system for processing streaming and batch data (Carbone et al. 2015). Flink was adopted to implement the architecture. For visualization, ThingsBoard[4] was used, which is an open-source IoT platform capable of managing devices and presenting graphics. MongoDB was used for the data storage.

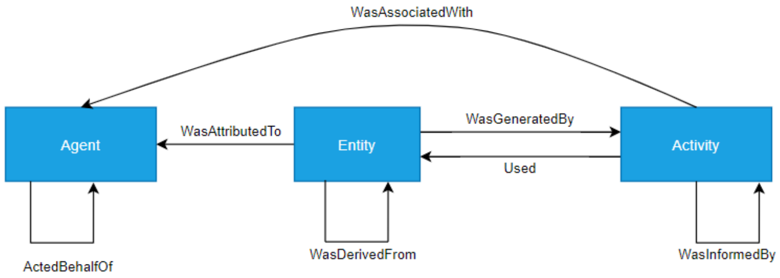

Figure 3. PROV Model[5]

According to the model presented in Figure 2, the sensor data are persisted in MongoDB, including context data. An example of context data can be seen in Figure 4. It includes external temperature, humidity, and pressure.

id	lat	lon	date	temperature	pressure	humidity	clouds	wind_speed
1	51.50853	-0.12574	2021-01-01 00:00:00	8.08	1034.848	78.61	0	2.959
2	51.50853	-0.12574	2021-01-01 03:00:00	6.34	1034.323	88.804	66	3.277
3	51.50853	-0.12574	2021-01-01 06:00:00	5.77	1034.504	89.5	33	3.575
4	51.50853	-0.12574	2021-01-01 09:00:00	6.02	1034.832	89.7	4	3.805
5	51.50853	-0.12574	2021-01-01 12:00:00	9.26	1035.297	82	2	4.254

Figure 4. Context Data

3.2. Evaluation

In order to verify the feasibility of the proposed architecture, we defined the scope based on GQM (from English Goal, Questions, Metrics): "**Analyze** the use of the architecture **from the point of view** of researchers for decision-making, **in the context of** the Compost Barn production system that focuses on livestock. The following question was defined: **RQ. How can the e-Livestock architecture support the decision-making process on farms, combining the diversity of data generated considering context and provenance data?**

The evaluation was carried out based on the collection and processing of data from a production system called Compost Barn (Embrapa Gado de Leite, 2020), located at EMBRAPA - Gado de Leite, experimental field. This system is part of a research project with Brazilian and international institutions related to improving the dairy cattle production system. Compost Barn aims to reduce maintenance costs of milk production, improve production and health of herds and enable the correct use of organic waste (feces and urine) from dairy activities. This space has sensors to monitor the temperature and humidity of the environment and lighting sensors. Continuous monitoring allows adjustments to the animals' living conditions, enabling increased production and quality of life. We used a set of data obtained from this real context. This data set contains the temperature and humidity inputs of the Compost Barn.

The Compost Barn production system has internal measuring equipment installed in the building and exhaust fans to control the temperature. The ideal is that the

[5] https://www.w3.org/TR/prov-dm/

internal temperature is 5 degrees less than the external temperature of the environment. As the temperature increases, more hoods are turned on to cool the environment. If a temperature exceeds the limit of 34º C, the system can communicate through the integration layer with external services and trigger an audible alarm. Figure 5 presents a set of internal temperature data from the environment, captured by the farms' sensors and used to verify the architecture's viability. The colors indicate a heat map, with colder blue, normal green, medium yellow, and high orange.

Dia	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Fazenda Araruana	17,2	13	12,4	11,9	12	11,8	12	12,2	14,1	12,6	14,1	15,5	16,6	16,1	16,8
Fazenda Esperança	17,6	12,7	12	11	11,1	11,3	11,5	12,2	12,6	12,8	12,8	15,1	16,2	15,8	16,1
Fazenda Maringá	16,5	12,3	12,6	10,6	10,9	11	11,1	12	12,3	12,9	12,4	14,7	15,7	15,2	15,9

Figure 5. Farm Dataset

This data is processed and sent to be integrated with external data sources and persisted in a database. The sensor data integrated with environmental context data are presented to the farmers through graphs and alert notifications, helping in the decision-making. Also, they have an overview of the environment being monitored. Through the ThingsBoard interface, as shown in Figure 6, it was possible to trigger alarms based on rules. For example, device "A" performs a temperature reading of 24º C that exceeds the defined limit. As a result, a "High temperature" alert is generated. Each alarm has a severity that can be defined as Critical, Main, Secondary, Warning, or Indeterminate (ranked by priority in descending order). Users can also receive alert notifications via SMS and email.

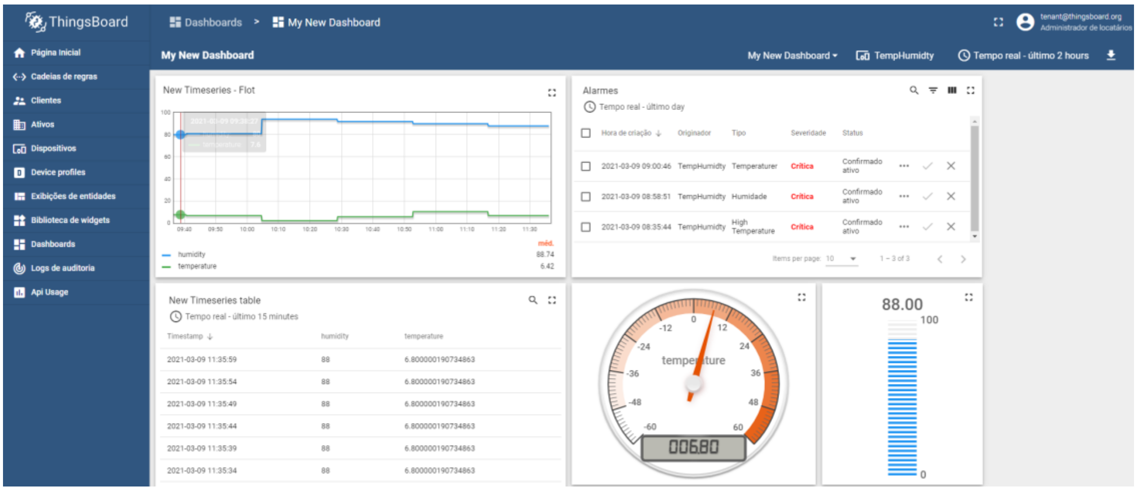

Figure 6. ThingsBoard with Evaluation Data

To collect the external temperature/humidity, the INMET website was used at the Compost Barn location. By analyzing context data, such as external temperature and climate forecast, farmers were able to make more sophisticated adjustments to the temperature, automating the process of starting exhaust fans. As a result, it was possible to avoid sudden changes that could affect the animals' production. Therefore, considering our research question (RQ), we have evidence that the e-Livestock architecture could meet the objective of processing Compost Barn data satisfactorily. Context data was integrated, following the model defined in MongoDB and provenance data, considering the PROV model. From the capture of provenance data, it was possible to track the decisions made and mitigate the alerts. Besides, decision-makers can be identified by providing a more transparent and reliable system. However, the results obtained cannot be generalized and additional evaluation should be conducted later.

4. Final Remarks

Decisions in agriculture need to be made considering the diversity of information and devices present in different contexts. Furthermore, context information is often not used in the decision-making process due to the complexity of managing a high volume of heterogeneous data. This work presents an architecture that aims to tackle the problems of collecting, processing, and visualizing data in real-time to support decision-making. Still, it was possible to support decisions with external information and data from other sources. By capturing the provenance data, we can mitigate events and track decisions made in the environment. Through the evaluation, it was possible to verify the architecture's feasibility considering context and provenance data applied in real rural environments.

For future work, we intend to integrate other types of data sources such as social networks and specify the provenance model for the agricultural domain. Also, to generate several e-Livestock architecture instances to create an ecosystem for agriculture, exploring aspects of collaboration, communication, and integration between farms to support the decision. Moreover, the combination and processing of additional data sources and sensors can lead to more accurate results, reduce costs, and maintain agribusiness sustainability.

ACKNOWLEDGMENTS

We would like to thank the researchers who participated in the evaluation and express thanks to the National Research Center of Dairy Cattle (Embrapa Dairy Cattle) of Brazilian Agricultural Research Corporation (Embrapa). This work was partially funded by UFJF/Brazil, CNPq/Brazil (grant 311595/2019-7), and FAPEMIG/Brazil (grant APQ-02685-17), (grant APQ-02194-18). This study was also financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

References

Akil, B., Zhou, Y. and Röhm, U., (2017) “On the Usability of Hadoop MapReduce, Apache Spark & Apache Flink for Data Science”, IEEE International Conference on Big Data (Big Data), p. 303-310.

Bahlo, C., Dahlhausac P., Thompsonac, H. and Trotterbc, M. (2019) “The role of interoperable data standards in precision livestock farming in extensive livestock systems: A review”, Computers and Electronics in Agriculture, p. 459-466.

Belciug, S. and Gorunescu, F. (2020) “Intelligent decision support systems–a journey to smarter healthcare”, Springer International Publishing.

Belhajjame, K., Cheney, J., Coppens, S., Cresswell, S., Gil, Y., Groth, P., Klyne, G., Lebo, T., McCusker, J., Miles, S., Myers J., Sahoo, S., Tilmes, C., Moreau, L., and Missier, P. (2013), “PROV-DM: The PROV Data Model”, W3C Recommendation REC-prov-dm-20130430, World Wide Web Consortium.

Buneman, P., Khanna, S. and Tan, C. (2001), "Why and where: A characterization of data provenance", 8th International Conference on Database Theory, p. 4-6.

Carbone, P., Katsifodimos, A., Ewen S., Markl, V., Haridi, S. and Tzoumas, K. (2015), "Apache Flink: Stream and Batch Processing in a Single Engine", Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, p. 28-38.

Embrapa Gado de Leite, (2020) "Brasil tem a primeira instalação de compost barn destinada à pesquisa – 2020". Available in: <https://www.embrapa.br/busca-de-noticias/-/noticia/53360675/brasil-tem-a-primeira-instalacao-de-compost-barn-destinada-a-pesquisa>. Accessed in: 03 mar 2021.

Fote, F., Mahmoudi, S., Roukh, A. and Mahmoudi, S. (2020), "Big Data Storage and Analysis for Smart Farming", 5th International Conference on Cloud Computing and Artificial Intelligence: Technologies and Applications (CloudTech), p. 1-8.

Jafarpour, H., Desai, R. and Guy, D. (2019) "KSQL: Streaming SQL Engine for Apache Kafka", International Conference on Extending Database Technology (EDBT), p. 524-533.

Ram, S. and LIU, J. (2007), "Understanding the Semantics of Data Provenance to Support Active Conceptual Modeling", Active Conceptual Modeling of Learning. Chen, P.P., Wong, L.Y. (eds.), p. 17-29.

Rittenbruch, M. (2002), "ATMOSPHERE: A Framework for Contextual Awareness", International Journal of Human-Computer Interaction 14 (2), 2002, pp. 159-180.

Roukh A., Fote F., Mahmoudi S. A. and Mahmoudi S. (2020), "Big Data Processing Architecture for Smart Farming", Procedia Computer Science, p. 78-85.

Sprague, R. (1980), "A framework for the development of decision support systems", MIS Quarterly 4 (4^{th} edition), p. 1–15.

Villa-Henriksen A., Edwards G., Pesonenc, L., Greenbd O. and Sørensena, C. (2020), "Internet of things in arable farming: Implementation, applications, challenges and potential", Biosystems Engineering, p. 60-84.

Wang, J., Yang, Y., Wang, T., Sherratt R. and Zhang J. (2020), "Big data service architecture: a survey", Journal of Internet Technology, p.393-405.

Xuan, S. and Nhat, L. (2019), "A dynamic model for temperature prediction in glass greenhouse", 6th NAFOSTED Conference on Information and Computer Science (NICS), p. 274-278.

Zhai Z., Martínez J., Beltran, V. and Martínez, N. (2020), "Decision support systems for agriculture 4.0: Survey and challenges", Computers and Electronics in Agriculture.