

This is the **published version** of the bachelor thesis:

Laura Rodriguez, Yelennis Brissey. *PigLF : Creació i gestió de bases de dades per animals*. Treball de Final de Grau (Universitat Autònoma de Barcelona), 2026 (Enginyeria Informàtica)

This version is available at <https://ddd.uab.cat/record/326563>

under the terms of the  license.

PigLF: Creació i gestió de bases de dades per a animals

Yelennis Brissey Laura Rodriguez

Resum— En aquest projecte s'ha dut a terme el desenvolupament d'una base de dades per a la recollida i la gestió de dades procedents de sensors en corrals de porcs durant el seu període d'engreix. El procés consisteix en la neteja i preparació de les dades obtingudes dels sensors, la creació de col·leccions específiques per tipus de sensor i la inserció de les dades prèviament sanejades. Així mateix, inclou la definició de consultes per centralitzar la informació i realitzar càlculs específics necessaris per als estudis dels investigadors, garantint la coherència i reutilització de la informació. El sistema es planteja inicialment en un entorn local i es proposa una solució escalable que permeti l'accés simultani de diversos investigadors. Els resultats obtinguts mostren que la base de dades és estructurada i útil per a futurs estudis científics.

Paraules clau— Base de dades, Sensors (Copeeks, iPERL, Autofom, Nedap, SoundTalks, Agritech), Engreix, Benestar animal, Condicions ambientals, Python, MongoDB, Consultes i agregacions.

Abstract— This project involved the development of a database for the collection and management of data from sensors in pig pens during the fattening period. The process included cleaning and preparing the sensor data, creating sensor-specific collections, and inserting the pre-processed data. It also involved defining queries to centralize the information and perform specific calculations required for the researchers' studies, ensuring data consistency and reusability. The system was initially designed for a local environment and proposes a scalable solution that allows simultaneous access for multiple researchers. The results demonstrate that the database is structured and suitable for future scientific studies.

Index Terms— Database, Sensors (Copeeks, iPERL, Autofom, Nedap, SoundTalks, Agritech), Fattening, Animal welfare, Environmental conditions, Python, MongoDB, Queries and aggregations.

1 INTRODUCCIÓ - CONTEXT DEL TREBALL

Els investigadors de la Facultat de Veterinària han estat recollint dades sobre creixement, benestar i condicions ambientals dels animals mitjançant diversos sensors instal·lats als corrals de porcs. Tanmateix, la gestió d'aquesta informació es realitzava de manera manual, amb descàrregues de fitxers i enviaments periòdics a través de serveis com WeTransfer. Aquest procediment era lent, propens a errors i dificultava la centralització i reutilització de la informació per a estudis científics i informes tècnics. Per això, aquest projecte té com a objectiu desenvolupar una base de dades específica per als sensors utilitzats en corrals de porcs, que permeti centralitzar la informació, realitzar consultes i càlculs específics, i facilitar l'accés simultani a diversos investigadors, establint així una eina fiable, eficient i reutilitzable per a futurs estudis científics.

2 ESTAT DE L'ART

Altres equips i institucions han desenvolupat solucions per a la recollida i gestió de dades del sector porcí, centrades principalment en el seguiment individual dels animals o en dades específiques de reproducció i producció durant les primeres etapes del cicle productiu. Una de les plataformes més utilitzades a Espanya és BDporc [1], gestionada per l'IRTA [2]. Aquesta base de dades registra informació sobre altes i baixes, parts, cobriments, avortaments i altres indicadors de reproducció i població durant la lactància i

el deslletament, permetent calcular diversos índexs de gestió i seguiment individual.

Encara que aquesta informació és molt completa en aquestes etapes inicials, BDporc no integra dades ambientals ni informació sobre el consum d'aliment i aigua ni sobre el comportament dels animals durant l'etapa final d'engreix, quan els animals passen de 20-30 kg a 100-120 kg de pes. Aquesta etapa és clau, ja que el benestar i el comportament dels animals durant l'engreix impacten directament en la seva productivitat i salut final.

En aquest context, el present projecte cobreix aquest buit, oferint una base de dades centralitzada i estructurada que recopila informació dels sensors ambientals, de consum i de comportament durant l'engreix, permetent consultes i agregacions automàtiques i facilitant l'accés simultani a diversos investigadors.

3. METODOLOGIA

El projecte s'ha estructurat en dues fases principals per tal de desenvolupar una eina completa que permeti gestionar les dades recollides dels sensors de manera eficient i coherent, separant la construcció de la base de dades de la seva gestió i accés per part dels investigadors.

3.1 Construcció de la base de dades

La primera fase es centra en la creació d'una base de dades centralitzada per emmagatzemar la informació procedent dels sis sensors: **Copeeks** [3], **iPERL** [4], **AutoFom** [5], **Nedap** [6], **SoundTalks** [7] i **AgriTech** [8].

3.1.1 Estructura de la base de dades

La base de dades conté atributs ben definits que permeten establir relacions entre els sensors i garantir la coherència i reutilització de les dades. Cada sensor disposa d'una col·lecció pròpia amb la seva estructura específica, pensada per facilitar la consulta i l'anàlisi diferenciada.

3.1.2 Relació entre col·leccions

Tot i que cada sensor disposa d'una col·lecció pròpia amb una estructura específica, la base de dades ha estat dissenyada per permetre la interrelació coherent de totes les col·leccions. Aquesta interconnexió no es fonamenta en l'ús de claus primàries i estrangeres pròpies del model relacional clàssic, sinó en un model lògic adaptat a una base de dades NoSQL orientada a documents.

Totes les col·leccions comparteixen l'atribut *engreix*, que actua com a identificador comú del període d'engreix i permet agrupar i relacionar la informació procedent dels diferents sensors dins d'un mateix estudi. Addicionalment, en funció del tipus d'anàlisi que es vulgui realitzar, s'utilitzen altres atributs comuns per establir relacions complementàries entre col·leccions, com ara:

- **Animal_number**, per a l'anàlisi individual dels animals (Nedap i Autofom).
- **Pen**, per a l'anàlisi a nivell de corral (Copeeks, iPERL i Agritech).
- **Date i Hour**, per a la sincronització temporal de les dades.

Cada document disposa d'un identificador únic *_id*, que garanteix la unicatat interna de cada registre. Les relacions entre col·leccions s'estableixen a nivell de consulta i agregació mitjançant aquests atributs comuns, fet que permet integrar de manera eficient informació ambiental, de comportament, de consum i de creixement, sense necessitat de duplicar dades ni definir relacions rígides.

A. Copeeks

La col·lecció Copeeks inclou dos tipus de documents: Environment i Behavior, diferenciats per l'atribut *type*.

Els documents d'ambient registren valors relacionats amb les condicions ambientals del corral. Els documents de comportament analitzen el moviment dels animals mitjançant càmeres, indicant distàncies, objectes en moviment i objectes estacionaris. Cal destacar que l'atribut *distance* presenta una *nested structure*, amb valors màxims i mitjans, la qual cosa permet concentrar tota la informació en un mateix camp, mantenint la taula més ordenada i coherent. La Fig. 1 mostra l'estructura de la informació de copeeks.

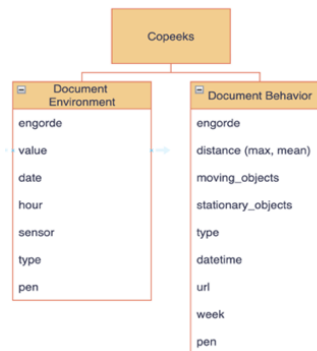


Figura 1. Estructura copeeks.

B. iPERL

La col·lecció iPERL inclou un tipus de document, aquests emmagatzemen dades de consum d'aigua per corral en litres. La Fig. 2 mostra l'estructura de la informació de iPERL.

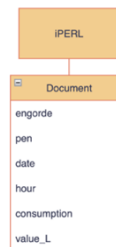


Figura 2. Estructura iPERL.

C. Autofom

La col·lecció AutoFom conté informació generada a l'escorxador per a cada animal. Els documents registren el pes durant l'engreix i de la canal, el percentatge de magre total i per seccions, així com prediccions de gruix de greix i llom. Cal destacar que els atributs, *pes*, *magre* i *mesures* tenen una estructura jeràrquica, la qual cosa concentra tota la informació relacionada en un mateix camp i facilita la seva posterior anàlisi. La Fig. 3 mostra l'estructura de la informació de Autofom.

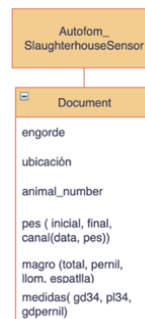


Figura 3. Estructura Autofom.

D. Nedap

La col·lecció Nedap conté informació sobre el comportament i el consum d'aliment dels animals durant l'engreix. La Fig. 4 mostra l'estructura de la informació de Nedap.



Figura 4. Estructura Nedap.

E. SoundTalks

La col·lecció SoundTalks recull informació ambiental i indicadors de benestar dels animals durant l'engreix. Els documents registren dades del lloc, la sala i el grup de porcs, així com informació sobre el monitor utilitzat, les mesures de temperatura i humitat i l'estat del risc respiratori (ReHS). Cal destacar que alguns atributs presenten una *nested structure*, com ara *Monitor* (nom i codi), *Temperature* (mínim, màxim i mitjana), *Humidity* (mínim, màxim i mitjana) i *ReHS* (estat i valor), la qual cosa permet concentrar tota la informació relacionada en un mateix document i facilita la seva anàlisi integrada amb les dades dels altres sensors. La Fig. 5 mostra l'estructura de la informació de SoundTalks.



Figura 5. Estructura SoundTalks.

F. Agritech

La col·lecció Agritech conté documents corresponents al tipus EnviroDetect, que emmagatzemen dades ambientals recollides mitjançant sensors durant l'engreix.

Els documents de tipus EnviroDetect registren mesures de temperatura (°C), humitat (%), concentració de pols (mg/m³), diòxid de carboni (CO₂, ppm), amoníac (NH₃, ppm) i taxa de ventilació (m³/h/kW). A més, cada registre inclou informació associada al lloc, al sensor, a la data i a l'hora de mesura, permetent una traçabilitat completa de les dades ambientals.

Inicialment, es contemplava també la incorporació de documents en format PDF associats a l'engreix. No obstant això, s'ha determinat que aquests documents no eren necessaris per als objectius de la investigació, motiu pel qual no s'han incorporat a la base de dades. La Fig. 6 mostra l'estructura de la informació de Agritech.

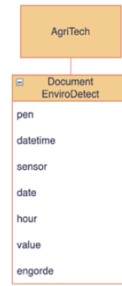


Figura 6. Estructura Agritech.

3.1.3 Sanejament de les dades

Un cop definides les col·leccions i la seva estructura, el següent pas consisteix en assegurar la qualitat i coherència de la informació abans d'inserir-les a la base de dades.

Les dades originals arriben en una carpeta proporcionada pel departament de Veterinària, organitzada en subcarpetes per sensor, i dins de cada sensor, subcarpetes corresponents a cada engreix. Inicialment, moltes de les dades poden contenir valors nuls, formats inconsistents o informació duplicada. Per això, es desenvolupa una funció de sanejament en Python [9] que realitza les següents operacions:

- **Unió de fitxers:** els fitxers de cada engreix es combinen en un únic resum per facilitar el processament.
- **Normalització de dates i hores:** totes les columnes amb informació temporal es converteixen a un format estàndard ISO8601 [10] per assegurar la coherència entre sensors.
- **Filtrat de valors nuls i inconsistents:** s'eliminen registres incomplets o erronis i es substitueixen els valors nuls per None.
- **Transformació d'atributs:**
 - Per Copeeks, es creen *nested structures* per agrupar les distàncies màximes i mitjanes dels animals.
 - Per Autofom i SoundTalks, s'organitzen les mesures en subdocuments jeràrquics per concentrar tota la informació relacionada en un mateix camp.
- **Generació de fitxers processats:** cada engreix disposa d'un fitxer Excel amb el sufix "*_processed.xlsx*" que conté les dades netes i estructurades, llestes per a la pujada a la base de dades.

3.1.4 Pujada de les dades

Un cop disponibles els fitxers processats, es procedeix a la seva inserció a la base de dades. Cada sensor disposa d'una col·lecció pròpia dins de **PigLF_BD** per mantenir les dades separades i estructurades segons l'objectiu de la investigació. Les operacions realitzades inclouen:

- **Recorregut de carpetes i fitxers processats:** s'identifiquen tots els fitxers amb el sufix "*_processed.xlsx*" dins de les subcarpetes corresponents a cada engreix.

- **Conversió de dates i hores:** totes les columnes temporals es reconverteixen a ISO8601 per garantir uniformitat.
- **Afegir metadades comunes:** s'afegeix l'identificador de l'engreix i el tipus de sensor (per Copeeks: Behavior o Environment) com a atributs de cada document.
- **Inserció massiva:** les dades es transformen en diccionaris i s'insereixen a la col·lecció corresponent. Les *nested structures* es mantenen per preservar la jerarquia de la informació.

Aquesta estratègia assegura que totes les dades dels sis sensors quedin centralitzades i disponibles per a consultes i anàlisis, amb una estructura coherent i reutilitzable per a càlculs, agregacions i visualitzacions posteriors.

3.2 Gestió i accés a la informació

Un cop les dades han estat processades i pujades a la base de dades, és necessari establir mecanismes eficients per consultar-les, agregar-les i exportar-les. Aquesta secció descriu les funcions implementades per accedir a la informació dels sensors, calcular indicadors productius i de benestar, i optimitzar les consultes per garantir un ús coherent i eficaç de la base de dades.

3.2.1 Consultes generals de les col·leccions

Les consultes generals de les col·leccions inclouen un conjunt de scripts pensats per accedir de manera directa a la informació emmagatzemada a la base de dades i realitzar càlculs bàsics o exportacions generals. Aquestes consultes constitueixen el nivell més genèric del sistema, ja que treballen directament amb les dades originals dels sensors i serveixen de base per a anàlisis i agregacions posteriors.

A. Obtenció de les dades d'una col·lecció

Per facilitar l'accés a la informació emmagatzemada a la base de dades, s'ha creat una funció general que permeten recuperar tots els documents d'una col·lecció determinada. Aquestes consultes poden retornar la totalitat de la informació d'un sensor o, si es desitja, filtrar només les dades corresponents a un engreix concret.

La funció principal estableix primer la connexió amb MongoDB [10] i selecciona la col·lecció del sensor desitjat. Si es proporciona un valor d'engreix, només es recuperen els documents corresponents a aquest engreix; en cas contrari, s'obtenen tots els registres de la col·lecció. La informació es retorna en forma de llista de documents, fet que en facilita la manipulació posterior. Addicionalment, s'ha implementat una funció complementària que permet exportar els resultats en formats Comma Separated Values (CSV) i Microsoft Excel (Excel), utilitzant els noms de columna corresponents als atributs dels documents, amb l'objectiu de facilitar la visualització i l'anàlisi de la informació.

B. Consultes de càlcul simple

Aquest conjunt de consultes agrupa scripts que realitzen càlculs simples i extraccions específiques a partir de les

dades emmagatzemades a les col·leccions dels sensors. Aquestes operacions s'apliquen directament sobre les dades originals i no impliquen processos d'agregació complexos.

- **Primera visita diària:** aquesta consulta selecciona, per a cada combinació d'animal i data, el primer registre temporal disponible, proporcionant una referència consistent per a l'anàlisi posterior dels patrons de comportament i consum.
- **Valors extrems diaris:** aquesta consulta permet obtenir els valors mínims i màxims diaris de variables com la temperatura, la humitat o la concentració de gasos, proporcionant una caracterització dels extrems ambientals als quals han estat exposats els animals durant el període analitzat.
- **Extracció i ordenació de variables específiques:** aquestes consultes permeten exportar variables específiques procedents del sensor Nedap, com la taxa de consum, o del sensor iPERL, com el consum d'aigua, organitzant els resultats per engreix, corral, data i hora. Aquestes consultes no modifiquen les dades originals, sinó que en faciliten l'accés i la reutilització en processos d'anàlisi posteriors.

Tots els scripts d'aquest grup generen els resultats en formats Comma Separated Values (CSV) i Microsoft Excel (Excel), facilitant la seva visualització i la integració amb altres processos d'anàlisi i informes.

3.2.2 Consultes i càlculs específics

Les consultes i càlculs específics constitueixen un nivell més avançat d'anàlisi sobre les dades emmagatzemades a la base de dades. A diferència de les consultes generals, que recuperen informació directa o realitzen càlculs simples, aquests processos integren dades de diferents sensors, apliquen mètodes de neteja i agregació i generen indicadors productius i de benestar més complexos.

A. Temperature-Humidity Index

Per monitoritzar el benestar dels animals dins del corral, s'ha implementat una funció que calcula el THI a partir de les dades ambientals recollides pel sensor Copeeks. Aquest indicador combina la temperatura i la humitat relativa per obtenir una mesura única del confort tèrmic dels animals.

Les dades utilitzades provenen de les mesures de temperatura interior (*inside_temperature_raw*) i humitat interior (*inside_moisture_raw*) dins del corral.

El càlcul del THI s'ha realitzat aplicant el següent procediment. En primer lloc, es pot filtrar la informació per un engreix concret o bé utilitzar totes les dades disponibles. A continuació, les mesures es consoliden agrupant-les per engreix, dia, hora i corral. Sobre aquests valors agregats s'aplica la següent fórmula:

$$THI = [(1,8 * inside_{temp}) + 32] - \left[(0,55 * \left(\frac{inside_{humidity}}{100} \right)) * [(1,8 * inside_{temp}) + 32] - 58 \right]$$

Finalment, els resultats es guarden en un arxiu Comma Separated Values (CSV) i Microsoft Excel (Excel), amb les columnes corresponents a engreix, dia, hora, corral, temperatura, humitat i THI, cosa que facilita la visualització i la posterior integració amb altres indicadors. La funció retorna també un DataFrame amb els valors agregats i el THI calculat. Cada engreix genera un arxiu separat, tot i que es pot obtenir un Comma Separated Values (CSV) i Microsoft Excel (Excel) global amb tots els engreixos. Aquest procediment permet obtenir una visió horària i per corral del confort ambiental dins dels espais d'engreix, constituint un pas previ essencial per a l'anàlisi del benestar animal.

B. Nedap: clean methods

Per garantir la fiabilitat dels càlculs derivats de les dades de consum d'aliment recollides pel sensor Nedap, s'aplica un procés de neteja que detecta i elimina outliers. Aquest procés no es realitza abans de pujar les dades a la base de dades, sinó que s'executa com a consulta posterior, ja que els líndars i criteris per considerar un valor com a outlier poden variar amb el temps i amb el conjunt de dades disponible. D'aquesta manera s'assegura que el càlcul sigui dinàmic i coherent en tot moment.

La funció de neteja permet utilitzar diversos mètodes, com el criteri IQR, Mahalanobis, Isolation Forest, filtres de desviació estàndard o criteris biològics. Per cada engreix, la funció processa les dades i retorna un DataFrame net, indicant els registres eliminats i el percentatge corresponent. Els resultats també es poden exportar a Comma Separated Values (CSV) i Microsoft Excel (Excel), per a la seva posterior utilització i anàlisis.

Aquest enfocament garanteix que qualsevol canvi en el criteri de neteja s'apliqui sempre sobre les dades més recents, sense dependre de la informació prèviament emmagatzemada a la base de dades.

C. Feed Conversion Ratio

El Feed Conversion Ratio (FCR) és un indicador productiu fonamental que permet avaluar l'eficiència amb què els animals converteixen el consum d'aliment en guany de pes durant el període d'engreix. En aquest projecte, el càlcul del FCR s'ha implementat mitjançant una funció específica que integra dades procedents del sensor Nedap i del sensor Autofom.

El procés s'inicia a partir de les dades de consum d'aliment recollides pel sensor Nedap, sobre les quals s'aplica prèviament el mètode de neteja d'outliers descrit en l'apartat anterior. Aquesta neteja es realitza dinàmicament en el moment de la consulta, assegurant que el càlcul del FCR es basi sempre en dades depurades segons els criteris seleccionats, sense alterar els registres originals de la base de dades.

Un cop obtingudes les dades netes, es calcula el consum total d'aliment per animal mitjançant l'agregació del feed intake al llarg de tot l'engreix. A continuació, es consulten les dades del sensor Autofom per obtenir el pes inicial i el pes final de cada animal. Aquestes dades es combinen utilitzant l'identificador de l'animal i l'engreix corresponent.

A partir d'aquesta informació, es calcula el guany de pes com la diferència entre el pes final i el pes inicial, i posteriorment el FCR com:

$$FCR = \frac{\text{total feed intake}}{\text{weight gain}}$$

El resultat és un conjunt de valors de FCR individualitzats per animal i engreix, que permeten analitzar l'eficiència alimentària de manera precisa.

Finalment, els resultats del càlcul del FCR es poden exportar en format Comma Separated Values (CSV) i Microsoft Excel (Excel), tant per a un engreix concret com de manera global per a tots els engreixos disponibles, facilitant així la seva visualització, comparació i anàlisi posterior dins del marc del projecte.

D. Consum individual d'aigua

A partir de les dades recollides pel sensor iPERL s'ha implementat el càlcul del consum individual d'aigua dels animals. Atès que el sensor registra el consum total d'aigua per corral, el consum individual s'estima repartint aquest valor entre el nombre d'animals presents en cada moment. Amb l'objectiu de facilitar diferents nivells d'anàlisi, aquest càlcul s'ha implementat en dues funcions diferenciades, una amb enfocament diari i una altra amb enfocament horari.

En ambdós casos, el procés pot aplicar-se a un engreix concret o bé al conjunt de tots els engreixos disponibles. El càlcul es realitza mitjançant consultes dinàmiques sobre la base de dades, sense modificar els valors originals emmagatzemats.

En el càlcul del consum individual diari, les dades d'iPERL s'agrupen per corral, dia i engreix, obtenint el consum total d'aigua del corral durant aquell dia. Posteriorment, mitjançant una consulta a la col·lecció Nedap, es determina el nombre d'animals presents en cada corral i dia. A partir d'aquesta informació es calcula el consum individual diari d'aigua dividint el consum total pel nombre d'animals detectats. El resultat s'assigna a cada animal del corral per a la data corresponent i es retorna com un conjunt de registres individuals. Els resultats poden exportar-se en format Comma Separated Values (CSV) i Microsoft Excel (Excel), facilitant la seva anàlisi posterior.

De manera anàloga, el càlcul del consum individual horari segueix el mateix principi però incorporant una resolució temporal més fina. En aquest cas, les dades d'iPERL

s'agrupen per corral, dia, hora i engreix, obtenint el consum total d'aigua per franja horària. A continuació, es consulta la col·lecció Nedap per identificar el nombre d'animals presents en el mateix corral, dia i hora. El consum individual horari s'obté dividint el consum total d'aigua entre el nombre d'animals detectats en aquell interval temporal. El resultat es genera de forma individualitzada per animal i es pot exportar també en format Comma Separated Values (CSV) i Microsoft Excel (Excel).

Aquest enfocament permet analitzar el consum d'aigua tant des d'una perspectiva global diària com amb un major detall horari, aportant informació rellevant sobre els patrons de consum dels animals durant l'engreix i facilitant la seva integració amb altres indicadors productius i de benestar.

E. Average Daily Gain

Com a indicador productiu específic, s'ha implementat el càlcul de l'ADG (Average Daily Gain), que permet mesurar el guany mitjà de pes diari de cada animal durant el període d'engreix. Aquest indicador s'obté a partir de les dades de pes recollides pel sensor d'escorxador Autofom i del temps total de permanència de l'animal a l'engreix.

El càlcul de l'ADG es realitza mitjançant una consulta d'agregació directament sobre la base de dades. En primer lloc, es poden filtrar les dades per un engreix concret o bé processar tots els engreixos disponibles. Per a cada animal, es recuperen el pes inicial i el pes final, així com les dates associades a aquestes mesures. A partir de les dates es calcula el nombre total de dies d'engreix mitjançant la diferència temporal entre la data inicial i la final.

Un cop determinats aquests valors, l'ADG es calcula com el quocient entre el guany total de pes i el nombre de dies d'engreix, segons l'expressió:

$$ADG = \frac{\text{weight gain}}{\text{fattening days}}$$

Aquest càlcul només s'aplica quan el nombre de dies d'engreix és superior a zero, evitant així valors no vàlids. El resultat s'associa a cada animal juntament amb el seu engreix corresponent i es retorna en forma de DataFrame. Finalment, els valors d'ADG es poden exportar a un arxiu Comma Separated Values (CSV) i Microsoft Excel (Excel), facilitant la seva anàlisi i la seva integració amb altres indicadors com el FCR o el consum d'aigua.

Aquest indicador permet disposar d'una mesura clara i comparable del rendiment productiu individual dels animals, sent un element clau per a l'avaluació de l'eficiència de l'engreix.

3.2.3 Agrupació i exportació visual

Per facilitar la visualització i l'anàlisi, les dades es consoliden en conjunts estructurats que permeten un accés clar i organitzat a la informació. En aquesta secció, es realitzen consultes que integren tota la informació relacionada,

proporcionant una visió completa i centralitzada dels registres a nivell individual, diari o horari.

A. Agrupació per identificador animal

El procés d'agrupació i consolidació genera un conjunt de dades a nivell individual, on cada fila correspon a un animal dins d'un engreix. El registre integra informació procedent de diferents fonts i càlculs, proporcionant una visió completa del rendiment productiu de cada animal.

S'inclouen els identificadors d'engreix i `animal_number`, el consum total d'aliment obtingut del sensor Nedap després d'aplicar els mètodes de neteja d'outliers, i les dades de pes i qualitat de canal del sensor Autofom, com el pes inicial i final, el pes de la canal i els valors de magre total i per parts (pernil, llom, espatlla).

A partir d'aquesta informació es calculen els indicadors `weight_gain`, FCR (Feed Conversion Ratio) i ADG (Average Daily Gain), així com les dates d'inici i final de l'engreix i el nombre de dies corresponent.

El procés identifica també la pen dominant per cada animal, és a dir, aquella on s'ha registrat més consum d'aliment, i assigna aquesta informació al registre final.

Els registres incomplets es mantenen amb valors buits per garantir la coherència de la taula i la correcta alineació de les dades.

El conjunt final s'exporta tant en Comma Separated Values (CSV) com Microsoft Excel (Excel), i es poden generar fitxers per engreix específic o per tots els engreixos de manera agregada. Aquesta estructura permet un anàlisi estadístic complet, la visualització de resultats i la comparació entre animals i engreixos de forma clara i centralitzada.

B. Agrupació diària de la sala per Copeeks i Soundtalks

El procés d'agrupació i consolidació de dades diàries de l'entorn genera un conjunt de dades a nivell de sala, engreix i dia, on cada fila correspon a un registre d'un engreix en un dia concret.

Les dades s'obtenen de dues fonts principals:

- **Copeeks (Environment Sensor):** inclou mesures de temperatura i humitat (tant dins com fora de la sala). Registra també els nivells de CO₂ i NH₃, i per cada variable es calculen els valors mitjà, màxim i mínim diaris.
- **SoundTalks (Room Sensor):** mesura la temperatura i humitat de la sala amb indicadors mínim, màxim i mitjà. Inclou també el valor de ReHS, indicant l'estat general de l'ambient.

El procés permet filtrar opcionalment per engreix si es vol analitzar un cas concret, i agrupa les dades de Copeeks i

SoundTalks per engreix i dia, combinant tota la informació en un únic registre diari. Finalment, ordena les columnes i les files per engreix i data per garantir una presentació clara i homogènia.

Els registres finals s'exporten en formats CSV i Excel, preparats per a l'anàlisi estadística, la visualització de resultats i el seguiment diari de l'entorn de cada engreix.

C. Agrupació diària per dades ambientals al Pen 10

El procés d'agrupació i consolidació de dades ambientals per Pen 10 genera un conjunt de dades a nivell diari i per engreix on cada fila correspon a un registre d'un engreix en un dia concret amb informació provinent de tres fonts principals de sensors:

- **Copeeks:** les dades inclouen la temperatura i humitat interiors de la sala així com els nivells de CO₂ i NH₃ per cada variable es calculen els valors mitjà màxim i mínim diaris.
- **SoundTalks:** les dades mesuren la temperatura i humitat de la sala amb indicadors mínim màxim i mitjà i inclouen el valor de ReHS que indica l'estat general de l'ambient.
- **AgriTech:** les dades corresponents al Pen 10 mesuren variables com la temperatura (T), la humitat relativa (R), l'índex temperatura/humitat (THI), la diferència de temperatura entre dins i fora (T_DIFF), la humitat exterior (R_OUT), la temperatura exterior (T_OUT), el CO₂ (C), l'amoníac (N), el polsim (D), l'emissió d'amoníac (EMISSION_N), l'emissió de polsim (EMISSION_D), la ventilació millorada (Ventilation_new), l'activitat dels animals (MOVEMENTS) i l'activitat humana dins del pen (Human Activity), calculant-se el valor mitjà diari de cada variable per a cada engreix.

El procés filtra les dades per Pen 10 i opcionalment per un engreix concret afegeix prefixos a les columnes segons la font copeeks agriTech soundTalks per evitar duplicats i realitza un merge horitzontal de les tres fonts per engreix i dia combinant tota la informació en un únic registre.

Les files s'ordenen per engreix i data i les columnes amb engreix i data apareixen primer. Les dades finals s'exporten en Comma Separated Values (CSV) i Microsoft Excel (Excel) preparades per a l'anàlisi estadística visualització de resultats i comparació de condicions ambientals diàries dins del Pen 10.

Aquest enfocament permet monitoritzar les condicions ambientals de forma detallada per engreix i dia comparar les dades entre diferents sensors detectar anomalies i preparar la informació per a informes i visualitzacions integrades.

D. Agrupació horària de dades ambientals de Copeeks i AgriTech

Aquest procés genera un conjunt de dades ambientals a nivell d'engreix i hora, on cada fila representa un registre horari corresponent a un engreix en un moment concret del dia. L'objectiu és consolidar informació procedent de diferents sistemes de monitoratge ambiental per obtenir una visió conjunta i detallada de l'estat de la sala al llarg del temps.

Les dades s'obtenen de dues fonts principals. Copeeks proporciona mesures ambientals de la sala, incloent la temperatura i la humitat tant interiors com exteriors, així com els nivells de CO₂ i NH₃. A partir de les dades originals, es construeix una marca temporal horària combinant la data i l'hora, i es calcula el valor mitjà de cada variable per a cada engreix i hora.

AgriTech aporta informació ambiental i de comportament animal, com la temperatura (T), la humitat relativa (R), l'índex tèrmic (THI), la diferència de temperatura entre interior i exterior (T_DIFF), la humitat exterior (R_OUT), la concentració de CO₂ (C), l'amoníac (N), les emissions d'amoníac (EMISSION_N), la ventilació (Ventilation_new) i el nivell d'activitat dels animals (MOVEMENTS). Igual que en el cas de Copeeks, les dades s'agrupen per engreix i hora i se'n calcula el valor mitjà horari.

Un cop obtingudes les dues taules, les dades de Copeeks i AgriTech es combinen mitjançant una unió per engreix i marca temporal, generant un únic registre horari que integra tota la informació disponible. El resultat final s'exporta en formats Comma Separated Values (CSV) i Microsoft Excel (Excel), preparat per a anàlisi temporal, comparació entre sensors i estudi detallat de l'evolució horària de les condicions ambientals i del comportament dins de cada engreix.

E. Agrupació per Pen de dades ambientals de Copeeks

Per tal de monitoritzar de manera efectiva les condicions ambientals dins dels espais d'engreix, s'ha implementat un procés d'agregació de les dades recollides pel sensor Copeeks. Aquesta agregació permet obtenir una visió consolidada de variables com la temperatura i la humitat (interior i exterior), així com els nivells de CO₂ i NH₃, calculant per a cada variable els valors mitjà, mínim i màxim.

S'han definit dos tipus de resolució temporal per a l'anàlisi:

- **Agregació diària:** aquesta agregació consolida les dades per engreix, pen i dia, generant un registre per cada pen en cada dia. La resolució diària permet una anàlisi general del comportament ambiental i facilita el seguiment longitudinal de les condicions dins dels espais d'engreix.

- **Agregació horària:** aquesta agregació consolida les dades per engreix, pen, dia i hora, obtenint un registre per cada pen en cada hora del dia. Aquesta resolució més fina permet identificar variacions i possibles pics o anomalies horàries en temperatura, humitat i concentració de gasos dins dels espais d'engreix.

En ambdós casos, els conjunts de dades generats es poden exportar tant en format Comma Separated Values (CSV) com Microsoft Excel (Excel), permetent la seva visualització, comparació i integració amb altres indicadors de benestar i productius, i facilitant l'anàlisi estadística i la preparació d'informes.

3.2.4 Creació d'índex a MongoDB

Per optimitzar les consultes i millorar el rendiment de la base de dades, s'han definit diversos índexs sobre els camps més rellevants de cada col·lecció. Alguns són compostos per combinar diversos atributs en una sola estructura, mentre que altres indexen un únic camp quan és suficient per agilitzar les operacions més freqüents.

A. Agritech

Per optimitzar les consultes i millorar el rendiment de la base de dades, s'han definit diversos índexs sobre els camps més rellevants de cada col·lecció. Alguns són compostos per combinar diversos atributs en una sola estructura, mentre que altres indexen un únic camp quan és suficient per agilitzar les operacions més freqüents. La *Taula 1* mostra els índex d'Agritech.

Nom	Tipus	Camps
engorde	compost	engorde

Taula 1. Índex Agritech

B. Autofom

Els índexs compostos faciliten les consultes que impliquen engreix i identificador de l'animal, així com la data inicial del pes. Aquesta estructura és útil per a càlculs d'indicadors productius a nivell individual. La *Taula 2* mostra els índex d'Autofom.

Nom	Tipus	Camps
engorde_1_pes.inicial.data_1_animal_number_1	compost	engorde, pes.inicial.data, animal_number
engorde_1_animal_number_1	compost	engorde, animal_number

Taula 2. Índex Autofom

C. Copeeks

L'índex compost sobre engreix permet filtrar ràpidament les dades ambientals recollides per cada engreix, essent fonamental per al càlcul d'indicadors de confort ambiental. La *Taula 3* mostra els índex de Copeeks.

Nom	Tipus	Camps
engorde	compost	engorde

Taula 3. Índex Copeeks

D. Nedap

Els índexs compostos permeten recuperar les dades d'alimentació per animal i per corral de manera eficient, tant a nivell diari com horari, facilitant la neteja d'outliers i el càlcul de FCR. La *Taula 4* mostra els índex de Nedap.

Nom	Tipus	Camps
engorde_1_location_1_date_1_Hour_1	compost	engorde, location, date, Hour
engorde_1_location_1_animal_number_1_date_1_visit_time_1	compost	engorde, location, animal_number, date, visit_time
engorde_1_animal_number_1	compost	engorde, animal_number

Taula 4. Índex Nedap

E. SoundTalks

L'índex per engreix facilita la selecció de les dades de so per cada engreix, permetent realitzar anàlisis de benestar animal per grup. La *Taula 5* mostra els índex de SoundTalks.

Nom	Tipus	Camps
engorde	compost	engorde

Taula 5. Índex SoundTalks

F. iPERL

Els índexs compostos permeten calcular el consum individual d'aigua per animal amb gran eficiència, tant per hores com per dies. L'índex sobre engreix proporciona un filtratge ràpid quan només es requereix seleccionar un engreix concret. La *Taula 6* mostra els índex d'iPERL.

Nom	Tipus	Camps
engorde_1_Pen_1_Date_1_Hour_1	compost	engorde, Pen, Date, Hour
engorde	compost	engorde

Taula 6. Índex iPERL

3.2.4 Presentació a l'usuari

Per facilitar l'ús del sistema, els usuaris disposen d'un document explicatiu en format Markdown (MD) que descriu la funcionalitat de cada script, el seu mode d'execució i els fitxers generats en cada cas. Aquesta documentació permet comprendre el funcionament del sistema i reproduir els diferents processos de manera controlada.

Com a resultat de l'execució dels scripts, les dades es retornen en dos formats complementaris: Comma Separated Values (CSV) i Microsoft Excel (Excel). El format CSV facilita el treball posterior amb les dades i la seva reutilització en altres eines d'anàlisi, mentre que el format Excel ofereix una visualització més clara i accessible dels resultats.

3.3 Estudi escalabilitat

Actualment, el sistema presenta limitacions, com l'accés restringit a un únic ordinador i la necessitat de processar manualment els fitxers dels sensors. Per això, s'ha realitzat aquest estudi d'escalabilitat per analitzar opcions que

permetin un creixement futur i millorin l'accessibilitat i la gestió de la informació.

3.3.1 Escalabilitat al núvol

Per garantir disponibilitat, escalabilitat i flexibilitat, es proposa migrar la base de dades a un servei de núvol, concretament a MongoDB Atlas [12] amb un clúster compartit (Shared Cluster).

Els principals beneficis són l'accés remot i concurrent des de múltiples dispositius, la possibilitat d'ampliar l'emmagatzematge i els recursos segons el creixement de dades i documents, i la seguretat i alta disponibilitat que ofereix Atlas, amb còpies de seguretat automàtiques, backups i xifratge de la informació. A més, el manteniment del servidor queda simplificat, ja que Atlas s'encarrega de les actualitzacions i gestió del servei.

MongoDB Atlas ens proporciona diferents plans adaptats a les necessitats del projecte. Donada la mesura actual de la base de dades i la taxa de creixement prevista, es recomana començar amb un clúster compartit M0 (gratuit) o M2 (2 GB d'emmagatzematge), ja que el volum total de dades i índexs actual és inferior a 40 MB. Això ofereix marge suficient per a diversos engreixos anuals sense necessitat d'escalar immediatament.

Si en el futur el sistema creix més del previst, Atlas permet escalar fàcilment a plans superiors com M5 o M10 sense interrupció del servei, augmentant tant la capacitat d'emmagatzematge com el rendiment de lectura i escriptura. Aquesta flexibilitat assegura que la base de dades pugui créixer amb el projecte, mantenint l'accessibilitat, la seguretat i la integritat de la informació.

3.3.2 Integració mitjançant APIs

Actualment, la gestió de la informació dels sensors es realitza de manera manual, amb processament local dels fitxers. Aquesta forma de treball limita la velocitat i l'eficiència del sistema, i dificulta l'escalabilitat.

Una possible via per millorar l'escalabilitat seria la implementació d'APIs proporcionades pels proveïdors dels sensors, que permetrien accedir de manera directa i automàtica a les dades. Això facilitaria la centralització i actualització contínua de la informació, reduint la càrrega manual i permetent que el sistema pugui créixer amb més sensors sense incrementar significativament la complexitat de la gestió.

3.3.3 Interfície web i visualització de dades

Com a evolució futura del sistema, es planteja la implementació d'una interfície web que permeti interactuar amb la base de dades de manera més intuïtiva. Aquesta capa de presentació facilitaria la realització de consultes i l'obtenció de resultats sense necessitat d'executar scripts manualment ni disposar de coneixements de programació. En aquest flux, el frontend no processa les dades, sinó que interactua amb el backend per sol·licitar informació i mostrar-la a l'usuari.

3.3.4 Diagrama de proposta d'escalabilitat

El diagrama de la *Figura 7* mostra la proposta d'arquitectura escalable del sistema **PigLF_BD** i gestió de dades. Representa com les dades recollides per diferents sensors (AgriTech, iPERL, Autofom, Nedap, SoundTalks i Copeeks) són enviades al backend per al seu processament i sanejament, emmagatzemades a la base de dades MongoDB Atlas, i finalment accessibles tant per a la manipulació de resultats a través del frontend com per a l'exportació en formats Comma Separated Values (CSV) i Microsoft Excel (Excel).

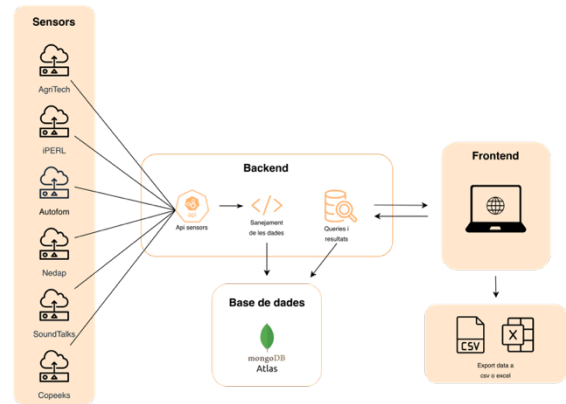


Figura 7. Diagrama arquitectura

4. RESULTATS

La divisió del projecte en dues fases ha demostrat ser efectiva, tant per a l'organització del desenvolupament com per al testeig individual de cada component.

Durant la fase de construcció de la base de dades, es van identificar diversos reptes relacionats amb la normalització i coherència de les dades. Les mesures procedents dels sensors presentaven formats diferents, valors nuls o inconsistències temporals, la qual cosa requeria un sanejament específic per a cada tipus de sensor. Per garantir la integritat de la informació, s'han aplicat operacions de filtrat, transformació de dates a un format estàndard ISO8601 i creació de nested structures per concentrar informació jeràrquica en un mateix document, especialment per a les dades de Copeeks, Autofom i SoundTalks. Aquesta estratègia ha facilitat la posterior agregació de dades i l'execució de consultes complexes.

Durant la fase de gestió i consultes, s'han implementat funcions que permeten calcular indicadors de benestar i productivitat com el Feed Conversion Ratio (FCR), Average Daily Gain (ADG), Temperature Humidity Index (THI) i el consum individual d'aigua. L'ús d'índexs composts a MongoDB ha millorat el rendiment de les consultes, permetent filtrar ràpidament per engreix, corral o identificador d'animal. Així, els càlculs es poden realitzar de manera dinàmica sobre dades actualitzades sense alterar els registres originals.

Els resultats més destacables mostren que la base de dades actualment està en local, però ha estat dissenyada de manera estructurada i escalable, preparada per a una futura migració a núvol si fos necessari. Els indicadors productius i de benestar es poden calcular de manera fiable per a cada animal i engreix. La implementació de nested structures i índexs composts ha estat clau per optimitzar consultes complexes, reduint significativament el temps de processament. Les funcions d'exportació permeten obtenir dades agregades i individuals en format Comma Separated Values (CSV) i Microsoft Excel (Excel), facilitant així l'anàlisi i la integració amb altres eines.

5. PLANIFICACIÓ

La planificació del projecte s'ha desenvolupat segons el ritme establert inicialment, seguint un format seqüencial que ha permès avançar de manera ordenada i controlada. Aquesta organització ha facilitat mantenir un bon ritme de treball, assegurant que cada etapa es completés abans de passar a la següent i deixant marge suficient per a la revisió i comprovació de dades. En conjunt, el caràcter seqüencial del projecte ha contribuït a gestionar el temps de manera eficaç, garantint que la base de dades i les consultes estiguin completament operatives i preparades per a la seva utilització futura. La Fig 8. Mostra el diagrama de planificació.

		SET					OCT					NOV					DES					GEN					FEB				
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22								
Construcció de la base de dades	Inici i plantejament	█																													
	Planificació i recollida dades	█																													
	Estructurar BD col·leccions inicials	█																													
	Implementació scripts i sanitització	█																													
Gestió i accés a la informació	Creació getters per obtenir dades																														
	Testejar scripts																														
	Estudi escalabilitat																														

Figura 8. Diagrama de planificació

6. CONCLUSIONS

El projecte ha aconseguit centralitzar i organitzar de manera eficient tota la informació recollida pels sensors, facilitant l'anàlisi integrada de les dades de creixement, benestar i condicions ambientals dels animals. La metodologia adoptada ha demostrat ser robusta i escalable, amb una estructura que permet actualment treballar en un entorn local i que pot ser migrada a un entorn de núvol.

El model de dades basat en *nested structures* i l'organització per sensor proporciona flexibilitat per incorporar nous sensors o variables sense afectar l'estructura global de la base de dades. Les operacions de neteja de dades, la conversió de dates a un format estàndard ISO i la gestió coherent dels valors nuls asseguren la fiabilitat, traçabilitat i reutilització de la informació.

A més, la col·laboració amb el departament de Veterinària ha garantit que les dades i les estructures implementades compleixin plenament els requisits dels estudis científics. L'organització de la base de dades i les col·leccions permet una futura integració amb APIs de sensors, fet que podria automatitzar completament la recollida de dades i millorar encara més l'eficiència del sistema.

En conjunt, el projecte estableix una base sòlida, fiable i escalable que, tot i estar actualment en local, permetria en el futur habilitar l'accés simultani de diversos investigadors, assegurar la coherència de la informació i facilitar l'anàlisi avançada de les dades, complint plenament els objectius establerts inicialment.

7. ÚS D'IA GENERATIVA

S'ha utilitzat IA per la millora de la qualitat d'aquest document per a fer-lo més clar i sintètic. L'ús de IA s'ha fet de la següent manera: primer s'ha redactat el contingut de la memòria i a continuació s'ha utilitzat la IA per a millorar-ne la qualitat del redactat, a continuació s'ha verificat el text generat per la IA ja que en ocasions aquest contenia algun error. La IA utilitzada ha estat Copilot [13].

BIBLIOGRAFIA

- [1] Inici ». (2022, octubre 3). IRTA. <https://www.irta.cat>
- [2] Documentación. (2022, setembre 29). BDporc <https://bdporc.irta.es/documentacion/>
- [3] Tell ELavage - Copeeks - Monitorize vos productions. (2022, marzo 29). Tell ELavage - Copeeks; COPEEKS. <https://copeeks.fr/en/>
- [4] Sensus Contadores de agua iPERL® (International). (s/f). Xylem.com. Recuperat el 14 de desembre de 2025, de <https://www.xylem.com/es-es/products--services/metrology-equipment-for-utilities/meters/iperl-international-water-meter/>
- [5] Sistema Automático de Clasificación de Canales de Porcino por Ultrasonidos. Frontmatec.com. Recuperat el 14 de desembre de 2025, de https://www.frontmatec.com/media/5126/frontmatec-instruments-autofom-iii_v1-0-2020-es_web.pdf
- [6] Nedap identification systems. (2022, diciembre 5). Nedap Identification Systems. <https://www.nedapidentification.com>
- [7] SoundTalks - AI-powered sound tech for healthier and happier animals. (s/f). Soundtalks.com. Recuperat el 14 de desembre de 2025, de <https://www.soundtalks.com/es/>
- [8] Technologies. (s/f). PLF Agritech. Recuperat el 14 de desembre de 2025, de <https://plfag.com/technologies/>
- [9] Download Python v3.13.0. Python.org. Recuperat el 4 d'octubre de 2025, de <https://www.python.org/downloads/>
- [10] ISO 8601. (2020). ISO. <https://www.iso.org/iso-8601-date-and-time-format.html>
- [11] MongoDB v8.2.0 Documentation - homepage. MongoDB.com; MongoDB. Recuperat l'1 d'octubre de 2025, de <https://www.mongodb.com/docs/>
- [12] MongoDB Atlas. MongoDB.com. Recuperat l'1 d'octubre de 2025, de <https://www.mongodb.com/docs/atlas/>
- [13] Herramientas de inteligencia artificial para organizaciones. (s/f). Microsoft.com. Recuperat el 9 de febrer de 2026, de <https://www.microsoft.com/es-es/microsoft-copilot/organizations>