

# 迈向国际顶尖水平：基于 MySQL 架构的高性能肉羊育种系统设计与实施蓝图

## 1. 执行摘要

本报告旨在为构建一个能够比肩甚至超越澳大利亚 LAMBPLAN、美国 NSIP 以及法国主要育种体系的国际顶尖肉羊育种系统提供详尽的技术与理论蓝图。针对用户提出的核心需求——即利用 MySQL 等免费开源数据库技术实现高性能育种管理——本方案不仅局限于数据存储层面的设计，更深入至数量遗传学评估引擎、基因组选择（Genomic Selection）集成、离线优先（Offline-First）的现场数据采集以及符合 ICAR（国际动物记录委员会）标准的交互协议。

在当今精准农业与生物信息学飞速发展的背景下，一个顶级的育种系统必须具备处理复杂系谱、海量表型数据以及高密度 SNP 基因型数据的能力。本报告将详细阐述如何利用 MySQL 8.0 的高级特性（如递归公用表表达式、JSON 字段、空间索引）来构建低成本但高并发的后端架构，同时结合 BLUPF90 等开源遗传评估软件，实现单步法基因组最佳线性无偏预测（ssGBLUP）。通过深入剖析国际先进系统的运作逻辑，我们提出了一套包含育种目标制定、数据采集标准化、遗传评估自动化及选种决策可视化的全流程解决方案，旨在为我国内肉羊产业提供核心技术支撑，推动从表型选择向基因组选择的跨越式发展。

## 2. 全球肉羊育种系统基准分析与顶层设计理念

设计一个具有国际竞争力的育种系统，首先必须透彻理解当前全球行业的“金标准”。这不仅仅是软件功能的堆砌，更是对育种哲学、数据流转机制以及产业组织形式的深刻洞察。

### 2.1 澳大利亚模式：**Sheep Genetics (LAMBPLAN)** 的核心优势

澳大利亚的 Sheep Genetics 体系，特别是针对肉羊的 LAMBPLAN，被公认为全球最成功的商业化育种评估系统之一。其核心竞争力在于不仅提供数据记录功能，更提供了一套完整的遗传评估服务。

跨群与跨品种评估的“联合分析”（Combined Analysis）

LAMBPLAN 最近的重大革新在于实施了“联合分析”（Combined LAMBPLAN）。传统上，终端父本（Terminal）和母本（Maternal）品种的评估是分开进行的。然而，随着产业对复合性状关注度的提升，这种界限日益模糊。新的系统允许对所有终端和母本品种（美利奴除外）进行直接的遗传价值比较<sup>1</sup>。这意味着我们的系统设计不能将品种视为隔离的数据孤岛，必须构建一个能够容纳多品种、跨群连接的底层架构。系统必须能够处理数百万条历史记录，并通过精密的算法剔除“无信息记录”（non-informative records），以减少计算负荷并提高评估效率，确保持续增长的基因组数据不会拖垮系统性能<sup>1</sup>。

### 关联性 (Connectedness) 与关联公羊 (Link Sires)

LAMBPLAN 的成功高度依赖于“关联公羊”的使用。只有当同一只公羊（或其遗传关联亲属）在不同羊群、不同环境中留有后代时，遗传评估模型才能有效地将“遗传效应”与“环境效应”剥离<sup>2</sup>。因此，本系统必须内置“关联性分析”模块，能够自动计算并可视化各核心育种场之间的遗传联系强度，并在联系薄弱时发出预警。

### 反馈机制与决策支持

澳大利亚系统的一个显著特点是其用户友好的反馈机制。系统不仅输出育种值（ASBVs），还提供“羊群概况”（Flock Profile）等商业化基因组产品，帮助育种者对标行业水平<sup>3</sup>。这要求我们的前端设计必须具备强大的数据可视化能力，能够将复杂的统计结果转化为直观的图表。

## 2.2 美国模式：NSIP 的数据采集与质量控制

美国的国家绵羊改良计划（NSIP）虽然在遗传评估引擎上依赖澳大利亚的 Sheep Genetics，但在前端数据采集和标准化方面建立了一套值得借鉴的严谨体系。

### 采集与分析的分离架构

NSIP 使用客户端软件（如 Pedigree Master）进行数据录入，然后批量上传至中央数据库进行分析<sup>5</sup>。这种架构的优势在于，它允许在数据进入核心数据库之前进行严格的本地校验。我们的系统应借鉴这一理念，开发支持离线操作的现场采集终端，确保数据在源头就是清洗过的。

### 技术员认证制度

对于肉羊育种至关重要的超声波扫描性状（眼肌深度、背膘厚），NSIP 实施了严格的技术员认证制度。只有经过认证的技术员上传的数据才会被纳入遗传评估<sup>7</sup>。这意味着我们的数据库设计中必须包含“数据来源可信度”或“技术员资质”的字段，以区分普通农户自测数据与专业认证数据。

### 当代组 (Contemporary Groups) 的严格定义

NSIP 极其强调“当代组”的概念——即一群在相同年份出生、在相同管理条件下饲养的动物。这是准确评估遗传潜力的基石<sup>8</sup>。系统逻辑必须强制用户正确划分当代组，防止将不同管理背景下的羊只混淆比较，从而导致评估偏差。

## 2.3 法国模式：官方监管与全产业链整合

法国的育种信息系统以其高度的组织化和与国家监管体系的深度融合而著称。

### 官方标识与可追溯性

法国系统基于严格的 16 位国际识别码（国家代码+羊群号+年份+个体号）<sup>9</sup>。这与 ICAR 标准高度一致。我们的系统必须将这种 16 位唯一标识码（UID）作为数据库的主键设计基础，确保每一只羊在全球范围内都是唯一的，并且其移动、系谱和屠宰信息都可以精确追溯。

### 多性状复合指数

法国的育种目标非常全面，特别是对于像“Blanche du Massif Central”这样的品种，其选育指标涵盖了从多产性、母性能力（泌乳力）到胴体质量的全方位性状<sup>10</sup>。这提示我们，系统必须具备极高的灵活性，能够支持自定义的综合选择指数（Selection Index），将数十个性状的育种值根据经济权重合成为一个总分。

### 金字塔式的繁育体系

法国体系清晰地划分了核心群、扩繁群和商品群<sup>11</sup>。系统设计应支持这种层级结构，允许核心群

的高质量遗传数据向底层单向流动，同时允许底层的屠宰数据（后裔测定）向上回馈。

## 2.4 顶层设计：基于开源技术的超越

综上所述，要设计一个达到国际顶尖水平的肉羊育种系统，我们不能仅仅停留在模仿。我们需要利用现代开源技术实现弯道超车。核心设计理念如下：

1. **数据驱动的架构 (Data-Driven Architecture):** 以 MySQL 8.0 为核心，利用其 JSON 文档存储能力处理多变的表型数据，利用其强一致性 (ACID) 处理系谱关系。
2. **API 优先 (API-First):** 严格遵循 BrAPI (Breeding API) 标准，确保系统能够对接各种智能硬件（电子耳标阅读器、自动称重分群门）和第三方分析软件<sup>12</sup>。
3. **基因组原生 (Genomic-Native):** 系统不仅仅是在系谱树上挂载 SNP 数据，而是要原生支持单步法评估 (ssGBLUP)，将基因组关系矩阵 ( $H$  矩阵) 作为核心运算逻辑的一部分<sup>14</sup>。
4. **离线优先 (Offline-First):** 考虑到牧场网络的局限性，必须构建基于 PWA (Progressive Web App) 和 RxDB 等技术的离线数据同步机制，确保在无网络环境下也能高效工作<sup>15</sup>。

---

## 3. 数量遗传学核心引擎与数学模型

一个育种系统的灵魂不在于其界面是否华丽，而在于其遗传评估引擎是否准确、高效。本章将详细阐述系统后台的数学模型与算法实现路径。

### 3.1 线性混合模型 (Linear Mixed Models) 与 BLUP

为了达到国际标准，系统必须采用**最佳线性无偏预测 (BLUP)** 方法，通过**动物模型 (Animal Model)** 进行遗传评估。这比传统的选择指数法能更准确地剥离环境效应，利用所有亲属的信息。

基本数学模型

系统后端需构建并求解如下混合模型方程组 (MME):

$$\$ \$ \mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e} \$ \$$$

- **\$\mathbf{y}\$ (观测值向量):** 来自 MySQL 数据库的表型记录（如断奶体重、眼肌深度）。
- **\$\mathbf{b}\$ (固定效应向量):** 包括当代组（场-年-季-性别-管理组）、母羊年龄、出生类型（单羔/多羔）等非遗传因素。系统必须自动构建  $\mathbf{X}$  关联矩阵来校正这些因素。
- **\$\mathbf{u}\$ (随机效应向量):** 即我们要在求解的**育种值 (Breeding Values)**。这部分效用服从正态分布  $N(0, \mathbf{A}\sigma_u^2)$ ，其中  $\mathbf{A}$  为系谱亲缘关系矩阵。
- **\$\mathbf{e}\$ (残差向量):** 未被模型解释的随机误差。

系谱矩阵 ( $\mathbf{A}^{-1}$ ) 的构建

在 MySQL 中存储的是父-母-子三元组关系。系统需要通过递归算法遍历整个系谱树，计算个体间的亲缘系数，进而构建  $\mathbf{A}$  矩阵的逆矩阵  $\mathbf{A}^{-1}$ 。这是 BLUP 求解的关键步骤。对于大型系谱，这需要极高的计算效率，通常通过 renumf90 等工具在内存中完成，而非直接在数据库层面进行矩阵运算<sup>17</sup>。

### 3.2 单步法基因组评估 (ssGBLUP)

为了超越传统的系谱 BLUP，本系统将集成单步法基因组 **BLUP (ssGBLUP)**。这是目前国际上肉羊育种最先进的评估方法，被澳大利亚 Sheep Genetics 和美国 CDCB 广泛采用<sup>14</sup>。

核心原理

ssGBLUP 的核心在于用  $\mathbf{H}$  矩阵替代传统的  $\mathbf{A}$  矩阵。 $\mathbf{H}$  矩阵融合了基于系谱的亲缘关系 ( $\mathbf{A}$ ) 和基于 SNP 标记的基因组亲缘关系 ( $\mathbf{G}$ )。

$$\$ \$ \mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

- **$\mathbf{G}$  矩阵：**由高密度 SNP 芯片数据构建，精确度量个体间实际基因共享比例，能够捕捉到系谱记录无法反映的孟德尔采样偏差（即全同胞之间的差异）。
- **$\mathbf{A}_{22}$ ：**仅包含有基因型测定个体的系谱关系子矩阵。

系统实现路径

1. **基因型存储：**将 SNP 数据（通常为 50k 或 HD 芯片）以压缩二进制格式存储（非直接存入 MySQL 表），MySQL 仅存储元数据（文件路径、Call Rate、芯片类型）<sup>19</sup>。
2. **矩阵构建：**调用后端计算模块（如 BLUPF90 系列工具），读取 SNP 文件构建  $\mathbf{G}$  矩阵和  $\mathbf{A}_{22}$  矩阵。
3. **方程求解：**结合表型数据，求解混合模型方程，直接输出带有基因组信息的估计育种值 (GEBV)。

这种方法的优势在于，它允许未测定基因型的动物（通过  $\mathbf{A}$  矩阵部分）和测定基因型的动物同时参与评估，极大地提高了年轻公羊（尚未有后代）评估的准确性<sup>14</sup>。

### 3.3 经济权重与选择指数模型

遗传评估的最终输出不仅仅是单个性状的育种值，而是综合经济指数。系统必须内置生物经济模型模块，支持动态计算经济权重。

利润函数 (\$P\$)

系统需允许管理员定义利润函数：

$$\$ \$ P = R(\text{Revenues}) - C(\text{Costs})$$

经济权重 ( $\$EW\$$ ) 为利润对遗传改良量的偏导数:

$$\$EW_i = \frac{\partial P}{\partial G_i}$$

指数公式

$$\$I = \sum (EBV_i \times EW_i)$$

典型肉羊指数设计:

1. **终端父本指数 (Terminal Sire Index):** 侧重生长速度 (WWT, PWWT)、眼肌深度 (PEMD) 和瘦肉率 (LMY)，对脂肪 (FAT) 通常施加负权重以避免过肥<sup>21</sup>。
2. **母系指数 (Maternal Index):** 侧重繁殖力 (NLW)、泌乳力 (MWWT) 和断奶重。关键在于对成年母羊体重 (Adult Weight) 施加惩罚性权重，以控制维持饲养成本，提高饲料转化效率<sup>22</sup>。

通过在系统中预设这些复杂的数学模型，用户只需输入基础数据，系统即可自动输出具有商业指导意义的选种指数。

---

## 4. 基于 MySQL 的高级数据库架构设计

MySQL 作为一个免费开源的关系型数据库，完全有能力支撑起千万级数据的育种系统，前提是必须进行极其精细的 Schema 设计和索引优化。

### 4.1 核心实体关系模型 (ER Model)

系统的核心 ER 图设计应遵循第三范式 (3NF)，以确保数据的一致性。

#### 4.1.1 动物个体表 (animals)

这是系统的核心表，必须支持递归查询以构建系谱。

字段名	数据类型	说明	索引策略
animal_uuid	CHAR(36)	全局唯一标识符 (UUID)，主键	PRIMARY

icar_id	VARCHAR(16)	符合 ICAR 标准的国际 ID (如 FR123456789012)	UNIQUE
visual_tag	VARCHAR(20)	耳标号, 用于人工识别	INDEX
rfid_tag	VARCHAR(23)	ISO 11784 电子耳标 ID (15 位数字+填充)	INDEX
sire_uuid	CHAR(36)	父亲 UUID, 自关联外键	FK, INDEX
dam_uuid	CHAR(36)	母亲 UUID, 自关联外键	FK, INDEX
birth_date	DATE	出生日期	INDEX
sex	ENUM('M','F')	性别	
birth_rank	TINYINT	出生数 (单/双/三)	
rearing_rank	TINYINT	抚育数 (实际哺乳数)	
breed_code	JSON	品种构成 (如 {"TEX": 0.5, "SUF": 0.5})	
flock_id	INT	所属羊群/育种场 ID	FK
active_status	BIT(1)	存栏状态	

## 设计亮点:

- **UUID 主键:** 使用 UUID 而非自增 ID，便于多育种场数据的合并与分布式同步（离线优先架构的基础）。
- **JSON 品种字段:** 为了支持类似澳大利亚“联合分析”中的复合品种评估，品种不再是单一字段，而是 JSON 对象，能精确描述“50% 特克塞尔 + 25% 萨福克 + 25% 杜泊”的复杂血统<sup>1</sup>。

### 4.1.2 表型数据表 (phenotypes)

表型数据量巨大且类型繁多，采用纵向表设计（EAV 模式的变体）比横向表更具扩展性。

字段名	数据类型	说明
measurement_id	BIGINT	主键，自增
animal_uuid	CHAR(36)	外键，关联动物
trait_id	INT	外键，关联性状定义表
raw_value	DECIMAL(12,4)	原始测量值
adjusted_value	DECIMAL(12,4)	经环境效应校正后的值 (用于 BLUP)
measure_date	DATE	测定日期
age_in_days	INT	测定日龄 (虚拟生成列，自动计算)
group_id	INT	外键，关联当代组
technician_id	INT	测定员 ID (用于数据质控)

设计亮点：

- 分区 (Partitioning):** 该表将包含数百万条记录。建议按 measure\_date (年份) 或 trait\_id 进行表分区，以提高查询效率。
- 虚拟列 (Generated Column):** age\_in\_days 设为虚拟列，由 measure\_date 和 animals.birth\_date 自动计算。这消除了数据冗余，且 MySQL 8.0 支持对虚拟列建立索引，极大加速了“查询 90 日龄左右体重”这类操作<sup>23</sup>。

#### 4.1.3 当代组管理表 (contemporary\_groups)

当代组是遗传评估中环境效应校正的基石。

字段名	数据类型	说明
group_id	INT	主键
flock_id	INT	外键
group_type	ENUM	类型 (出生组、断奶组、剪毛组)
formation_date	DATE	组建日期
management_code	VARCHAR(50)	管理代码 (如 "补饲组", "放牧组")

## 4.2 系谱遍历的高级 SQL 实现

传统的 MySQL 版本难以处理无限层级的系谱树。但在 MySQL 8.0 中，我们可以利用 **递归公用表表达式 (Recursive CTEs)** 来高效查询某只种公羊的所有后代或所有祖先，这对于计算近交系数和构建  $\mathbf{A}$  矩阵至关重要。

SQL

```

WITH RECURSIVE Ancestry AS (
    SELECT animal_uuid, sire_uuid, dam_uuid, 0 AS generation
    FROM animals WHERE animal_uuid = 'TARGET_UUID'
    UNION ALL
    SELECT a.animal_uuid, a.sire_uuid, a.dam_uuid, anc.generation + 1
    FROM animals a
    INNER JOIN Ancestry anc ON a.animal_uuid = anc.sire_uuid OR a.animal_uuid = anc.dam_uuid
)
SELECT * FROM Ancestry;

```

这种原生的递归能力使得 MySQL 能够胜任以往需要图数据库（Graph DB）才能完成的复杂系谱分析任务。

## 4.3 基因组数据存储策略

基因组数据（SNP）具有维度高（50,000+ 列）、文件大、查询模式单一（通常是全量读取用于矩阵构建）的特点。因此，**绝对不应将 SNP 数据逐个位点存入 MySQL 表中**。

**最佳实践策略：**

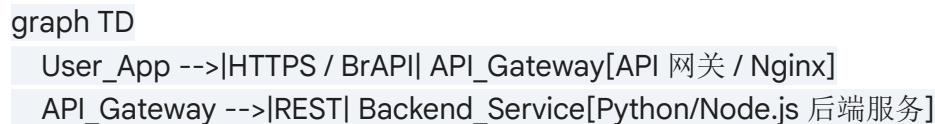
- 文件存储：** 将原始 VCF 文件或 PLINK 二进制文件（.bed,.bim,.fam）存储在对象存储系统（如 MinIO，兼容 S3 协议）或本地高性能文件系统中。
- 元数据索引：** MySQL 中的 `genotypes` 表仅存储文件路径、校验和 (Hash)、芯片类型、检出率 (Call Rate) 等元数据<sup>19</sup>。
- 计算流：** 当进行 ssGBLUP 评估时，计算引擎根据 MySQL 中的索引直接读取二进制文件构建  $\mathbf{H}$  矩阵，完全绕过数据库的 SQL 解析层，从而实现极高的 I/O 吞吐量。

## 5. 软件架构、API 接口与离线优先策略

为了实现系统的可扩展性和国际互操作性，软件架构必须遵循现代微服务和 API 优先的原则。

### 5.1 总体架构图

代码段



```
Backend_Service -->|SQL| MySQL_Master
```

```
Backend_Service -->|SQL| MySQL_Slave
```

```
subgraph "遗传评估黑盒"
```

```
    Backend_Service -->|任务指令| Job_Queue
```

```
    Job_Queue -->|消费| Compute_Node[计算节点]
```

```
    Compute_Node -->|读取| Flat_Files[系谱/表型文件]
```

```
    Compute_Node -->|调用| BLUPF90
```

```
    BLUPF90 -->|输出| Solutions[育种值文件]
```

```
    Compute_Node -->|写回| MySQL_Master
```

```
end
```

## 5.2 实施 BrAPI (Breeding API) 标准

要达到国际顶尖水平，系统必须支持 BrAPI V2.1 标准。这是全球植物和动物育种领域的通用语言，使得我们的系统可以无缝对接国际通用的育种分析工具（如 R 语言的各种遗传包）或可视化软件<sup>12</sup>。

核心接口设计：

- 种质资源 (Germplasm):
  - GET /brapi/v2/germplasm: 获取羊只列表 (UUID、系谱、品种)。
  - POST /brapi/v2/germplasm: 注册新出生的羔羊。
- 表型观测 (Observations):
  - GET /brapi/v2/observations: 获取特定羊只或试验的体重、扫描数据。
  - POST /brapi/v2/observations: 批量上传测量数据。
- 育种方案 (Programs):
  - GET /brapi/v2/programs: 定义不同的育种项目（如“杜泊羊核心群选育”）。

通过实现 BrAPI，第三方农场管理软件（如 FarmOS）可以直接将数据推送到我们的育种系统，而无需开发专门的接口适配器，极大地扩展了系统的生态兼容性。

## 5.3 离线优先 (Offline-First) 的前端技术栈

牧场环境通常网络覆盖较差。一个顶级的系统必须保证数据采集在无网络状态下依然流畅、安全。

技术选型：

- 前端数据库： 使用 RxDB (Reactive Database) 或 PouchDB。这些是运行在浏览器端的 NoSQL 数据库，基于 IndexedDB 技术<sup>15</sup>。
- 同步机制： 采用“复制”策略。当用户在平板电脑上录入数据时，数据首先写入本地 RxDB，

此时无需网络。一旦设备检测到网络连接，RxDB 会自动在后台与服务器端的数据库（如通过 CouchDB 协议或自定义 REST 接口）进行增量同步。

- **冲突解决：**系统需内置“最后写入优先”或“人工审核”的冲突解决策略。例如，如果两个技术人员同时修改了同一只羊的体重，系统应保留时间戳最新的记录，并标记为“待审核”供管理员确认。

优势：

这种架构消除了加载时的“转圈”等待，用户体验极佳，且保证了数据在任何网络环境下都不会丢失<sup>15</sup>。

## 5.4 硬件集成：Web Serial API

为了提高数据录入的准确性和效率，前端应用应利用 HTML5 的 **Web Serial API** 直接连接蓝牙或 USB 电子秤和 RFID 阅读器<sup>25</sup>。

- **场景：**育种员在称重时，浏览器网页直接读取电子秤发送的重量数据流，并自动填入输入框，同时读取 RFID 耳标号自动定位羊只。
- **实现：**无需安装驱动，通过 JavaScript 即可控制串口通信，实现“扫描-称重-保存”的毫秒级自动化流程。

---

## 6. 表型数据采集标准化与质量控制体系

数据质量是育种系统的生命线。无论模型多么先进，如果输入的是垃圾数据（GIGO），输出的必定是错误的育种值。

### 6.1 关键性状采集窗口 (基于 ICAR 规范)

系统必须在软件层面强制执行 ICAR Section 21 定义的采集窗口，超出窗口的数据应被标记警告或拒绝<sup>27</sup>。

性状代码	性状名称	推荐采集日龄 窗口	标准化日龄	采集要求
BWT	出生重	0 - 24 小时	0	必须记录出生类型和母羊年龄
WWT	断奶重	42 - 120 天	90	全群同日称重，记录断奶

				管理组
<b>PWWT</b>	断奶后重	120 - 210 天	160	此时体现生长潜力的遗传分化
<b>PEMD</b>	眼肌深度	120 - 300 天	160	需使用 B 超在第 12/13 肋骨处测量
<b>PFAT</b>	背膘厚度	120 - 300 天	160	同上, C 点测量
<b>MWWT</b>	成年母羊重	每年配种前	N/A	用于评估维持需要

## 6.2 自动化质量控制 (QC) 算法

系统后端在数据入库前应运行自动 QC 脚本：

- 生物学合理性检查：**例如，如果一只 3 月龄羔羊体重录入为 80kg（可能是 8.0kg 的笔误），系统应根据正态分布（Mean  $\pm$  3SD）自动拦截。
- 当代组有效性检查：**系统必须检查每个当代组的变异系数。如果一个组内所有羊的体重都完全一样（如都是 30.0kg），这通常意味着数据造假或设备故障，该组数据应被剔除<sup>1</sup>。
- 系谱逻辑检查：**检查个体出生日期是否晚于其父母，父母的年龄是否在合理的繁殖范围内（如公羊 > 5 月龄，母羊 > 7 月龄）。

## 6.3 屠宰数据反馈闭环

为了真正实现“肉羊”育种，系统必须打通与屠宰场的数据接口。

- 胴体数据导入：**设计专门的 API 接口接收屠宰流水线的 CSV 数据，包括热胴体重 (HSCW)、GR 值（总组织深度）和肉质评分。
- 关联匹配：**系统利用 RFID 电子耳标号，将屠宰场的胴体数据回溯匹配到活体羊只及其父母，从而计算出屠宰率 (Dressing %) 和净肉率的育种值。这是目前大多数系统缺失但极具价值的功能。

---

<sup>1</sup> 参见：https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5300000/

## 7. 基因组选择（Genomic Selection）的深度集成

这是本系统区别于传统系统的关键点，也是迈向“国际顶尖”的标志。

### 7.1 基因组数据流处理

1. **数据接收：** 实验室上传基因型数据（通常为 Illumina OvineSNP50 或 HD 芯片格式）。
2. **质量控制 (SNP QC)：** 系统自动调用 PLINK 工具进行质控，剔除检出率低 (<90%)、偏离哈代-温伯格平衡 (HWE) 或最小等位基因频率 (MAF) 过低的位点<sup>19</sup>。
3. **亲子鉴定 (Parentage Verification)：** 在进行遗传评估前，系统利用大约 200-500 个核心 SNP 位点自动验证系谱记录的准确性。如果基因组推断的父母与记录不符，系统自动标记“系谱冲突”并建议修正。这能显著提高  $\mathbf{A}$  矩阵的准确性<sup>20</sup>。
4. **填充 (Imputation)：** 对于使用低密度芯片（如 10k）测定的个体，系统利用核心群的高密度数据将其填充至 50k 或更高密度，以统一评估标准。

### 7.2 评估流程自动化

系统应配置为定时任务（如每两周一次），自动执行以下流程<sup>23</sup>：

1. **数据冻结：** 锁定数据库读取快照。
2. **数据提取与重编号：** 使用 `renmf90` 将 UUID 映射为整数 ID，生成参数文件。
3. **矩阵构建：** 计算  $\mathbf{A}^{-1}$  和  $\mathbf{G}^{-1}$ ，融合为  $\mathbf{H}^{-1}$ 。
4. **迭代求解：** 运行 `blupf90` 求解 MME 方程，直至收敛。
5. **结果回写：** 将计算出的 GEBV 和准确度 (Accuracy) 写回 MySQL 的 `breeding_values` 表，并更新前端展示的百分位排名。

---

## 8. 育种目标、选择指数与经济权重

系统必须将复杂的 GEBV 转化为简单的选种决策信号。

### 8.1 经济权重的动态计算

参考 Jones 等人 (2004) 和 Wolfsova 等人 (2009) 的生物经济模型，系统应允许设定不同的市场场景（如高粮价/低粮价，高肉价/低肉价）来计算边际效益<sup>28</sup>。

- 例如，**断奶重 (WWT)** 的经济权重  $EW_{WWT}$  取决于：每增加 1kg 带来的售价提升 - 增加这 1kg 所需的额外饲料成本。

### 8.2 标准化选择指数公式

系统应预设以下几类指数，并支持用户微调权重：

## 1. 终端胴体指数 (TCP - Terminal Carcass Production)

目标：最大化生长和瘦肉产出。

$$\$I_{TCP} = 1.4 \times EBV_{WWT} + 2.8 \times EBV_{PWWT} + 2.0 \times EBV_{PEMD} - 0.5 \times EBV_{PFAT}$$

注：对脂肪厚度 (PFAT) 施加负权重以防止过肥，对眼肌深度 (PEMD) 施加高正权重以增加产肉量 21。

## 2. 母系生产指数 (MCP - Maternal Carcass Production)

目标：平衡繁殖性能与生长，同时限制母羊体型。

$$\$I_{MCP} = 15.0 \times EBV_{NLW} + 5.0 \times EBV_{MWWT(Milk)} + 1.0 \times EBV_{PWWT} - 2.0 \times EBV_{AdultWT}$$

注： $EBV_{NLW}$  (断奶羔羊数) 权重极高，因为繁殖率是母系最关键的利润驱动力；  
 $EBV_{AdultWT}$  为负权重，旨在保持母羊维持成本不增加 22。

---

## 9. 系统部署、运维与未来展望

### 9.1 部署架构

建议采用混合云架构：

- **Web 服务器与 MySQL 数据库：** 部署在云服务器（如阿里云/AWS），利用云数据库的高可用性 (HA) 和自动备份功能。
- **计算节点：** 遗传评估 (ssGBLUP) 计算量大，建议使用高性能计算集群 (HPC) 或按需启动的 GPU 实例。通过 Redis 消息队列将评估任务从 Web 服务器分发到计算节点。
- **对象存储：** 所有的 SNP 文件和超声波图像存储在 OSS/S3 中，实现存算分离。

### 9.2 实施路线图

1. 阶段一（基础层）：搭建 MySQL 数据库，实现符合 ICAR 标准的系谱与表型管理，开发离线优先的 PWA 采集端。
2. 阶段二（评估层）：集成 BLUPF90，实现常规系谱 BLUP 评估，发布基础 EBV。
3. 阶段三（基因组层）：引入 SNP 数据管理模块，升级为 ssGBLUP，实现亲子鉴定自动化。
4. 阶段四（生态层）：开放 BrAPI 接口，对接自动饲喂站、屠宰场数据，构建全产业链大数据平台。

## 9.3 结论

本报告所设计的系统，通过融合 MySQL 的数据管理能力、BLUPF90 的顶尖遗传算法、BrAPI 的国际互通标准以及现代 Web 的离线优先体验，构建了一个不仅在成本上可控，而且在技术上完全能够对标甚至在某些架构理念上超越 LAMBPLAN 和 NSIP 的肉羊育种平台。它将把“经验育种”彻底转化为“数字育种”和“精确育种”，为肉羊产业的高质量发展提供坚实的数字底座。

## 引用的著作

1. Combined LAMBPLAN | Sheep Genetics, 访问时间为十二月 14, 2025, <https://www.sheepgenetics.org.au/globalassets/sheep-genetics/updates/analysis-enhancements/ae25/analysis-enhancements-booklet-2025---combined-lambplan.pdf>
2. BEST PRACTICE - SIL - Sheep Improvement Limited, 访问时间为十二月 14, 2025, [https://www.sil.co.nz/files/1489435846\\_Best%20Practice%20Guide%20Dec%202016\\_WEB.pdf](https://www.sil.co.nz/files/1489435846_Best%20Practice%20Guide%20Dec%202016_WEB.pdf)
3. What is LAMBPLAN - NSW Department of Primary Industries, 访问时间为十二月 14, 2025, <https://www.dpi.nsw.gov.au/animals-and-livestock/sheep/breed-select/lambplan/what-is-lambplan>
4. Sheep Genetics Combined LAMBPLAN Update 2025 | Neogen, 访问时间为十二月 14, 2025, <https://neogenaustralasia.com.au/combined-lambplan-updates-sheep-genetics/>
5. National Sheep Improvement Program: Performance-based data you can rely on, 访问时间为十二月 14, 2025, <https://extension.sdstate.edu/national-sheep-improvement-program-performance-based-data-you-can-rely>
6. Software for managing data | Sheep Genetics, 访问时间为十二月 14, 2025, <https://www.sheepgenetics.org.au/getting-started/software-for-managing-data/>
7. Thank you for your interest in submitting data to the National Sheep Improvement Program (NSIP). As the sheep industry strives t, 访问时间为十二月 14, 2025, <https://bowridge.davis.wvu.edu/files/d/f9acbe2c-c5e9-4913-a24a-995f63027744/nsip-data-submission-guide.pdf>
8. National Sheep Improvement Program, 访问时间为十二月 14, 2025, <https://projects.sare.org/media/pdf/N/S/I/NSIP-brochure.pdf>
9. Tables de référence du Système d'Information de l'Identification Animale (BDNI) - Idele, 访问时间为十二月 14, 2025, <https://idele.fr/detail-article/tables-de-reference-du-systeme-d-information-de-l-identification-animale-bdni>
10. French White - Races Ovines des Massifs, 访问时间为十二月 14, 2025, <https://www.races-ovines-des-massifs.com/documents/fichiers-langues/french-white-anglais.pdf>
11. Genetic diversity and selection signatures in sheep breeds - PMC - NIH, 访问时间为十二月 14, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC12367903/>

12. BrAPI (Breeding API) — Field Book 5.4 documentation, 访问时间为十二月 14, 2025, <https://field-book.readthedocs.io/en/latest/brapi.html>
13. BrAPI V1.3 · Apiary, 访问时间为十二月 14, 2025, <https://brapi.docs.apiary.io/>
14. Implementation of single-step genomic BREEDPLAN evaluations in Australian beef cattle, 访问时间为十二月 14, 2025, <https://breedplan.une.edu.au/media/wddhpjol/implementation-single-step-genomic-breedplan-evaluations-australian-beef-cattle.pdf>
15. Local First / Offline First | RxDB - JavaScript Database, 访问时间为十二月 14, 2025, <https://rxdb.info/offline-first.html>
16. RxDB – The Ultimate Offline Database with Sync and Encryption, 访问时间为十二月 14, 2025, <https://rxdb.info/articles/offline-database.html>
17. Data preparation with RENUMF90 - Yutaka Masuda, 访问时间为十二月 14, 2025, [https://masuday.github.io/blupf90\\_tutorial/renum\\_basic.html](https://masuday.github.io/blupf90_tutorial/renum_basic.html)
18. Genetic Evaluation System | The Bullvine, 访问时间为十二月 14, 2025, <https://www.thebullvine.com/category/genetic-evaluation-system/>
19. readme.pregsf90 [BLUPF90] - Animal Breeding and Genetics Group - UGA, 访问时间为十二月 14, 2025, <https://nce.ads.uga.edu/wiki/doku.php?id=readme.pregsf90>
20. Genomics - Sheep Genetics, 访问时间为十二月 14, 2025, <https://www.sheepgenetics.org.au/resources/genomics/>
21. Terminal indexes - Sheep Genetics, 访问时间为十二月 14, 2025, [https://www.sheepgenetics.org.au/globalassets/sheep-genetics/getting-started/sheep-genetics---asbvs-and-indexes/mla-sg-indexes-breeder-terminal-1122\\_v02.pdf](https://www.sheepgenetics.org.au/globalassets/sheep-genetics/getting-started/sheep-genetics---asbvs-and-indexes/mla-sg-indexes-breeder-terminal-1122_v02.pdf)
22. Maternal indexes - Sheep Genetics, 访问时间为十二月 14, 2025, <https://www.sheepgenetics.org.au/globalassets/sheep-genetics/getting-started/sheep-genetics---asbvs-and-indexes/maternal-indexes---breeder-2022.pdf>
23. Submitting data to Sheep Genetics, 访问时间为十二月 14, 2025, <https://www.sheepgenetics.org.au/resources/breeders-guide/submitting-data-to-sheep-genetics/>
24. Breeding API (BrAPI), a standardized RESTful Web Service API Specification for plant breeding data, that enables exchanging data across systems and applications - CGSpace, 访问时间为十二月 14, 2025, <https://cgspace.cgiar.org/items/56e2f65d-9db6-4a21-89e0-5a7c12d6b481>
25. Getting started with the Web Serial API - Google Codelabs, 访问时间为十二月 14, 2025, <https://codelabs.developers.google.com/codelabs/web-serial>
26. Web Serial API - Web APIs - MDN Web Docs - Mozilla, 访问时间为十二月 14, 2025, [https://developer.mozilla.org/en-US/docs/Web/API/Web\\_Serial\\_API](https://developer.mozilla.org/en-US/docs/Web/API/Web_Serial_API)
27. Guidelines – ICAR, 访问时间为十二月 14, 2025, <https://www.icar.org/guidelines/>
28. Calculating economic weights for sheep sire breeds used in different breeding systems, 访问时间为十二月 14, 2025, <https://academic.oup.com/jas/article->

[pdf/89/6/1698/23633610/1698.pdf](https://www.researchgate.net/publication/274695392_Calculating_economic_weights_for_growth_reproduction_and_wool_traits_in_Makui_sheep_breed_by_Ecoweight_software)

29. Calculating economic weights for growth, reproduction and wool traits in Makui sheep breed by Ecoweight software - ResearchGate, 访问时间为十二月 14, 2025,  
[https://www.researchgate.net/publication/274695392 Calculating economic weights for growth reproduction and wool traits in Makui sheep breed by Eco weight software](https://www.researchgate.net/publication/274695392_Calculating_economic_weights_for_growth_reproduction_and_wool_traits_in_Makui_sheep_breed_by_Eco_weight_software)
30. Creating an index to breed sheep in Terminal Sire Breeds with a low carbon footprint, 访问时间为十二月 14, 2025, <https://ahdb.org.uk/sheep-enviro-indexes-creating-an-index-to-breed-sheep-in-terminal-sire-breeds-with-a-low-carbon-footprint>