

問3 農業用機器メーカーによる観測データ分析システムの SQL 設計、性能、運用に関する次の記述を読んで、設問に答えよ。

ハウス栽培農家向けの農業用機器を製造・販売する B 社は、農家の DX を支援する目的で、RDBMS を用いたハウス栽培のための観測データ分析システム（以下、分析システムという）を構築することになり、運用部門の C さんが実装を担当した。

[業務の概要]

1. 顧客、^{ほじょう}圃場、農事日付
 - (1) 顧客は、ハウス栽培を行う農家であり、顧客 ID で識別する。
 - (2) 圃場は、農家が農作物を育てる場所の単位で、圃場 ID で識別する。圃場には一つの農業用ハウス（以下、ハウスという）が設置され、トマト、イチゴなどの農作物が1種類栽培される。
 - (3) 圃場の日出時刻と日没時刻は、圃場の経度、緯度、標高によって日ごとに変わるが、あらかじめ計算で求めることができる。
 - (4) 日出時刻から翌日の日出時刻の1分前までとする日付を、農事日付という。農家は、農事日付に基づいて作業を行うことがある。
2. 制御機器・センサー機器、統合機器、観測データ、積算温度
 - (1) 圃場のハウスには、ハウスの天窗の開閉、カーテン、暖房、^{かんすい}灌水などを制御する制御機器、及び温度（気温）、湿度、水温、地温、日照時間、炭酸ガス濃度などを計測するセンサー機器が設置される。
 - (2) 顧客は、圃場の一角に設置した B 社の統合環境制御機器（以下、統合機器という）を用いて、ハウス内の各機器を監視し、操作する。もし統合機器が何か異常を検知すれば、顧客のスマートフォンにその異常を直ちに通知する。
 - (3) 統合機器は、各機器の設定値と各センサー機器が毎分計測した値を併せて記録した1件のレコードを、B社の分析システムに送り、蓄積する。分析システムは、蓄積されたレコードを観測データとして分析しやすい形式に変換し、計測された日付ごと時分ごと圃場ごとに1行を“観測”テーブルに登録する。
 - (4) 農家が重視する積算温度は、1日の平均温度をある期間にわたって合計したもので、生育の進展を示す指標として利用される。例えば、トマトが開花して

から完熟するまでに必要な積算温度は、1,000~1,100℃といわれている。

(5) 分析システムの目標は、対象にする圃場を現状の 100 圃場から段階的に増やし、将来 1,000 圃場で最長 5 年間の観測データを分析できることである。

[分析システムの主なテーブル]

C さんが設計した主なテーブル構造を図 1 に、主な列の意味・制約を表 1 に示す。

また、“観測” テーブルの主な列統計、索引定義、制約、表領域の設定を表 2 に示す。

顧客 (顧客 ID, 顧客名, 連絡先情報, …)
圃場 (圃場 ID, 圃場名, 顧客 ID, 緯度, 経度, 標高, …)
圃場カレンダー (標準日付, 圃場 ID, 日出時刻, 日没時刻, 日出方位角, 日没方位角)
観測 (観測日付, 観測時分, 圃場 ID, 農事日付, 分平均温度, 分日照時間, 機器設定情報, …)

図 1 テーブル構造 (一部省略)

表 1 主な列の意味・制約

列名	意味・制約
標準日付	1 日の区切りを、0 時 0 分 0 秒から 23 時 59 分 59 秒までとする日付
観測日付, 観測時分	圃場内の各種センサーが計測したときの標準日付と時分。時分は、0 時 0 分から 23 時 59 分までの 1 分単位
農事日付	1 日の区切りを、圃場の日出時刻から翌日の日出時刻の 1 分前までとする日付
分平均温度	ハウス内の温度 (気温) の 1 分間の平均値

表 2 “観測” テーブルの主な列統計、索引定義、制約、表領域の設定 (一部省略)

列名	列値個数	主索引 (列の定義順)	副次索引 (列の定義順)	表領域の設定
観測日付		1		表領域のページ長 : 4,000 バイト ページ当たり行数 : 4 行 / ページ
観測時分	1,440	2		
圃場 ID	1,000	3	1	
農事日付			2	
...	...			
制約	外部キー制約	FOREIGN KEY (観測日付, 圃場 ID) REFERENCES 圃場カレンダー (標準日付, 圃場 ID) ON DELETE CASCADE		

注記 網掛け部分は表示していない。

[RDBMS の主な仕様]

1. 行の挿入・削除, 再編成

- (1) 行を挿入するとき、表領域の最後のページに行を格納する。最後のページに空き領域がなければ、新しいページを表領域の最後に追加し、行を格納する。
- (2) 最後のページを除き、行を削除してできた領域は、行の挿入に使われない。
- (3) 再編成では、削除されていない全行をファイルにアンロードした後、初期化した表領域にその全行を再ロードし、併せて索引を再作成する。

2. 区分化

- (1) テーブルごとに一つ又は複数の列を区分キーとし、区分キーの値に基づいて表領域を物理的に分割することを、区分化という。
- (2) 区分方法には次の2種類がある。
 - ・レンジ区分：区分キーの値の範囲によって行を区分に分配する。
 - ・ハッシュ区分：区分キーの値に基づき、RDBMS が生成するハッシュ値によって行を一定数の区分に分配する。区分数を変更する場合、全行を再分配する。
- (3) レンジ区分では、区分キーの値の範囲が既存の区分と重複しなければ区分を追加でき、任意の区分を切り離すこともできる。区分の追加、切離しのとき、区分内の行のログがログファイルに記録されることはない。
- (4) 区分ごとに物理的に分割される索引（以下、分割索引という）を定義できる。区分を追加したとき、当該区分に分割索引が追加され、また、区分を切り離したとき、当該区分の分割索引も切り離される。

[観測データの分析]

1. 観測データの分析

分析システムは、農家の要望に応じて様々な観点から観測データを分析し、その結果を農家のスマートフォンに表示する予定である。Cさんが設計した観測データを分析するSQL文の例を表3のSQL1に、結果行の一部を後述する図2に示す。

表3 観測データを分析するSQL文の例（未完成）

SQL	SQL文の構文（上段：目的，下段：構文）
SQL1	圃場ごと農事日付ごとに1日の平均温度と行数を調べる。
	WITH R (圃場 ID, 農事日付, 日平均温度, 行数) AS (SELECT a , COUNT(*) FROM 観測 GROUP BY b) SELECT * FROM R

2. SQL 文の改良

顧客に表 3 の SQL1 の日平均温度を折れ線グラフにして見せたところ、知りたいのは日々の温度の細かい変動ではなく、変動の傾向であると言われた。そこで C さんは、折れ線グラフを滑らかにするため、表 4 の SQL2 のように改良した。SQL2 が利用した表 3 の SQL1 の結果行の一部を図 2 に、SQL2 の結果行を図 3 に示す。

表 4 改良した SQL 文

SQL	SQL 文の構文 (上段: 目的, 下段: 構文)
SQL2	指定した圃場と農事日付の期間について、日ごとの日平均温度の変動傾向を調べる。
	WITH R (圃場 ID, 農事日付, 日平均温度, 行数) AS () SELECT 農事日付, AVG(日平均温度) OVER (ORDER BY 農事日付 ROWS BETWEEN 2 PRECEDING AND CURRENT ROW) AS X FROM R WHERE 圃場 ID = :h1 AND 農事日付 BETWEEN :h2 AND :h3

注記 1 ホスト変数の h1 には圃場 ID を、h2 には期間の開始日 (2023-02-01) を、h3 には終了日 (2023-02-10) を設定する。

注記 2 網掛け部分は、表 3 の SQL1 の R を求める問合せと同じなので表示していない。

圃場 ID	農事日付	日平均温度	...	農事日付	X
〇〇	2023-02-01	9.0	...	2023-02-01	
〇〇	2023-02-02	14.0	...	2023-02-02	
〇〇	2023-02-03	10.0	...	2023-02-03	11.0
〇〇	2023-02-04	12.0	...	2023-02-04	12.0
〇〇	2023-02-05	20.0	...	2023-02-05	c
〇〇	2023-02-06	10.0	...	2023-02-06	14.0
〇〇	2023-02-07	15.0	...	2023-02-07	d
〇〇	2023-02-08	14.0	...	2023-02-08	13.0
〇〇	2023-02-09	19.0	...	2023-02-09	e
〇〇	2023-02-10	18.0	...	2023-02-10	17.0

注記 日平均温度は、小数第 1 位まで表示した。

注記 1 X は、小数第 1 位まで表示した。

注記 2 網掛け部分は表示していない。

図 2 SQL1 の結果行の一部

図 3 SQL2 の結果行 (未完成)

3. 積算温度を調べる SQL 文

農家は、栽培している農作物の出荷時期を予測するために積算温度を利用する。

Cさんが設計した積算温度を調べるSQL文を、表5のSQL3に示す。

表5 積算温度を調べるSQL文（未完成）

SQL	SQL文の構文（上段：目的，下段：構文）
SQL3	指定した農事日付の期間について、圃場ごと農事日付ごとの積算温度を調べる。
	WITH R(圃場 ID, 農事日付, 日平均温度, 行数) AS (<input type="text"/>) SELECT 圃場 ID, 農事日付, SUM(<input type="text"/> f) OVER (PARTITION BY <input type="text"/> g ORDER BY <input type="text"/> h ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW) AS 積算温度 FROM R WHERE 農事日付 BETWEEN :h1 AND :h2

注記1 ホスト変数のh1とh2には積算温度を調べる期間の開始日と終了日を設定する。

注記2 網掛け部分は、表3のSQL1のRを求める問合せと同じなので表示していない。

〔“観測”テーブルの区分化〕

1. 物理設計の変更

Cさんは、大容量になる“観測”テーブルの性能と運用に懸念をもったので、次のようにテーブルの物理設計を変更し、性能見積りと年末処理の見直しを行った。

- (1) 表領域のページ長を大きくすることで1ページに格納できる行数を増やす。
- (2) 圃場IDごとに農事日付の1月1日から12月31日の値の範囲を年度として、その年度を区分キーとするレンジ区分によって区分化する。
- (3) 新たな圃場を追加する都度、当該圃場に対してそのときの年度の区分を1個追加する。

2. 性能見積り

表5のSQL3について、表2に示した副次索引から100日間の観測データ144,000行を読み込むことを仮定した場合の読み込みに必要な表領域のページ数を、区分化前と区分化後のそれぞれに分けて見積もり、表6に整理して比較した。

表6 区分化前と区分化後の読み込みに必要な表領域のページ数の比較（未完成）

比較項目	区分化前	区分化後
ページ当たりの行数（ページ長）	4行（4,000バイト）	16行（16,000バイト）
読み込み行数	144,000行	144,000行
読み込みページ数	144,000ページ	<input type="text"/> ア ページ

3. 年末処理の見直し

5 年以上前の不要な行を効率よく削除し、表領域を有効に利用するための年末処理の主な手順を、区分化前と区分化後のそれぞれについて検討し、表 7 に整理した。

表 7 区分化前と区分化後の年末処理の主な手順の比較（未完成）

	区分化前	区分化後
期限	特になし	元日の日出時刻
手順	1. “圃場カレンダー” に翌年の行を追加する。 2. <input type="text" value="イ"/> 3. “圃場カレンダー” を再編成する。 4. <input type="text" value="ウ"/>	1. “圃場カレンダー” に翌年の行を追加する。 2. “観測” に翌年度の区分を追加する。 3. <input type="text" value="エ"/> 4. <input type="text" value="オ"/> 5. <input type="text" value="カ"/>

注記 二重引用符で囲んだ名前は、テーブル名を表す。

設問 1 「観測データの分析」について答えよ。

- (1) 表 3 中の , に入れる適切な字句を答えよ。
- (2) SQL1 の結果について、1 日の行数は、1,440 行とは限らない。その理由を 30 字以内で答えよ。ただし、何らかの不具合によって分析システムにレコードが送られない事象は考慮しなくてよい。
- (3) 図 3 中の ~ に入れる適切な数値を答えよ。
- (4) 表 5 中の ~ に入れる適切な字句を答えよ。

設問 2 「観測」テーブルの区分化について答えよ。

- (1) C さんは、区分方法としてハッシュ区分を採用しなかった。その理由を 35 字以内で答えよ。
- (2) 表 6 中の に入れる適切な数値を答えよ。
- (3) 区分化前では、副次索引から 1 行を読み込むごとに、なぜ表領域の 1 ページを読み込む必要があるか。その理由を 30 字以内で答えよ。ただし、副次索引の索引ページの読み込みについては考慮しなくてよい。
- (4) 区分化後の年末処理の期限は、なぜ 12 月 31 日の 24 時ではなく元日の日出時刻なのか。その理由を 35 字以内で答えよ。
- (5) 表 7 中の ~ に入れる手順を、それぞれ次の①~⑤

の中から一つ選べ。①～⑤が全て使われるとは限らない。ただし、バックアップの取得と索引の保守については考慮しなくてよい。

- ① “圃場カレンダー” から古い行を削除する。
- ② “圃場カレンダー” を再編成する。
- ③ “観測” から古い行を削除する。
- ④ “観測” を再編成する。
- ⑤ “観測” から古い区分を切り離す。