

# 製造実績データの検索高速化手法

勝沼 聡<sup>†</sup> 室 啓朗<sup>†</sup>

<sup>†</sup> 日立製作所研究開発グループ 〒185-8601 東京都国分寺市東恋ヶ窪 1-280

E-mail: <sup>†</sup> {satoshi.katsunuma,hb,keiro.muro.vn}@hitachi.co.jp

**あらまし** 製造現場では、IoT システムにて製造工程や 3M に関する大量の製造実績データを収集・蓄積し、トレーサビリティ管理や品質要因分析を日々実施している。しかしながら、生産の拡大でデータ量は増大し、これら業務で多用される長期の蓄積データに対する検索や、複数の工程間を跨るデータの検索にかかる時間が問題となる。本報告では、製造実績データを製造工程、ワーク、日時を組み合わせた単位でテーブル分割し、さらに同一ワークの複数製造工程をデータ統合し保持することで、検索性能を改善する方式を提案する。評価の結果、テーブル分割より最大 93%削減、またデータ結合により最大 56%の検索時間削減を確認できた。

**キーワード** 製造業 IoT, データ管理, 製造実績データ, テーブル分割, データ統合

## 1. はじめに

製造現場では IoT システムの導入に伴い、PLC(Programmable Logic Controller)等の現場のシステムから設備のセンサデータ等の実績値を収集したり、また MES(Manufacturing Execution System)で管理している工程実績データと紐づけ、製造実績データとして統合管理されている[2]。製造実績データ用いた主要業務である、トレーサビリティ管理では、不良等の要因分析のために、長期保存された蓄積データを対象としたデータ検索を実施する。同じく製造実績データを用いる、品質要因分析では、ライン全体から品質の劣化の要因となっている工程を突き止めるなど、多数工程の蓄積データを対象にデータ検索を実施する。

MES の標準データ形式である ISA-95[1]では製造実績データのモデルを定義している。ISA-95 は、汎用的なデータモデルであり、様々な製造形態、管理形態の製造現場への対応している。しかし ISA-95 は、データ格納や検索方法を規定していない。

製造実績データは、出荷後の製品のトレーサビリティ管理のために 10 年以上の長期保存が求められている。また長期保存に加え、工場の生産規模の拡大に伴い、蓄積される製造実績データは増大する一方となっている。データ量が増大すると、テーブルのデータサイズが増大し検索性能が劣化する。また多数工程データなど、広範囲なデータを取得する際にも、データの結合処理コストが大きくなり、検索性能が劣化する。その結果、分析や調査の長期化による機会損失や生産性低下の問題が発生する。従って製造実績データの増大に拠らず検索を高速に実行することが求められている。

本研究では、データの長期保存に対応するために、製造実績データの検索に用いる製造工程、ワーク、日時を組み合わせた単位でテーブルを水平分割することでデータ量増加時の検索性能を維持する。また同一ワ

ークの複数の製造工程のデータを、予めワーク ID をキーにデータ結合し保持する垂直統合により、多数工程データ抽出時の検索性能を向上する。

本稿の構成は以下のようになっている。まず 2 章で、製造実績データ管理に求められる要件及び、従来手法の課題について述べる。3 章で、前記要件を満たす製造実績データ管理手法を提案する。4 章で、提案方式の性能評価結果を述べる。5 章で関連研究を述べ、最後に 6 章でまとめと今後の課題を述べる。

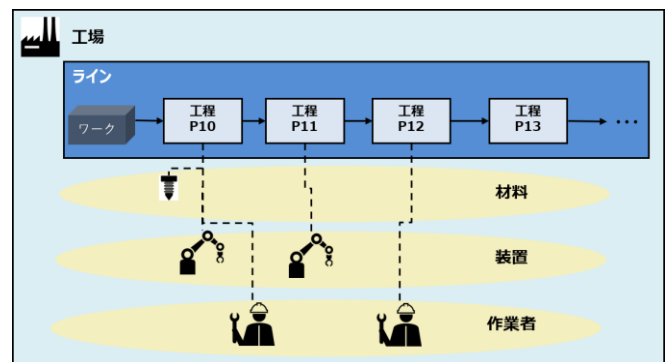


図 1：製造現場のモデル

## 2. 製造実績データ管理における課題

### 2.1. 製造実績データ

製造実績データとは、工場の製造現場における工程及び、製造の基本要素である 3M(Man/Machine/Material)に関する実績値である。MES の標準データ形式である ISA-95 で 3M の実績値を定義しており、また近年では 3M に方法(Method)を加えた 4M のデータを管理対象とすることも多い[2]。本稿では、ISA-95 に併せて以降では 3M として説明する。

前提とする製造現場のモデルを図 1 に示す。以下ではライン生産方式を想定する。工場の製造ラインには複数の工程(P10～P13)が順に並んでおり、ワークを順

に処理する。ワークとは、製造工程にある原料、仕掛品、完成品の総称であり、個品生産の場合はワークは個品単位であり、ロットやバッチ生産の場合は、ロットやバッチ単位となる場合もある。各工程では、加えられている材料や、製造に用いる装置、担当している作業員など、3Mが紐づけられる。

製造実績データは、前述の工程や3Mにおける実績値である。製造実績データのイメージ図を図2に示す。製造実績データは、各工程の製造日時や、工程ID、処理するワークID、合・不合格などの工程の実績値を持つ。また工程に紐づく3Mの実績値を持つ。3Mの実績値の種類は、工程毎に異なり、多種多様であり、例えば、装置のセンサ値や、材料の型番、作業員の名前や作業記録などがある。また製造実績データの検索方法としては、単一ワークの実績データを抽出(S#1)、単一工程のq特定期間の実績データの抽出(S#2)、複数工程の特定期間の実績データ抽出(S#3)、工程や3Mの実績値を条件にした実績データ抽出(S#4)が挙げられる。

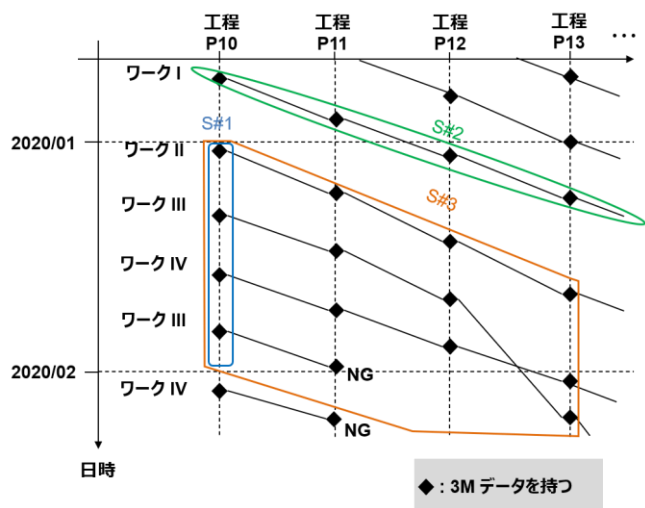


図 2：製造実績データ

### 2.1.1. 製造実績データのユースケース

本節では、前節で述べた製造実績データを活用するユースケースを述べる。まず複数ユースケースで共有する製造実績データの検索方法を述べ、次に個別ユースケース UC#1～3 について説明する。

#### □ 製造実績データの検索方法

表 1 に典型的な製造実績データの検索方法 S#1～S#3 を示す。表 1 には各検索方法における抽出対象の実績データとその抽出条件及び、後述の、対応するユースケース及び、提案方式での対策方針を示す。

S#1 は、単一のワーク ID を指定して、そのワークを実行した全工程の実績データを取得する検索である。

S#2 は、単一の工程と期間を指定して、指定した工程における指定期間における複数ワークの実績データを取得する検索である。S#3 は複数ワークの複数工程の実績データを取得する検索である。複数工程と期間を指定して、指定した一つの工程の、指定した期間の実績データと、ワーク ID で紐づく他工程の実績データを取得する。

表 1：製造実績データの検索方法

ID	抽出データ	ユースケース			テーブル	
		#1	#2	#3	分割	統合
S#1	同一ワークの複数工程実績データ	x	x		x	
S#2	単一工程の特定期間のデータ	x	x		x	
S#3	単一工程と特定期間のデータと、そのデータにワーク ID で紐づく他工程のデータ	x	x	x	x	x

#### □ UC#1：製造及び製品管理

日々の製造業務において、従来、紙面や Excel 等で管理していたデータを、蓄積した製造実績データを用いてシステムで閲覧可能としたり、グラフ等で可視化することで業務を効率化する。工程管理のために、各工程で製造した仕掛品の良・不良の確認や、その工程の設備のセンサ値やそのトレンドを管理する。また出荷時の最終製品管理のために、該当製品の各工程における製造実績データの履歴を確認する。

本ユースケースに必要な製造実績データの保存期間は、製品の出荷前迄のデータが必要であるため、各工場の製品リードタイムに依存するが1年以内であることが多い。データの抽出条件としては、工程管理においては、担当の工程と、その抽出期間を指定してデータを抽出する(S#2)。また担当工程だけでなく前後の数工程含めて抽出することがある。前後の工程を取得する場合には、担当工程と紐づけて確認するため、担当工程と前後の工程の実績データ取得する(S#3)。最終製品管理において、該当最終製品に相当するワークを指定して、該当ワークの各工程の実績データを抽出する(S#1)。

#### □ UC#2：製品出荷後のトレーサビリティ管理

製品出荷後のトレーサビリティ管理のために製造実績データを保存する。顧客からの場外不良の問合せ時など、リコール対応のために、該当製品の実績データを調査し、また必要に応じて、同時期や同装置で製造した他製品なども調査する。

必要な製造実績データの期間は、製品出荷後の管理のため、他ユースケースと比較して長期間に渡ることが特徴である。工場によっては 10 年間保存など実施される。データの抽出条件としては、不良の可能性がある製品のワークを指定してその各工程の実績データを取得したり、特定工程、場合によっては複数の工程に対し、不良製品が出た期間を指定し、他のワークの実績データを抽出する(S#1～S#3)。

#### □ UC#3：品質要因分析

製造ラインの垂直立上げや、製造カイゼン活動として、設計部門や、製造部門の班長レベルの人により製造実績データを用いた品質要因分析が実施される。品質改善のために不良の原因となっている工程を突き止めるために、製造ライン内の全工程のデータを用いて、相関を取るなどの分析が実施される。

工場によっては定期的に製造ラインの構成変更が実施され、また基本的には直近発生した問題に対処するため、分析のために用いる実績データは直近 1 年以内であることが多い。データ抽出条件としては、他のユースケースと同様にワークをキーに複数工程の実績データを紐づけたデータが必要となる(S#3)。但し本ユースケースでは、製造ライン全体の工程等、抽出する工程数が多いことが特徴である。また統計分析も実施するため、一定量のデータが必要となる。そのため抽出期間も 1 カ月以上など長期間となることが多い。

#### 2.1.2. 製造実績データ管理の要件

本節では、製造実績データ活用のユースケースから得られた、製造実績データ管理に求められる要件を述べる。

##### □ 要件 1：膨大な蓄積データに対する検索性能維持

トレーサビリティ管理(UC#1)では、長期間保存したデータに対する検索が求められる。検索対象として、製造ライン全体や、複数製造ラインを跨る工場全体、或いは複数工場を跨る原料から製品までのサプライチェーン全体の実績データが必要となる場合がある。そのため保存年数の増大や、生産規模の拡大により、蓄積データ量が膨大になる。そしてそのような膨大な蓄積データ量のデータに対し、S#1～S#3 のワークや期間、工程を指定した検索の性能を維持する必要がある。

##### □ 要件 2：多工程、長期間に渡る広範囲なデータの抽出の検索性能維持

品質要件分析(UC#2)のために、製造ライン全体など、多数工程の、1 カ月以上の長期間の製造実績データが求められる。そのため、S#3 の検索において、特に多

数工程、長期間のデータ抽出に対して検索性能を維持する必要がある。

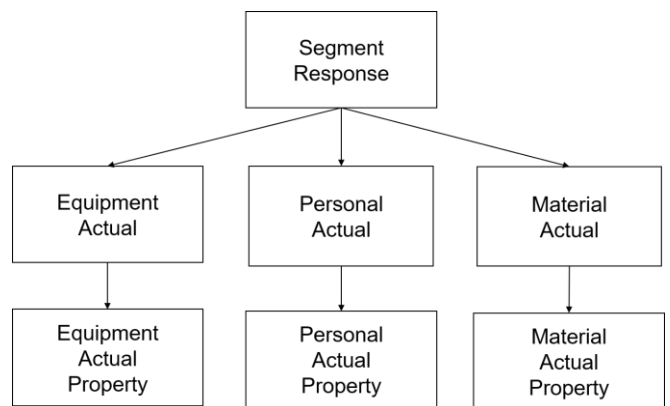


図 3：ISA-95 における製造実績データのモデル

#### 2.2. 従来の製造実績データ管理手法の課題

従来の製造データ管理手法として、MES の標準データ形式である ISA-95 における製造実績データモデル(図 3)について述べる。

##### 2.2.1. ISA-95

ISA-95 では、製造実績データを、製造実行管理 MOM(Manufacturing Operations Management)における Operation Performance データとして定義される。Operation Performance データでは、Segment Response データに、工程実績データを保持し、3M すなわち、設備、材料、作業者の実績データは、各々 Equipment Actual, Personnel Actual, Material Actual データとして保持する。そして Segment Response データに各 Actual データのリンクを持たせる。また設備のセンサデータ値等の 3M のパラメータ値は、Equipment Actual Property データなどの Property データにパラメータの ID と値のセットで保持する。そして Equipment Actual データ等の Actual データにその Property データのリンクを持たせる。

ISA-95 は、様々な製造形態や管理形態に対応できる標準データモデルとして定義されている。しかしデータの管理やその検索方法を提供していない。

ISA-95 の各データをそのままデータベースのテーブルとして実装した場合、Segment Response のデータには全工程の実績数分のデータが保持される。また Equipment Actual のデータは全設備の実績数分のデータを持ち、Equipment Actual Property は設備の各パラメータの実績値分のデータが格納される。工程や 3M の実績データのサイズは、保存年数や生産規模の拡大に比例して増大するため、実績データのサイズが増大した場合に、要件 1 で挙げた検索性能を維持することは難しい。また Segment Response に工程の実績データを

保持し、複数工程の実績データを紐づける場合には、Segment Response とリンクする Material Actual データからワーク ID を抽出する。そしてワーク ID で異なる Segment Response データを紐づける JOIN 処理が必要となる。そのため要件 2 で述べた広範囲のデータ抽出時に JOIN 処理コストが増加し、性能を維持するのは難しい。

### 3. 製造実績データ管理手法の提案

本章では、2.1.2 節で述べた要件 1, 2 を満たす製造実績データ管理手法を提案する。

#### 3.1. 要件への対応方針

本研究で提案する製造実績データ管理手法では、要件 1 を満たすために、長期保存する製造実績データに対し、異なるレコードを複数のテーブルに分割する、テーブル分割を実施する。また要件 2 を満たすために、直近の製造実績データを抽出し、複数のレコードを統合するデータ統合を実施する。

前提として、図 4 に示すように、製造実績データを格納するテーブルでは、製造時刻と、製造したワーク ID、工程 ID 及び、その 3M の実績値を保持する。テーブル分割では、長期保存した大量データに対して、前述のテーブルを、検索キーに従って分割することで、要件 1 で挙げた、大量データ格納時の検索性能を維持する。またデータ統合では、単一ワークの複数工程の実績データを予め JOIN しテーブルに格納することで、要件 2 で挙げた、多工程、長期間に渡る広範囲なデータの検索性能を維持する。

以下、各方針について詳細に述べる。

Time	Work piece	Process	3M data #1	...	3M data #10
2019/9/3 9:10:10	III	P10	"20.3"	...	"Tanaka"
2019/9/3 9:20:05	III	P11	"True"	...	"0.00021"
2020/9/3 9:11:05	IV	P10	"21.0"	...	"Yamada"
2020/9/3 9:20:55	IV	P11	"False"	...	"0.00022"

図 4：製造実績データ格納テーブル

#### 3.2. テーブル分割

##### 3.2.1. ワークと工程の異なる軸で分割

テーブル分割では、表 1 の S#1～S#3 の検索を想定する。そしてテーブル分割の基本方針として、図 5 に示すように、テーブルを複製し、ワークと工程の異なる軸で分割したテーブルを用意する。そしてワーク単位で分割したテーブルを対して、ワークをキーに検索する S#1 を実施する。また工程単位で分割したテーブルに対して、工程をキーに検索する S#2, S#3 を実施する。

ワークと工程の異なる軸でテーブル分割するため、テーブルの総データサイズは約 2 倍となるが、ワークと工程のいずれか一方でテーブルを分割した場合には、もう片方のデータをキーに検索した場合、検索性能が劣化するため、二つの軸で分割する。

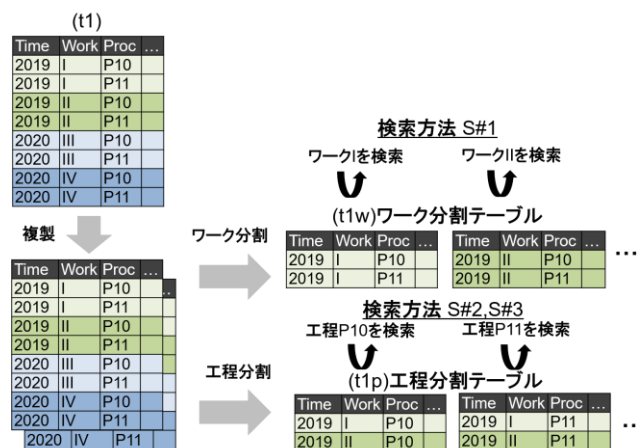


図 5：ワーク及び工程軸でのテーブル分割

	工程 P10	工程 P11	...
2020年	(t1p.P10.2020) Time Work Proc ... 2020 III P10 2020 IV P10 2020 V P10	(t1p.P11.2020) Time Work Proc ... 2020 III P11 2020 IV P11 2020 V P11	
2019年	(t1p.P10.2019) Time Work Proc ... 2019 I P10 2019 II P10	(t1p.P11.2019) Time Work Proc ... 2019 I P11 2019 II P11	
⋮			

図 6：工程と日時を組み合わせたテーブル分割

##### 3.2.2. 製造年と組み合わせた分割

テーブル分割では、工程に加えて製造年を組み合わせたテーブル分割を実施する。

図 6 に示すように、工程単位で分割するテーブルは、工程と年によりテーブル分割し、同一年の同一工程の製造実績データを同一のテーブルに割り当てる。例えば図 6 (t1p. P10. 2020) のテーブルには、2020 年の工程 P10 のデータを割り当てる。工程と期間で検索する S#2, S#3 では、年でテーブルを分割することで検索範囲をさらに絞ることが可能である。

#### 3.3. データ統合

テーブル分割では、単一工程の実績データを単一行に格納するため、S#3 の複数工程データを抽出する検索時には、ワーク ID をキーに複数工程の実績データの JOIN が必要となる。そのため工程数やその抽出期間が増えると、JOIN 対象のデータ数が増え、CPU への



負荷が増加し検索性能が劣化する。そのためデータ統合では、予めワーク ID で同一の製造ラインに属する複数工程の実績データを JOIN したデータをテーブル(以下、複数工程データ統合テーブル)に格納することで、検索時に JOIN データ数を削減し、工程数、抽出期間増加時にも検索性能を維持する。

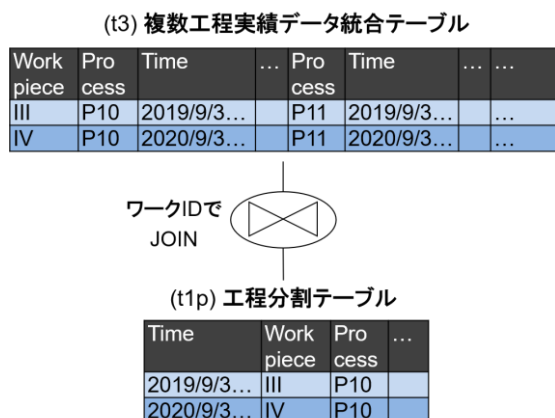


図 7：複数工程データ統合テーブルとその検索方法

### 3.3.1. 複数工程データ統合テーブル

複数工程データ統合テーブルの例を図 7 (t3)に示す。図 7では、ワーク III の工程 P10, P11 のデータを一つに行に格納し、また同様にワーク IV の同データを 1 行に格納する。本方式では、ライン生産方式を想定する。ライン生産方式では、異なるワークが、製造ラインにおいて同一の工程を同一の順序で処理される。従って製造ライン内の工程を列、ワークを行とすることで、製造ラインにおける全工程の実績データを格納することができる。

### 3.3.2. 複数工程データ統合テーブルの検索方法

複数工程データ統合テーブルにおける検索方法として、S#3 の検索では、複数工程データ統合テーブルと、図 7 (t1p)に示す工程分割テーブルを組み合わせた検索を実施する。すなわち工程分割テーブルの日時を用いて抽出条件を判定し、抽出条件にマッチしたワーク ID の実績データを複数工程統合テーブルから抽出する。これにより複数工程データ統合テーブルと工程分割テーブルの JOIN 処理が必要になるが、複数工程データ統合テーブルにおいてワーク ID 以外の索引の定義、あるいは複数工程データ統合テーブルのスキンを不要とすることができるため、採用する。

## 4. 製造実績データ管理手法の評価

本章では、提案した製造実績データ管理手法の検索時間の評価結果を述べる。提案手法のテーブル分割及

び、データ統合に関して、2.1.3 節で挙げた要件 1, 2 を満たすか評価を実施する。

### 4.1. 評価環境

評価環境は、データベースのテーブルを格納するディスクとして、内蔵 SATA の内蔵 HDD (以下、HDD 環境) 及び USB3.0 の外付け SSD (以下、SSD 環境) の二通りとし各々のディスクで評価する。データベースは、RDB(Relational Database)を用いる。RDB の検索クエリ (SQL) は、単一 CPU コア上で単一スレッドで実行させる。工程実績データとしては、50 工程、2 年分のデータを用いる。各工程において毎月平均 40 万ワーク分のデータがあり、データのサイズは平均 200byte とした。

### 4.2. 蓄積データ量増加時の検索時間評価

要件 1 で述べた蓄積データ量増加時における検索時間の評価結果を述べる。評価に用いた検索クエリは、表 1 の S#1~S#3 の検索である。S#1 をワーク単位でテーブル分割、S#2, S#3 を工程と年単位でテーブル分割した方式を適用する。

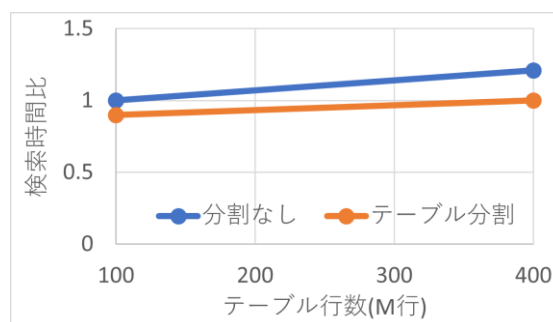


図 8：HDD 環境におけるワーク検索時間

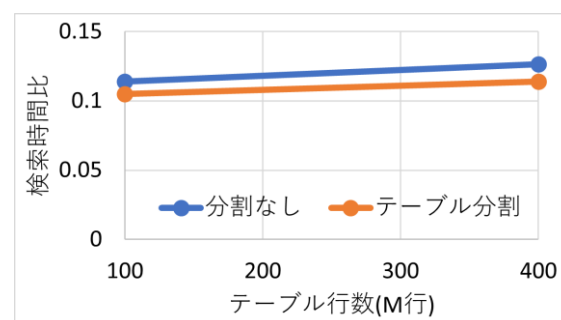


図 9：SSD 環境におけるワーク検索時間

#### 4.2.1. ワーク検索(S#1)

ワークを指定した工程実績データの検索 (表 1 S#1) に対して、ワーク単位のテーブル分割方式を適用した検索時間の評価結果を述べる。

図 8 に HDD 環境, 図 9 に SSD 環境における評価結果を示す。検索対象の製造実績データのテーブル行数が 100M 行と 400M 行の場合における, テーブル分割しない方式と, ワーク単位テーブル分割方式の検索時間を評価した。ワーク単位テーブル分割方式では, ワーク ID のハッシュ値に従って, 4 つのテーブルに分割した。検索時間は, 絶対値でなく, テーブル分割しない方式における HDD 環境での 100M 行のテーブルの検索時間との比で示す。

評価の結果, テーブル行数 400M 行の場合に, テーブル分割により, HDD 環境では 17%, SSD 環境では 10%検索時間を削減することができ, テーブル行数 100M 行から 400M 行の検索時間の増加を 10%未満に留めることが確認できた。

#### 4.2.2. 単一工程検索(S#2)

単一工程と期間を指定した工程実績データの検索(表 1 S#2)に対して, 工程と年単位のテーブル分割を適用した検索時間の評価結果を述べる。

図 10 に HDD 環境, 図 11 に SSD 環境における評価結果を示す。製造実績データのテーブル行数が 4M 行(1 工程 1 年分), 8M 行(1 工程 2 年分), 400M 行(50 工程 2 年分)における, テーブル分割しない方式と, 工程と年単位のテーブル分割方式の検索時間を評価した。検索の抽出工程数を 1 工程, 抽出期間を 1 カ月とした。検索時間は, テーブル分割しない方式における HDD 環境での 4M 行のテーブルの検索時間との比で示す。

評価の結果, テーブル行数 400M 行の場合に, テーブル分割により HDD 環境で 60%, SSD 環境で 42%検索時間が削減し, 4M 行から 400M 行へのテーブル行数増加時に HDD 及び SSD の環境において検索時間が増加しないことを確認できた。

#### 4.2.3. 複数工程検索(S#3)

複数工程と期間を指定した工程実績データの検索(表 1 S#3)に対して, 工程と年単位のテーブル分割を適用した検索時間の評価結果を述べる。

図 12 に HDD 環境, 図 13 に SSD 環境における評価結果を示す。製造実績データのテーブルの総行数が 40M 行(10 工程 1 年分), 80M 行(10 工程 2 年分), 400M 行(50 工程 2 年分)における, テーブル分割しない方式と, 工程と年単位でテーブル分割する方式の検索時間を評価した。検索の抽出工程数を 10 工程, 抽出期間を 1 カ月とした。検索時間は, 4.2.2 節で述べた検索時間比と同様である。

評価の結果, テーブル行数 400M 行の場合に, テーブル分割により HDD 環境で 91%, SSD 環境で 26%の検索時間が削減し, 40M 行から 400M 行へのテーブル

行数増加時に HDD 及び SSD の環境においても検索時間が増加しないことを確認できた。

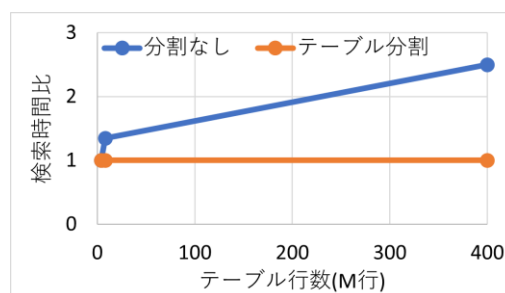


図 10 : HDD 環境における単一工程検索時間

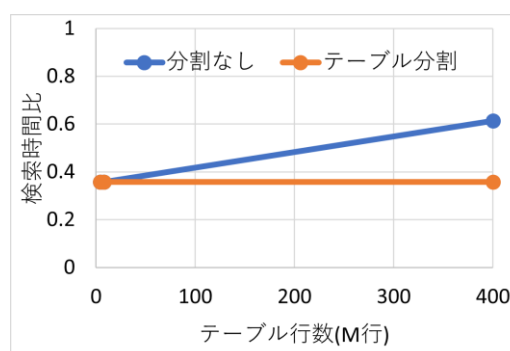


図 11 : SSD 環境における単一工程検索時間

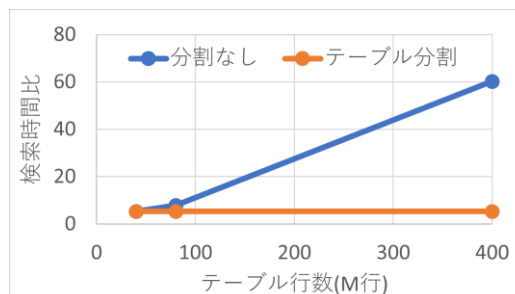


図 12 : HDD 環境における 10 工程検索時間

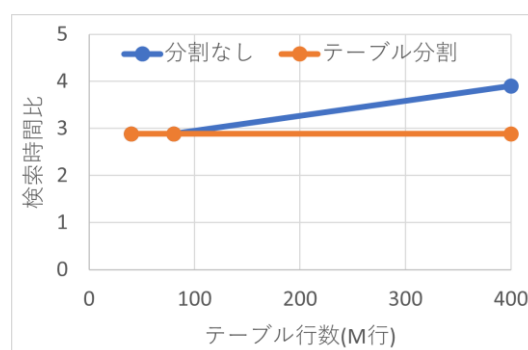


図 13 : SSD 環境における 10 工程検索時間

#### 4.2.4. 評価結果纏め

蓄積データ量増加時における検索時間の評価結果を纏める。評価結果としては、いずれの検索においても、テーブルの行数が 100M 行以下から 400M 行に増加した際の検索時間の増加 10%未満に留め、HDD 及び SSD の環境において要件 1 で挙げたデータ量増加時に検索時間を維持を達成できる見込みを得た。特に工程と期間を指定した検索では、400M 行のテーブルの検索において、テーブル分割により SSD 及び HDD 環境で各々最大 42%、91%検索時間を削減し、効果を確認できた。

#### 4.3. 抽出工程数増加時の検索時間評価

要件 2 で述べた抽出工程数増加時における検索時間評価結果を述べる。評価に用いた検索クエリは、要件 2 で求められる、複数工程と期間を指定した検索(表 1S#2)検索である。S#2 の検索時間を、データ統合方式と、テーブル分割しない方式及びテーブル分割方式と比較する。評価に用いたデータ統合方式では、最大 15 工程を 3M 実績データを含めテーブルの 1 レコードに格納する。テーブルには、品質要因分析(UC#2)を想定し直近 1 年分の実績データを格納する。

図 14 に HDD 環境、図 15 に SSD 環境における評価結果を示す。HDD 環境では抽出工程数が 1 工程から 20 工程、また SSD 環境では抽出工程数が 1 工程から 38 工程の検索時間を評価した。抽出期間は全ての検索において 1 カ月で固定した。検索時間は、4.2.2 節で述べた検索時間比と同様である。

評価の結果、HDD 環境では 10 工程以上、SSD 環境では 5 工程以上で、データ統合方式において、テーブル分割方式に対して検索時間を削減し、特に 10 工程以上の検索において、HDD 環境で平均 24%、最大 45%、SSD 環境では、平均 48%、最大 56%の検索時間を削減し、抽出工程数増加時の検索時間削減効果を確認できた。その結果、HDD 環境で抽出工程数を 1 工程から 20 工程に増やした場合に、13 倍の検索時間の増加となり、また SSD 環境で、抽出工程数を 1 工程から 20 工程に増やした場合に 12 倍、1 工程から 38 工程に増やした場合には 20 倍の検索時間の増加となった。従って HDD 及び SSD の環境において、検索時間の増加率を工程数の増加率をよりも低く抑えることができ、要件 2 で挙げた多数工程の検索に対応できると考えられる。

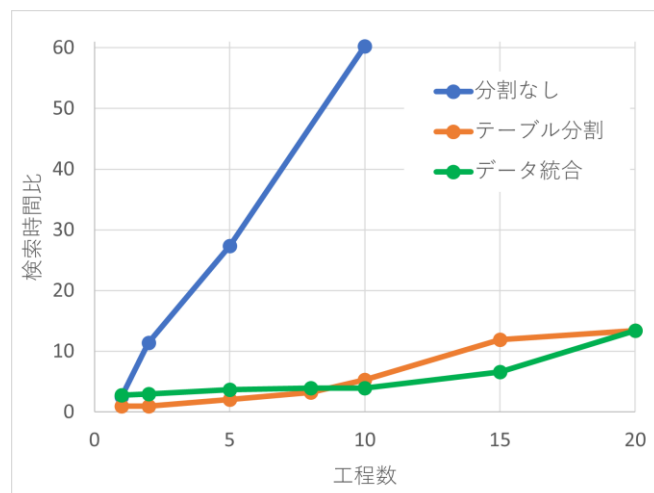


図 14：HDD 環境における工程数と検索時間の関係

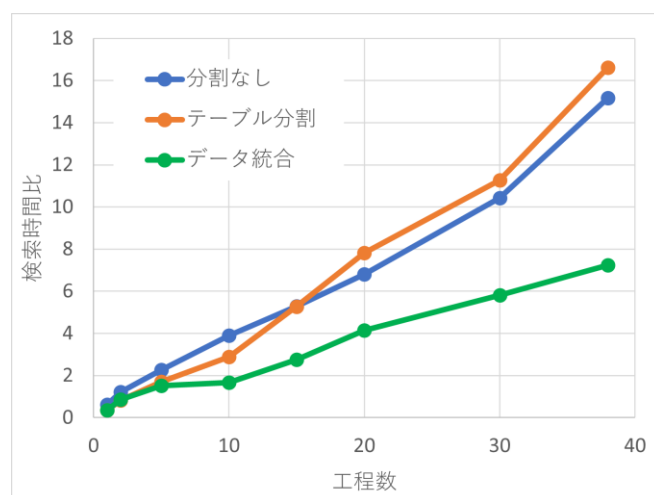


図 15：SSD 環境における工程数と検索時間の関係

#### 5. 関連研究

関連研究としては、データベースのテーブルパーティショニングの研究[3][4][5][6][7][8][9][10][11]が挙げられる。テーブルパーティショニングとしては、異なるレコードを異なるテーブルを分割する水平分割と、単一レコードのカラムを複数のテーブルに分割する垂直分割に分けられる。例えば、S. Agrawal らの手法[3]では性能の運用性を考慮して、水平分割及び垂直分割と、テーブルの索引付け、マテリアルビュー生成を考慮した、最適な物理テーブル設計を算出する。本研究で提案する、工程、ワーク、日時でのテーブル分割は、水平分割にあたり、また複数工程データ統合は、垂直分割または統合にあたり、本研究は、製造実績データの検索における水平及び垂直分割の最適化手法と位置付けられる。

## 6. おわりに

### 6.1. まとめ

製造現場では、IoT システムにおいて製造工程や 3M に関する大量の製造実績データを収集・蓄積し、トレーサビリティ管理や品質要因分析を日々実施している。しかしながら、生産の拡大でデータ量は増大し、これら業務で多用される長期の蓄積データに対する検索や、複数の工程間を跨るデータの検索にかかる時間が問題となる。

本研究では、長期保存する製造実績データに対して、その検索を用いる製造工程、ワーク、日時を組み合わせ単位でテーブル分割することで、データ量増加時に検索性能を維持する手法を提案した。また直近 1 年以内等、特定期間の製造実績データに対して、同一ワークの複数の製造工程のデータを、予めワーク ID をキーにデータ統合し保持することで、多数工程データ抽出時の検索性能を維持するデータ統合手法を提案した。

提案手法を評価した結果、テーブル分割により検索時間を最大 91%削減し、その結果、テーブルの行数が 100M 行以下から 400M 行に増加した際の検索時間の増加を 10%未満に抑えることを確認し、大量の蓄積データの検索に対応できる見込みを得た。またデータ統合により、10 工程以上の検索において最大 56%の検索時間削減を確認し、抽出工程数を 1 工程から 38 工程に増やした場合の検索時間の増加率を 20 倍に留め、多工程を抽出する検索の対応できる見込みを得た。

### 6.2. 今後の課題

今後の課題としては、検索のバリエーションを増やした場合における検索高速化や、データ挿入やバックアップ等のデータベースのオペレーション全体を考慮したテーブル設計が挙げられる。

## 参 考 文 献

- [1] AMERICAN NATIONAL STANDARD, ANSI/ISA-95.00. 02-2018, "Enterprise-Control System Integration - Part 2: Objects and Attributes for Enterprise-Control System Integration", 2018.
- [2] 伊藤大輔, 石田仁志, "製造現場を改善し続けるデジタルツイン技術", 日立評論 vol.102 No.3, 2020.
- [3] Sanjay Agrawal, Vivek R. Narasayya, Beverly Yang, "Integrating Vertical and Horizontal Partitioning Into Automated Physical Database Design", SIGMOD Conference 2004.
- [4] Brian Stein and Alan Morrison, "The enterprise data lake: Better integration and deeper analytics", Technology Forecast: Rethinking integration Issue 1, 2014.
- [5] Hong Yin, Shuqiang Yang, Hui Zhao, Zhikun Chen, "Partial query optimization Techniques for partitioned table", IEEE 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2012) .
- [6] N. Polyzotis, "Selectivity-based Partitioning: A Divide-and-union Paradigm for Effective Query Optimization", CIKM, 2005.
- [7] Ladjel Bellatreche, Kamalakara Karlapalem, "What can Partitioning Do For your Data Warehouses and Data Marts?", IDEAS'00, Proceedings of the 2000 International Symposium On Database Engineering and Applications.
- [8] Chun-Hung Cheng, Wing-Kin Lee, And Kam-Fai Wong, "A Genetic Algorithm-Based Clustering Approach For Database Partitioning", IEEE Transactions On Systems, Man, And Cybernetics-Part C-Applications and Reviews, Vol.32, No.3, August 2002.
- [9] Camille Furtado, Alexandre A. B. Lima, Esther Pacitti, Patrick Valduriez and Marta Mattoso, "Physical and Virtual Partitioning in OLAP Database Clusters, Proceedings of the 17th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD'05) .
- [10] Wen Qi, Jie Song, Yu-bin Bao, "Near-uniform Range Partition Approach for Increased Partitioning Large Database", IEEE, 2010.
- [11] Lisbeth Rodríguez and Xiaohu Li, "A Dynamic Vertical Partitioning Approach for Distributed Database System", IEEE 2011.