SQL

Announcements

Databases

Data is very, very powerful!

AI is made of data…

DOI:10.1145/3448247

Industry experiences on the data challenges of AI and the call for a data ecosystem for industrial enterprises.

BY CHRISTOPH GRÖGER

There Is No AI Without Data

ARTIFICIAL INTELLIGENCE (AI) has evolved from hype to reality over the past few years. Algorithmic advances in machine learning and deep learning, significant increases in computing power and storage, and huge amounts of data generated by digital transformation efforts make AI a game-changer across all industries.[8] AI has the potential to radically improve business processes with, for instance, real-time quality prediction in manufacturing, and to enable new business models, such as connected car services and self-optimizing machines. Traditional industries, such as manufacturing, machine building, and automotive, are facing a fundamental change: from the production of physical goods to the delivery of AI-enhanced processes and services as part of Industry 4.0.[35] This paper focuses on AI for industrial enterprises with a special emphasis on machine learning and data mining.

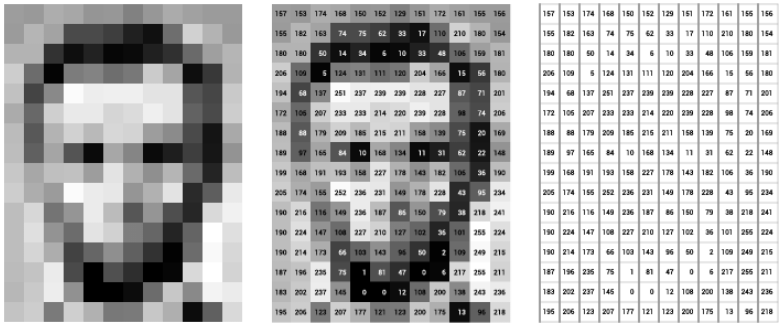
Despite the great potential of AI and the large investments in AI technologies undertaken by industrial enterprises, AI has not yet delivered on the promises in industry practice. The core business of industrial enterprises is not yet AI-enhanced. AI solutions instead constitute islands for isolated cases—such as the optimization of selected machines in the factory—with varying success. According to current industry surveys, data issues constitute the main reasons for the insufficient adoption of AI in industrial enterprises.[27,35]

In general, it is nothing new that data preparation and data quality are key for AI and data analytics, as there is no AI without data. This article profoundly investi-

» key insights

- **Despite AI's great potential, the business of industrial enterprises is not yet AI-enhanced. AI is done in an insular fashion, leading to a polyglot and heterogeneous enterprise data landscape that limits the comprehensive application of AI.**
- **Data challenges, such as data management, data democratization, and data governance, constitute the major obstacles to leveraging AI and go far beyond ensuring data quality, comprising diverse aspects such as metadata management, data architecture, and data ownership.**
- **The presented data ecosystem for industrial enterprises addresses these challenges and comprises data producers, data platforms, data consumers, and data roles for AI.**

(Source: https://cacm.acm.org/research/there-is-no-ai-without-data/)



digital images are made out of data…

To many of the biggest, most powerful corporations in the world…

Meta

ByteDance

Alphabet

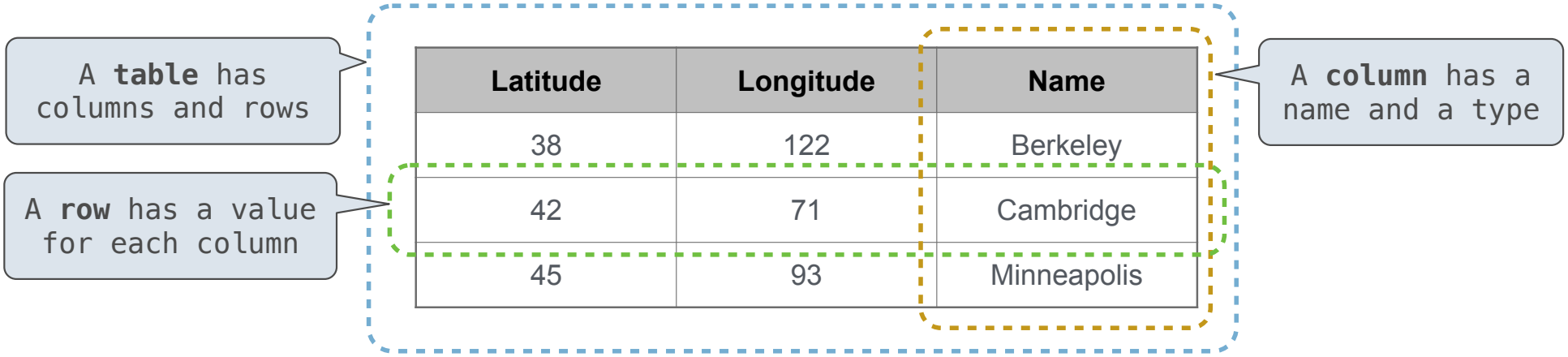
…data about ***us*** is their most prized resource!

(Source: https://www.researchgate.net/publication/330591504_Visual_Udder_Detection_with_Deep_Neural_Networks)

Database Management Systems

Database management systems (DBMS) are important, heavily used, and interesting!

A table is a collection of records, which are rows that have a value for each column



The Structured Query Language (SQL) is perhaps the most widely used programming language

SQL is a *declarative* programming language

Programming Paradigms

Programming Paradigms

• **Paradigm** (Merriam Webster): a typical example or pattern of something; a model. Example: "there is a new paradigm for public art in this country”

• **Programming Paradigm** (Joe Turner, Clemson University): “A programming paradigm is a general approach, orientation, or philosophy of programming that can be used when implementing a program.” You might call this a "style"

Many Different Approaches

There is no universally agreed upon taxonomy of human programming styles.

One possible list:

- Imperative
- Functional
- Array-based
- Object-Oriented
- Declarative

These terms are a bit fluid, and as you'll see if you read more on wikipedia, there is substantial disagreement about these terms.

Some Examples

Example, very different approaches to squaring list:

```
  lst = []
  for i in range(5):
      lst += [ i*i ]
```

```
map(lambda x: x*x, range(5))
```

```
[ x * x for x in range(5) ]
range(5).square_nums() # Only theoretically, e.g assume `def square_nums(self)` exists
```

```
np.sum(
  np.array([0, 1, 2, 3, 4]) *
  np.array([0, 1, 2, 3, 4])
)
```

```
np.sum(np.array([0, 1, 2, 3, 4]) ** 2)
```

Declarative Programming

In **declarative programming:**

- A "program" is a description of the desired result
- The interpreter figures out how to generate the result

Imperative Programming
is like…
"Add 2 teaspoons of salt and 2 teaspoons of yeast.
Add 3 cups of flour.
Add 2 tablespoons of olive oil.
Add 1/4th a cup of water;
Start mixing the ingredients together.
Put the dough ball on a surface…"

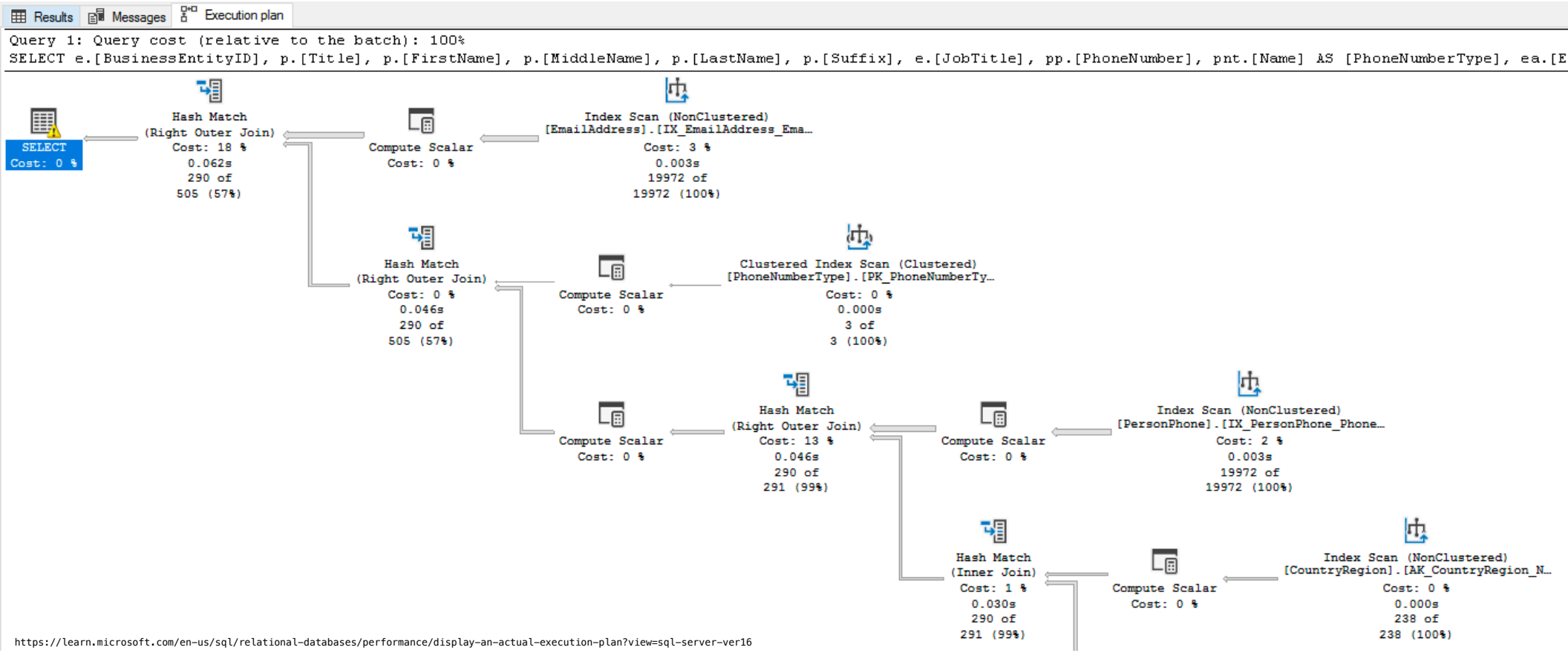
Declarative Programming
is like…
"I would a like pizza"
"16 inches, with pepperoni"

Declarative Programming

In **declarative programming:**

- A "program" is a description of the desired result
- The interpreter figures out how to generate the result

SQL Server Query Plan:



https://learn.microsoft.com/en-us/sql/relational-databases/performance/display-an-actual-execution-plan?view=sql-server-ver16

Structured Query Language (SQL)

Naming Tables

A **select** statement creates a new table and displays it.

A **create table** statement names the result of a **select** statement.

```
create table [name] as [select statement];
```

Here's how I might create a table of some of my most-listened-to spotify tracks in SQL:

```
create table songs as
  select "WILDFLOWER" as track, "Billie Eilish" as artist union
  select "BIRDS OF A FEATHER" , "Billie Eilish"           union
  select "360"                , "Charli xcx"              union
  select "Pasilyo"            , "Sunkissed Lola"          union
  select "Cinderella"         , "Remi Wolf"               union
  select "Good Luck Babe!"    , "Chappell Roan"           union
  select "Meow"               , "Anamanaguchi";
```

songs:

track	artist
WILDFLOWER	Billie Eilish
BIRDS…	Billie Eilish
360	Charli xcx
Pasilyo	Sunkissed Lola
Cinderella	Remi Wolf
Good Luck Babe!	Chappell Roan
Meow	Anamanaguchi

Select Statements Project Existing Tables

A **select** statement can specify an input table using a **from** clause

A subset of the rows of the input table can be selected using a **where** clause

An ordering over the remaining rows can be declared using an **order by** clause

Column descriptions determine how each input row is projected to a result row

```
select [expression] as [name], [expression] as [name], ...;
select [columns] from [table] where [condition] order by [order];
select track from songs where artist = "Billie Eilish";
select track from songs where track < artist;
```

songs:

track	artist
360	Charli xcx
BIRDS…	Billie Eilish
Cinderella	Remi Wolf
Good Luck Babe!	Chappell Roan
Pasilyo	Sunkissed Lola
Meow	Anamanaguchi
WILDFLOWER	Billie Eilish

Optional Example:
UC Salary Data / Your Own Data

SOURCES: https://ucannualwage.ucop.edu

The University is a public institution, so it is supported to an extent by California taxpayers through an allocation by the state government. In the past, generous state support allowed UC Berkeley to operate while keeping costs to students low. While still an important revenue source, the state's financial support of the university has diminished significantly. Thirty years ago, 50 percent of the university's revenue came from the state, but today, the state provides just 14 percent of the university's revenue.