The GQL Standard is Published! Now what?

Keith W. Hare
  Convenor, ISO/IEC JTC1 SC32 WG3 Database Languages

August 30, 2024
Eighteenth LDBC TUC Meeting
Guangzhou China

Photo from the June 2024 WG3 meeting, Sanya China

Who Am I?

- JCC Consulting, Inc.
  - President since August 2019
  - Senior Consultant 1985 – 2019
  - Specialize in
    - High performance database systems
    - Data replication and migration
    - Database Administration

- Standards – SQL and GQL
  - Convenor, ISO/IEC JTC1 SC32 WG3 Database Languages
  - Vice Chair, ANSI INCITS Data Management

- Neo4j Languages, Standards, and Research (LANGSTAR)
  - Standards Diplomacy
  - GQL Strategy
  - Developing the GQL standard

Introduction

- What is a standard?
- What is the GQL standard and who created it?
- What does it mean to conform to the GQL standard?
- What products have announced GQL implementations?
- What happens after GQL V1?
- Summary

What is a standard?

- Standards are agreements between participants on how to do something

  It could be about making a product, managing a process, delivering a service or supplying materials – standards cover a huge range of activities. Standards are the distilled wisdom of people with expertise in their subject matter and who know the needs of the organizations they represent – people such as manufacturers, sellers, buyers, customers, trade associations, users or regulators. https://www.iso.org/standards.html

- **Standards benefit businesses and consumers**

  - **For businesses – easier and cheaper to cooperate and compete**

  - **For consumers – increased choices, quality, and safety, and decreased costs**

Standards Organizations (incomplete list)

- ISO – International Organization for Standardization
- IEC – International Electrotechnical Commission – electronics
- IETF – Internet Engineering Task Force – specifies how the internet works
- IEEE – Institute of Electrical and Electronics Engineers
- ITU – International Telecommunications Union – telephones

Standards Example – Shipping Containers

- Defined by ISO 668:2020 *Series 1 freight containers — Classification, dimensions and ratings*

- Reduce labor to load and unload ships

- Reduce the cost of shipping goods world wide

- Size:
  - 8 feet wide
  - 20 or 40 feet long

- Transported by
  - Ships
  - Trucks
  - Trains

Photo by Ian Taylor

Standards Example – Wireless networks

- IEEE 802.11

  *IEEE Standard for Information technology--Telecommunications and information exchange between systems Local and metropolitan area networks--Specific requirements - Part 11: **Wireless** LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*

- Wireless internet access

- Revised and enhanced over multiple generations
  - 802.11 – 1997
  - 802.11b – 1999
  - 802.11a – 1999
  - 802.11g – 2003
  - 802.11n – 2008
  - 802.11ac – 2014
  - 802.11ax – 2019/2020
  - 802.11be – (2024?)
  - 802.11bn – (2028?)

Database Language Standards

- ISO/IEC 9075 *Information technology — Database languages — SQL*
- ISO/IEC 39075 *Information technology — Database languages — GQL*

SQL Standards – a brief history

ISO/IEC 9075 Database Language SQL

- SQL-87 – Transactions, Create, Read, Update, Delete

- SQL-89 – Referential Integrity

- SQL-92 – Internationalization, etc.

- SQL:1999 – User Defined Types, triggers

- SQL:2003 – XML & OLAP

- SQL:2008 – XML expansion, "instead of" triggers

- SQL:2011 – Temporal

- SQL:2016 – JSON, RPR, PTF, MDA (2019)

- SQL:2023 – Property Graphs in SQL, JSON enhancements
  - Published June 1, 2023
  - 11 parts including "Part 16: Property Graph Queries (SQL/PGQ)"

GQL Standard – a brief history

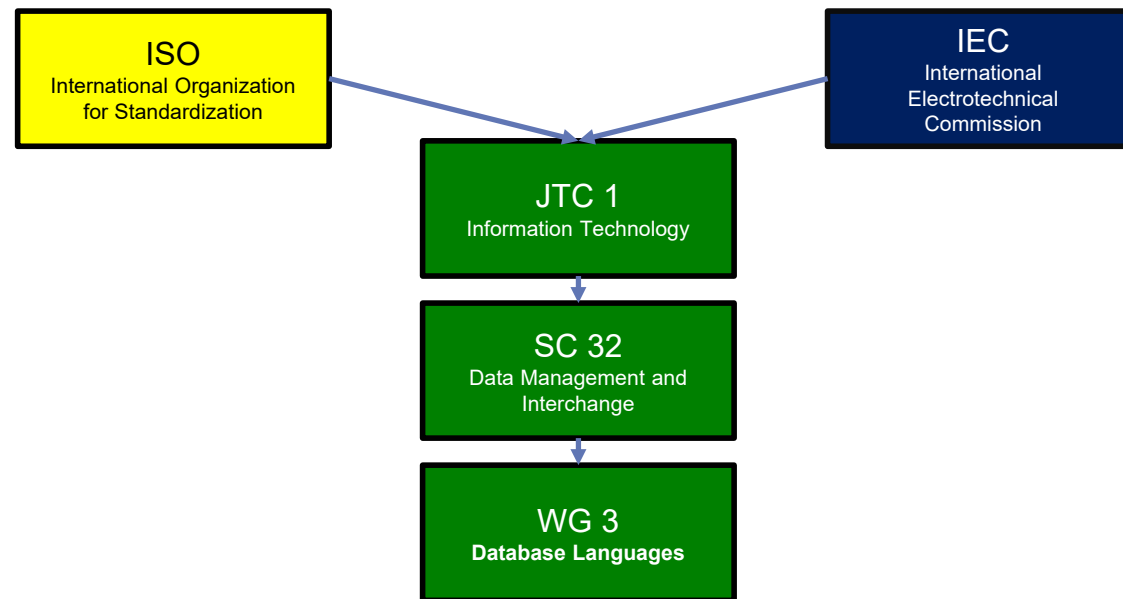ISO/IEC 39075 *Information technology — Database languages — GQL*

- Property Graph Database Language

- Project approved in September 2019

- We did a bit of work

- Standard officially published April 12, 2024!

- Full database language

  - Create graphs

  - Insert, update, delete, and query nodes and edges using graph patterns

  - Graph Pattern Matching language is common with SQL/PGQ Graph Pattern Matching

GQL is a full database language

- DML – Create, Read, Update, Delete
- DDL – Supports schema-fixed and schema-flexible variants
  - Create Graph,
  - Create Graph Type,
  - Create Graph From Graph Type
- Graph Pattern Matching queries
- Leverages common foundation from
  - SQL
  - Property graph languages
  - Other standards
    - Unicode character set
    - IEEE Floating Points
    - ISO 8601 Dates & Times

Who created the GQL Database Language Standard?

- Developed by ISO/IEC JTC/1 SC/32 WG/3 Database Languages
  - ISO – International Organization for Standardization
  - IEC – International Electrotechnical Commission
  - JTC 1 – Joint Technical Committee 1 – Information Technology standards
  - SC 32 – Sub Committee 32 – Data Management and Interchange
  - WG 3 – Working Group 3 – Database Languages

```
┌──────────────────────┐                    ┌──────────────────────┐
│         ISO          │                    │         IEC          │
│ International         │                    │ International        │
│ Organization         │                    │ Electrotechnical     │
│ for Standardization  │                    │ Commission           │
└──────────────────────┘                    └──────────────────────┘
              ┌──────────────────────┐
              │        JTC 1         │
              │ Information Technology│
              └──────────────────────┘
              ┌──────────────────────┐
              │        SC 32         │
              │ Data Management and  │
              │ Interchange          │
              └──────────────────────┘
              ┌──────────────────────┐
              │         WG 3         │
              │ Database Languages   │
              └──────────────────────┘
```

Who participates in SC/32 WG/3?

SC/32 WG/3 "Database Languages" is an international committee made up of "individual experts" delegated by various national body standards groups from:

- China
- Denmark
- Finland
- Germany
- Japan
- Korea
- Netherlands
- Sweden
- United Kingdom
- USA – INCITS Data Management
- Liaison participants from LDBC

How is a standard developed?

Creating the GQL database language standard required a **lot** of effort.

- Lots of papers – over 430 papers for GQL v1
- Lots of meetings – creating the GQL standard took:
  - Eleven face-to-face meetings
  - Twenty-seven web conferences.
- Lots of time – Five years – 2019 to 2024

Who uses the GQL standard?

The GQL standard is a computer language standard. The primary groups that use the standard directly are likely to be:

- Property graph database vendors
- Open source projects such as LDBC GQL implementation working group
- Academic researchers

Property graph database users will use property graph database products that implement the GQL standard.

So, what happens now?

GQL Implementations!

- Implementation aids:
  - GQL Artifacts
  - LDBC GQL Implementation Working Group
- GQL Conformance statements
- Vendors beginning to release GQL implementations

Next, GQL V2

GQL Artifacts

- GQL has a number of artifacts available on the ISO web site
- Generated from the source standards documents as a part of building the PDFs
  - BNF in TXT and XML
  - Exception Conditions
  - Optional Features
  - Implementation Defined Elements
  - Implementation Dependent Elements
- Useful for GQL implementers
- GQL https://standards.iso.org/iso-iec/39075/ed-1/en/

(Similar artifacts exist for SQL)

LDBC GQL Implementation Working Group

The LDBC GQL implementation working group has been developing tools to assist GQL implementations.

- Open GQL Language Tools – https://ldbcouncil.org/pages/opengql-announce/
  - ANTLR grammar repository
  - GQL Code Editor Web UI (and repository)
  - Railroad Diagrams Web Page (and repository)

Conforming to the GQL Standard

The GQL standard includes a specification of what is required to claim conformance to the standard.

- Implement a coherent subset of the GQL standard, including
  - Mandatory features
    - Everything that is not optional
    - No nomenclature for mandatory features
    - Some options, support either A or B.
  - Some optional features – specified in conformance rules (mostly)
  - Decide and document values for implementation-defined elements
- GQL implementations need to document their conformance in a conformance statement.
- Conformance statements are an important tool for users

GQL Subclause 24.2, "Minimum Conformance"

Minimum conformance is defined as meeting the requirements of the data model and all syntax and semantics not explicitly identified as belonging to an optional feature.

A claim of minimum conformance shall also include:

1) A claim of conformance to at least one of:
   a) Feature GG01, "Graph with an open graph type".
   b) Feature GG02, "Graph with a closed graph type" and a claim of conformance to at least one of:
      i) Feature GG20, "Explicit element type names".
      ii) Feature GG21, "Explicit element type key label sets".
      iii) Feature GG22, "Element type key label set inference".
      iv) Feature GG23, "Optional element type key label sets".

2) A claim of conformance to a specific version of The Unicode® Standard and the synchronous versions of Unicode Technical Standard #10, Unicode Standard Annex #15, and Unicode Standard Annex #31. The claimed version of The Unicode® Standard shall not be less than "13.0.0".

3) A claim of conformance to the set of all value types that are supported as the types of property values. At minimum, this set shall include:
   — The character string type specified by STRING or VARCHAR.
   — The Boolean type specified by BOOLEAN or BOOL.
   — The signed regular integer type specified by SIGNED INTEGER, INTEGER, or INT.
   — The approximate numeric type specified by FLOAT.

Conformance Statement Examples

- SQL database vendors have been writing conformance statements for decades

- GQL conformance statements can be modeled on SQL statements

- SQL Examples:

  - Postgres Documentation: *Appendix D. SQL Conformance*

    - https://www.postgresql.org/docs/current/features.html

  - Oracle Documentation: *SQL Language Reference Appendix C Oracle and Standard SQL*

    - https://docs.oracle.com/en/database/oracle/oracle-database/19/sqlrf/Oracle-and-Standard-SQL.html#GUID-330DEBBB-006E-4B35-A516-5C0AEFFE06B9

GQL Implementations

Several companies have announced some level of conformance to the GQL standard.

- Neo4j documented their conformance to the GQL standard in *Cypher Manual / Appendix / GQL Conformance*
  - https://neo4j.com/docs/cypher-manual/current/appendix/gql-conformance/
- Google Spanner released a version with some elements of SQL/PGQ mixed with some elements of GQL
  - https://cloud.google.com/blog/products/databases/announcing-spanner-graph
  - Google has not published a conformance statement for Google Spanner.
- Vesoft announced that NebulaGraph V5.0 will support GQL.
  - This version of Nebula was released to beta test in June.
  - No conformance statement has been published for NebulaGraph
- At the June 2024 SIGMOD conference, Ricky Sun of Ultipa said that Ultipa would support GQL in Q4.

What happens after GQL V1? GQL V2!

- Lots of ideas
  - Bug fixes
  - Schema information graph – LDBC Extended Schema working group
  - Vector data types and operations – under discussion for SQL
  - Cheapest path – under discussion for SQL/PGQ
  - Conditional execution
  - Etc.
- GQL users are likely to produce more requirements
- Timing and content of GQL V2 depends on:
  - Vendor ability implement GQL capabilities in products
  - User ability to absorb GQL capabilities delivered by products
  - Standards participants ability to write new papers

Summary – Database Language Standards

- GQL Standard – new in 2024
  - Full property graph database language
  - Published in April, 2024
  - GQL implementations are starting to appear
    - Need to include conformance statement
  - Already thinking about GQL V2 and later

Questions?

Photo from the June 2024 WG3 meeting, Sanya China