Bachelor-/Master-Forum 2022

Information Systems and Quantum Computing (ISQC)

Institute of Information Systems (IFIS)

4.11.2022

Professor Dr. rer. nat. habil. Sven Groppe
https://www.ifis.uni-luebeck.de/index.php?id=groppe

Institute of Information Systems (IFIS)

- 5 Professors/PDs
- currently ≈ 15 PhD students, 4 Postdocs, 3 External PhD students
- Labs
  - Information Systems and Quantum Computing (ISQC)
    - Prof. Dr. Sven Groppe
  - Cyber-Physical Medical Systems (CPMS)
    - Prof. Dr.-Ing. Jörg-Uwe Meyer
  - Foundations of AI (FAI)
    - Prof. Dr. Ralf Möller
  - Intellectics
    - PD Dr. Özgür Özçep
  - Human-Aware AI (HAI)
    - Prof. Dr. Nele Russwinkel

Information Systems and Quantum Computing (**ISQC**)

- Head: Prof. Dr. rer. nat. habil. Sven Groppe
- Projects and Research Assistants
  - **QC4DB**: Accelerating Relational Database Management Systems via Quantum Computing (BMBF)
    - Umut Çalikyilmaz
    - Tobias Winker
    - N.N.
  - **QualityOnt**: High Quality Knowledge Graphs from recent English, French and German Emergent Trends with the example of COVID-19 (DFG/ANR)
    - Hanieh Khorashadizadeh
  - **Semantic Data** Integration and Analysis (Bosch)
    - Simon Paasche (External PhD Student)
  - **BigSIoT**: Big Data Management for the Semantic Internet of Things (DFG)
    - Benjamin Warnke
  - **Hybrid**^{2}-Index Structures for Main Memory Databasess (DFG)
    - Tobias Groth

Supervision of Bachelor/Master Thesis & Result

- Often co-supervision of Prof. Groppe together with PhD student
  - meetings regularly and on request, typical:
    - weekly meetings with PhD student
    - monthly meetings with Prof. Groppe
- Experience
  - 91 supervised thesis (bachelor/master/student research project/Diploma/PhD)
- Publications based on bachelor/master thesis
  - improves visibility of student's contribution
  - improves chances for good job (in academia and industry)
  - 43 (out of 141) publications (of ISQC lab) are co-authored by a bachelor/master student (being a student at time of writing)
    - 30% of the publications

Typical Outline of Bachelor/Master Thesis

1. Introduction/Einführung
   1.1. Motivation
   1.2. Tasks of the Thesis/Aufgabenstellung
   1.3. Organization/Organisation der Arbeit
2. Basics/Grundlagen
   2.1. ...
   2.2. Further Related Work/Weitere wissenschaftliche Literatur
3. Concept/Konzept
4. Realization/Realisierung
5. Evaluation
6. Summary and Conclusions/Zusammenfassung und Ausblick

- Latex Template available, e.g.: **In Moodle**
- FAQs on Bachelor's and Master's Theses (from examination board for MINT): in **English**/**German**

DVD in addition to Thesis & for IFIS Archive

Please do not forget to burn DVDs for each of the thesis, content:

- Readme-file with installation instructions
- source code with documentation
  - additionally push to thesis repository in IFIS-Gitlab
- all necessary third-party-libraries
- all data sets for reproducing evaluation in thesis
- PDF of the thesis
- source files of the thesis (Word-file/latex-folder)

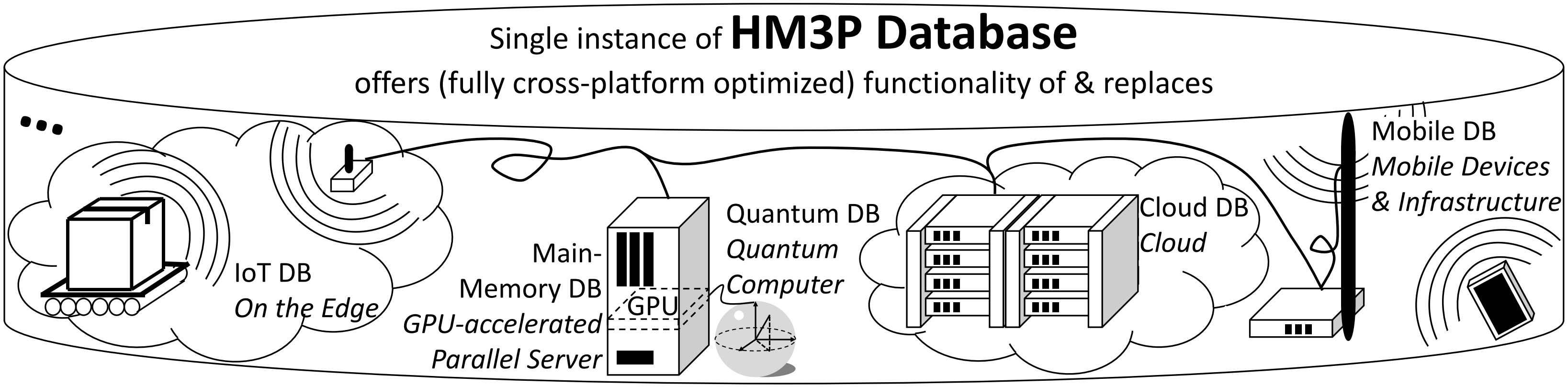
On the day of defense please deliver DVD for the IFIS-archive:

- Content as above
- PDF of the presentation for the defense
- source files of the presentation for the defense (Powerpoint file/latex-folder etc.)

Typical Defense of Thesis

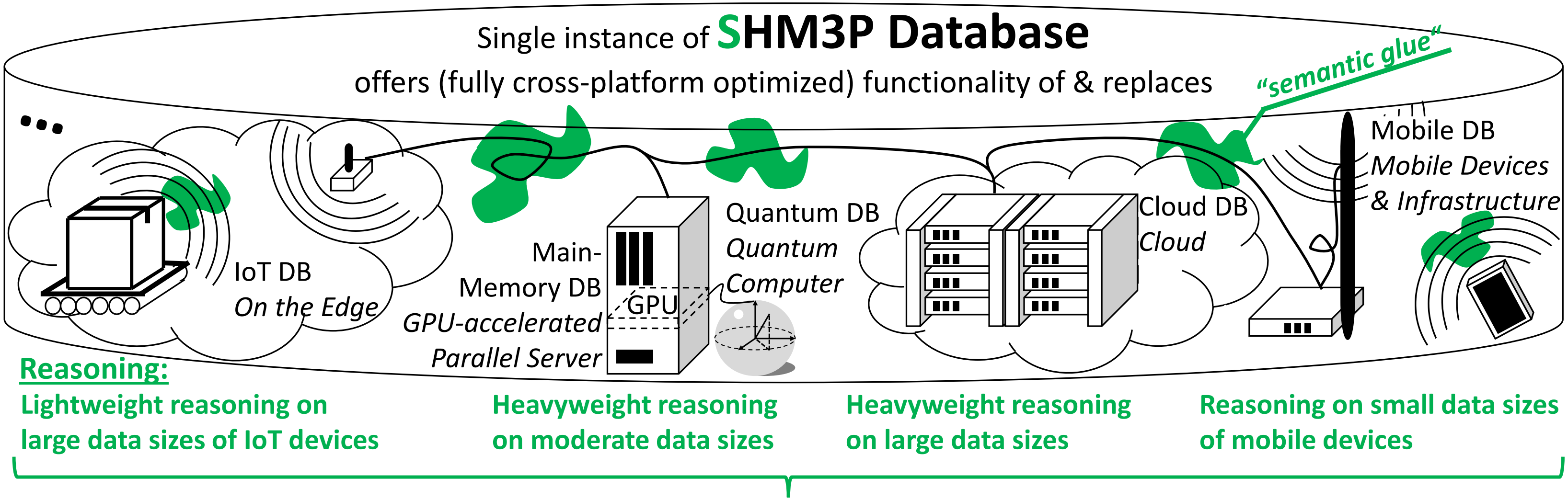
- 20-25 minutes presentation
  - often similar structure like thesis, but without 4. Realization
- maybe with succeeding short demonstration ($\approx$ 5 minutes) of developed software (dependent on thesis)
- Afterwards questions of reviewers and listeners
- Reviewers discuss alone in room about result
- Reviewers announce score to student and explain the reasons for the score
- In total: up to 1 hour

Hybrid Multi-Model Multi-Platform (HM3P) Database



- full and uniform data integration at database level
- performance: fully optimized across different data models
- transparent fault-tolerance
- SQL standards: relational ('87), XML ('03), temporal ('11), JSON ('16), Multi-dimensional Arrays ('19), schemaless ('19), streams ('20?), property graphs ('21?)
- **features of different types of databases running on different platforms can be used**

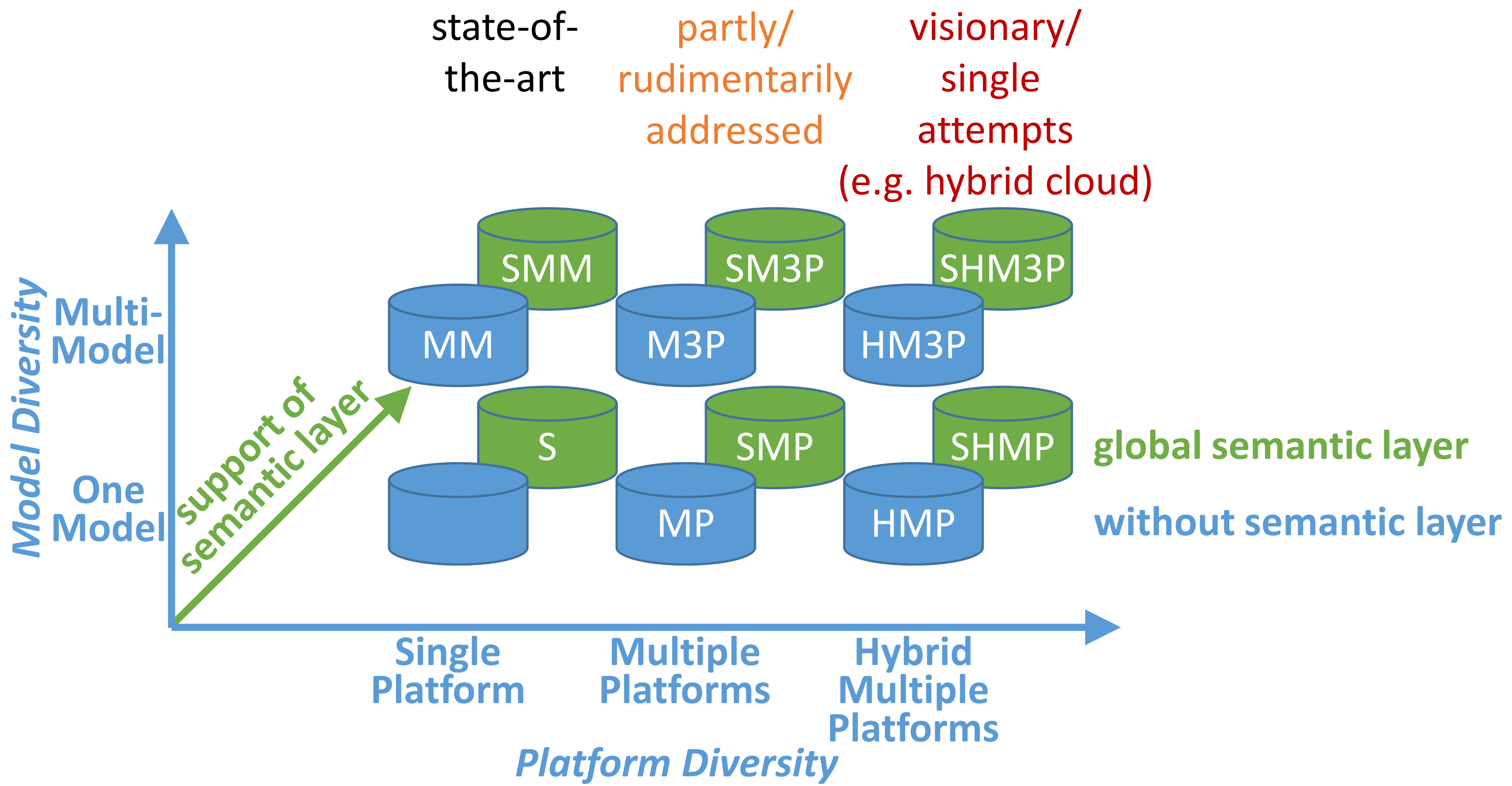
Variant: Semantic HM3P (SHM3P) DB



- Semantic Layer as glue between other models and platforms
- new challenges like integrating different types of reasoners in a transparent global reasoner

+ Features of HM3P databases
+ Easier data integration
- Performance issues may occur due to semantic layer

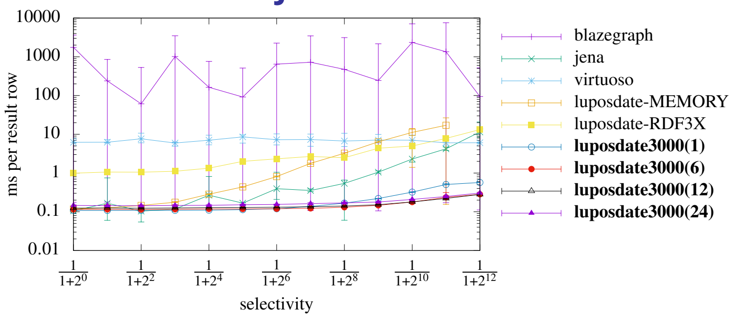
Types of DBMS



Legend: S: Semantic MP: Multi-Platform MM: Multi-Model M3P: MM MP H: Hybrid

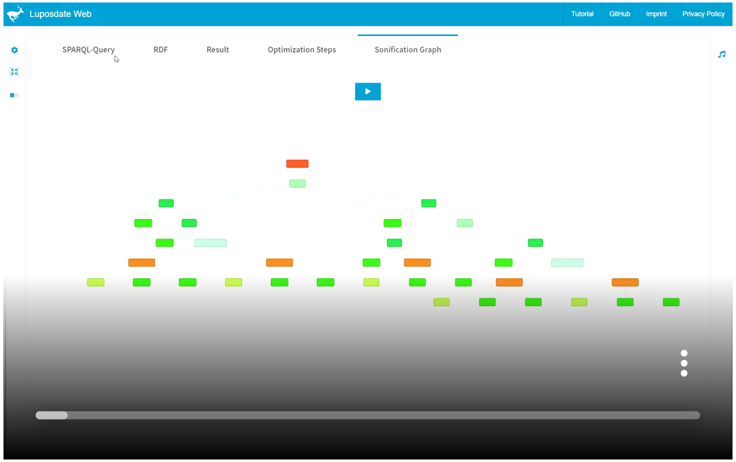
The Power of Multi-Platform: LUPOSDATE3000

- ultra-fast in jvm...



B. Warnke, M.W. Rehan, S. Fischer, S. Groppe: Flexible data partitioning schemes for parallel merge joins in semantic web queries in: BTW'21

- ...but also enabling web demos running completely in the browser!

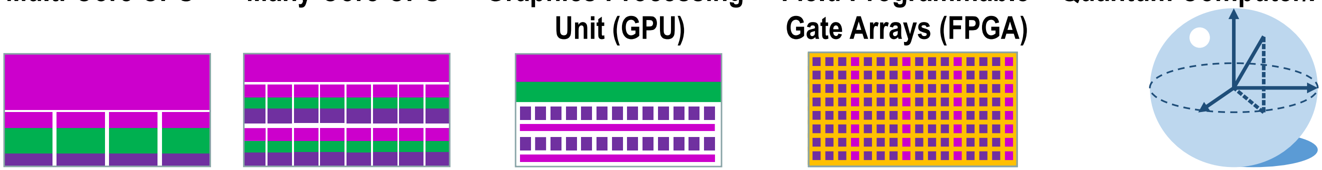


S. Groppe, R. Klinckenberg, B. Warnke. Sound of Databases: Sonification of a Semantic Web Database Engine. PVLDB, 14(12), 2021

Thesis supervised by Prof. Groppe (S)IoT Database LUPOSDATE3000

- Project BigSIoT (DFG, together with ITM) (Benjamin Warnke)
- External PhD student at Bosch (Simon Paasche)
- Combining (e.g., geo) routing protocols with query processing
- Compress data (in storage/messages) according to SHACL definitions, reformulate queries to directly work on compressed data
- Continuous Queries (e.g., redundant processing, efficient recovery after crashes, new types of windows for consistency checks)
- Distributed RDFS Inference
- Digital twins of machines in production
- Green Computing for digital twins/monitoring
- Automatic Testing of Databases: randomized testing, generation of test queries, automatic minimization of test cases, ...
- Semantic Web Layer for NebulaStream (co-supervision with DIMA/Berlin)

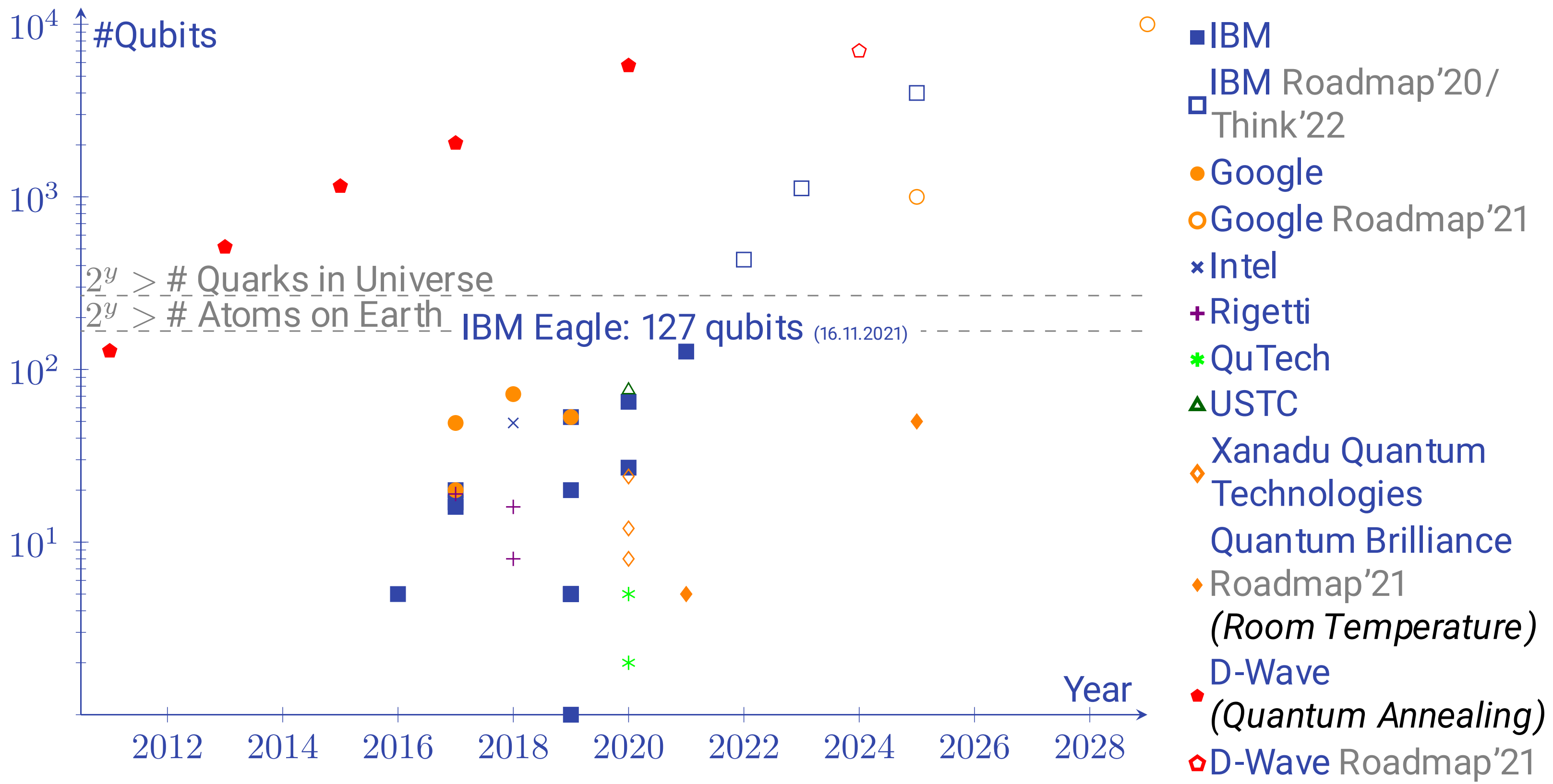
Architectures of Emergent Hardware



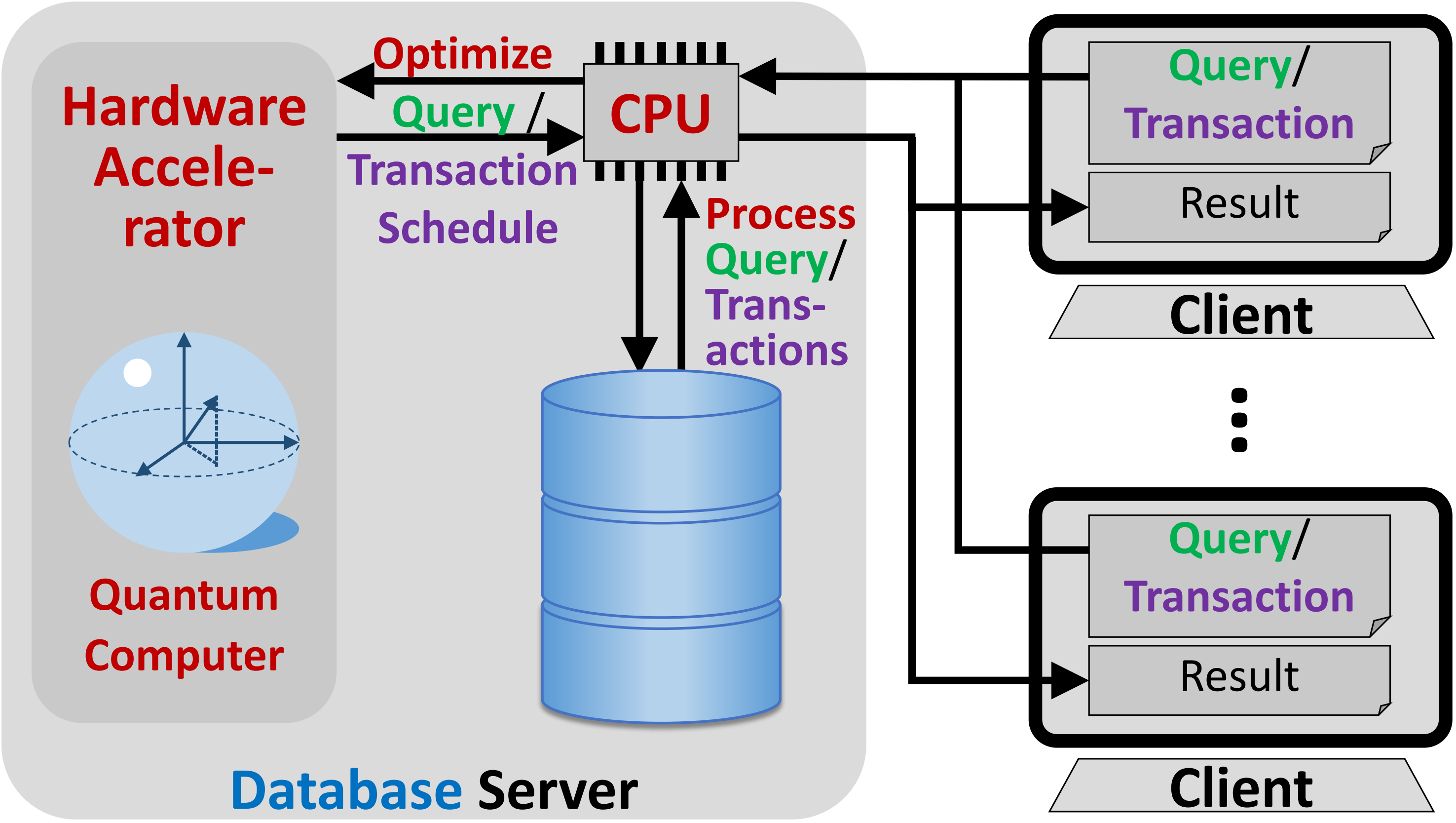
	Multi-Core CPU	Many-Core CPU	Graphics Processing Unit (GPU)	Field Programmable Gate Arrays (FPGA)	Quantum Computer/Annealer
Cores	~10	~100	~1000	~100 000	~100/~5000 qubits
Core Complexity	Complex (optimized for single thread performance)			Simple	Complex
Computational Model	MIMD + SIMD		SIMD	Data-Flow	Universal/Adiabatic Quantum computing
Parallelism	Thread and Data Parallel		Data Parallel	Arbitrary	Quantum Logic Gates/ Fluctuation (glob. Min.)
Memory Model	Shared	Distributed			Quantum Superposition
Power	150 W	200 W	250 W	50 W	25 KW
Database Op.	Query Optimization (Enumeration of Plans), Concurrency Control				
	Efficient Processing Of				
	Query Processing				

Legend: Computational Unit   Execution Controller   Interconnection Network   On-Chip Memory

Timeline of Quantum Computers



Using Hardware Accelerator for optimizing Queries / Transaction Schedules



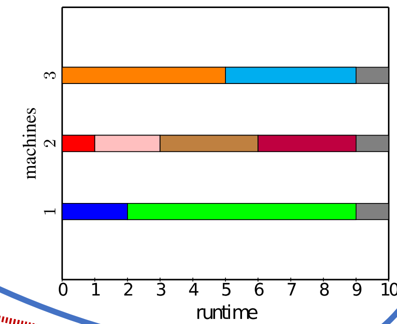
Approaches for Query/Transaction Schedule Optimization

Query Optimization:

$$\bowtie_{i=1}^{n} R_i \rightarrow ? \quad (R_1 \bowtie R_2) \cdots \bowtie R_n \quad \vdots \quad (R_1 \bowtie R_n) \bowtie (\cdots)$$

Transaction Schedule Optimization:

$$\{T_1, \ldots, T_m\} \rightarrow ?$$



Open Source Relational Database Management System (RDBMS),
e.g. PostgreSQL, MySQL

Dynamic Programming
Random Walk
Linear Program-ming
Machine Learning
...
Simulated Annealing
Genetic Algorithm

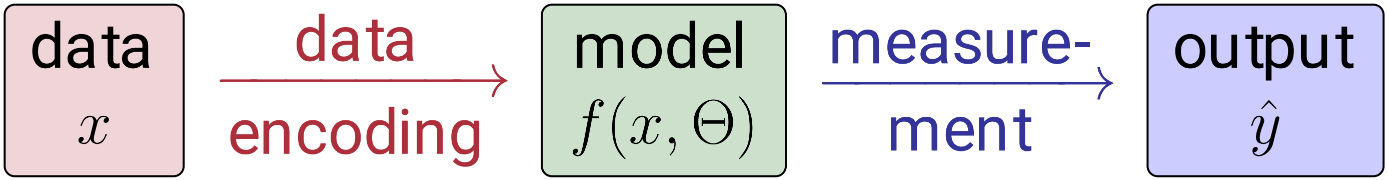
Algorithms (used e.g. in Query Optimization) and their Quantum Counterparts

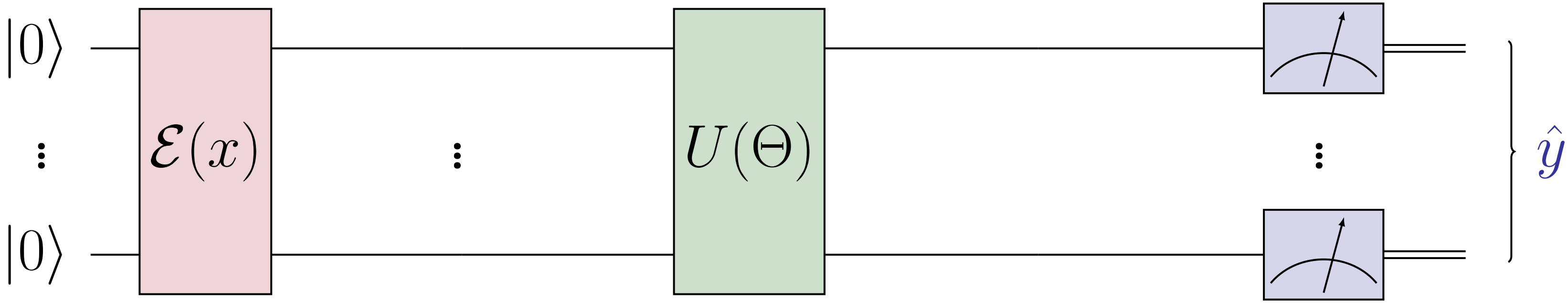
Query Optimization Approach	Basic Algorithm	Quantum Computing Counterpart
[S+79]	Dynamic Programming [E04]	[R19] [A+19]
[IW87], QA: [TK16]	Simulated Annealing [KGV83]	[J+11]
[MP18] [Y+20] [W+19] [O+19]	Reinforcement Learning [BSB81]	[S+21] [DCC05]
[GPK94]	Random Walk [BN70]	[ADZ93] [A+01]
[BFI91]	Genetic Algorithm [H92]	[W+13]
[TC19]	Ant Colony Optimization [CDM91] [DBS06]	[WNF07] [G+20]
[TK17]	Mixed Integer Linear [BGG+71] Programming [D02]	[HHL09] [A12] [CKS17] [SSO19] [AL22] [AL22]

This list is not complete...

- Please check my lecture about quantum computing: https://www.ifis.uni-luebeck.de/~groppe/lectures/qc

Quantum Machine Learning - Data encoding and Quantum Model

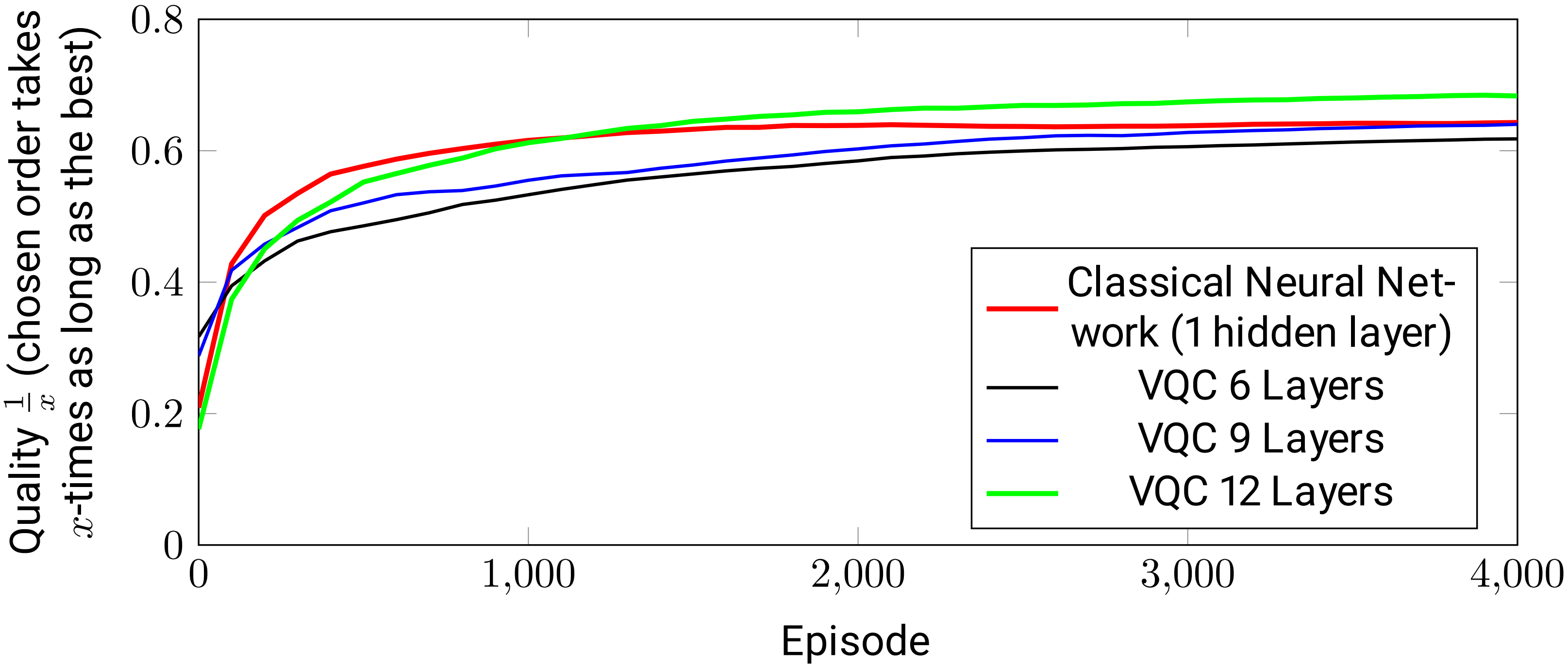




Relations to Join

Join Order

QML 4 Join Order Optimization



- Variational quantum circuits (VQCs) beat classical neural networks for join order optimization
- Extensions to this work $\Rightarrow$ bachelor/master thesis

Optimizing Transaction Schedules via Quantum Annealing

- Experiments on real Quantum Annealer (D-Wave 2000Q cloud service)
  - first minute free
    (afterwards too much for our budget)
- Versus Simulated Annealing on CPU
- Preprocessing time/Number of QuBits: $O((n \cdot k \cdot R)^2)$

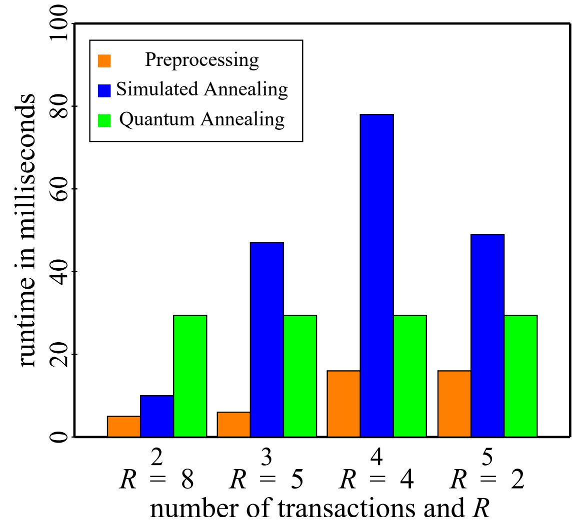


Fig.	k	n	R	O	$l_1, \ldots, l_n$	$r_1, \ldots, r_n$	**req. var.**
11	2	2	8	$\{\}$	8, 4	0, 4	8
		3	5	$\{(t_1, t_3)\}$	4, 5, 1	1, 0, 4	10
		4	4	$\{(t_2, t_4)\}$	3, 2, 1, 2	1, 2, 3, 2	16
		5	2	$\{(t_1, t_2), (t_4, t_5)\}$	1, 1, 1, 1, 1	1, 1, 1, 1, 1	10

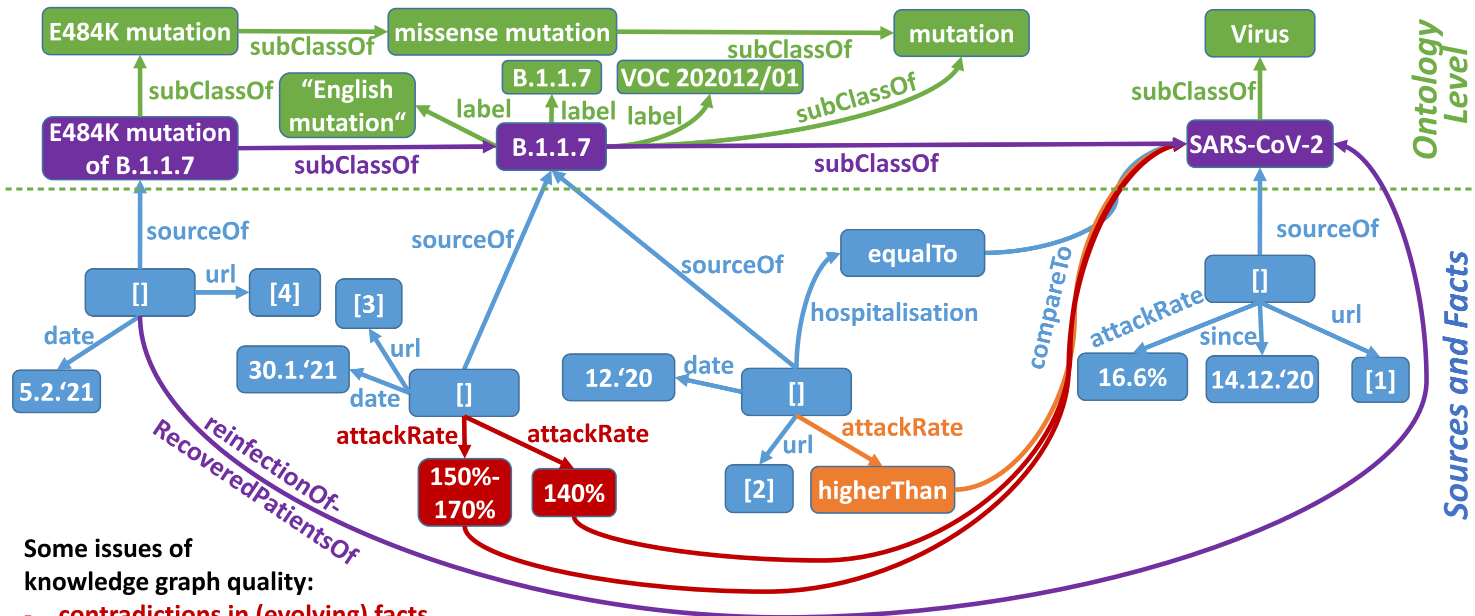
Open Challenges for QC for Databases/ Topics for Thesis

- Replacing basic algorithms with their QC counterparts in query optimizations for speeding up databases
  - Query Optimization: Tobias Winker
  - Transaction Schedule Optimization: Umut Çalikyilmaz
- What should be the properties of a quantum computer (e.g. #qubits, latencies of gates) to achieve certain speedups?
- How to combine classical and quantum computing algorithms to achieve good speedups with few qubits?
  (...for running database optimizations on current available quantum computers...)
- What other (database) domains besides query and transaction schedule optimizations benefit from quantum computers?
  (In short: those based on mathematical optimization problems, but also other...?)

QC4DB: Accelerating Relational Database Management Systems via Quantum Computing

Name:	QC4DB: Accelerating Relational Database Management Systems via Quantum Computing	
Proj. Web:	Project Website@Quantentechnologien	
Funded by:	BMBF, Fördermaßnahme Anwendungsnetzwerk für das Quantencomputing	
Duration:	3 years	
Volume:	1.8M Euros	
Topics:	Optimizing an open source relational database management system • Queries • Transaction Schedules	
Partners:	UNIVERSITÄT ZU LÜBECK INSTITUT FÜR INFORMATIONSSYSTEME (Coord.)	QUANTUM BRILLIANCE
Expertises:	Hardware-Acceleration of Databases	Room Temperature Diamond Quantum Accelerators/qbOS
Website:	https://www.ifis.uni-luebeck.de/~groppe/	https://quantumbrilliance.com/

COVID-19 Knowledge Graph (KG)



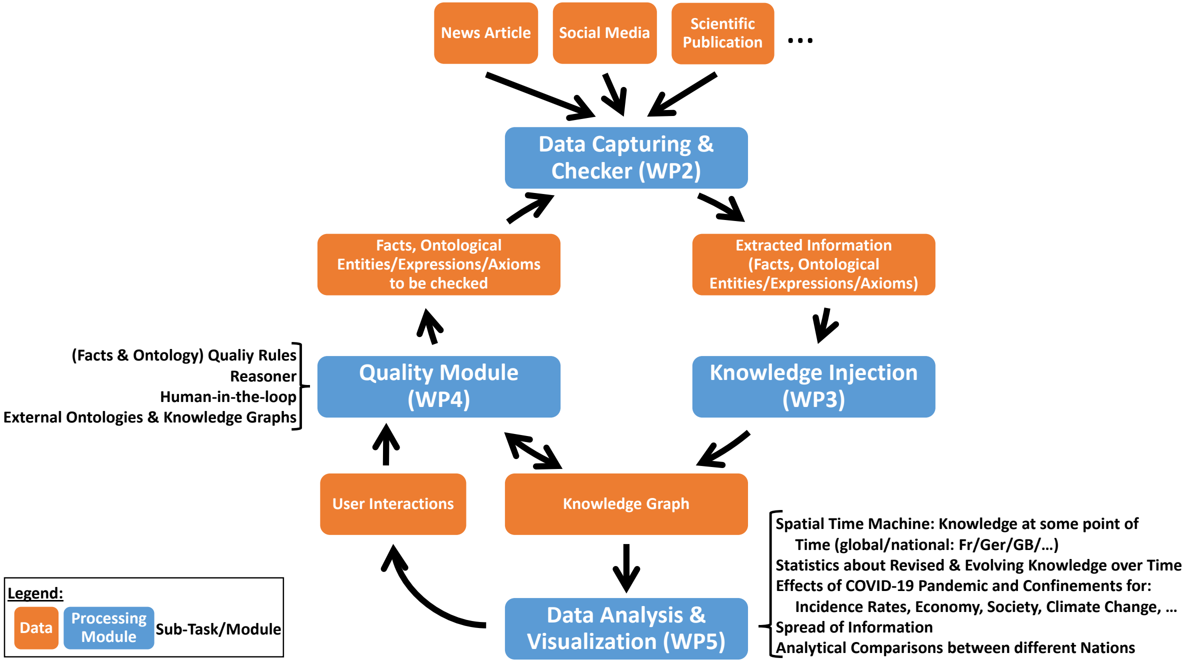
Some issues of knowledge graph quality:

- contradictions in (evolving) facts
- checking vague formulations and compare them with other given information
- errors with ambiguity: E484K mutation of B.1.1.7 is a SARS-CoV-2 virus, but reinfection with E484K is only possible for patients recovered from SARS-CoV-2 virus not mutating E484K (→ introduce class "SARS-CoV-2 without E484K mutation")

[1] https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2774102
[2] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959361/Technical_Briefing_VOC202012-2_Briefing_2.pdf
[3] https://www.ruhr24.de/service/corona-britische-mutation-neue-studie-mutante-virus-toedlicher-sterberate-deutschland-90184403.html
[4] https://www.bmj.com/content/372/bmj.n359

High Quality KGs from recent English, French & German Emergent Trends with the example of COVID-19

- Project with partners in Paris & Toulouse (Hanieh Khorashadizadeh)
- Thesis in the areas of
  - data capturing,
  - visualization & analysis,
  - detection of contradictions in KG, ...



Information Systems and Quantum Computing (**ISQC**)

- Head: Prof. Dr. rer. nat. habil. Sven Groppe
- Projects and Research Assistants
  - **QC4DB**: Accelerating Relational Database Management Systems via Quantum Computing (BMBF)
    - Umut Çalikyilmaz
    - Tobias Winker
    - N.N.
  - **QualityOnt**: High Quality Knowledge Graphs from recent English, French and German Emergent Trends with the example of COVID-19 (DFG/ANR)
    - Hanieh Khorashadizadeh
  - **Semantic Data** Integration and Analysis (Bosch)
    - Simon Paasche (External PhD Student)
  - **BigSIoT**: Big Data Management for the Semantic Internet of Things (DFG)
    - Benjamin Warnke
  - **Hybrid**^{2}-Index Structures for Main Memory Databasess (DFG)
    - Tobias Groth