Gurjeet Singh

gurjeet@singh.im

- Database internals/Systems Software Engineer and Architect with over 20 years of experience.
- Proficient in C, Database Internals, PostgreSQL internals, GNU/Linux and Windows.
- Proficient in Docker, Containers, Kubernetes, AWS, MS Azure, Google Cloud.
- Fair exposure to C++, NodeJS, GoLang, Java, Oracle, MS-SQL Server, Solaris OS.
- Engineering leader with a strong sense of ownership in everything I do.

I have spent most of my career developing RDBMS internals, including my work on Oracle's Cost Based Query Optimizer and active major contributions to the Postgres open source community; see the "Detailed Work Experience" section for details.

I have spoken/presented at various Conferences; see the "Speaking Engagements" section.

I have 8 years of experience as a Postgres Database Architect, Administrator, Trainer, and Consultant. In these roles I helped customers learn Postgres development, Postgres internals, helped tune their databases, migrated them from other databases to Postgres, as well as custom solutions development, and Postgres data recovery.

Qualifications:

1. Agile/Scrum Master
When: January, 2015
Institution: Scrum Alliance

2. Post Graduate Diploma in Advanced Computing
Duration: Feb 2001-July 2001
Institution: C-DAC (Center for Development of Advanced Computing)

3. Bachelor of Engineering (Mechanical)
Duration: 1996-2000
Institution: Gulbarga University, India.

Fields of Interest:

OS internals, Compilers, RDBMS, Distributed Systems, Linux Containers, Kubernetes, RISC-V.

Skills:

Programming Languages: C, Go lang, C++, Shell-scripting (Windows NT and Unix), and some assembly.

Operating Systems: Windows 9x, Windows NT, Sun Solaris, GNU-Linux.

Database Management Systems: Oracle, MS-SQL server, Sybase, SQLite, and an expert in Postgres RDBMS.

Web Technologies: NodeJS, AngularJS, XML, HTML, JavaScript, Web 2.0, PHP.

Work Experience (summary):

1. ***Open Source***
   Designation: Contributor / Hacker / Developer
   Duration: 2005 – Now

   Description: Contributed features and bug-fixes to Postgres project. Developed projects of my own. Speaker at conferences. See "Detailed Work Experience" section for details.

2. ***Supabase Inc.***
   Designation: Fellow Engineer
   Duration: Sep. 2021 – now

   Description: Contribute to Postgres community, and initiate & work on company's long-term, multi-year projects.

3. ***GE Digital***
   Designation: Principal Engineer
   Duration: May 2016 – Aug. 2021

   Description: Architected, developed, owned, and operated Postgres-as-a-Service at Predix Platform on AWS, as well as air-gapped Kubernetes platform. Service Owner of Blobstore service.

4. ***Amazon Web Services***
   Designation: Software Development Engineer
   Duration: May 2015 – May 2016

   Description: Architect and develop Amazon Aurora for Postgres. Help with Amazon RDS Postgres.

5. ***Thales e-Security, San Jose, CA., USA***
   Designation: Senior Engineer, Database Specialist
   Duration: Nov. 2014 – May 2015

   Description: Design and develop an HSM (Hardware Security Module) Manager using service-oriented architecture. Also, migrate a long-standing market-leading hardware key-manager from embedded MySQL to Postgres.

6. ***EnterpriseDB Corporation, Boston, MA., USA***
   Designation: Database Architect
   Duration: Feb. 2011 – Oct. 2014

   Designation: Principal Systems Engineer
   Duration: Feb. 2010 – Jan. 2011

   Description: Worked on database internals and contributed to Open Source Postgres database. My work as a primary/sole contributor includes Index Advisor, Reindex Primary Key Indexes, Postgres Hibernator, DBYardstick Benchmark development. I also co-lead the Postgres Plus Cloud Database and Postgres Enterprise Manager projects.

   In consulting role, I helped customers and partners perform Architectural Health Checks of Postgres databases, perform database migrations, conduct trainings, develop custom

replication solutions to other database, etc. while also participate in Postgres community contributing code patches and features.

7. ***EnterpriseDB Software India Pvt. Ltd., Pune, India.***
   Designation: Lead Engineer.
   Duration: July 2006 – Jan. 2010

   Description: Researching new features and implementing them in the EnterpriseDB RDBMS (which is based on Postgres). Act as Trainer, Consultant and Remote-DBA for customers and partners, on different occasions.

8. ***CommVault Systems (India) Pvt. Ltd., Hyderabad, India.***
   Designation: Software Engineer.
   Duration: May 2005 – July 2006

   Description: Develop and bug-fix the Report Generation Server, developed in C++, which interacts with MS-SQL Server for its data storage needs.

9. ***Oracle India Private Ltd., Bangalore, India.***
   Designation: Member Technical Staff.
   Duration: Nov 2003 – May 2005

   Description: Suggesting and implementing new features, and fixing defects in the *Cost Based Optimizer* module of the Oracle RDBMS, on Solaris and Linux platforms.

10. ***Aztec Software and Services Ltd., Bangalore, India.***
    Designation: Developer.
    Duration: Sept 2003 – Nov 2003

    Description: Designed *External Stored Procedures* feature for an RDBMS product, which was being developed for a customer.

11. ***Vaman Technologies (R&D) Ltd, Mumbai, India.***
    Designation: Junior Software Engineer.
    Duration: July 2001 – Sept 2003

    Description: Worked on various components of an RDBMS, including the TPC-C benchmarking utility, the Query Parser, Procedural Language support, OS independent kernel, and the OS abstraction layer.

Detailed Work Experience:

1. ***Open Source***
   Duration: 2005 – now

   Projects:
   - ***Postgres RDBMS***
     Language: *C*
     Role: *Contributor*

     Description: Postgres is the most advanced Open Source RDBMS

     Contributions:
     - Found and fixed a Security Vulnerability (commit, discussion, docs)
     - Replace Primary Key index of a table (commit, submission, discussion, docs)
     - Index Advisor extension (Git repository)
     - Postgres Hibernator extension (Git repository)
     - Improve interaction with Linux OOM Killer (commit, submission, discussion, docs)
       It was very difficult to convince Tom Lane that this idea had merit.
     - The \ir command for psql utility (commit, submission, review, docs)
     - Fix a Windows-only race condition (commit, submission)
     - Improve pgbench to support large databases (commit, submission, discussion)
     - pgbench: add client_id variable (commit, submission)
     - List of all my contributions: link to search results
     - Postgres-hackers mailing-list participation: link to search results

   - ***Speaking Engagements at Conferences***
     Role: *Speaker*

     - ***Postgres Open***
       2015 - Postgres Inside Out
       2014 - How Databases Work - For Developers, Accidental DBAs and Managers
       2013 - How Postgres is different from (read BETTER than) your database
       2011 - Introducing Postgres Enterprise Manager
       2011 - Why you should NOT move away from Oracle
       2011 - Postgres Plus Cloud Database

     - ***PGCon***(Postgres Hackers' annual conference)
       2011 - Best Practices with Ora2Pg (link)

   - ***Other Open Source Projects, and Work***

     - DBYardstick - Lightweight TPC-C Benchmark in NodeJS (Git repo)
     - ZFS File System driver for Docker containers (Git repo)
     - Postgres blogs: Current, Random Neural Firings, Left-Brain Dump

2. ***Supabase Inc.***
   Designation: Fellow Engineer
   Duration: Sep. 2021 – now

   Description: The company develops and provides Backend-as-a-Service, built using Postgres. It provides User Management, Authentication, Authorization, and automatic generation of REST APIs for the tables in the database.

   Contributions: My primary responsibilities are to contribute to the Postgres community, in the way of developing patches, reviewing patches, and participating in community discussions. My company-facing role involves initiating and working on projects that will be helpful to the company's future 3-5 years down the line.

3. ***GE Digital, San Ramon, CA, USA***
   Designation: Principal Engineer
   Duration: May 2016 – Aug. 2021

   Apart from the projects listed below, I am a Bar Raiser at the interviews. Along with another Principal Engineer, I also helped define the interviewing process to reduce false positives in phone interviews.

   Projects:
   Postgres-2.0 Service
   Language: Golang
   Team Size: 6
   Role: Creator, Service Owner

   Description: GE Digital's Predix platform helps Industrial companies collect and analyze data, and develop new applications to get new insights from their data. This project provides Postgres-as-a- Service to the Predix platform's customers.

   Contributions: With one team member, I started and released this project to replace the previous Postgres-as-a-Service since that had many architectural problems. This service provisions and manages Amazon RDS instances for customers. As a Service Owner, I am responsible for design, development, support, and maintenance of this service. I now lead the team of 6 engineers.

   With one team member I am developing a Kubernetes based Postgres-as-a-Service, as the next generation of this service.

   I'm also the Service Owner of the Blobstore service, and am responsible for feature enhancements, support, and maintenance of the service.

4. ***Amazon Web Services, Palo Alto, CA, USA***
   Designation: Software Development Engineer
   Duration: May 2015 – May 2016

   Projects:
   *i. **Amazon Aurora for Postgres***
   Language: C, C++
   Team Size: 8
   Role: Senior Engineer

Description: Amazon RDS is a market leader in providing Relational Databases in the Cloud. This project developed Postgres compatible Amazon Aurora database.

Contribution: I am one of the founding team of 3. I am the primary contributor to the Postgres Transaction Log implementation. I helped team with knowledge of other Postgres internals like shared- memory operations and disk-IO subsystems.

ii. ***Postgres RDS***
Language: C
Team Size: ~10
Role: Senior Engineer, Postgres Internals Specialist

Description: Amazon RDS is a market leader in providing Relational Databases in the Cloud. RDS Postgres team develops and maintains the Postgres engine within the Amazon RDS umbrella.

Contribution: Although I was hired for Amazon Aurora Postgres, due to my past experience with Postgres internals I was engaged with this team and quickly became a component owner of the Postgres code. I am involved in code reviews of the changes made to the Postgres database core and extensions, and I ensure the stability and security aspects of the Amazon-custom changes to the Postgres code.

5. ***Thales e-Security, San Jose, CA, USA.***
Designation: ***Senior Engineer, Database Specialist***
Duration: Nov. 2014 – till date

Projects:
i. Central Manager
Language: TypeScript, JavaScript, C++; NodeJS, AngularJS 1.x
Team Size: 8
Role: Senior Engineer, Database Specialist, Technology Generalist

Description: The company is a leader in Hardware Security Modules. Central Manager is being developed to monitor and manage the large estates of these devices.

Contribution: The project is in infancy, with aggressive release dates. I have been instrumental in strongly advocating and getting a buy-in for an array of technologies we'll use to develop this product: Vagrant and TypeScript needed a hard-sell, and also had to vote down some bad choices being made.

ii. KeyAuthority
Language: C++
Team Size: 3
Role: Senior Engineer, Database Specialist

Description: keyAuthority is a FIPS compliant Encryption Key Management (KMIP/TKLM) product.

Contribution: After years of success, the product has hit a performance wall in terms of load the embedded MySQL database can serve. I am responsible for migrating the product to Postgres, while keeping the custom secure replication and HA features intact, and improve the database performance.

6. ***EnterpriseDB Corporation, Boston, MA.***

Designation: Database Architect
Duration: Feb. 2011 – till date

Designation: Principal Systems Engineer
Duration: Feb. 2010 – Jan. 2011

Projects:

i. ***Enable Postgres to Reindex Primary Keys***
Language: C
Team Size: 1
Role: Architect and Engineer

Description: Reindexing the index of a primary key constraint in Postgres requires requires dropping and recreating the constraint, which also requires an exclusive lock for a prolonged period. This is impractical in production environment if the table is huge. This feature allows users to pre-create an index concurrently (which doesn't require exclusive lock), and then swap this index with the index behind the primary key in one quick command: ALTER TABLE ADD UNIQUE/PRIMARY KEY USING INDEX.

Contribution: I came up with the idea, wrote the code and submitted the patch. The maintainers of Postgres project reviewed it before applying the patch to the core Postgres.

ii. ***Postgres Hibernator***
Language: C
Team Size: 1
Role: Architect and Engineer

Description: After a restart, the database loses all the shared-memory cache, and this causes the application performance to be very slow until that cache is reconstructed over a long time. The Postgres Hibernator extension rebuilds this cache automatically after the restart so that the application performance after the restart is the same as before the restart.

Contribution: I came up with the idea, wrote the code and developed the extension.

iii. ***Found and Fixed a Security Bug in Postgres***
Language: C
Team Size: 1
Role: Architect and Engineer

Description: I discovered a security bug in Postgres which allows a normal database user to gain superuser privileges. I developed a bug-fix that prevents such escalation of privileges. This security vulnerability was assigned CVE number: CVE-2009-4136.

Contribution: Discovered and fixed the bug.

iv. ***DBYardstick: Lightweight Database Benchmark Application***
Language: JavaScript, TypeScript, NodeJS
Team Size: 1
Role: Architect and Engineer

Description: Running an OLTP benchmark requires a huge investment in hardware, and a large fraction of that is spent to buy many high-end workstations to run just the client applications. I implemented a lightweight OLTP/TPC-C benchmark that can run thousands of client applications on a mere laptop.

This implementation can reduce the cost of running TPC-C benchmarks significantly, and these savings can be used to buy better hardware to run the database system under test.

Contribution: Idea, research, implementation and scalability testing.

v. Postgres Plus Cloud Database

Language: Java
Team Size: 3-4
Role: Architect and Engineer

Description: PPCD is an application for database infrastructure provisioning, management, scalability and fault tolerance. It allows the user to create Postgres databases in the cloud (Amazon AWS, OpenStack, etc.), create read-replicas of that database, perform load balancing among replicas, increase storage/disk as needed, and perform automatic failover in case of master node failure.

Contribution: Design and architect the whole project in collaboration with Java application architects. Developed various components in a cloud-provider agnostic way, so that more cloud providers can be easily supported.

vi. Postgres Enterprise Manager

Language: SQL, C++
Team Size: 3-4
Role: Architect and Engineer

Description: PEM is an application for managing, monitoring and forecasting a large-scale installation of Postgres databases. It enables users to create custom alert notifications on abnormal conditions reported by the database agents, forecast the resource usage so that upgrades and maintenance operations can be planned.

Contribution: I designed and implemented extremely flexible and customizable alert notification subsystem. It allows the user to define new *types/templates* of alerts, and create alerts based on those templates, all without the need to wait for PEM developers to implement a new feature request.

vii. Enable pgbench to test large databases

Language: C
Team Size: 3
Role: Architect and Engineer

Description: pgbench is a TPC-B like benchmark utility, and it is the standard tool the Postgres project uses to test performance improvements in the database. pgbench was not capable of running tests for databases initialized with a scale-factor larger than 21474.

Contribution: I, along with other community members, improved the code to handle that case, while maintaining the validity of pgbench results obtained in the past.

viii. Enable psql to process package of scripts

Language: C
Team Size: 1
Role: Architect and Engineer

Description: psql is the de facto command-line utility/shell of the Postgres project. The feature enables the user to create a set of related SQL scripts that can be processed as one unit/package irrespective of the files' location in the filesystem.

Contribution: I came up with the idea and developed the feature.

ix. Miscellaneous

Other projects, customer-facing or otherwise, that I worked on for cannot be listed here for brevity, and for restrictions imposed by customers' Non Disclosure Agreements. In brief, though:

a) Migrated an ERP system from Sybase to Postgres
   This included full-text search migrations as well.
b) Performed Architectural Health Checks (AHC) of Postgres databases.
   Developed the first AHC code, procedures and documents; still in use today at EDB.
c) Recovering data from corrupted databases.
d) Slony/Postgres replication setup.
e) Auditing custom Postgres builds/packages.
f) RemoteDBA services on-boarding of customers.
g) Advising a customer on Postgres Internals, who chose to fork Postgres code.
h) Conducted DBA and Developer training.
i) Oracle to Postgres migration assessments.
j) Designed and developed a custom replication solution for Postgres-to-Oracle.
k) Presented talks at Postgres conferences.

7. EnterpriseDB Software India Pvt. Ltd., Pune, India.

Designation: Lead Engineer.
Duration: July 2006 - till date

Projects:

i) CRC Slicing-by-8
Language: *C*
Team Size: *1*
Role: *Developer*

Description: This is a novel fast CRC generation algorithm developed by Intel. This algorithm runs at under 3 cycles per byte (vs. 7) and thus triples the performance of existing CRC implementations. Slicing-by-8 uses an 8KB cache footprint.

Contribution: Integrated the 'Slicing-by-8' algorithm developed by Michael E. Kounavis, into the Postgres code and benchmarked the performance. Although it proved to be better than the standard CRC calculation, it was not accepted by the Postgres community since the performance of the algorithm heavily depended on the CPU cache's hotness with the algorithm's data.

ii) Optimizer Hints
Language: *C*
Team Size: *2*
Role: *Design and developement*

Description: This feature provides a way for the user to instruct the Query Optimizer which access paths to choose when the user knows that she has better knowledge about the data pattern than the Optimizer does.

Contribution: Did the Low Level Design of the feature, for implementing it as cleanly as possible in the Postgres' Scanner, Parser and Query Optimizer. Faced some issues with the Lex and Yacc implementations, hence implemented a separate hand-coded scanner-cum-parser just for recognizing the hints in the query.

Designed and implemented the FULL, INDEX, NO_INDEX, ALL_ROWS and FIRST_ROWS hints. Co-implemented and reviewed the USE_NL, NO_USE_NL, USE_HASH, NO_USE_HASH, USE_MERGE, NO_USE_MERGE hints. Designed and implemented Global Hints feature, which can be used to provide hints for relations/queries inside sub-queries and inline-views.

iii) Index Adviser
Language: *C*
Team Size: *1*
Role: *Developer*

Description: Index Adviser is a plugin for Postgres that leverages Postgres' capability to make use of external Query optimizers. It injects an additional step into the query optimization phase and creates virtual indexes based on the WHERE and GROUP BY clauses of the query. Based on the estimated difference in cost of the two phases, it recommends new indexes.

Contribution: Migrate the code from Postgres version 7.4 to 8.2; in the process, performed a major cleanup. Add additional features to take into consideration the new Bitmap Index Scans. Add the ability to store the recommendations in a user-definable physical relation. Submit it to the Postgres community, and follow it up till it is accepted by the community.

iv) Training

Conducted training for Customers and Partners, to train them on Postgres and EnterpriseDB on the following topics: SQL and PL/PGSQL programming for developers, internals and management tasks for DBAs. The training lasts between 3 to 5 days.

v) Consulting

Provide consultancy for Customers, wherein I provided help and support in Database development, administration and maintenance tasks.

vi) RemoteDBA

I acted as Remote-DBA for various clients, for periods ranging from 3 to 6 months. As a Remote-DBA I have been involved in architecture review, schema and data migration from other RDBMSs, do maintenance work like dumps/reloads, maintaining cron jobs, etc.

8. CommVault Systems (India) Pvt. Ltd., Hyderabad, India.

Designation: Software Engineer.
Duration: May 2005 – July 2006

<u>Modules worked on:</u>

i) QNet
Duration: May 2005 – July 2006
Language: *C++*, SQL, T-SQL
Team Size: *2*
Role: *Development and Maintenance*

Description: The QNet server is responsible for taking a client's requests, and generating XML report for the clients' consumption. The reports contain description of the various Backup and Restore tasks performed by the Flagship product CommVault-QiNetix. The data is stored in MS-SQL Server, and the front-end clients were written in Java.

Contribution: I added new features for one release and managed the whole code during the upgrade from one release to the other.

9. Oracle India Private Ltd., Bangalore.

Designation: Member Technical Staff.
Duration: Nov 2003 – May 2005

<u>Modules worked on:</u>

i) Cost Based (Query) Optimizer (CBO):
Duration: Nov 2003 – May 2005
Language: C
Team Size: 5
Role: Diagnostics and Defect Resolution

Description: The Cost Based Optimizer (CBO) tries many transformations with the query; some based on heuristics and some based on cost. Heuristics include view merging and subquery unnesting. The cost based decisions are made for access path analysis. Finally, the best plan is chosen based on the least cost.

Contribution: Resolving performance issues and bugs encountered in CBO. Also, suggested a new enhancement for optimization, but couldn't finish it by the time I left.

10. Aztec Software and Services Ltd., Bangalore.

Designation: Developer.
Duration: Sept 2003 – Nov 2003

<u>Projects Undertaken:</u>

i) Extended Stored Procedures (ESPs):
Duration: Sept 2003 – Nov 2003
Language: C++
Team Size: 2
Role: Developer

Description: ESP will enable the developers of Pervasive SQL RDBMS to call their external stored procedure, implemented in a language of their choice, from within the database.

Contribution: Low Level Design and Implementation.

11. Vaman Technologies (R&D) Ltd, Mumbai

Designation: Junior Software Engineer.
Duration: July, 2001-Sept, 2003

Company Profile: Developing an RDBMS named VAMAN, using internal-multithreading kernel (threads not visible to the OS, hence OS independent), and Finite State Automata. ANSI C is used as the programming language of choice, for its strength and portability. MS-VC++ 6.0 (on Windows-NT) and GCC (on GNU-Linux) were the tools being used for the development of the server and related software.

<u>Projects Undertaken:</u>

i) Extended Procedural Language (EPL)
Duration: Dec 2001 – Sept 2003 (18 months approx.)
Language: C
Team Size: 6
Role: Jr. Software Engineer (Design and Implementation)

Description: This module adds the procedural language support to the SQL based RDBMS server. It provides various 'High Level programming Language' (HLL) constructs such as 'IF-ELSE', 'FOR', 'WHILE', 'EXCEPTIONS', 'CURSORS', 'PACKAGES', etc.

This, otherwise an independent module, runs as a part of the server and is responsible for communicating with 'Data Manipulation Language' (DML) engine, for performing the data manipulation as desired by the 'EPL programmer'.

EPL also allows 'COMMIT'ing or Rolling back the transactions by communicating with the Transaction manager.

Contribution: I have been instrumental in implementing the 'Extended Procedural Language' module since its inception. I was responsible for implementing the 'CURSOR' construct, which fetches data from the DML engine and allows the EPL programmer to iterate through the result-set row-by-row, and perform various operations depending on the contents of the result-set.

I was also responsible for the communication of EPL with the Transaction manager.

ii) Windows-NT Services
Duration: 1 month
Language: C
Team Size: 1
Contribution: Integrated the RDBMS server with Windows-NT services. It allows the server to run with 'System' privileges as a part of the OS, also allowing the RDBMS server to be started and shutdown in sync with the Operating System.

iii) Kernel
Language: C
Task: Optimizations

Description: Vaman's Kernel provides cooperative multithreading, and it provides OS-independent and hardware-independent threading model, hence it is capable of providing multithreading even in a non-multitasking OS like MS-DOS. it provides various modern OS features, without relying on the underlying OS. These features include: Priority based scheduling, Event-driven architecture, semaphores, thread synchronisation, etc.

Contribution: Tried various optimisations, in various parts of the Kernel, like the scheduling algorithm. Optimised the way kernel used to handle the semaphores. Now, instead of iterating though all the semaphores created so far, it has to iterate through only those that have just been signaled.

iv) Resource Library (OSLib)
Language: C
Task: Improvements and bug removal

Description: This library provides OS independence. It is a collection of subroutines, written in C, that hide the differences in operating systems from the programmer, hence making the code portable. It consists of various modules, such as : Memory-Management, Process-Creation, Network-Communication, File-Handling, etc.

Contribution: Added the feature of assigning priorities to the new processes being created under various OSs. Also, removed a potential bug from the Memory manager.

v) SQL Query Parser
Duration: 2 months
Language: C
Team Size: 10
Role: Team Member

> ***Description:*** It is an object-based parser. It parses the input according to a dictionary, which dictates the rules of occurrences of keywords and identifiers, and fills an interface in the 'Object-Operation-Option-Value' format. This interface is then forwarded to the RDBMS server for execution.
>
> ***Contribution:*** Implemented the 'Grant/Revoke' commands in the Parser module. It prepares the interface describing the permissions a client wants to be Granted to or Revoked from another user. This interface is then used by the server to actually grant or revoke privileges from a user.

vi) ODBC Driver Development
Duration: 2 months

Language: C
Team Size: 3
Role: Jr. Software Engineer
Contribution: Implemented the 'SQLExecDirect' function in the ODBC Driver.

vii) TPC benchmark-C (TPC-C)
Duration: 1 month
Language: C, SQL
Team Size: 1

Contribution: Implemented the 'TPC benchmark-C' as described by 'Transaction Processing Council'. It was used for populating the database with the data in conformance with the TPC benchmark-C.

Also designed a primitive HTML based client, implemented as a DLL, which could be used for performing OLTP transactions on the populated database, in conformance with the benchmark.

viii) RDBMS Views
Duration: 1 month
Language: C
Team Size: 1

Contribution: Completely re-designed and re-implemented View-creation module, which was left mid-way by an ex-employee.