

Asif Hussain Shahid

25+ years of software development spanning low latency Distributed Systems, Querying Engines, Apache Spark Internals, with proficiency in Java / Scala

📍 Beaverton, OR USA 📞 +1 9712698126 📧 asif.shahid@kwikquery.com 🌐 <https://www.kwikquery.com>

Summary

27 years of experience in core java spanning high performance, concurrent access, low latency distributed in-memory data management, OQL (Object Query Language) & SQL querying engine development & index management , understanding of cost based optimizer of querying engine , **in depth knowledge of Apache Spark Catalyst and SQL layer.**

Implemented Http 1.1 protocol for page fetching, cookie management, using raw java sockets for page fetching(handling Get & Post requests with capability of authentication, Http 401 response code handling) for building web collaboration tool.

Knowledge of ANTLR & Parboiled based parser development , generation of AST (Abstract Syntax Tree) and Compiled Trees.Proficiency in the implementation of multi threaded, lock free & optimally locking, low latency java applications.

Proficiency in Java & Scala platforms. Good understanding of C & C++ fundamentals.

Experience

KwikQuery LLC

July 2025

Founder and Principal Engineer

Beaverton, OR USA

🌐 <http://www.kwikquery.com>

Started my own venture KwikQuery LLC, whose product TabbyDB is an extremely high performant fork of Apache Spark.

It solves complex issues like uncapped query tree size, blowing up of Constraints, impact of rules like DeduplicateRelations etc, all of which can result in compile time of complex queries to run into hours.

It also improves the runtime performance of queries by implementing dynamic file pruning for Broadcast Hash Joins.

TabbyDB can bring order of magnitude performance improvement in real world use cases.

TPCDS benchmark shows 35 % improvement compared to stock spark, as tested internally on 1 TB and 3 TB data, using Hive based parquet tables.

Cloudera Inc

Sep 2022 - Aprl 2025

Principal Engineer

Beaverton OR, USA

🌐 <https://www.cloudera.com>

Worked on Apache Spark SQL, Core and catalyst layers, to fix the functional & performance issues in Spark Engine. Opened multiple jira tickets and github PRs fixes, with the open source. Debugged the performance and functional issues in the customer environment like BOFA, Amazon, JPMC, Axis Bank, Nokia etc.

Worked on Iceberg core and Spark-Iceberg layer.

Workday Inc

Aug 2020 - Sep 2022

Principal Engineer

Beaverton OR USA

🌐 <https://www.workday.com>

Optimized the catalyst layer of apache spark fork being used.

1) Implemented a new algorithm for Constraint Propagation rule of the Optimizer which can speed up compilation time by a factor of 10 -100 in problematic cases.

Proposed an improvement Pull Request for the same with the Open Source Spark.

<https://github.com/apache/spark/pull/33983>

2) Presented the above solution in Databricks Spark Summit of 2021

https://databricks.com/session_na21/optimizing-the-catalyst-optimizer-for-complex-plans

3) Other optimizations to reduce the compile phase time of queries, like avoiding re-application of expensive optimizer rules, in part of trees which are guaranteed to remain unchanged, over optimization phase cycle.

Tibco (via SnappyData acquisition)

March 2019 - March 2020

Principal Architect

Beaverton, OR USA

🌐 <https://www.tibco.com>

Worked on computeDB, previously known as SnappyData.

Implemented the optimized (memory efficient & much faster) version of HashAggregateExec of Spark, which uses serialized data of Group By Keys in a byte based HashMap. Works for aggregate functions of fixed length DataType.

Acquired relatively good knowledge of Spark's code generation architecture for each stage, catalyst parser, Analyzer , Optimizer, Spark Plan generation & conversion of SparkPlan into java code for each stage.

SnappyData (incubated from Pivotal)

June 2017-March 2019

Staff Engineer

Beaverton, OR USA

Implemented approximate query processing (AQP) & error estimation by enhancing catalyst querying framework of apache spark for aggregate queries on stratified sample table.

Enhanced the spark's querying engine to apply Poisson Bootstrap analysis and Closed Form analysis as required, for error estimation.

Developed a connector for GemFire (Apache Geode) for exposing regions as external tables in SnappyData with automatic schema detection.

Implemented Policy based Row Level Security feature in SnappyData by filtering the rows as per security criteria specified in Policy Filter.

Pivotal Labs (GemStone/ Apache Geode)

September 2015 - June 2017

Staff Engineer

Beaverton OR USA

Worked on new generation OLAP/OLTP sql based database on top of Apache Spark, which became SnappyData

Pivotal Labs (GemStone Systems / Apache Geode)

November 2012 - September 2015

Staff Engineer

Utrecht, Netherlands

Software development in Core Java & Scala.

Added support for off heap memory storage of the data of GemFireXD (Relational distributed database of ours), ensuring complete data integrity & concurrency using lock free techniques as far as possible.

Deep integration of spark & spark streaming with GemFireXD. Enhanced the RDDs & DStream to provide support for continuous querying semantics on GemFireXD & collocating arbitrary DStream data with GemFireXD tables.

Modification of spark engine to use GemFireXD as the cluster manager with the spark executors & spark driver running inside the GemFireXD VMs. Implemented Scheduler & Executor framework for the same.

Exposing GemFireXD tables as RDD for data manipulation & taking advantage of colocation of Tables in the join operations of RDD to avoid shuffle.

Implemented support for aggregate functions & User Defined Aggregate Functions, Group By & Order By Clause for OQL (Object Querying Language) engine of GemFire (Apache Geode) .

eBuddy

September 2011 - November 2012

Sr. Java Backend Developer

Amsterdam, Netherlands

Enhanced the performance of eBuddy chat to be 6 times faster by optimal synchronization & efficient handling of byte data , optimizing the encoding/decoding conversions.

Reduced the memory foot print of image data handled by the client by 50%

Learnt & worked on Android

GemStone Systems (subsidiary of VMware group)

September 2004 - September 2011

Manager R & D

Pune India

Architecting the design, managing day to day working of team and implementing. Actively involved in coding in core java & completely hands on with it.

As a part of Gemstone's GemFire (Apache Geode) implemented the JTA spec for Transaction Manager, enhanced the Object Query Engine significantly with intelligent usage of indexes and increased the disk based persistence performance using Oplogging technique.

Wrote multi threaded , thread safe code for GemFire with judicious use of synchronization & other locking structures without sacrificing performance.

Worked on developing new product SQLFire which is a highly available, scalable and distributed relational database engine.

eGain Corporation

September 2000 - September 2004

Team Leader

Pune India

🌐 <https://www.egain.com>

In the capacity of team leader architected the design and development of collaboration proxy server for co-navigation of customer and agent representatives which involved understanding of http protocol, parsing of webpage, management of cookies, and retrieving webpages on behalf of web browsers using URLConnection objects and custom protocol where necessary. It also involved understanding the SSL communication and implementation of Trust Manager.

Implemented the pushlet mechanism for messaging with clients establishing nailed socket connections with the web server and , the web servers pushing data asynchronously to the client unlike the usual polling mechanism.

Designed the Chat server using the Pushlet mechanism.

Led the effort for successful completion of Suggestion Approval system & External Content Search &

import for the Knowledge Base module of the eCRM offering.

Onward Technologies Ltd

1999

Sr. Engineer Tech Support

Pune India

Learnt the PIMS system of Aspen Tech which is Linear programming package for identifying refinery throughputs & yields based on profit maximization functions.

Indian Oil Corporation

1997 - 1999

Production Engineer

Vadodara, India

Handled the FPU (Feed Preparation Unit) which provides the feed to the FCC (Fluidized Catalytic Cracking Unit) . Oversaw functioning of Distillation Column , Furnaces.

Education

Indian Institute of Technology Bombay

1993 - 1997

Chemical Engineering

B. Tech

7.57

Volunteering

Cat Adoption Team

2023

Hospital Shelter Assistant

Beaverton OR USA

🌐 <https://www.catadoption.org>

Cleaning of the kennels.

TAP

2023

Mentor

🌐 <https://www.tapcareers.io>

Teaching Algorithm development.

Skills

Apache Spark

SQL, Catalyst layer

●●●●○

Apache Iceberg

●●○○○

R&D

●●●●○

Debugging / Problem Solving

●●●●○

Awards

Tech Wizard

December 2021

WorkDay Inc

🌐 <https://www.workday.com>

For conceptualising and implementing alternate algorithm for Constraints management and propagation rule of Apache spark which reduced the compilation time of query plans from 8 hours to under 10 minutes

Performance Recognition Award

October 2014

Pivotal Labs

For development of GemFireXD (SQL based distributed relational database)

Performer of the Year

June 2006

GemStone Systems

For development of OQL (Object Query Language Engine), Disk based Persistence, HA Queues etc)

Special Recognition Award

April 2002

eGain Communications

🌐 <https://www.egain.com>

For development of eGain Interact (Co - browsing Proxy Server)

Languages

Java

●●●●○

Scala

●●●○○

C, C++

●●○○○