

# Avishek Sen Gupta

## Principal Consultant, ThoughtWorks

**Avishek Sen Gupta** works with ThoughtWorks as a **Principal Consultant** with over 20 years of experience in the creation of large-scale distributed systems involving integration of disparate components, legacy and greenfield. He consults with clients on design, architecture, technical choices, patterns, testing techniques and development methodologies. He has worked with C#/. NET development, J2EE technologies, Ruby on Rails, NodeJS, R for statistical analysis, Python for ML and Large Language Model-based solutions, and Objective C for mobile (iOS) development. His current interests include compiler engineering, integrating symbolic AI with connectionist AI, logic programming with Prolog/Datalog, and OCaml.

## Domains

- Automotive
- Payment Processing
- Retail
- Leasing
- Banking
- Healthcare Solutions
- Pharmacy Information Services
- Derivatives Trading
- Rail Transport
- Medical Liability Insurance
- Foreign Currency Exchange
- Textile Supply Chain
- Contact Management
- E-Learning Systems

## Open Source

**Cobol-REKT** <https://github.com/avishek-sen-gupta/cobol-rekt>: An evolving toolkit of capabilities helpful for reverse engineering legacy Cobol code. These capabilities range from building flowcharts of the underlying code to translating COBOL into a hybrid Intermediate Representation (graph- and instruction-based) potentially suitable for transpilation to other languages.

Engineers can pick components from this library to incorporate into their reverse engineering analysis workflows, or more cohesive products suitable for wider audiences like analysts.

**Tape/Z** <https://github.com/avishek-sen-gupta/tape-z>: A toolkit for analysing mainframe HLASM (High Level Assembler) code. The library provides capabilities for working with mainframe assembler code, including parsing, control flow

graph building, dependency tracing, and flowchart visualization capabilities.

**IRIS** [<https://github.com/asengupta/IRIS>] : Machine Vision / Machine Learning Library written in C++

**Duck-Angular** [<https://github.com/asengupta/duck-angular>]: A container for bootstrapping and testing AngularJS views and controllers in memory: no browser or external process needed.

**Exo** [<https://github.com/asengupta/Exo>] : A lightweight AOP framework for .NET, with some bundled aspects for immediate use.

**Basis-Processing** [<https://github.com/asengupta/Basis>] : A library for easily plotting and transforming arbitrary non-orthogonal coordinate systems in Ruby=Processing.

**Jquery-Jenkins-Radiator** [<https://github.com/asengupta/jquery-jenkins-radiator>] : A jQuery-based build radiator for Jenkins, with support for multiple radiators, independent refreshes and timeouts.

## Experience

### AI-assisted Reverse Engineering Product

Avishek Sen Gupta is currently engaged as a Principal in developing an internal AI-assisted product for reverse and forward engineering legacy codebases in multiple stacks including COBOL, HLASM, Java, C#, DB2 SQL, PL-SQL and T-SQL. The tool uses a combination of code graphs and Large Language Models to reverse engineer requirements from legacy code, and feed into an agent graph to build modernised code. This information is also used to extract higher-level domain capabilities from the system. This product has been successfully used in both PoCs and production systems, and continues to be under active development.

### AI-assisted Reverse + Forward Engineering of Mainframe Assembly+COBOL Legacy Systems at a US Financial Group offering insurance services

Avishek has been involved in building capabilities for reverse engineering requirements from mixed High Level Assembly Language and COBOL code for a legacy insurance policy administration system, and forward engineering it into modernised Spring-Java services.

### AI-assisted Reverse Engineering of COBOL-based Legacy Systems at a leading German Automotive Manufacturer

Avishek Sen Gupta has worked as a Principal Engineer on a team building a Proof of Concept - and ultimately productionising said concept - for the client. The legacy system of this client consists of several million lines of legacy COBOL code. Avishek worked with the team in understanding the problem space, working hands-on on the techniques to enable the solution to take full

advantage of the latent structure of the COBOL code, so as to generate more meaningful and accurate responses from Large Language Models. As part of this work, he has also released an open source library called Cobol-REKT, which is an evolving toolkit of capabilities helpful for reverse engineering COBOL source code, with optional integration with Large Language Models.

The client intends to scale this solution to other departments across the organisation facing similar challenges in legacy modernisation.

### **EU Data Evangelist**

Avishek Sen Gupta has worked as the Data Evangelist for the EU Region. His responsibilities during this time included;

- Building a solid technical foundation of data and ML for the EU markets to tap into
- Laying a solid consulting foundation that the EU markets can tap into for proposals
- Making EU Data a training ground for Data/AI practitioners for proposals and projects
- Growing Data capabilities as a core skill for the majority of developers in all regions of the EU
- Work with business leaders to materialise bets in new areas / domains involving Data
- Implement important ML papers as part of his technical learning journey in AI/ML. Examples include:
  - Transformers: An implementation of the original 2017 paper on Transformer architecture, which underlies Large Language Models
  - Plenoxels : An implementation of the Plenoxels paper in the field of NERFs (Neural Radiance Fields), a research area in Computer Vision

### **Market Technical Principal**

Avishek has worked as the Technical Principal for the UK region. His responsibilities included, but were not limited to:

- Working with stakeholders on new business pursuits, and performing technical due diligence on responses to RFPs (Request for Proposal)
- Working with multiple big-label accounts in the region to drive technical excellence, while maintaining technical standards in teams on said accounts through a combination of informal and formal channels of collaboration with associated technical leaders and the Head of Technology.
- Intervening directly in a hands-on manner (both technical and non-technical) on any accounts which needed consultation in areas not readily available within their teams. These include working with client CTOs, Heads of Engineering, and architects to empower both client and ThoughtWorks teams. A few examples of collaboration include:
  - Working with the CTO of Turkey’s leading modest fashion platform to help rearchitect and modernise the existing legacy workflow of their

warehouse processes to scale up to a multi-region warehouse strategy, as part of the company's growth plans.

- Consulting and working with software delivery leaders at one of UK's leading retail giants to improve their state of the art in truck load optimisation, which is crucial for maintaining the optimum consumption of transport resources when supplying stores from warehouses.
- Consulting with the Head of Technology and the CTO of one of UK's leading supplier of knowledge management systems for thousands of schools to support teams doing a rewrite and modernisation of their existing platform.
- Mentoring future Technical Principals and Tech Leads
- Working on Market Strategy
  - New models of client engagement
  - Identifying and growing current capabilities in the market to fit the needs of short-term and projected capabilities
  - Internal initiatives to lead and sustain a thriving Tech Forum to afford technical leaders on accounts to showcase and share technical learnings at a regional level.
  - Leading and sustaining a Tech Council on the market comprised of technical leads on accounts to promote and sustain technical excellence and assurance on accounts in a decentralised fashion.
- Working on exploring and promoting ideas and techniques to articulate value to clients using grounded, well-established economics techniques. These are documented in:
  - Every Software Engineer is an Economist
  - Every Software Engineer is an Accountant
  - Economic Factors in Software Architectural Decisions

#### **Client: Leading Latin America-based Retail Giant**

Avishek Sen Gupta has worked as one of the Account Technical Principals for a large Latin America-based retail giant with businesses in Chile, Peru, Colombia and Argentina.

In addition, he has worked as Architect and Technical Lead for building out a Payment Processor Integration Platform for a large Latin America-based retail giant with businesses in Chile, Peru, Colombia and Argentina. The important business features of this Platform include: - **Payment as a Service:** Will provide standardised API's for Business Units to integrate with multiple Payment Service Providers, and provide applicable payment options and workflows. - **Multiple Payment Options:** Payment Domain will support all the forms of payment, such as: CMR Card, External Credit/Debit card, Cash, Pay at store/agency, CMR points, Gift Cards, Transferencia, Sales Finance etc. - **Tenant-specific Configurations:** for Business Units to select and prioritise Payment modes and which Payment Service Providers they want to integrate with. - **Fraud Validation:** for Business Units to perform Fraud Validation with Cybersource as well as other configurable Fraud Validation Service Providers.

- **Flexible Out-of-Process Integration:** Can be used by multiple Business Units systems - e.g. Authorization can be invoked from ecommerce domain, and payment can be invoked from Point of Sale systems. Other system like OMS/WMS etc can also integrate with Payment.

The important technical features of this platform include: - Multiple Payment Options for a single Payment Intention to support Split Payments with Saga support (allowing distributed transactions with optional automatic rollback) - Ability to configure operations to occur in sync mode (wait for Provider to respond) or async mode (results are notified asynchronously) through simple configuration. - Independent fine-grained control over lifecycle of each Payment Intention Method - Independently configurable deployments of Payment workers (local, REST, Google PubSub, Kafka) - Support for Idempotent Requests - Durable, database-backed, event log which may be used for auditing - Event streams for all Payment Intention Method operations

This initiative involved concepting out the architecture of the platform, doing proofs of concept, analysis of APIs for integration with different Payment Service Providers. The following are the highlights of this platform. The technologies used are:

- NodeJS
- Kubernetes hosted on Google Cloud Platform
- Google PubSub for messaging
- Istio for Service Mesh capabilities (rate limiting, routing)

#### **Client: Leading US/UK Bank**

Avishek Sen Gupta has worked as part of a platform team for a Retail Bank. The responsibility of the Platform team was to build out deployment-level and application-level infrastructure for multiple product teams to rapidly build and deploy their customised products used by the bank's internal users as well as external customers. This initiative involved building out the platform. The following are the highlights of this platform. - CQRS (Command Query Responsibility Segregation) based read write consistency - Pure Event Sourcing using the Axon Event Framework (internally using the LMAX Disruptor architecture) - Event Replay - Compensating Events - Deploying and maintaining an in-house Kubernetes cluster with templates for use in rapid application bootstrapping - Maintaining legacy Puppet scripts

The technologies used are:

- Spring Boot
- Kubernetes
- Axon

#### **Client: Leading US Retail Giant**

Avishek Sen Gupta worked as Technical Principal for the client for a broad retail platform initiative. The client owns various brands of stores and wants a web

platform built that will allow for the brands to have their own flavor of the white labelled front end. This platform aims to modernise and integrate with various aspects of the client's business systems.

#### **Client: Leading US Provider of Pharmacy Solutions**

Avishek Sen Gupta has worked for the client as the Technical Principal, Technical Lead and Architect on the Pharmacy team, as part of a broader retail platform initiative. The client owns various brands of stores and wants a web platform built that will allow for the brands to have their own flavor of a white-labelled front end. This platform, which resembles a CMS, will enable the sites to share common features as well as allow for customization specific to a site (w.r.t content, look and feel and such). The platform also integrates with legacy systems that are also used by the client's other legacy software systems.

The technologies used are:

- AngularJS (including RequireJS, Mocha)
- Spring MVC for web interfaces

#### **Client: Global Non-Profit Microfinancing Organisation (Data Analysis)**

Avishek has worked as a Data Analyst for the client. The work involved analysing performance of various CKW (Community Knowledge Worker) initiatives across Africa to determine which CKW's were providing the maximum impact to the farming and animal husbandry communities, so as to optimise microfinance investment initiatives. As part of this work, Avishek built models to analyse the performance data, as well as visualisations to intuit the performance of these initiatives, based on various factors like geographical regions.

#### **Client: Global Non-Profit Microfinancing Organisation (Data Analysis)**

Avishek has worked as a Data Analyst for the client. The work involved analysing test scores across various districts in the state of Karnataka, India, conducted under the Anganwadi initiative (which aims to uplift children of underprivileged workers through education). The aim was to analyse the effectiveness of the intervention initiatives, as well as potentially detect exam fraud. Avishek built several models to explain the data, as well as predictors for potential future performance. This included: - Decision Trees - Naive Bayes Density Estimators as predictor of performance - Correlating geography with test performance - Statistics involving the distribution of data (Chi-Square Tests, etc.) - Visualisations like Box Plots, Parallel Coordinates, Probability Distributions

#### **Client: Trusted advisor and consultant firm to many of the world's most influential businesses and institutions**

Avishek has been involved in the development of a mobile dashboard app (for the iPad) for one of the clients serviced by the client. The app is responsible

for displaying projections, profiles and other interactive statistical visualisations. The libraries/technologies used were:

- AngularJS
- RequireJS
- Cordova (PhoneGap)

**Client: Leading Designer and Manufacturer of Networking Equipment**

Avishek has been involved in leading the effort to build a platform for the client's e-learning initiatives. The specific need was to supplant paper reading material for the client's proprietary networking courses with an iPad e-reader with helpful features like notes, annotations and full-text search.

The technologies and libraries used were:

- Node.js
- Q.js for promises
- Mocha for unit testing
- Sequelize for ORM
- Solr

Avishek has also worked on spikes for the offline version of the same product. As part of that, he has evaluated and is familiar with various Javascript dependency management frameworks like:

- Inject (LinkedIn)
- RequireJS

**Client: Leading community development microfinance organisation**

Avishek is currently working with the client as a data analyst, using surveys to evaluate the impact of the client's Knowledge Worker initiative, as well as inform future investment strategy in specific areas of the client's field operations.

This also involves cleaning and validating the existing data, which in many cases, is incomplete and/or corrupted. Avishek is using R and Ruby on this assignment.

**Client: Leading Telecom/Data Communications systems provider**

Avishek has consulted with the Data Analysis team within the client's center. Starting with a 3-day Object Bootcamp to introduce and reinforce the team's understanding of basic OOP concepts, he has worked with the team on the ground, looking into their legacy Hadoop-based ETL workflow system. He has been instrumental in suggesting, spiking, and showcasing ways to allow ETL workflows to be tested outside the Hadoop container, as well as incorporating logging and tracing workflows. In addition, he has worked with the team to identify specific static analysis rulesets which will provide maximally useful diagnostic information about the legacy codebase.

**Client: Leading US Healthcare Solutions provider**

Avishek has consulted with multiple development teams within the client's IT operations. Beginning with assessments of the the teams' delivery capability and the identification of major bottlenecks, he has been instrumental in coming up with technical recommendations required to streamline and speed up the delivery of the product.

Continuing beyond the assessment phase, he has worked on the ground with one such early adopter team, coaching team members on an individual basis, as well as introducing unfamiliar concepts to the team, and assisting them in evaluating options to help them accelerate their delivery timelines. In addition, he has spiked out specific solutions like Git-Clearcase bridge and Nant environment initialisation scripting to streamline the developers' daily activities.

**Client: Leading Business consulting organization**

Avishek has led development of an iPad newsletter for the client. The newsletter exposes current thinking and opinions on the latest trends in various business domains. This is built using native iOS components and libraries, and supports text, images and video to provide a pleasing user experience.

The text experience was designed using CoreText.

This project resulted in a simple, extensible framework for creating iPad publications.

Interesting and useful patterns discovered in the project are being documented in <http://www.avishek.net/blog/?cat=11>.

**Client: Leading Business consulting organisation**

Avishek has worked on building a contact management system for the client. This system is capable of tracking contacts created or modified by consultants in the field around the world; additionally, provides deduplication, custom conflict resolution according to specific rules. Additionally the platform supports syncing information with Microsoft Outlook; an addin for Outlook allows for custom information entry as well as in-app resolution of conflicting contact data.

Avishek was responsible for developing crosscutting components of the platform, like user-role based security and assistant access (where a business partner's corporate assistant can proxy as the partner).

He also introduced an interesting way of reducing timing uncertainties in the existing functional testing framework. He has written and OpenSourced an IL-rewriting AOP framework called Exo (available at <https://github.com/asengupta/Exo>)

**US-based daily, independent, award-winning news organisation**

This project involved parallelised adaptive bitrate encoding of video streams to speed up transcoding between video formats.



Avishek has worked on a custom Map-Reduce based pipeline for encoding videos for delivery through adaptive streaming protocols like HLS (HTTP Live Streaming) and Microsoft Silverlight's ismc format. This involved multiple jobs running as daemons, with the actual transcodes/encodes being done using ffmpeg/x264/mp4box/mp4split.

The platform is a Ruby backend with an RoR frontend.

**Client: Leading online UK derivatives trading corporation**

Avishek has been involved in working on the existing core account management system for the client. The system also provides services to other consumers. Technologies involved were:

- J2EE with Spring Remoting for interservice communication.
- Java ServerFaces for UI (MyFaces)
- Oracle 10g
- iBatis (DB access)
- Spring for dependency injection and service location.

Original responsibility included leading the team to build up a long-anticipated slew of features related to country-specific Taxation in the product.

He was also involved in driving the migration of the existing tightly-coupled Spring Remoting approach to a looser REST-style interface (POX). In this he:

- Worked on draft versions of requests and responses.
- Coordinating with consumers (and tech leads of those teams) to refine the approach.
- Spike on candidate REST server and client side libraries. Spikes were done on Apache CXF, following standard Rails idioms.

**Client: Leading online UK train ticket retailer**

Avishek has been involved in developing a portal-based solution for the client. The source code is inherited from a former vendor and requires significant refactoring. The project facilitates rapid deployment of new portal sites, a high degree of customisation for each portal. The implementation is in .NET 2.0 and uses DotNetNuke as the basic presentation layer, using a mixture of autogenerated code and XSLT for allowing customisation. The system interacts with multiple third party systems for checking journey availability, booking reservations, validating card payments, etc. Biztalk orchestrations handle crucial workflow steps like payment, and an Oracle DB serves as the backing store. His major areas of contribution so far include:

1. refactoring old DNN modules to use XSLT controls
2. working on a generic exception handling framework at the presentation layer
3. setting up and working on Cruise (Continuous Integration), refactoring build scripts.

4. Patching Nant to identify, collect and profile bottlenecks and hotspots in the build from a central Rails-based server.
5. Working on end-to-end delivery of new key functionality like online delivery of tickets, including coordinating with TWI and third party dev teams to achieve this.
6. Enabling good practices, holding sessions on testing application code as well as hard-to-touch ASP.NET framework code.

#### **Client: High Profile US Textile Converter**

Avishek has been involved in developing the inventory management system for the client. The technology employed is Ruby on Rails 1.2. Responsibilities included assuming a mentoring role to ramp up new team members and providing technical direction to the project.

The tools used were:

- Source control system: SVN.
- IDE :IntelliJ IDEA, Ruby plugin
- Unit Testing: RUnit
- Persistence: ActiveRecord
- User Interface: ActionView
- Framework: Ruby on Rails

#### **Client: Major UK Pharmaceutical Information Services Provider**

Avishek has been involved in providing technical input for the inception/discovery phase of a data sourcing engine. The engine is to be the backbone for sourcing enormous quantities of heterogeneous data and assimilate them into a central repository. Subscription to this repository will let commercial users find peer-reviewed, authenticated, statistically relevant data on which to base their future directions of research.

The project would mostly involve J2EE technologies, and use scripting languages as necessary.

His major area of contribution was developing spikes (focused proof-of-concept chunks of code) to demonstrate potential solutions and roadblocks in terms of scalability and data integrity.

#### **Client: Canada-based Foreign Currency Exchange Company**

Avishek has been involved on a project developing multiple products for a foreign exchange transactions system. The applications are both thick client (WinForms) as well as web applications (ASP.NET 1.1). The engagement involves not only development, but also close coordination with the end users and the client business users to understand the continuously evolving software requirements.

The project is a cosourced project, and involves enablement and delivery, and is distributed across three locations. The total team size is at least 30 people, distributed across all locations.

The tools used were:

- Source control system: CVS.
- IDE : Visual Studio.NET 2003
- Unit Testing: NUnit
- Persistence: Custom Mappers
- User Interface: WinForms, ASP.NET 1.1

The key developments of this project are WebTofu, a web testing tool for test-driven user interface development.

#### **Client: US-based Leasing client**

Avishek has been involved on a project developing an end-to-end lease management system. The application architecture consists of a Swing front end using XML/HTTP to talk to a Servlet/EJB server, which used JDBC to talk to the database. Specific products used include Web Logic 8.1, Oracle 9i, and JDK 1.4.

This was a distributed, cosourced project with at least 70 developers at its peak. The methodology followed was XP. In order to make following the practices easier in the large team, it was split into five teams, each focusing on a particular area of the code base. Frequent team rotations and tech stand ups were done to facilitate information transfer.

The tools used were:

- Source control system : StarTeam, Perforce.
- IDE : IntelliJ IDEA
- UI Testing Suite: Abbot/Frankenstein
- Unit Testing: JUnit, nFIT, FitNesse
- Persistence: Hibernate
- User Interface: Swing

#### **Client: US-based Medical Liability Insurance Company**

Avishek has been involved on the project involving development of an insurance rating engine and associated rates management system on the .NET platform using Visual Studio, specifically C#.

The total team size was around 14 developers at its peak. The project was a co-sourced project and involved distributed development. The methodology used was XP, with its practices tailored for the distributed nature of the project.

He has also been involved in developing .NET addins (for integration with the application) for Microsoft Office 2003 (specifically Outlook, Word and Excel), using COM Interop classes available in .NET for document management.

The key outcomes of this project were a new OpenSource WinForms testing tool called SharpRobo, and a tool called SharpWiki to make writing FIT tests into JSPWiki easier.

The tools used were:

- Source control system: CVS, Subversion
- IDE : Visual Studio.NET 2003
- Testing: nFIT, NUnit
- Persistence: NHibernate
- User Interface: WinForms

### **Training Experience**

Avishek has driven several developer workshops called Object Bootcamps in university campuses, external companies, as well as inside ThoughtWorks to enable attendees imbibe key TW developer values. He has also driven role-independent TWI induction programs called Immersions.

These are in addition to the several ad-hoc and planned sessions he has done on projects.

He has also spoken in internal ThoughtWorks technical conferences.

### **Technologies worked on**

- **Programming languages:** Java, Javascript, C, C++, C#, Ruby, R, Python, JavaScript, OCaml, Prolog, Objective C
- **Python:** PyTorch, NetworkX
- **Java:** Spring Boot, Swing, JavaServer Faces
- **.NET:** .NET 3.5, WinForms, ASP.NET, Silverlight
- **Ruby:** Ruby on Rails
- **JavaScript:** Node.js, AngularJS
- **Databases:** MS SQL Server, Oracle, MySQL, PostgreSQL
- **Testing:** JUnit, NUnit, FIT, Selenium/RC, Mocha, Jest, Tap
- **Build tools:** Ant, NAnt, Rake, Grunt
- **Infrastructure:** Kubernetes
- **IDE:** Eclipse, Visual Studio.NET, IntelliJ IDEA, Arachno Ruby IDE, XCode
- **Source control systems:** CVS, Subversion, Perforce, ClearCase, Git
- **GUI libraries:** AngularJS, FLTK (C++), Swing, WinForms (.NET), GTKmm (C++), FOX, SDL (Simple Directmedia Layer), iOS Core Foundation
- **ORM Tools:** Hibernate, NHibernate, iBatis, Sequelize, Knex
- **Continuous Integration:** CruiseControl.NET, CruiseControl.rb, Cruise Enterprise, Gitlab, Jenkins
- **Business Workflow:** Microsoft BizTalk 2006

### **Education**

- **Master of Science in Artificial Intelligence / Machine Learning,** Liverpool John Moores University, UK
- **Bachelor of Engineering, Electronics and Communications,** PES Institute of Technology. (Vishveswaraiah Technological University), India