Rahul Jain

Head of Business Intelligence and Data Engineering at Omio

rahulj51@gmail.com • +49.1744884738 • Berlin • Germany

 rahulj51.github.io/ • linkedin.com/in/rahul-jain-83055b45 • twitter.com/rahulj51

Summary

I am a technical leader with 20 years of experience in building and running software products for various domains. For the last three years, I have led the Business Intelligence and Data Engineering function at Omio. In my current role, I lead teams of smart data engineers, analysts and product owners to build a strong business intelligence and data engineering foundation at Omio.

Some of my responsibilities include

a. Developing a Business Intelligence / data-platform strategy for the organization and setting a road map for its evolution.

b. Managing the data ecosystem and platform for Omio. This includes multiple data-sources, ingestion systems, storage (data warehouse and lake) as well as analytics/visualization.

c. Managing data governance and GDPR within the purview of the analytics data org.

d. Driving data-ops and self-servability of analytics at Omio in order to enable teams to take better data-driven decisions without depending on a central BI team.

e. Working with multiple stakeholders including across the organizations to provide reliable data for their rapidly growing needs.

f. Providing technical leadership with respect to system design and architecture, managing tech debt and future planning and prototyping to keep our data engineering systems up-to-date.

g. Growing a motivated team with active participation in their career development.

h. Managing data infrastructure cost.

Under my leadership, my team has evolved the practice from a tiny data warehouse into a sophisticated big-data platform based on modern engineering principles. We now run several data pipelines processing hundreds of terabytes of data each day in batch as well as real-time.

I am an engineer at heart and fairly hands-on with coding. I am also active in many open source communities in the data space.

Outside work, I am passionate about giving back to the community by mentoring budding engineers and engineering managers across the world. I offer free mentoring over hangouts as a member of a mentoring club. Occasionally, I provide consulting to startups in Europe and India on their data engineering setup.

I write frequently on data strategy, engineering and leadership on my blog, LinkedIn and Twitter.

Featured Publications

Data Team Roadmap A detailed post about my experience creating a 2021 strategy roadmap for my team at Omio. [https://rahulj51.github.io/data/planning/roadmap/2020/10/15/data-roadmap.html]

Guiding principles for a data engineering team A suggestion on a set of guiding principles that a Data Team could use as a north star in decision making. [https://rahulj51.github.io/data-engineering/principles/2020/08/29/Guiding-principles-for-a-data-engineering-team.html]

Migrating a data warehouse across Cloud providers My experience and approach as I led my team to migrate our entire data warehouse from Redshift to BigQuery . [https://medium.com/omio-engineering/how-we-migrated-our-data-warehouse-from-redshift-to-bigquery-89f33988e1b5]

Why hiring funnels don't work and what hiring managers can do about it. [https://linkedin.com/posts/rahul-jain-83055b45_the-reason-why-funnel-based-hiring-is-broken-activity-6738053882510917632-xqTR/]

Why data lakes are an anti-pattern A short twitter thread about problems with uncurated data-lakes. [https://twitter.com/rahulj51/status/1332007621814087680]

Work Experience

Omio (formerly GoEuro)   BERLIN, GERMANY
Principal Engineering Manager, BI   *Nov 2017 – Present (3 years 1 month +)*
Responsibilities include:
  – Enabling the data engineering team to meet their short and long term goals.
  – Providing a vision to the team for the evolution of our data platform and data strategy.
  – Providing technical leadership to my team with respect to system design and architecture, managing
    tech debt and future planning and prototyping to keep our engineering systems up-to-date.
  – Working with multiple stakeholders including BI analysts, data scientists and other business
    functions to provide data for their rapidly growing needs.
  – Working with the engineering leadership to establish technical best practices in the area of data
    engineering
  – Growing a motivated team with regular feedback cycles and participation in their career develop-
    ment.
  – Managing data infrastructure cost
  – Managing data governance around privacy (GDPR), meta-data management and data quality
    management.

Key accomplishments so far:
  – Evolved the data engineering architecture from an early stage batch heavy system to a state of the
    art data platform based on **real time streaming systems**.
  – Stabilized the data pipelines from an MTBF of 10 days to 90 days.
  – Significantly improved timeliness, quality and reliability of analytics data.
  – Led the development of self serve, **real-time data ingestion platform based on Kafka and Apache
    Spark** that enabled any team at Omio to setup their own ETL pipeline to send data to warehouse.
  – Grew the BI engineering team from 4 members to 10.
  – Executed a complex data warehouse migration project to move from Redshift to BigQuery. This
    migration resulted in an annual saving of $200K in cloud cost.

The Mentoring Club   BERLIN, GERMANY
Mentor   *Apr 2020 – Present (8 months +)*
Offering guidance and coaching to engineers and new engineering managers on varied topics such as
career growth, engineering leadership. - `https://www.mentoring-club.com/the-mentors/rahul-j`
`ain`

Thoughtworks   BANGALORE, CHENNAI & BOSTON
Lead Consultant and Tech Lead   *Oct 2012 – Nov 2017 (5 years 3 months)*
Key responsibilities included:
  – Technical Leadership on Thoughtworks projects.
  – Responsible for end to end delivery of product features.
  – Architecture and Design of software systems.
  – Implementing agile technical practices and processes
  – Agile Consulting

Built a pricing decision making system for a Thoughtworks client to enable pricing managers in retail
organisations make informed decisions about their pricing strategies. As the technical lead for this
product, I was involved in
  – Architecture and design decisions for the Big-data toolset (HDFS/Spark/ElasticSearch). The
    product ingested and processed very high volume data (1 few hundred million rows per week).
  – Performance engineering and tuning of ElasticSearch cluster for approx. a few hundred million
    writes (ETL) and complex reads.
  – Development and support of the platform's various business processes as micro-services (using
    Golang and Scala)
  – Developing the front-end of this product using React.js/Redux

Earlier Experience:
  – Creating a set of RESTful APIs (Dropwizard/Jersey/Mulesoft/Moodle) to allow external e-learning
    systems to integrate with a learning management portal.
  – Enhanced a data-ingest pipeline for a learning management portal by replacing ActiveMQ with
    Apache Kafka and Apache Camel for better reliability and performance. The pipeline manages
    asynchronous processes like search-indexing and transformation for the portal.
  – Created a set of Restful APIs for a very large US based travel portal. These APIs allowed adver-

tising partners of the client to download various reports of financial and commercial value. This application was built on top of the client's Apache Hadoop platform (using Apache Hive and Oozie workflows) and integrated tightly with other information systems used by them.

Credit Suisse   Singapore

Senior Analyst and Java Development Lead   *Jul 2007 - Sep 2012 (5 years 3 months)*

I was mainly responsible for writing/enhancing/maintaining a multitude of java/j2ee based processes used for ingesting large volumes of financial market data (balance sheets, stock prices, earnings estimates etc.) into our software systems and developing several downstream processes (applications, reporting frameworks, ETL tools etc.) that use this data. I also developed web based dashboard to help our production support team monitor all production processes from a single platform.

Infosys   Bangalore, Hyderabad and Boston

Software Developer & Project Lead   *Jan 2000 - Jun 2007 (7 years 6 months)*

Worked on various projects as a senior developer and later as a project lead. Handled multiple responsibilities including

– Software development and programming
– SoftwareDesign
– Testing/Debugging
– Performance testing
– Onsite coordination at client site
– Technical Leadership

Please refer to my LinkedIn profile for a more complete list of work experience along with recommendations.

Education

Indian Institute of Technology (ISM)   Dhanbad, India

Bachelor of Technology   *1995 – 1999*

Main project on creating a simulation model of the grinding efficiency of a ball-mill efficiency based on feed properties.

Skills

Technical expertise:

– Engineering Leadership, strategy, planning and execution

– Agile software development processes

– Software design principles and enterprise integration patterns

– Microservices and service-oriented architecture

– Data engineering principles and DataOps

– CI/CD pipelines and modern software release processes

– Programming languages - Java, Python, Scala

– Distributed processing and storage systems - Kafka, ElasticSearch, Apache Spark

– Cloud platforms - AWS and GCP

– MPP datawarehouses - Redshift and BigQuery

– Data Modeling and data warehousing. Data lake Architecture

– ETL orchestration frameworks like Airflow

– DevOps Toolchain - Virtualization, Containers (Docker) and Kubernetes