

Thiago da Mata

Big Data Software Engineer

Thiago Henrique Ramos da Mata

+61 0413 065 180

e-mail: thiago.henrique.mata@gmail.com

Twitter: [@thiagomata](https://twitter.com/thiagomata)

<http://thiagomata.com>

<https://www.linkedin.com/in/thiagomata/>

<https://github.com/thiagomata>

<https://stackoverflow.com/users/456164>

Australian Resident - Visa Subclass 186

Summary

I started working in development as a software engineer in 2001. I have held many positions, including Front-end Developer, Database Administrator, Process Analyst, Developer Lead, Data Engineer, and Chief Technology Officer. These different perspectives helped me better understand each area's particular needs and integrate them properly.

In recent years, I have specialised in Big Data, Cloud Computing and Machine Learning. I have worked with real-time systems, clusters, distributed systems, distributed storage, distributed search engines, functional programming, data streams, data processing pipelines and key-value and graph databases.

I am always searching for a balance between product demands and strategic needs by observing the complexity of implementation and the value proposition. I can combine the technical and academic framework, design and analysis aptitude, and communication skills to interact with the tech team, investors, clients, and executives, using prototyping, experimentation, data, and facts to aid decision-making.

I prefer a job where I can work and solve problems creatively. I seek opportunities to gain valuable skills while using my knowledge and capability to improve the company by developing and creating impactful, useful, reliable, elegant, clear, and thrifty products.

Education

Master's Degree, Biomedical Engineering

2011 - 2013

Universidade de Brasília - UnB with 4 publications.

Graduate Degree, Computer Science

2002 - 2008

Universidade de Brasília - UnB.

Experience

Senior Data Engineer - ROKT

<https://rokt.com/>

Sydney, Australia

2019 - Present

Initially, I worked in the Customer Data Platform, a real-time distributed system to organise, clean, validate and enrich customer data. Then, we created the Identity system that groups the customer's interactions and helps the company understand their behaviour better. We also created the Attribution system that combines customer interactions with advertiser conversion events to provide campaign success metrics based on customer behaviour. These metrics are essential to train and evaluate the machine learning models. Currently, I'm working on the Traffic Resilience team, which uses real-time, offline analysis and statistical inference to monitor and safeguard all company traffic, apply rate limits, ensure local residency, and detect anomalies and bots. These evaluations run in the Kubernetes Edge Cluster using Istio, Envoy Proxy integrated with our services via GRPC.

5 years

Git, Java, Scala, Python, Go, Gremlin, Spark, Kafka, AWS SQS / SNS, AWS Redshift, AWS Neptune, AWS EC2, AWS EKS, AWS CloudFormation, AWS CloudWatch, AWS DynamoDB, Terraform, Elasticsearch, Redis, HBase, Grafana, Prometheus, DataDog, CICD, Buildkite, Docker, Helm, Kubernetes, Istio, Envoy and GRPC.

Senior Data Engineer - SLI Systems

<https://www.sli-systems.com/>

Christchurch, New Zealand

2017 - 2019

On the Analytics Team, we worked on adding new features to current products, creating new products with a high volume of data, prototyping, and online and offline testing new machine learning strategies using Cloud Computing. I translated the models from the Data Scientist to SQL and Spark Streaming. We also moved many company services from Apache Spark to Apache Beam, improving speed and cost.

1 year and 8 Months

Git, Java, Python, Google BigQuery, Apache Beam, Google DataFlow, Google Cloud Storage, Google Spanner, Google Pub/Sub, ElasticSearch, CICD, Jenkins, Docker, Helm and Kubernetes.

Chief Technology Officer - Ideia Big Data

<http://ideiabigdata.com>

Brasília, Federal District, Brazil

2014 - 2017

I guided the company's strategy and vision in designing, modelling, and developing its product for research, microtargeting, communication, and statistics using publicly available demographic data, distributed systems, machine learning, and streaming on the cloud. We had to deal with Data Cleaning, Data Mining, Feature Extraction and Data Visualization. I also improved the company development process and training policy, defined the career ladder, reported the company strategy and vision to investors, improved the observability and monitoring, evaluated product existing and new features with clients, measured their engagement and feedback, and optimised the cloud infrastructure for scalability and reliability while reducing the cost.

3 years

Git, PHP, Python, R, Scala, Node.JS, Postgres, Hadoop, Apache Spark, Redis, Memcached, Grafana, MongoDB, AWS EC2, AWS SQS / SNS, AWS S3, AWS Elastic Beanstalk, Cloudera, Elasticsearch, Vagrant, Docker, Angular, Laravel, Cordova, Bootstrap, JQuery, D3.js and PDF.js, Karma and Selenium.

| | |
|--|-------------------------|
| <p>Senior Programmer - Adjetiva http://www.adjetiva.com.br/ Brasília, Federal District, Brazil</p> | 2014 - 2014 |
| <p>Here, I created the brand channel of Embratur, mobile applications for the World Cup Events, educational games, and the Fiat EcoDrive desktop application. We also developed the SAT-JF14, a mobile application winner of 3 Lion Cannes.</p> | 10 months |
| <p>Java, Node.js, Angular 2, PHP, Laravel, Postgres, UML, Scrum, Cordova, Node Webkit and Phaser.</p> | |
| <p>Co-Founder - OtimiCar http://www.otimicar.com Brasília, Federal District, Brazil</p> | 2013 - 2014 4 months |
| <p>Otimicar is a hardware and app solution that converts any car into a smart car. In Otimicar, I worked with the first version of the server's specification and development. This server communicates with the hardware devices and some partners to sync the data about the cars and send recommendations for services and maintenance to the car owners. It also created reports to the partners about their customers and their products.</p> | |
| <p>Git, Java, Node.js, Rest, Soap, MongoDB, UML, BPMN, Postgres, Google Web App and IoT.</p> | |
| <p>Senior Coordinator - Grupo 954 http://www.grupo954.com Brasília, Federal District, Brazil</p> | 2012 - 2013 |
| <p>The Grupo 954 provided clients with many data analyses, reports, and prediction models, including General Motors, one of Brazil's most significant car companies. I have modelled the internal business processes for estimating, specification and development for the company. Most of all, I worked developing web applications for the company's clients.</p> | 1 year and 6 months |
| <p>Git, PHP, Oracle Database, MsSQL Server, Rest, UML, BPMN and Scrum.</p> | |
| <p>Process Analyst - INEP http://www.inep.gov.br/ Brasília, Federal District, Brazil</p> | 2009 - 2012 |
| <p>Here, I worked on many software systems with high availability, high data volumes, and a high number of concurrent users, around hundreds of millions, including the E-MEC, which is software for educational accreditation and quality assurance for all universities in the country, and Revalida, the National Examination for the Revalidation of Medical Diplomas Issued by Foreign Higher Education Institutions. I have also worked on the specification and development of many internal systems.</p> | 3 years and 1 month |
| <p>PHP, Laravel, Zend, Java, JBoss Seam, Memcached, Oracle, Redis, BPMN and Scrum.</p> | |
| <p>Teacher - Studio On-Line http://www.studioonline.com.br/ Brasília, Federal District, Brazil</p> | 2004 - 2010 |
| <p>Here, I taught PHP, MySQL, HTML, CSS and Javascript.</p> | 6 years |
| <p>PHP, MySQL, HTML, CSS, Javascript</p> | |

| | |
|---|---------------------|
| <p>Analyst and Web Developer - SECOM http://www.secom.gov.br/ Brasília, Federal District, Brazil</p> | 2008-2009 |
| <p>I worked with a Survey of Requirements, Modeling, Prototyping and Web Development there.</p> | 7 months |
| <p>PHP, Oracle Database, Postgres, Zend and Javascript.</p> | |
| <p>Database Administrator - Bancoob http://www.bancoob.com.br/ Brasília, Federal District, Brazil</p> | 2007 - 2008 |
| <p>Here, I developed the system for portfolio credit transfers with a significant volume of financial transactions and ensured the accounting closing process. I used storage procedures, memory tables, transactions and bulk loading to deal with heavy data manipulation.</p> | 1 year and 2 months |
| <p>Visual Basic, MsSQL Server, Java, Crystal Reports, and UML.</p> | |
| <p>Web Developer - Ministério da Educação http://www.mec.gov.br/ Brasília, Federal District, Brazil</p> | 2006 - 2007 |
| <p>We developed the System of Information and Supervision of Professional Education and Technology and the System of Integrated Planning, Budget and Finance of Brazil's Ministry of Education. Those systems deal with a high volume of real-time and batch data and provide fast and dynamic reports.</p> | 1 year |
| <p>SVN, PHP, Zend, Postgres, Javascript, and SVG.</p> | |
| <p>Senior Developer - AgênciaClick - Isobar https://www.isobar.com Brasília, Federal District, Brazil</p> | 2005 - 2006 |
| <p>Here, we developed the SISAC - System for Monitoring and Controlling Communication Actions to the Presidency of the Republic of Brazil, the BrazilTour site, and the RMA - Network of NGOs of the Atlantic Forest site.</p> | 1 year |
| <p>PHP, Postgres, Oracle and Javascript.</p> | |
| <p>Developer - SAMU Brasília, Federal District, Brazil</p> | 2005 - 2005 |
| <p>The SAMU is a critical call centre software for emergency calls that connects the patient, ambulance driver, and medical team in one call. There, I worked as a front-end Software Developer to create the Internal System of Ambulance Request Manager.</p> | 6 months |
| <p>Java, JSTL, CSS, Javascript, Struts and XUL.</p> | |
| <p>Developer - M.S.D. Brasília, Federal District, Brazil</p> | 2003 - 2004 |
| <p>Here, we developed some computer-based training using Adobe Authorware, Adobe Flash, Javascript and ToolBook.</p> | 11 months |
| <p>Javascript, ToolBook, Adobe Authorware, Computer-Based Training and Adobe Flash.</p> | |

Knowledge and Skills

| | | | |
|------------------------------|-----------------------|------------------------|---------------------|
| <i>PHP</i> | <i>Javascript</i> | <i>Python</i> | <i>Node.js</i> |
| <i>Java</i> | <i>Scala</i> | <i>Go</i> | <i>Visual Basic</i> |
| <i>C#</i> | <i>Dafny</i> | <i>Haskell</i> | <i>R</i> |
| <i>Spark</i> | <i>Hadoop</i> | <i>Apache Beam</i> | <i>Kafka</i> |
| <i>MsSQL Server</i> | <i>Postgres</i> | <i>Oracle Database</i> | <i>MySQL</i> |
| <i>MongoDB</i> | <i>Redis</i> | <i>ElasticSearch</i> | <i>HBase</i> |
| <i>AWS EMR</i> | <i>AWS Athena</i> | <i>AWS DynamoDB</i> | <i>AWS Redshift</i> |
| <i>AWS Elastic Beanstalk</i> | <i>AWS Lambda</i> | <i>AWS SQS/SNS</i> | <i>AWS EC2</i> |
| <i>AWS CloudFormation</i> | <i>AWS CloudWatch</i> | <i>AWS CW Logs</i> | <i>AWS Neptune</i> |
| <i>Google Big Table</i> | <i>Google Spanner</i> | <i>Oracle Cloud</i> | <i>Terraform</i> |
| <i>Jenkins</i> | <i>BuildKite</i> | <i>Git</i> | <i>Bash</i> |
| <i>Kubernetes</i> | <i>Docker</i> | <i>Istio</i> | <i>Envoy</i> |

Languages

| | | | |
|----------------|-------------------|-------------------|---------------|
| <i>English</i> | <i>71/100 PTE</i> | <i>Portuguese</i> | <i>Native</i> |
|----------------|-------------------|-------------------|---------------|

Certifications

| | |
|---|------|
| <u>Parallel programming</u> | 2024 |
| École Polytechnique Fédérale de Lausanne | |
| <u>Programming with Google Go</u> | 2023 |
| University of California, Irvine | |
| <u>Solutions Architect Associate</u> | 2022 |
| Amazon Web Services | |
| <u>Functional Program Design in Scala</u> | 2021 |
| École Polytechnique Fédérale de Lausanne | |

2020

[Introduction to Graph Databases, Cypher, and Neo4j](#)

Pluralsight

[Practical Machine Learning](#)

2019

Johns Hopkins University

[AWS Fundamentals: Going Cloud-Native](#)

2019

Amazon Web Services

[Regression Models](#)

2019

Johns Hopkins University

[Deep Learning in Python Course](#)

2017

DataCamp

[Leveraging Unstructured Data with Cloud Dataproc](#)

2017

Google Cloud Platform

[Big Data and Machine Learning Fundamentals](#)

2017

Google Cloud Platform

[Apache Spark 2.0 + Scala: DO Big Data Analytics & ML](#)

2017

Udemy - V2 Maestros

[Reproducible Research](#)

2017

Johns Hopkins University

[Functional Programming Principles in Scala](#)

2016

École Polytechnique Fédérale de Lausanne

[Exploratory Data Analysis](#)

2016

Johns Hopkins University

[Getting and Cleaning Data](#)

2016

Johns Hopkins University

[R Programming](#)

2016

Johns Hopkins University

[Introduction to Probability and Data](#)

2016

Duke University

[The Data Scientist's Toolbox](#)

2016

Johns Hopkins University

[Advanced Track of Introduction to Artificial Intelligence](#)

2011

Stanford University

Publications

- [*Critérios de Usabilidade e Acessibilidade para Avaliação de Aplicações Educacionais 3D na Web*](#) 2015
CBIE-LACLO 2015
- [*Definition of Computational Criteria to Evaluation of 3D Anatomical Web Manikins for Teaching in Health Science and for Medical Applications*](#) 2013
PAHCE – Pan American Health Care Exchanges
- [*Accessibility for Web 3D Anatomical Manikins*](#) 2013
**SETI - First Symposium on Education and Inclusive Technology
 Regional Integrated University of High Uruguay and Missions**
- [*Definition of criteria for evaluating 3D anatomical manikins on the web for education in healthcare or medical applications*](#) 2012
Extensio: Electronic Journal of Extension

Honours & Awards

- Honourable Mention in Ranking of the Best Published Articles - Conferência Latino-Americana de Objetos e Tecnologias de Aprendizagem - LACLO* 2015
- [*Gold Lion - 1º Place - Category Media, Entertainment and Leisure*](#) 2014
Cannes Lions - Campaign SAT-JF14. Record Collection Music Album - John Frusciante
- [*Silver Lion - 2º Place - Category Use of Digital Platforms*](#) 2014
Cannes Lions - Campaign SAT-JF14. Record Collection Music Album - John Frusciante
- [*Bronze Lions - 3º Place - Category New Product Launch*](#) 2014
Cannes Lions - Campaign SAT-JF14. Record Collection Music Album - John Frusciante