Curriculum Vitae - Miguel Mascarenhas Filipe

Lisbon, Portugal
E-mail: miguel.filipe+blog@gmail.com
Web: blog.mfilipe.eu | linkedin.com/in/miguelmfilipe | 13 US patents
European Union citizen

Summary

I'm a Principal Engineer with +20 years of experience. My focus spans databases, query engines, control planes and other distributed systems. I spent five years at AWS and Microsoft, including DynamoDB's launch, and more than a decade in start-ups and scale-ups of 20 to 250 people. I take substantial systems from an unclear problem through design, implementation and production. I have launched new products and platforms, re-architected critical pipelines, and delivered major improvements in performance, reliability and cost. I lead teams and multi-engineer workstreams while staying close to the code, operations and users.

I work from first principles, measure before deciding, and design simple systems that can be understood and operated end to end. I am drawn to hard problems and hold myself and the teams around me to a high standard. If something is worth doing, it is worth doing well. Problems are waiting to be solved.

Professional Experience

Principal Software Engineer April 2022 - Present

Dune, the source of truth for onchain data: 115+ chains, 1.5M+ datasets, petabyte-scale lakehouse.

- Led DuneSQL, Dune's MPP OLAP query engine, from scope and architecture through launch and evolution across multiple engineers. Later led an efficiency initiative that reduced infrastructure costs by 26% and moved roughly half of production queries to a single-node deployment.
- Led the metadata platform from the catalog service into DuneCP, an event-driven control plane governing health, maintenance and lifecycle work across roughly one million lakehouse tables per day. Designed change-driven replication between data lakes.
- Built the public Dune API and Query Execution Service from the first commit through launch, including metered, tiered execution and streaming access to large Parquet results.

Principal Software Engineer February 2019 - April 2022

Unbabel, an AI-powered human translation platform.

- Chief architect for Unbabel's translation pipeline and AI cluster; highest-ranked engineer in a 110-person R&D division, reporting to the CTO
- Designed and led the re-architecture of the distributed translation pipeline, including *Flowrunner*, its workflow engine
- Led the LanguageOS APIs and raised engineering rigour through RFCs, runbooks, shared metrics and automated end-to-end testing; served as interim SRE lead

Lead Software Engineer February 2016 - February 2019

Citymapper, an award-winning urban mobility mobile app.

- Lead programmer and chief architect for SmartRide, a bus-taxi hybrid transit service
- Led distributed systems design, development and operations; formerly led SRE

Notable achievements:

- Individual Contributor:
  - Designed, led and co-authored the SmartRide system and its APIs
  - Designed and co-authored a distributed, fault-tolerant HTTP proxy for isolation from external faults
  - Removed numerous single points of failure with a distributed request-submission framework built on DynamoDB, with service discovery, retries and backoff
  - Automated and codified the main infrastructure in CloudFormation, covering dozens of dynamically scaled, sharded microservices
- Leadership:
  - Launched SmartRide from idea to vehicles picking up passengers on London roads in 5 months, with a team of 6 engineers
  - Established SRE practices: better production metrics and visibility, with an emphasis on postmortems
  - Introduced Go as a core programming language; mentored and promoted engineers from junior to senior

Senior Software Engineer November 2014 - January 2016

Microsoft, Application & Systems Group, Aria team (large-scale data ingestion and analytics).

- Developed Aria's distributed data ingestion pipeline, with a focus on scalability and robustness

Notable achievements:

- Delivered data compression at rest and in transit in 4 months, saving more than $12 million per year
- Improved pipeline scalability by more than 2x while reducing end-to-end latency by 90%

Software Engineer - II February 2011 - October 2014

Amazon Web Services, NoSQL Database Services, DynamoDB.

- Contributed to all major components of DynamoDB as a member of the launch team

Notable achievements:

- Main author and implementer of DynamoDB's provisioned throughput
- Co-author of 13 US patents on admission control, workload and performance metering, and throughput maximisation with quality of service (QoS)
- Won the AWS Operational Excellence award for reducing latency and fleet costs through garbage collection tuning (tp99 latency reduced by 33%)

Team Lead, Software Engineer April 2008 - January 2011

SAPO, a Portuguese web portal.

- Led seven engineers working on SAPO's contextual advertising system

Notable achievements:

- Redesigned the database schema with zero downtime and worked extensively with MySQL replication

Software Engineer August 2006 - March 2008

Altitude Software, a provider of call-centre software.

- Developed C# and C++ software for a distributed interactive voice response (IVR) framework

Software Developer May 2005 - July 2006

Master's thesis project, Center for Petroleum Reservoir Modeling, Technical University of Lisbon.

- Developed parallel seismic-modelling algorithms for Linux high-performance computing (HPC) clusters using Message Passing Interface (MPI)

Unix Systems Administrator February 2003 - April 2005

Technical University of Lisbon, Computer and Software Engineering Department.

- Administered Solaris and Linux servers with about 7,000 accounts, plus 150 Linux workstations

Education

September 1999 - July 2006
Master of Engineering (MSc) in Computer and Software Engineering
Instituto Superior Técnico (https://tecnico.ulisboa.pt/en/) (Technical University of Lisbon)

Talks

Slides and recordings at blog.mfilipe.eu/papers

- ACM DEBS 2026 (keynote): DuneCP, an event-driven control plane for a blockchain data platform
- BUIDL Europe 2025: Building with Dune ECHO
- DuckCon #5 (2024): Delighting users with RESTful APIs and DuckDB
- Trino Fest 2023: DuneSQL, a query engine for blockchain data
- 2018: DynamoDB, a view under the hood
- Earlier: Linux Internals (2014), Linux File Systems (2012), How to serve 2500 requests per second (2010)

Technical Experience

- Solid experience designing, developing and operating large-scale distributed systems (10,000+ nodes)
- Professional experience in Go, Python, C/C++, Java and SQL; previously C#
- Databases, engines and table formats: DynamoDB, PostgreSQL, FoundationDB, CockroachDB, Trino, Iceberg, DuckDB and MySQL
- Storage: object-storage-backed lakehouses; tiered NVMe caches and eviction policies; durability and end-to-end integrity
- Good knowledge of Linux kernel internals (VFS, BIO, scheduler, memory manager, filesystems)
- Good knowledge of virtualisation and container technologies (Kubernetes, cgroups, KVM, Docker)
- Good knowledge of hardware architectures and topologies, IP networking and network protocols
- Observability and capacity engineering with Prometheus, Thanos and Grafana; infrastructure as code with Terraform
- Knowledge of x86 assembly

Technical Interests

- High-performance, large-scale distributed applications
- Distributed data stores and filesystems; consensus, replication and their failure modes
- Highly reliable and highly durable storage systems; deterministic simulation testing
- Full end-to-end systems operations: from applications and the software stack through the OS and hardware topology to data-centre design
- Operating-system internals, computer hardware design and architecture