Thomas Harper

Profile

Software engineer currently working in data infrastructure and customer data privacy at Spotify. I develop tools and infrastructure for ensuring customer data privacy in compliance with the EU General Data Protection Regulation for both batch and realtime applications. I have also worked on high-volume event delivery as well as scheduling and orchestration. Previously, I worked both professionally and in a research capacity in functional programming using Erlang and Haskell. Broadly speaking, I am interested in building my career as an engineer and am especially but not exclusively interested in applications of functional programming.

Education

Doctor of Philosophy in Computer Science 2014
Oriel College, University of Oxford, Oxford, United Kingdom
Dissertation: *Theory and Practice of Shortcut Fusion*
Supervised by Prof Ralf Hinze

Master of Science in Computer Science 2008
St Anne's College, University of Oxford, Oxford, United Kingdom
Dissertation: *Fusion on Haskell Unicode Strings*
Supervised by Prof Oege de Moor

Bachelor of Science *cum laude* **in Computer Science** 2007
Syracuse University, Syracuse, New York, United States

Skills

Programming Languages

- ◇ Extensive experience with Java (including Java 8), Erlang
- ◇ Extensive research experience with Haskell
- ◇ Some experience with Python, Scala
- ◇ Basic familiarity with bash scripting, C

Other Development Tools and Technologies

- ◇ Extensive experience with Google Cloud Platform, basic experience on Amazon Web Services
- ◇ Microservices: Apollo, an open-source Java framework
- ◇ Databases: Cassandra, Cloud Spanner, Cloud BigTable, PostgreSQL
- ◇ Distributed data processing: Scio (Cloud Dataflow), BigQuery, MapReduce (Hadoop)
- ◇ Message queues: RabbitMQ, Kafka, Cloud PubSub
- ◇ Configuration and deployment: Puppet, Jenkins
- ◇ Data serialisation: Avro, JSON, Protobuf

Languages
Native English, Fluent Swedish, Fluent Spanish, Conversational Icelandic, Elementary Mandarin

Experience

Staff Engineer, Data Infrastructure 2015—2019
Spotify AB Stockholm, Sweden

Spotify is a leading online music-streaming service with 180 million active users in 65 markets worldwide with mobile and desktop applications. All roles included operational responsibilities and on-call duty rotations.

Data Privacy Infrastructure

- ◇ Defined Spotify's strategy for dealing with the European Union's General Data Protection Regulation, using per-user, centrally managed encryption keys to encrypt all personal data.
- ◇ Formed and lead the team that built a key management system that provides realtime key derivation and issuing in Java, using Cassandra to provide global replication to three different sites on three different continents.
- ◇ Wrote clients and encryption/decryption libraries for use in realtime microservices as well as batch processing on Cloud Dataflow, using either Java (Apache Beam) or Scala (Scio, an open-source implementation of Beam written and maintained at Spotify).
- ◇ Automated the encryption and decryption of data on analysis platforms to hide such complexity from internal customers.
- ◇ Continuously worked to optimise these solutions to decrease overhead and reduce the associated costs incurred from cloud resource usage.

- Evaluated technologies from Google to replace non-cloud products currently in use, notably Cloud Spanner.

Event Delivery

- Migrated from a cross-site, Kafka-based event delivery system to a Google Cloud-based system using Cloud PubSub that processes over 2 million messages per second and is responsible for collection, deduplication, and batching of events.
- Automated the setup and teardown of all infrastructure associated with delivering a given event type, allowing other developers and analysts at Spotify to declare new event types in YAML files and automatically create all necessary infrastructure within 1 hour.
- Regularly monitored and improved scalability metrics as demand grew with respect to number of event types and total event volume.

Scheduling and Orchestration

- Developed and maintained Luigi, an open-source job orchestration framework in Python.
- Participated in the initial stages of the data infrastructure migration to Google Cloud Platform, including meetings with teams at Google to set requirements and strategy.
- Helped build the next-generation scheduling platform at Spotify in Java using Kubernetes to run user-defined Docker images and data jobs executed on Hadoop and Cloud Dataflow.
- Lead the project to make data infrastructure and royalty reporting compliant with IT General Controls as part of Spotify's Sarbanes-Oxley compliance.

Deputy Development Manager 2014—2015
Klarna AB Stockholm, Sweden

Klarna is an e-commerce company active in Europe and the US that provides credit-based payment options to customers on participating online stores with no sign-up using realtime credit and fraud decisions.

- Line manager for the customer identification team.
- Responsible for the delivery and development of a five-person team.

Developer 2012—2014
Klarna AB Stockholm, Sweden

Identification Team

- Broke out customer identification functionality from an Erlang monolith into a separate Erlang service with cross-site replication.
- Added support for new data models as required for new market launches
- Migrated services from on-premise systems to AWS as part of a company-wide shift to the cloud.

Risk Team

- Developed and maintained realtime risk assessment systems used by risk analysts to make credit and fraud decisions for online shoppers in Erlang and Java.
- Migrated from a risk decision engine inside an Erlang monolith to a Java-based third-party product in a zero-downtime, phased migration.
- Automated deployment tasks required for regulatory compliance, allowing risk analysts to deploy automatically several times daily instead of once weekly.

Finance Team

- Wrote integrations with financial institutions in various markets to support local payment methods in Erlang.
- Refactored large parts of the legacy payment batch processing system to allow for parallel processing, providing significant speedups of nightly jobs.

Doctoral Thesis 2008—2014
Department of Computer Science, University of Oxford Oxford, United Kingdom

- Conducted research into using fusion of structural recursion as an automated, optimising program transformation in Haskell.
- Designed and implemented purely functional data structures and adapted fusion techniques for them.
- Modelled fusion transformations in mathematical settings to prove correctness.
- Analysed intermediate compiler outputs to investigate low-level implications of transformations on performance.
- Implemented a compiler feature that allowed for semi-automated fusion user-defined recursive data structures in the Utrecht Haskell Compiler.

Master's Dissertation 2007—2008
Department of Computer Science, University of Oxford Oxford, United Kingdom

- Designed and implemented a high-performance, Unicode-compliant string library in Haskell by applying stream fusion to byte array structures.
- Examined compiler core dumps to find and optimise low-level inefficiencies.
- Created a functional, high-level interface on top of hand-optimised code that exploited low-level memory management to achieve significant performance improvements over Haskell's built-in string type.
- Presented the results in a dissertation that earned a Distinction-level mark.
- This appeared as the initial version of the `Data.Text` library, still maintained today.

Summer of Code Participant June 2007—August 2007
Google Summer of Code

- Optimised Mlucas, a Lucas-Lehmer primality tester under the supervision of a mentor from Sun Microsystems, Inc. using OpenMP to create the (then) most scalable primality tester for the Great Internet Mersenne Prime Search on Sun servers.

Research Aid June 2006—August 2006
United States Air Force Research Lab/Rome Research Site Rome, New York, United States

- Conducted machine translation research for the Information Exploitation division on building statistical models for limited domain translations using small training sets
- Built a proof-of-concept in C++ using the QT framework and open-source machine translation tools.

Information Systems Intern June 2005—September 2005
Anheuser-Busch, Inc. Baldwinsville, New York, USA

- Supported a 24/7/365 brewery as a member of the Information Systems department, supporting end-users and live process control systems in a Microsoft environment.