AHMED M. ALY

1600 Amphitheatre Pkwy, Mountain View, CA, 94043
aaly@google.com ◇ (+1) 765-413-5681

EDUCATION

Ph.D. Computer Science (2015)
Purdue University, West Lafayette, IN, USA

M.Sc. Computer Science (2009)
Alexandria University, Alexandria, Egypt

B.Sc. Computer Science (2006)
Alexandria University, Alexandria, Egypt

EMPLOYMENT

- **Google Inc.** - Mountain View, CA, USA
  Software Engineer, September 2015 - present
  Software Engineering Intern, Summer 2014
- **Purdue University** - West Lafyatte, IN, USA
  Research Assistant, 2009 - 2015
  Teaching Assistant, Spring 2015
- **Turn Inc.** - Redwood City, CA, USA
  Software Engineering Intern, Summer 2013
- **Microsoft Research** - Redmond, WA, USA
  Research Intern, Summer 2012
- **Facebook Inc.** - Palo Alto, CA, USA
  Software Engineering Intern, Summer 2010

PUBLICATION RECORD

- **Ahmed M. Aly**, Hazem Elmeleegy, Yan Qi, Walid G. Aref: *Kangaroo: Workload-Aware Processing of Range Data and Range Queries in Hadoop.* **WSDM 2016** *(Full Paper).*
- Mohamed S. Hassan, Walid G. Aref, **Ahmed M. Aly**: *Graph Indexing for Shortest-Path Finding over Dynamic Sub-Graphs.* **SIGMOD 2016** *(Full Paper).*
- Ahmed S. Abdelhamid, Mingjie Tang, **Ahmed M. Aly**, Ahmed R. Mahmood, Thamir Qadah, Walid G. Aref, Saleh Basalamah: *Cruncher: Distributed In-Memory Processing for Location-Based Services.* **ICDE 2016** (*Demo*).
- Amr Magdy, **Ahmed M. Aly**, Mohamed F. Mokbel, Sameh Elnekity, Yuxiong He, Suman Nath, Walid G. Aref: GeoTrend: Spatial Trending Queries on Realtime Microblogs. **SIGSPATIAL 2016** *(Full Paper).*
- Ahmed R. Mahmood, Walid G. Aref, **Ahmed M. Aly**, Mingjie Tang: Atlas: On the Expression of Spatial-Keyword Group Queries Using Extended Relational Constructs . **SIGSPATIAL 2016** *(Full Paper).*
- **Ahmed M. Aly**, Ahmed R. Mahmood, Mohamed S. Hassan, Walid G. Aref, Mourad Ouzzani, Hazem Elmeleegy, Thamir Qadah: *AQWA: Adaptive Query-Workload-Aware Partitioning of Big Spatial Data.* **VLDB 2015** (*Full Paper*).

- **Ahmed M. Aly**, Ahmed S. Abdelhamid, Ahmed R. Mahmood, Mohamed S. Hassan, Walid G. Aref, Hazem Elmeleegy, Mourad Ouzzani: *A Demonstration of AQWA: Adaptive Query-Workload-Aware Partitioning of Big Spatial Data.* **VLDB 2015** (*Demo*).
- Ahmed R. Mahmood, **Ahmed M. Aly**, Thamir Qadah, Amgad Madkour, Ahmed S. Abdelhamid, Mohamed S. Hassan, Walid G. Aref, Saleh Basalamah: *Tornado: A Distributed Spatio-Textual Stream Processing System.* **VLDB 2015** (*Demo*).
- **Ahmed M. Aly**, Walid G. Aref, Mourad Ouzzani: *Cost Estimation of Spatial k-Nearest-Neighbor Operators.* **EDBT 2015** (*Full Paper*).
- **Ahmed M. Aly**, Walid G. Aref, Mourad Ouzzani: *Spatial Queries with kNN and Relational Predicates.* **SIGSPATIAL 2015** (*Full Paper*).
- Ahmed R. Mahmood, Walid G. Aref, **Ahmed M. Aly**, Saleh Basalamah: *Indexing Recent Trajectories of Moving Objects.* **SIGSPATIAL 2014** (*Poster Paper*).
- **Ahmed M. Aly**, Walid G. Aref, Mourad Ouzzani, Hosam M. Mahmoud: *JISC: Adaptive Stream Processing Using Just-In-Time State Completion.* **EDBT 2014** (*Full Paper*).
- Amr Magdy, **Ahmed M. Aly**, Mohamed Mokbel, Sameh Elnekity, Suman Nath, Yuxiong He: *Mars: Real-time Spatio-temporal Queries on Microblogs.* **ICDE 2014** (*Demo*).
- **Ahmed M. Aly**, Walid G. Aref, Mourad Ouzzani: *Spatial Queries with Two kNN Predicates.* **VLDB 2012** (*Full Paper*).
- **Ahmed M. Aly**, Asmaa Sallam, Bala M. Gnanasekaran, Long-Van Nguyen-Dinh, Walid G. Aref, Mourad Ouzzani, Arif Ghafoor: *M^3: Stream Processing on Main-Memory MapReduce.* **ICDE 2012** (*Demo*).
- Yasin N. Silva, **Ahmed M. Aly**, Walid G. Aref, Per-Ake Larson: *SimDB: A Similarity-Aware Database System.* **SIGMOD 2010** (*Demo*).

INDUSTRIAL EXPERIENCE

Google Inc. - *Mountain View, CA, USA*

- Backend Engineer - Google Analytics (current role)
  Working on two distributed systems: 1) Realtime Analytics, a system that brings down the freshness of web analytics and reports to seconds, and 2) Data Studio 360, a system that enables the users to interactively create and share reports over multiple diverse data sources that can be SQL-based, e.g., MySQL, Oracle, or non-SQL-based, e.g., spreedsheets.
- Backend Engineer - Google Display Ads Infrastructure
  worked on several components of Google's infrastructure for Display Ads. My role varied between adding new features, doing integration between different components, testing, and performance optimization.
- Intern - Google Search
  Worked on the research project of link-based recommendation in the Web. The goal of that project was to recommend for each web-host (e.g., www.cnn.com) a set of hyperlinks based on the similarity of the hyperlinks between the different web-hosts. The motivation is that both the existing and recommended link structures of the web-hosts provide a rich set of information that can enhance the quality of the web search. I designed and implemented **SHRec**, a scalable MapReduce-based solution that determines link recommendations for every webpage on the Internet. SHRec operates on all the web documents crawled by Google (i.e., billions of webpages), and is an order of magnitude faster than the state-of-the-art alternatives (e.g., locality sensitive hashing).

Turn Inc. - *Redwood City, CA, USA*

Designed and implemented **Kangaroo**, a workload-aware system for partitioning and processing temporal range data on Hadoop. Kangaroo has been accepted for publication as a full paper at the 9th ACM International Conference of Web Search and Data Mining (WSDM 2016). Moreover, Kangaroo has been patented by Turn Inc.:
Ahmed M. Aly, Hazem Elmeleegy, Yan Qi: *Horizontal Interval-Based Data Partitioning and Indexing for Large Clusters*. **United States Patent 8903803**; Turn Inc. December 2014.

Facebook Inc. - *Palo Alto, CA, USA*

Contributed to the open source project: 'Hive' by adding functionality to support table/partition-level statistics for Hive tables. More details about this feature can be found at the wiki-page: `https://cwiki.apache.org/confluence/display/Hive/StatsDev`.

RESEARCH EXPERIENCE

Purdue University - *West Lafayette, IN, USA*
Research Assistant: 2009 - 2015
PhD Dissertation: *'Towards Efficient Processing of Big Spatial Data'*

Contributed to the following research projects:

- **AQWA:**
  Designed and implemented AQWA, a system that is built on top of Hadoop to support workload-aware processing of spatial queries over big spatial data. Unlike other existing systems for processing big spatial data which apply static data-partitioning, AQWA has the ability to acquire new batches of data, and update the partitioning of the data according to the workload in an incremental manner. AQWA has been published in VLDB 2015 as a full paper and as a demo.
- **Optimization of k-Nearest-Neighbor queries:**
  Conducted three key research efforts towards realizing a spatial query optimizer that can process spatial k-nearest-neighbor queries: 1) processing queries with more than one k-nearest-neighbor predicate (published in VLDB 2012), 2) estimating the cost of the k-nearest-neighbor operators (published in EDBT 2015), and 3) processing queries with relational and k-nearest-neighbor predicates (published in SIGSPATIAL 2015).
- **JISC:**
  Designed and implemented JISC, a technique that can efficiently migrate the execution of a continuous query from one query-evaluation-plan to another. JISC is an online technique that seemlessly performs the migration without interrupting the input or output data streams. JISC has been published in EDBT 2014 as a full paper.
- **EDP:**
  Contributed to EDP, a system for graph indexing and shortest-path finding. EDP can answer queries like: find the shortest-path to a certain destination while avoiding highways. EDP efficiently updates its index structures according to graph updates, e.g., adding/dropping roads. In addition to road networks, EDP extends to biological and social networks, while maintaining superior performance compared to the state-of-the art systems. EDP has been accepted for publication in SIGMOD 2016 as a full paper.
- **Tornado:**
  Contributed to Tornado, a distributed system that is built on top of Storm to process spatio-textual queries on microblog data streams, e.g., Twitter. Tornado has been published in VLDB 2015 as a demo.

- **Cruncher:**
  Contributed to Cruncher, a system that is built on top of Spark to support distributed processing of spatial queries in main-memory. Cruncher has been accepted for publication in ICDE 2016 as a demo.
- **M^3:**
  Contributed to M^3, a system that supports main-memory distributed execution of continuous queries over data streams. M^3 has been published in ICDE 2012 as a demo.
- **SimDB:**
  Contributed to SimDB, a similarity-aware DBMS that supports various similarity-aware operations (e.g., similarity group-by and similarity join). SimDB extends PostgreSQL, the open-source DBMS, and has been published in SIGMOD 2010 as a demo.

Microsoft Research - *Redmond, WA, USA*
Research Intern: Summer 2012

Worked on the research problem of finding the frequent/trending items in a spatio-temporal data stream. Designed and implemented **GeoTrend**, a system that can find the top-k frequent (or trending) keywords at any given spatial region. Having access to the Twitter firehose, Geotrend has been deployed to find the top-k frequent hashtags at any given region. In collaboration with Bing Geospatial team, Geotrend has been integrated within Bing search engine. Geotrend has been demonstrated in ICDE 2014 as part of a bigger system, Mars, that can answer a large variety of spatial queries on microblog data streams.

SERVICE

- Serving as a publicity chair for ACM SIGSPATIAL 2016, 2017.
- Served as a reviewer for the International Conference on Mobile Data Management (MDM, the VLDB Journal, the ACM Transactions on Database Systems (TODS), the ACM Transactions on Spatial Algorithms and Systems (TSAS), and the Information Systems Journal.
- Served as an external reviewer for TKDE, VLDB, ICDE, SIGMOD, EDBT, SSTD, and CIKM.