Hosting, collaboration, query, analytics, and visualization for public and private datasets.

Anant Bhardwaj | Aaron Elmore | Justin Anderson | Albert Carter

datahub.mit.edu

At the LivingLab, we transform research into demo-and-development ready products for our data ecosystem.

**DataHub**

**Architecture**

**Apps**

- **DataHub Web**
  - DataQ: A visual SQL query builder (H. Subramanyam)
  - DataTables: Manipulate, sort, and filter SQL tables like MS Excel (H. Subramanyam)
  - dbWipes: Sophisticated algorithms for quickly visualizing data and finding outliers (E. Wu)
  - Notebook: Post-query processing of data using Python, R, or other languages (A. Bhardwaj)
  - Refiner: Convert poorly structured text into structured data (H. Subramanyam & A. Bhardwaj)
  - Sentiment: Sentiment analysis of text data (H. Subramanyam)
  - Vis2: Data Visualization (A. Bhardwaj)

- **3rd Party**
  - Confer: Personal Schedule Builder for Conferences (A. Bhardwaj)
  - GetFit: Fitness Tracking with aggregate statistics and personal data stores (A. Carter)
  - Notebook: Post-query processing of data using Python, R, or other languages (A. Bhardwaj)

- **DataHub Core**
  - Roles + Access Control: Grant users privileges on to schemas, tables, or rows
  - Machine Learning/Analytics: Refine semi-structured data into structured data
  - Versioning + Provenance: Git-model branching, versioning, and provenance tracking of data
  - File Handler: Ingest and store files of up to 2 gb

**DBMS Cluster**

- Postgres Instances

**Used By**

- **Confer** (a 3rd party web app) used DataHub to help users build personal schedules for conferences.

- **GetFit** (iOS) used DataHub to help MIT community members track their personal fitness levels.

- DataHub's allowed end users to own and directly access their data; while simultaneously allowing GetFit to perform aggregate computations.

- The NY State Moorland Commission uses DataHub's collaboration model to investigate public corruption.

- DataHub served as an out-of-the-box database for Confer's browser application.