INSIGHT

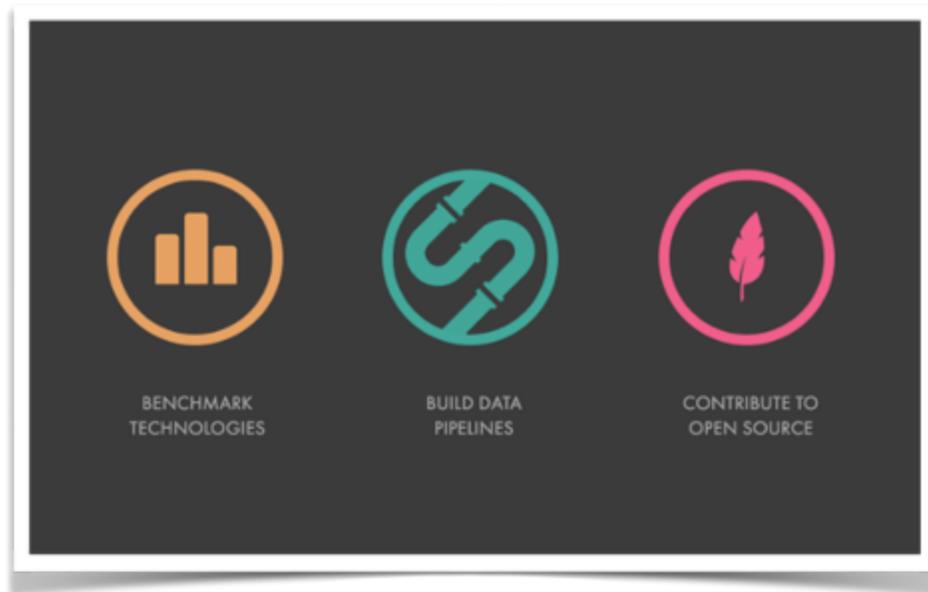## DATA ENGINEERING FELLOWS PROGRAM

Insight Data Engineering Fellows Program is an intensive, seven week professional training fellowship. The program enables software engineers and academic programmers to gain experience with the industry-specific tools needed to work in the growing field of data engineering at leading companies in Silicon Valley, New York, Boston, Seattle, and Los Angeles.

Your Bridge to a Career in Data Engineering

The Insight Data Engineering Fellows Program is a professional training fellowship that bridges the experience gap between computer science fundamentals and data engineering. Insight is a seven week, intensive program that enables experienced engineers and programmers to apply their existing software development skills to transition to a career in data engineering. Fellows learn by building an efficient and scalable data platform in a self-directed, collaborative, and hands-on approach.

Fellows are provided guidance and resources from industry and open source leaders in Silicon Valley, New York, Boston, Seattle, and Los Angeles. Additionally, Fellows gain access to a network of Insight alumni. Immediately following the program, Fellows interview with data engineering teams at leading companies, and receive continued guidance and support until they receive a full-time offer as a data engineer.

Insight Data Engineering Fellowship in a Nutshell:

1. 7 week, full-time, professional data engineering training in San Francisco, New York City, Boston, Seattle, or Los Angeles.

2. Self-directed, hands-on learning using cloud computing resources under the guidance of industry-leading data engineers.

3. Alumni network of Insight Fellows who are now working as data engineers and data scientists in Silicon Valley, New York, Boston, Seattle, and Los Angeles.

4. Meet dozens of top companies, present your work to teams you're interested in, and interview immediately following the program.

5. Need-based scholarships to help cover living expenses.

What is Data Engineering?

For companies of all sizes and sectors, data is among the most important tools for understanding the world. Whether collecting user activity on popular mobile apps, tracking customer engagement at established media outlets, or processing genome sequences at bioinformatic firms, data is essential for making an impact. As the amount of data has increased in size and complexity, companies have formed dedicated teams to leverage this wealth of information.

At the core of these data teams is the critical task of efficiently collecting, processing, and accessing data. To solve this foundational problem, a specialized role has emerged - the data engineer. Data engineers build the tools and platforms to answer questions with data, using software engineering best practices, computer science fundamentals, core database principles, and recent advances in distributed systems.

> *"Data engineers are specialized software engineers that enable others to answer questions on datasets, within latency constraints"* [1]
> -Nathan Marz, Inventor of Apache Storm and the Lambda Architecture, Author of *Big Data*, and Insight Advisor

In addition to building the infrastructure and "pipelines" for handling data, the best data engineers use generalization and abstraction to improve the productivity of their data team. This involves anticipating needs and removing bottlenecks from their systems. In some cases, data engineers may even develop new technologies that are tailored to a specific use case. For example, after understanding the fundamental design decisions behind relational databases, data engineers have developed many specialized "NoSQL" databases that sacrifice complex queries in favor of optimized performance at specific tasks.

Challenges in Data Engineering

Data engineers work on some of the most interesting challenges in software engineering and computer science:
- Processing and storing data consistently and efficiently across a distributed cluster of dozens (or hundreds) of servers.
- Ingesting complex, semi-structured data from a variety of real-world sources that dynamically evolve.
- Evaluating the technical trade-offs of tools to build simple, yet robust, architectures that survive hardware failures.
- Monitoring systems in real-time to automate decisions within milliseconds of incoming events.

---

[1] Paraphrased from "Demystifying Data Engineering," July 2015

- Working alongside data scientists to productionalize advanced computer science algorithms like machine learning and natural language processing.

The Demand for Data Engineers

Since data engineering is at the foundation of nearly all data teams, the number of positions requiring experience with data engineering technologies like Hadoop, Spark, and Kafka has steadily increased. This trend, which has risen over the past several years, will continue as more companies realize the full potential of data and develop their own data teams.
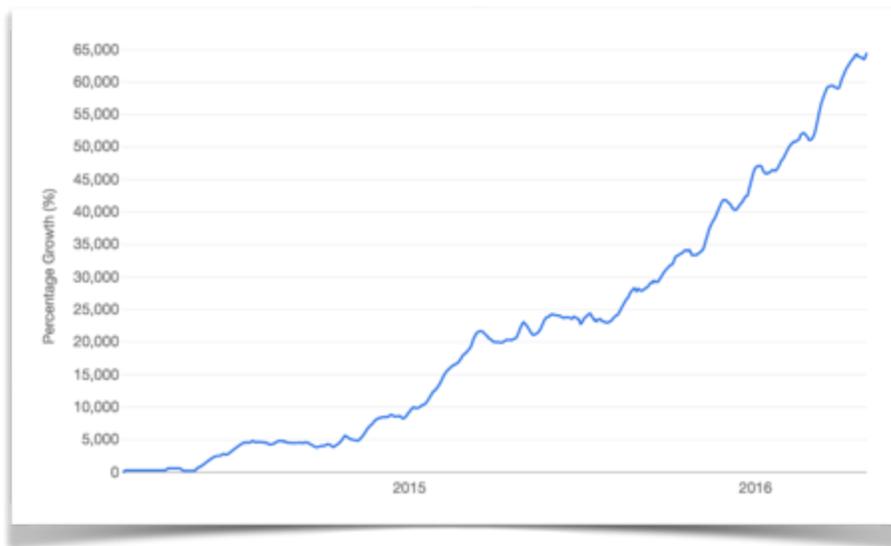
Job Trends on indeed.com for Hadoop, Spark, and Kafka

However, the demand for skilled data engineers far exceeds the current supply. One reason for this gap is the advanced software engineering and computer science skills required for these roles. These are fundamental notions and key concepts like data locality, consistency, caching, hash tables, and tree-based indexes.

> *"Too often people look to data scientists and ignore the fact that, in order to be successful in data science, you need to have an effective data platform. We actually see the biggest skill gap is in high quality data engineers who can build these new data applications and organize the data."* [2]
> - Ron Bodkin, CEO of Think Big Analytics

---

[2] Datanami, "Rise of the Big Data Engineer," September 2014

However, there is another crucial issue: building sophisticated data platforms often requires experience with the relevant technologies. There are an overwhelming number of new tools in the data engineering ecosystem, and practically no one has direct experience with all of them. Since these are specialized technologies, tailored for select use cases, even experienced engineers seldom get the chance to experiment with them. **Many companies are hesitant to hire candidates without direct data engineering experience, but many candidates with the necessary fundamentals cannot get that experience in their current roles.**

This is the problem that Insight set out to solve.

About Insight

In 2012, Insight developed a new model for education: we bring together hardworking and ambitious candidates who have the fundamental skills, and enable them to make a transition into an advanced field by gaining hands-on experience with the tools used in industry.

Learning by Building

Since our Fellows enter the program with advanced software engineering skills and already know the core concepts of computer science, there are no classes or lectures on these foundations. Instead, the main focus of the program is to learn by building a data platform using the same technologies and best practices as leaders in the data engineering community.

> *"The essential thing here is to base the decision on as much real programming as you can. Even a mere week working with a technology can tell you things you'd never learn from a hundred vendor presentations."* [3]
> - Pramod J. Sadalage and Martin Fowler, Principal Consultant and Chief Scientist at ThoughtWorks

In addition to providing experience with specific tools, the platform serves as a portfolio piece to demonstrate that Fellows can successfully transition into data engineering roles. Rather than just listing skills on a resume, Fellows can provide a concrete example of their experience that reduces the risk of hiring for data teams.

For more detailed examples of platforms from previous Fellows, visit our Insight blog.
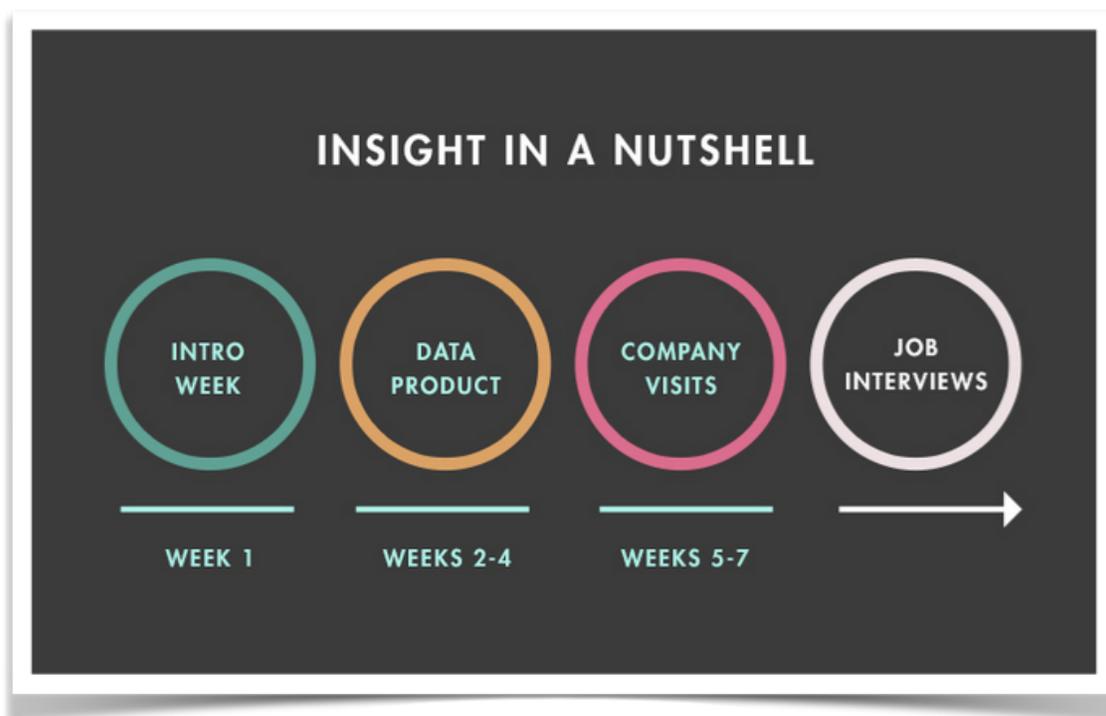
Building a Data Platform

Depending on their interests and company preferences, each Fellow builds a platform that addresses one of these industry tasks:

---

[3] "NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence," 2013

- **Adding Features**: A production-quality contribution that adds functionality to a single technology, such as an open source project. This demonstrates that Fellows can deeply understand and improve an existing system, which companies with mature infrastructure appreciate.
- **Benchmark Technologies**: A well-conceived benchmark of two or three technologies, resulting in recommendations for a given use case. Thoughtful comparisons show a thorough understanding of each tool and the technical trade-offs they make, which is appealing to companies that are evaluating or expanding their current tech stack.
- **Building Pipelines**: An end-to-end pipeline that ingests and processes data so that it can be queried efficiently, and possibly visualized using dashboards. This requires a general understanding of 4-6 technologies, but still examines a few challenges in-depth. This approach emulates building an entire infrastructure, similar to a start-up or "greenfield" project.

Weekly Breakdown of the Program

Insight is intended for advanced programmers who are motivated to quickly build on their engineering fundamentals. Fellows gain meaningful experience with data engineering technologies and design patterns in just a few weeks. This high rate of learning mirrors the iterative nature of development in the tech industry, and is appreciated by the partnering companies.

- **Week 1: Intro** - Learn the data engineering ecosystem and begin building your data platform.
- **Week 2: MVP** - Build the minimum viable platform with the core functionality.
- **Week 3: Scaling** - Scale your platform to handle more data and edge cases.
- **Week 4: Perfect and Practice** - Finish your platform and practice presenting it.
- **Weeks 5-7: Present and Prep** - Present your platform to the companies you are interested in, and begin preparing for the interviews that will follow.
- **Week 8+: Interview** - Interview at the companies that call you back from your presentations.

Collaboration and Mentorship

Though each Fellow builds their own data platform based off of their individual interests and abilities, the strength of Insight is rooted in a collaborative environment. Fellows accelerate their learning by working together to solve common problems and leveraging the diverse backgrounds of one another.

Mentors at Insight

- **Your fellow Fellows** - Ambitious engineers with common goals and a diverse set of skills that complement yours. Some have years of experience with databases and writing production quality code, while others studied sophisticated machine learning topics like natural language processing and recommendation. Learn the way you do in industry, by collaborating and working through challenges with your peers.
- **Network of Alumni** - Previous Fellows who have been through this transition, and are now at top companies in the field. They provide individualized guidance and practice with interviews.
- **Program Managers & Directors** - Insight staff who offer continued guidance throughout the entire process. They point you to the right resources to help troubleshoot tough issues.
- **Mentor Companies** - Over 20 data teams at leading companies of various sectors and sizes that will share the problems they're currently solving. This helps you better understand the current challenges in the industry and decide which teams you want to join.
- **Industry and Open Source Leaders** - Pioneers at the forefront of Data Engineering that help you learn the best practices and newest trends in the industry.
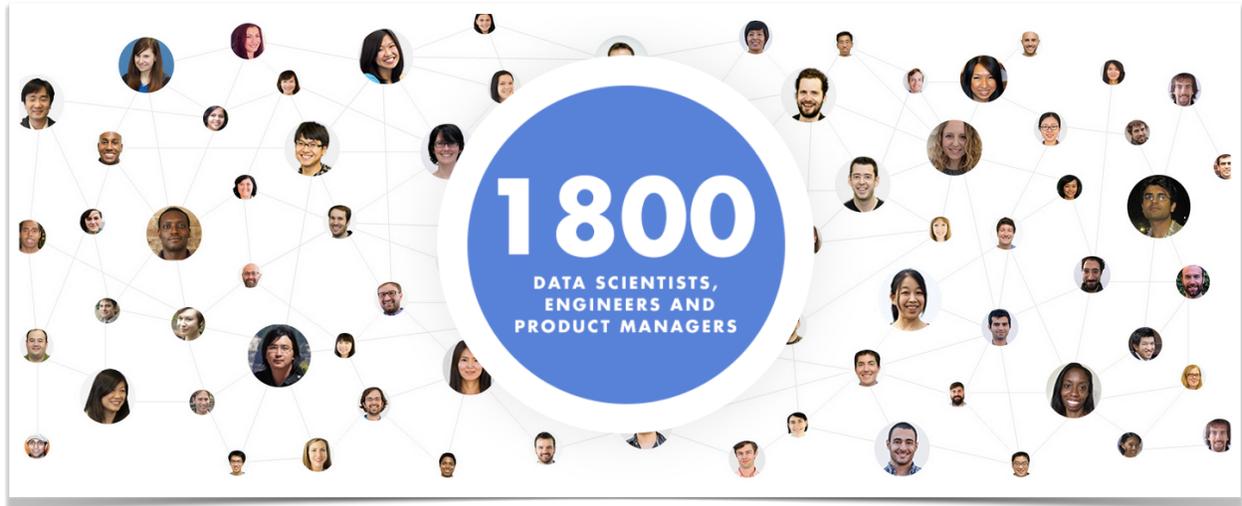
Benefits

As an Insight Fellow, you will be provided with the resources and connections needed to transition to a career in data engineering. Specifically, you will:

- Need-based scholarships to help cover expenses are also available.
- Get a reserved desk space at the Insight office, with full-time access to a library of relevant resources.
- Be provided with dedicated cloud computing resources to build and maintain your data platform for the duration of the program.
- Receive guidance from alumni and industry professionals throughout your data engineering platform and interview preparation.
- Be introduced to dozens of companies, present your work to teams you're interested in, and interview immediately following the program.
- Develop a professional network of Fellows who are recognized as top data engineers and data scientists in-industry.
- Receive personalized advice for negotiating and accepting offers with companies.

"The data engineering program seemed like a fantastic opportunity for any programmer to sharpen their skills and get an awesome job, whether a new grad or someone experienced looking to transition their career." [4]

- Nathan Marz

---

[4] "The Entrepreneur Who Captivated Me", Thoughts from the Red Planet

Responsibilities

In return for our commitment to you, we ask that Insight Fellows agree to:

- Actively and thoughtfully contribute to group activities and sessions during program hours.
- Take a leave of absence, if applicable, from your current responsibilities (e.g. current employment, research, or studies) in order to participate at Insight. Both the 7-week portion and the subsequent interview process require a full-time focus.
- Commit to self-directed learning and complete a challenging data engineering platform during the program while giving and receiving constructive feedback.
- Interview for full-time data engineering positions with Mentor Companies in your program's location upon completion of the program.
- Plan to keep coming into the office during the interview weeks (when not interviewing) to participate in interview prep sessions with other Fellows.
- Support future Fellows by providing mentorship and guidance once you become a data engineer in the industry.

Frequently Asked Questions

How much are data engineers compensated?
The starting salary for a data engineer typically ranges from $100,000 to $130,000 and increases significantly with professional experience, with most companies matching or exceeding previous salaries. Additionally, companies often provide full benefits, stock options, and 5-15% annual bonuses.

How long will it take to get an offer as a data engineer?
The median time for our Fellows to complete the interview process and receive an offer is 5 weeks after the program, and approximately 85% receive an offer after 3 months of interviewing. We cannot guarantee an offer since Fellows must complete their own interview process, but we do continue to support and work with 100% of Fellows to get them a great role as a data engineer.

How can I best prepare for the Insight Data Engineering Program?
Insight is designed for advanced programmers with strong computer science and software engineering fundamentals. We recommend familiarizing yourself with distributed systems, cloud computing resources, the Linux environment, functional programming, and the many open source technologies that we commonly discuss on our blog.

I'm open to multiple locations, how should I choose which location to apply for?
Fellows who participate in a particular session location will be introduced to data teams in that location (e.g. Silicon Valley Program Fellows will interview for Silicon Valley companies). Thus

we ask that you consider each location and submit your application to the location where you are most interested in working. Admissions for all program locations are equally competitive, so your choice should be based on where you prefer to live and work.

I'm interested in both the Data Science and Data Engineering Programs. How should I choose which program to apply for?
Data engineering focuses primarily on building robust and scalable platforms to answer questions with large and complicated data sets. On the other hand, data science has an emphasis on answering open-ended questions with sophisticated statistical methods and machine learning. We encourage you to investigate these fields in our blog post Data Science vs Data Engineering, and apply to the field that you prefer.

Please see our FAQ page for other frequently asked questions related to our programs.

Apply Now or Learn More

If you're interested in applying to Insight Data Engineering, please visit:
http://insightdataengineering.com/apply

You can learn more about us by visiting our blog or asking additional questions at:
info@insightdataengineering.com