US AI and Big Data Talent Assessment

May 2018
AGENDA

01 | Demand Analysis

02 | Location Deep Dive Analysis

KEY HIGHLIGHTS

- Methodology for estimating Demand for AI and BD&A
- Employable AI & Big Data talent across G500 companies, start-ups and Universities
- Unique job roles analysis across locations
- Job progression analysis
**DRAUP Analysis Track:** One of the key components of our methodology involves analysis of installed Big Data and AI workforce employed in G500* organizations, Start-ups and Service Providers.

### Demand Side: Analysis Tracks

1. **G500* Companies**
   - Skill Maturity Analysis
   - Job progression and maturity assessment
   - Workloads

2. **Start-Up**
   - Talent distribution by Geo
   - Skill characteristics

3. **Service Providers**
   - Installed Talent Pool
   - Skill Distribution by Geography

---

*G500 Companies (Top 500 R&D spenders)*

Source: DRAUP Methodology
**DRAUP Methodology:** The current research papers in global AI and Big Data talent trends do not provide enough transparency in the methodology. Draup attempts to change this trend.

---

**Step 1: Skill Analysis**
- List of skills associated with AI/BD obtained from primary interviews and internal Zinnov research.
- All AI/BD employees and their companies globally obtained by mining job platforms.

**Step 2: Company Analysis**
- All companies extracted are tagged with size, location, and industry vertical.
- Coverage ratios estimated for size groups and industry verticals based on primary interviews.

**Step 3: Estimation of Talent Pool**
- Installed Talent Pool = Sum of talent installed in all companies.

---

[1] Job platforms include LinkedIn, Naukri, Monster and Indeed
[2] Location includes both the city and the corresponding hotbeds across globe
[3] There are 3 size groups: Small (1-200 employees), Mid (201-1,000 employees) and Large (>1,000 employees)
[4] List of industry verticals include Automotive, Enterprise Software and Consumer Internet, BFSI and Retail
**Mining Open Positions for Companies Globally**

**Demand Growth Rate Estimation**

**Projecting Future Demand**

**DRAUP Methodology:** The current research papers in global AI and Big Data talent trends do not provide enough transparency in the methodology. Draup attempts to change this trend.

**TOTAL TALENT DEMAND** = **[A] INSTALLED TALENT POOL** + **[B] OPEN POSITIONS ANALYSIS**

**[A] INSTALLED TALENT POOL**

- List of skills associated with AI/DS obtained from primary interviews and internal Zinnov research

**[B] OPEN POSITIONS ANALYSIS**

- Global AI/DS open positions in the past 3 years obtained by mining job platforms
- Open positions for all companies are tagged with size group, location and industry vertical
- Quarterly growth for open positions captured per size group, location and vertical (Parameters 1, 2, 3)
- Parameters 1, 2, 3 are validated and benchmarked with primary interviews
- Estimating Future growth in Job openings and estimating the Demand in AI/DS technology area

---

{1} Job platforms include LinkedIn, Naukri, Monster and Indeed
{2} There are 3 size groups: Small (1-200 employees), Mid (201-1,000 employees) and Large (>1,000 employees)
{3} Location includes both the city and the corresponding country
{4} List of industry verticals include ITeS, ER&D and SPD
Unique job roles: Definitions for AI & Big Data / Analytics roles

Roles have been segregated in categories: AI, Big Data/Analytics and Auxiliary roles

7 Big Data / Analytics roles
- Analyst – Data Quality Management
- Analyst – Business Intelligence
- Data Warehouse Engineer
- Big Data / Hadoop Administrator
- Data Architect
- Data Steward
- Visualization Specialist

3 Artificial Intelligence roles
- Data Scientist
- Applied Data Scientist – Vision
- Applied Data Scientist – Speech

7 Auxiliary roles
- Solutions Architect
- Infrastructure Engineer
- Product Manager
- Software Engineer – Autonomous Systems
- Systems Integration Engineer – Autonomous Systems
- Test Engineer
- Information Security - Analyst

[AI Research Scientist merged with Data Scientist]
[DevOps Engineer Redefined]
[Hardware Engineer Redefined]

Roles analysed
Not analysed in current study

Skill Analysis Methodology
Unique job roles: Methodology to determine unique job roles

**Phase-1**
- Harvesting relevant job descriptions
- Extracting job descriptions
- Filtering job descriptions
- Creating title library

**Phase-2**
- Extracting technical & conceptual skills
- Using existing skill library to fine-tune spaCy model
- Using spaCy to locate relevant section(s) within JD

**Phase-3**
- Clustering extracted skills to identify unique roles
- Estimating the semantic similarities between each skill using a word2vec model
- Developing clusters of skills commonly found together using k-means clustering

**Phase-4**
- Primary validation
- Identifying unique job roles
- Outcomes
- Outreach
- Conducting primary interviews to validate findings

**QC and validation checkpoints**
- Manual QC on output
- Manual QC on output
- Primary validation
- Primary validation

**Primary validation**

**Note:** DRAUP Methodology
**Unique job roles:** Developing a JD corpus for analysis

<table>
<thead>
<tr>
<th>Step No.</th>
<th>Harvesting Funnel</th>
<th>Steps Involved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step-1: JD Extraction</td>
<td>28 million</td>
<td>• Total number of job descriptions extracted (Sources: Indeed, LinkedIn &amp; Glassdoor)</td>
</tr>
<tr>
<td>Step-2: Language Filtering</td>
<td>15 million</td>
<td>• Job descriptions filtered on the basis of language • Only JDs typed in English were filtered out from the total set</td>
</tr>
<tr>
<td>Step-3: Title Library Creation</td>
<td>TITLE LIBRARY$^2$</td>
<td>Titles unique to data science + Titles similar to data science - Titles that may create noise</td>
</tr>
<tr>
<td></td>
<td>Manual QC to test exhaustiveness of title library</td>
<td>Core Data Science Titles: (Data Scientist, Data Engineer, Data Architect, Applied Scientist, ADAS Engineer, etc.) Adjacent / Intersecting Titles: (Business Analyst, Data Analyst, Financial Analyst, Risk Analyst, Marketing Analyst, etc.) Negative Titles: (Data Entry Operator, Data Center Technician, EHRS, Clinical Scientist, etc.)</td>
</tr>
<tr>
<td>Step-4: JD Corpus Creation</td>
<td>~600,000</td>
<td>• Relevant job descriptions filtered using a random forest algorithm that classified job descriptions on the basis of the title corpus</td>
</tr>
</tbody>
</table>

---

2 Title library developed using data extracted from LinkedIn, Naukri, Indeed in addition to NASSCOM data
Unique job roles: 18 unique roles were discovered in AI and Big Data/Analytics

Data and Analytics Process

- **Capture**
  - Organize: Identify syntax and semantics
  - Collect: Identify data sources
  - Prepare: Enable access to data sources

- **Integrate**
  - Integrate: Aggregate data into data lakes

- **Leverage**
  - Analyse: Develop models to analyse data
  - Visualize: Create visualizations
  - Decide: Make decisions using data

- **Deploy**
  - Develop: Develop data-driven apps

**Breakdown of Roles by Process**

- **Chief Data Officer**
  - Analyst – Data Mgmt.
  - Hadoop Administrator
  - Data Warehouse Engineer
  - Analyst – BI
  - Visualization Specialist
  - Data Scientist
  - Applied Data Scientist – Speech
  - Applied Data Scientist – Vision
  - Data Architect

- **Auxiliary Roles**
  - Solution Architect
  - Software Engineer – Autonomous Systems
  - Autonomous Systems Integration Engineer
  - Infrastructure Engineer
  - Test Engineer
  - Product Manager
  - Analyst – Information Security

* Business intelligence analysts can be part of varied functions, e.g. ops, finance, marketing, or project planning
** Auxiliary roles include those which may not form part of the data science function, but are essential to productize data platforms or autonomous systems

Source: 1 H. Gilbert Miller, Peter Mork, “From Data to Decisions”, IT Professional, vol 15, no.1, pp. 57-59, Jan-Feb 2013, 2 Indeed, Naukri, LinkedIn, StackExchange, GitHub
## Unique job roles: 7 Primary Skills

The unique job roles and job roles were shortlisted and analysed within Big Data & Data Science engineering.

<table>
<thead>
<tr>
<th>Unique Roles</th>
<th>Titles</th>
<th>Technical or Conceptual Skills</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Analyst - Data Management</td>
<td>Database Analyst, Data Management Analyst, Database Developer, Database Administrator</td>
<td>database architecture, relational databases, data cleaning, data manipulation, tableau, power bi, excel</td>
<td>Maintains and manages the database; Responsible for performing quality checks on datasets; Ensures correct data schema and syntax; Filters and cleans data</td>
</tr>
<tr>
<td>2. Data Architect</td>
<td>Data Architect, Tech Lead Data Platform, Tech Lead Data Modelling</td>
<td>apache, azure, distribute systems, flume, google cloud, gradle, integrations, java j2ee,</td>
<td>Designs and implementing the technical architecture; Defines and designs data systems, services and technology solutions; Implements and administers data infrastructure</td>
</tr>
<tr>
<td>3. Big Data / Hadoop Administrator</td>
<td>Hadoop Administrator, Data Administrator, Big Data DevOps Engineer, Hadoop Platform Engineer</td>
<td>hadoop, flume, YARN, mongodb, dynamodb, mapreduce, devops, hbase, hdfs, AWS</td>
<td>Supports the Hadoop infrastructure and ensures availability; Responsible for node-cluster configuration, deployment and capacity planning; Monitors and maintains clusters and tunes performance; Responsible for administering YARN and providing support for running and monitoring MapReduce jobs</td>
</tr>
<tr>
<td>3. Database Engineering</td>
<td>Data Engineer, Data Warehouse Engineer, Data Warehousing Specialist, Data Developer, Hadoop Developer, Spark Developer, Hadoop Engineer, Spark Engineer, Scala Engineer, Scala Developer</td>
<td>hbase, amazon web service, kafka, spark, cassandra, dynamodb, flume, gradle, graph, hadoop, jmeter, json</td>
<td>Creates data pipelines to move and transform data; Responsible for performing transformations to aggregate disparate data volumes into data lakes; Manages data from different sources; Provides support, maintenance, monitoring and troubleshooting for data warehouse processes</td>
</tr>
</tbody>
</table>

**Note:** DRAUP’s proprietary talent module was used to analyse jobs by job roles and skill type.
Unique job roles: 7 Primary Skills and job roles were shortlisted and analysed within Big Data & Data Science engineering.

<table>
<thead>
<tr>
<th>Unique Roles</th>
<th>Titles</th>
<th>Technical or Conceptual Skills</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Scientist</td>
<td>Data Scientist, Applied Scientist, Data Researcher, Applied Researcher,</td>
<td>Classification, clustering, decision trees, dimensionality reduction,</td>
<td>Analyses and interprets data (both structured and unstructured) and generates prescriptive and predictive insights; Responsible for generating insights from raw data using inferential and predictive models; Responsible for developing new analytical models for the organization.</td>
</tr>
<tr>
<td></td>
<td>Data Modeling Scientist, Data Mining Scientist, Data Mining Engineer,</td>
<td>logistic regression, SVM, natural language process, predictive analytics,</td>
<td></td>
</tr>
</tbody>
</table>
Precisely estimating the talent suited for AI and Big Data involves several interviews and deeper understanding of technical stacks across resumes and profiles.

### Global Demand in 2018

1. **Total Employed AI & BD/A talent pool**
   - G500 companies
   - Global Start-Ups
   - Service Providers

2. **Unmet AI & BD/A talent demand in 2018**

### Global Big Data & AI Talent pool Demand

- **~1.2 M**
- **~650,000**
  - AI and Big data/ Analytics relevant installed talent
- **~265,000**
  - Installed Talent in G500 companies
- **~270,000**
  - Installed Talent in Start-ups
- **~120,000**
  - Installed Talent in ITeS Service Providers (SPs)
- **~515,000**
  - Job openings across G500, Start-ups and SPs

*G500 Companies (Top 500 product organization by R&D spend)*

Note: DRAUP Methodology, Numbers of graduating students are estimates and will vary by year
Note: Other engineering skills include Electronics, Electrical, communication, mechanical and engineering skills employed across Semiconductors, Telecommunication, IT & computer peripheral industry
Big Data and AI talent in US, China and Israel is predominantly employed by **US based Tech Giants**; Talent in Canada and UK talent is largely employed by **start-ups**; India has high talent pool installed among **Service Providers**

### Big Data & AI Installed Talent 2018 split by company type

<table>
<thead>
<tr>
<th>Country</th>
<th>Talent</th>
<th>G500 Companies</th>
<th>Start-Ups Companies</th>
<th>Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Israel</td>
<td>14,000</td>
<td>0%</td>
<td>21%</td>
<td>79%</td>
</tr>
<tr>
<td>USA</td>
<td>295,000</td>
<td>5%</td>
<td>34%</td>
<td>61%</td>
</tr>
<tr>
<td>India</td>
<td>152,000</td>
<td>3%</td>
<td>20%</td>
<td>52%</td>
</tr>
<tr>
<td>China</td>
<td>67,000</td>
<td>9%</td>
<td>32%</td>
<td>60%</td>
</tr>
<tr>
<td>UK</td>
<td>28,000</td>
<td>5%</td>
<td>48%</td>
<td>48%</td>
</tr>
<tr>
<td>Canada</td>
<td>18,000</td>
<td>5%</td>
<td>48%</td>
<td>48%</td>
</tr>
</tbody>
</table>

### Insights

- **AI & BD talent is largely engaged by** Start-ups such as BrainQ, Iguazio etc. and **Tech giants** such as Microsoft, Samsung and Google. **Automotive OEM's** such as GM and Ford have engineering focus on autonomous and connected driving.

- US AI and Big Data talent is predominantly split between large companies and start-ups. Companies such as Google, Microsoft, Facebook IBM, Etc. have large installed talent bases for AI & Big Data in US, led by Google with nearly “doing almost all its Big Data Analytics work out of US.” Notable small and medium companies have come up such as Splunk, Cloudera, MongoDB who are becoming key infrastructure developers.

- India’s **Big Data talent is predominantly employed in large service provider companies** involved in solutions deployment and support (IBM, Infosys, TCS etc.). Amazon, Microsoft and IBM Watson have significant AI talent pool. Notable medium and small companies such as InMobi, Musigma use Big Data Analytics for Performance tracking of Mobile Ads and are key employer for analytics talent.

- China’s AI and Big Data Talent is employed predominantly in large companies, both local and MNCs, such as Baidu, Tencent, Alibaba, EBay, Amazon etc. Engineering teams with-in Deep learning start-ups such as I-CorbonX, Sensetime, Face++ have rapidly scaled during last 2 years.

- UK AI & Big Data talent is significantly available in Start-ups largely focussed on industry applications. Tech giants such as Google (Deepmind), IBM(Watson healthcare) and Microsoft (Big Data Analytics for Bing and Skype) etc. large scale in-house engineering team grown through acqui-hires.

- Canada’s AI and Big Data talent pool has a large presence of start-ups and niche mid-sized companies such as Algolux, Instaclick(webpage and ad analytics). Large companies also employ a significant portion of the AI and big data talent such as Microsoft (Healthcare AI), IBM Watson (big data banking and finance back-end solutions) and EA Sports (gaming analytics),
Installed talent across G500 organizations: Tier-1s US locations have nearly ~44% of total employed Big Data and AI talent pool largely consolidated within Tech giants.

Employed big data analytics and AI talent across global G500 companies

~250K
Installed Big Data & AI talent in G500 companies

35%
of the 250K employees are working for Tech Giants

~110K
Installed Big Data & AI talent in US

Note: DRAUP G500 skills mapping modeler – 2017, 2018
Generic AI Talent pool not considered as there is noise in the data
Installed talent across G500 organizations: Tier-2 locations have high talent scalability potential due to increased government spending to strengthen infrastructure, socio-economic factors and maturing start-up and university ecosystem.

- **EUROPE**
  - Brexit’s impact – Shift to Germany
  - Job Demands in Eastern Europe
  - Companies setting up new EU HQs

- **AFRICA**
  - Social Start-ups’ Impact
  - iHub maturity
  - Investment by China

- **SOUTH AMERICA**
  - Brazil tier-2 cities
  - Corruption impact in Brazil
  - Mexico stepping up tech impact

- **CHINA**
  - Megalopolis initiative
  - One-child policy impact
  - Migration towards west

- **USA/Canada**
  - Higher real estate cost in East
  - Rewiring of Auto Industry
  - Less regulation in South
  - Move towards south and west
  - Positive view about Canada

Source: Zinnov Global Machine Learning talent forecasting modeler
Zinnov analysis of university programs, fresh ML graduates, digital initiatives of Government and Enterprises, Startups
Generic Data Science Talent pool not considered as there is noise in the data
**Start-up Talent Overview**: Indian and Chinese AI & Big Data start-ups have attracted late stage investments from global VCs and thus have rapidly scaled their engineering teams during last 2 years.

- **Chinese Deep learning** start-ups such as iCarbonX, Face++, Sensetime have attracted large late stage investments from global VCs during last 2 years.
- **Indian** start-ups such as Netrdyne, niki.AI and FluidAI have also raised mid staged rounds and have been ramping up their engineering team size.

### Key AI & Big Data Start-ups across global locations

- **Massively scaled deep learning**
- **AI based consumer robotics start-up**
- **Data cataloguing and cleaning**
- **ML based threat detection**
- **NLP API**
- **Vision based advanced assistance system**
- **ML based recruitment solution**
- **ML for personalised healthcare**
- **ML for retail**

**Note**: DRAUP’s Talent Simulation Module analysed 10,000+ start-ups globally to identify top ML and Big Data based start-ups.
Global IT and Engineering Services giants have mature Big Data talent pool and have been upskilling employes to develop AI capability; Indian players hold up ~58% of total talent

Geo-Wise Talent Distribution across skills

<table>
<thead>
<tr>
<th>Location</th>
<th>Installed Big Data and ML/AI talent</th>
<th>G500 Companies</th>
<th>Start-ups</th>
<th>Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>India</td>
<td>~110K</td>
<td>~100K</td>
<td>~58K</td>
<td>~50K</td>
</tr>
<tr>
<td>US &amp; Canada</td>
<td>~11K</td>
<td></td>
<td>~11K</td>
<td>~8K</td>
</tr>
<tr>
<td>China</td>
<td>~2.5K</td>
<td></td>
<td>~2.5K</td>
<td>~9K</td>
</tr>
<tr>
<td>Europe</td>
<td>~23K</td>
<td></td>
<td>~23K</td>
<td>~3K</td>
</tr>
</tbody>
</table>

- India accounts for more than 50% of the available talent in service providers.
- Service providers like TCS are setting up CoE in collaboration with Intel to speed up adoption of AI.
- TCS and Infosys have developed their proprietary AI platforms to serve global customers.
- Infosys provides mandatory training on Artificial Intelligence to new joiners.
- Tech Mahindra has tied up with Edx to reskill 117K employees in areas like Big Data, IoT, Machine Learning etc.
- Capgemini bagged the European Commission data infrastructure project.
- Insurance firm Direct Line Group appointed Capgemini for IT restructure.
- Atos has partnered Google Cloud to create secure enterprise business solutions in Artificial Intelligence, Machine Learning, Hybrid Cloud, data analytics and the digital workplace.
- Atos and Siemens have committed 100M pounds for R&D in AI, Big Data, IoT and Cybersecurity.
- DXC Technology launched an Agile Process Automation (APA), which combines cloud and robotic process automation (RPA) with embedded artificial intelligence (AI) to enhance a company’s business processes.
- Cognizant has tied up with Goodwill University to impart training in Digital technologies and other IT courses.
- Avanade has tied up with Microsoft to create new AI-based solutions.
- Avande collaborated with Hortonworks to provide big data solutions to enterprises.
- Pactera launched an innovation outpost called Moonshot to lead global clients through the next era of digital products with a heavy emphasis on artificial intelligence, data and continuous software delivery paired with next generation human-centered experience design.

Note: Data curated by DRAUP engineering services deals database updated in April, 2018
US, India, China, and UK have the most number of job openings in AI & Big Data/Analytics

<table>
<thead>
<tr>
<th>Job Title</th>
<th>ML</th>
<th>Big Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML - Data Scientist, Revolution Analytics, Test Architect, Research Scientist-Computer Vision</td>
<td>Flipkart, Amazon, Accenture, Intel, Citi, Amazon</td>
<td>Google, EY, Deloitte, Mark's, Synopsys, Uber, CapitalOne</td>
</tr>
<tr>
<td>Big Data - Applications Developer, Cloud Architect</td>
<td>JP Morgan, Intel Amazon, Google, Microsoft, NVIDIA, Facebook</td>
<td>IBM, Google, Microsoft, JP Morgan, Amazon Home Depot, McLane Company, Ring Central, Sirius XM</td>
</tr>
</tbody>
</table>

AI and Big data Analytics Global Job Opening distribution

<table>
<thead>
<tr>
<th>~515,000</th>
<th>~72,000</th>
<th>~67,300</th>
<th>~32,100</th>
<th>~5,000</th>
<th>~15,700</th>
<th>~310,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>India</td>
<td>77%</td>
<td>79%</td>
<td>81%</td>
<td>76%</td>
<td>83%</td>
<td>80%</td>
</tr>
<tr>
<td>China</td>
<td>23%</td>
<td>21%</td>
<td>19%</td>
<td>24%</td>
<td>17%</td>
<td>20%</td>
</tr>
<tr>
<td>UK</td>
<td>ML</td>
<td>Big Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>ML</td>
<td>Big Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>US</td>
<td>ML</td>
<td>Big Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Data curated from global job portals such as LinkedIn, Indeed, Glassdoor etc.
About 1 Million jobs are expected to be created in AI and Big Data/Analytics roles in 2021

- Globally, Job Creation for AI and Big Data Analytics roles will reach **960K** in 2021 with an average **CAGR of 23%**
- India is expected to grow at a **faster rate (~25%)** compared to the rest of the world
- The Job Creation for Big Data/Analytics roles will grow at a much **lower rate (~7%)** compared with AI over the next 3 years

![Graph showing job openings from 2018 to 2021](image)

**Note:** Data curated from global job portals such as LinkedIn, Indeed, Glassdoor etc.
A detailed analysis of over 200,000 resumes and job opening corpus shows that the AI and Big Data talent pool in Global 500 companies is around 250,000.

<table>
<thead>
<tr>
<th>Roles</th>
<th>Talent Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Analyst - Data Management</strong></td>
<td>33%</td>
</tr>
<tr>
<td><strong>2. Data Architect</strong></td>
<td>10%</td>
</tr>
<tr>
<td><strong>3. Database Engineering</strong></td>
<td></td>
</tr>
<tr>
<td>Big Data / Hadoop Administrator</td>
<td>7%</td>
</tr>
<tr>
<td>Data Warehouse Engineer</td>
<td>12%</td>
</tr>
<tr>
<td><strong>4. Data Scientist</strong></td>
<td>30%</td>
</tr>
<tr>
<td><strong>5. Applied AI</strong></td>
<td></td>
</tr>
<tr>
<td>Applied Data Scientist - Vision</td>
<td>5%</td>
</tr>
<tr>
<td>Applied Data Scientist - Speech</td>
<td>3%</td>
</tr>
</tbody>
</table>

**Total Directly relevant talent pool:** ~250,000

- Global Data Science and Applied Data Science talent is estimated at ~95,000 engineers employed across G500 organizations, start-ups and Service Providers
- Applied Data Science role is currently consolidated across tech giants and few platform start-ups

Note: DRAUP’s proprietary talent module was used to analyse jobs by job roles and skill type
Note: DRAUP’s Talent Simulation Module analysed jobs of ~2000 peer companies on the basis of skill type, salaries, talent pool, years of experience and employers
Majority of G500 organizations have employed Big Data talent across geographies; Computer Vision & NLP talent footprint is employed majorly by tech giants.

**AI and bigdata analytics Global G500 Talent distribution**

**Geo-Wise G500 Talent Distribution across skills**

- **West Coast**: ~85,000
  - Analyst-Data Management: 35%
  - Data Architect: 31%
  - Database Engineering: 16%
  - Data Scientist: 7%
  - Applied ML: 10%

- **East Coast**: ~25,000
  - Analyst-Data Management: 42%
  - Data Architect: 34%
  - Database Engineering: 24%
  - Data Scientist: 9%
  - Applied ML: 13%

- **Canada**: ~11,000
  - Analyst-Data Management: 55%
  - Data Architect: 23%
  - Database Engineering: 11%
  - Data Scientist: 1%
  - Applied ML: 1%

- **UK**: ~9,000
  - Analyst-Data Management: 41%
  - Data Architect: 26%
  - Database Engineering: 24%
  - Data Scientist: 8%
  - Applied ML: 1%

- **Germany**: ~15,000
  - Analyst-Data Management: 54%
  - Data Architect: 23%
  - Database Engineering: 15%
  - Data Scientist: 4%
  - Applied ML: 4%

- **Israel**: ~11,000
  - Analyst-Data Management: 44%
  - Data Architect: 29%
  - Database Engineering: 22%
  - Data Scientist: 4%
  - Applied ML: 2%

- **India**: ~43,000
  - Analyst-Data Management: 56%
  - Data Architect: 20%
  - Database Engineering: 16%
  - Data Scientist: 5%
  - Applied ML: 3%

- **China**: ~45,000
  - Analyst-Data Management: 34%
  - Data Architect: 16%
  - Database Engineering: 16%
  - Data Scientist: 4%
  - Applied ML: 4%

- **Singapore**: ~2,400
  - Analyst-Data Management: 45%
  - Data Architect: 11%
  - Database Engineering: 11%
  - Data Scientist: 4%
  - Applied ML: 1%

**Note**: DRAUP Talent Module

**Traditional Hubs for Engineering for the Tech Giants** – Google, Amazon, Facebook, Apple, and Microsoft hold ~35% of global G500 Big Data Machine learning to NLP & Computer vision talent. Driverless Cars, Drones, Predictive medicine, Cyber Security are the hot areas.

**MNCs such as Google DeepMind group, Microsoft and IBM Watson** group have large scaled AI and Big Data labs in Canada. Automotive OEMs such as Ford, GM have set-up large AI labs in Montreal and Toronto to expand autonomous driving research.

**UK AI & Big Data talent is significantly is engaged in large companies such as Google DeepMind group, IBM Watson Health, and Microsoft (Big Data Analytics for Bing and Skype) etc.**

**Microsoft Israel technical centre works on Medical Imaging** for predictive Eyecare.

**Automotive OEMs like Renault, Volkswagen partnering with Autonomous start-ups like Mobileye**

**Presence of heavyweights like IBM, GE, Microsoft, Amazon, etc. has helped create an AI/BD ecosystem in India. Amazon India, Walmart Labs have invested heavily on analysing Big Data sets generated by customer interactions across Retail, Seller Services and leverage NLP algorithms to predict customer behaviour.**

**China’s AI & Big Data Talent is employed predominantly in large companies, both local and MNCs, such as Baidu, Tencent, Alibaba, Ebay, Amazon etc. Baidu is investing heavily in Vision for Autonomous Driving and fleet route optimization by analyzing Mn of vehicle datasets.**
Over 60% of the demand is distributed across Enterprise Software, Consumer Electronics and BFSI.
**Data Scientist Job Progression**: Business Analyst and Algorithms Engineer are the key roles which have progressed into Data Scientist role.

### Data Scientist Job Progression Roadmap

<table>
<thead>
<tr>
<th>Past Role</th>
<th>Previous Role</th>
<th>Current Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business/Data Analyst 1</td>
<td>Business/Data Analyst 1</td>
<td>Data Scientist</td>
</tr>
<tr>
<td>Algorithms Developer/Engineer</td>
<td>Data Warehouse Engineer</td>
<td>Data Scientist</td>
</tr>
<tr>
<td>Algorithms Developer/Engineer</td>
<td>Machine learning specialist</td>
<td>Data Scientist</td>
</tr>
<tr>
<td>Statistician</td>
<td>Statistician</td>
<td>Data Scientist</td>
</tr>
<tr>
<td>Statistician</td>
<td>Area/Operations Manager</td>
<td>Data Scientist</td>
</tr>
<tr>
<td>Research &amp; Consulting</td>
<td>Marketing Programs Manager</td>
<td>Data Scientist</td>
</tr>
<tr>
<td>IT Admin</td>
<td>Data Warehouse Engineer</td>
<td>Data Scientist</td>
</tr>
</tbody>
</table>

~24,000 Data Scientist profile analysed across 10 Tech Giants*  

~85% profiles  

~15% profiles  

**~24,000 Data Scientist profile analysed across 10 Tech Giants**: Google, Facebook, Apple, Microsoft, Intel, Amazon, IBM, SAP, Salesforce, Cisco


Tech Giants*: Google, Facebook, Apple, Microsoft, Intel, Amazon, IBM, SAP, Salesforce, Cisco
**Talent cost by roles:** India and China have cost arbitrage compared to US Tier-1 locations such as Bay area and Seattle. But this difference is expected to shrink based on our SME interviews.

| Note: DRAUP’s Talent Simulation Module. Median cost is derived from analysis of ~2000 tech company employers globally |

---

**Median Salary (USD) per annum**

<table>
<thead>
<tr>
<th>$30,000.00</th>
<th>$50,000.00</th>
<th>$70,000.00</th>
<th>$90,000.00</th>
<th>$1,10,000.00</th>
<th>$1,30,000.00</th>
<th>$1,50,000.00</th>
<th>$1,70,000.00</th>
<th>$1,90,000.00</th>
</tr>
</thead>
</table>

**FTE costs across global locations (in USD)**

- **Big Data/Analytics**
- **AI**

**Insights**

- The average FTE for talent (AI & Big Data/Analytics) in the Silicon Valley Bay Area costs about 50% more than that of talent in Massachusetts.
- Globally, Machine Learning talent costs around 15-25% more than Big Data/Analytics talent.
- For both AI and Big Data/Analytics, talent in India costs only around a third of the talent in Bay Area.
AGENDA

01 Demand Analysis

02 Location Deep Dive Analysis

KEY HIGHLIGHTS

• Location Characteristics: Key hotspots, cost analysis and top employers

• Workloads across G500, start-ups and service providers

• Talent characteristics of G500 and Start-Ups employers

• Talent profile deep dive
US: Employed Talent consolidation is across Tech giants and platform start-ups having large scale engineering labs focussed on cross industry application; Universities provide mature homegrown talent

<table>
<thead>
<tr>
<th>Mature AI talent pool concentrated across Tech giants, Start-ups and Universities</th>
</tr>
</thead>
</table>

1. **Tech giants** employ large scale AI/BD talent pool concentrated in tier-1 locations
   - Google
   - Facebook
   - Amazon
   - Apple
   - Microsoft
   - ~25,000 AI/BD engineers employed
   - 100+ Start-up Acquisitions made globally
   - 25 bn+ AI/BD Spend annually

2. **Mature platform start-ups**: 100 Mn+ investment start-up are concentrated in US. Focus area is cross-functional industry platform
   - 300+ Platform start-ups
   - 40% Deep Learning
   - 20% NLP
   - 10% Computer Vision
   - 10% Robotics

3. **Mature University Ecosystem**: MIT, NYU and Stanford produce ~2500+ fresh AI/ML talent pool individually
   - 20+ Tiet-1 Universities with PhD programmes in AI/Big Data
   - Facebook - NYU
     - Facebook hires NYU deep learning expert to run its AI lab
   - Amazon - UC Irvin
     - Amazon hires UC Irvine principal scientist to run its Deep learning AWS lab
US: Bay Area & Seattle have talent consolidation across Tech giants and start-ups; Central and Eastern region’s talent is largely spread across start-ups

### Supply Drivers

<table>
<thead>
<tr>
<th>Location</th>
<th>Talent Quality</th>
<th>Talent Scalability</th>
<th>Cost Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bay Area</td>
<td>It has high quality Data science and ML talent</td>
<td>Average scalability due to average fresh university talent supply</td>
<td>High talent cost due to scarce AI talent</td>
</tr>
<tr>
<td>Boston</td>
<td>It is the emerging hub for AI and Big data analytics</td>
<td>High scalability fueled by researchers at local universities</td>
<td>High Talent cost in Boston for the given talent quality</td>
</tr>
<tr>
<td>Dallas</td>
<td>It has low AI/ML talent and average Big data/ analytics talent</td>
<td>Average scalability due to average fresh university talent supply</td>
<td>Low talent cost as most of Dallas talents are fresh talents</td>
</tr>
<tr>
<td>Seattle</td>
<td>It has most number installed DS/ Big data talents after Bay area</td>
<td>High scalability options in Seattle due to availability of university talent pool</td>
<td>Talent cost in Seattle is high due to intense demand for ML expertise</td>
</tr>
<tr>
<td>Phoenix</td>
<td>It is an emerging location for AI tech companies</td>
<td>Average scalability due to average fresh university talent supply</td>
<td>Low talent cost as most of Dallas talents are fresh talents</td>
</tr>
</tbody>
</table>

###AI/Big Data Talent

- **Seattle**: 63K+ Talent
  - Tech Companies: 43%
  - Start-ups: 57%
- **Bay Area**: 110K+ Talent
  - Tech Companies: 48%
  - Start-ups: 52%
- **Phoenix**: 9K+ Talent
  - Tech Companies: 33%
  - Start-ups: 67%
- **Dallas**: 17K+ Talent
  - Tech Companies: 29%
  - Start-ups: 71%
- **Austin**: 10K+ Talent
  - Tech Companies: 60%
  - Start-ups: 40%
- **Boston**: 35K+ Talent
  - Tech Companies: 24%
  - Start-ups: 76%
- **Others**: 14%

Note: DRAUP’s Talent Simulation Module. We have analysed ~2,000 tech companies and ~10,000 start-ups.
Bay Area: Hotspot for mature AI & Big Data talent pool in Data Science and Data Management roles

- **San Francisco Bay Area**
  - ~110,000 Demand talent
  - $195K Median Cost

**AI Talent Hubs**
- San Francisco
- Palo Alto
- Mountain View
- Santa Clara
- Sunnyvale
- San Jose
- San Mateo

**Headcount Distribution (By Skills)**
- Data Scientist: 35%
- Analyst: Data Management: 36%
- Database Engineering: 11%
- Data Architect: 8%
- Applied AI: 10%

**Key Employers**
- G500 companies: 56K+
- AI/Big Data start-ups: 60K+
- Universities: ~9 - 11

**Top Verticals**
- Enterprise & Consumer Software, Semi-Conductor, Consumer Electronics

**Total Employable Talent**
- ~ 1,500

**Total Number**
- 60K+

Note: DRAUP’s Talent Simulation Module. Median cost is derived from analysis of ~2000 tech company employers globally.
Bay Area – Data Science is the most employed role employed across top players with nearly every engineering priority focussed on building cross industry AI platform.

Top 5 tech companies AI/Big Data head count ~12000

- **Google**
  - ~4600
  - Data Scientist: 66%
  - Database Engineer: 16%
  - Applied AI: 3%
  - Data Architect: 4%
  - Analyst – Data Management: 11%
  - Responsible for key AI-based product development activities across major business units such as Tensor Flow, Waymo, DeepMind, Android, Youtube, Chrome, Maps etc.

- **Facebook**
  - ~2000
  - Data Scientist: 64%
  - Database Engineer: 25%
  - Applied AI: 2%
  - Data Architect: 4%
  - Analyst – Data Management: 5%
  - Facebook key AI product priorities such across DeepText, Facial recognition, Oculus Advanced Development Group, Instagram & WhatsApp etc are based out of Bay Area.

- **Apple**
  - ~3000
  - Data Scientist: 56%
  - Database Engineer: 16%
  - Applied AI: 6%
  - Data Architect: 6%
  - Analyst – Data Management: 16%
  - Apple’s HQ - 80% of R&D talent is out of this center. Apple’s core ML teams for Siri, Developer’s platform, iOS, iCloud, etc are all based out of this center.

- **Oracle**
  - ~1000
  - Data Scientist: 21%
  - Database Engineer: 10%
  - Applied AI: 1%
  - Data Architect: 10%
  - Analyst – Data Management: 58%
  - Oracle AI Platform Cloud Services team is based out of the Bay Area centre.

- **Uber**
  - ~1200
  - Data Scientist: 78%
  - Database Engineer: 15%
  - Applied AI: 1%
  - Data Architect: 2%
  - Analyst – Data Management: 4%
  - Uber’s major share of talent pool responsible for Fleet management services, location intelligence and autonomous driving group are based out of Bay Area.

Note: DRAUP’s Talent Simulation Module. We have analysed ~2,000 tech companies and ~10,000 start-ups.
**Bay Area** – Google and Facebook have a high number of experienced engineers laterally hired from peer tech giants and graduated from Tier-1 US Universities such as CMU and Stanford University.

### Google

<table>
<thead>
<tr>
<th>Work Profile</th>
<th>Education</th>
<th>Past Companies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Maximum number of employees hired from</strong></td>
<td>Stanford University, University of California, CMU</td>
<td></td>
</tr>
<tr>
<td><strong>Most of the employees have worked in</strong></td>
<td>Microsoft, Yahoo</td>
<td></td>
</tr>
<tr>
<td><strong>Work Experience</strong></td>
<td>&gt; 10 years</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Open Source Big Data Developer Advocate</th>
<th>Work Experience in years 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Past Companies</td>
<td><a href="#">Amazon</a>, <a href="#">Linux</a>, <a href="#">Apache</a></td>
</tr>
<tr>
<td>Education</td>
<td><a href="#">University of Waterloo</a>, <a href="#">MIT</a>, <a href="#">Stanford</a></td>
</tr>
</tbody>
</table>

### Facebook

<table>
<thead>
<tr>
<th>Work Profile</th>
<th>Education</th>
<th>Past Companies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Maximum number of employees hired from</strong></td>
<td>Stanford University, University of California, CMU</td>
<td></td>
</tr>
<tr>
<td><strong>Most of the employees have worked in</strong></td>
<td>Microsoft, Google</td>
<td></td>
</tr>
<tr>
<td><strong>Work Experience</strong></td>
<td>&gt; 10 years</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Senior Data Scientist, Applied Machine learning</th>
<th>Work Experience in years 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Past Companies</td>
<td><a href="#">Instagram</a>, <a href="#">Coinbase</a>, <a href="#">Hyperplane</a></td>
</tr>
<tr>
<td>Education</td>
<td><a href="#">University of Toronto</a>, <a href="#">UCSD</a>, <a href="#">Scripps</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Engineering Lead at Facebook</th>
<th>Work Experience in years 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Past Companies</td>
<td><a href="#">Uber</a>, <a href="#">Yahoo</a>, <a href="#">Facebook</a></td>
</tr>
<tr>
<td>Education</td>
<td><a href="#">Scripps</a>, <a href="#">UCLA</a>, <a href="#">UCSD</a></td>
</tr>
</tbody>
</table>

Note: DRAUP’s Talent Simulation Module. We have analysed ~2,000 tech companies and ~10,000 start-ups.
Bay Area – Similarly Apple & Uber have recruited high end engineering roles from Microsoft; Majority of engineers are graduates from University of California Berkley and Caltech

<table>
<thead>
<tr>
<th>Maximum number of employees hired from</th>
<th>Stanford University, University of California, CMU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most of the employees have worked in</td>
<td>Microsoft, Yahoo</td>
</tr>
<tr>
<td>Work Experience</td>
<td>&gt; 10 years</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Maximum number of employees hired from</th>
<th>Stanford University, University of California, CMU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most of the employees have worked in</td>
<td>Microsoft, Google</td>
</tr>
<tr>
<td>Work Experience</td>
<td>6 - 10 years</td>
</tr>
</tbody>
</table>

**Work Profile**

### Engineering Lead – Siri Natural language and Machine learning platform

- **Work Experience in years**: 15
- **Past Companies**
  - IBM
  - LinkedIn
- **Education**
  - Stanford University

### Director of Engineering, Mobile Apps & Mobile Commerce

- **Work Experience in years**: 10
- **Past Companies**
  - TOSHIBA
  - Intuit
- **Education**
  - Stanford University

### Leading Multiple Data Science and Machine Learning Organizations

- **Work Experience in years**: 10
- **Past Companies**
  - LinkedIn
  - Intuit
- **Education**
  - Stanford University

### Senior Data Science Manager for Safety Data Science

- **Work Experience in years**: 13
- **Past Companies**
  - Goldman Sachs
  - Intuit
- **Education**
  - Stanford University

**Note**: DRAUP’s Talent Simulation Module. We have analysed ~2,000 tech companies and ~10,000 start-ups.
## Total Start-Up Talent pool

<table>
<thead>
<tr>
<th>Top Verticals/Tech focus</th>
<th>Key Start-ups</th>
<th>Key Job profiles</th>
<th>Previous Experience</th>
<th>Workloads/ Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core AI</td>
<td>sentient</td>
<td>ML Research Scientist</td>
<td>Berkeley University of California</td>
<td>• Use extensive experience and knowledge in deep learning, natural language processing, graphical models, and/or computer vision to help people and businesses make better decision</td>
</tr>
<tr>
<td>Computer Vision</td>
<td>vicarious.</td>
<td>Senior Roboticist</td>
<td>Honeywell, Carnegie Mellon University</td>
<td>• Combine data science and ML to develop solutions using Python, PySpark, Hadoop, Matlab</td>
</tr>
<tr>
<td>NLP</td>
<td>figure eight</td>
<td>ML Scientist, Computer Vision</td>
<td>UCI, UC Berkeley</td>
<td>• Anomaly detection, extending ML algo like neural networks</td>
</tr>
<tr>
<td>Data Science</td>
<td>gumgum</td>
<td>Computer vision scientist</td>
<td>NANTMOBILE, UCLA</td>
<td>• Training deep learning models for image classification, object detection and categorization and object segmentation for natural images, drone</td>
</tr>
<tr>
<td></td>
<td>conversica</td>
<td>Sr Data Scientist</td>
<td>Intel</td>
<td>• Large scale object detection and bounding box regression using convolutional neural networks on social media images, deployed as a web service</td>
</tr>
<tr>
<td></td>
<td>SoundHound</td>
<td>Machine Learning Engineer-Language Modelling</td>
<td>AYASDI</td>
<td>• Build products using machine learning, deep learning, reinforcement learning, and natural language processing techniques</td>
</tr>
<tr>
<td></td>
<td>Palantir</td>
<td>Machine Learning Engineer</td>
<td>Google</td>
<td>• ML skills and knowledge of MapReduce/Spark, speech recognition technologies, Lucene, Nutch to build Statistical Language Models</td>
</tr>
<tr>
<td></td>
<td>AppZen</td>
<td>Lead data scientist</td>
<td>GEORGIA</td>
<td>• Build time series search, propensity modeling, record linkage, entity extraction and other ML efforts, for applications in fraud, insurance, retail, cyber etc.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Working on analyzing mobile and web data. More specifically, analyzing terabytes of digital data for understanding trends and patterns</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Using Big Data technology for building machine learned models and statistical models for predicting characteristics for user</td>
</tr>
</tbody>
</table>

### Bay Area – Hotspot for multi-million dollar platform startups, employing small scale teams comprised of Sr. engineering AI/BD roles laterally hired from Tech giants

- **Total Number of AI/Big Data Start-ups**: ~1500
- ** Median Salary**: $195–200K

**Note**: DRAUP’s Talent Simulation Module. We have analysed ~2,000 tech companies and ~10,000 start-ups.
Seattle: 2nd largest talent hotspot in US with majority of talent in Data management and Data Scientist role employed across G500 employers and vertical specific start-ups.

Headcount Distribution (By Skills):
- **35%** Data Scientist
- **30%** Analyst: Data Management
- **20%** Database Engineering
- **5%** Data Architect
- **10%** Applied AI

Key Employers:
- **G500 companies**: 27K+
- **AI/Big Data start-ups**: 36K+
- **AI/Big Data Universities**: 4

Total Number:
- **Total Employable Talent**: 43K+
- **Top Verticals**: Enterprise & Consumer software and services, BFSI
- **Tech companies**
- **Start-ups**

Note: DRAUP's Talent Simulation Module. We have analysed ~2,000 tech companies and ~10,000 start-ups.
Seattle – Microsoft Azure ML, Facebook Deep Learning and Amazon Core NLP teams are based out of Seattle; Brick and Mortar players like Starbucks employ data analytics teams

<table>
<thead>
<tr>
<th>ML/Big Data Headcount</th>
<th>Data Scientist</th>
<th>Database Engineer</th>
<th>Applied ML</th>
<th>Data Architect</th>
<th>Analyst – Data Management</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Microsoft</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>~2100</td>
<td>61%</td>
<td>23%</td>
<td>9%</td>
<td>5%</td>
<td>2%</td>
</tr>
<tr>
<td><strong>Amazon</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>~2000</td>
<td>51%</td>
<td>18%</td>
<td>27%</td>
<td>3%</td>
<td>1%</td>
</tr>
<tr>
<td><strong>Expedia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>~300</td>
<td>24%</td>
<td>44%</td>
<td>15%</td>
<td>17%</td>
<td>0%</td>
</tr>
<tr>
<td><strong>Facebook</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>~200</td>
<td>66%</td>
<td>10%</td>
<td>20%</td>
<td>2%</td>
<td>4%</td>
</tr>
<tr>
<td><strong>Starbucks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>~200</td>
<td>22%</td>
<td>29%</td>
<td>42%</td>
<td>7%</td>
<td>0%</td>
</tr>
</tbody>
</table>

- **Microsoft**
  - Develop new NLP capabilities and text understanding APIs in Text Analytics Cognitive Service. Extract insights from unstructured data and build predictive solutions for core NLP problems.

- **Amazon**
  - Build Deep Learning tools and APIs, and contribute to open source frameworks such as MXNet and Keras. Help build industry-leading conversational technologies and machine learning systems that powers Alexa.

- **Expedia**
  - Design, build, scale, and optimize the data infrastructure as a highly innovative service that enhance and innovate the Expedia e-Commerce ecosystems. Build full-stack data from multiple data sources leveraging cloud systems, and structured/unstructured data.

- **Facebook**
  - Core Machine Learning Team in Seattle develops and optimizes various algorithms including Neural Networks, Boosted Decision Trees, Sparse Linear Models, and Deep Learning for several ranking teams including Ads, Feed, Search, Instagram and others.

- **Starbucks**
  - Starbucks AI and analytics team in Seattle is using customers’ spending patterns data to develop insights that would help in generating personalized product promotions such as user based reward cards.

Note: DRAUP’s Talent Simulation Module. We have analysed ~2,000 tech companies and ~10,000 start-ups.
Seattle – Microsoft and Amazon employee mid level engineering talent hired from IT giants such as TCS, IBM and Accenture

<table>
<thead>
<tr>
<th>Microsoft</th>
<th>Amazon</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Maximum number of employees hired from</strong></td>
<td>University of Washington, Carnegie Mellon University</td>
</tr>
<tr>
<td><strong>Most of the employees have worked in</strong></td>
<td>Amazon, IBM, Intel</td>
</tr>
<tr>
<td><strong>Average Work Experience</strong></td>
<td>7-10 years</td>
</tr>
</tbody>
</table>

**Work Profile**

<table>
<thead>
<tr>
<th>Machine Learning Scientist</th>
<th>Data Scientist</th>
<th>Machine Learning Scientist</th>
<th>Senior Data Engineer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Work Experience in years</strong></td>
<td>11</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Past Companies</td>
<td>SAMSUNG</td>
<td>amazon</td>
<td>SAMSUNG</td>
</tr>
<tr>
<td>NorthWestern University</td>
<td>Pfizer</td>
<td>SAMSUNG</td>
<td>INTEL</td>
</tr>
<tr>
<td>Purdue University</td>
<td></td>
<td></td>
<td>AMEX</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td>Education</td>
</tr>
<tr>
<td>Georgia Institute of Technology</td>
<td></td>
<td></td>
<td>W University of Washington</td>
</tr>
<tr>
<td>Princeton University</td>
<td></td>
<td></td>
<td>Stanford University</td>
</tr>
</tbody>
</table>

Note: DRAUP's Talent Simulation Module. We have analysed ~2,000 tech companies and ~10,000 start-ups.
Seattle: Majority of start-up talent pool is with Core platform start-ups; Most of the engineers have migrated from Microsoft and other Enterprise Software giants

### Total Start-Up Talent pool

- **~ 120** Total Number of AI/Big Data Start-ups
- **~$100K** Median Salary

### Top Verticals/ Tech focus
- Core AI
- Computer Vision
- NLP
- Data Science

### Key Start-ups
- KenSci
- Utrip
- XNOR.ai
- Mighty AI
- Saykara
- Signal genomics

### Key Job profiles
- Data Scientist
- Machine Learning Engineer
- Research Engineer
- NLP engineer
- Senior Machine Learning Engineer
- Bioinformatics Data Scientist

### Previous Experience
- University of Washington
- Dell
- Google
- Facebook
- Amazon
- University of Washington

### Workloads/ Skills

- **Data Scientist**
  - Created a verbose set of data structures for representing a variety of healthcare data points coming from multiple source healthcare data formats (CCLF, eCW, EDI, etc.).
  - Data-driven destination discovery and trip planning platform. Leveraging artificial intelligence, Utrip helps partners deliver highly personalized travel experiences to their end customers.

- **Machine Learning Engineer**
  - Accelerating AI and deep learning adoption in consumer and business devices through enterprise AI algorithms and self-serve developer platform.
  - Work within Ruby codebase to maintain and extend existing systems, integrations, and infrastructure.

- **Research Engineer**
  - Build large scale analytics engines for offline and online analysis of healthcare documents.
  - Design and implement new features and expand existing ones.

- **NLP engineer**
  - Development and optimization of data analysis pipelines for NGS assays, in production or research environments.
  - Develop new technologies/protocols for improving and scaling genomic data production.

---

Note: DRAUP’s Talent Simulation Module. We have analysed ~2,000 tech companies and ~10,000 start-ups.
Boston: AI/BD employed talent is consolidated in start-ups from Healthcare and Retail industries; Boston Universities such as MIT and North-eastern University provide mature engineering talent pool.

Key Employers

- G500 companies: 8.4K+
- AI/Big Data start-ups: 26K+
- AI/Big Data Universities: ~300
- Total Employable Talent: 26K+
- Top Verticals: Healthcare, Retail

Headcount Distribution (By Skills)

- Data Scientist: 40%
- Analyst: Data Management: 30%
- Database Engineering: 15%
- Data Architect: 8%
- Applied AI: 7%

Total Number

- ~300

Note: DRAUP’s Talent Simulation Module. We have analysed ~2,000 tech companies and ~10,000 start-ups.
Boston – Along with Tech giants, traditional Retail and Banking players are also key employers of Data Science skills primarily working for digital transformation of backend and frontend operations.

Key focus areas are Computational Biology, Computer Science, Cryptography, Machine Learning, Systems and Security.

Wayfair launched “Search with Photo” a new feature that leverages artificial intelligence to assist consumers in the product buying process.

Creating AI/ML driven products that combine natural language understanding with batch and real time sales and service models with the goal of improving the omni-channel customer experience in a measurable way.

The R&D team in Boston area develops software automation infrastructure for Amazon’s Kiva robotic systems in an integrated service-oriented cloud computing environment.

Embedded Systems and Mobile Apps for Android are some of primary activities carried out Boston Area.

Microsoft has about ~72 Data Scientists, ~6 Database Engineers, ~0 Applied ML specialists, ~5 Data Architects, and ~19 Analysts for Data Management.

Google has about ~30 Data Scientists, ~22 Database Engineers, ~3 Applied ML specialists, ~0 Data Architects, and ~13 Analysts for Data Management.

Amazon has about ~130 Data Scientists, ~13 Database Engineers, ~10 Applied ML specialists, ~2 Data Architects, and ~5 Analysts for Data Management.

Wayfair has about ~130 Data Scientists, ~10 Database Engineers, ~0 Applied ML specialists, ~2 Data Architects, and ~16 Analysts for Data Management.

Fidelity has about ~190 Data Scientists, ~15 Database Engineers, ~0 Applied ML specialists, ~13 Data Architects, and ~55 Analysts for Data Management.

Note: DRAUP’s Talent Simulation Module. We have analysed ~2,000 tech companies and ~10,000 start-ups.
Boston – IBM, Nokia and CVS Health are among the top lateral hiring prospects for Microsoft and Amazon.

**Microsoft**
- **Maximum number of employees hired from**: MIT, Northeastern university, Harvard university
- **Most of the employees have worked in**: IBM, Nokia
- **Work Experience**: > 10 years

**Amazon**
- **Maximum number of employees hired from**: MIT, Northeastern University, Boston University
- **Most of the employees have worked in**: Microsoft, IBM
- **Work Experience**: > 10 years

**Data Scientist II**
- **Work Experience in years**: 6
- **Past Companies**:
  - CVS Health
  - Landmark
  - Predixion
- **Education**:
  - Wellesley University
  - Northeastern University

**Partnering Director of Engineering, Machine learning**
- **Work Experience in years**: 20
- **Past Companies**:
  - Landmark
  - Predixion
  - CVS Health
- **Education**:
  - Cornell University
  - Northeastern University

**Software Development Engineer, Alexa Machine learning Data Platform**
- **Work Experience in years**: 13
- **Past Companies**:
  - Microsoft
  - Amazon
- **Education**:
  - UMass Amherst
  - Boston University

**Software Quality Assurance Engineer**
- **Work Experience in years**: 8
- **Past Companies**:
  - Bullhorn
  - Huawei
- **Education**:
  - Emory University

Note: DRAUP's Talent Simulation Module. We have analysed ~2,000 tech companies and ~10,000 start-ups.
Boston – Healthcare and Retail industries have large number of start-up talent pool majorly employed from Tier-2 Eastern US Universities

Total Start-Up Talent pool

- Healthcare
- Retail
- Deep Learnings
- Robotics

~ 300
Total Number of AI/Big Data Start-ups

~$90-100K
Median Salary

Key Start-ups
- Data Scientist
- Software Engineer
- Senior Data Scientist
- Senior Computer Vision Researcher
- Data Scientist
- Software Engineer
- Embedded Software Engineer
- Deep learning Engineer

Key Job profiles
- Data Scientist
- Software Engineer
- Senior Data Scientist
- Senior Computer Vision Researcher
- Data Scientist
- Software Engineer
- Embedded Software Engineer
- Deep learning Engineer

Previous Experience
- Data processing and statistical analysis in the field of precision medicine, biomarker identification and patient stratification
- Developed, maintained, and deployed internal and consumer facing API machine learning services utilizing tools such as Nvidia-Docker, Kafka, Kubernetes and DC/OS
- Architecting the ML Platform enabling ingestion, ETL, feature extraction, model training and serving real time/offline predictions
- Developed, maintained, and deployed internal and consumer facing API machine learning services utilizing tools such as Nvidia-Docker, Kafka, Kubernetes and DC/OS
- Knowledge of Data Science techniques, Spark / YARN distributed machine learning libraries, Hadoop Ecosystem to develop key components of platform
- Build a web application which consumes data from REST API Service using AngularJS and developed the RESTful Web Service using C#.NET.
- Develop prototype embedded systems for self-driving cars, or optimize our AD system for GPU architectures
- Designing Software to make consumer-level robots easier and more fun to use

Workloads/ Skills

- Data processing and statistical analysis in the field of precision medicine, biomarker identification and patient stratification
- Developed, maintained, and deployed internal and consumer facing API machine learning services utilizing tools such as Nvidia-Docker, Kafka, Kubernetes and DC/OS
- Architecting the ML Platform enabling ingestion, ETL, feature extraction, model training and serving real time/offline predictions
- Developed, maintained, and deployed internal and consumer facing API machine learning services utilizing tools such as Nvidia-Docker, Kafka, Kubernetes and DC/OS
- Knowledge of Data Science techniques, Spark / YARN distributed machine learning libraries, Hadoop Ecosystem to develop key components of platform
- Build a web application which consumes data from REST API Service using AngularJS and developed the RESTful Web Service using C#.NET.
- Develop prototype embedded systems for self-driving cars, or optimize our AD system for GPU architectures
- Designing Software to make consumer-level robots easier and more fun to use

Note: DRAUP’s Talent Simulation Module. We have analysed ~2,000 tech companies and ~10,000 start-ups.
**Location Deep Dive:** Austin has niche availability of talent employed across retail and healthcare giants and vertical specific startups.

- **Headcount Distribution (By Skills):**
  - **30%** Data Scientist
  - **45%** Analyst: Data Management
  - **15%** Database Engineering
  - **10%** Data Architect

- **Key Employers:**
  - G500 companies: 6K+
  - AI/Big Data start-ups: 4K+
  - AI/Big Data Universities: 4 - 6

- **Total Number of Employable Talent:**
  - Retail, BFSI, Healthcare: ~70

- **Total Number of Employable Talent:**
  - AI/Big Data start-ups: ~70

- **Median Cost:** $150K
- **Demand talent:** ~10,000
- **Location:** Austin

Note: DRAUP’s Talent Simulation Module. We have analysed ~2,000 tech companies and ~10,000 start-ups.
Austin – GM’s autonomous ride-sharing facility is based out of Austin; Dell, HomeAway and Oracle have engineering focus on Chatbots and messaging platforms.

Top 5 tech companies AI/Big Data head count: 880

<table>
<thead>
<tr>
<th>ML/Big Data Headcount</th>
<th>Data Scientist</th>
<th>Database Engineer</th>
<th>Applied ML</th>
<th>Data Architect</th>
<th>Analyst – Data Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>~140</td>
<td>24%</td>
<td>10%</td>
<td>0%</td>
<td>11%</td>
<td>55%</td>
</tr>
<tr>
<td>~200</td>
<td>20%</td>
<td>30%</td>
<td>0%</td>
<td>18%</td>
<td>33%</td>
</tr>
<tr>
<td>~300</td>
<td>37%</td>
<td>12%</td>
<td>0%</td>
<td>13%</td>
<td>38%</td>
</tr>
<tr>
<td>~100</td>
<td>16%</td>
<td>6%</td>
<td>0%</td>
<td>10%</td>
<td>68%</td>
</tr>
<tr>
<td>~140</td>
<td>30%</td>
<td>35%</td>
<td>0%</td>
<td>6%</td>
<td>29%</td>
</tr>
</tbody>
</table>

Applying AI and ML technology to turn data into intelligent insights, drive a faster time to market and achieve better outcomes.

General Motors has introduced its Maven-car sharing platform in Austin. The Maven fleet is staged at the GM IT Innovation Center.

Has partnered with General Motors to add its artificial intelligence into the latter’s cars, in order to sell the customer stuff while driving.

Created Oracle chatbots by taking advantage of advances in AI and ML, that enables one to build and employ intelligent chatbots that optimize customer engagement.

Uses the first AI powered Secure messaging platform developed by TRACK.

Note: DRAUP’s Talent Simulation Module. We have analysed ~2,000 tech companies and ~10,000 start-ups.
### Austin
Most of the recruits for IBM and GM are from University of Texas; Dell, Intel and HP are their top lateral hiring prospects

<table>
<thead>
<tr>
<th>Tech companies</th>
<th>Start-ups</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Texas at Austin, Georgia Institute of Technology</td>
<td>University of Texas at Austin, Texas A&amp;M, Texas State University</td>
</tr>
</tbody>
</table>

| Most of the employees have worked in | Intel, Dell | Dell, HP |

| Work Experience | > 10 years | > 10 years |

### Work Profile

<table>
<thead>
<tr>
<th>Global Cyber Resiliency &amp; Security Architect</th>
<th>Senior Data Architect</th>
<th>Artificial Intelligence/ Machine Learning Scientist</th>
<th>Lead Engineer Machine Learning Design</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Work Experience in years</strong></td>
<td>17</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td><strong>Past Companies</strong></td>
<td><strong>Past Companies</strong></td>
<td><strong>Past Companies</strong></td>
<td><strong>Past Companies</strong></td>
</tr>
<tr>
<td>ORACLE, McAfee Institute</td>
<td>Intel, eSO</td>
<td>HDR, IBM</td>
<td>T58 Technologies, Dell, IBM</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td><strong>Education</strong></td>
<td><strong>Education</strong></td>
<td><strong>Education</strong></td>
</tr>
<tr>
<td>RIT, Texas A&amp;M</td>
<td>Texas, RIT</td>
<td>Texas</td>
<td>Texas</td>
</tr>
</tbody>
</table>
Austin: Small engineering teams in healthcare and e-commerce employ AI & Big Data talent

<table>
<thead>
<tr>
<th>Total Start-Up Talent pool</th>
<th>Top Verticals/ Tech focus</th>
<th>Key Start-ups</th>
<th>Key Job profiles</th>
<th>Previous Experience</th>
<th>Workloads/ Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>~ 70</td>
<td>Ecommerce</td>
<td>BIGCOMMERCE</td>
<td>Database Engineer</td>
<td>Allegro, b</td>
<td>Key workloads include Performance Tuning and Query Optimization, DEVOps/Chatops (Slack) global collaboration, ETL Pipeline Design and Implementation.</td>
</tr>
<tr>
<td></td>
<td>Healthcare</td>
<td>HomeAway</td>
<td>Data Scientist</td>
<td>&amp;</td>
<td>Research and experiment on a variety of machine learning models for use in customer engagement and search/sort analysis.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KINNSNER</td>
<td>Database Engineer</td>
<td></td>
<td>Maintenance, customization and management of company information system with focus on supporting implementation of quality standards in design and development.</td>
</tr>
<tr>
<td></td>
<td>AdTech</td>
<td>KASASA</td>
<td>Sr Data Architect</td>
<td></td>
<td>Development of financial services products using MS Azure, SQL Server, Hadoop and other data analytics techniques</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Banker's Toolbox</td>
<td>Machine Learning Engineer</td>
<td></td>
<td>Key Skills required include NLP, Advanced Big Data Analysis, Reinforcement Learning, Bayesian Machine Learning, Python, Matlab, SQL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Outbound Engine</td>
<td>Big Data Engineer</td>
<td>Google, Ericsson</td>
<td>Workloads include development of data models using tools such as Hadoop, SQL for Search Engine Optimization.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Data Scientist</td>
<td>Asuragen</td>
<td>Develop pipelines and algorithms to facilitate the development and analysis of customer insights using latest data analytics and Machine learning technologies/techniques.</td>
</tr>
</tbody>
</table>

Note: DRAUP’s Talent Simulation Module. We have analysed ~2,000 tech companies and ~10,000 start-ups.

~ 70 Total Number of AI/Big Data Start-ups

~$90K Median Salary

Tech companies Start-ups

Data Scientist
- Develop pipelines and algorithms to facilitate the development and analysis of customer insights using latest data analytics and Machine learning technologies/techniques.

Data Architect
- Build products using latest Machine learning/Artificial Intelligence Models, Data clearing and development and administration of databases.

Sr Data Architect
- Development of financial services products using MS Azure, SQL Server, Hadoop and other data analytics techniques

Machine Learning Engineer
- Key Skills required include NLP, Advanced Big Data Analysis, Reinforcement Learning, Bayesian Machine Learning, Python, Matlab, SQL

Big Data Engineer
- Workloads include development of data models using tools such as Hadoop, SQL for Search Engine Optimization.

Database Engineer
- Key workloads include Performance Tuning and Query Optimization, DEVOps/Chatops (Slack) global collaboration, ETL Pipeline Design and Implementation.

Data Scientist
- Research and experiment on a variety of machine learning models for use in customer engagement and search/sort analysis.

Database Architect
- Build products using latest Machine learning/Artificial Intelligence Models, Data clearing and development and administration of databases.

Sr Data Architect
- Development of financial services products using MS Azure, SQL Server, Hadoop and other data analytics techniques

Machine Learning Engineer
- Key Skills required include NLP, Advanced Big Data Analysis, Reinforcement Learning, Bayesian Machine Learning, Python, Matlab, SQL

Big Data Engineer
- Workloads include development of data models using tools such as Hadoop, SQL for Search Engine Optimization.

Database Engineer
- Key workloads include Performance Tuning and Query Optimization, DEVOps/Chatops (Slack) global collaboration, ETL Pipeline Design and Implementation.

Data Scientist
- Research and experiment on a variety of machine learning models for use in customer engagement and search/sort analysis.