Oracle's DataScience.com acquisition fills key gap with managing data science teams
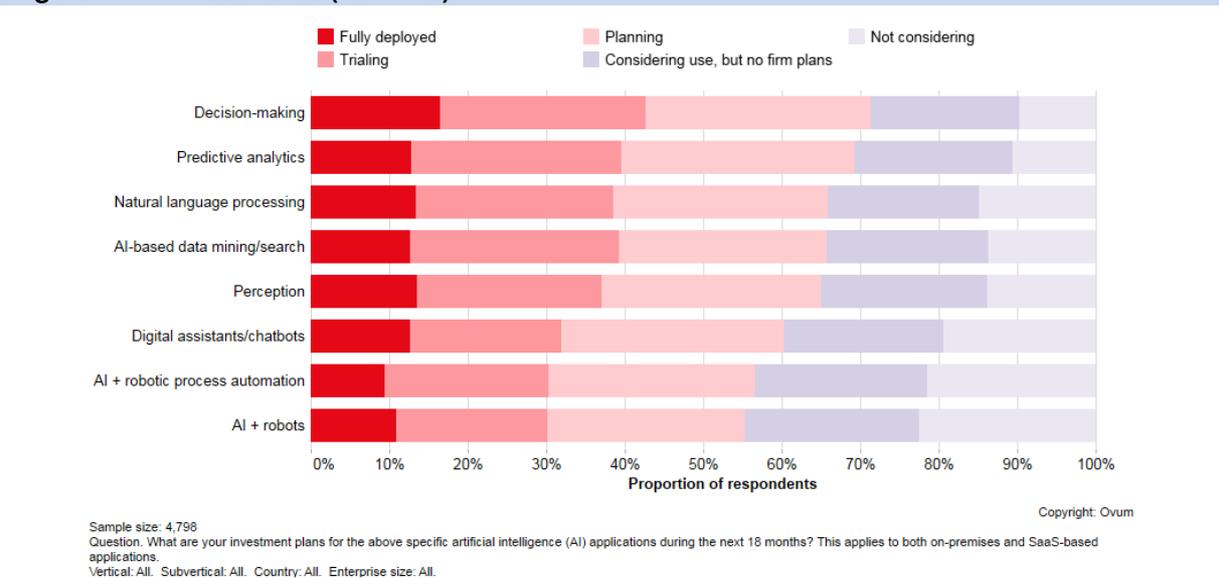
Ovum view

Summary

Oracle has announced its intention to acquire DataScience.com for an undisclosed amount in a deal that targets a key gap in Oracle's analytics portfolio. For managing data science collaboration, DataScience.com has the piece that Oracle until now has lacked: providing a platform where teams can check in projects, develop code, and manage collaboration and deployment throughout the full lifecycle. This is a market where Oracle's primary rivals, including IBM and all major cloud platform providers (not to mention a growing ecosystem of startups), have already staked their claims. Oracle's move comes not a moment too soon, given the high priority that enterprises are already putting on artificial intelligence and data science. Oracle has hardly been absent with AI, but until now, the brunt of its investments have been in smart applications and autonomous cloud PaaS infrastructure. This acquisition fits neatly into Oracle's platform-as-a-service portfolio, with the devil being in the details of when (and how) it will integrate the new DataScience.com offerings into the Oracle Analytics and Big Data clouds.

Figure 1: AI investments (2017–18)

Sample size: 4,798
Question. What are your investment plans for the above specific artificial intelligence (AI) applications during the next 18 months? This applies to both on-premises and SaaS-based applications.
Vertical: All. Subvertical: All. Country: All. Enterprise size: All.

Source: Ovum

AI matters, and data science is the key prerequisite

Ovum's enterprise insights data illustrates the extent that AI has moved to the front burner. In our latest global findings, well over half of enterprises surveyed are at some phase of AI investment (see Figure 1).

Given all the hype, it is natural to conflate data science and AI. Roughly five years ago, a well-cited article in *Harvard Business Review* termed data scientist "the sexiest job of the 21st century." Five years on and the business and media spotlight has shifted to AI. The two are not synonymous, but neither are they mutually exclusive. Data science is the practice of identifying data sets and using scientific method to identify the optimal model through rigorous testing, comparison, and re-evaluation

of multiple models. AI builds on that with algorithms and frameworks that are used to construct models that allow the computer to perceive and learn from patterns to make decisions; as such, AI requires data science as a prerequisite.

AI has emerged from its long winter thanks to growing abundance of data that is essential for building and validating models; scalable, commodity infrastructure for storage, connectivity, and compute; accessibility through a growing universe of open source programming languages, frameworks, tools, and algorithms; and the cloud, which makes the necessary compute and storage resources affordable to large audiences of developers and data scientists (and organizations) who would not otherwise have ready access to large compute clusters. It's not surprising that at big data conferences these days, the largest group of startups has something to do with AI.

Many of Oracle's rivals have already geared up for addressing demand. Examples include the following:

- IBM is already a step ahead of Oracle here with Watson Studio, itself a reworking of the Watson Data platform and Data Science Experience. Watson Studio provides a visually integrated experience that bundles formerly à la carte capabilities like Data Refinery (for data preparation) and cataloging while automating key steps of preparing and testing models.
- Cloudera offers its own Data Science Workbench geared toward bridging the notebook-to-deployment gap.
- Each of Oracle's major cloud rivals has also jumped into the fray, but mostly with point services ranging from chatbot conversational offerings to speech to text (and vice versa), image and facial recognition, language translation, and more general-purpose machine learning and deep learning services.

Figure 2: DataScience.com in the Oracle Public Cloud
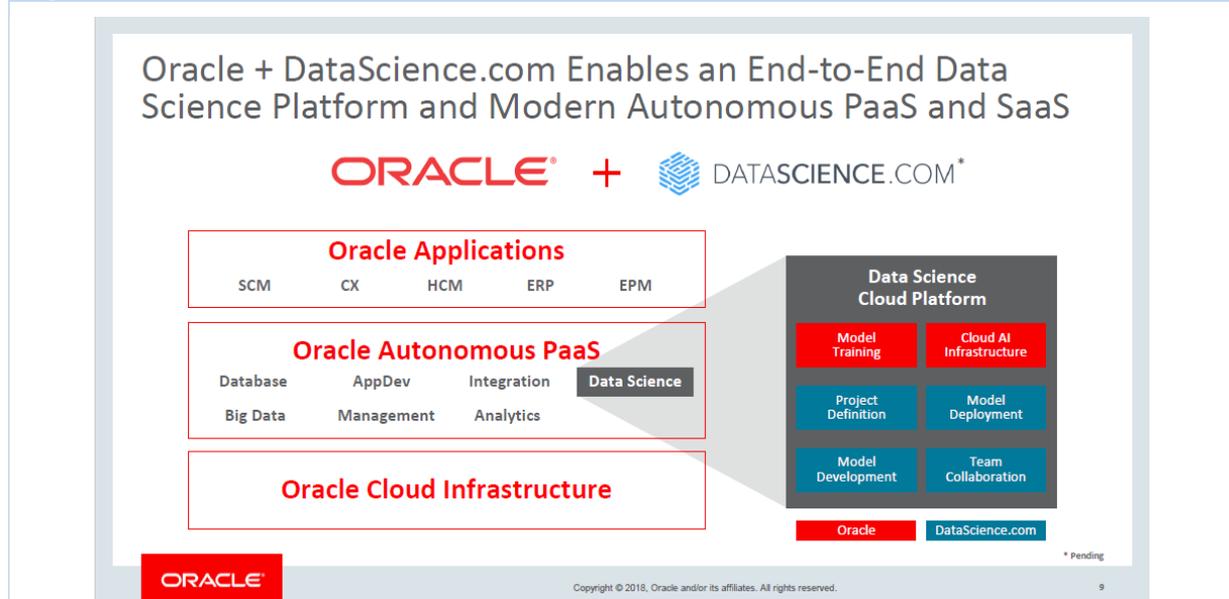
Source: Oracle

Additionally, there is a large emerging ecosystem of providers focusing on project management, team collaboration, and lifecycle management of data science and AI projects, of which DataScience.com is one. With Oracle's base as an enterprise software provider, its logical market is not necessarily with

machine learning algorithms or frameworks, where developers and data scientists have plenty of choices (and preferences) – although Oracle could, as an inducement to analytics PaaS customers, expose some of the algorithms it has developed for its own internal use. Instead, the place that Oracle needs to be is at the lifecycle management, collaboration, and deployment levels.

Where DataScience.com fills the gap

Oracle is focusing its big data analytics offerings in the cloud – both the Oracle Public Cloud and Oracle Cloud at Customer. For big data analytics and data science, the Oracle Cloud offers services including big data platforms and related services for database, business analytics, and integration (see Figure 2). But there are no explicit services for managing data science and AI projects – that's where DataScience.com comes in.

Before being acquired by Oracle, DataScience.com was one of a growing population of solution providers aimed at managing data science teams and enabling them to collaborate and integrate with the business. Compared to others in the ecosystem, such as DataRobot, Dataiku, Domino Data Lab, and H2O.ai, DataScience.com's capitalization was quite modest at $28m. Despite that capitalization, DataScience.com's ambitions were quite expansive, with its offering of a hub that requires tight integration to external data platforms such as Hadoop.

DataScience.com offers a hub that allows data scientists to use the languages, tools, notebooks, and other libraries and packages of their choice; in spite of its dot-com name, its offering was originally designed to run either on premises or in the cloud. It offers a project-based interface that consolidates all the assets (e.g., tools, code, data sets) associated with a data science project; employs Git (and integrates with all major Git providers) for version control; and provides access to team members through the project console and search utility. It supports knowledge sharing through annotations and documentation (e.g., inclusion of "Read.me" pages). Access is provided through role-based access and single sign-on. On the deployment side, DataScience.com allows teams to manage the running of models by running scripts as scheduled jobs and supporting microservices deployed through APIs. To support deployment of models, the system provides IT administrators with a dashboard to oversee production. The company also offers consulting and quick-start advisory services to help clients jump-start their data science hubs.

Given that its hub requires significant integration to multiple back-end data platforms and, in some cases, professional services to jump-start customers, DataScience.com in the long run would have required significant additional funding, a partner with deep pockets (and personnel), or an exit. So, while DataScience.com fills a key gap for Oracle, Oracle in turn could provide the resources to scale DataScience.com's ambitious offering.

How DataScience.com could fit into the Oracle Cloud PaaS portfolio

As Figure 2 illustrates, DataScience.com fills an empty space in the Oracle PaaS portfolio for which there is little if any functional duplication. While DataScience.com was originally available for both cloud and on-premises deployment, Ovum believes this type of collaboration hub is best suited to a cloud PaaS.

There are some parallels with IBM's Watson Studio, but that focuses more on the data scientist development workflow and includes capabilities for data preparation; the major area of overlap is that

both tools provide a collaboration workspace. By contrast, DataScience.com focuses exclusively on managing project workflow and deployment, leaving development to the notebooks and open source tools of choice to the data scientist.

As the deal has yet to close, Oracle was not able to detail a roadmap. But a number of touchpoints in its PaaS portfolio could benefit from integration with DataScience.com. The most obvious is Oracle's Big Data Cloud, which includes a Hadoop and Spark implementation, plus options for graph and spatial data management extensions. There should also be synergy with Oracle Data Integration Platform Cloud Service (which includes oracle Big Data Preparation), as preparing data is often a large percentage of the legwork performed in a data science project (it would make sense to embed this capability in the DataScience.com offering). Likewise, the outputs of data science or AI modeling projects could generate analytics that are surfaces in the Oracle Analytics Cloud. There is a data governance angle with DataScience.com's project lineage capability that could link to governance tools from Oracle's Data Management Cloud services.

While Oracle is on-ramping DataScience.com to its cloud portfolio, we hope that product development will not slow down. That encompasses original plans for enhancing model management capabilities for tracking A/B testing – a key function for identifying models and then monitoring them for "drift" that is inevitable over time as business challenges or data sets mutate. It has been working on Grunion, a patent-pending query optimization tool that may have synergies with Oracle Big Data SQL. And as we noted in our original analysis, a logical next step will be adding optional workflow management for project communications.

Appendix

Further reading

On the Radar: DataScience.com helps enterprises overcome the silos of data science, INT002-000067 (February 2018)

Author

Tony Baer, Principal Analyst, Information Management

tony.baer@ovum.com

Ovum Consulting

We hope that this analysis will help you make informed and imaginative business decisions. If you have further requirements, Ovum's consulting team may be able to help you. For more information about Ovum's consulting capabilities, please contact us directly at consulting@ovum.com.

Copyright notice and disclaimer

The contents of this product are protected by international copyright laws, database rights and other intellectual property rights. The owner of these rights is Informa Telecoms and Media Limited, our affiliates or other third party licensors. All product and company names and logos contained within or appearing on this product are the trademarks, service marks or trading names of their respective

CONTACT US

ovum.informa.com

askananalyst@ovum.com

INTERNATIONAL OFFICES

Beijing

Dubai

Hong Kong

Hyderabad

Johannesburg

London

Melbourne

New York

San Francisco

Sao Paulo

Tokyo