SMU
Big Data

By Shaohua

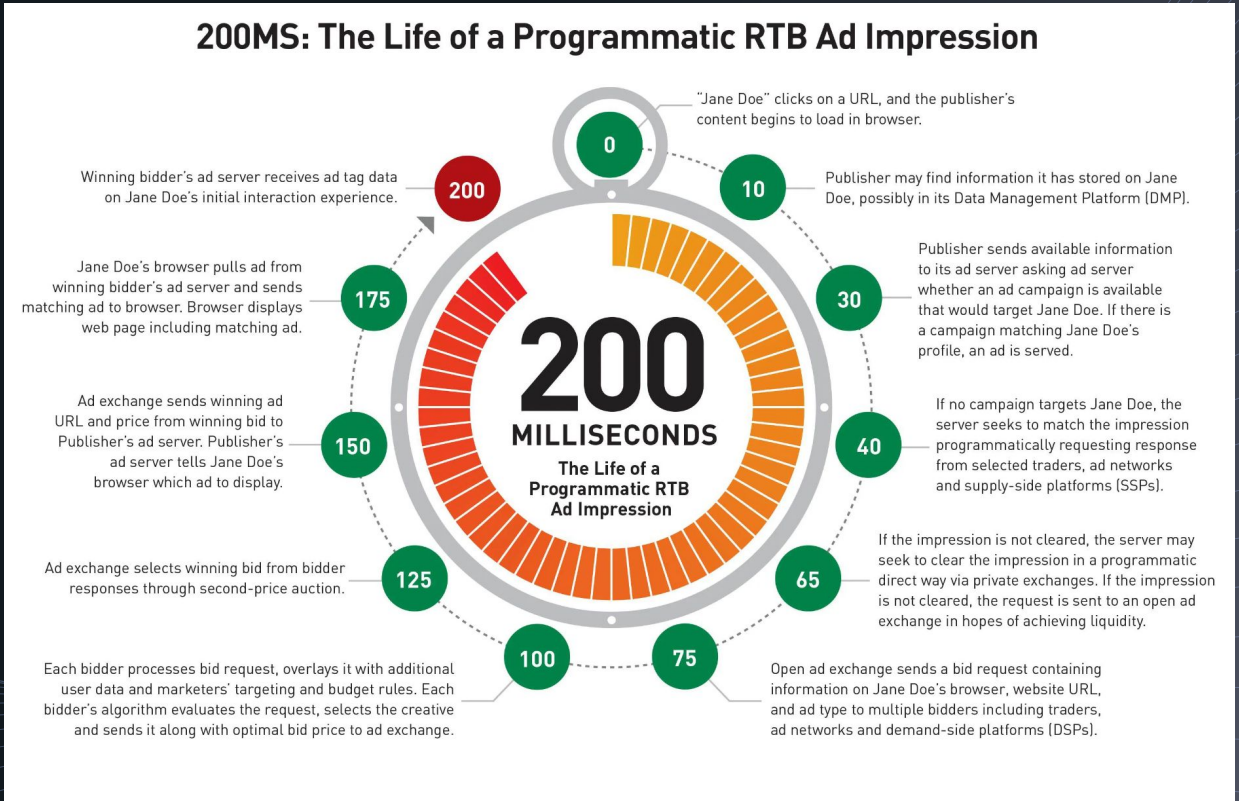
1. Programmatic Real-time bidding
2. DSP / DMP
3. Structured Vs. Unstructured Data
4. SaaS, PaaS, and IaaS
5. SaaS
6. PaaS
7. IaaS
8. AWS Benefits
9. AWS Services 1
10. AWS Services 2

What is Programmatic Direct?

Programmatic Direct is a term used to describe the process of automating a direct sale of guaranteed advertising between an advertiser and a publisher.

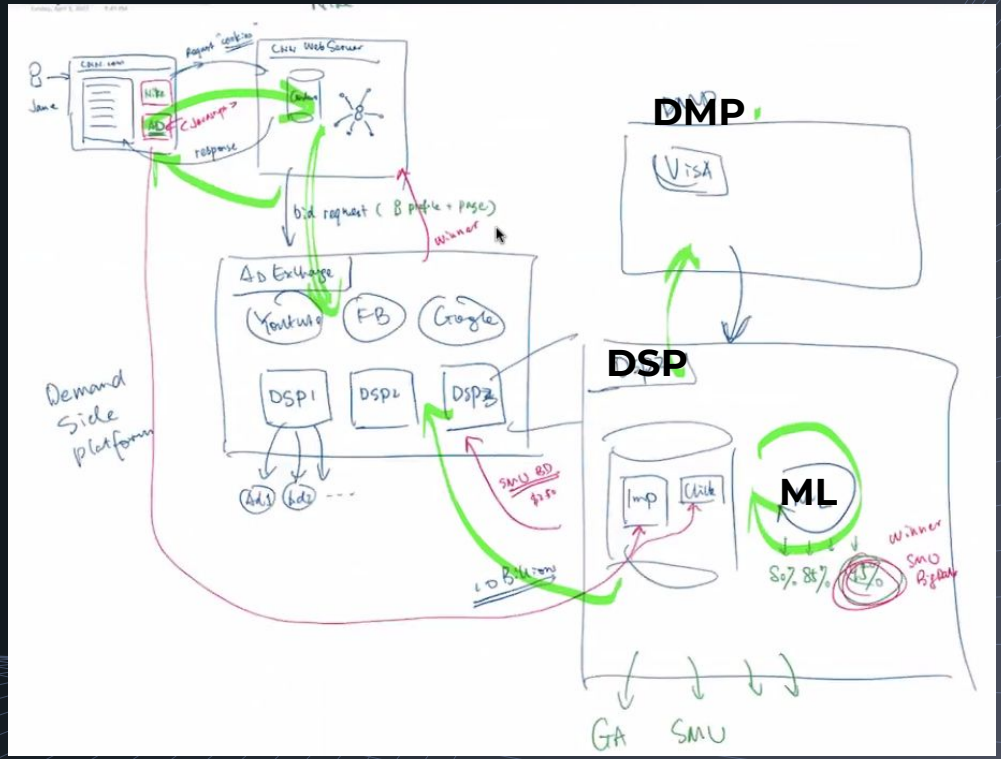
What is Real-Time Bidding (RTB)?

Real-Time Bidding is a term used to describe the buying and selling of online ad inventory that happens through automated auctions in real-time.



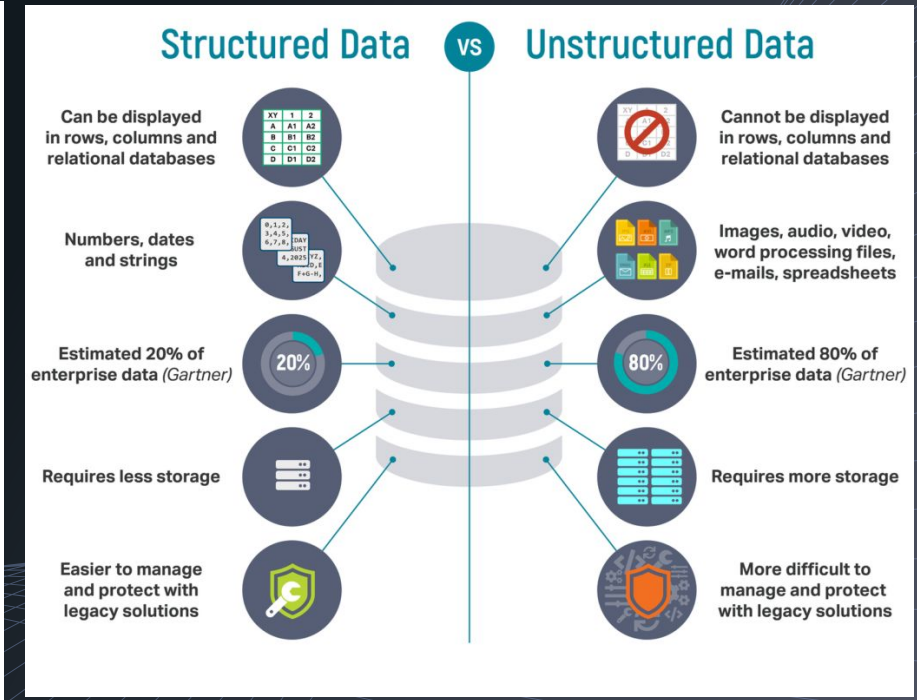
A digital signal processor (**DSP**) is a specialized microprocessor chip, with its architecture optimized for the operational needs of digital signal processing

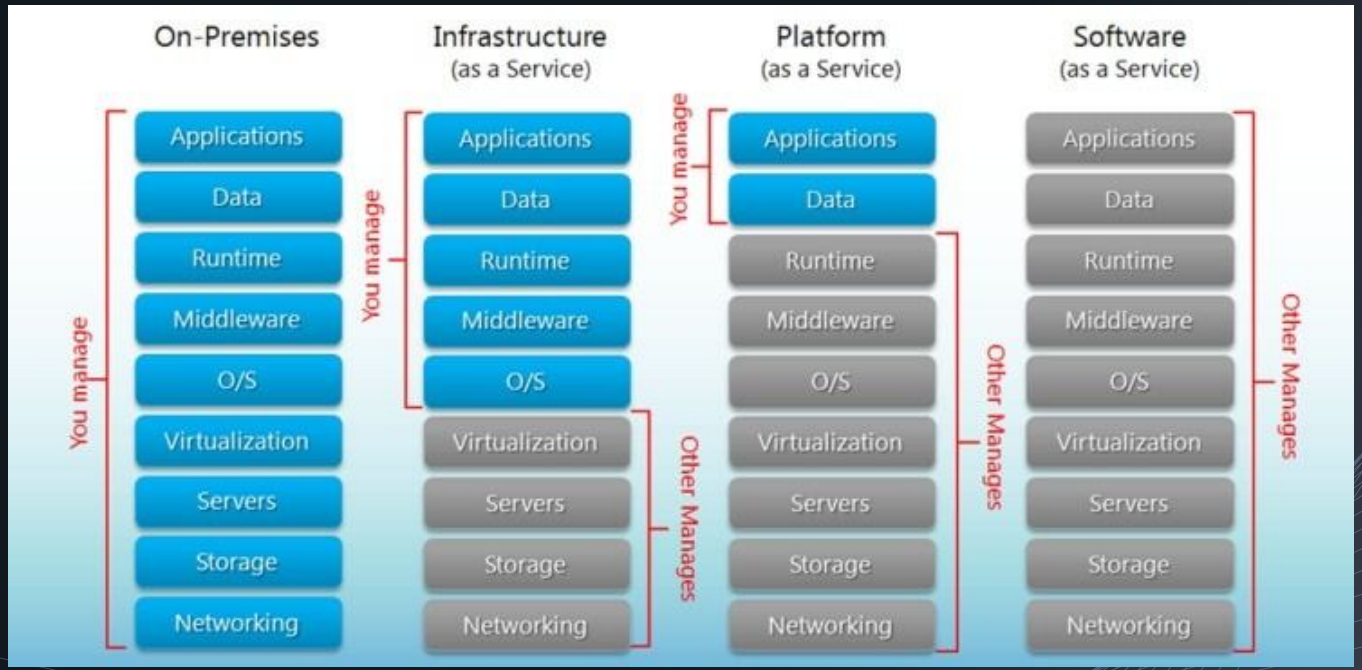
A data management platform (**DMP**) is a software platform used for collecting and managing data.



Structured Vs. Unstructured Data

	Structured Data	Unstructured Data
Characteristics	• Pre-defined data models • Usually text only • Easy to search	• No pre-defined data model • May be text, images, sound, video or other formats • Difficult to search
Resides in	• Relational databases • Data warehouses	• Applications • NoSQL databases • Data warehouses • Data lakes
Generated by	Humans or machines	Humans or machines
Typical applications	• Airline reservation systems • Inventory control • CRM systems • ERP systems	• Word processing • Presentation software • Email clients • Tools for viewing or editing media
Examples	• Dates • Phone numbers • Social security numbers • Credit card numbers • Customer names • Addresses • Product names and numbers • Transaction information	• Text files • Reports • Email messages • Audio files • Video files • Images • Surveillance imagery





IaaS: cloud-based services, pay-as-you-go for services such as storage, networking, and virtualization.

PaaS: hardware and software tools available over the internet.

SaaS: software that's available via a third-party over the internet.

On-premise: software that's installed in the same building as your business.

IaaS (Infrastructure as a Service)

IaaS businesses offer services such as pay-as-you-go storage, networking, and virtualization. IaaS gives users cloud-based alternatives to on-premise infrastructure, so businesses can avoid investing in expensive on-site resources.

IaaS platforms are: Highly flexible and highly scalable. Accessible by multiple users. Cost-effective.

When to Use IaaS: IaaS is beneficial to businesses of all shapes and sizes, as it allows complete control over your infrastructure, and operates on a pay-as-you-use model, so it fits into most budgets. With most IaaS platforms, you get access to ongoing support and have the option of scaling up your requirements at any time.

IaaS Non-Ecommerce Example: A good example of IaaS is AWS EC2. EC2 provides scalable infrastructure for companies who want to host cloud-based applications. EC2 users do not own the physical servers; AWS provides virtual servers.

IaaS Ecommerce Example: Magento 1 Enterprise Edition can be either on-premise or IaaS depending on how the merchant chooses to host their store. In the case of IaaS, the merchant is paying Magento for the licensing of the software and then using a third party vendor for the best web hosting such as Rackspace. Merchants are able to pay for a hosting plan that meets their own needs without the cost of maintaining their own physical servers. The merchant is still responsible for installing and managing updates to their Magento software.

PaaS (Platform as a Service)

A PaaS vendor provides hardware and software tools over the internet, and people use these tools to develop applications. PaaS users tend to be developers.

PaaS platforms are: Accessible by multiple users. Scalable – you can choose from various tiers of resources to suit the size of your business. Built on virtualization technology. Easy to run without extensive system administration knowledge.

When to Use PaaS: PaaS is often the most cost-effective and time-effective way for a developer to create a unique application. PaaS allows the developer to focus on the creative side of app development, as opposed to menial tasks such as managing software updates or security patches. All of their time and brainpower will go into creating, testing, and deploying the app.

PaaS Non-Ecommerce Example: A good example of PaaS is AWS Elastic Beanstalk. Amazon Web Services (AWS) offers over 100 cloud computing services such as EC2, RDS, and S3. Most of these services can be used as IaaS, and most companies who use AWS will pick and choose the services they need.

PaaS Ecommerce Example: Magento Commerce Cloud (also known as Magento Enterprise Cloud Edition) is the most common example of PaaS for ecommerce. This enables the merchant to bundle their hosting as part of their package with Magento. Merchants evaluating Magento go through a scoping process to determine their hosting needs which is then bundled into their monthly plan. Merchants still have full access to edit the source code of their Magento store and can fully customize the application.

SaaS (Software as a Service)

SaaS platforms make software available to users over the internet, usually for a monthly subscription fee.

SaaS platforms are: Available over the internet. Hosted on a remote server by a third-party provider. Scalable, with different tiers for small, medium, and enterprise-level businesses. Inclusive, offering security, compliance, and maintenance as part of the cost.

When to Use SaaS: Take your email server, for example. You want to know that you'll continue to send and receive emails without needing to fiddle with your email settings or worry about updates. Imagine if your email server went under because you forgot to update it and you went days without email? That's simply not an option in today's marketplace. If you use a SaaS platform to run your email inbox, the chances of something going wrong are very small.

SaaS Ecommerce Example: Shopify is an example of a SaaS ecommerce platform. The Shopify platform also has regular updates that automatically roll out for users, and all the software licenses, upgrades, and hosting costs are covered in the monthly subscription fee.

AWS Benefits

- **Easy to use**: AWS is designed to allow application providers, ISVs, and vendors to quickly and securely host your applications – whether an existing application or a new SaaS-based application. You can use the AWS Management Console or well-documented web services APIs to access AWS's application hosting platform.
- **Flexible**: AWS enables you to select the operating system, programming language, web application platform, database, and other services you need. With AWS, you receive a virtual environment that lets you load the software and services your application requires.
- **Cost-Effective**: You pay only for the compute power, storage, and other resources you use, with no long-term contracts or up-front commitments.
- **Reliable**: With AWS, you take advantage of a scalable, reliable, and secure global computing infrastructure, the virtual backbone of Amazon.com's multi-billion dollar online business that has been honed for over a decade.
- **Scalable and high-performance**: Using AWS tools, Auto Scaling, and Elastic Load Balancing, your application can scale up or down based on demand. Backed by Amazon's massive infrastructure, you have access to compute and storage resources when you need them.
- **Secure**: AWS utilizes an end-to-end approach to secure and harden our infrastructure, including physical, operational, and software measures. For more information, see the AWS Security Center.

AWS Services 1

- **EC2**: Provides secure, resizable compute capacity in the cloud. It makes web-scale cloud computing easier for developers.
- **Lightsail:** easiest way to launch & manage a virtual private server with AWS. An easy-to-use cloud platform that offers everything need to build an application or website.
- **ECS**: Highly secure, reliable, & scalable way to run containers.
- **ECR**: Easily store, manage, & deploy container images
- **Lambda**: Run code without thinking about servers. Pay only for the compute time you consume..
- **S3**: is the storehouse for the internet i.e. object storage built to store & retrieve any amount of data from anywhere.
- **Glacier**: S3 Glacier and S3 Glacier Deep Archive are a secure, durable, & extremely low cost S3 cloud storage classes for data archiving & long-term backup.
- **Transfer Family**: The Transfer Family provides fully managed support for file transfers directly into & out of S3.
- **Snow Family**: Highly-secure, portable devices to collect & process data at the edge, and migrate data into and out of AWS.
- **RDS**: is a web service that makes it easier to set up, control and scale a relational database in the cloud.
- **Redshift**: The most popular and fastest cloud data warehouse

AWS Services 2

- **Kinesis:** Kinesis makes it easy to collect, process, & analyze real-time, streaming data so one can get timely insights.
- **EMR**: is the industry leading cloud big data platform for processing vast amounts of data using open source tools such as Apache Spark, Hive, HBase, Flink, Hudi, & Presto
- **Athena**: is an interactive query service that makes it easy to analyze data in S3 using standard SQL.
- **SageMaker**: is a fully managed service that provides every developer & data scientist with the ability to build, train, & deploy machine learning (ML) models quickly.
- **QuickSight**: is a ast, cloud--powered business intelligence service that makes it easy to deliver insights to everyone in organization.
- **Glue:** is a serverless data integration service that makes it easy to discover, prepare, and combine data for analytics, machine learning, and application development. AWS Glue provides all the capabilities needed for data integration so that you can start analyzing your data and putting it to use in minutes instead of months.