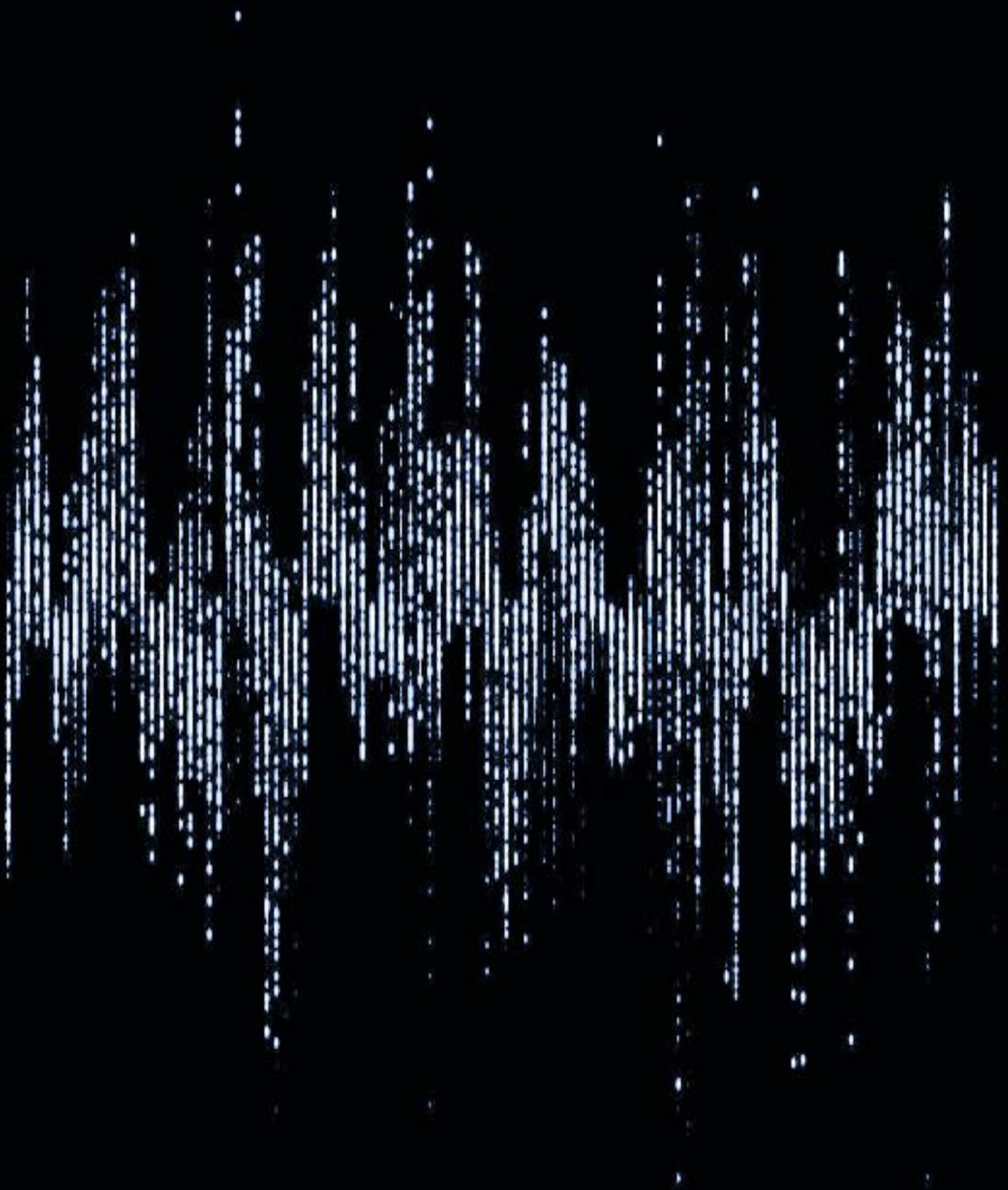# Enterprise-Scale Automatic Speech Recognition, SaaS versus On-Premise Deployment

Abstract

This whitepaper compares and contrasts the characteristics of Automatic Speech Recognition (ASR) delivered as a Cloud-Hosted (SaaS) offering, with ASR capabilities delivered through on-premise licensing. Both approaches have merit and are being widely adopted, both also have unique characteristics and limitations, which may mean one approach is better suited to addressing your needs than the other.

The topics investigated include - whether or not you have a choice, whether changes in privacy, compliance and security regulation plays a role? if requirements for real-time processing (vs. batch) is relevant? And what to consider when calculating the Total Cost of Ownership.

Do you have the technical IT infrastructure in place?

If you're developing a solution which includes an ASR capability (transcription), then deciding whether to use a SaaS ASR offering or licensing an ASR capability into your solution "on-premise" is a key decision.

The first question to ask is if you have the technical infrastructure in place (or plan to have in place) to deliver the end-to-end solution. This not only includes the ASR solution, but the surrounding infrastructure and organization, acquiring or producing the audio streams or recordings, performing the speech recognition and transcriptions, managing and storing the content while it is being processed or archived.

If you do not have this infrastructure in place or have no plans to put it in place, then it's very likely that a SaaS ASR offering is best suited for your needs. Typically, ASR SaaS offerings are better suited to smaller companies and startups that do not have this infrastructure. With that infrastructure in place and security ascertained, then licensing on-premise ASR solution may be the best fit.

KEY CONSIDERATIONS
● Do you have the IT infrastructure in place to support the end-to-end solution which incorporates an ASR capability?

Do you have the organizational infrastructure in place to support regulatory compliance requirements such as GDPR, CCPA, HIPAA, PCI, FISMA, SOC2 Type II ?

In an increasingly security conscious world, with new emerging regulations playing an ever-important role in protecting individual's privacy, any business managing and operating an ASR service, needs to be fully aware and willing to disclose, the implications of processing and storing such confidential and sensitive data through their business. This is particularly true of specific industries such as Financial

Services, Healthcare and Legal where you will need to determine which emerging standards are likely to affect you most.

The ASR provider must offer full transparency to their customers, including solution developers, providing documented visibility into the organizational processes, policies and procedures for securing such data for today and for supporting emerging regulations in the future.

Ignorance and not understanding how the customer data your business is entrusted with is managed, is not an excuse. Your business may be non-compliant with much of the pending regulation as a result of how you currently engage your third-party ASR hosting provider today and how in turn they manage the sensitive data you provide them in using their ASR service. Worse still you may be breaking the law. While many of the larger Cloud-hosted ASR (SaaS) service offerings are striving to adhere to these new regulations and assure a level of security and compliance, some new entrant offerings may not. Regardless of whether or not the ASR service is offered by an established brand, additional complexity is introduced into the operational compliance process. This can be a considerable cost to the overall business in delivering a new service offering.

For on-premise or in-house developed/deployed ASR services, the compliance concerns will typically fall under the same general business IT concerns, as seen with ERP, CRM etc. (vs. dealing with a third party for voice only), which reduces the cost in ensuring overall operational compliance and simplifies the effort in bringing the solution offering to market.

Ultimately, based on support and disclosure by your ASR vendor, you will need to decide if you're comfortable with your customer's sensitive data (and meta data) residing in your ASR vendors data center or if you believe that is simply too great a risk and data you need to bring in-house and securely manage as part of your business.

KEY CONSIDERATIONS
● Will the Solution you're developing be used by companies who will need to comply with emerging security regulations such as GDPR, HIPAA, CCPA, PCI, FISMA or SOC2 Type II?   ● Does your company have a compliance program, including written policies, procedures and regular training and report up to an Executive or senior leader?   ● How does your organization perform risk assessments, security monitoring, and third-party audits?

Do you need transcripts in real-time?

Delivering an accurate real-time transcript from a Cloud-Hosted ASR (SaaS) offering is inherently a challenging technical problem. From a speaker on a phone, to an ASR engine in a remote data center processing the audio stream and at the same time delivering a transcript, there are a number of variables and risks that could impact the Quality of Service (Real-time). This is one reason why many of the ASR service offerings in the market, through configurable parameters, enable you to trade off accuracy for speed, to terminate the speech recognition process (accuracy) in lieu of sacrificing the Quality of Service.

One of the first questions to ask is whether or not the requirements dictate real-time transcription? is it absolutely "real-time" or is "near real-time" good enough for your needs, where rapid batch processing may suffice.

Following this is the question as to which type of API your ASR solution vendor supports. If it's a RESTful based API (based on the HTTP protocol), then "real-time" performance due to the nature of the technology and protocol, is at best going to be a goal.

If on the other hand, the API is WebSockets based, a communications protocol providing full-duplex communication channels over a single TCP connection or gRPC based, providing bidirectional streaming and flow control, then real-time is achievable, albeit still a challenge and one where you should ensure to understand the Quality of Service guarantees if any, before making a decision.

The bottom line, many of these risks can be reduced or avoided with an on-prem ASR solution. Having the ASR capability local to the media stream and native to the solution, reduces the risk with ensuring Quality of Service.

KEY CONSIDERATIONS
● Does your solution need to offer real-time transcription or is "near" real-time (low latency batch) good enough?   ● Where does the speech you plan to transcribe originate? Is it from live phone conversations or archived recordings?

How about a Hybrid approach?

For ASR, a hybrid approach leveraging both Cloud Hosted assets (e.g. non-real-time functionality, such as analytics or reporting / dashboards) along with on-premise-based assets (e.g. speech recognition and production of transcripts) may be worth considering and help reduce risk, but it will depend very much on your specific use case.

Other hybrid approaches may also be worth considering, such as a Cloud Hosted ASR service acting as a backup or hot-standby for an on-prem ASR service should it go down or become unavailable or vice versa, although redundancy such as this will drive up cost.

Understanding from your ASR vendor whether they support hybrid models today or have planned in the future is important, as is understanding how they partition and offer the functionality and if it's relevant for you, in addressing your needs.

KEY CONSIDERATIONS
● Does your solution lend itself to being partitioned across Cloud and non-Cloud infrastructure?   ● Do you have a need and a backup plan should your main ASR provider (external or in-house) go down?

Total Cost of Ownership

Calculating the Total Cost of Ownership (TCO) for your ASR solution will depend on the approach you have taken, whether that's with a SaaS service offering or on-premise licensing. Deciding on either option largely depends on the state of your company and the direction you want your organization to take in the future, whether you have the IT infrastructure in place or not, whether new regulations around management of sensitive personal data is a concern now or in the future for your business.

While the "pay for what you use" model the majority of ASR SaaS vendors support is attractive for small businesses, startups or companies with minimal transcription needs. With sufficient transcription volume, the break-even as compared with on-premise licensing can be surprising, with the benefits that you have greater control of the solution and less overhead and risk when it comes to management of sensitive data that falls under the prevue of regulation.

KEY CONSIDERATIONS
● Determine the best approach for your business needs, be that on-premise licensing or SaaS ASR.
● For on-premise licensing, include the cost of IT infrastructure and staffing to operate and manage in your calculations.
● For on-premise licensing ensure to ask you vendor what commercial options they support beyond just charging based on volume of data transcribed e.g. processor or capacity-based licensing.
● What other initiatives does your company have underway which in the future may also have a need for ASR.