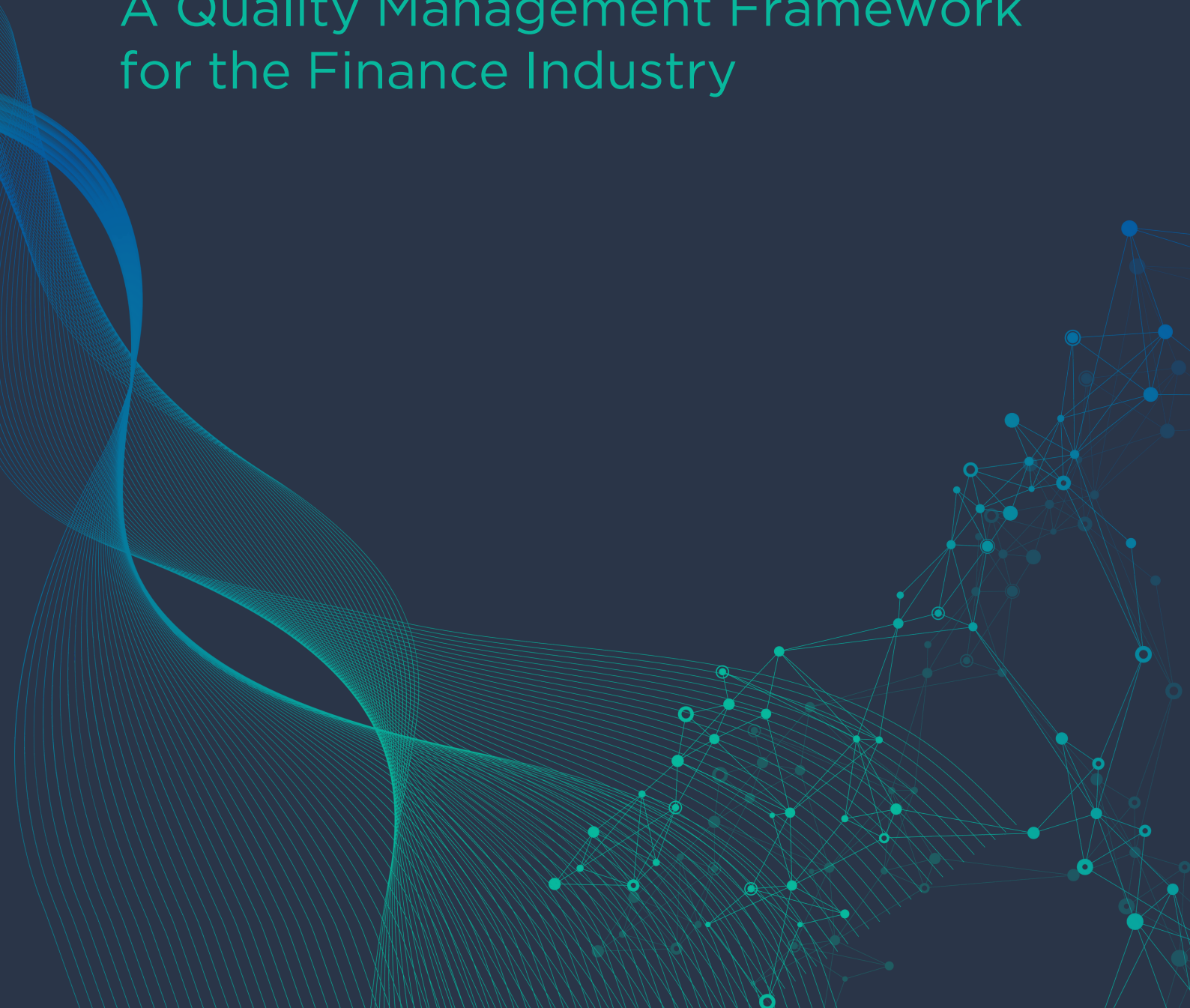
WHITE PAPER

cloudfactory

Outsourced Data Processing

A Quality Management Framework for the Finance Industry



Who this paper is For

This paper is for you if…

You're the director of operations, chief operating officer, vice president of operations, head of operations, operations manager, or in any other operational role in a financial technology (fintech) or services (finserv) company.

It's for you if you're charged with finding **better, faster, lower-cost ways of working.**

It's for you if you're a visionary leader looking to **maintain the pace of innovation** or **personalize the customer experience** to remain competitive.

It's for you if you're seeking a greater focus on agility and efficiency to **optimize your business operations.**

This paper is for you if you're an operational leader looking for a long-term partner to serve as an extension of your team, acting as the **behind-the-scenes machine powering your data processing operations** in an agile, scalable way.

Most of all, **if you care about quality data processing**, this paper is for you.

cloudfactory

Table of Contents

Figures, Tables, and Features

Figures and Tables

Featured Client Profiles

Introduction

> *Quality is not just a goal, process, or deliverable. Instead, it's a mindset we hold while serving our clients.*
>
> *-- CloudFactory Handbook of Quality Management*

Let's talk quality.

What does quality mean to you?

In the world of the **operations leaders** we work with, a focus on quality runs deep. Perhaps to you, quality means efficiency—or maximizing value right down to the source. Or maybe it means creating effective operational practices that set you apart from your competition. Or perhaps to you, quality means orchestrating the inflow and outflow of resources in win-win-win ways.

In the world of the **operational data leaders** we work with, quality usually boils down to a set of dimensions, some of which will match those set forth by researchers Richard Wang and Diane Strong in their seminal paper, "Beyond Accuracy: What Data Quality Means to Data Consumers."[1]

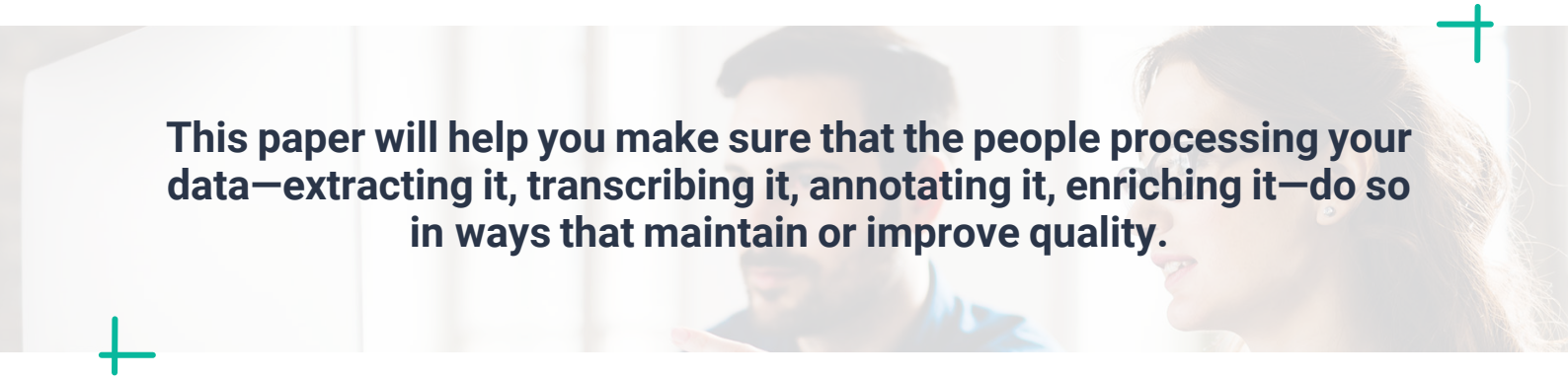
Wang and Strong's 15 Dimensions of Data Quality

- Accuracy
- Relevancy
- Value-added
- Accessibility
- Ease of understanding
- Completeness
- Timeliness
- Appropriate amount
- Reputation
- Representational consistency
- Objectivity
- Believability
- Access security
- Interpretability
- Concise representation

For companies in the financial technology and financial services industries, **quality often equates to accuracy** because, when it comes to numbers, the accuracy of those trailing decimal points matters.

[1] https://dl.acm.org/doi/10.1080/07421222.1996.11518099

The Data Management Body of Knowledge, an ever-evolving, definitive collection of data management principles and practices, provides another perspective on **data quality**. It says that data quality refers to whether data is **fit for consumption and meets the needs of data consumers[2].**
Only you know the needs of your consumers. Only you and your organization can determine what fit for consumption looks like.



This paper will help you make sure that the people processing your data—extracting it, transcribing it, annotating it, enriching it—do so in ways that maintain or improve quality.

This paper will also help you answer the questions that will inevitably arise as you consider outsourcing your data and document processing, and even your customer service and other back-office tasks:

- *How do I frame discussions about quality with prospective outsourcing firms?*
- *How do I balance the need for quality with variables like throughput and cost?*
- *How do I know that the people doing the work care about the quality of our data?*

Since 2010, CloudFactory has partnered closely with visionary companies in the fintech and financial services spaces to accelerate AI initiatives and optimize business processes through managed, human-powered workforces.

Whether we're indexing receipts to create an expense report, enriching firmographic data, or serving as your 24/7/365 front-line support team, our workforce becomes the machine that powers your business, making possible fantastic customer experiences and competitive advantages.

After processing hundreds of millions of documents for hundreds of clients, we've learned a thing or two about what it takes to recruit, train, manage, and sustain a workforce that acts as an extension of your team.

We've also learned what it takes to transform unstructured data into quality, valuable assets that fuel business growth. In this paper, we share our quality management framework for doing just that.

The framework encompasses:

1. Quality Definitions and Documentation.
2. Quality Assurance.
3. Quality Control.
4. Continuous Improvement.

We'll also share several client profiles and a few other resources to help you on your journey, including a list of questions to ask prospective outsourcing partners.

Let's get started.

[2] https://www.dama.org/cpages/body-of-knowledge

CloudFactory's 4-Part Quality Management Framework

Quality Management Framework Part 1

Quality Definitions and Documentation

In the definitions and documentation phase of implementing your quality framework, you'll set forth objectives, expected tasks, guiding rules and instructions, and metrics. This phase also provides learning time for your data analysts, who have opportunities to practice the work, build confidence, and become experts at the tasks at hand.

To ensure accuracy, a key output of this phase is a **quality management plan**, which defines the process and activities to come.

A typical quality management plan captures information about:

- The workstream and tasks involved in it.
- Standard operating procedures for completing the tasks.
- Instructions and training for analysts, if applicable.
- Quality assurance objectives, protocols, workflow model, reporting hierarchies, and where they fit into the process (more on quality assurance soon).
- Accuracy measures, including calculations for metrics.
- Quality control objectives, protocols, sampling rate (more on quality control later).
- Error types likely to occur and corrective actions for each.
- Roles and responsibilities for the quality assurance and control activities.
- Tools for measuring quality, including reports and dashboards.
- Lessons learned—captured as the work progresses.
- Additional resources and documentation, usually in an appendix.

Client Profile:
Expense Management

Experiencing double-digit growth, Expensify needed a reliable partner to process large volumes of receipts according to complex business rules. Today, Expensify customers upload photos of receipts through the company's app, and CloudFactory analysts process that data in near real-time.

- A team of 1,000+ data analysts processes more than three million receipts each month.
- **Expensify customers can access their data 80% faster.**
- Expensify leadership can focus on innovation rather than back-office operations.

> *CloudFactory helped us scale with accuracy and speed during this time of dramatic growth!*
>
> *~Matt McNamara, Director of Engineering*

Learn more about our work with Expensify.

Expensify

Quality Management Framework Part 2

Quality Assurance

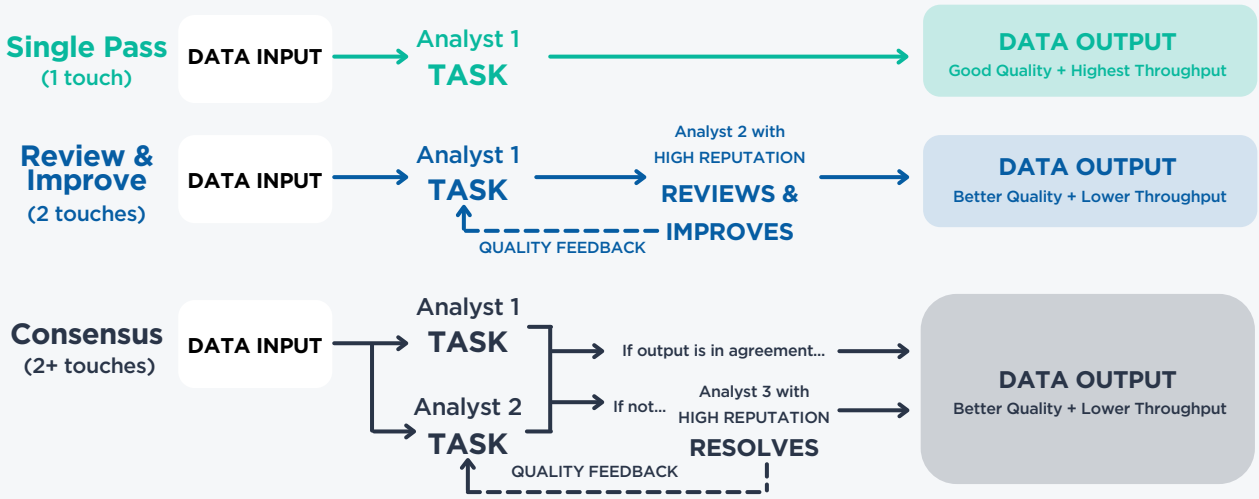
According to the American Society for Quality, **the goal of quality assurance is confidence**—internal confidence for management, and external confidence for clients, regulators, and other stakeholders that quality requirements have been fulfilled.[3]

To fulfill those quality requirements, you need processes for delivering quality outputs and pre-planned actions for evaluating and correcting quality issues.

For our clients in the finserv and fintech spaces, we rely primarily on three quality assurance workflow models: Single Pass, Review and Improve, and Consensus.

The following graphic illustrates each method at a high level.

3 Common Quality Assurance Workflows



Next, we'll look more closely at each of the workflows.

[3] https://asq.org/quality-resources/quality-assurance-vs-control

The Single Pass QA Method

The Single Pass method, which involves no formal quality assurance at all, maximizes throughput by having a data analyst complete each task once; all data outputs enter directly into your system or process.

Single Pass is best for simple, objective use cases where subject matter experts or experienced data analysts complete the tasks. Companies also use the Single Pass approach when they need a large quantity of data processed irrespective of quality.

Example financial services use cases for the Single Pass method

- **Data classification:** *Is it a receipt or an invoice? Is the amount positive or negative?*
- **Data verification:** *Is a receipt or invoice available or not? Is the URL valid or not?*
- **Data enrichment:** *Given the company name, what is the URL?*

A note on quality, edge cases, and OCR technology

Companies across the financial services and financial technology spaces use optical character recognition (OCR) to process and extract data from a wide range of images and documents:

- Bank statements
- Mortgage statements
- Insurance dec pages
- Powers of attorney
- Court orders
- Trust certifications
- Paystubs
- Business cards
- Photo IDs
- Property tax bills
- Pension award letters
- Insurance policies
- Tax returns
- Invoices
- Receipts
- Divorce decrees
- Lease agreements
- Purchase orders

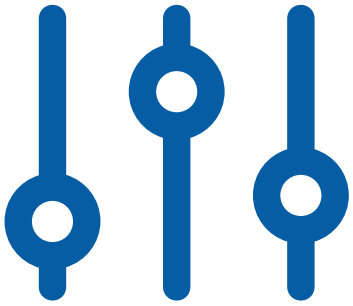
Most of our finserv and fintech clients use OCR to optimize operational efficiency. But given that OCR technology is notoriously prone to error, the human-powered work we do for those clients is to provide quality assurance and quality control on OCR outputs and handle edge cases that do not fit the technology's criteria.

The Review and Improve QA Method

In the Review and Improve QA method, a senior data analyst or subject matter expert reviews the outputs of junior analysts before those outputs enter your system or process. Accurate outputs pass through. Inaccurate outputs flow to the reviewer for correction or back to the junior analyst with feedback. The junior analyst then corrects their errors and resubmits the data for another round of review. Iterating in this way can lead to any number of Review-Improve cycles for a single task or data output.

The Review and Improve method leads to higher accuracy for most use cases. It's appropriate for tasks that involve some ambiguity and subjectivity and when data quality is paramount. Review and Improve is also the method we recommend our clients consider first.

The Review and Improve approach gives you multiple levers to pull, including time, throughput, and accuracy. Because two analysts must touch the data at least once before a task is complete, throughput is lower, and accuracy is higher. Depending on the task complexity and type, the QA portion of the work can take more, less, or the same time as the initial task.



Fine tune levers like time, throughput, and quality.

You can also limit how much work to review. According to our client delivery experts, reviewing 10% of work is common, especially as an analyst group improves. It's also common to review the work of new junior analysts as they get up to speed.

Generally, expect QA to take less time than the initial task. For data enrichment, web scraping, and NLP use cases, expect QA time closer to the initial task time. And for more straightforward tasks, such as data verification or image transcription, expect QA time to be less than the initial task time.

Also, note that as the number of cycles of Review and Improvement increases, throughput will decrease, and quality will increase.

A note on quality assurance and performance APIs

CloudFactory clients gain access to our secure, cloud-native Workforce Management Platform, powered by AWS. The platform makes it possible for analysts to process unstructured data with high accuracy, speed, and reliability using virtually any data tool and for clients to control workforce capacity, collaborate with analysts, and gain complete visibility into workforce performance.

To help us monitor analyst performance, CloudFactory provides a ThroughPut API, which captures how many tasks a given analyst performs in a workstream over time, and a Quality API, which captures how accurately an analyst performs a task over time.

Example financial services use cases for the Review and Improve method

- **Receipt transcription:** *From any receipt, transcribe the store name, amount, credit card details, date, and store location.*
- **Data enrichment:** *Complete the location-based matrix by determining whether a business belongs to a specific site.*
- **Image transcription:** *From scanned images of invoices, extract data for given fields.*

Client Profile:
AP & AR Automation

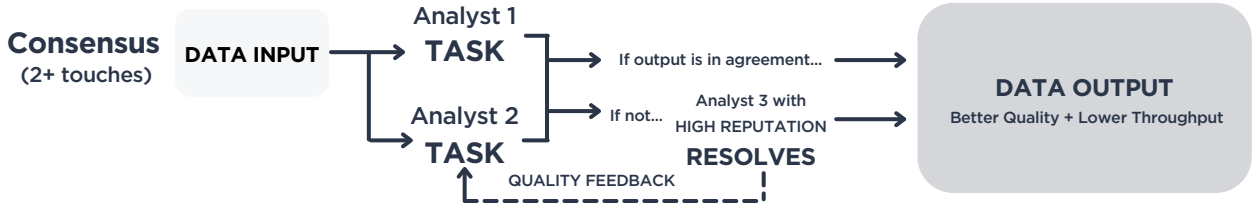
Bill.com is a financial technology company that offers consumers the tools they need to manage personal finance activities like processing invoices and paying bills.

Since 2015, CloudFactory's data analysts have been at work behind the scenes, cleaning and entering invoice data into Bill.com's platform. And to avoid any backlogs, we also maintain the resources and support to provide coverage during seasonal volume fluctuations.

Our partnership has resulted in an 85% reduction in processing time and accuracy rates of 98% or better, exceeding Bill.com's expectations.

bill.com

The Consensus QA Method



In the Consensus method, two or more data analysts complete the same task, and the final result flows into your system or process only after the analysts reach consensus. Discrepancies are placed back into the work pool and resolved by a subject matter expert or senior data analyst.

This "wisdom of the crowd" approach, or inter-analyst/inter-annotator agreement, ensures higher accuracy as compared to the accuracy achieved by a single analyst. You can use this agreement to:

- Identify the least and most reliable analysts in a pool.
- Calculate an analyst's consistency over time.
- Create reference standards, sometimes called gold standards, benchmarks, or ground-truth answers.
- Escalate complex or confusing tasks.

Although multiple variations of the Consensus QA method exist, all involve the same input being processed more than once by two or more analysts.

Example financial services use cases for the Consensus method

- **Invoice transcription:** *Split, sort, and transcribe data for a niche invoice automation platform.*
- **Data categorization:** *Extract and categorize a set of information from varying document types.*
- **Sentiment analysis:** *Identify whether a sentiment is neutral, positive, or negative.*

Comparison: The Single Pass, Review and Improve, and Consensus QA Methods

	Single Pass	Review and Improve	Consensus
Summary	A fast workflow with low accuracy; best for simple use cases and when high throughput is needed.	Flexible enough for most use cases; provides the benefits of review, expert feedback, and analyst improvement over time.	Best when a ground-truth answer exists, and accuracy is paramount.
Throughput	●●●	●●○	●○○
Accuracy	●○○	●●○	●●●
Best use cases	• Transcription. • Simple tasks. • Objective tasks; analyst subjectivity is not a factor.	• Annotation. • Segmentation. • Accuracy is a priority. • Subjective tasks; some analyst subjectivity is at play.	• Natural language processing. • High accuracy is a requirement. • Highly subjective tasks; analyst subjectivity is a major factor.

Quality Management Framework Part 3

Quality Control

Quality assurance focuses on the processes that ensure data quality in real-time—complete a task, check it, complete another task, check it. In contrast, quality control looks forward to predicting where quality may be headed and, more importantly, to what you can do to steer it in a positive direction.

Think of quality control like the alert in your car warning you that your brakes may be failing; repeated measurements show that the discs are getting thinner and thinner, which raises an alert to let you know action is needed.

Quality control involves the scientific method of sampling a statistically significant number of processed outputs. This post from Qualtrics—Calculating Sample Size: a Quick Guide[4] —provides an overview of the approach along with a sample size calculator to experiment with.

A simple example of quality control for a hypothetical CloudFactory client follows.

Example: Applying Quality Control to a Data Processing Workstream

Given these facts:

- Task population size: 10,000 tasks
- Confidence level: 95%
- Margin of error: 5%

...the sampling calculator indicates that we must sample a minimum of 370 tasks.

If we have four analysts on the job, and if our client requires that we sample 5% of each analyst's tasks, we will sample 500 tasks in all, meeting all requirements.

The following table shows how the per-analyst breakdown plays out.

	# Completed Tasks	% Sample Required	# Tasks Sampled
Analyst 1	5,000	5%	250
Analyst 2	2,000	5%	100
Analyst 3	2,000	5%	100
Analyst 4	1,000	5%	50
	= 10,000 tasks		= 500 tasks

[4] https://www.qualtrics.com/blog/calculating-sample-size/

A note on quality control and performance APIs

CloudFactory clients may use our QC Task Output API to submit task outputs to our Workforce Management Platform. The API provides a continuous measurement of data accuracy, complete visibility into accuracy information, timely feedback to data analysts so they can learn from their mistakes and improve, and a mechanism for tying feedback to performance metrics and a reward system.

In terms of your data, QC sampling can alert you to slight deviations in quality so you can address them before they become large enough to affect the accuracy of future data processing work. It can also allow you to provide further training to analysts whose accuracy fails to meet your requirements.

CloudFactory clients may use our QC Task Output API to submit task outputs to our Workforce Management Platform. The API provides a continuous measurement of data accuracy, complete visibility into accuracy information, timely feedback to data analysts so they can learn from their mistakes and improve, and a mechanism for tying feedback to performance metrics and a reward system.

The following table compares quality assurance and quality control and provides additional details you might find helpful.

Comparison: Quality Assurance and Quality Control

	Quality Assurance	Quality Control
Method	QA is a method of defect prevention.	QC is a method of defect detection.
Timing	QA is performed before data enters your system or process.	QC is performed on a sample of processed data.
Purpose	Improve the processes you use to assure accuracy. Define actions for evaluating and correcting data quality.	Evaluate the accuracy of previously processed data. Ensure that data deliverables continue to meet your requirements.
Data analyst	Improve the analyst's skills and adherence to standards.	Evaluate an analyst's performance. Improve an analyst's performance.
Feedback loops	Frequent feedback (could be daily or more often). Feedback from subject matter experts and senior analysts.	Less frequent feedback (could be weekly, bi-weekly, or monthly). Feedback from data consumers, subject matter experts, and senior analysts.
Communication	Communicate frequently to analysts via alerts or reporting bursts.	Communicate periodically to motivate and uphold continuous improvement.
Goal	Meet accuracy targets. Train and improve lower-performing analysts.	Continually meet or exceed accuracy targets. Prevent and mitigate performance issues.
Duration	Execute multiple cycles until the threshold is met.	Execute throughout the project lifecycle.
Reporting	No specific reporting, but study and analyze to see patterns and trends.	Standard and ad-hoc monitoring reports. Personalized performance reports for each analyst.

Quality Management Framework Part 4

Continuous Improvement

Continuous improvement is a method of monitoring, evaluating, and improving your data processing operations and data products and services.

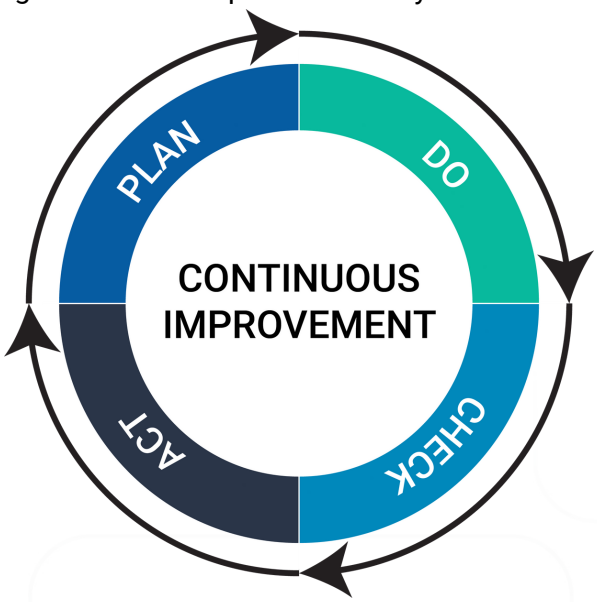
It is a never-ending quest for greater accuracy and ongoing improvement.

At CloudFactory, we use two major approaches for achieving continuous improvement, both from the Lean methodology: the Plan-Do-Check-Act cycle (PDCA) and root cause analysis.

Plan-Do-Check-Act Cycle

The PDCA cycle is the most common model for bringing continuous improvement to your data processing practice.

- **Plan**: Establish the objectives, processes, and KPIs necessary to obtain your expected results.
- **Do**: Carry out the plan as a pilot or a limited study.
- **Check**: Compare the results achieved to the expected results. We typically recommend using error analysis for this phase.
- **Act**: If the analysis reveals a successful pilot, roll out the changes and begin planning for the next cycle and even more improvement. If not, work through the cycle again and continue until you achieve results worth rolling out on a broader scale.



One of our AR & AP Automation clients requires an error rate under 400 defects per million opportunities (DPMO). Their team (and ours) is delighted that our error rates routinely hit 220 DPMO.

Client Profile:
AI-Driven InsurTech

To improve insurance outcomes, our client in the insurance technology space offers a business intelligence platform that uses AI to digitize risk for commercial property owners.

More than 100 CloudFactory data analysts contribute 8,000 hours each month to support two unique tasks: 1) pre-cleansing documents by identifying incomplete, incorrect, inaccurate data and modifying data where necessary, and 2) geocoding documents by pinning the correct location on a digital map.

Root Cause Analysis

Root cause analysis is an iterative process by which you drill down into a problem until you uncover the root cause of an error.

This practice supports CloudFactory's mission of equipping data analysts with the skills they need to succeed so they may earn, learn, and serve their way to becoming Leaders Worth Following. While we will not hesitate to offboard analysts who do not perform, we prefer to first get at the root cause of issues and then coach analysts via a performance improvement plan.

We also conduct root cause analysis for our finance industry clients whenever concerns or changes arise, for instance scaling ramps, quality concerns, throughput concerns, the launch of a new use case, or a request for overage hours.

The steps, generally, are as follows:

1. Define the problem or change.
2. Collect data.
3. Identify causal factors.
4. Drill down to the root.
5. Recommend and implement solutions.

Root cause analysis helps you root out underlying issues rather than just treating surface symptoms and putting out fires. It's beneficial when accuracy is below target or at risk.

Keep the following guidelines in mind when performing root cause analysis:

- Beware of acting on superficial symptoms.
- Realize there can be multiple root causes.
- Identify the Who, What, Where, When, Why, and How for a given defect.
- Educate analysts by providing information, training, and example scenarios for defects and corrective actions.

Client Profile:
Financial Document Digitization

This long-term client—a financial technology company changing the way accountants and bookkeepers work—leverages cloud computing and machine learning to convert financial documents into data customers can use.

CloudFactory's data analysts perform real-time invoice transcription to the tune of more than one million invoices and 20,000 hours of work each month. This transcription work, which resulted in a 100-times processing capacity increase, powers our client's platform while playing an integral role in augmenting their machine-learning model, primarily through handling automation exceptions and quality control.

The CloudFactory Approach to Quality

Pulling the right levers in the right ways and at the right times to put a managed workforce engine in motion is a complex effort. But over the years, CloudFactory has developed a three-pronged, strategic approach for making it—and quality data—possible: People, Process, and Technology.

People

Your team: Workforce, project manager, client success manager

Technology

Data processing tools and a technology platform for communicating with your workforce

Process

Operations, such as business rules, communication protocols, and quality control

PEOPLE

At CloudFactory, quality begins with people. That's why we vet our data analysts for personal character, technical skills, and cultural awareness through a rigorous screening process. Each analyst joins a small team, where their strengths are known and valued by leaders.

A dedicated project manager leads your analyst team, serves as your day-to-day contact, and optimizes the analysts' outputs.

You'll also work with a client success manager who:

- Advocates for you.
- Tracks outcomes to ensure we meet target metrics.
- Troubleshoots as needed.

Learn more about our people.

PROCESS

CloudFactory begins working for you even before you become a client. As part of the proposal process, we perform a thorough (and complimentary) expert analysis of your project requirements and deliver a complete playbook with task feedback, QA workflow recommendations, and an outline of the ideal skills your workforce will need.

Our solid, agile processes mean you can quickly scale across all types of use cases and achieve better outcomes during our work together.

Learn more about our process.

TECHNOLOGY

CloudFactory is tool agnostic.

Instead of forcing you to work in a tool we created, we partner with best-of-breed tooling providers while also leaving the door open to work using your tools.

This approach maximizes quality outputs and frees us to invest our focus and resources where we excel—the managed workforce.

As our client, you receive access to our workforce management platform for real-time productivity tracking, quality control checks, and team communication.

Learn more about our technology.

Typical Tasks We Handle for Fintech and Finserv Companies

At CloudFactory, we specialize in document and data processing work for companies in multiple spaces, including, but not limited to, accounting software, AP and AR automation, expense management, insurance software, insurtech, mortgagetech, banking, consumer lending, loan processing, legal documents, supply chain and logistics, signature compliance, tax filing and processing, and recruitment and HR technology.

The typical tasks we handle for these companies fall into four categories: Transcription, enrichment, moderation, and data tagging.

Transcription	Enrichment	Moderation	Data Tagging
• Receipt transcription • Invoice transcription • Document transcription • Image to text • PDF extraction • Data reformatting	• Web research • Data mapping • Metadata creation • Lead enrichment • Data normalization • Data entry • Website migration • Data classification	• Quality control • Sentiment analysis • Image verification • Data verification • Call recording audits • Exception handling • Customer support	• Image tagging • Video tagging • Object identification • Audio tagging • Product identification • Image classification

What tasks can we support you with? **Let us know.**

Key Questions to Ask Prospective Data Processing Partners

- Can you help us address process inefficiencies?
- How can we diversify our sub-processors to build in redundancy and mitigate risk?
- Can you help us improve the customer experience and grow the business?
- In terms of data processing, what can we do to stay competitive?
- What are our peers doing in terms of outsourced data processing?
- What is your client onboarding process like?
- How do you screen and select your workforce?
- What is your analyst training process like?
- Will we work with the same analysts over time?
- What is the average tenure of your data analysts?
- How do you transfer context and domain expertise to new data analysts?
- How do you manage changes or iterations from our team?
- How do you report on quality metrics?
- What happens if you don't meet quality metrics?
- How involved does my team need to be?
- Do you offer dedicated project managers?
- Do you offer customer success managers?
- How will our team communicate with your analysts?
- Where are your office locations?
- What time zones do you work in?

Ready to uplevel or scale your document processing operations?

So are we. Let's talk.

contact@cloudfactory.com

cloudfactory

UK | USA | NEPAL | KENYA

CloudFactory's strategic blend of people, process, and technology helps financial services and financial technology companies scale and optimize business operations, maintain the pace of innovation, deliver exceptional customer experiences, and stay competitive.

Using virtually any tool, even the ones you build, our managed teams of data analysts can process data with up to 25% higher accuracy than crowdsourced teams. Trusted by 500 companies, we process hundreds of millions of documents each day for innovators like Expensify and Ibotta.

What documents or data can we process for you?