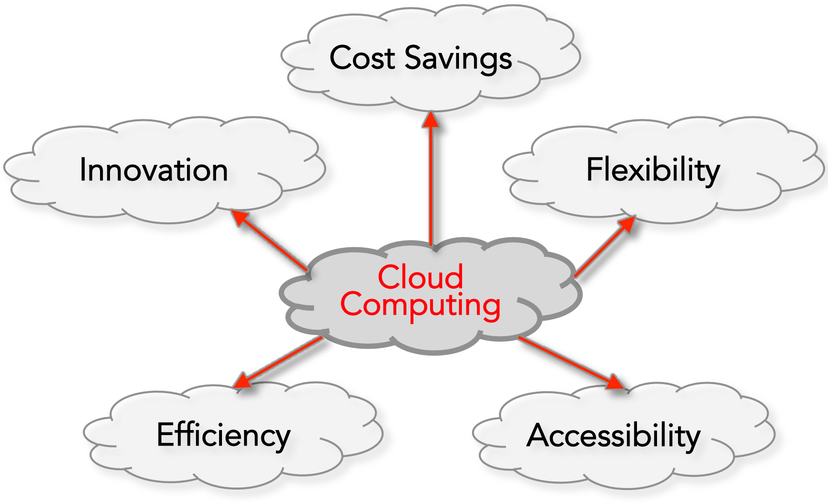
Scalability for Post-Trade Processing: a Benchmark study

思新 SERISYS

www.serisys.com

Introduction

Capital markets are changing rapidly as the benefits of Cloud Computing are being realised.



> "Cloud Computing offers the promise of massive cost savings combined with increased IT agility."
>
> Source: National Institute of Standards and Technology

Today's technologies offer opportunities to capital markets participants

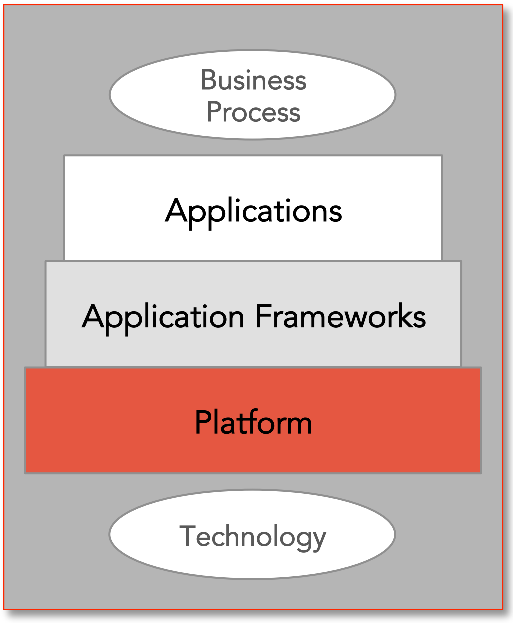
- to be more responsive to business changes
- to reduce post-trade processing costs
- to accommodate new distributed ledger business models (blockchains)
- to respond to regulatory demands and deadlines for greater transparency and control.

This brochure explains how ADYPT CLOUD's adaptive process engine is able to handle unlimited volumes of post-trade processing in a cost-effective manner by exploiting elastic cloud environments. On the following pages, we'll:

- examine a benchmark scenario mimicking post-trade processing at increasing volumes
- prove that the ADYPT CLOUD platform processes over 500,000 trades per hour with twenty 4vCPU servers with 7.5 GiB memory, and
- extrapolate to over one million trades per hour with fewer than 50 such servers.

ADYPT CLOUD is scalable to the volumes of global banks

Post-Trade Cloud Processing



ADYPT CLOUD's overall architecture of Applications, Application Frameworks and Platform decouples *Business Process* changes (e.g. adding a new product, workflow or business unit) from *Technology* changes (e.g. program code or database).

The business behaviour of the application is similarly totally decoupled from the environment into which it is deployed. It doesn't know, or need to know, how many servers business processes might be executing on, nor where these physically exist.

This insulation from the physical and network characteristics of the underlying hardware is guaranteed by the ADYPT CLOUD Runtime Engine, a part of the Platform

ADYPT CLOUD's business benefits

Supports Innovation:

- Built for the Cloud
- Enables innovations such as blockchain clearing

Provides Cost Efficiency:

- Processes all financial instruments in capital markets
- Enables the lowest cost of operations
- Supports the lowest cost of changes

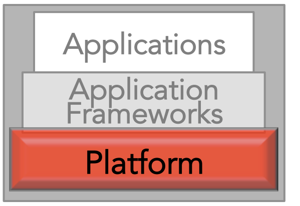
Provides Agility:

- Supports the fastest implementation
- Scalable to the volumes of global banks

Control and Compliance:

- Provides high security
- Enables all activities to be audited in real time and in context
- Uses open standards throughout

Platform



ADYPT CLOUD RUN TIME ENGINE

ADYPT CLOUD's models deploy into the ADYPT CLOUD Run Time Engine and drive the execution of business concepts across the platform services - workflows, audit and authorisation, integration points etc.

The Run Time Engine is a confederation of processing units which are:

- Self-governing, self-managing, self-contained, self-sufficient
- Decoupled – minimal risk of fault propagation
- Choreographed not orchestrated
- Model driven: processing units "know" what to do.

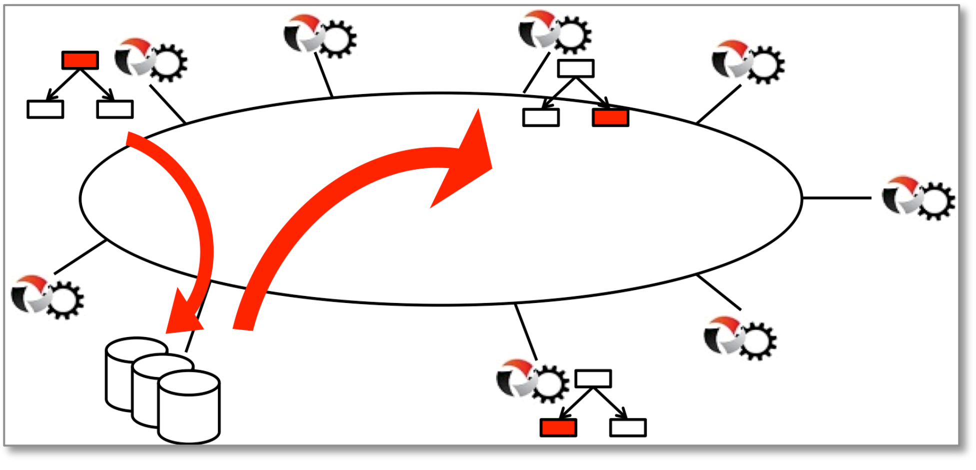
Every processing unit performs all functions, therefore each processing unit serves as a redundant backup for every other processing unit.

Processes continue seamlessly elsewhere in the event of a failure of a processing unit because all state information is stored in the database.

Recovery from whole datacentre failure is determined simply by the database configuration with no dependence on ADYPT CLOUD itself. All inter-process communication is via the committed state of processes in the database.

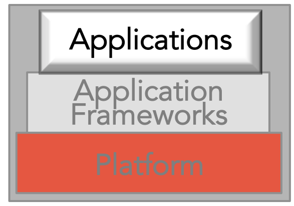
- Runs on any network and on any number of commodity machines, whether on the Cloud, an internal Cloud or a server farm
- Manages capacity easily because the performance of each processing unit is predictable
- Continues business processing seamlessly even if some components of the technology infrastructure fail

Zero Configuration – the system functions optimally on elastic computing environments with no manual intervention



ADYPT

Post-Trade Scalability



SCALABILITY

ADYPT CLOUD's Run Time Engine is built for elastic compute environments such as the Cloud. Additional processing power is provisioned without the need for reconfiguration.

As the pool of units grow and shrink elastically to meet demand, they self-balance their loads with idle processes which pick up outstanding work items, unconstrained by function. The utilisation of the pool of processing units is readily measured, quantified and predicted.

SCALABILITY BENCHMARK

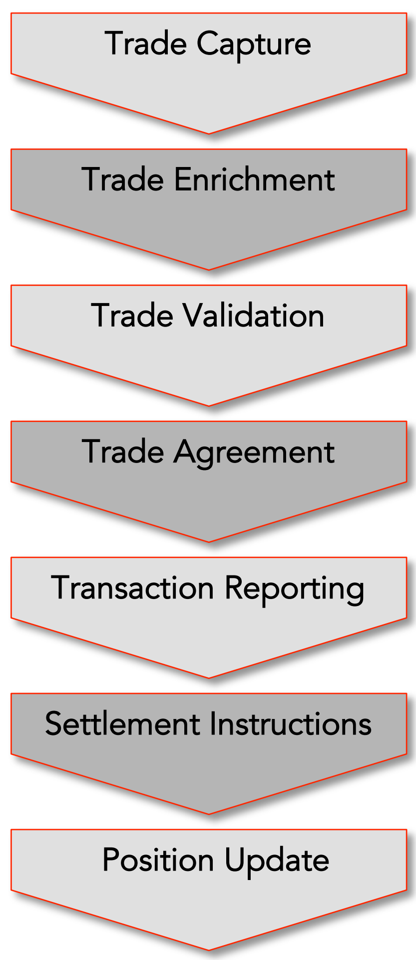
ADYPT CLOUD's ability to optimally run in a cloud environment is regularly assessed and tested. The most recent test aimed to set a benchmark for the system's capacity to process broker equity trades in the cloud.

To assess this, batches of randomly generated trade messages were fed into ADYPT CLOUD, as if they had been received from an upstream front office or order capture system. The trade processes then progressed through the whole range of activities performed on TD+0. This includes:

- validation and enrichment of the trade message
- creation and configuration of the trade object in accordance with the specific details of the trade message (for example, addition of any extra required cross-currency behaviours if multiple currencies are specified in the message)
- matching & calculation of key amounts (e.g. market specific charges, countervalues)
- generation of obligations on parties to exchange assets
- matching and enrichment of Confirmation information
- matching and enrichment of Settlement information
- generation of Settlement instructions in SWIFT format
- matching of position rules and generation of balanced position updates.

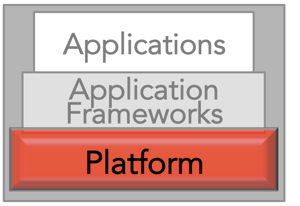
- ADYPT CLOUD uses *Xtreme Modelling*™ to process a wide variety of business processes but is particularly well suited to those we encounter in the financial markets, with lifetimes of a few milliseconds to several hours, days or years, and with many complex interactions involving other processes, systems, and users

- We assess the scalability using the primary processing scenario for ADYPT CLOUD, which is the middle and back office activities around post-trade capture., where ADYPT CLOUD enables the organisation to achieve exceptional rates of STP



ADYPT

Scalability Benchmark

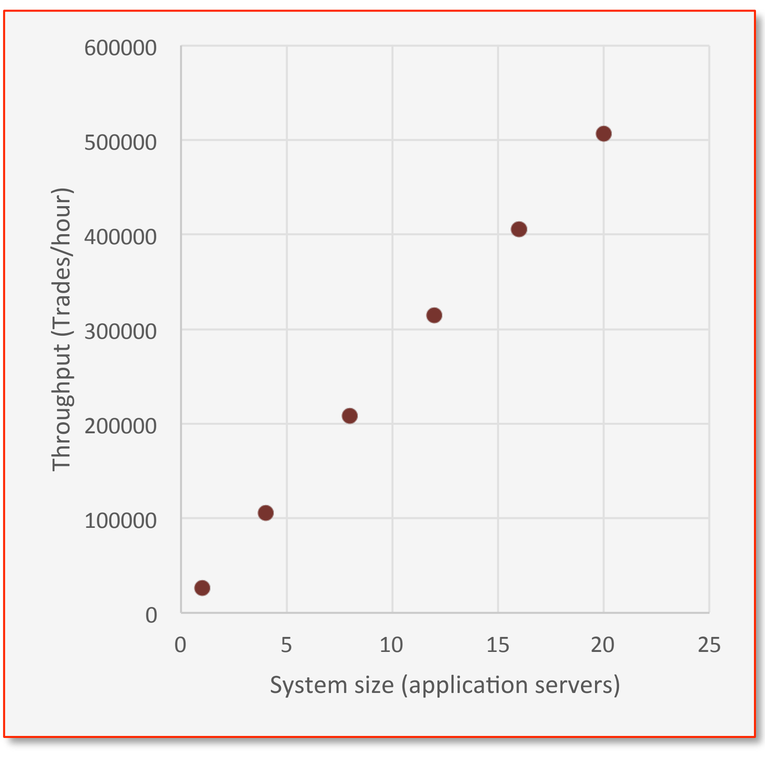


SCALABILITY BENCHMARK

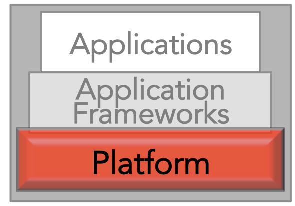
Trade messages were fed to ADYPT CLOUD at high rate to ensure that ADYPT CLOUD was always 100% utilized. The throughput of the system was then measured in terms of trade messages processed per hour for varying sizes of deployment – by varying the number of ADYPT CLOUD processing units executing within the cloud environments.

Our goal in ADYPT CLOUD is that both the application and the runtime engine should be wholly scalable: there's no built in limit to the volumes of trades that can processed, and there is no special "high-volume" edition of the system needed to handle the volumes encountered by global players in the markets.

The scalability graph below demonstrates how ADYPT CLOUD can grow from one through ten to twenty servers, increasing the throughput of trades processed in proportion to the size of the system.



- It's important to remember that Scalability is a function and not a number. Scalability describes how well the system is able to adapt to handle increased load. In a system exhibiting ideal scalability, the efficiency of the system is constant irrespective of the load. This is hard to achieve in practice!
- ADYPT CLOUD's mechanism of scaling is to gear up in size to handle greater volumes of trade: the number of cloud servers is seamlessly increased to meet increasing demand. We can therefore quantify the scalability by measuring the throughput of the system as we vary the number of active processing units
- Plotting a graph of throughput versus system size gives us the *scalability curve*. In a system with ideal scalability this is linear: doubling the size of the system results in a doubling of the throughput.
- For operational planning purposes, the scalability curve gives a form of cost-benefit analysis.



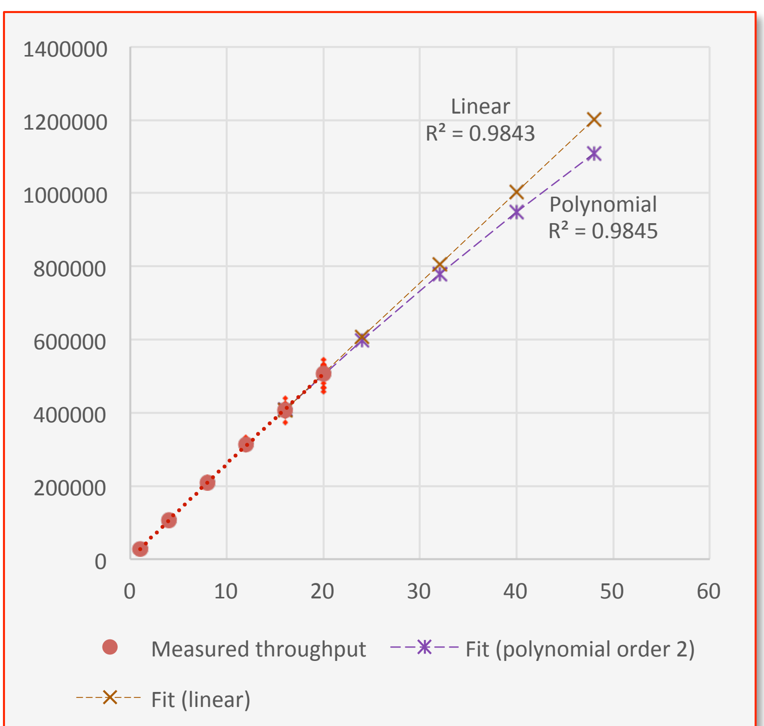
ADYPT Scalability Benchmark

KEY OBSERVATIONS

The throughput of ADYPT CLOUD was measured for various sizes of system up to twenty processing units, each of which equated to a 4vCPU server within the Amazon Elastic Compute cloud. For this exercise we restricted the system to using a single database server, running a standard Amazon Oracle image.

The system demonstrated processing volumes in excess of 500,000 trades per hour with 20 servers with 7.5 GiB memory. The scalability graph clearly indicates that adding further servers would enable the system to service increasingly higher volumes to over one million trades per hour with fewer than 50 such servers.

At the high end of system size we detected a small loss of efficiency of around one-quarter of one percent (0.25%) per processing unit added. We can relate this to file access in the Oracle database, relating to database commits (Oracle's "redo logs"). Reconfiguration of the database's files and the media on which they were hosted showed that this effect could be reduced or increased according to how the database was set up. In a production system the database administrator would monitor and adjust for this effect - amongst many others.



- ADYPT CLOUD processing units operate in parallel and independently of each other. In a scenario such as this benchmark, the single database is the common shared element in the stack, and its capabilities must be shared between all ADYPT CLOUD processes
- Potentially then ADYPT CLOUD processes could find themselves waiting for something in the database, be it CPU, a row lock, or for file access to be performed. This waiting is described by *Queue Theory* and is a classic problem in multi-processor systems.
- However, the benchmark showed zero lock contention in the database. And ADYPT CLOUD's use of the database was just as scalable as its use of cloud application servers: adding additional resource to the database enabled it to productively handle load in a near-linear fashion.
- Future benchmarks are planned which will demonstrate the use of multiple, smaller database servers to partition the load.

Based on the recorded data, we can extrapolate to show the increase in throughput far beyond the size of system studied in the benchmark. Both the linear and polynomial fits describe the data well. The polynomial fit shows the deviation from linear in the event that the DBA leaves the log file access uncorrected.

www.serisys.com