

LIBRO BLANCO | ABRIL DE 2016

Gestión de datos de prueba

Cómo este método puede convertirse en la única de implementar la entrega continua

Huw Price
CA Technologies



Índice

Sección 1:	3
Introducción a la entrega continua	
<hr/>	
Sección 2:	3
La TDM deficiente impide la implementación de la entrega continua	
<hr/>	
Sección 3:	6
Enfoque de TDM basado en requisitos	
<hr/>	
Sección 4:	8
Resumen	
<hr/>	
Sección 5:	9
Referencias	
<hr/>	
Sección 6:	9
Acerca del autor	

Sección 1

Introducción a la entrega continua

Cada vez se escucha más el término “entrega continua” dentro del contexto del desarrollo de software. Muchos distribuidores prometen que pueden conseguir que las organizaciones implementen la entrega continua, por lo que ofrecen sus herramientas para acabar con las causas tradicionales de los fallos y retrasos en los proyectos. Afirman que, al adoptar estas herramientas, las organizaciones podrán innovar y entregar software de calidad de manera continua, puntualmente y conforme con el presupuesto.

Es comprensible que la entrega continua sea un concepto que llame la atención. En la economía de aplicaciones moderna en que vivimos, las organizaciones confían en el software para ofrecer valor a sus clientes. Por consiguiente, cada vez las exigencias de los departamentos de TI y los negocios están más en sintonía. Asimismo, la posición de una empresa en un mercado dependerá de su capacidad para ofrecer valor al consumidor día tras días. Los equipos de TI deben ser capaces de responder con rapidez a las cambiantes expectativas de los consumidores y el mercado desarrollando software que se base en las necesidades fundamentales para los negocios y, al mismo tiempo, reduciendo los costes y las horas invertidos en realizar pruebas.

Si bien muchas herramientas afirman poder implementar la entrega continua, son fundamentalmente logísticas y solo pueden implementarse en las fases finales del ciclo de vida de desarrollo. Además, la metodología de DevOps solo se contempla en términos de operaciones, ya que estas herramientas aprovechan el software ya diseñado y, posteriormente, posibilitan la realización de pruebas de desarrollo y regresión. Los problemas sistemáticos que surgen a partir de la fase de recopilación de requisitos en adelante siguen existiendo y causan costosos cuellos de botella y retrasos en los proyectos que hacen que sea imposible implementar con éxito la entrega continua.

En concreto, se pasa por alto la importancia de la gestión de datos de prueba (TDM) —o entregar los datos correctos en el lugar y momento adecuados—. Si el diseño de las pruebas es deficiente y se aprovisionan datos de baja calidad de una forma ineficaz, los equipos de pruebas no dispondrán de los datos que necesitan para probar un sistema en su totalidad. Es decir, se compromete la calidad para poder entregar el software a tiempo y conforme con el presupuesto.

Sección 2

La TDM deficiente impide la implementación de la entrega continua

Para la mayoría de las organizaciones y los distribuidores, la TDM empieza y acaba con los procesos de copia, enmascaramiento y, posiblemente, segmentación de los datos de producción. Estas estrategias de TDM son puramente logísticas, ya que se centran exclusivamente en mover los datos, que no se tienen en cuenta a la hora de diseñar los requisitos de los proyectos. Los datos de producción que se copian y migran a un entorno de desarrollo, de pruebas o de control de calidad suelen denominarse “copia definitiva”; es decir, un conjunto de datos perfecto con el que los comprobadores pueden realizar todas las pruebas que sean necesarias para probar un sistema en su totalidad.

Una buena forma de empezar y de conseguir corregir, en parte, los puntos débiles tradicionales en materia de conformidad y costes de infraestructura es aplicando procesos de enmascaramiento y segmentación. No obstante, *per se*, el enmascaramiento y la segmentación de una base de datos de producción generan muchos de los problemas inherentes al uso de este tipo de datos. Todos estos problemas propician que sea imposible implementar con éxito la entrega continua.

Tiempo desaprovechado

La mayoría de las organizaciones carecen de un servicio de aprovisionamiento o de un equipo de TDM centralizado, de modo que son los equipos locales quienes tienen que localizar o crear los datos. Los comprobadores pueden invertir hasta la mitad de su tiempo buscando datos, y se desperdicia hasta un 20 % de todo el SDLC esperando a que se finalice esta fase. Esto se traduce en tener que probar cuellos de botella que obstaculizan implementar con éxito la agilidad y la entrega continua. En una instancia, Grid-Tools (que ahora parte de CA Technologies) detectó que un equipo tenía que trabajar en iteraciones de tres semanas, pero que tuvo que invertir cuatro semanas en preparar los datos para las iteraciones.

Como consecuencia del enmascaramiento y la segmentación de los datos de producción, se precisa contar con equipos dedicados a buscar los datos que se utilizarán, lo que constituye una tarea ardua y lenta, que se acentúa por el hecho de almacenar los datos de forma incoherente en hojas de cálculo no controladas. Por ejemplo, podría agregarse la información que permite la identificación personal (PII) en una columna llamada "Notas" si tenemos tres columnas para tarjetas de crédito y un cliente utiliza cuatro. Por otro lado, los datos deben ser relacionamente intactos. Cuantos más complejos sean, más complicado será de lograr esto. Las horas y el esfuerzo que se invierten en enmascarar los datos compensan el tiempo que habrían dedicado los comprobadores si tuvieran que haber empezado de cero.

Crear los datos desde cero manualmente puede ser igual de laborioso. Además, como los datos se suelen crear teniendo en cuenta pruebas específicas, enseguida se vuelven obsoletos e irrelevantes, ya que el entorno real cambia de forma constante. Un ejemplo que se suele utilizar es el de las tasas de cambio de divisas y las tendencias comerciales, donde los datos quedan desfasados a diario. Cuando esto sucede, tienen que actualizarse (otro retraso que deriva en una gran pérdida de tiempo) o, lo que suele hacerse, descartarse. Lo mismo ocurre cuando se cambia una versión o una persona solicita que se actualicen los datos, o cuando hay que crear datos cada vez que se realizan cambios en un entorno virtual.

Rara vez se pueden compartir y reutilizar los datos que crean de forma manual los equipos. El hecho de no poder aprovechar el esfuerzo dedicado durante fases anteriores a crear los datos genera más trabajo y obliga a las organizaciones a tener que, lamentablemente, decidirse entre reducir los costes y el plazo de comercialización o entregar aplicaciones de software valiosas con todas las funciones requeridas.

Restricciones de dependencia

Los comprobadores desperdician una gran parte de su tiempo esperando a que se aprovisionen los datos o a que estén disponibles. Las organizaciones y los distribuidores suelen entender el SDLC como una serie de etapas lineales: se avanza hacia la siguiente fase cuando un equipo finaliza su tarea. Por tanto, los equipos tienen que esperar a que los equipos de la fase anterior pongan a disposición los datos adecuados. Además, podrían encontrarse con la situación de que no pueden utilizar los datos que necesitan, ya que el otro equipo está trabajando en el mismo conjunto de datos.

Esto contrasta con los principios del desarrollo paralelo ágil del eje de la entrega continua, donde se aprovecha al máximo el tiempo y las capacidades de los recursos para innovar. Además, los cambios que realice un equipo podrían afectar a los datos de tal forma que, cuando otro equipo los utilice, las pruebas de aplicaciones fallen sin motivo aparente.

Si los equipos tienen que crear los datos manualmente o esperar días o semanas a que estén disponibles, no se podrá responder con rapidez a los requisitos cambiantes ni entregar software que se haya probado en su totalidad.

Baja calidad

El verdadero problema que existe con los datos de producción en entornos que no son de este tipo es la calidad. Normalmente, los datos de producción solo proporcionan una cobertura funcional de entre el 10 y el 20 %, de modo que es poco probable que los métodos de muestreo ofrezcan datos que superen todas las pruebas necesarias para crear un nuevo subsistema. Una gran parte de los datos de producción son muy similares, ya que proceden de las transacciones empresariales habituales, así que, por su propia naturaleza, se limpian con el objetivo de excluir los datos incorrectos que causarían daños en los sistemas. Por tanto, las pruebas suelen realizarse para poner más énfasis en las rutas previsibles que en las pruebas negativas o no funcionales.

No obstante, las pruebas negativas deben abarcar alrededor del 80 % de todas ellas, ya que son estos escenarios atípicos y limitados los que provocan que se colapsen los sistemas. Siempre que las pruebas se centren en las rutas previsibles, los defectos llegarán inevitablemente a la fase de producción, con lo que se producirán repeticiones de tareas, retrasos importantes, aumento de costes y posibles fallos en los proyectos. Las investigaciones del sector demuestran que se tarda el doble en corregir un error durante la fase de pruebas que en identificarlo en la de requisitos¹, con los consiguientes retrasos e incapacidad para responder a los cambios, que obstaculizan la implementación de la entrega continua.

La cobertura funcional de los datos de producción puede complementarse con la creación manual, pero se trata de un método inexacto, lento y acientífico. No hay forma de comprobar si se ha alcanzado esa cobertura máxima, ni de saber si los datos siguen estando referencialmente intactos.

Costes elevados

Finalmente, además de los costes elevados asociados a las repeticiones de tareas y a las entregas tardías, el proceso de copia de datos de producción es excesivamente lento y costoso. El informe revela que algunas organizaciones se encuentran con hasta 20 copias de una sola base de datos, lo que acarrea un importante desembolso en hardware, licencias y soporte.

Además, el almacenamiento de una gran cantidad de datos puede resultar muy costoso. Una investigación relacionada sobre almacenamiento de datos de terceros señala que, al mover los datos a niveles con un coste más reducido —entre ellos, la nube— se reduce temporalmente el coste de almacenamiento, pero no soluciona ni aborda el problema fundamental del aumento del volumen de los datos. Es decir, cualquier política de TDM cuya fase final consista en migrar, segmentar y enmascarar los datos no podrá abordar el problema de cómo almacenar de manera rentable los datos que se procesan con aplicaciones cada vez más complejas y modernas.

También nos encontramos con el riesgo de no cumplir las normativas, que es más costoso y no se soluciona aunque se enmascaren los datos de producción. Un estudio independiente reveló que en 2014 aumentó la media del coste de las fugas de datos en un 15 % (3,5 millones de dólares por registro). Por otra parte, la nueva directiva de protección de datos de la Unión Europea (UE), que entrará en vigor en 2016, sancionará con multas de hasta 100 millones de euros² o el 5 % de la facturación, lo que suponga un importe más elevado para el negocio.

Con el enmascaramiento de los datos de producción no se garantiza la conformidad, ya que los riesgos reales siguen siendo de naturaleza humana, donde más del 50 %³ de las fugas de datos pueden estar relacionadas con el comportamiento del personal interno. Asimismo, al realizar este proceso, la integridad referencial de los datos debe permanecer intacta; de lo contrario, podrían fallar las pruebas de aplicaciones. Sin embargo, cuantos más complejos sean los datos, más fácil será que se produzcan errores, ya que habrá más fragmentos de información que pueden estar correlacionados.

La única forma real de evitar estos costes de infraestructura elevados, así como de garantizar que no se produzcan fugas de datos confidenciales, consiste en no utilizar datos de producción. Aunque estos gastos no sean inherentes de organizaciones que desean implementar un marco de entrega continua, en realidad, contrastan con sus principios. Además, evitan que las organizaciones puedan entregar de forma rápida y conforme con el presupuesto software de calidad basado en las cambiantes demandas empresariales.

Sección 3

Enfoque de TDM basado en requisitos

Las organizaciones que desean implementar la entrega continua deben dar un giro de 180 grados a sus procesos de desarrollo y prueba en lugar de reestructurarlos simplemente: tienen que replantearse su enfoque de TDM. Si adoptan un enfoque completo e integral de gestión de datos de prueba basado en requisitos, las organizaciones podrán resolver en origen las pruebas, mitigar los riesgos y minimizar la generación de defectos, por lo que podrán entregar software de calidad de forma más rápida y con un coste menor.

En lugar de procesar los datos prueba por prueba, las organizaciones deberían ver los datos en términos de decisiones de diseño (es decir, de los propios requisitos) con el fin de crear pruebas con datos que estén asociados directamente a ellas. Si se adaptan los datos de pruebas a los requisitos, se garantiza que estos sean adecuados. Además, si se pueden zaprovisionar rápidamente a los equipos de pruebas, podrán responder con rapidez a las cambiantes demandas empresariales.

En este sentido, la entrega continua no puede iniciarse en las últimas fases del SDLC: debe comenzar con una idea (los propios requisitos) para que los equipos de desarrollo y pruebas puedan responder con rapidez a las cambiantes exigencias empresariales y trabajar a partir de esta idea, que va cambiando a lo largo de todas las fases. Los equipos tienen que validar y comprobar de forma sencilla los requisitos creando las pruebas, que están asociadas a los datos virtuales y resultados que se prevé obtener, en función de dichos requisitos. Este modelo basado en pruebas es a lo que llamamos “resolver en origen” y “entregar software de forma continua”. El objetivo consiste en condensar todo el trabajo del ciclo de vida de desarrollo en la fase de recopilación de requisitos; todas las tareas posteriores deben derivar de esta etapa, aunque los requisitos cambien.

Creación de requisitos más eficaces

Los requisitos tienen que incluir toda la información cualitativa de un sistema que deba probarse, de modo que los comprobadores puedan basar los casos de uso y las pruebas en la idea inicial. Además, se debe poder realizar un seguimiento de dichos requisitos. Esto será necesario si los comprobadores tienen que actualizar rápidamente las pruebas cuando se modifiquen los requisitos.

Con las herramientas adecuadas, el modelado de diagramas de flujo de trabajo puede generar un diagrama de flujo de trabajo que contenga toda la información cualitativa sobre el sistema que debe probarse, a pesar de lo sencillo que resulta diseñar el propio flujo. CA Agile Requirements Designer (anteriormente, Agile Designer de Grid-Tools) ofrece un diagrama de flujo con toda la lógica funcional necesaria para obtener de forma automática la cantidad más reducida de pruebas con la máxima cobertura, evitando interrupciones en la actividad empresarial, como suele ocurrir con los enfoques de comparación y modelado de causa-efecto.²

Además, con los diagramas de flujo se aumentan las probabilidades de que los requisitos sean completos e inequívocos, ya que dividen los diagramas poco prácticos y dispares que suelen crear los requisitos en conjuntos de datos pequeños y claros. Estos procesos reflejan la lógica de causa-efecto de un sistema generando una serie de respuestas a hipótesis.

Esto no solo ayuda a reducir el 56 %⁴ de los defectos que surgen como consecuencia de las ambigüedades de los requisitos, sino que también obliga al equipo encargado de los requisitos a pensar en términos de modelado de sistemas; es decir, en limitaciones, restricciones y condiciones marco. A partir de estos requisitos modelados como un diagrama de flujo, pueden identificarse todas las rutas posibles de un sistema. Por tanto, pueden obtenerse pruebas que ofrecen una cobertura funcional completa, de modo que cubren todas las rutas posibles de un sistema, incluidas las negativas y los resultados no previstos.³

Con un diagrama de flujo activo, estas pruebas pueden vincularse posteriormente a estadísticas de complejidad, datos virtuales, datos de pruebas, scripts de automatización, resultados esperados y tareas pendientes, con lo que los esfuerzos del SDLC se concentran en la fase de recopilación de requisitos.

Aprovisionamiento de los datos adecuados

La automatización de los procesos de diseño de pruebas con herramientas como CA Agile Requirements Designer permite a los comprobadores generar las pruebas adecuadas que se precisan para probar requisitos de forma óptima y satisfactoria. Esto nos lleva hasta el argumento central de este libro blanco: se requiere una TDM más eficaz para implementar con éxito la entrega continua. Los comprobadores necesitan acceder a los datos adecuados (entregados en el lugar correcto y en el momento oportuno) que puedan ayudarlos a alcanzar la cobertura completa de las pruebas.

Calidad

Tal y como hemos comentado, los datos de producción no pueden proporcionar la cobertura necesaria para probar un sistema en su totalidad. En cambio, la generación sintética de los datos de pruebas puede generar pequeños pero completos conjuntos de datos que cubran todos los escenarios posibles, incluso cuando un evento no ha ocurrido anteriormente. Cada escenario real puede verse como otro punto de datos y, por tanto, pueden crearse datos sintéticos, incluso para escenarios nuevos y futuros. CA Test Data Manager (anteriormente, CA Data Finder o Data Maker de Grid-Tools) utiliza técnicas de creación de perfiles de datos inteligentes para obtener una perspectiva precisa de un modelo de datos, con lo que se generan datos completos y sofisticados que proporcionan todas las variaciones funcionales basadas en dichos datos.

Con un enfoque basado en requisitos, los datos generados se asocian a una prueba, de modo que se garantiza que sean adecuados y que satisfagan todas las exigencias de los comprobadores. Gracias a esta generación de datos de pruebas, será más probable que los equipos detecten defectos a la primera, evitando las repeticiones de tareas que tanto tiempo requieren y que impiden que pueda implementarse con éxito la entrega continua.

Tiempo

El tiempo que se invierte en los procesos manuales de creación o manipulación de datos, o en esperar a que estos estén disponibles, contrasta con los principios de la entrega continua, tal y como hemos comentado.

En cambio, las herramientas automatizadas de creación de datos, como CA Test Data Manager, pueden trabajar directamente con las capas de API de RDBM o ERP. De este modo, los usuarios podrán generar los datos con la rapidez que permita la potencia de procesamiento. Los scripts de procesos masivos pueden duplicar la cantidad de datos que posee una organización con la misma rapidez que permita la infraestructura de la base de datos. Gracias a esto, los datos adecuados pueden generarse en cuestión de horas en lugar de semanas, por lo que los comprobadores dispondrán, en una iteración, de los datos que necesitan en el momento oportuno.

Además, con la potente funcionalidad de asociación de pruebas, si los datos ya existen, pueden identificarse y extraerse de varias fuentes de origen diverso antes de asignarse y asociarse a las pruebas adecuadas. Está demostrado que esta funcionalidad reduce en un 95 % el tiempo invertido en buscar y aprovisionar datos, en comparación con los procesos manuales. Además, permite a los equipos conocer qué pruebas fallarán debido a errores antes de ejecutarse.

El almacenamiento de los datos de pruebas en un almacén de datos centralizado puede evitar otros cuellos de botella que se producen cuando los equipos esperan a que estén disponibles los datos. El portal de datos de pruebas a petición de CA Test Data Manager permite crear formularios dinámicos, por lo que los usuarios podrán seleccionar qué clase de datos quieren en función de criterios específicos, como el tipo de tarjeta de crédito o la ubicación geográfica. Este formulario asignará los datos necesarios a las pruebas, o bien creará nuevos datos si no se encuentra ninguno. Esto significa que los ingenieros tienen más horas para corregir los defectos expuestos por las pruebas, en lugar de invertir la mitad de su tiempo en esperar a que se aprovisionen los datos. Asimismo, las organizaciones pueden centralizar la propiedad de los datos en el equipo de seguridad de TI poniendo solamente los datos confidenciales a disposición del personal autorizado que los solicite.

Coste

Finalmente, si se aplica una política de TDM más eficaz, se reducirán los riesgos de sobrecostes. Al detectar y solucionar antes los defectos, se podrá reducir la generación de defectos en hasta un 95 %⁵ y ahorrar más de 50 000 \$⁵ por defecto. Además, como se emplean subconjuntos de datos más pequeños y completos, se podrán reducir los costes de infraestructura en hasta 50 000 \$⁵ por base de datos, y como se llevan a cabo menos pruebas pero de mayor calidad, también disminuyen significativamente los costes y el tiempo asociados a las pruebas. Al emplear datos sintéticos, también se acaba con el riesgo de que aparezcan las costosas no conformidades, ya que la información confidencial no sale de los entornos de producción.

Sección 4

Resumen

Confiar exclusivamente en los procesos de copia de datos de producción en entornos que no están preparados para ello impide que se pueda implementar con éxito la entrega continua. Es probable que los equipos de pruebas no puedan utilizar los datos adecuados cuando los precisen para probar el software en su totalidad. Por tanto, no podrán responder con rapidez a los cambiantes requisitos empresariales. Si bien las herramientas de enmascaramiento y segmentación corrigen algunos puntos débiles, los datos de producción no tendrán la calidad suficiente como para detectar defectos y solucionarlos.

Si se quiere implementar con éxito la entrega continua, resultará esencial adoptar un enfoque integral basado en requisitos para desarrollar software y tratar los datos de pruebas como un activo reutilizable y valioso. Si se trabaja a partir de una idea en todas las fases, los equipos podrán responder con rapidez a las cambiantes exigencias de los negocios utilizando los datos adecuados (entregados en el lugar correcto y en el momento oportuno) con el fin de probar el software en su totalidad con el menor número de pruebas posibles.

Gracias a este enfoque, las organizaciones podrán aprovisionar los datos de pruebas de manera eficaz al mismo tiempo que mitigan los riesgos asociados a los retrasos, las repeticiones de tareas y los aumentos de costes que hacen que sea imposible implementar con éxito la entrega continua. De esta forma, podremos entregar de manera continua, puntual y con un coste menor software valioso basado en las exigencias fundamentales de los negocios.

Sección 5

Referencias

- 1 <http://www.softwaretestingclass.com/why-testing-should-start-early-in-software-development-life-cycle/>
 - 2 <http://www.agile-designer.com/resources/test-case-generation-bloor-market-report/>
 - 3 Consulte “A Critique of Testing Primer” de Llyr Wyn Jones para leer un tratamiento conceptual y matemático de las ventajas del modelado de diagramas de flujo en <http://www.agile-designer.com/resources/critique-testing-primer/>
-

Sección 6



Acerca del autor

Cofundador de Grid-Tools y vicepresidente de CA Technologies

Huw Price, elegido como el director de TI del año 2010, ha ostentado el cargo de arquitecto técnico responsable en varias empresas de software de EE. UU. y Europa. Especializado en las herramientas de automatización de pruebas, ha lanzado al mercado numerosos productos innovadores que han reformulado el modelo de pruebas utilizado en el sector del software.



Comuníquese con CA Technologies en ca.com/es



CA Technologies (NASDAQ: CA) crea software que impulsa la transformación de las empresas y les permite aprovechar las oportunidades que brinda la economía de las aplicaciones. El software se encuentra en el corazón de cada empresa, sea cual sea su sector. Desde la planificación hasta la gestión y la seguridad, pasando por el desarrollo, CA trabaja con empresas de todo el mundo para cambiar la forma en que vivimos, realizamos transacciones y nos comunicamos, ya sea a través de la nube pública, la nube privada, plataformas móviles, entornos de mainframe o entornos distribuidos. Para obtener más información sobre nuestros programas de éxito de clientes, visite ca.com/customer-success. Para obtener más información, visite ca.com/es.

1 Estudio de Ponemon Institute sobre el coste de las fugas de datos en 2014

2 Normativa de protección de datos propuesta por la UE

3 Informe de investigaciones de fugas de datos de 2013

4 Informe de Bender sobre los requisitos basados en pruebas

5 Estadísticas recopiladas de la experiencia de implementación de Grid-Tools