

# Hur hittar du rätt i datagruset?



*Det går fort att fylla hårddisken med test-data, så gör systemet skalbart och välordnat från början.*



**Av Stephanie Orci,  
National Instruments**

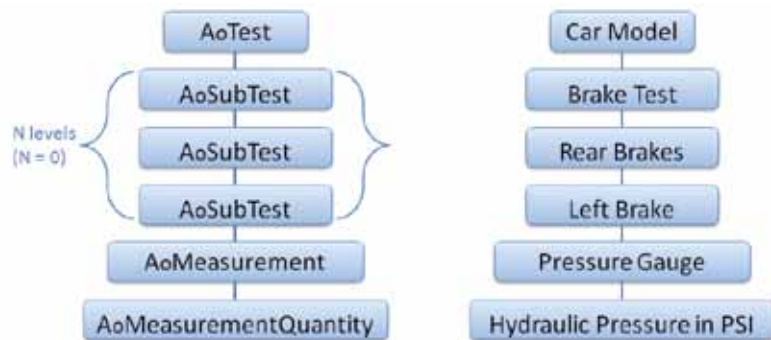
Stephanie Orci har studerat biomedicinsk teknik vid University of Texas i Austin. Efter examen började hon som applikationsingenjör med inriktning på realtidsapplikationer hos National Instruments. För närvarande är hon produktchef för DIAdem.

När man för ett decennium sedan testade en mobiltelefon handlade det om att testa sändning och mottagning av signaler. Idag är telefonen inte bara en telefon utan en komplicerad apparat som kan skicka textmeddelanden, anslutas till Internet, fotografera och mycket mer därtill. Alla funktionerna måste hela tiden uppdateras och testas på allt kortare tid om man ska behålla konkurrenskraften.

Komplexiteten och behovet av korta testtider kräver intelligentare datahanteringssystem som kan skalas upp eller ner när behov uppkommer. Den här artikeln handlar om dagens trender när det gäller att ordna och spara data inför den framtida datahanteringen:

**I TEST- OCH MÄTAPPLIKATIONER** kan man samla in väldiga mängder data. Under varje sekund som den stora Large Hadron Collider vid Cern är igång genererar den 40 Tbyte data. Vid en färd över Atlanten kan en fyrmotorig jumbojet skapa 640 Tbyte driftinformation. Om man multiplicerar det med de över 25000 flygturer som görs varje dag, ger det en uppfattning om vilka enorma datamängder det blir.

Hur man ska kunna dra noggranna och meningsfulla slutsatser från så stora datamängder är ett växande problem. Det är det som uttrycket "Big data" syftar på. De stora datamängderna innebär nya problem inom dataanalys, sökning, dataintegration, rapportframställning och systemunderhåll som måste lösas om man ska kunna hålla jämna steg med den exponentiella datatillväxten. Teknikforskningsföretaget IDC genomförde nyligen en studie av digitala data – inklusive mätfiler, video, musikfiler och så vidare där man uppskattade att mängden tillgängliga data fördubblas vartannat år. Bara under 2011 skapades 1,8 Zbyte (1 021 byte).



En förenklad representation av en basmodell i ASAM ODS.

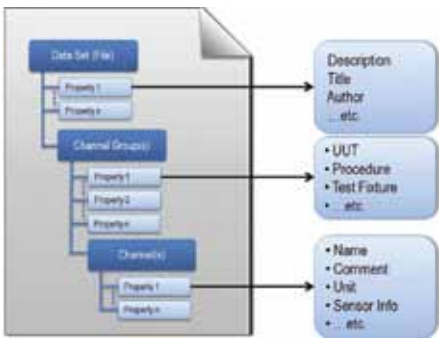
Ger de verktyg som du nu använder möjlighet att gräva igenom de stora datamängder som du samlar in? Stöter du på begränsningar för hur mycket data du kan ta in åt gången, eller får du prestandaproblem när du arbetar med stora datamängder?

**DET FINNS MYCKET** som försvårar optimalt utbyte av teknisk information. Vanligast är olämplig informationsinsamling under testning eller simulering. Alltför ofta lagras uppgifterna utan beskrivning, i olika format och utspridda på ett flertal olika datorer, vilket resulterar i en informationsmassa där det är oerhört svårt att hitta en viss datauppsättning och dra slutsatser från den. När informationen inte går att hitta, måste tester och simuleringar göras om. Följden blir att många företag blir mindre effektiva och får kraftigt ökade utvecklingskostnader. För att klara de här utmaningarna är det många företag som letar efter lösningar som garanterar att data lagras välordnat och skalbart och på effektivaste möjliga sätt.

Inom många branscher har man försökt klara utmaningen genom att implementera egna standarder för lagring och hantering

av data. Inom fordonsbranschen har man på senare tid haft stora framgångar med standarden ASAM ODS, som skapades för att förenkla den universella tolkningen av data som erhållits från test-, evaluerings- och simuleringsapplikationer. Att standarderna blivit så vanliga idag beror till stor del på dess uppbyggnad; modellen är stabil men kan ändå anpassas och utökas till många olika test- och mätapplikationer med skilda behov.

**PÅ MYCKET HÖG NIVÅ** definierar ASAM ODS-standarderna en generisk basdatamodell för universell tolkning av data. Själva modellen definierar ett antal strukturer för lagring av test- och mätdata och de inbördes relationerna mellan strukturerna. Den här uppbyggnaden är medvetet generell för att kunna passa till behoven i alla mät- och automatiseringsmiljöer och säkerställa standardiserad datatolkning från olika applikationsområden inom testautomatisering och mätning. Grundmodellen utformades för att vara anpassningsbar, eftersom den ska kunna användas som utgångsmodell när man skapar konkreta modeller som är specifika för de olika applikationerna.



Varje TDMS-fil innehåller beskrivande information på nivåerna fil, grupp och kanal.

Idén med standardiserade gränssnitt för lagring och hämtning av data är inte specifik för ASAM ODS. Många databaser och filformat har liknande APler som tydligt specificerar hur data ska skrivas i eller läsas från de fysiska lagringsformaten. Ett exempel är filformatet TDMS (Technical Data Management Streaming), som är baserat på några av organisatoriska elementen i ASAM ODS. Det binära TDMS-filformatet är ett lättläst, välstrukturerat filformat för snabb överföring som i kombination med ett effektivt program för datasökning blir snabbt sökbar. TDMS-formatet är strukturerat i en hierarki med tre nivåer: fil, grupp och kanal. Filnivån kan innehålla ett obegränsat antal grupper, och varje grupp kan innehålla ett obegränsat antal kanaler. Den här grupperingen av kanaler gör att man kan välja hur man vill ordna data så att de blir lättare att förstå. Till exempel kan man ha en grupp för rådata och en annan grupp för analyserade data i samma fil. Eller så kan man ha flera grupper som motsvarar givartyper eller platser.

På varje nivå i hierarkin kan man lagra ett obegränsat antal numeriska egenskaper. Varje nivå kan ha ett obegränsat antal egendefinerade attribut så att datafilerna kan bli väldokumenterade och klara för sökning. När dokumentationskraven ökar behöver man inte konstruera om applikationen, utan man bara utökar modellen så att den omfattar de nya kraven. Ju fler parametrar man anger när man dokumenterar sina mätdata, desto lättare hittar man dem i framtiden med hjälp av avancerade sökverktyg.

EN STUDIE HAR VISAT att bara ungefär tio procent av alla insamlade data analyseras. Resten är det ingen som tittar på. Vanligen försöker man efter hand lagra och ordna hundratals eller tusentals datafiler på disk. Filerna lagras ofta på olika sätt och i olika format, och de kan till och med lagras på olika platser i en lokal maskin eller någonstans i ett nätverk. Om man förser sina data med beskrivande information, såsom i TDMS, och använder lämpliga verktyg, blir det lättare att hitta just de data man vill analysera.

För att urskilja trender i många datafiler och hitta de tio procent av data man vill analysera för att snabbt kunna fatta viktiga beslut kräver vissa programverktyg, till exempel Microsoft Excel, att man öppnar varje enskild datafil, kopierar de aktuella kolumnerna, klistrar in dem i en samlingsfil och fortsätter på samma sätt med nästa datafil. En noggrann trendanalys av likartade datakanaler genom hundratals enskilda datafiler kan kräva dagar av värdefull arbetstid.

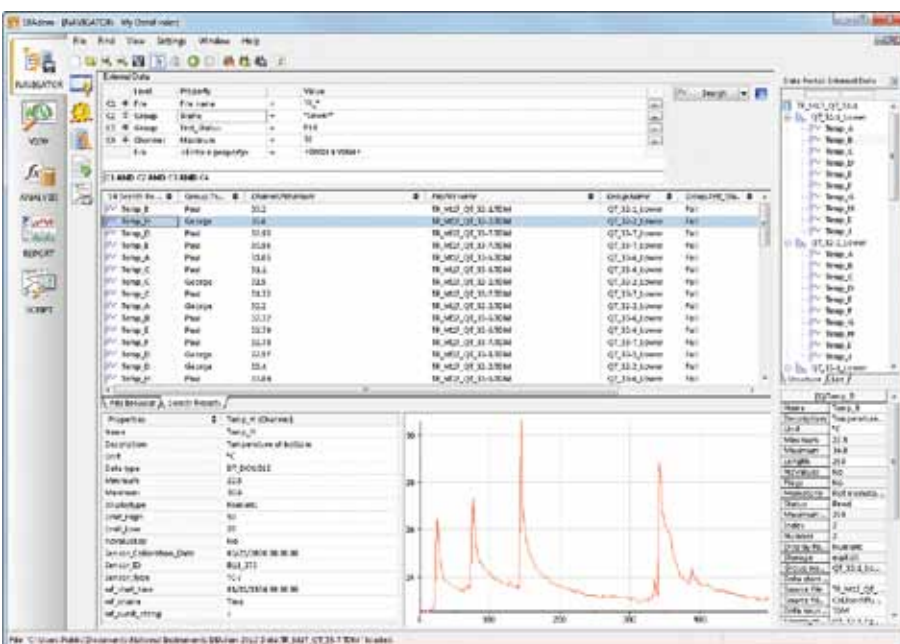
Å andra sidan börjar verktyg som Natio-

nal Instruments DIAdem bli mer och mer avancerade när det gäller att hitta de data man behöver. Med DIAdem kommer en teknik som kallas NI My DataFinder som gör att man snabbt hittar och får fram precis de data man är intresserad av, även om de ligger i olika filer eller mappar.

My DataFinder skapar automatiskt ett index över den beskrivande informationen som finns i datafilerna och gör den sökbar i DIAdem. Med tekniken i DIAdem och My DataFinder hittar man snabbt alla datafiler som skrevs av en viss operatör, alla underkända testresultat och till och med godtyckliga datakanaler från en viss sensortyp i alla lagrade filer. Ju mer beskrivande information man lägger in i datafilerna, desto fler möjligheter får man när man ska söka fram specifika uppgifter som är lagrade någonstans i många filer i olika mappar och/eller filformat.

FÖRETAGET TAKATA ANVÄNDE My DataFinder vid utvecklandet av ett säkerhetssystem för fordonspassagerare. Takata genomförde en serie med tio tester som genererade hundratals filer som sparades i många olika mappar. Ingenjörerna vid Takata fann att de med My DataFinder kunde fatta viktiga beslut snabbare. Enligt ingenjören Jeff Blackburn "kan vi lagra våra genomförda analyser och lättare få fram gamla data med My DataFinder. Vi kan söka igenom våra gamla testdata efter samtliga frontalkrockstester och en hastighet av 40 km/h med en femte percentil kvinnliga testdockor i förarsätet och där skullskadekriteriet var lägre än 700. Med den här möjligheten att titta bakåt i tiden kan vi snabbt utesluta konstruktionsidéer och konfigurationer som inte håller måttet när vi utvecklar nya produkter, och vi sparar både tid och pengar. Med det gamla arbetssättet behövde en ingenjör ungefär fyra timmar för att samla ihop data, verifiera noggrannheten, göra analysen och få fram en rapport. Med det nya systemet, som är baserat på DIAdem, kan han göra samma arbete på mindre än fem minuter. Eftersom vi genomför tusentals tester varje år sparar vi årligen tusentals mantimmar."

NATIONAL INSTRUMENTS genomförde nyligen en kundundersökning där ingenjörer och forskare fick säga vad som var det allra svåraste när de arbetade med sina nuvarande programverktyg. Det helt dominerande svaret var att data – allt från lagring, formatering, sökning, rapportframtagning och analys – var den största huvudvärken i deras nuvarande projekt. När hårdvarutestning och lagringsutrymme blir billigare och billigare ökar kravet på att samla in och analysera stora datamängder. Om man har en skalbar och välordnad lösning blir arbetet med data enklare. ■



I den här databasfrågan har My DataFinder hittat datakanaler i samtliga datafiler som samlats in, har filnamnet TR\_\*, testresultatet underkänt och ett maximum över 30° C.