Mc ích, vn cn gii quyt

Chuyn i s là s dng công ngh s và da trên d liu s. Trong quá trình trin khai ng dng CNTT, Chính ph in t và chính quyn in t ã hình thành mt s c s d liu. Tuy nhiên, nhiu c s d liu hin có tn ti ri rác, c bit là không y , cha m bo cht lng và tính duy nht s dng hiu qu, không c chia s s dng chung cho toàn thành ph. Kho d liu c xây dng gii quyt vn ó, ng thi thc hin chc nng thu nhn ngay d liu s ca các ng dng xây mi, phát sinh mi bo m tp hp, s dng d liu s c quy c trong thi gian n.

T nm 2020, thành ph ã tin hành nghiên cu, hình thành nn tng Kho d liu dùng chung (giai on 1) tng quát, linh hot, có kh nng thu nhn, hp nht d liu n t nhiu ngun, h tr làm sch, chun hóa d liu phân tích, x lý d liu ln, ra quyt nh da trên d liu; cng nh là u mi duy nht chia s d liu cho các c quan khác khai thác, s dng, cung cp d liu m cho ngi dân và doanh nghip.

Kt qu t c n nay nh sau:

- ã tích hp v Kho d liu dùng chung 9 ngun d liu: CSDL công dân; CSDL nhân, h khu; CSDL doanh nghip; CSDL bo him xã hi; CSDL cp phép xây dng; CSDL cp phép lái xe; CSDL du lch; CSDL t ai; CSDL môi trng. Chi tit: 112 bng d liu; 1.149 trng d liu; 7.669.915 bn ghi.

- ã xây dng 58 chun, 252 quy tc d liu (ví d: CMND/CCCD ch có 9 s hoc 12 s; mã quc gia theo TCVN 7217-1: 2007; mã n v hành chính theo quy chun 124/2004/Q-TTg;...).

- Thông qua b lc d liu; ã phát hin khong 10% bn ghi d liu không áp ng quy tc d liu. S Thông tin và Truyn thông ã phi hp vi các c quan ch qun ngun d liu làm sch d liu và a vào Kho d liu chun hóa.

Trên c s Kho d liu c làm sch, chun hóa, các c quan, a phng ã tích cc khai thác, s dng cung cp dch v s, to ra giá tr mi.

Thành ph à Nng ã trin khai thí im s dng d liu in t (CSDL t ai, CSDL nhân h khu, CSDL doanh nghip) thay th thành phn h s th tc hành chính phi np ti hu ht các s, ngành có liên quan và UBND cp qun, phng.

Thành ph à Nng cng ã phân tích, khai thác, s dng d liu phc v công tác phòng chng dch COVID-19 ch ng, hiu qu nh khai thác CSDL bo him xã hi, CSDL doanh nghip cp Giy i ng QRCode; phân tích d liu khai báo y t bit, cnh báo sm ngi có nguy c (ho, st, ngi v t vùng dch, tip xúc vi F0)...

T Kho d liu, ã h tr phân tích d liu kinh t - xã hi phc v công tác ch o, iu hành, ra quyt nh ca lãnh o thành ph, c bit là trong phòng, chng dch COVID-19 va qua. T ó góp phn nâng cao s tin tng ca lãnh o các cp a phng v tính hiu qu ca d liu s, to thói quen xây dng và chia s d liu s gia các c quan, n v.

T Kho d liu, thành ph à Nng ã m d liu ca c quan nhà nc; n nay ã cung cp gn 600 d liu m cho ngi dân, doanh nghip khai thác thông qua API, SMS, Zalo, web.

Hình 1: Giao din H thng Kho d liu dùng chung thành ph à Nng

Kho d liu dùng chung (ti a ch https://khodulieu.danang.gov.vn) là mt trong nhng hp phn ca Nn tng ô th thông minh (Smart City Platform) theo hng dn ca B Thông tin và Truyn thông ti Công vn s 4176/BTTTT-THH ngày 22/11/2019. Hin nay, Kho d liu dùng chung (giai on 1) áp dng i vi d liu có cu trúc, ang tip tc m rng áp dng vi d liu phi/bán cu trúc.

Cách làm, hiu qu mang li

1. Thc trng

Vi s xut hin ca hàng triu thit b thông minh nh in thoi di ng, camera, cm bin..., d liu sinh ra ngày

càng phc tp, khi lng d liu ln, tc tng trng nhanh, yêu cu x lý theo thi gian thc, a dng v chng loi gm có cu trúc, bán cu trúc, phi cu trúc. S phc tp này khin cho các phn mm qun lý d liu và báo cáo nghip v thông thng không còn phù hp vi d liu ln.

Ti thành ph à Nng, d liu còn cát c, ri rc, cha c tích hp, qun lý thng nht h tr vic ra quyt nh. N lc trin khai trc tích hp d liu và các dch v d liu cng ch gii quyt c vn cung cp d liu n l. S thiu sót và thiu tính chun xác, s sai lch, không khp nhau v lc d liu (data schema), không toàn din v d liu khin cho vic s dng d liu hp nht t nhiu ngun gp nhiu khó khn. ây cng là vn mà th gii ang tp trung gii quyt.

Gii pháp phù hp cho vn nêu trên là xây dng mt nn tng d liu (Data Platform) tng quát, linh hot có kh nng thu nhn, hp nht, chun hóa d liu n t nhiu ngun, h tr lu tr, phân tích, x lý d liu ln.

Do ó, m bo trin khai hiu qu thành ph thông minh, thành ph à Nng ã phát trin nn tng Kho d liu dùng chung nhm h tr lãnh o thành ph ra quyt nh chính xác, ti u, cung cp tri thc, hiu bit v quá trình hot ng ca ô th, d báo s phát trin kinh t xã hi thành ph.

Vic xây dng nn tng Kho d liu dùng chung thành ph à Nng là mt trong nhng nhim v quan trng, cp thit và phù hp vi xu hng chuyn i s, xây dng thành ph thông minh hin nay ca Trung ng và thành ph.

2. Mc tiêu

Hình thành Nn tng Kho d liu dùng chung có kh nng x lý a dng các loi d liu (có cu trúc, bán cu trúc, phi cu trúc), hp nht d liu t nhiu ngun d liu khác nhau, làm sch, chun hóa phc v khai phá, phân tích d liu ln, h tr ra quyt nh da trên d liu, s dng d liu to ra giá tr mi, phát trin các sn phm/dch v mi.

3. Mô t h thng

a) Quy trình x lý d liu

Nn tng Kho d liu dùng chung thành ph à Nng gm các thành phn lõi sau ây (c phân tích, thit k, và hoàn toàn làm ch v công ngh):

- ETL (Extract-Transform-Load): T phát trin công c ETL m bo kh nng x lý a dng các loi d liu.

- Data Lineage - lu vt ngun gc d liu: Lp h s v ngun d liu c thu nhn và ánh x t d liu ngun n d liu trong h thng.

- Data Governance - qun lý các lung thu nhn d liu cng nh d liu ã c thu nhn vào h thng. D liu tn ti trong h thng c phân làm 2 loi: (1) d liu không hp chun,, c lu kho cha DQS (Data Quality Storage) và (2) d liu ã chun hóa (Normalized Data Storage).

- Data Firewall - áp dng các quy tc kim tra tính hp chun ca d liu. Các quy tc kim tra tính hp chun ca d liu c xây dng da trên tp hp các tiêu chun cho các loi trng d liu.

- Data Flow Designer - nh ngha lung thu nhn d liu, cho phép ngi dùng la chn ngun d liu, ích n ca d liu khi thu nhn vào h thng, xây dng các biu thc chuyn i ánh x d liu u vào, thit lp các tiêu chí ánh giá kh nng trùng lp d liu.

- Data Flow Executor - thc thi lung thu nhn d liu: hp nht d liu, kim tra tính trùng lp d liu, kim tra tính bo toàn tham chiu d liu trc khi lu vào h thng. Quá trình thc thi c ghi nht ký có th truy vt khi cn.

Hình 2: Quy trình x lý d liu

b) Mô hình tng th Nn tng Kho d liu dùng chung thành ph à Nng

Nn tng Kho d liu dùng chung xây dng theo kin trúc micro-service, kin trúc hng dch v (SOA); có tính mô-un hóa (modularity) m bo kh nng phân tách thành các phân h thun li trong qun lý kin trúc h thng; da trên các tiêu chun m (open standards) h tr n gin hóa vic tích hp vi các nn tng và h thng khác; có kh nng chu li (fault-tolerance), có kh nng m rng mà không phi thay i kin trúc h thng; m bo hiu nng cao vi c ch hot ng song song, a lung trong môi trng phân tán, truy xut d liu nhanh khi d liu ngày càng ln theo thi gian, m bo s lng giao dch nhiu ngi cùng lúc. Tùy theo mc , quy mô d liu ca thành ph, h thng có th c nâng cp h tng phn cng mt cách linh hot.

Hình 3: Mô hình tng th Kho d liu dùng chung thành ph à Nng

c) Chc nng h thng

- Nn tng Kho d liu dùng chung là h thng cho phép thu thp, trích xut, chuyn i, làm sch, chun hóa, lu tr, qun lý; t ó phc v trin khai Chính quyn in t, chính quyn s, thành ph thông minh:

+ Chia s d liu, dùng chung cho các ng dng ca các c quan nhà nc;

+ Cung cp d liu cho Cng d liu m thành ph (https://congdulieu.vn hoc https://opendata.danang.gov.vn) công khai thông tin, chia s d liu cho ngi dân, cho doanh nghip h tr khi nghip, sáng to;

+ Cung cp d liu phân tích, x lý, khai phá và h tr ra quyt nh;…

- Mt s im ni bt ca Kho d liu thành ph à Nng nh sau:

+ Da trên công c ETL (Extract-Transform-Load) t phát trin.

+ Làm vic vi a dng loi d liu (có cu trúc, bán cu trúc, phi cu trúc) n t nhiu ngun khác nhau (excel, CSV, d liu camera, mng xã hi, logfile, thit b IoT, d liu GIS...).

+ Cho phép nh ngha tùy bin lung thu nhn, kim tra, hp nht và chun hóa d liu mt cách linh hot.

+ Cung cp C ch “tng la d liu" (Data Firewall) giúp lc d liu không hp chun, cho phép ngi qun tr d liu thc hin chun hóa. Tng la d liu óng vai trò là b lc tùy bin c cho tng lung thu nhn d liu c th. Cho phép ngi dùng tinh chnh tham s phát hin s trùng lp d liu trong quá trình thu nhn.

+ Cung cp các c ch x lý d liu song song (parallel processing), x lý theo lô (batch processing), x lý theo thi gian thc (streaming processing).

+ Cung cp các c ch qun tr d liu (Data Governance), kim soát d liu và siêu d liu nhm m bo luôn cha các d liu sch và chun hóa: Qun lý siêu d liu (Metadata); Qun lý danh mc d liu (Data Catalog); Qun lý cht lng d liu (Data Quality), h tr vic nh ngha các quy tc cht lng d liu (rules and standards), cung cp c ch làm sch (cleaning) và hiu chun (transformation and standardization) da trên các quy tc cht lng d liu; Qun lý vòng i ca d liu; Qun lý ngun d liu (Data Lineage); Bo mt d liu (Data Security).

+ Tích hp vi Trc tích hp liên thông thành ph LGSP, H thng thông tin Chính quyn in t thành ph à Nng; sn sàng kt ni vi Nn tng tích hp chia s quc gia (NDXP),... Cung cp dch v chia s d liu (bao gm d liu thô, d liu chun hóa, d liu phân tích) cho Trung tâm giám sát iu hành thông minh và các h thng ng dng khác.

d) Công ngh s dng

H thng s dng các công ngh D liu Ln (Big Data) xây dng và qun lý H d liu (Data Lake) cho phép thu nhn, x lý hp nht d liu. S dng các công c phân tích x lý d liu ln mã ngun m nh Kafka, Hadoop, Hive,

Spark..., hiu chnh, khp ni to thành h thng nn tng d liu toàn din.

) Kh nng kt ni, h tr ti các nn tng và h sinh thái liên quan

- Thành ph à Nng ã xây dng Trc tích hp liên thông LGSP và kt ni, tích hp vi Trc liên thông quc gia NGSP. Theo Khung kin trúc chính quyn in t và thành ph thông minh ca thành ph à Nng, Kho d liu ã kt ni vi Trc liên thông LGSP kt ni n các CSDL nn (công dân, doanh nghip, nhân h khu, t ai) và các CSDL chuyên ngành ca các s, ban, ngành, qun huyn nhm thu thp d liu. ng thi, sn sàng kt ni, tích hp vi Nn tng tích hp chia s d liu quc gia NDXP và Trc liên thông quc gia NGSP thu nhn d liu t các CSDL quc gia và CSDL B, ngành Trung ng.

- Thành ph à Nng ã xây dng Cng d liu m (opendata.danang.gov.vn hoc congdulieu.vn) nhm cung cp các tp d liu m qua nhiu kênh tra cu, khai thác nh web, API, SMS, Zalo phc v các t chc, cá nhân khai thác, s dng. Kho d liu ã kt ni vi Cng d liu m, qua ó cho phép m d liu t các c quan thành ph (t các CSDL nn và chuyên ngành) a lên Cng d liu m công khai ra bên ngoài.

- Thành ph à Nng ang hình thành Trung tâm Giám sát iu hành thông minh (IOC) và nn tng phân tích d liu thông minh (Artificial Intelligence Platform). Kho d liu là mt trong nhng hp phn ca H thng Nn tng thành ph thông minh (Smart City Platform), cung cp toàn b d liu cho Trung tâm IOC trình din, hin th; ng thi cung cp d liu cho nn tng phân tích d liu phc v khai phá, phân tích thông minh.

e) Tính bo mt và quyn riêng t

- H thng ã xây dng sn sàng phân h Bo mt d liu (Data Security) nhm qun lý, m bo tính toàn vn d liu và m bo rng d liu không th truy cp c bi các bên trái phép và làm hng d liu. Bo mt d liu bao gm mã hóa d liu, mã thông báo và qun lý khóa bo v d liu trên tt c các ng dng và nn tng. H thng s dng chc nng xác thc tp trung và ng nhp mt ln ca H thng thông tin Chính quyn in t thành ph.

- H thng m bo kh nng phân quyn truy cp n tng bng, trng d liu. Mi hành ng truy cp, c/ghi, thêm mi, sa i, hy b,.. u c ghi li nht ký và cung cp giao din trc quan ngi qun tr kim soát.

- Áp dng các tiêu chun bo mt cng nh các k thut nhm m bo tính bo mt cho các h thng, nhm tránh các li bo mt có th phát sinh nh SQL Injection, XSS, CSR.

- Kim th mã ngun, ng dng trong quá trình phát trin và trc khi a vào s dng.

- Xây dng phng án phòng chng DDOS ngay t u nh: Áp dng proxy trong vào kin trúc ng dng; s dng Firewall lc các truy cp; áp dng bin pháp chng DDOS tng lp trình.

- S dng giao thc https và xác thc ngi dùng khi s dng h thng.

- Thit k h thng theo hng phân tán, tách bit h thng qun tr d liu và lu tr d liu.

- Thit lp các firewall và hn ch các cng truy xut vào C s d liu.

- Xây dng c ch tracking thay i và thông tin cho ngi qun tr.

- Giám sát tình trng vn hành h thng (Server, RAM, Storage).

- nh k sao lu d liu d phòng.

- H thng c cài t ti Trung tâm d liu thành ph à Nng, do ó k tha các gii pháp bo m an toàn thông tin chuyên dng nh Tng la External/Internal, thit b phát hin và phòng chng xâm nhp trái phép (IPS/IDS), thit b cân bng ti, h thng giám sát an toàn thông tin tp trung (SIEM),...

g) S n nh và tin cy

- Cht lng d liu thu nhn vào h thng c giám sát, qun lý và m bo thông qua: (1) h thng chun ràng buc d liu c nh ngha rõ ràng, (2) các tiêu chí ánh giá trùng lp d liu c ngi dùng thit lp, tinh chnh c th, (3) quá trình chun hóa d liu bng tay hoc t ng u c lu vt. Chính vì vy, cht lng d liu luôn c m bo.

- Lung thu nhn d liu c lp h s qun lý. Quá trình thc thi ca lung thu nhn d liu c theo dõi, ghi nht ký, qua ó bo m c tính rõ ràng và tin cy v ngun gc d liu.

- D liu ca h thng, ngoài vic có th sao lu-phc hi còn c ánh phiên bn.

- Các thành phn lõi c phân tích, thit k, cài t và làm ch hoàn toàn nên h thng có kh nng bo hành, bo trì cao và d dàng m rng, phát trin.

h) Kh nng tùy bin, m rng

- H thng xây dng theo hng lp trình tùy bin các chc nng hin ti. Mt s chc nng trong h thng có th c m rng bi ngi dùng chuyên nghip thông qua vic cung cp các on mã lp trình thêm bao gm:

+ B sung các hàm chuyn i d liu (transformation function) trong công on nh ngha lung thu nhn d liu.

+ B sung ràng buc cho các kiu d liu, nh ngha các quy tc áp dng ràng buc cho kiu d liu phù hp vi yêu cu ca c quan, n v.

- H thng có th c tùy bin phù hp vi quy mô d liu cng nh tính cht, hot ng nghip v ca các c quan, n v:

+ V h tng phn cng, s lng máy trong cm máy ch d liu, máy ch ng dng có th c tùy bin.

+ Chc nng chính ca h thng - thc thi lung thu nhn d liu - có th c lp lch linh hot áp ng vi nhu cu thc t ca các c quan.

- H thng có kh nng m rng mà không phi thay i kin trúc h thng; m bo hiu nng cao vi c ch hot ng song song, a lung trong môi trng phân tán, truy xut d liu nhanh khi d liu ngày càng ln theo thi gian, m bo s lng giao dch nhiu ngi cùng lúc.

4. T chc trin khai

a) V chính sách:

- Trc tiên, S Thông tin và Truyn thông là c quan ch trì tham mu UBND thành ph ban hành Quy ch chia s d liu s trên a bàn thành ph ti Quyt nh s 846/Q-UBND ngày 21/02/2019 (trc khi Chính ph ban hành Ngh nh s 47/2020/N-CP) xác nh rõ các danh mc d liu dùng chung; trách nhim ca c quan ch qun d liu, qun lý, cp nht, chia s, s dng, khai thác d liu.

- S Thông tin và Truyn thông ch trì ban hành các vn bn gi các c quan, a phng trên a bàn thành ph chia s các c s d liu chuyên ngành thuc phm vi qun lý v Kho d liu dùng chung thành ph; t chc tp hun, hng dn s dng.

- Tham mu khung hng dn trin khai xây dng, tip nhn các ng dng mi phi có hp phn to, cp nht d liu s; c bit là kt ni, chia s d liu v Kho d liu dùng chung ca Thành ph.

b) V truyn thông:

- Qung bá qua các kênh truyn thông in t ca thành ph à Nng: Tng ài 1022 à Nng: qua kênh thoi, tin nhn, fanpage mng xã hi, chatbot, Zalo ca Tng ài 1022; Cng Thông tin in t thành ph, các trang

thông tin in t ca các s, ngành, a phng trên a bàn thành ph.

- Thng xuyên gii thiu Kho d liu dùng chung ti các hi ngh, hi tho quc t và trong nc v xây dng thành ph thông minh, chuyn i s …

5. Tn ti, hn ch

a) Các CSDL quc gia v dân c, t ai cha c chia s cho các a phng hình thành d liu nn/d liu tham chiu.

b) Các c quan Trung ng nh Công an, Thu, Thng kê,... cha thc hin chia s d liu cho các a phng do vng các quy nh pháp lut chuyên ngành trái vi quy nh ti Ngh nh s 47/2020/N-CP ngày 09/4/2020 ca Chính ph; do vy các c quan trên ch sa i Lut chuyên ngành mi trin khai chia s d liu.c) Các B, ngành Trung ng cha ban hành các chun d liu phc v kt ni, chia s. Các CSDL chuyên ngành ca thành ph trin khai, hình thành trong quá trình hot ng cha c thu thp y , không cht lng, chun hóa; do ó tính kh dng ca d liu thp, cn có thi gian rà soát, hoàn thin.

Bài hc rút ra

a) Thng nht v nguyên tc, quan im ch o (trong các vn bn liên quan ca Thành y, UBND thành ph): D liu c to ra trong quá trình hot ng ca c quan nhà nc là tài sn thuc s hu ca nhà nc, không phi ca riêng c quan/cá nhân nào; do ó d liu phi c chia s, s dng chung gia các c quan nhà nc.

b) Xây dng Kin trúc d liu làm nh hng, kim ch nam cho xây dng, to lp, kt ni, chia s, khai thác d liu ca thành ph. Kin trúc d liu ca thành ph m bo phù hp vi Kin trúc d liu quc gia và Kin trúc d liu ca các b, ngành, a phng.

c) Phi ban hành trc Quy ch chia s d liu s trên a bàn thành ph.

d) Xây dng Kho d liu dùng chung phân k nhiu giai on kp thi s dng, tip tc m rng, hoàn thin; không cu toàn, trin khai quy mô quá ln.

) Các ng dng mi phi có hp phn to, cp nht d liu s; c bit là kt ni, chia s d liu v Kho d liu dùng chung ca Thành ph; c bit có quy ch vn hành, cp nht d liu cho ng dng ó bo m d liu luôn c y , cht lng và duy nht; bo m cho chia s và s dng d liu.

e) Ly ngi dân, doanh nghip làm trung tâm, ch th phc v; thng xuyên rà soát, ánh giá cht lng d liu và tình hình s dng, tái cu trúc quy trình s dng hiu qu hoc k tha d liu s, xem vic cung cp dch v s cho ngi dân, doanh nghip là mt kt qu, u ra, là “giá tr” ca d liu s.

Theo "https://dx.danang.gov.vn/"