



دليل بيانات دبي

# تحميل البيانات

الإصدار 1.2 (نوفمبر 2020)



دبي الذكية  
SMART DUBAI

# تحميل البيانات

<p>تصف هذه الوثيقة المتطلبات وتوصيات الممارسات الجيدة المتعلقة بالطريقة التي سيتم استخدامها عند نشر بيانات دبي المفتوحة أو المشتركة. تتمثل أهمية هذه الوثيقة في كونها تضمن اختيار الطريقة المناسبة لتحميل البيانات على منصة دبي بالس.</p>	<p>الغرض:</p>
<p>أخصائيو البيانات، الذين يخضعون لإشراف إداري البيانات مباشرة.</p>	<p>على من تقع المسؤولية الرئيسية عن تنفيذ هذا المعيار في كل جهة حكومية:</p>
<p>سوف تطبق مؤسسة بيانات دبي هذا المعيار بعد أن تقوم الجهات الحكومية <a href="#">بفهرسة</a> الدفعة الحالية من مجموعات البيانات ذات الأولويات المحددة، وقبل <a href="#">تحميل</a> البيانات على منصة دبي بالس. وينبغي أن يتم ذلك بالتزامن مع الإجراءات المتعلقة <a href="#">بالتصنيف</a>، و <a href="#">أشكال البيانات والبيانات الوصفية وجودة البيانات</a> وقبل <a href="#">تحميل</a> البيانات على منصة دبي بالس.</p>	<p>متى تستخدم هذه الوحدة من دليل بيانات دبي</p>
<p>مؤسسة بيانات دبي</p>	<p>مالك الوثيقة</p>
<p>يكون للمصطلحات الواردة في هذا المعيار نفس التعاريف الواردة في <a href="#">قاموس مصطلحات دليل بيانات دبي</a>.</p>	<p>التعاريف:</p>
<p>نشرت هذه الوثيقة تحت شروط <a href="#">رخصة المشاع الإبداعي الدولية 4.0</a> من أجل تسهيل إعادة استخدامها من قبل الحكومات الأخرى والجهات الخاصة. وباختصار، هذا يعني أنه سيكون لكم مطلق الحرية في مشاركة المواد والتعديل عليها، بما في ذلك لأغراض تجارية، شريطة أن يتم تقديم الضمان المناسب لمؤسسة بيانات دبي بصفتها المالك لها مع عدم الأخذ في اعتباركم أن مؤسسة بيانات دبي ستجيز استخدامكم لهذه المواد.</p>	<p>الترخيص:</p>
<p>الإصدار 1.2، الصادر بتاريخ 11 نوفمبر 2020.</p>	<p>الإصدار</p>

# نظرة عامة

تؤثر الطريقة التي يتم من خلالها تحميل البيانات على منصة دبي بالس بصورة كبيرة على جعل الإجراءات الحالية أكثر كفاءة وسلاسة وتوليد مخرجات قوية جديدة في ذات الوقت.

فيما يلي المبادئ الأساسية المطبقة أثناء تحديد طريقة تحميل البيانات:

- 1- بذل أقل قدر ممكن من الجهد من قبل الجهات الحكومية
- 2- الحصول على البيانات في شكلها الأصلي
- 3- عدم وجود متطلبات محددة لمعالجة البيانات
- 4- تطبيق إجراءات الأمن والخصوصية عند نقل البيانات

## المبادئ

### بذل أقل قدر ممكن من الجهد من قبل الجهات الحكومية

يتمثل المبدأ التوجيهي لمشاركة البيانات وتحميلها على منصة دبي بالس في بذل أقل قدر ممكن من العمل من قبل الجهات المزودة للبيانات. ومن أجل إنفاذ هذا المبدأ، اعتمد مكتب دبي الذكية نهج المعايير المفتوحة للبيانات الضخمة استناداً إلى إطار أباتشي هادوب (Apache Hadoop) للحلول. وقد ساهمت هذه المعايير في تخفيض تكلفة توزيع البيانات وتخزينها وتحليلها بمقدار الثلثين، مقارنة بالطرق التقليدية على مدى السنوات الخمس الماضية.

ورغم أنه لن تكون هناك حاجة إلى أدوات أو بنية تحتية لمشاركة البيانات على المنصة، إلا أن منصة الذكاء دبي توفر أدوات عدة يمكن أن تستخدمها الجهات المانحة للبيانات في تحميل البيانات بأنفسهم ودون مساعدة خارجية. ومن المهم أن تتجنب الجهات المساهمة بالبيانات الحاجة إلى إعداد نسخ محددة من قواعد البيانات أو استخدام أجهزة مخصصة لأغراض تبادل البيانات، إلا إذا كانت هذه الموارد متاحة أو تغطي احتياجات القدرة الاستيعابية/المتطلبات الأمنية لأمين حفظ البيانات، ذلك أن إعداد هذه النسخ من قواعد البيانات أو استخدام هذه الأجهزة المخصصة سيكون أكثر استهلاكاً للوقت أو أكثر كلفة بوسائل أخرى. وحينما ينفذ مثل هذا العمل لغرض مشاركة البيانات، فينبغي أن يدعم المشاركة المستقبلية لكافة مجموعات البيانات التي يديرها المساهم حالياً والتي تصنف على أنها تتطلب المشاركة.

### الحصول على البيانات في شكلها الأصلي

لقد بات الحصول على البيانات الآن في شكلها [الأصلي](#) ممكناً وأمناً بل وأقل تكلفة أيضاً باستخدام حلول النظام الإيكولوجي للبيانات الضخمة، كما بات ممكناً تحويل أو تحليل البيانات داخل مستودع بيانات مركزي بعد تحميلها على المنصة. ومن شأن هذا أن يمكن دبي من جعل النظم والإجراءات الحالية أكثر كفاءة، فضلاً عن تحقيق نتائج جديدة لم تكن ممكنة قبل أن تكون البيانات مرئية في المستودع المركزي. وأحد هذه النتائج هو وجود تحليلات تنبؤية دقيقة، تتطلب أكبر عدد ممكن من النقاط المرجعية للبيانات لتعزيز الدقة.

## عدم وجود متطلبات محددة لمعالجة البيانات

لن تكون هناك حاجة إلى أشكال ذات تصميم خاص للبيانات أو محتوى له خصائص معينة ذلك أن منصة دبي بالس تعتمد على معايير/أشكال البيانات المفتوحة والمستخدمة على نطاق واسع (مثل ملفات البيانات بشكل (csv)، أو (XLS) أو (XML) أو (JSON)...) ولن تكون هناك حاجة سوى إلى الحد الأدنى من البيانات الوصفية باعتبارها أفضل الجهود للمساعدة في فهم البيانات.

## الأمن والخصوصية

سيتم الاتفاق مع كل مزود للبيانات على الإجراءات الأمنية للولوج إلى منطقة تجميع البيانات بمنصة دبي بالس وكذلك إجراءات إرسال البيانات، وذلك بالاستناد إلى ترتيباتهم الحالية المتعلقة بالربط الإلكتروني وعوامل أخرى مثل حساسية البيانات أو وتيرة الإرسال. والهدف من ذلك هو تعزيز الترتيبات الحالية قدر الإمكان. ويمكن عند الاقتضاء، تنفيذ عمليات إخفاء وحجب البيانات على منصة دبي بالس.

# عملية تحميل البيانات

## 1. وضع خطة لتحميل البيانات

بمجرد أن تقوم الجهة بإعداد المسودة الأولية لقائمة جرد البيانات، تبدأ كل من مؤسسة بيانات دبي ومزود البيانات عملية تصنيف البيانات وتحديد البيانات ذات الأولوية كل حسب اختصاصه (مؤسسة بيانات دبي على مستوى المدينة ومزود البيانات على مستوى الجهة).

تقوم مؤسسة بيانات دبي ومزود البيانات بإتمام تصنيف مجموعات البيانات هذه على أنها إما مفتوحة أو مشتركة، باستخدام نموذج تصنيف بيانات دبي أو مجموعة الأدوات. الهدف من هذه الخطوة هو تصنيف البيانات بشكل صحيح قبل إدخالها في منصة دبي بالس لضمان الاستخدام الآمن والمعالجة الصحيحة للبيانات في جميع المراحل.

بعدها تمر مجموعات البيانات عبر معايير تحديد البيانات ذات الأولوية، والتي تستند إلى مجموعتين: الفائدة المرجوة والجاهزية. هذه الخطوة مهمة لتحديد البيانات ذات الأهمية القصوى، بما في ذلك مجموعة البيانات الأكثر طلباً من أصحاب المصلحة المختلفين مما يمنح مؤسسة بيانات دبي حافزاً أكبر لإعطاء الأولوية لمجموعة البيانات هذه على مجموعات البيانات الأخرى للنشر.

يجب تصنيف قائمة خصائص البيانات في مجموعة البيانات بشكل صحيح من قبل مزود البيانات. تتم بعدها مراجعة تصنيف مجموعة البيانات والموافقة على التصنيف من قبل مؤسسة بيانات دبي. تأخذ مؤسسة بيانات دبي أعلى تصنيف لقائمة خصائص البيانات وتعتبرها التصنيف الرسمي لمجموعة البيانات.

قد تقرر مؤسسة بيانات دبي أيضاً إتاحة جزء من مجموعة البيانات كبيانات مفتوحة إذا كانت مجموعة البيانات سرية ولكنها تتضمن خصائص مفتوحة تم تحديدها على أنها ذات أهمية / فائدة عالية لمقدم الطلب، فيمكن تنفيذ منطق الانقسام عن طريق إنشاء نسختين مختلفتين من مجموعة البيانات، واحدة تتضمن الخصائص المفتوحة فقط وستبقى الأخرى شاملة ومع التصنيف الأصلي. في هذه الحالات، ستقوم مؤسسة بيانات دبي بإبلاغ مزود البيانات بذلك.

لإكمال خطة التحميل، ترسل مؤسسة بيانات دبي نموذج اتفاقية البيانات لمزود البيانات لتعبئته. يوضح هذا النموذج جميع المعلومات الضرورية التي تحتاجها مؤسسة بيانات دبي لتحميل البيانات، بما في ذلك البيانات الوصفية لمجموعة البيانات وطرق التحميل والتفاصيل الفنية وأنواع الملفات والتنسيق، إلخ.

يقوم بعدها فريق مؤسسة بيانات دبي وفريق منصة دبي الذكية بمراجعة اتفاقية البيانات للتأكد من أن البيانات مفيدة من حيث الجودة والمحتوى والتوافق وتقييم ما إذا كانت جاهزة للتحميل في منصة دبي بالس.

## 2. تحديد موعد لتحميل البيانات

تبدأ مؤسسة بيانات دبي بعد ذلك في تحديد موعد لتحميل البيانات على دفعات الشهرية بناءً على عملية تحديد البيانات ذات الأولوية السابقة، والطلب المفترض، وتوافر مجموعة البيانات، والمشاريع المهمة الجارية. تغطي كل دفعة تحميل مجموعات بيانات معينة من جهات مختلفة. وفي بعض الأحيان، بسبب التعقيد والعدد الهائل من مجموعات البيانات، قد يتم تحويل تحميل بعض مجموعات البيانات إلى دفعة ثانية في المستقبل.

## 3. تحميل مجموعات البيانات

بمجرد الموافقة على اتفاقية البيانات من قبل مؤسسة بيانات دبي، يبدأ فريق منصة دبي الذكية عملية الربط مع مزود البيانات لضمان نقل البيانات بشكل صحيح إلى منصة دبي بالس. بعد ذلك، يتم إجراء اختبار قبول المستخدم لضمان عرض البيانات دون أي أخطاء على منصة دبي بالس، والتحقق مما إذا تم إنشاء الترابط. بشكل صحيح، والتأكد من تنفيذ جميع الخطوات السابقة بشكل صحيح قبل نشر البيانات.

للقيام بذلك، قد يطلب فريق البيانات عينات من البيانات لمقارنتها بالاتفاقية الأصلية للتأكد من عدم وجود اختلافات.

بالإضافة إلى ذلك، ستقوم مؤسسة بيانات دبي بالحفاظ على هيكل الخدمات والاصناف، والذي يحدد كيفية عرض كل مجموعة بيانات على منصة دبي بالس، والتخطيط العام للبيانات.

بمجرد موافقة مؤسسة بيانات دبي على اختبار قبول المستخدم الناجح وتوفير التصنيف والخدمات، يبدأ فريق منصة دبي الذكية بعد ذلك مرحلة النشر حيث يقومون بنشر المجموعة بأكملها.

## 4. نشر مجموعات البيانات

بمجرد احتمال جميع الخطوات السابقة لتحميل البيانات، تُعلم مؤسسة بيانات دبي مزود البيانات بأنه تم نشر البيانات وأن الجهة تحتاج إلى البدء في دفع / تعبئة البيانات وفقاً للطريقة المتفق عليها. سيقوم فريق منصة دبي الذكية بعد ذلك بنشر البيانات على منصة دبي بالس وإتاحتها كمجموعات بيانات قابلة للاستخدام بالكامل، مع متطلبات الوصول وفقاً لنموذج متطلبات الوصول الخاص بمؤسسة بيانات دبي.

## 5. الحفاظ على مجموعات البيانات

بعد نشر البيانات، يتعين على كل من فريق منصة دبي الذكية ومزود البيانات التأكد من تحديث البيانات في الوقت المحدد وفقاً لاتفاقية البيانات لضمان وصول المستخدمين إلى أحدث البيانات وأكثرها دقة. إذا كانت هناك حاجة إلى تغيير هيكل في مجموعة البيانات، فيجب على مزود البيانات إبلاغ مؤسسة بيانات دبي عن التغيير مسبقاً لضمان إخطار جميع الأطراف بشكل مناسب. تشمل هذه التغييرات:

- أي تغييرات في هيكل البيانات
- أي تغييرات على واجهة التطبيق
- أي تغييرات على بنود اتفاقية البيانات
- أي تغييرات على عمليات تحديث البيانات

بعد إخطار مؤسسة بيانات دبي بهذا التغيير، يجب على مزود البيانات تحديد النطاق وخطة التغيير وإرسال طلب خطي إلى مؤسسة بيانات دبي مع نطاق التغيير وخطة التنفيذ.

# التقنيات

فيما يلي التقنيات المحتملة لمشاركة البيانات مع منصة دبي بالس.

## 1. (دفع أو سحب) الملفات

من شأن توافر الملفات من الجهات الحكومية (السحب) أو توفير منطقة تجميع للبيانات من قبل منصة دبي بالس (الدفع) أن يتيح مشاركة البيانات بطرق متعددة، في أي فترة زمنية تناسب مجموعة البيانات:

- يمكن إتاحة الوصول إلى أي نقطة تخزين فيها الجهة المانحة للبيانات الملفات المراد مشاركتها. يمكن تشغيل عملية السحب عبر بروتوكول نقل الملفات (SFTP)
- يمكن دفع الملفات إلى منطقة تجميع الملفات بمنصة دبي بالس عبر بروتوكول نقل الملفات (SFTP)

## 2. البيانات شبه المنظمة أو غير المنظمة

يمكن تحميل البيانات ذات الأشكال شبه المنظمة (مثل ملفات السجل أو ملفات وسائل التواصل الاجتماعي) أو ذات الأشكال غير المنظمة (مثل المستندات المخصصة لأغراض معينة)، وذلك بالصورة التي تكون عليها وبشكل الملف الموجود حالياً.

## 3. واجهات برمجة التطبيقات (APIs)

إذا كانت واجهات برمجة التطبيقات موجودة بالفعل أو في حال كان الفريق الفني للجهات المساهمة بالبيانات يتفهم جيداً كيفية إنشاء واجهات برمجة التطبيقات أو كان إنشاء هذه الواجهات يستغرق بضعة أيام (بما في ذلك المدة اللازمة لإنهاء إجراءات الموافقة واتخاذ القرار)، فيمكن في هذه الحالة استخدام واجهات برمجة التطبيقات لمشاركة البيانات مع منصة دبي بالس. ورغم أن واجهات برمجة التطبيقات قد تكون مفيدة في عملية إخفاء البيانات أو حجبها، إلا أن إنشاء واجهات برمجة التطبيقات يعني أن قراراً قد اتخذ بشأن فائدة أو استخدام مجموعة فرعية معينة من البيانات لغرض محدد، وبالنظر إلى أن الهدف من منصة دبي بالس هو الوصول إلى أوسع مجموعة ممكنة من البيانات، فإنه يمكن تمكين حالات الاستخدام المستقبلية هذه (التي لا تتوافر لها تصورات معينة حتى الآن) دون بذل مزيد من الجهود من جانب مزود البيانات.

## الأساليب

فيما يلي الأساليب المحتملة التي قد تتبناها الجهات التابعة لحكومة دبي لتمكين مشاركة البيانات مع الجهات الحكومية الأخرى أو إتاحتها كبيانات مفتوحة عبر منصة دبي بالس. ورغم أن هذه الأساليب لم تعتمد بعد في ترتيب الأولويات، لكن السياسة المنشورة ستشتمل على شرح مفصل لها بمجرد الموافقة عليها.

- الوصول إلى مجموعات البيانات الموجودة بالفعل بحالتها الراهنة من خلال آليات الربط القائمة والتي يمكن أن تشمل:
  - برامج استنساخ الملفات القائمة
  - واجهات برمجة التطبيقات القائمة
  - لوحات منصة دبي بالس
- مشاركة البيانات من خلال نموذج إذن البيانات المبني على منصة دبي بالس