

Data Warehouse Design & Develop



Masoud Mirzakhani
Senior DW/ ETL/ BI Architect

Microsoft SQL Server 2019 Design & Develop

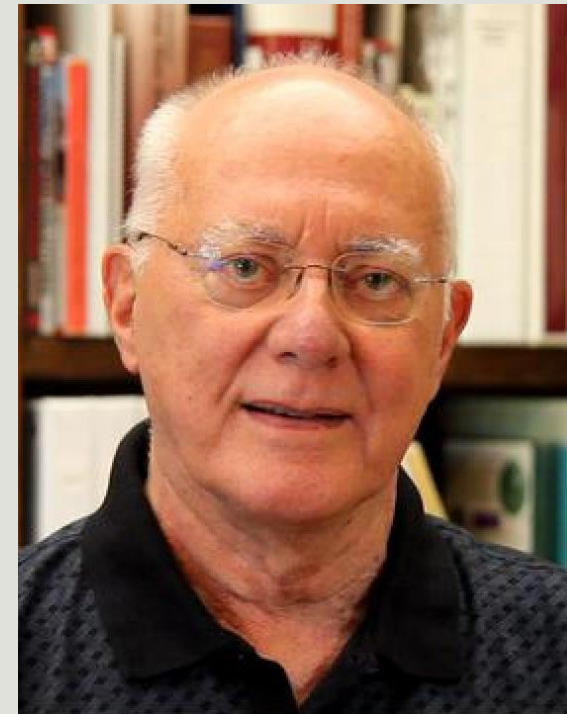
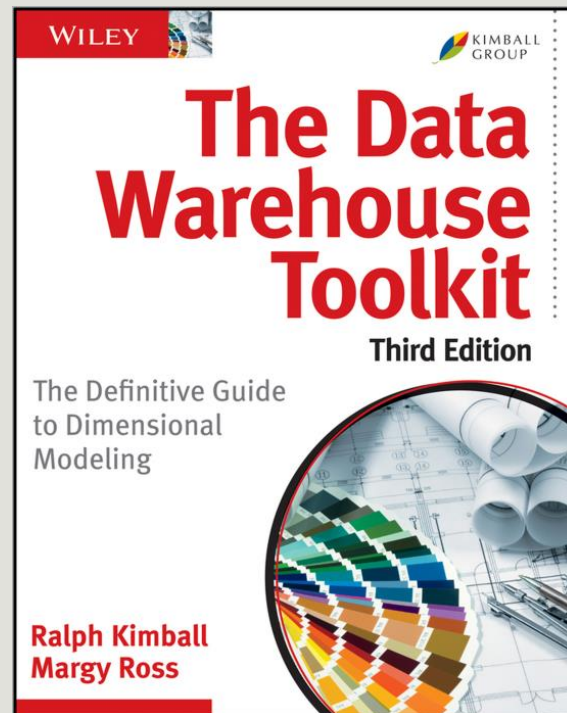


Masoud Mirzakhani
Senior DW/ ETL/ BI Architect

- **Master of Science in Information Technology**
- **Bachelor of Science in Information Technology**

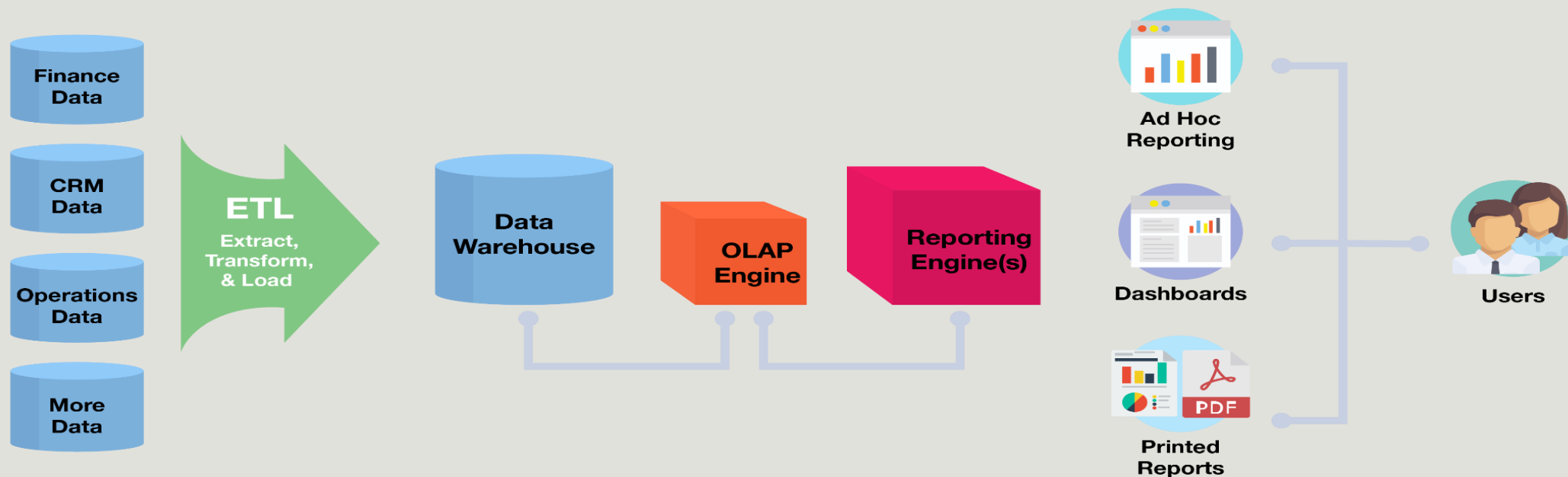
- md.mirzakhani@gmail.com
- [@MasoudMirzakhani](#)
- [linkedin.com/in/masoudmirzakhani](https://www.linkedin.com/in/masoudmirzakhani)





معماری سیستم BI

مجموعه ای از راهکارهای نرم افزاری و سخت افزاری است، که داده های خام و انبوه موجود در سازمانها و شرکت ها را به دانش تصمیم ساز تبدیل می کند.



نگرانی های مدیران

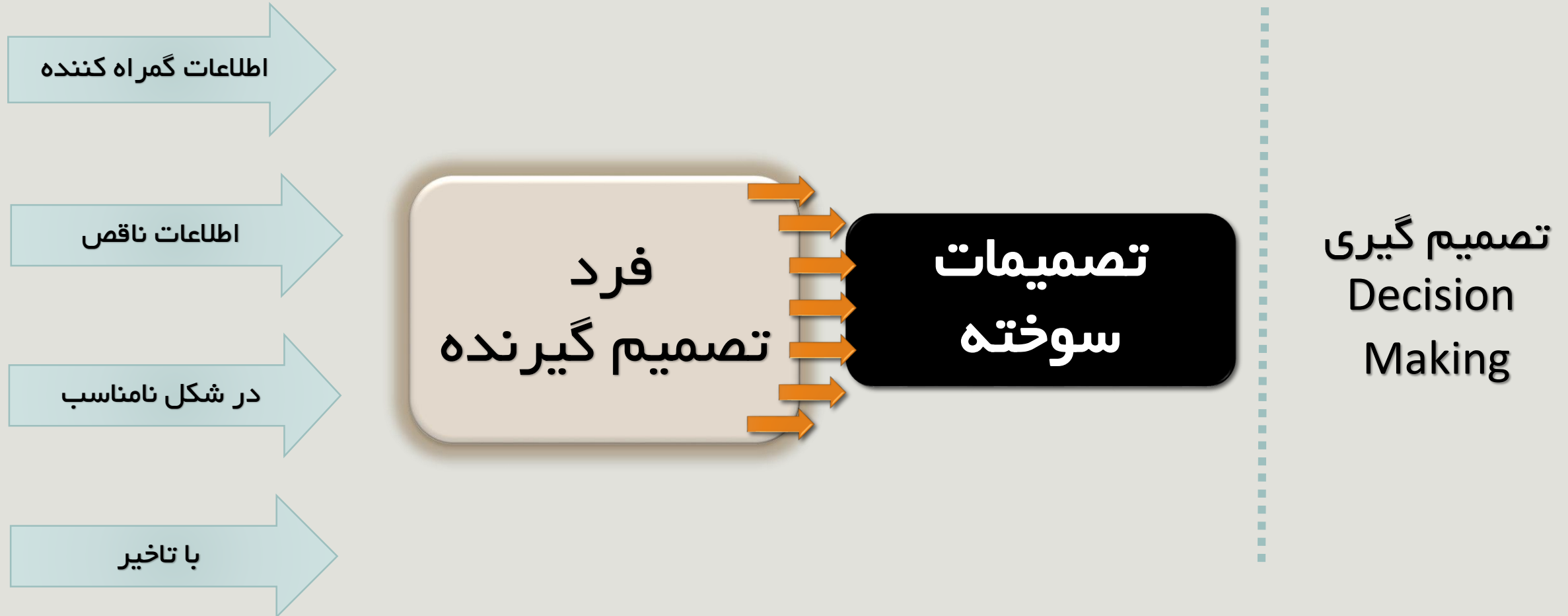
-
- ❖ میلیون ها رکورد اطلاعات داریم، ولی به آن ها دسترسی نداریم.
 - ❖ اطلاعات به راحتی به دست ما نمی رسد.
 - ❖ من فقط نیاز دارم اطلاعات مهم را داشته باشم.
 - ❖ در جلسات به جای تصمیم گیری، بر روی صحت اطلاعات بحث می کنیم.

نیازمندی مدیران



-
- ❖ نمایش اطلاعات به صورت یکپارچه
 - ❖ به روز رسانی اطلاعات در زمان قابل قبول
 - ❖ اطلاعات بایستی قابل اعتماد و صحیح باشند.
 - ❖ بستر ارایه اطلاعات بایستی امن باشد.

تصمیم گیری



تصمیم گیری



معرفی مدل چند بعدی

- ❖ این مدل به عنوان روشی جهت نمایش اطلاعات تحلیلی مورد پسند افراد فنی می باشد. زیرا:
 - ❖ اطلاعات را در قالب قابل فهم به مدیران ارائه می کند.
 - ❖ کویری ها از کارایی بالایی برخوردار هستند.
- ❖ این مدل روشی قدیمی جهت ساده سازی پایگاه های داده اطلاعاتی می باشد.
- ❖ هر چند که مدل چند بعدی در محیط پایگاه داده رابطه ای پیاده سازی می شود، متفاوت از مدل های نرمال سازی (مانند 3NF) می باشد.
- ❖ اختلاف اساسی بین این دو مدل، در درجه نرمال سازی جداول می باشد.

دنیای متفاوت سیستم های عملیاتی و تحلیلی

یکی از مهمترین دارایی های هر سازمان اطلاعات آن است.

این دارایی تقریباً همیشه برای دو هدف استفاده می شود:

- ثبت و نگهداری داده های عملیاتی
- تصمیم گیری تحلیلی
- سیستم های عملیاتی جایی هستند که داده ها را در آن قرار می دهید.
- سیستم انبار داده جایی است که شما داده ها را بیرون می آورید.

سیستم عملیاتی

کاربران یک سیستم عملیاتی چرخ های سازمان را می چرخانند.

- گرفتن سفارش
 - ثبت نام مشتریان جدید
 - نظارت بر وضعیت فعالیت های عملیاتی
 - ورود اطلاعات شکایات مشتریان به سیستم
-
- سیستم های عملیاتی برای پردازش سریع تراکنش ها بهینه شده اند.
 - این سیستم ها تقریباً همیشه با یک رکورد از تراکنش ها در یک زمان سر و کار دارند.
 - آنها بطور پیش فرض کارهای عملیاتی مشابه را بارها و بارها انجام می دهند.
 - سیستم های عملیاتی به طور معمول تاریخ را حفظ نمی کنند
 - آن ها اطلاعات را به روز می کنند تا وضعیت فعلی را منعکس کنند.

چالش های دیتابیس نرمال

- برای نگهداری اطلاعات ریز و جزئی طراحی شده است
- کارایی مناسبی برای ذخیره سازی داده های تراکنشی دارد.
- از یکپارچگی داده ها به خوبی پشتیبانی می کند.
- در نتیجه این طراحی تعداد زیادی جدول ایجاد می شود.
- راه حل مناسبی برای مدیریت داده های روزانه است.

Normalized Structure Challenges



-
- معمولا کارایی مناسبی در خواندن اطلاعات ندارد.
 - نیاز به JOIN کردن تعداد زیادی جدول دارد.
 - معماری مناسبی برای تحلیل داده ها ارائه نمی دهد.
 - داده ها در ریزترین حالت خود ذخیره می شوند.
 - با زیاد شدن حجم داده، کارایی Query ها کاهش پیدا می کند.

سیستم انبار داده

- کاربران یک سیستم انبار داده، به چرخش چرخ های سازمان نظارت می کنند.
- سفارشات جدید را شمرده و با سفارشات هفته گذشته مقایسه می کنند.
- مشتریان از چه چیزی شکایت کرده اند.
- نگران این هستند که آیا فرآیندهای عملیاتی به درستی کار می کنند.
- برای پشتیبانی از سؤالات مداوم در حال تغییر، به داده های دقیق نیاز دارند.
- کاربران انبار داده تقریباً هیچ وقت با یک تراکنش سروکار ندارند.
- این سیستم ها برای نمایش داده های با کارایی بالا بهینه می شوند.
- سؤالات کاربران غالباً نیاز به جستجو در بین صدها یا میلیون ها تراکنش دارد.
- کاربران یک سیستم انبار داده معمولاً خواستار حفظ اطلاعات تاریخی برای ارزیابی دقیق عملکرد سازمان در طول زمان هستند.

مقایسه سیستم عملیاتی با سیستم انبار داده

سیستم انبار داده	سیستم عملیاتی	
مدیران	کاربران عمومی	کاربران
پشتیبانی تصمیم	عملیات روزانه	کارکرد
خاص منظوره / تحلیل	عملیات تکرار شونده	کاربرد
پرس وجوهای پیچیده	تراکنش های ساده و کوتاه	واحد کاری Working Unit
میلیون ها رکورد	ده ها رکورد	تعداد رکوردهای مورد دسترسی
صدها کاربر	هزاران کاربر	تعداد کاربران
پشتیبانی از تصمیم گیری مدیریتی	گزارش گیری روزانه از کسب و کار	هدف
Denormalized	Normalized	معماری
Relational Database (OLTP)	Relational Database (OLTP)	تکنولوژی

اهداف طراحی سیستم انبار داده

❖ دسترسی آسان به میلیون ها رکورد داده ی جمع آوری شده

❖ امکان نفوذ به عمق داده ها

❖ امکان گزارش گیری چند بعدی از داده ها

❖ تمرکز بر روی اطلاعات مهمتر

❖ امکان تحلیل اطلاعات

❖ پشتیبانی از تصمیم گیری مدیران

باورهای اشتباه در مورد انبار داده

- مدل چند بعدی تنها اطلاعات را به صورت خلاصه شده نگهداری می کند.
- مدل چند بعدی بر اساس دپارتمان های سازمان طراحی شده و در کل سازمان فراگیر نمی باشد.
- مدل چند بعدی مقیاس پذیر نیست.
- مدل چند بعدی تنها کاربرد پیش بینی دارد.
- مدل چند بعدی، قابلیت یکپارچگی ندارد.

گام های طراحی مدل چند بعدی

- تعیین فرایندها
- تعیین سطح ریزدانگی
- تعیین ابعاد
- تعیین اندازه ها (Fact ها)

گام های طراحی مدل چند بعدی

فرایندهای کسب و کار

- فرایندها، فعالیت های عملیاتی یک سازمان هستند، مانند:
- ثبت سفارش، پردازش یک ادعای خسارت، ثبت نام دانشجو یا تایید حساب های ماهانه.
- از روی اتفاق های داخل یک فرایند؛ متریک های اندازه گیری عملکرد ساخته می شود.
- متریک ها تبدیل به Fact یا Measure می شوند.
- فرایندها تبدیل به جداول Fact می شوند.
- هر فرایند متناظر با یک سطر از Enterprise DW Bus Matrix می باشد.

گام های طراحی مدل چند بعدی

سطح ریزدانگی Granularity

- سطح ریز دانگی مشخص می کند که هر سطر از جدول Fact، نشان دهنده چه چیزی باشد.
- سطح ریزدانگی بایستی قبل از شناسایی جداول Dimension و Fact تعیین شود.
- بیشترین جزئیات در پایین ترین سطح از ریزدانگی مشاهده می شود.

Atomic Grain •

- توصیه می شود که از این سطح از ریزدانگی استفاده شود.

گام های طراحی مدل چند بعدی

بعد یا Dimension

- فرایند را با اطلاعاتی نظیر "چه زمانی؟ چه کسی؟ چه چیزی؟ کجا؟ چرا؟ چگونه؟" توصیف می کند.
- از ابعاد برای فیلتر کردن و گروه بندی کردن اطلاعات در نرم افزارهای BI استفاده می شود.
- اطلاعات توصیفی، لیبل ها و عناوین در جداول بعد نگهداری می شود.

گام های طراحی مدل چند بعدی

فکت (Fact) یا اندازه ها (Measure) یا متریک

- اندازه گیری اتفاق های یک فرایند کسب و کار از طریق Fact صورت می گیرد.
- این اندازه ها؛ معمولا به صورت عددی می باشند.
- هر سطر از جدول Fact متناظر با یک اتفاق در فرایند می باشد.

طراحی مدل چند بعدی

Bus Matrix

ابزار طراحی انبار داده

سطرها مشخص کننده فرایندهای کسب و کار هستند.

سطرها مشخص کننده جداول Fact هستند.

ستونها مشخص کننده جداول Dimension هستند.

BUSINESS PROCESSES	COMMON DIMENSIONS						
	Date	Product	Warehouse	Store	Promotion	Customer	Employee
Issue Purchase Orders	X	X	X				
Receive Warehouse Deliveries	X	X	X				X
Warehouse Inventory	X	X	X				
Receive Store Deliveries	X	X	X	X			X
Store Inventory	X	X		X			
Retail Sales	X	X		X	X	X	X
Retail Sales Forecast	X	X		X			
Retail Promotion Tracking	X	X		X	X		
Customer Returns	X	X		X	X	X	X
Returns to Vendor	X	X		X			X
Frequent Shopper Sign-Ups	X			X		X	X